[8] ——, (2002) Optimality of universal Bayesian prediction for general loss and alphabet. Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA), Manno (Lugano), Switzerland. [Online]. Available: ftp://ftp.idsia.ch/pub/techrep/IDSIA-02-02.ps.gz. Tech. Rep. IDSIA-02-02

[9] R. J. Solomonoff, "Complexity-based induction systems: Comparisons and convergence theorems," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 422–432, July 1978.

[10] P. D. Grünwald, "The minimum description length principle and reasoning under uncertainty," Ph.D. dissertation, Univ. Amsterdam, Amsterdam, The Netherlands, 1998.

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, ser. Wiley Series in Telecommunications. New York: Wiley, 1991.

[12] A. A. Borovkov and A. Moullagaliev, *Mathematical Statistics*. New York: Gordon & Breach, 1998.

[13] A. R. Barron, "Limits of information, Markov chains, and projection," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Sorrento, Italy, June 2000, p. 25.

[14] M. Hutter, "New error bounds for Solomonoff prediction," *J. Comput. Syst. Sci.*, vol. 62, no. 4, pp. 653–667, 2001. [Online]. Available: ftp://ftp.idsia.ch/pub/techrep/IDSIA-11-00.ps.gz.

[15] ——, "An open problem regarding the convergence of universal a priori probability," in *Proc. 16th Conf. Computational Learning Theory (COLT 2003)*, 2003, Available: [Online] http://www.hutter1.de/ai/mlconv.htm.

[16] N. Cesa-Bianchi *et al.*, "How to use expert advice," *J. ACM*, vol. 44, no. 3, pp. 427–485, 1997.

[17] A. R. Barron, B. S. Clarke, and D. Haussler, "Information bounds for the risk of Bayesian predictions and the redundancy of universal codes," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, 1993, pp. 54–54.

[18] M. Hutter, "Self-optimizing and Pareto-optimal policies in general environments based on Bayes-mixtures," in *Proc. 15th Annu. Conf. Computational Learning Theory (COLT 2002) (Lecture Notes in Artificial Intelligence)*. Sydney, Australia: Springer, 2002, pp. 364–379. [Online]. Available: ftp://ftp.idsia.ch/pub/techrep/IDSIA-04-02.ps.gz.

Multiple Description Quantization by Deterministic Annealing

Prashant Koulgi, *Student Member, IEEE*, Shankar L. Regunathan, *Member, IEEE*, and Kenneth Rose, *Fellow, IEEE*

***Abstract*—The design of vector quantizers for diversity-based communication over two or more channels of possibly differing capacities and failure probabilities, is considered. The crucial dependence of current design techniques on initialization, especially of index assignment, is well recognized. Instead, we propose to pursue a deterministic annealing approach which is independent of initialization, does not assume any prior knowledge of the source density, and avoids many poor local minima of the cost surface. The approach consists of iterative optimization of a random encoder at gradually decreasing levels of randomness as measured by the Shannon entropy. At the limit of zero entropy, a hard multiple description (MD) quantizer is obtained. This process is directly analogous to annealing processes in statistical physics. Via an alternative derivation, we show that it may also be interpreted as approximating the minimum rate sums among points on the convex hull of the MD achievable rate-distortion region of El Gamal and Cover, subject to constraints on the sizes of the reproduction alphabets. To illustrate the potential of our approach, we present simulation results that show substantial performance gains over existing design techniques.**

***Index Terms*—Deterministic annealing, multiple descriptions, quantization, source coding, vector quantization.**

I. Introduction

We consider the design of multiple description vector quantizers (MDVQs) for use in a diversity-based communication system (hereafter referred to as a diversity system). A diversity system provides several channels for communication between the transmitter and the receiver. The MDVQ encoder encodes a fixed-length block of source samples into individual indexes for transmission over each of the channels, subject to separate rate constraints. Each of these channels may fail independently, and the decoder reconstruction is based on information received from the subset of channels that are in working order. Applications of multiple description source codes are currently being pursued in speech and video coding over packet-switched networks and fading multipath channels [18], [19], [21]. MDVQs designed for asymmetric channels are strongly motivated by packet-switched networks with priority classes. Finally, scalable quantizer design may be viewed as a special case of MDVQ design.

For simplicity, we will restrict the discussion to diversity systems with two channels. The two channels may have differing capacities and failure probabilities. (We call the special case when the two channels allow the same rates and have identical failure probabilities the case of *balanced descriptions*). When both the channels function reliably, the distortion achieved with the joint description is the "central" distortion;

Manuscript received August 1, 2001; revised March 5, 2002. This work was supported in part by the National Science Foundation under Grants EIA-9986057 and EIA-0080134, the University of California MICRO Program, Dolby Laboratories, Inc., Lucent Technologies, Inc., Mindspeed Technologies, Inc., and Qualcomm, Inc. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, Sorrento, Italy, June 2000.

P. Koulgi and K. Rose are with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106-9560 USA (e-mail: rose@ece.ucsb.edu; prashant@ece.ucsb.edu).

S. L. Regunathan is with Microsoft Corporation, Redmond, WA 98052 USA (e-mail: shanre@microsoft.com).

Communicated by R. Zamir, Associate Editor for Source Coding.

Digital Object Identifier 10.1109/TIT.2003.814493

when one of the channels fails, the distortion achieved with the received single-side description is the corresponding "side" distortion.

The study of practical multiple description systems was inaugurated by Vaishampayan in [18]. Considering the design of balanced multiple description scalar quantizers (MDSQs), he derived an iterative design algorithm (closely related to Lloyd's algorithm for quantizer design [13]) that minimizes a weighted sum of the expected central and side distortions. (Here, the weights may be determined by the channel failure probabilities, while the codebook sizes may be fixed by the channel capacities.) This algorithm is guaranteed to find a locally optimal solution. In [7], an extension of this algorithm was proposed for the design of unstructured MDVQs for arbitrary weights and rate constraints. This extension continues to guarantee the local optimality promised by the original algorithm.

The occurrence of poor local minima on the cost surface, and the consequent sensitivity of Lloyd's locally optimal quantizer design algorithm to initialization, is widely recognized. While one approach to overcoming this problem has concentrated on clever heuristic initializations, an alternate approach has considered the development of algorithms which do not require initialization. See [8] for a summary of both these approaches. In the case of multiple description (MD) quantizers, the problem of poor local minima is exacerbated by the presence of a more complicated cost function. Vaishampayan already recognized this problem in [18]; an important contribution of that paper was the proposal of asymptotically good initializations for balanced MDSQ design for the Gaussian source. But such initializations have not been forthcoming for the design of unstructured MDVQs for arbitrary source distributions with arbitrary distortion and rate constraints.

Pursuing the alternate approach, in this correspondence, we propose an unstructured-MDVQ design algorithm which does not require initialization. This algorithm is an extension of the deterministic annealing (DA) approach to single-description quantizer design, introduced in [16], which avoids many poor local minima via an annealing procedure. Our algorithm may be especially useful in training-based design of unstructured nonbalanced MD quantizers, where the search for good initializations is hindered by the lack of knowledge of the geometry of the source distribution, as well as the continuum of possible rate and distortion constraints.

Certain chemical systems can be driven to their low energy states by annealing, which is a gradual reduction of temperature, spending a long time in the vicinity of phase transition points. Analogously, we randomize the encoding rule of the multiple description system and seek to minimize the expected distortion cost subject to a specified level of randomness measured by the Shannon entropy. This problem can be formulated as the minimization of a Lagrangian functional that is analogous to the Helmholtz free energy of chemical systems. The degree of randomness is parameterized by the "temperature" of the configuration. We start at a high degree of randomness, where we, in fact, maximize the entropy. Here, the globally minimum configuration requires that all code vectors be coincident at the centroid of the source distribution; no initialization of codebook or index assignment is necessary. We then track the minimum at successively lower levels of entropy, by recalculating the optimum locations of the reproduction points and the encoding probabilities at each stage. Coincident codevectors split at certain critical temperatures (which can be calculated), thus increasing the effective codebook sizes. This phenomenon is analogous to the phase transitions of statistical physics. At the limit of zero randomness, the algorithm directly minimizes the expected distortion cost, and a deterministic encoder is obtained.

