Robust Distributed Source Coder Design by Deterministic Annealing

Ankur Saxena, *Member, IEEE*, Jayanth Nayak, and Kenneth Rose, *Fellow, IEEE*

***Abstract*—This paper considers the design of efficient quantizers for a *robust* distributed source coding system. The information is encoded at independent terminals and transmitted across separate channels, any of which may fail. The scenario subsumes a wide range of source and source-channel coding/quantization problems, including multiple descriptions and distributed source coding. Greedy descent methods depend heavily on initialization, and the presence of abundant (high density of) "poor" local optima on the cost surface strongly motivates the use of a global design algorithm. We propose a deterministic annealing approach for the design of all components of a generic robust distributed source coding system. Our approach avoids many poor local optima, is independent of initialization, and does not make any simplifying assumption on the underlying source distribution. Simulation results demonstrate a wide spread in the performance of greedy Lloyd-based algorithms, and considerable gains are achieved by using the proposed deterministic annealing approach.**

***Index Terms*—Deterministic annealing, distributed source coding, sensor networks.**

I. Introduction

In a distributed network of sensors, different sensors may be designed to observe various physical quantities, e.g., temperature, humidity, pressure, light, and sound. We may be interested in efficient reconstruction of one or more physical entities measured at different, spatially separated locations. Typically, the data communicated by networks of sensors exhibit a high degree of correlation. Since the encoders at each sensor location function independently, the system will not, in practice, achieve the performance of optimal joint lossy compression of the sources. A related issue is that of estimation of a source from another correlated source. For example, if a sensor (or a transmission channel) fails, then to obtain an estimate of data being (or that would be) measured by the sensor, we can only utilize information acquired from the other sensors (or channels). To achieve the dual objectives of obtaining the best possible compression efficiency from independent encoders and attaining system robustness, it is necessary that the code *design* at all the terminals be performed jointly for a robust distributed source coding system (see Fig. 1).

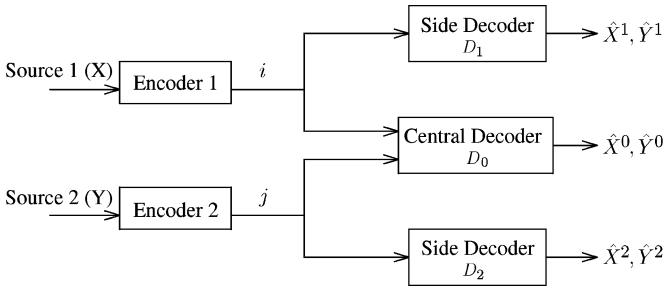

Fig. 1. Block diagram for robust distributed source coding.

The robust distributed source coding model was first proposed and studied in [1] and later in [2] and [3]. As pointed out in [3], the model subsumes a variety of source coding problems ranging from distributed source coding [4], [5] and the CEO problem [6] to multiple description coding. Estimating a source from another correlated source (see, e.g. [7] and [8]) is another special case of the robust distributed coding problem. A good design for the robust distributed coding system should be able to take into account the correlation between the sources as well as the possibility of a component failure.

Constructive and practical code design techniques for distributed coding using source and channel coding principles were proposed, e.g., by Pradhan and Ramchandran in [9]. Existing distributed coding system research can be roughly categorized into two "camps": one adopting ideas from channel coding (see, e.g., [10] and [11]), some of which use turbo/low-density parity check codes with large block-lengths and exploit long delays to achieve good performance ([12]–[15]), and another building directly on source coding methodologies [16]–[18]. The channel coding approaches can conceivably be leveraged to address robust distributed vector quantizer (RDVQ) design. However, these approaches appear most suitable when the sources can be modeled as noisy versions of each other, where the noise is unimodal in nature, and are of limited use wherever the simplifying assumptions do not apply. An illustrative example is when, say, temperature and humidity are drawn from a mixture of joint Gaussian densities, where the mixture components are due to varying underlying conditions such as the time of day, pressure, etc. Further note that the channel-coding strategies for RDVQ design will create a long delay in the system and, not surprisingly, get closer to the asymptotic bounds. However, the proposed source coding schemes in this paper will have practically zero delay. Hence, the channel coding and source coding methodologies will actually solve substantially different problems.

Manuscript received November 18, 2008; accepted August 04, 2009. First published September 01, 2009; current version published January 13, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Hongbin Li. The work was supported in part by the National Science Foundation under Grants IIS-0329267 and CCF-0728986, the University of California MICRO program, Applied Signal Technology Inc., Cisco Systems Inc., Dolby Laboratories Inc., Qualcomm Inc., and Sony Ericsson, Inc. This paper was presented in part at the IEEE Data Compression Conference, Snowbird, UT, March 2006.

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

J. Nayak is with Mayachitra Inc., Santa Barbara, CA 93111 USA (e-mail: nayak@mayachitra.com).

Digital Object Identifier 10.1109/TSP.2009.2031283

The source coding methodologies to design RDVQ can be based on Lloyd's algorithm [19], but they will suffer from the presence of numerous "poor" local minima on the distortion-cost surface, and thus will be critically sensitive to initialization. Clever initialization as proposed, for example, in the context of multiple description scalar quantizer design [20] can help mitigate this shortcoming. But such initialization heavily depends on symmetries or simplifying assumptions, and no generalizations are available to vector quantization nor to more complicated scenarios such as RDVQ. Alternatively, a global optimization scheme—i.e., a powerful optimization tool that provides the ability to avoid poor local optima and is applicable to sources exhibiting any type of statistical dependencies, such as deterministic annealing—can eliminate or substantially mitigate these shortcomings.

In [21], it has been shown that a deterministic annealing-based approach offers considerable gains over extensions of Lloyd-like iterative algorithm and various schemes employing heuristic initialization for the case of generic multiple description vector quantizer design. Numerous other applications where deterministic annealing outperforms greedy iterative algorithms can be found in a tutorial paper [22] and references therein. In this paper, an iterative greedy algorithm for RDVQ design is described that will underline the need for a global optimization approach. We then derive and propose a deterministic annealing approach for optimal RDVQ design (our preliminary results appeared in [23]).

Deterministic annealing (DA) is motivated by the process of annealing in statistical physics but is founded on principles of information theory. It is independent of the initialization, does not assume any knowledge about the underlying source distribution, and avoids many poor local minima of the distortion-cost surface [22]. In DA, a probabilistic framework is introduced via random encoding where each training sample of the input source is assigned to a reproduction value *in probability*. The optimization problem is recast as minimization of the expected distortion subject to a constraint on the level of randomness as measured by the Shannon entropy of the system. The Lagrangian functional can be viewed as the free energy of a corresponding physical system and the Lagrangian parameter as the "temperature." The minimization is started at a high temperature (highly random encoder) where, in fact, the entropy is maximized and hence all reproduction points are at the centroid of the source distribution. The minimum is then tracked at successively lower temperatures (lower levels of entropy) by recalculating the optimum locations of the reproduction points and the encoding probabilities at each stage. As the temperature approaches zero, the average distortion term dominates the Lagrangian cost and a hard (nonrandom) encoder is obtained.

Further note that the channel-coding strategies for RDVQ design will create a long delay in the system and, not surprisingly, get closer to the asymptotic bounds. However, the proposed source coding schemes in this paper (both based on the Lloyd approach and deterministic annealing approach) will have practically zero delay. Hence, the channel coding and source coding methodologies will actually solve substantially different problems, and their comparison may be misleading.

The rest of this paper is organized as follows. In Section II, we state the problem formally, establish the notation, and describe an iterative greedy method based on Lloyd's algorithm for multiple prototype coder design. This will underline the need of a global approach. In Section III, we derive the DA approach to RDVQ design, provide its update formulas (necessary optimality conditions), and discuss the phenomenon of phase transition in DA approach for RDVQ. Simulation results are given in Section IV, followed by conclusions in Section V.

II. The RDVQ Problem and Iterative Greedy Methods

A. Problem Statement and Design Considerations

Consider the robust distributed source coding scenario in Fig. 1. For brevity, we will restrict the analysis to the case of two sources, but the model can be extended in a straightforward fashion to an arbitrary number of sources. Here (X, Y) is a pair of continuous-valued, independent identically distributed (i.i.d.), correlated (scalar or vector) sources, which are independently compressed at rates R_1 and R_2 bits per sample, respectively. The encoded indexes i and j are transmitted over two separate channels, which may or may not be in working order, and the channel condition is not known at the encoders. The end-user tries to obtain the best estimate of the sources depending on the descriptions received from the functioning channels. Let $(\hat{X}^0, \hat{Y}^0)$ denote the reconstruction values for sources (X, Y), which are produced by the central decoder D_0, i.e., when information is available from both channels. If only channel 1 (or 2) is working, then side decoder D_1 (or D_2) is used to reconstruct $(\hat{X}^1, \hat{Y}^1)$ (or $(\hat{X}^2, \hat{Y}^2)$). The objective of the RDVQ is to minimize the following overall distortion function given rate allocations of R_1 and R_2:

$$\begin{aligned} D_{RDVQ} = E\Big\{ & \lambda_0 \left[\alpha_0 d(X, \hat{X}^0) + (1-\alpha_0) d(Y, \hat{Y}^0)\right] \\ & + \lambda_1 \left[\alpha_1 d(X, \hat{X}^1) + (1-\alpha_1) d(Y, \hat{Y}^1)\right] \\ & + \lambda_2 \left[\alpha_2 d(X, \hat{X}^2) + (1-\alpha_2) d(Y, \hat{Y}^2)\right] \Big\} \quad (1) \end{aligned}$$

where $d(\cdot, \cdot)$ is an appropriately defined distortion measure and $\alpha_n \in [0, 1]$ $\{n = 0, 1, 2\}$ governs the relative importance of the sources X and Y at decoder n. The first two terms in the RDVQ cost of (1) contribute to the central distortion when both channels work. Similarly, the remaining terms correspond to the distortions for side decoders 1 and 2 when only one channel is in working condition. The central distortion is weighted by λ_0, while the side distortions are weighted by λ_1 and λ_2, whose specific values depend on the importance we wish to give to the side distortions as compared to the central distortion. In a practical system, λ_0, λ_1, and λ_2 will often be determined by the channel failure probabilities.

The RDVQ problem comprises the design of mappings from the sources X and Y to indexes at the respective encoders and of the corresponding reconstruction values at the three decoders. To minimize the overall distortion for given transmission rates, the correlation between the sources must be exploited. This may be done by sending the same index for many, possibly noncontiguous regions of the source alphabet on a channel and then

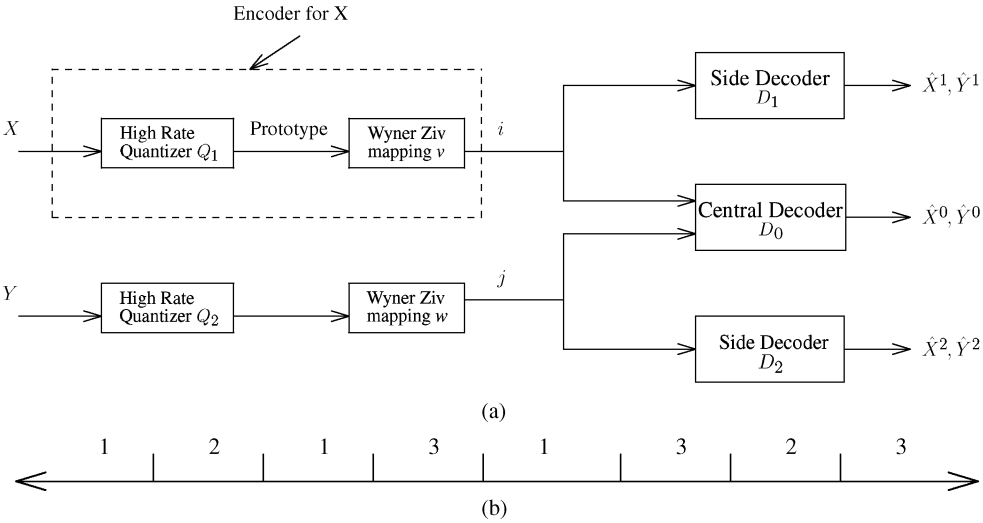

Fig. 2. (a) Breakup of encoder in robust distributed coding and (b) an example of Wyner–Ziv mapping from prototypes (Voronoi regions) to indexes.

using the information from the other source to distinguish between index-sharing regions. In the case that only one channel is functioning, the RDVQ problem reduces to estimating a signal from another correlated source. On the other hand, if both the channels work and the central decoder is used, the problem reduces to that of correlated source coding. Locally optimal quantizer design techniques for general networks (which encompass the RDVQ model as well) and correlated source coding have been proposed in the literature in [16]–[18], respectively. We next adopt this framework and describe a locally optimal algorithm using multiple prototypes (MP) for the design of a generic RDVQ system. The MP approach can be viewed as combining histogram or kernel-based techniques for source distribution estimation and quantizer design.

Specifically, we have a training set $\mathcal{T}$, which consists of N data pairs for (possibly scalar or vector) correlated sources (X, Y). Each source is assumed to be i.i.d. We design a high-rate vector quantizer Q_1 for X using a standard VQ design algorithm such as Lloyd's algorithm [19] or DA [22]. Q_1 assigns training set data points to one of the $\mathcal{K}$ regions, C_k^x. The disjoint Voronoi regions C_k^x span the source space, and a prototype x_k is associated with each of them. Next, each Voronoi region is mapped to one of the $\mathcal{I} = \{1, \ldots, I\}$ indexes via a mapping $v(k) = i$, to which we refer as Wyner–Ziv (WZ) mapping (the name loosely accounts for the fact that the scenario involves lossy coding with side information whose asymptotic performance bound was given in [5]). The index i is then transmitted across the channel. An example of WZ mapping for a scalar source X with $\mathcal{K} = 8$ and $\mathcal{I} = 3$ is given in Fig. 2. The region associated with index i is denoted $R_i^x = \bigcup_{k:v(k)=i} C_k^x$.

We similarly define quantizer Q_2, regions C_l^y, R_j^y, and prototypes y_l in the Y domain. Here, the $\mathcal{L}$ Voronoi regions are mapped to $\mathcal{J}$ indexes via WZ mapping $w(l) = j$. At the central decoder, we receive indexes in $\mathcal{I} \times \mathcal{J}$ and generate reconstruction values $\hat{x}_{ij}^0$ and $\hat{y}_{ij}^0$ (where $\hat{x}_{ij}^0 \in \mathcal{X}^0$, $(i, j) \in \mathcal{I} \times \mathcal{J}$ etc.). At the side decoder 1 (or 2), the received index is in $\mathcal{I}(\mathcal{J})$, and reconstruction values are $\hat{x}_i^1(\hat{x}_j^2)$ and $\hat{y}_i^1(\hat{y}_j^2)$. Note that we use uppercase letters for a random variable and lowercase letters to denote their particular realization.