The DA approach may also be interpreted within the rate-distortion framework. Thus, we consider calculation of the minimum rate sums at different sections (parameterized by the corresponding distortion triples) of the convex hull of an MD achievable rate-distortion region (attributed to El Gamal and Cover in [22]). We cast this problem as the unconstrained minimization of a Lagrangian functional over certain probability densities and output maps, and show that this functional is identical to the free energy used in the DA algorithm for MDVQ design. Further, we prove that in several cases (including the important one of squared error distortion and compactly supported source alphabet) the supports of the output maps minimizing the above functional are discrete and finite. We then show that the DA algorithm for MDVQ design simulates the calculation of the above functional, subject to constraints on the maximum sizes allowed for the ranges of the optimum output maps. We also interpret the phase transitions of the DA algorithm within the context of the calculation of the minimum rate sums. Note that a similar DA-based approach was previously proposed in [17] for the calculation of the rate-distortion function of a (continuous) source.

While our focus here is on MD systems based on unstructured quantizers, several other approaches to the construction of MD systems have been studied in the literature. The design algorithm of Diggavi *et al.* in [5] for lattice MDVQs with arbitrary rate and distortion constraints, and its asymptotic analysis therein, represents an important contribution to the problem of structured-MDVQ design. MD systems based on overcomplete expansions were considered in [2], [4], and [9], while methods based on optimizing transforms and predictors were presented in [10], [14], and [20]. For results on the characterization of the MD achievable region (in the Shannon-theoretic sense) see [6], [22], and [15].

We formulate the problem of MDVQ design and establish notation in the next section. In Section III, we describe the DA approach to this problem. Necessary conditions for optimality are then used to derive an iterative MDVQ design algorithm. We conclude the section by describing the "mass-constrained" form of our algorithm, which is our preferred implementation. In Section IV, we rederive the algorithm by considering calculation of the convex hull of an MD achievable region. In Section V, we present simulation results and comparisons with existing approaches. Phase transition analysis, including derivation of critical temperatures at which the size of the reproduction set increases, is considered in the Appendix.

II. The MDVQ Problem and Design Considerations

We are interested in encoding a real-valued source represented by a stationary and ergodic random process X distributed as $p(\cdot)$. Let a single-letter distortion measure d be given. Consider a diversity system with two channels capable of transmission of information at rates R_1 and R_2 bits per source sample (bpss), respectively. Each channel may or may not be in working order, and its condition is not known at the encoder. The encoder sends a different description over each channel, and the decoder forms the best estimate of the source output from the descriptions received via the channels that were functioning reliably. An MDVQ maps an n-dimensional source vector x to the n-dimensional reproduction vectors $\hat{x}^0$, $\hat{x}^1$, and $\hat{x}^2$, which take values in the codebooks $\hat{\mathcal{X}}^0 = \{\hat{x}^0_{jk},\ (j,\ k) \in J \times K\}$, $\hat{\mathcal{X}}^1 = \{\hat{x}^1_j,\ j \in J\}$, and $\hat{\mathcal{X}}^2 = \{\hat{x}^2_k,\ k \in K\}$, respectively. Here, $J = \{1,\ 2,\ \ldots,\ 2^{nR_1}\}$ and $K = \{1,\ 2,\ \ldots,\ 2^{nR_2}\}$.

The MDVQ encoder is the mapping $e\colon \mathcal{R}^n \mapsto J \times K$. Given source vector x, it selects an index pair $e(x) = (j,\ k)$. Each index is sent over its respective channel. The MDVQ decoder $l = (l_0,\ l_1,\ l_2)$ is, in fact, a bank of three switched decoders each performing a lookup operation: the central decoder $l_0\colon J \times K \mapsto \mathcal{R}^n$ takes in a double index $(j,\ k)$ and produces the code vector $l_0(j,\ k) = \hat{x}^0_{jk}$. The side decoders $l_1\colon J \mapsto \mathcal{R}^n$ and $l_2\colon K \mapsto \mathcal{R}^n$ take in the single indexes i and j to produce the code vectors $l_1(i) = \hat{x}^1_i$ and $l_2(j) = \hat{x}^2_j$, respectively.

For given values of R_1, R_2, D_1, and D_2 we wish to find an MDVQ which minimizes the *expected distortion cost*

$$D(e, l) = E\{d(X, \hat{X}^0)\} + \lambda_1 E\{d(X, \hat{X}^1)\} + \lambda_2 E\{d(X, \hat{X}^2)\} \tag{1}$$

over e and l. The specific choice of λ_1 and λ_2 in a practical system is determined by the weights we wish to assign to the side distortions relative to the central distortion, which could depend on the channel failure probabilities.

In the subsequent discussions, we will often use the concise notation

$$\mathcal{D}(x, y_0, y_1, y_2) = d(x, y_0) + \lambda_1 d(x, y_1) + \lambda_2 d(x, y_2) \tag{2}$$

to denote the weighted distortion cost of x relative to the triple composed of the central codevector y_0 and the side codevectors y_1 and y_2. This will enable us to tidy up several long equations.

Note that the expected distortion cost depends on the code vector locations. Further, it also depends on the indexes assigned to codevectors, since they determine which pair of side vectors are mapped to each central code vector. Locally optimal multiple description quantizer design algorithms [18], [7] must be initialized with codebooks and index assignment. The choice of initial index assignment constrains the algorithm to a part of the cost surface, so that "good" initial index assignment is crucial to the performance of the algorithm. But good heuristics for choosing the initial index assignment are elusive, since they depend on the particular rate and distortion constraints, and some knowledge of the source distribution. Strategies for initial index assignment are discussed in detail in [18], where good heuristics are presented for the special case of balanced descriptions and scalar quantizer. Moreover, it was shown that heuristic index assignment is asymptotically optimal (in the sense of high resolution) for the particular case of memoryless Gaussian source and balanced MDSQs. However, these heuristics do not seem to generalize to the case of unstructured, unbalanced MDVQs. It may here be noted that the corresponding index assignment problem for unbalanced *lattice* MDVQ design was recently resolved in [5].

III. Derivation of the DA Algorithm

A formal derivation of DA can be based on principles of information theory, by considering optimization of a random encoder subject to constraints on its degree of randomness. An alternate, but equivalent, derivation appeals to Jaynes' principle of maximum entropy for statistical inference [11]. We shall use the former approach here.

A. Encoding Probabilities and Reproduction Points

Let us begin by assuming that the three codebooks, $\hat{\mathcal{X}}^0 = \{\hat{x}^0_{jk}\}$, $\hat{\mathcal{X}}^1 = \{\hat{x}^1_j\}$, and $\hat{\mathcal{X}}^2 = \{\hat{x}^2_k\}$ are given. We use a random encoding rule, and assign input source vector x to the index pair (j, k) with probability $q(j, k|x)$. The central decoder and the two side decoders output $\hat{x}^0_{jk}$, $\hat{x}^1_j$, and $\hat{x}^2_k$ when presented with indexes (j, k). We can rewrite the expected distortion cost of (1) for a random encoder as

$$D = \int dx\, p(x) \sum_{j,k} q(j, k|x) \mathcal{D}(x, \hat{x}^0_{j,k}, \hat{x}^1_j, \hat{x}^2_k) \tag{3}$$

where we drop the arguments of D for notational simplicity.