The distortion for a data pair (x, y) and corresponding index pair (i, j) is given by

$$\begin{aligned} D_{\text{net}}(x, y, i, j) = {} & \lambda_0\alpha_0 d\left(x, \hat{x}_{ij}^0\right) + \lambda_1\alpha_1 d\left(x, \hat{x}_i^1\right) \\ & + \lambda_2\alpha_2 d\left(x, \hat{x}_j^2\right) + \lambda_0(1-\alpha_0) d\left(y, \hat{y}_{ij}^0\right) \\ & + \lambda_1(1-\alpha_1) d\left(y, \hat{y}_i^1\right) \\ & + \lambda_2(1-\alpha_2) d\left(y, \hat{y}_j^2\right). \end{aligned} \tag{2}$$

The net distortion in (1) that we seek to minimize simply averages the distortion from all the source data points. In the next subsection, we outline an iterative greedy strategy for the design of a RDVQ system. The design strategy is based on the multiple prototype framework and is similar in spirit with the algorithms presented in [16]–[18] for various versions of correlated source coding.

B. Greedy Iterative Design Strategy

The high-rate quantizers Q_1 and Q_2 for X and Y may be designed using a standard quantizer design algorithm such as Lloyd's algorithm [19] or DA [22] (to minimize the distortion between the source and the prototypes). Note that the actual objective is to minimize the distortion between the sources and their reconstruction values, and the primary task of the high rate quantizers is to discretize the source. As long as the output rate of these quantizers is sufficiently high (in comparison to the transmitted rate), the performance loss due to such discretization will be marginal. Although the output of the high-rate quantizer is not directly transmitted over the channel, a large number of prototypes can incur a significant overhead in terms of the processing and storage complexity of the encoder. This limits the allowable rate of these quantizers in practice. In such circumstances, careful design of the quantizer modules will be critical for the overall system performance. A design strategy for the case of limited encoder-storage/processing complexity where the quantizer modules are optimized for the distributed source coding scenario was presented in [24].

In this paper, we focus on the setting where storage at the encoders is not a critical issue, and the quantizer modules Q_1 and Q_2 may simply have high rate. Given fixed Q_1 and Q_2 (see Fig. 2), the WZ mappings v and w, as well as the reconstruction

values at various decoders, can be optimized iteratively by using a Lloyd-like iterative algorithm. The equations for updating the various entities are as follows.

1) *WZ Mapping for X:* For $k = 1, \ldots, \mathcal{K}$, assign k to index i, such that

$$v(k) = i = \arg\min_{i'} \sum_{\substack{(x,y)\in\mathcal{T};\\ x\in C_k^x}} D_{\text{net}}(x, y, i', j). \tag{3}$$

2) *WZ Mapping for Y:* For $l = 1, \ldots, \mathcal{L}$, assign l to index j, such that

$$w(l) = j = \arg\min_{j'} \sum_{\substack{(x,y)\in\mathcal{T};\\ y\in C_l^y}} D_{\text{net}}(x, y, i, j'). \tag{4}$$

3) *Reconstruction Values for X:* For all $i = 1, \ldots, \mathcal{I}$ and $j = 1, \ldots, \mathcal{J}$, find $\hat{x}_{ij}^0$, $\hat{x}_i^1$, and $\hat{x}_j^2$ such that

$$\hat{x}_{ij}^0 = \arg\min_{a_0} \sum_{\substack{(x,y)\in\mathcal{T}; x\in R_i^x,\\ y\in R_j^y}} d(x, a_0) \tag{5}$$

$$\hat{x}_i^1 = \arg\min_{a_1} \sum_{(x,y)\in\mathcal{T}; x\in R_i^x} d(x, a_1) \tag{6}$$

$$\hat{x}_j^2 = \arg\min_{a_2} \sum_{(x,y)\in\mathcal{T}; y\in R_j^y} d(x, a_2). \tag{7}$$

The corresponding update equations for the reconstruction values of Y have not been reproduced here but can be trivially obtained by symmetry.

At this point, we reemphasize that it is the WZ module that exploits the correlation between the quantized versions of source. The above technique optimizes the WZ mappings from prototypes to indexes for X and Y and the final reconstruction values at the various decoders in an iterative manner. We will thus refer to the above design algorithm as the Lloyd approach (LA). LA inherits from the original Lloyd's algorithm the interrelated shortcomings of getting trapped in poor local minima and dependence on initialization. The suboptimality of LA will be observed experimentally in the results section. These issues call for the use of a global optimization scheme, such as DA. We next present the DA algorithm and the necessary conditions for optimality in RDVQ design.

III. The Deterministic Annealing Approach

A. Derivation

A formal derivation of the DA algorithm is based on principles borrowed from information theory and statistical physics. Here the standard deterministic encoder is replaced by a random encoder, and the expected distortion is minimized subject to an entropy constraint that controls the "randomness" of the solution. By gradually relaxing the entropy constraint, we obtain an annealing process that seeks the minimum distortion solution. More detailed derivation and the principle underlying DA can be found in [22].

Given the RDVQ setup, we separately design quantizers Q_1 and Q_2 for the two sources using DA [22]. As mentioned earlier in Section II-B, the rationale for this separate design is that as long as the number of prototypes per index is large, the correlation between the quantized versions of the sources can be fully exploited within the WZ mapping modules of the encoders. This means that efficient WZ mappings from prototypes to indexes is crucial for the overall system performance. The DA approach for RDVQ optimizes these mappings and the reconstruction values jointly, is independent of the initialization, and converges to a considerably better minimum.

The high-rate quantizer Q_1 for source X assigns each data point in the training set for source X to a prototype x_k. We define binary variables that specify the deterministic quantizer rule

$$c_{k|x} = \begin{cases} 1, & \text{if } Q_1(x) = k \\ 0, & \text{otherwise} \end{cases}. \tag{8}$$

The random WZ mapping is specified by the probability variables $r_{i|k} = \Pr[i|k] = \Pr[x_k \in R_i^x]$, i.e., the probability that the kth prototype x_k falls in the (random) cell R_i^x. The effective probability that a point x belongs to the random cell R_i^x is thus given by

$$p_{i|x} = \Pr[x \in R_i^x] = \sum_k r_{i|k} c_{k|x}. \tag{9}$$

Similarly, in the Y domain, we define

$$c_{l|y} = \begin{cases} 1, & \text{if } Q_2(y) = l \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

$r_{j|l} = \Pr[j|l] = \Pr[y_l \in R_j^y]$ and $p_{j|y} = \Pr[y \in R_j^y] = \sum_l r_{j|l} c_{l|y}$. Note that

$$\sum_k c_{k|x} = 1 \quad \text{and} \quad \sum_l c_{l|y} = 1 \tag{11}$$

since a data point is associated with only one prototype.

The probabilistic equivalent of the distortion function D_{RDVQ} in (1) that we seek to minimize is

$$D = \frac{1}{N} \sum_{(x,y)\in\mathcal{T}} \sum_{i,j} p_{i|x} p_{j|y} D_{\text{net}}(x, y, i, j) \tag{12}$$

$$= \frac{1}{N} \sum_{(x,y)\in\mathcal{T}} \sum_{k,l,i,j} c_{k|x} c_{l|y} r_{i|k} r_{j|l} D_{\text{net}}(x, y, i, j) \tag{13}$$

subject to a constraint on the *joint entropy* H of the system. Here N is the number of data points in the training set. This is equivalent to the following Lagrangian minimization:

$$\min_{\{r_{i|k}\},\{r_{j|l}\},\{\hat{x}_{ij}^0\},\{\hat{y}_{ij}^0\},\{\hat{x}_i^1\},\{\hat{y}_i^1\},\{\hat{x}_j^2\},\{\hat{y}_j^2\}} \{L = D - TH\} \tag{14}$$

where the "temperature" T plays the role of Lagrange parameter.

The joint entropy of the system is $H = H(X, Y, K, L, I, J) = H(X, Y) + H(K, I|X) + H(L, J|Y)$, since by construction, the source variables X and Y, prototypes K and L, and transmitted indexes I and J form a Markov chain: $J - L - Y - X - K - I$. Also, $H(X, Y)$ is the source entropy and is unchanged by the encoding decisions for a given training set. The solution will therefore depend on

the conditional entropy terms $H(K,I|X)$ and $H(L,J|Y)$. $H(K,I|X)$ is given by

$$H(K,I|X) = \frac{-1}{N} \sum_{(x,y)\in\mathcal{T}} \sum_{k,i} c_{k|x} r_{i|k} \log(c_{k|x} r_{i|k}) = \frac{-1}{N} \sum_{(x,y)\in\mathcal{T}} \sum_{k,i} c_{k|x} r_{i|k} \log(r_{i|k}) \quad (15)$$

using the fact that $c_{k|x}$ in (8) can take values zero and one only. Here the base of logarithm is two. Similarly, $H(L,J|Y)$ is given by

$$H(L,J|Y) = \frac{-1}{N} \sum_{(x,y)\in\mathcal{T}} \sum_{l,j} c_{l|y} r_{j|l} \log(r_{j|l}). \quad (16)$$

Next, we derive the necessary conditions for minimizing the Lagrangian cost in (14).

B. Update Equations for RDVQ Design

At a fixed temperature T, the objective function in (14) is convex in terms of the probabilities $r_{i|k}$ and $r_{j|l}$. The optimal expressions for $r_{i|k}$ and $r_{j|l}$ are given by

$$r_{i|k} = \frac{e^{-D_{ki}/T}}{\sum_{i'} e^{-D_{ki'}/T}} \quad \text{and} \quad r_{j|l} = \frac{e^{-D_{lj}/T}}{\sum_{j'} e^{-D_{lj'}/T}} \quad (17)$$

where

$$D_{ki} = E\left[D_{\text{net}}(X,Y,i,J)|X \in C_k^x\right] \quad D_{lj} = E\left[D_{\text{net}}(X,Y,I,j)|Y \in C_l^y\right]. \quad (18)$$

The distortion term D_{ki} can be interpreted as the average distortion for the data points that belong to the kth Voronoi region (for source X) and are being mapped to the ith transmitted index. The encoding probability $r_{i|k}$ follows a Gibbs distribution. At a particular temperature T, the kth Voronoi region will be most associated with the ith index for which the average distortion D_{ki} is minimum (for a fixed k, $r_{i|k}$ will be maximum for the ith index when $D_{ki} < D_{ki'}, \forall i' \neq i$). Note that the kth Voronoi region is still associated with the other indexes but at lower probabilities. However, at the limit $T \to 0$, these association probabilities become either one or zero, and a hard mapping rule is obtained.

We next give the expressions for the reconstruction values in the case of the squared-error distortion measure. The general approach is clearly not restricted to this choice of distortion measure

$$\hat{x}_{ij}^0 = E\left[X|X \in R_i^x, Y \in R_j^y\right], \quad \hat{x}_i^1 = E\left[X|X \in R_i^x\right] \quad \hat{x}_j^2 = E\left[X|Y \in R_j^y\right]. \quad (19)$$

These update rules are relatives of the standard centroid rule and are simply weighed by the various association probabilities. Also note that side decoder 2 does not have access to X, and the reconstruction of X is done solely based on the information received from source Y. By the symmetry in the problem, the decoding rules for Y can be trivially obtained and will not be reproduced here.

In the annealing process, we begin at a high temperature and track the optimum at successively lower temperatures. At high temperature, all the reproduction points are at the centroid of the source distribution, and a prototype is associated with all the indexes with equal probability. More specifically, at high temperature, minimizing the Lagrangian L implies maximizing the entropy H. This is achieved by assigning all the reproduction points to the centroid of source distribution (which results in maximum randomness and hence maximum entropy), and thus the global minimum is achieved at high temperature. As the temperature is lowered,[1] a bifurcation point is reached, where the existing solution is no longer an "attractor" solution, in the sense that small perturbation may trigger the discovery of a new solution where reproduction points are now grouped into two or more subsets. Intuitively, at this particular temperature, the original system configuration (which was a minimum at higher temperatures) becomes a saddle point. To minimize the Lagrangian cost, it is therefore beneficial to move to a newer minimum by slightly perturbing the reproduction points. We refer to this process of bifurcation as the first phase transition in analogy to statistical physics. The corresponding temperature is called "critical temperature." The subsets of reconstruction points further bifurcate at lower temperatures, and each bifurcation can be considered as a phase transition that occurs at the corresponding critical temperature. The expression for the critical temperature for the first phase transition is derived in Appendix A. This generalizes the critical temperature results for the case of a) multiple-description vector quantizer [21] and for b) single-source vector quantizer [22]. The overall DA algorithm for design can be summarized by the following steps.

1) Start at a high temperature (above critical temperature for first phase transition).
2) At a fixed temperature T, perturb the reconstruction values and minimize the free energy in (14) by iterating between the following two steps until convergence:
   a) fix the reconstruction values in (19) to compute the encoding probabilities using (17);
   b) fix the encoding probabilities and optimize the reconstruction values using (19).
3) Reduce the temperature $T \leftarrow \delta T$.
4) If $T < T_{\text{threshold}}$, stop; else goto 2).

Note that both steps 2a) and b) are monotone nonincreasing in the cost. In case T is not a critical temperature and the original state is already a minimum, the perturbed reconstruction values return to their respective original positions after step 2). Finally, at the limit of zero temperature, the algorithm will reduce to the locally optimal algorithm for RDVQ design described in Section II-B.

While the method is motivated by the ability of annealing procedures in physics/chemistry to find the global minimum (or ground state), it is not a stochastic procedure, such as "simulated annealing" [25]. The costly computation involved in simulating the random evolution of the system is replaced by minimization of an expected functional, namely, the free energy. This is, in fact, a deterministic procedure.

IV. Simulation Results

We give examples for various settings in an RDVQ system to demonstrate the gains of the deterministic annealing approach

[1]In our simulations, we used the exponential cooling schedule $T \leftarrow \delta T$, $\delta < 1$.

TABLE I
RDVQ SETTINGS FOR FIRST FOUR SIMULATION RESULTS

	Source Model	Vector Dimension	$\{\lambda_0, \lambda_1, \lambda_2\}$	$\{\alpha_0, \alpha_1, \alpha_2\}$
1.	Jointly Gaussian	1	{1,0.01,0.01}	{0.5,1,0}
2.	Jointly Gaussian	1	{1,0.005,0.01}	{0.5,0.5,0.5}
3.	Jointly Gaussian	2	{1,0,0}	{0.5,-,-}
4.	Mixture of Gaussians	1	{1,0,0}	{0.5,-,-}

over the iterative greedy method described in Section II-B. The greedy method is referred to as LA since it inherits the interrelated shortcomings of getting trapped in poor local minima and dependence on initialization, similar to the original Lloyd's algorithm [19] and its vector extension [26] for quantizer design. To avoid any potential fairness issues, we decided to design the high-rate quantizers Q_1 and Q_2 using DA for both competing approaches. This design could obviously have been done using Lloyd's algorithm for the LA contender, but we prefer to eliminate concerns regarding poor minima in the quantizer design. The focus of the paper is on Wyner–Ziv mappings optimization (and reconstruction values) given fixed high-resolution quantizers. In all the simulations, the LA algorithm was run 20 times with different initializations, while DA was run only once (DA is independent of initialization). The training set consisted of 4000 samples, while the test set had size 40 000. We first briefly enlist the important parameters in the first four RDVQ settings in Table I.

In the first three examples, X and Y are assumed to be drawn from a jointly Gaussian source with zero means, unit variances, and correlation coefficient 0.9. In the last example, the source are assumed to come from a mixture of Gaussians.