We seek the distribution $q(j, k|x)$ which minimizes D subject to a specified level of randomness, which is measured by the Shannon entropy

$$H(J, K, X) = -\int dx\, p(x) \sum_{j,k} q(j, k|x) \log p(x) q(j, k|x). \tag{4}$$

The corresponding Lagrangian to minimize is

$$F = D - TH. \tag{5}$$

The Lagrangian functional F is analogous to the Helmholtz free energy of a physical system where D is the energy, H is the entropy, and the Lagrangian multiplier T is the temperature. Minimizing F corresponds to seeking isothermal equilibrium of the system.

Now note that $H(J, K, X) = H(J, K|X) + H(X)$, and the source entropy $H(X)$ is independent of the encoding rule. We may therefore drop the constant $H(X)$ from the Lagrangian definition. Minimizing F with respect to the encoding probabilities $q(j, k|x)$ yields

$$q(j, k|x) = \frac{\exp\left[-\left(\frac{1}{T}\right)\mathcal{D}(x, \hat{x}^0_{jk}, \hat{x}^1_j, \hat{x}^2_k)\right]}{\sum_{j,k} \exp\left[-\left(\frac{1}{T}\right)\mathcal{D}(x, \hat{x}^0_{jk}, \hat{x}^1_j, \hat{x}^2_k)\right]}. \tag{6}$$

The corresponding minimum of F is obtained by plugging (6) into (5)

$$\begin{aligned} F^* &= \min_{q(j,k|x)} F \\ &= -T \int dx\, p(x) \log \sum_{j,k} \exp\left[-\frac{\mathcal{D}(x, \hat{x}^0_{jk}, \hat{x}^1_j, \hat{x}^2_k)}{T}\right]. \end{aligned} \tag{7}$$

We now find the optimal sets of reproduction vectors $\hat{\mathcal{X}}^0$, $\hat{\mathcal{X}}^1$, and $\hat{\mathcal{X}}^2$ which minimize F^* for this random encoder. These vectors satisfy the following necessary conditions:

$$\frac{\partial}{\partial \hat{x}^0_{jk}} F^* = \frac{\partial}{\partial \hat{x}^1_j} F^* = \frac{\partial}{\partial \hat{x}^2_k} F^* = 0. \tag{8}$$

Substituting for F^* from (7), we have, for all $\hat{x}^0_{jk} \in \hat{\mathcal{X}}^0$, $\hat{x}^1_j \in \hat{\mathcal{X}}^1$, and $\hat{x}^2_k \in \hat{\mathcal{X}}^2$

$$\begin{aligned} \int dx\, p(x|j, k) \frac{\partial}{\partial \hat{x}^0_{jk}} d(x, \hat{x}^0_{jk}) &= \int dx\, p(x|j) \frac{\partial}{\partial \hat{x}^1_j} d(x, \hat{x}^1_j) \\ &= \int dx\, p(x|k) \frac{\partial}{\partial \hat{x}^2_k} d(x, \hat{x}^2_k) = 0. \end{aligned} \tag{9}$$

Here, $p(x|j, k)$, $p(x|j)$, and $p(x|k)$ denote the posterior probabilities calculated using Bayes' rule.

For the squared error distortion case $d(x, y) = \|x - y\|^2$, the above equations reduce to the centroid rules

$$\begin{aligned} \hat{x}^0_{jk} &= \int dx\, p(x|j, k) x \\ \hat{x}^1_j &= \int dx\, p(x|j) x \\ \hat{x}^2_k &= \int dx\, p(x|k) x. \end{aligned} \tag{10}$$

Note that the encoding probabilities of (6) and the centroid rules of (10) are just random relatives, respectively, of the encoding and decoding rules for optimal MD quantizer design for squared error distortion ([18], (9), and (18)–(20)).

Our algorithm consists of minimizing F^* with respect to the code vectors starting at a high temperature and tracking the minimum while decreasing the temperature.[1]

The central iteration itself is composed of the following two steps:

1) fix the codebooks and use (6) to compute the encoding probabilities;
2) fix the encoding probabilities and optimize the codebooks according to (9).

Clearly, this procedure is monotone nonincreasing in F^*. Note that the algorithm reduces to the known locally optimal MDVQ design algorithm of [18] at the limit of zero temperature.

At very high temperatures, the global minimum configuration consists of all the (central and side) code vectors coincident at the centroid of the source distribution. As the temperature is reduced from the initial high value, the set of coincident code vectors bifurcates into subsets for the first time at some lower temperature. We call this bifurcation the first phase transition, and the corresponding temperature the first critical temperature, in analogy with the phase transitions seen during the annealing of physical systems. As the temperature is lowered further, these subsets again bifurcate, and each such bifurcation is a subsequent phase transition with its corresponding critical temperature. We analyze conditions for phase transitions and derive the critical temperatures in the Appendix. A study of this phenomenon of phase transitions provides insight into the annealing process. Further, since the phase transitions are the critical points of the process, knowledge of the critical temperatures allows us to accelerate the annealing between phase transitions.

B. Mass-Constrained DA (MCDA)

The observation of the phenomenon of phase transitions enables us to recast our algorithm in a more efficient form. Since all the (central and side) code vectors are coincident at high temperatures, they can be viewed as belonging to a single cluster, and this entire cluster can effectively be represented by a single index pair without affecting the expected distortion cost. When the codebooks bifurcate at the critical temperatures, the effective number of clusters increases. Each of these clusters should be represented by a different index pair. We use this observation to derive the "mass-constrained" implementation of our algorithm.

Let us assume an unlimited supply of code vectors and index pairs. The fraction of code vectors of the first side codebook $\hat{\mathcal{X}}^1$ which are coincident at some point can be assigned a common first index for transmission over one of the channels. Let this common index be j; the corresponding fraction of code vectors is labeled $r(j)$ (the cluster "prior" or "mass"), and the point where the code vectors are coincident is $\hat{x}_j^1$. Similarly, we assign a common index k to the mass $s(k)$ of code vectors of the second side codebook coincident at $\hat{x}_k^2$. Consequently, a fraction $r(j)s(k)$ of all index pairs are assigned the index pair (j, k), and the central reproduction corresponding to this index pair is $\hat{x}_{jk}^0$. We can recast the expression for the encoding probability $q(j, k|x)$ in (6) as

$$q(j, k|x) = \frac{r(j)s(k)\exp\left[-\left(\frac{1}{T}\right)\mathcal{D}(x, \hat{x}_{jk}^0, \hat{x}_j^1, \hat{x}_k^2)\right]}{\sum_{j,k} r(j)s(k)\exp\left[-\left(\frac{1}{T}\right)\mathcal{D}(x, \hat{x}_{jk}^0, \hat{x}_j^1, \hat{x}_k^2)\right]}. \quad (11)$$

[1]The performance of the algorithm is insensitive to the choice of the cooling schedule. In our simulations we use the schedule $T \leftarrow \alpha T (\alpha < 1)$.

The corresponding minimum of the free energy over encoding probabilities is (cf. (7))

$$F^* = -T\int dx\, p(x)\, \log \sum_{j,k} r(j)s(k)\exp\left[-\frac{\mathcal{D}(x, \hat{x}_{jk}^0, \hat{x}_j^1, \hat{x}_k^2)}{T}\right]. \quad (12)$$

F^* is to be minimized under the obvious constraints on the masses: $\sum_j r(j) = 1$ and $\sum_k s(k) = 1$. This yields the following update rules:

$$r(j) = \int dx\, p(x) \sum_k q(j, k|x)$$
$$s(k) = \int dx\, px \sum_j q(j, k|x). \quad (13)$$

In other words, the distribution of masses on the indexes is identical to the probability distribution induced on the indexes via the encoding rule.