Example 1: A scalar RDVQ is designed in the first example. The distortion weighting parameters λ_1 and λ_2 for the side decoders are both set to 0.01, while λ_0 is set to one. The rates R_1 and R_2 are 3 and 4 bits, while the number of prototypes for X and Y is 64 and 128, respectively. The source weight parameters are $\alpha_0 = 0.5, \alpha_1 = 1$, and $\alpha_2 = 0$, i.e., each side decoder reconstructs its corresponding source; decoder 1 reconstructs X and decoder 2 reconstructs Y, while at the central decoder, both the sources are reconstructed with equal importance. The results depicting optimization performance on the training set are shown in Fig. 3. Here DA outperforms the best solution obtained by LA by $\sim$1.3 dB. The difference between the best and worst distortions of LA is $\sim$2.9 dB, which illustrates the fact that greedy methods are heavily dependent on initialization and are highly likely to get trapped in a local minimum. For the test set, the net distortion obtained by the best LA (by best we mean the initialization which led to the best training set data performance) versus single run DA was -15.18 and -15.95 dB (gain of 0.77 dB), respectively.

Example 2: Here, a scalar RDVQ is designed. The distortion weighting parameters λ_0, λ_1 and λ_2 are set to be 1, 0.005, and 0.01, while the rates R_1 and R_2 are 2 and 3 bits, respectively. The number of prototypes for both X and Y is 64. The source weight parameters are $\alpha_0 = \alpha_1 = \alpha_2 = 0.5$ to give equal importance to each source at all the decoders. The results are shown in Fig. 4. The net distortion obtained for the test set for best LA versus single run DA was -12.06 and -13.08 dB (gain of 1.02 dB), respectively.

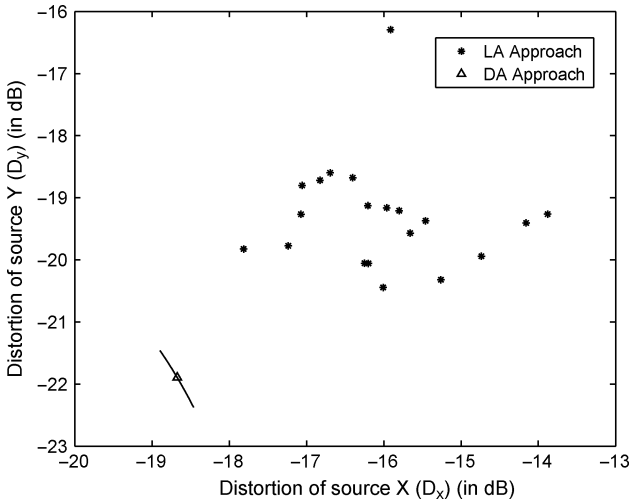

Fig. 3. Comparison between LA and DA approaches for $R_1 = 3$, $R_2 = 4$, $\mathcal{K} = 64$, $\mathcal{L} = 128$, $\alpha_0 = 0.5$, $\alpha_1 = 1$, $\alpha_2 = 0$, $\lambda_0 = 1$, $\lambda_1 = \lambda_2 = 0.01$. Net distortion from DA is -16.98 dB, while LA gives best and worst distortion as -15.69 and -12.77 dB, respectively. For ease of comparison, a line along which constant $D_{\text{net}} = -16.98$ dB is drawn.

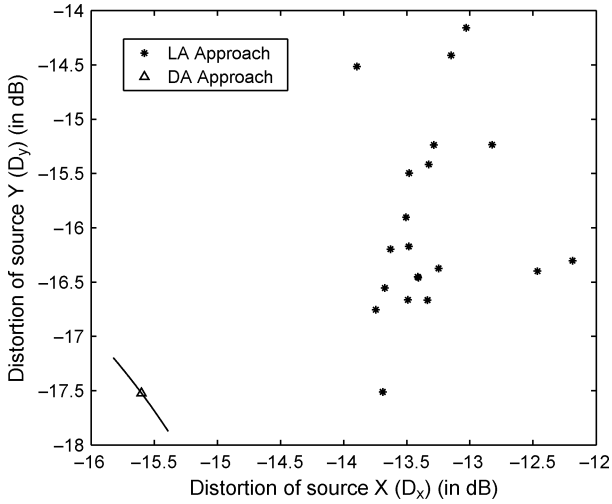

Fig. 4. Comparison between LA and DA approaches for $R_1 = 2$, $R_2 = 3$, $\mathcal{K} = \mathcal{L} = 64$, $\alpha_0 = \alpha_1 = \alpha_2 = 0.5$, $\lambda_0 = 1$, $\lambda_1 = 0.005$, $\lambda_2 = 0.01$. Net distortion from DA is -13.44 dB, while LA gives best and worst distortion as -12.18 and -10.54 dB, respectively. For ease of comparison, a line along which constant $D_{\text{net}} = -13.44$ dB is drawn.

Example 3: A distributed quantizer of dimension two is designed in this example (i.e., $\lambda_0 = 1$ and $\lambda_1 = \lambda_2 = 0$, implying that both the channels function and only the central decoder is used at the receiver). Both the sources are transmitted at rate of 2 bits and given equal importance (i.e., $\alpha_0 = 0.5$). The simulation result is given in Fig. 5. The distortion achieved by DA and best run LA approach are -12.75 and -10.85 dB, respectively.

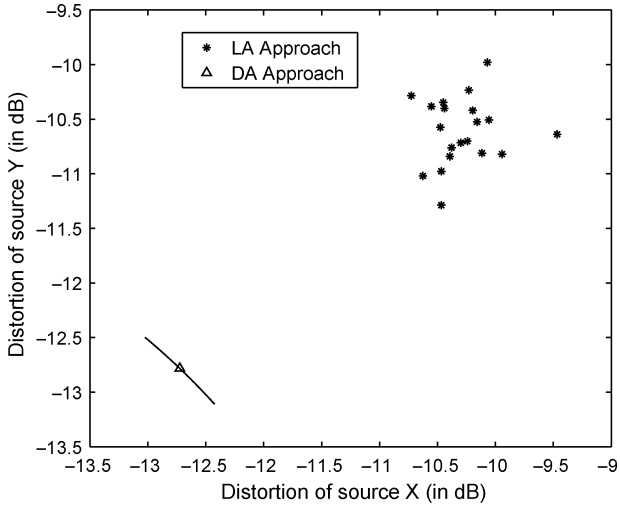

Fig. 5. Comparison between LA and DA approaches for a distributed vector quantizer of dimension two. $R_1 = R_2 = 2$ bps, $\mathcal{K} = \mathcal{L} = 128$, $\alpha_0 = 0.5$, $\lambda_0 = 1$, $\lambda_1 = \lambda_2 = 0$. Net distortion from DA is -12.75 dB, while LA gives best and worst distortion as -10.85 and -10.01 dB, respectively. For ease of comparison, a line along which constant $D_{\text{net}} = -12.75$ dB is drawn. Achievable distortion as promised in [27] is -15.61 dB.

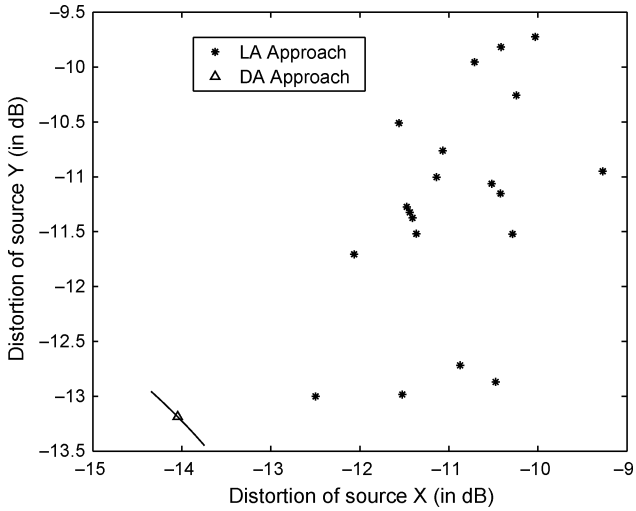

Fig. 6. Comparison between LA and DA approaches for a distributed vector quantizer for sources coming from a Gaussian mixture model. $R_1 = R_2 = 3$ bps, $\mathcal{K} = \mathcal{L} = 64$, $\alpha_0 = 0.5$, $\lambda_0 = 1$, $\lambda_1 = \lambda_2 = 0$. Net distortion from DA is -13.59 dB, while LA gives best and worst distortion as -12.74 and -9.87 dB, respectively. For ease of comparison, a line along which constant $D_{\text{net}} = -13.59$ dB is drawn.

The theoretically achievable (asymptotic) distortion at the corresponding rates and correlation coefficients as promised in [27] is -15.61 dB.[2] Here the DA approach is roughly 2.86 dB away from the asymptotic bound of the distortion, and the greedy LA approach is a further 1.9 dB away. Note that the distortion from the LA and DA approaches can be further reduced if entropy coding is employed or the dimension of the quantizers is increased.

Example 4: In the next example (see Fig. 6), X and Y are drawn from a mixture of four joint Gaussians. Such a situation can arise, for example, when sources correspond to the temperature and humidity readings and the different mixture components are due to varying underlying conditions such as the time of day, pressure, etc. Here $\lambda_1 = \lambda_2 = 0$, $\alpha_0 = 0.5$, and the source rates are 3 bits/sample. In our simulations, the mixtures components are assumed to be equiprobable. The means for X, Y and correlation coefficients for the four components are taken as $\{0, 0, 0.87\}$, $\{1, 0.5, 0.9\}$, $\{-1, 1, -0.92\}$, and $\{2, -1, -0.95\}$ respectively. The variance of X and Y in all the components of the mixture was taken to be one. The distortion values achieved by DA and from the best and worst LA algorithm are -13.59, -12.74, and -9.87 dB (DA gains 0.85 and 3.72 dB over best and worst LA), respectively.

The next simulation result (see Fig. 7) depicts the variation in weighted distortion for the LA (best of 20 runs) and DA approaches for a scalar RDVQ system with the number of prototypes for the sources. Here $\lambda_0 = 1$, $\lambda_1 = \lambda_2 = 0.01$, and $\alpha_0 = \alpha_1 = \alpha_2 = 0.5$ and the source rates R_1 and R_2 are kept fixed at 3 bits. As the number of prototypes is increased, the WZ mappings can possibly combine more noncontiguous regions together and utilize the intersource correlation more efficiently.

[2]To calculate the distortion bounds from [27], we have assumed that the individual source distortions will be approximately the same and hence equal to the average distortion, since both sources have similar statistics, are encoded at the same rate, and are given equal importance at the decoder.

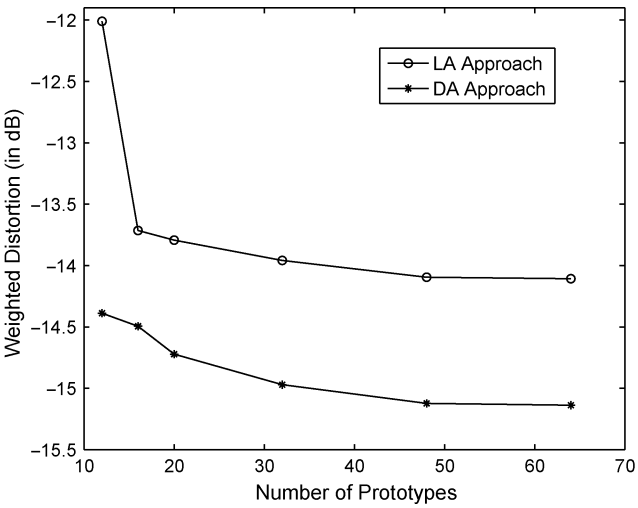

Fig. 7. Comparison between LA and DA approaches when the number of source prototypes is varied for $R_1 = R_2 = 3$ bps, $\alpha_0 = \alpha_1 = \alpha_2 = 0.5$; $\lambda_0 = 1, \lambda_1 = \lambda_2 = 0.01$.

Note that even for a large number of prototypes, the greedy LA approach underperforms the DA approach, justifying the use of a global optimization tool for a robust distributed quantizer design. Also, after a point, increasing the number of prototypes does not lead to reduction in the distortion cost. This implies that only a sufficiently large number of prototypes (in comparison to the transmitted indexes) is required for achieving a good system performance. Further, note both the greedy LA and deterministic annealing-based methods are extendible to incorporate entropy coding, but such extension is omitted for brevity.

Finally, a note on system complexity. The design complexity of DA-based algorithm is higher than that of the LA approach. In our simulations, the DA approach took on average 20–25 times longer than for a single run of LA approach. The run time of the DA algorithm can be further reduced by simple schemes outlined in [22], but this is tangential to the work presented in this paper. For

completeness, we just outline a simple procedure to accelerate the DA algorithm. In DA, almost all the interesting activity happens near the phase transitions, when the codevectors split and move to different locations to minimize the cost. In between the phase transitions, the codevectors remain at the same locations, and the changes in distortion cost are insignificant. Thus, the cooling between phase transitions can be done in a rapid fashion without actually compromising the algorithm performance. We have not pursued the above idea for accelerating the DA approach in between the phase transitions in this paper and have used the simple exponential cooling schedule for DA. Further, instead of starting from a high temperature, the DA algorithm can be initialized from a temperature slightly above the critical temperature for first phase transition, since above this temperature, there is only one global minimum on the cost surface (see the result for critical temperature for first phase transition in the Appendix). Note that the design complexity of DA is a one-time cost only. During operation, *hard* quantizers are used, and both the DA and LA approaches have the same operational complexity.

V. Conclusions

We have proposed a multiple prototype-based deterministic annealing approach for the design of quantizers for a robust distributed source coding system. The approach is general and is applicable to a wide gamut of coding and quantization problems such as multiple descriptions, distributed source coding, CEO problem, etc. This approach assumes no prior knowledge about the underlying probability distribution of the sources, eliminates the dependence on good ad-hoc initial configurations, and avoids many poor local minima of the distortion cost surface. The necessary conditions (and update equations) for system design are derived and presented. Simulation results comparing DA with an iterative Lloyd-like algorithm are shown. Significant improvements confirm the advantage of using a global optimization scheme such as DA for robust distributed vector quantizer design.

Appendix

Recall that DA finds the trivial global optimum at "high temperature," where all the reproduction points coincide at the center of mass of the source distribution. The first "phase transition" corresponds to the bifurcation of the reproduction points into subsets. The temperatures at which various phase transitions occur are called the critical temperatures. Here we derive the expression for the critical temperature corresponding to the first phase transition for RDVQ. The result will be a generalization of the critical temperature for special cases such as multiple-description vector quantizer, single source vector quantizer, etc.

Without loss of generality, we assume that the phase transition occurs for code vectors corresponding to index i (representing source X), and the number of code vectors increases from one to two. (There can be a phase transition to more than two code vectors, but the necessary condition for bifurcation can be obtained by assuming that the number of code vectors increases to only two.) At high temperature (greater than the critical temperature for the first phase transition), all the association probabilities are equal (uniform) and the code vectors for both sources will be located at their respective centroids.