Minimizing F^* with respect to the central and side reproduction points gives the update formulas

$$\int dx\, p(x|j, k)\, \frac{\partial}{\partial \hat{x}_{jk}^0} d(x, \hat{x}_{jk}^0) = \int dx\, p(x|j)\, \frac{\partial}{\partial \hat{x}_j^1} d(x, \hat{x}_j^1) = \int dx\, p(x|k)\, \frac{\partial}{\partial \hat{x}_k^2} d(x, \hat{x}_k^2) = 0 \quad (14)$$

with the encoding probabilities of (11) used to calculate the posterior probabilities.

MCDA increases the effective number of index pairs only when it is needed, i.e., at a phase transition. Thus, it is computationally more efficient than the earlier "unconstrained" approach. At the limit of low temperatures, the two approaches converge to the same descent process for the expected distortion cost, since their encoding probabilities are identical at the limit (they assign each data point to a single index pair (j, k) with probability 1. In the next section, we will provide an alternate interpretation of MCDA. Note that we use MCDA for all our simulations.

IV. DA and an MD Achievable Region

In this section, we make the following additional assumptions: X is an independent and identically distributed (i.i.d.) random variable distributed over the alphabet $\mathcal{X} \subseteq \mathcal{R}^n$, and the reproduction alphabet is the same as the source alphabet.

In [22], the following sufficient conditions for an MD achievable rate-distortion region are attributed to El Gamal and Cover.[2]

The quintuple $(R_1, R_2, D_0, D_1, D_2)$ is achievable if there exist random variables J and K (defined on the spaces $\mathcal{J}$ and $\mathcal{K}$, respectively) jointly distributed with X such that

$$R_1 \geq I(X; J), \quad R_2 \geq I(X; K), \quad \text{and} \quad R_1 + R_2 \geq I(X; J, K) + I(J; K) \quad (15)$$

and deterministic functions $\hat{x}^0(J, K)$, $\hat{x}^1(J)$, and $\hat{x}^2(K)$ such that

$$E[d(X, \hat{x}^1(J))] \leq D_1, \quad E[d(X, \hat{x}^2(K))] \leq D_2, \quad \text{and} \quad E[d(X, \hat{x}^0(J, K))] \leq D_0. \quad (16)$$

[2]Note that these sufficient conditions are in general not tight, as shown in [22]. But Ozarow showed in [15] that they are tight for the important special case of the Gaussian source and squared-error distortion. Tight characterization of the MD achievable region is not known for any other interesting example.

Denote by $\mathcal{S}$ the convex hull of the region characterized by (15) and (16). Define

$$R_{\rm sum}(D_0,\, D_1,\, D_2) = \inf_{J,\, K:\, \exists \hat{x}^0,\, \hat{x}^1,\, \hat{x}^2 \text{ satisfying (16)}} I(X;\, J,\, K) + I(J;\, K). \tag{17}$$

(Note that we do not impose the side rate constraints $R_1 \geq I(X;\, J)$ and $R_2 \geq I(Y;\, K)$.)

In the following lemma, we will show that $R_{\rm sum}(D_0,\, D_1,\, D_2)$ is the minimum rate sum of the section of $\mathcal{S}$ at the distortion triple $(D_0,\, D_1,\, D_2)$.

Lemma 1: For fixed $(D_0,\, D_1,\, D_2)$, there exists a quintuple $(R_1,\, R_2,\, D_0,\, D_1,\, D_2)$ on $\mathcal{S}$ with $R_1 + R_2$ not less than, but arbitrarily close to, $R_{\rm sum}(D_0,\, D_1,\, D_2)$.

Proof: It is clear from the definition of $R_{\rm sum}(D_0,\, D_1,\, D_2)$ that, for quintuples $(R_1,\, R_2,\, D_0,\, D_1,\, D_2)$ on $\mathcal{S}$

$$R_1 + R_2 \geq R_{\rm sum}(D_0,\, D_1,\, D_2).$$

Now pick $(J,\, K)$ in (17) such that $I(X;\, J,\, K) + I(J;\, K)$ is sufficiently close to $R_{\rm sum}(D_0,\, D_1,\, D_2)$. We then have

$$\begin{aligned} &I(X;\, J,\, K) + I(J;\, K) - I(X;\, J) - I(X;\, K) \\ &\quad = I(X;\, K|J) + I(J;\, K) - I(X;\, K) \\ &\quad = I(X;\, K|J) + H(K|X) - H(K|J) \\ &\quad = H(K|X) - H(K|J,\, X) = I(K;\, J|X) \geq 0 \end{aligned}$$

by the nonnegativity of mutual information. This shows that the rate pair (R_1, R_2) with $R_1 = I(X;J)$ and $R_2 = I(X;J,K) + I(J;K) - I(X;J)$ is on the section of $\mathcal{S}$ at (D_0, D_1, D_2), and we are done. □

The determination of $R_{\rm sum}(D_0,\, D_1,\, D_2)$ in (17) may be recast as an unconstrained minimization via a Lagrangian formulation. The corresponding Lagrangian functional is

$$F = \inf_{q, \hat{x}^0, \hat{x}^1, \hat{x}^2} \int dx\, p(x) \int dj\, dk\, q(j,\, k|x) \left[\log \frac{q(j,\, k|x)}{q(j)q(k)} + \frac{1}{T}\left\{d(x,\, \hat{x}^0_{jk}) + \lambda_1 d(x,\, \hat{x}^1_j) + \lambda_2 d(x,\, \hat{x}^2_k)\right\}\right]. \tag{18}$$

Here, $q(j,\, k|x)p(x)$ denotes the joint distribution of J, K, and X, and $q(j)$ and $q(k)$ are the corresponding marginal distributions. T, λ_1, and λ_2 are (nonnegative) Lagrangian multipliers. We now convert (18) into a three-way minimization, in the spirit of the double minimization of the Blahut–Arimoto algorithm [3], [1].

Lemma 2:

$$F = \inf_{\hat{x}^0, \hat{x}^1, \hat{x}^2, r, s, q} \int dx\, p(x) \int dj\, dk\, q(j,\, k|x) \left[\log \frac{q(j,\, k|x)}{r(j)s(k)} + \frac{1}{T}\, \mathcal{D}(x,\, \hat{x}^0_{jk},\, \hat{x}^1_j,\, \hat{x}^2_k)\right] \tag{19}$$

where r and s are probability measures over $\mathcal{J}$ and $\mathcal{K}$, respectively, and $\mathcal{D}(\cdot,\, \cdot,\, \cdot,\, \cdot)$ was defined in (2).

Proof: Denote the argument of the infimum above by $F(q,\, r,\, s)$. For fixed q, $F(q,\, r,\, s)$ is minimized by

$$r_q(j) = \int dx\, p(x) \int dk\, q(j,\, k|x)$$

and

$$s_q(k) = \int dx\, p(x) \int dj\, q(j,\, k|x) \tag{20}$$

while for fixed r and s, $F(q,\, r,\, s)$ is minimized by

$$q_{rs}(j,\, k|x) = \frac{r(j)s(k) \exp\left[-\frac{1}{T}\, \mathcal{D}(x,\, \hat{x}^0_{jk},\, \hat{x}^1_j,\, \hat{x}^2_k)\right]}{\int dj\, r(j) \int dk\, s(k) \exp\left[-\frac{1}{T}\, \mathcal{D}(x,\, \hat{x}^0_{jk},\, \hat{x}^1_j,\, \hat{x}^2_k)\right]}. \tag{21}$$

These facts follow from the easily verifiable identities

$$\begin{aligned} F(q,\, r,\, s) &= F(q,\, r,\, s_q) + D(r_q \| r) \\ &= F(q,\, r_q,\, s_q) + D(r_q \| r) + D(s_q \| s) \\ F(q,\, r,\, s) &= F(q_{rs},\, r,\, s) + \int dx\, p(x) D(q \| q_{rs}). \end{aligned}$$

The result now follows, since the divergence $D(.\|.)$ is non- negative.□

For fixed r and s, substituting $q = q_{rs}$ in (19), we are left with

$$F = \inf_{r, s, \hat{x}^0, \hat{x}^1, \hat{x}^2} - \int dx\, p(x) \cdot \left\{\log \int dj\, r(j) \int dk\, s(k) \exp\left[-\frac{1}{T}\, \mathcal{D}(x,\, \hat{x}^0_{jk},\, \hat{x}^1_j,\, \hat{x}^2_k)\right]\right\}. \tag{22}$$

Consider now the necessary conditions for optimality of the maps $\hat{x}^0$, $\hat{x}^1$, and $\hat{x}^2$. These are derived via a standard procedure from the calculus of variations: we require

$$\frac{\partial}{\partial \epsilon}\, F(\hat{x}^i + \epsilon \eta_i)|_{\epsilon = 0} = 0, \qquad i = 0,\, 1,\, 2 \tag{23}$$

for all admissible[3] perturbation functions $\eta_0(j,\, k)$, $\eta_1(j)$, and $\eta_2(k)$. After some manipulations, we obtain

$$\int dx\, p(x) q_{rs}(j,\, k|x)\, \frac{\partial}{\partial \hat{x}^0}\, d(x,\, \hat{x}^0(j,\, k)) = 0 \tag{24}$$

$$\int dx\, p(x) q_{rs}(j|x)\, \frac{\partial}{\partial \hat{x}^1}\, d(x,\, \hat{x}^1(j)) = \int dx\, p(x) q_{rs}(k|x)\, \frac{\partial}{\partial \hat{x}^2}\, d(x,\, \hat{x}^2(k)) = 0 \tag{25}$$

$r(j)$—and $s(k)$—almost everywhere. Here, $q_{rs}(j|x)$ and $q_{rs}(k|x)$ are the marginals derived from $q_{rs}(j,\, k|x)$.

We will now show that the ranges of the optimum maps $\hat{x}^0$, $\hat{x}^1$, and $\hat{x}^2$ are often discrete and finite. Note that, in this case, $\mathcal{J}$ and $\mathcal{K}$ may be assumed to be discrete and finite as well (i.e., they may be replaced by finite sets of indexes). If the points y_0, y_1, and y_2 are in the ranges of $\hat{x}^0$, $\hat{x}^1$, and $\hat{x}^2$, respectively, (24) imposes the following requirement on $(y_0,\, y_1,\, y_2)$:

$$\int dx\, p(x) Z(x) \cdot \exp\left\{-\frac{1}{T}\left[d(x,\, y_0) + \lambda_1 d(x,\, y_1) + \lambda_2 d(x,\, y_2)\right]\right\} \cdot \frac{\partial}{\partial y_0}\, d(x,\, y_0) = 0 \tag{26}$$

where

$$Z(x) = \left[\int dj\, r(j) \int dk\, s(k) \exp -\frac{1}{T}\, \mathcal{D}(x,\, \hat{x}^0_{jk},\, \hat{x}^1_j,\, \hat{x}^2_k)\right]^{-1}$$

depends only on x.

Lemma 3: The ranges of the optimum maps $\hat{x}^0$, $\hat{x}^1$, and $\hat{x}^2$ are discrete and finite in the following cases:

[3]Since we derive a necessary condition, we do not need to be too careful about how restrictive our definition of admissibility is. Hence, we simply require that admissible functions be measurable, that the required integrals exist, and that changing the order of integration and differentiation (where needed) is allowed.

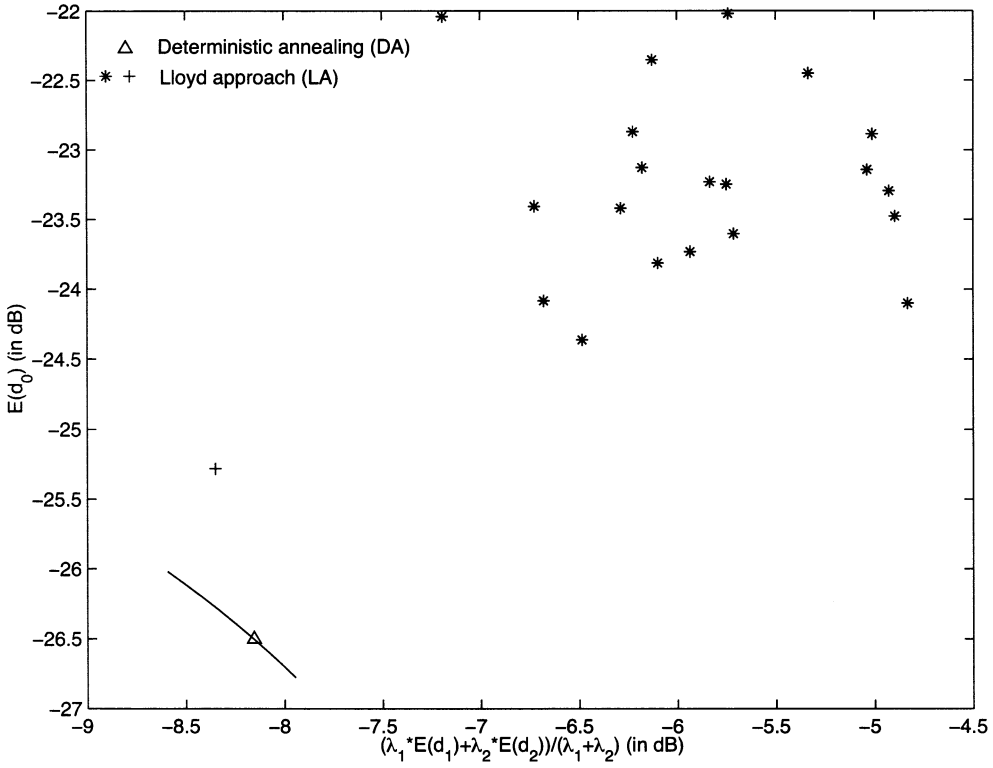

Fig. 1. MDSQ for Gaussian source. $\boldsymbol{R}_1 = \boldsymbol{R}_2 = 3$ bpss, $\boldsymbol{\lambda}_1 = \mathbf{0.006}$, $\boldsymbol{\lambda}_2 = \mathbf{0.012}$. Minimum and maximum $\boldsymbol{D}$ for LA are $-$22.52 dB and $-$19.55 dB. $\boldsymbol{D}$ for DA $= -$23.02 dB, achieved by the distortion triple $(\boldsymbol{D}_0, \boldsymbol{D}_1, \boldsymbol{D}_2) = (-$26.50 dB, $-$4.60 dB, $-$12.51 dB). (For $(\boldsymbol{R}_1, \boldsymbol{R}_2, \boldsymbol{D}_1, \boldsymbol{D}_2) = (\mathbf{3}, \mathbf{3}, -$4.60 dB, $-$12.51 dB), $\boldsymbol{D}$ promised by [15] is $-$25.18 dB.) For ease of comparison, a line along which $\boldsymbol{D} = -$23.02 dB is drawn. Design with initialization from [18] is marked by $+$.

1) $d(x, y)$ is a real, monotonic entire function of x and y, and the source alphabet $\mathcal{X}$ is discrete and finite;
2) $d(x, y) = \|x - y\|^2$ is the squared-error distortion, and the source alphabet $\mathcal{X}$ is compact.

Proof:

Case 1): If $\mathcal{X}$ is discrete and finite, the integral over $p(x)$ in (26) may be replaced by a finite sum. The left-hand side of (26) is then a real entire function of y_0, y_1, and y_2, since the entire functions are closed under differentiation, composition, and finite summation. The monotonicity of d ensures that the ranges of $\hat{x}_0$, $\hat{x}_1$, and $\hat{x}_2$ are within $\mathcal{X}$. Liouville's theorem then demands that these ranges be discrete and finite.