The expression of the Lagrangian cost in (14) that needs to be minimized is

$$L = D - TH \tag{20}$$

$$= \frac{1}{N}\sum_{x,y\in\mathcal{T}}\left[\sum_{k,l,i,j} c_{k|x}c_{l|y}r_{i|k}r_{j|l}D_{\text{net}}(x,y,i,j) + T\left\{\sum_{k,i} c_{k|x}r_{i|k}log(r_{i|k}) + \sum_{l,j} c_{l|y}r_{j|l}log(r_{j|y})\right\}\right] \tag{21}$$

where $D_{\text{net}}(x,y,i,j)$ is given in (2) and the last two terms are for the source entropies $H(K,I|X)$ and $H(L,J|Y)$, respectively, defined in (15) and (16). Since we are assuming only one possible value for index j (phase transition occurs for code vectors corresponding to index i), the second entropy term is zero. Also $\sum_l c_{l|y} = 1$ from (11) since a training set point for Y will map only to one out of $\mathcal{L}$ possible prototypes. Hence the above expression reduces to

$$L = \frac{1}{N}\sum_{x,y\in\mathcal{T}}\left[\sum_{k,l,i} c_{k|x}c_{l|y}r_{i|k}D_{\text{net}}(x,y,i,j) + T\left\{\sum_{k,i} c_{k|x}r_{i|k}log(r_{i|k})\right\}\right] \tag{22}$$

$$= \frac{1}{N}\sum_{x,y\in\mathcal{T}}\left[\sum_{k,i} c_{k|x}r_{i|k}D_{\text{net}}(x,y,i,j) + T\left\{\sum_{k,i} c_{k|x}r_{i|k}log(r_{i|k})\right\}\right]. \tag{23}$$

Next we make a simplifying assumption that the number of prototypes (output of the high-rate quantizer Q_1, see Fig. 2) is large, and there are as many prototypes as the number of data points. Hence there is one-to-one correspondence between a data point and a prototype and

$$r_{i|k} = \Pr\left[x_k \in R_i^x\right] \approx \Pr\left[x \in R_i^x\right] = p_{i|x}. \tag{24}$$

Using (24) and $\sum_k c_{k|x} = 1$ from (11), the expression of free energy in (23) can be rewritten as

$$L = \frac{1}{N}\sum_{x,y\in\mathcal{T}}\left[\sum_i p_{i|x}D_{\text{net}}(x,y,i,j) + T\left\{\sum_i p_{i|x}log(p_{i|x})\right\}\right]. \tag{25}$$

We assume squared-error distortion measure for further analysis. We further write D_{net} as D_i^{pt} to explicitly indicate that the distortion at first phase transition (pt) is only affected by index i (the only possible value of j is one). The expression for D_i^{pt} can be simplified as

$$D_i^{pt} = \lambda_0\left\{\alpha_0\left(x-\hat{x}_{ij=1}^0\right)^2 + (1-\alpha_0)\left(y-\hat{y}_{ij=1}^0\right)^2\right\} + \lambda_1\left\{\alpha_1\left(x-\hat{x}_i^1\right)^2 + (1-\alpha_1)\left(y-\hat{y}_i^1\right)^2\right\} + \lambda_2\left\{\alpha_2\left(x-\hat{x}_{j=1}^2\right)^2 + (1-\alpha_2)\left(y-\hat{y}_{j=1}^2\right)^2\right\}. \tag{26}$$

The reconstruction values for central and side decoder 1 are the same at high temperature at the source centroid, i.e., $\hat{x}_{ij}^0 =$

$$H_L = \begin{bmatrix} \beta_1\left(I-\frac{1}{T}\beta_1 C_{xx}\right) & \frac{(\beta_1)^2}{T}C_{xx} & -\frac{\beta_1\beta_2}{T}C_{xy} & \frac{\beta_1\beta_2}{T}C_{xy} \\ \frac{(\beta_1)^2}{T}C_{xx} & \beta_1\left(I-\frac{1}{T}\beta_1 C_{xx}\right) & \frac{\beta_1\beta_2}{T}C_{xy} & -\frac{\beta_1\beta_2}{T}C_{xy} \\ -\frac{\beta_1\beta_2}{T}C_{xy}^t & \frac{\beta_1\beta_2}{T}C_{xy}^t & \beta_2\left(I-\frac{1}{T}\beta_2 C_{yy}\right) & \frac{(\beta_2)^2}{T}C_{yy} \\ \frac{\beta_1\beta_2}{T}C_{xy}^t & -\frac{\beta_1\beta_2}{T}C_{xy}^t & \frac{(\beta_2)^2}{T}C_{yy} & \beta_2\left(I-\frac{1}{T}\beta_2 C_{yy}\right) \end{bmatrix} \tag{30}$$

$\hat{x}_i^1$ (since j takes only one value) and similarly for Y. Using this and combining terms, D_i^{pt} reduces to

$$\begin{aligned} D_i^{pt} = &(\lambda_0\alpha_0+\lambda_1\alpha_1)\left(x-\hat{x}_i^1\right)^2 \\ &+\{\lambda_0(1-\alpha_0)+\lambda_1(1-\alpha_1)\}\left(y-\hat{y}_i^1\right)^2 \\ &+\lambda_2\alpha_2\left(x-\hat{x}_{j=1}^2\right)^2+\lambda_2(1-\alpha_2)\left(y-\hat{y}_{j=1}^2\right)^2. \end{aligned} \tag{27}$$

We define the covariance matrices for the source data as follows:

$$\begin{aligned} C_{xx} &= \frac{1}{N}\sum_{x,y\in\mathcal{T}}(x-\mu_x)(x-\mu_x)^t \\ C_{xy} &= \frac{1}{N}\sum_{x,y\in\mathcal{T}}(x-\mu_x)(y-\mu_y)^t \\ C_{yy} &= \frac{1}{N}\sum_{x,y\in\mathcal{T}}(y-\mu_y)(y-\mu_y)^t \end{aligned} \tag{28}$$

where μ_x and μ_y are the respective source means. For notational convenience, we define $\beta_1 = \lambda_0\alpha_0 + \lambda_1\alpha_1$ and $\beta_2 = \lambda_0(1-\alpha_0) + \lambda_1(1-\alpha_1)$.

At the phase transition, the code vectors $\hat{x}_{i=1}^1$ and $\hat{x}_{i=2}^1$ for X (and similarly for Y) will separate and move to new respective different locations. At the critical temperature for phase transition, the system solution changes from a minimum to a saddle point. Equivalently, the Hessian matrix of the free energy (L) with respect to the code vectors ($\hat{x}_{i=1}^1, \hat{x}_{i=2}^1, \hat{y}_{i=1}^1$, and $\hat{y}_{i=2}^1$) will no longer be positive definite, and its determinant will vanish.

For the calculation for the Hessian matrix, we first compute the association probabilities $p_{i|x}$ from (8), (9), and (17) or by directly minimizing the free energy L with respect to $p_{i|x}$. The association probability $p_{i|x}$ is given by

$$p_{i|x} = \frac{e^{-D_i^{pt}/T}}{\sum_{i'} e^{-D_{i'}^{pt}/T}} \tag{29}$$

and can be substituted in (23). It can be shown by straightforward derivation that the Hessian matrix is given by (30) as shown at the top of the page, where I is the identity matrix and superscript t denotes matrix transposition.

Setting the Hessian matrix determinant to zero yields

$$\det\left[\left(I-\frac{2}{T}\beta_1 C_{xx}\right)\left(I-\frac{2}{T}\beta_2 C_{yy}\right)-\frac{4\beta_1\beta_2}{T^2}C_{xy}C_{xy}^t\right]=0. \tag{31}$$

The above equation is implicit in the critical temperature T. We next obtain and interpret an explicit solution for special cases of RDVQ.