Case 2): It follows from [17, proof of Theorem 2] that the solution sets $\{y_0\}$, $\{y_1\}$, and $\{y_2\}$ for an integral equation of the form of (26), with $d(x, y) = \|x-y\|^2$, cannot have an accumulation point. Since $\mathcal{X}$ is compact, this implies that the solution sets are, in fact, discrete and finite. □

Fix the values of the Lagrangian multipliers T, λ_1, and λ_2, and consider the corresponding triple (D_0, D_1, D_2). The above discussion then shows that the minimum rate-sum of the section of $\mathcal{S}$ at (D_0, D_1, D_2) may be calculated using (24), (25), and (21) in the minimization of (22). If it is now assumed that the conditions of Lemma 3 are satisfied, then $\mathcal{J}$ and $\mathcal{K}$ may be taken to be discrete and finite, so that the integrals over these spaces in (24), (25), (21), and (22) may be replaced by finite sums. In this case, note that the Lagrangian functional of (22) is identical to the free energy of MCDA at the temperature T, (12). Similarly, the transition probability distribution of (21) is identical to the transition probabilities (11) of MCDA. The necessary conditions for optimality of reproduction points of (24) and (25) are identical to the codevector update rules of (14). Thus, the MCDA iterations at a fixed temperature T, and for fixed λ_1 and λ_2, are identical to the calculations of (24), (25), (21), and (22) to determine the minimum-rate sum on $\mathcal{S}$ corresponding to (D_0, D_1, D_2).

Consider the MCDA algorithm for MDVQ design, with prespecified constraints on the sizes of the various codebooks. These translate to corresponding constraints on the sizes of the solution sets of (24) and (25). As T is decreased from an initial high value, the MCDA algorithm "crawls up" the convex hull $\mathcal{S}$. Note that the correct number of code vectors to use at any temperature arises naturally in the MCDA algorithm, with code vectors splitting into new code vectors at the phase transitions.[4] Once the constraints on the solution set sizes are reached, the algorithm may be interpreted as calculating the best approximation with fixed reproduction alphabet sizes to the minimum rate sums on $\mathcal{S}$. Finally, at the limit of zero temperature, the algorithm produces an MDVQ with the desired codebook sizes, as explained in the previous section.

V. Simulation Results

The proposed DA-based design algorithm may be used to design unstructured MDVQs with unequal rate and distortion constraints on the two channels. We illustrate the wide applicability of the DA algorithm by considering three examples: 1) scalar quantizer design for unequal rate and equal distortion constraints, 2) scalar quantizer design for equal rate and unequal distortion constraints, and 3) two-dimensional vector quantizer design for equal rate and distortion constraints.

[4]An additional approximation may be introduced since we ignore the possibility of phase transitions where new code vectors grow continuously from zero mass. See [17] for a more detailed discussion of such phase transitions.

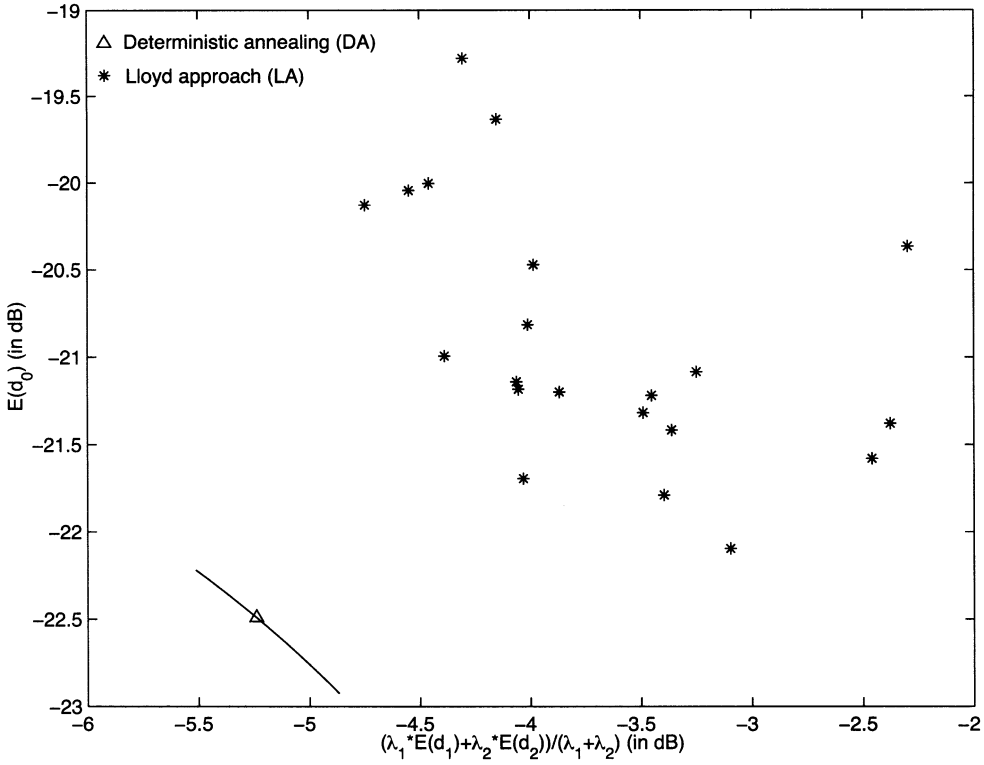

Fig. 2. MDSQ for Gaussian source. $\boldsymbol{R}_1 = 3$ bpss, $\boldsymbol{R}_2 = 2$ bpss, $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_2 = \mathbf{0.01}$. Minimum and maximum $\boldsymbol{D}$ for LA are $-$18.34 dB and $-$16.78 dB. $\boldsymbol{D}$ for DA $= -$19.35 dB, achieved by the distortion triple $(\boldsymbol{D}_0, \boldsymbol{D}_1, \boldsymbol{D}_2) = (-$22.49 dB, $-$6.31 dB, $-$4.38 dB). (For $(\boldsymbol{R}_1, \boldsymbol{R}_2, \boldsymbol{D}_1, \boldsymbol{D}_2) = (\mathbf{3}, \mathbf{3}, -$6.31 dB, $-$4.38 dB), $\boldsymbol{D}$ promised by [15] is $-$21.45 dB.) For ease of comparison, a line along which $\boldsymbol{D} = -$19.35 dB is drawn.

For comparison, we consider the performance of the existing iterative MDVQ design technique [18], [7], which we call the "Lloyd approach" (LA) as it is directly based on Lloyd's algorithm for conventional scalar quantizer design [13] and its vector extension [12]. Recall that the performance of LA depends heavily on the initialization. We use 20 different *random* initializations for the LA in our simulations. The initialization proposed in [18] is for MDSQ design with equal rate and distortion constraints. In particular, this initialization does not generalize to vectors or for unequal rate constraints. However, as an additional comparison, we used this initialization for MDSQ design with equal rate but unequal distortion constraints.

In all three examples, the quantizer designed by DA is seen to yield a significantly lower expected distortion cost than LA with random/heuristic initializations. Further, the wide variation in performance of the quantizers designed by LA illustrates and emphasizes the significance of the problem of local minima even for simple low-rate quantizers.

In Fig. 1, we present the results for the design of scalar quantizers with unequal distortion constraints for a unit-variance Gaussian source. The constraints were: $R_1 = R_2 = 3$ bpss and $\lambda_1 = 0.006$, $\lambda_2 = 0.012$. The training set consisted of 5000 samples. The quantizers produced by LA with different random initializations show wide variation in performance (the best and the worst of these designs differ by $\sim$3 dB in terms of the expected distortion cost). Note that LA initialized with the heuristic proposed in [18] yields significant gains over random initialization, and demonstrates the benefits of a good heuristic. However, MDSQ designed via the proposed DA approach outperforms by $\sim$0.5 dB the best of all LA initializations.

In Fig. 2, we present results for scalar quantizer design under unequal rate constraints. The constraints were: $R_1 = 3$ bpss, $R_2 = 2$ bpss, and $\lambda_1 = \lambda_2 = 0.01$. The training set consisted of 5000 samples of a unit-variance Gaussian source. The DA design is compared with randomly initialized designs of LA. The DA design gains $\sim$1 dB over the best of the latter in terms of the expected distortion cost. Note that the heuristic index assignment cannot be extended to this case.

Recall that Ozarow determined, in [15], the MD achievable region for the memoryless Gaussian source. We also compared the DA-based designs of Figs. 1 and 2 with the theoretical benchmark provided by [15]. While the precise comparisons are noted in the captions of the corresponding figures, here it must only be noted that the expected distortion costs of the MDSQs designed by DA are within $\sim$2.2 dB of the respective optimal expected distortion costs (in the limit of infinite block lengths) promised by the results of Ozarow.

In Fig. 3, we present the results for the design of two-dimensional vector quantizers for a Gauss–Markov source with autocorrelation coefficient $\rho = 0.9$ and unit variance per dimension. A training set of 5000 vectors was used. The rate and distortion constraints were $R_1 = R_2 = 1.5$ bpss (i.e., each side codebook has eight two-dimensional code vectors) and $\lambda_1 = \lambda_2 = 0.01$. We compare the performance of DA design with quantizers produced by LA for 20 different random initializations. The distortion cost of the MDVQ designed by DA is $\sim$0.6 dB below the distortion of the "best" quantizer produced by random initializations of LA. Note that the heuristic index assignment proposed in [18] cannot be generalized to this case.

Finally, note that the DA-based algorithm has a longer running time than LA, but the ratio of the running times of the two algorithms is a constant, influenced by the choice of the cooling schedule for the former.

VI. Conclusion

Deterministic annealing is proposed for the design of MDVQs when the two channels need not have identical capacities or failure probabilities. This approach eliminates the dependence on initial configuration

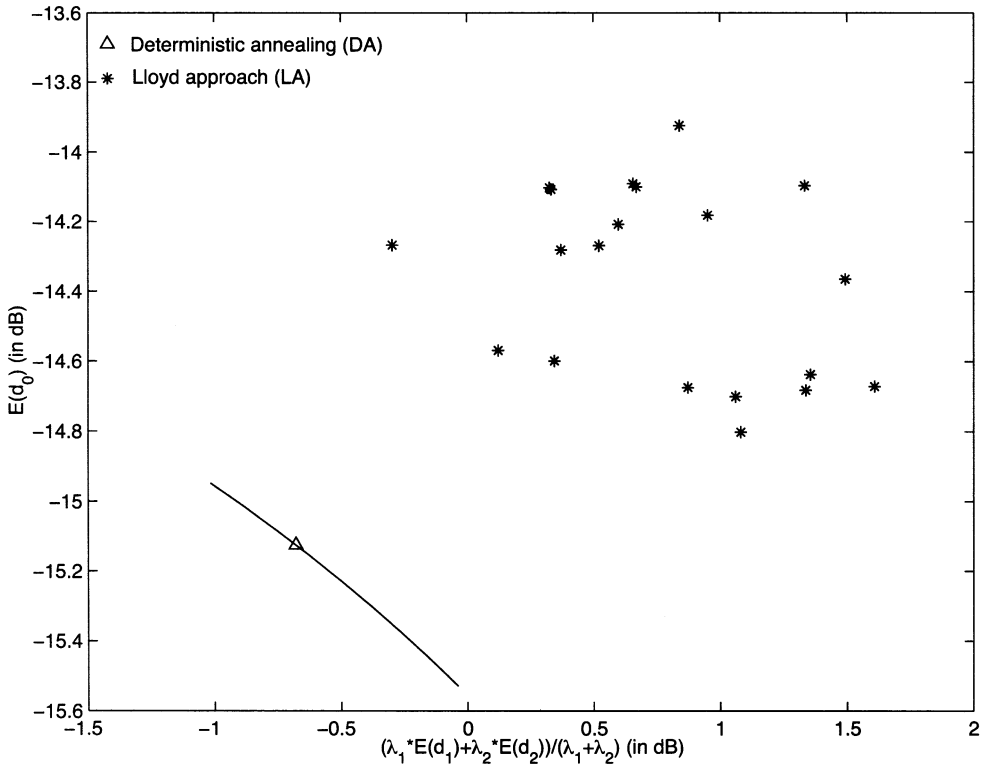

Fig. 3. Two-dimensional MDVQ for Gauss–Markov source $\boldsymbol{\rho = 0.9}$. $\boldsymbol{R_1 = R_2}$ = 1.5 bpss, $\boldsymbol{\lambda_1 = \lambda_2 = 0.01}$. Minimum and maximum $\boldsymbol{D}$ for LA are $-$12.56 dB and $-$11.79 dB. $\boldsymbol{D}$ for $\boldsymbol{DA = -}$13.20 dB, with corresponding distortion triple $\boldsymbol{(D_0, D_1, D_2) = (}-15.13$ dB, -1.76 dB, 0.19 dB). For ease of comparison, a line along which $\boldsymbol{D = -}$13.20 dB is drawn.

and avoids many poor local minima of the cost surface. Further, no knowledge is assumed on the underlying probability distribution of the source. DA is motivated by analogy to statistical physics and is derived from principles of information theory. A random encoding rule is used, and the encoding probabilities are determined by minimization of the expected distortion cost at a specified level of entropy. The algorithm starts at the global minimum at high temperature and tracks the minimum while lowering the temperature. An MD quantizer is obtained at the limit of low temperature. We compared our approach with existing methods, and obtained consistent, substantial improvements.

Appendix

A continuous phase transition occurs when the temperature is reduced below a critical value, if the existing solution changes from a minimum of the Lagrangian functional F^* of (7) to a saddle point or a local maximum. We use this condition, and variational calculus, to derive an expression for the critical temperatures.

Let us consider the perturbed central and side codebooks given by

$$\hat{\mathcal{X}}^0 + \epsilon\Psi^0 = \{\hat{x}^0_{jk} + \epsilon\psi^0_{jk},\ (j,\ k) \in J \times K\}$$
$$\hat{\mathcal{X}}^1 + \epsilon\Psi^1 = \{\hat{x}^1_{j} + \epsilon\psi^1_{j},\ j \in J\}$$

and

$$\hat{\mathcal{X}}^2 + \epsilon\Psi^2 = \{\hat{x}^2_{k} + \epsilon\psi^2_{k},\ k \in K\}$$

where ψ^0_{jk}, ψ^1_j, and ψ^2_k are the perturbation vectors, and the nonnegative scalar ϵ is used to scale the perturbations. We denote $(\Psi^0,\ \Psi^1,\ \Psi^2)$ by Ψ and the vector of concatenated perturbations $(\psi^0_{jk}\ \ \psi^1_j\ \ \psi^2_k)$ by ψ_{jk}. Further, we define the concatenation of central and side error vectors as $e_{jk} = ((x - \hat{x}^0_{jk})\ (x - \hat{x}^1_j)\ (x - \hat{x}^2_k))$. Here, and in the subsequent derivations, all vectors (unless transposed) are row vectors.