1) *Single Source Vector Quantizer (say for X):* Here, only one channel will be present, i.e., $\lambda_1 = 1$; $\lambda_0 = \lambda_2 = 0$ and only source X will be of interest, i.e., $\alpha_1 = 1$. Therefore, we have $\beta_1 = 1$ and $\beta_2 = 0$, and (31) reduces to $\det[I - (2/T)C_{xx}] = 0$. This implies that the critical temperature for the first phase transition will be at $T = 2\gamma_x$, where γ_x is the largest eigenvalue of C_{xx}, and matches the basic DA result in [22].
2) *Multiple Descriptions Vector Quantizer:* Here the two sources are identical, i.e., $Y = X$. The expression in (31) reduces to $\det[I - (2/T)(\beta_1 + \beta_2)C_{xx}]$. Also $\beta_1 + \beta_2 = \lambda_0 + \lambda_1$. The critical temperature for the first phase transition will be $2\gamma_x(\lambda_0 + \lambda_1)$, which was also derived in [21].
3) *Jointly Gaussian Scalar Sources:* For zero-mean sources X and Y with respective variances σ_x^2 and σ_y^2 and correlation coefficient ρ, the condition in (31) reduces to

$$\left(1-\frac{2\beta_1}{T}\sigma_x^2\right)\left(1-\frac{2\beta_2}{T}\sigma_y^2\right)-4\frac{\beta_1\beta_2}{T^2}\rho^2\sigma_x^2\sigma_y^2=0. \tag{32}$$

The expression for T_{crit} can be found by solving the above equation. If $\sigma_x^2 = \sigma_y^2 = \sigma^2$ and both sources are given equal importance during reconstruction ($\beta_1 = \beta_2$), we have $T_{\text{crit}} = (\lambda_0 + \lambda_1)\sigma^2(1 + |\rho|)$. When the sources are perfectly correlated ($\rho = 1$), this reduces to the multiple description case for scalar sources as expected ($Y = X$). For the case when $\rho = 0$ (uncorrelated sources), T_{crit} reduces to $\{\lambda_0 + \lambda_1\}\sigma^2$. This can be interpreted as follows: when X and Y are perfectly correlated ($\rho = 1$), the sources are spread along one direction only (in the $X - Y$ plane). On the other hand, as ρ decreases from one to zero, the sources (in $X - Y$ space) are spread in an isotropic fashion along all the directions. Thus, there is more symmetry in the system, and it will take longer for the codevectors to split as we lower the temperature during the annealing process. Therefore, the critical temperature decreases to a lower value as ρ decreases (analysis for negative values of ρ is similar).

REFERENCES

[1] P. Ishwar, R. Puri, S. S. Pradhan, and K. Ramchandran, "On compression for robust estimation in sensor networks," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun.–Jul. 2003, p. 193.

[2] J. Chen and T. Berger, "Robust coding schemes for distributed sensor networks with unreliable sensors," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun.–Jul. 2004, p. 115.

[3] J. Chen and T. Berger, "Robust distributed source coding," *IEEE Trans. Inf. Theory*, vol. 54, pp. 3385–3398, Aug. 2008.

[4] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, pp. 471–480, Jul. 1973.

[5] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side-information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, pp. 1–10, Jan. 1976.

[6] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem [multiterminal source coding]," *IEEE Trans. Inf. Theory*, vol. 42, pp. 887–902, May 1996.

[7] A. V. Rao, D. J. Miller, K. Rose, and A. Gersho, "A generalized VQ method for combined compression and estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 1996, vol. 4, pp. 2032–2035.
[8] A. Gersho, "Optimal nonlinear interpolative vector quantization," *IEEE Trans. Commun.*, vol. 38, pp. 1285–1287, Sep. 1990.
[9] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Inf. Theory*, vol. 49, pp. 626–643, Mar. 2003.
[10] X. Wang and M. T. Orchard, "Design of trellis codes for source coding with side information at the decoder," in *Proc. IEEE Data Compression Conf.*, Mar. 2001, pp. 361–370.
[11] P. Mitran and J. Bajcsy, "Coding for the Wyner-Ziv problem with turbo-like codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2002, p. 91.
[12] J. Garcia-Frias and Y. Zhao, "Compression of correlated binary sources using turbo codes," *IEEE Commun. Lett.*, vol. 5, pp. 417–419, Oct. 2001.
[13] A. D. Liveris, Z. Xiong, and C. Eorghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Commun. Lett.*, vol. 6, pp. 440–442, Oct. 2002.
[14] Y. Yang, V. Stankovic, Z. Xiong, and W. Zhao, "On multiterminal source code design," in *Proc. IEEE Data Compression Conf.*, Mar. 2005, pp. 43–52.
[15] Y. Yang, V. Stankovic, Z. Xiong, and W. Zhao, "On multiterminal source code design," *IEEE Trans. Inf. Theory*, vol. 54, pp. 2278–2302, May 2008.
[16] M. Fleming, Q. Zhao, and M. Effros, "Network vector quantization," *IEEE Trans. Inf. Theory*, vol. 50, pp. 1584–1604, Aug. 2004.
[17] J. Cardinal and G. V. Assche, "Joint entropy-constrained multiterminal quantization," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2002, p. 63.
[18] D. Rebollo-Monedero, R. Zhang, and B. Girod, "Design of optimal quantizers for distributed source coding," in *Proc. IEEE Data Compression Conf.*, Mar. 2003, pp. 13–22.
[19] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, pp. 129–137, Mar. 1982.
[20] V. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inf. Theory*, vol. 39, pp. 821–834, May 1993.
[21] P. Koulgi, S. L. Regunathan, and K. Rose, "Multiple descriptions quantization by deterministic annealing," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2067–2075, Aug. 2003.
[22] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems," *Proc. IEEE*, vol. 86, pp. 2210–2239, Nov. 1998.
[23] A. Saxena, J. Nayak, and K. Rose, "On efficient quantizer design for robust distributed source coding," in *Proc. IEEE Data Compression Conf.*, Mar. 2006, pp. 63–72.
[24] A. Saxena, J. Nayak, and K. Rose, "A global approach to joint quantizer design for distributed coding of correlated sources," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2006, vol. 2, pp. 53–56.
[25] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, May 1983.
[26] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. COM-28, pp. 84–95, Jan. 1980.
[27] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic gaussian two-encoder source-coding problem," *IEEE Trans. Inf. Theory*, vol. 54, pp. 1938–1961, May 2008.

Ankur Saxena (S'06–M'09) was born in Kanpur, India, in 1981. He received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Delhi, in 2003 and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of California, Santa Barbara, in 2004 and 2008, respectively.

He was an Intern with Fraunhofer Institute of X-Ray Technology, Erlangen, Germany, and NTT DoCoMo Research Labs, Palo Alto, CA, in the summers of 2002 and 2007, respectively. He is currently a Postdoctoral Researcher with the University of California, Santa Barbara. His research interests span source coding, image, and video compression and signal processing.

Dr. Saxena received the President Work Study Award during his Ph.D.

Jayanth Nayak received the Ph.D. degree in electrical and computer engineering from the University of California, Santa Barbara, in 2005.

He is currently a Research Staff Member with Mayachitra, Inc., Santa Barbara, CA. His research interests include information theory, pattern recognition, and computer vision.

Kenneth Rose (S'85–M'91–SM'01–F'03) received the Ph.D. degree from the California Institute of Technology, Pasadena, in 1991.

He joined the Department of Electrical and Computer Engineering, University of California, Santa Barbara, where he is currently a Professor. His main research activities are in the areas of information theory and signal processing, and include rate-distortion theory, source and source-channel coding, audio and video coding and networking, pattern recognition, and nonconvex optimization. He is interested in the relations among information theory, estimation theory, and statistical physics and their potential impact on fundamental and practical problems in diverse disciplines.

Prof. Rose was a corecipient of the 1990 William R. Bennett Prize Paper Award from the IEEE Communications Society and the 2004 and 2007 IEEE Signal Processing Society Best Paper Award.