In terms of the Lagrange functional of (7) evaluated with the perturbed codebooks $F^*(\hat{\mathcal{X}}^0 + \epsilon\Psi^0,\ \hat{\mathcal{X}}^1 + \epsilon\Psi^1,\ \hat{\mathcal{X}}^2 + \epsilon\Psi^2)$, we can write the necessary condition for the optimality of the codebooks $(\hat{\mathcal{X}}^0,\ \hat{\mathcal{X}}^1,\ \hat{\mathcal{X}}^2)$ as

$$\frac{\partial}{\partial\epsilon}F^*(\hat{\mathcal{X}}^0 + \epsilon\Psi^0,\ \hat{\mathcal{X}}^1 + \epsilon\Psi^1,\ \hat{\mathcal{X}}^2 + \epsilon\Psi^2)|_{\epsilon=0} = 0$$

for all choices of finite perturbation Ψ. (Note that this leads directly to the centroid rules of (9).) We must also require a condition on the second derivative to ensure the minimum is stable

$$\frac{\partial^2}{\partial\epsilon^2}F^*(\hat{\mathcal{X}}^0 + \epsilon\Psi^0,\ \hat{\mathcal{X}}^1 + \epsilon\Psi^1,\ \hat{\mathcal{X}}^2 + \epsilon\Psi^2)|_{\epsilon=0} \geq 0 \qquad (27)$$

for all choices of finite perturbation Ψ. A necessary condition for bifurcation is equality in (27). Applying straightforward differentiation, we obtain the following condition for equality in (27):

$$T\int dx\ p(x)\left[\sum_{jk}\left(\frac{2}{T}\right)q(j,\ k|x)e_{jk}L^2\psi^t_{jk}\right]^2$$
$$+2\sum_{jk}q(j,\ k)\psi_{jk}L\left[I_{3n} - \left(\frac{2}{T}\right)LC_{x|j,\,k}L\right]L\psi^t_{jk} = 0 \qquad (28)$$

where $q(j,\ k|x)$ is given by (6). Here, I_{3n} is the $(3n \times 3n)$ identity matrix

$$L = \begin{pmatrix} I_n & 0 & 0 \\ 0 & \sqrt{\lambda_1}I_n & 0 \\ 0 & 0 & \sqrt{\lambda_2}I_n \end{pmatrix}$$

where I_n is the $(n \times n)$ identity matrix, and

$$C_{x|jk} = \sum_x p(x|j,\ k)e^t_{jk}e_{jk}$$

is the covariance matrix of the posterior distribution $p(x|j,\ k)$ of the cluster corresponding to the index pair $(j,\ k)$.

We claim that the left-hand side of (27) is positive for all perturbations iff the second term of (28) is positive. The "if" part is trivial since the first term of (28) is nonnegative. We prove the "only if" part. Consider a subset of index pairs, $\mathcal{C}$, with coincident central and side code vectors. This subset bifurcates if the matrix $I_{3n} - (\frac{2}{T})LC_{x|jk}L$ loses positive definiteness, in which case the second term on the left-hand side of (28) can be nonpositive. (Note that $C_{x|jk}$ is the same for all $(j, k) \in \mathcal{C}$.) We now show a particular perturbation that makes the first term vanish in this case. In fact

$$\psi_{jk} = 0, \qquad \forall (j, k) \notin \mathcal{C}, \quad \text{and} \quad \sum_{(j,k)\in\mathcal{C}} \psi_{jk} = 0$$

works. So the subset $\mathcal{C}$ bifurcates at temperature T if the conditional distribution $p(x|jk)$ satisfies the condition

$$\det\left[I_{3n} - \left(\frac{2}{T}\right) LC_{x|jk}L\right] = 0. \tag{29}$$

The above condition is implicit in the critical temperature. The critical temperature for the first phase transition (i.e., when the code vectors coincident at the centroid of the source distribution move apart for the first time) can be explicitly evaluated, giving

$$T_{c_1} = 2(1 + \lambda_1 + \lambda_2)\alpha_{\max} \tag{30}$$

where $\alpha_{\max}$ is the largest eigenvalue of the source covariance matrix. This critical temperature may be compared with the critical temperature for the first phase transition when DA is used for single-description VQ design [16]: $T_{c_1}^{\mathrm{SDVQ}} = 2\alpha_{\max}$, and

$$T_{c_1}^{\mathrm{MDVQ}} = (1 + \lambda_1 + \lambda_2)T_{c_1}^{\mathrm{SDVQ}}.$$

This result is consistent with the observation that the MDVQ design algorithm degenerates to the DA algorithm for single description VQ design [16] if $\lambda_1 = \lambda_2 = 0$.

ACKNOWLEDGMENT

The authors wish to thank Ajit Rao for fruitful discussions and insightful comments at an early stage of this work.

REFERENCES

[1] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 14–20, Jan. 1972.

[2] R. Balan, I. Daubechies, and V. A. Vaishampayan, "The analysis and design of windowed Fourier frame based multiple description source coding schemes," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2491–2536, Nov. 2000.

[3] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 460–473, July 1972.

[4] P .A. Chou, S. Mehrotra, and A. Wang, "Decoding of overcomplete expansions using projections onto convex sets," in *Proc. IEEE Data Compression Conf.*, Snowbird, UT, Mar. 1999, pp. 72–81.

[5] S. N. Diggavi, N. J. A. Sloane, and V. A. Vaishampayan, "Asymmetric multiple description lattice vector quantizers," *IEEE Trans. Inform. Theory*, vol. 48, pp. 174–191, Jan. 2002.

[6] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 851–857, Nov. 1982.

[7] M. Fleming and M. Effros, "Generalized multiple description vector quantization," in *Proc. IEEE Data Compression Conf.*, Snowbird, UT, Mar. 1999, pp. 3–12.

[8] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston, MA: Kluwer, 1992.

[9] V. K. Goyal, J. Kovacevic, and M. Vetterli, "Quantized frame expansions as source-channel codes for erasure channels," in *Proc. IEEE Data Compression Conf.*, Snowbird, UT, Mar. 1999, pp. 326–335.

[10] A. Ingle and V. A. Vaishampayan, "DPCM system design for diversity systems with applications to packetized speech," *IEEE Trans. Speech Audio Processing*, vol. 1, pp. 48–58, Jan. 1995.

[11] E. T. Jaynes, "Information theory and statistical mechanics," in *Papers on Probability, Statistics and Statistical Physics*, R. D. Rosenkratz, Ed. Dordrecht, The Netherlands: Kluwer, 1989.

[12] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. COM-28, pp. 84–95, Jan. 1980.

[13] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 129–137, Mar. 1982.

[14] M. Orchard, Y. Wang, V. A. Vaishampayan, and A. Reibman, "Redundancy rate distortion analysis of multiple description coding using pair-wise correlating transforms," in *Proc. IEEE Int. Conf. Image Processing*, Oct. 1997.

[15] L. Ozarow, "On a source coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, pp. 1909–1921, Dec. 1980.

[16] K. Rose, E. Gurewitz, and G. C. Fox, "Vector quantization by deterministic annealing," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1249–1257, July 1992.

[17] K. Rose, "A mapping approach to rate-distortion computation and analysis," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1939–1952, Nov. 1994.

[18] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inform. Theory*, vol. 39, pp. 821–834, May 1993.

[19] V. A. Vaishampayan and S. John, "Balanced interframe multiple description video compression," in *Proc. IEEE Int. Conf. Image Processing (ICIP1999)*, vol. 3, Oct. 1999, pp. 812–816.

[20] V. A. Vaishampayan and A. A. Siddiqui, "Speech predictor design for diversity communication systems," in *Proc. 1995 IEEE Speech Coding Workshop*, Sept. 1995.

[21] J. Wenqing and A. Ortega, "Multiple description speech coding for robust communication over lossy packet networks," in *Proc. IEEE Int. Conf. Multimedia Expo. (ICME2000)*, vol. 1, July 2000, pp. 444–447.

[22] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 502–521, July 1987.