Predictive Vector Quantizer Design Using Deterministic Annealing

Hosam Khalil, *Member, IEEE,* and Kenneth Rose, *Fellow, IEEE*

***Abstract*—A new approach is proposed for predictive vector quantizer (PVQ) design, which is inherently probabilistic, and is based on ideas from information theory and analogies to statistical physics. The approach effectively resolves three longstanding fundamental shortcomings of standard PVQ design. The first complication is due to the PVQ prediction loop, which has a detrimental impact on the convergence and the stability of the design procedure. The second shortcoming is due to the piecewise constant nature of the quantizer function, which makes it difficult to optimize the predictor with respect to the overall reconstruction error. Finally, a shortcoming inherited from standard VQ design is the tendency of the design algorithm to terminate at a locally, rather than the globally, optimal solution. We propose a new PVQ design approach that embeds our recent asymptotic closed-loop (ACL) approach within a deterministic annealing (DA) framework. The overall DA-ACL method profits from its two main components in a complementary way. ACL is used to overcome the first difficulty and offers the means for stable quantizer design as it provides an open-loop design platform, yet allows the PVQ design algorithm to asymptotically converge to optimization of the closed-loop performance objective. DA simultaneously mitigates or eliminates the remaining design shortcomings. Its probabilistic framework replaces hard quantization with a differentiable expected cost function that can be jointly optimized for the predictor and quantizer parameters, and its annealing schedule allows the avoidance of many poor local optima. Substantial performance gains over traditional methods have been achieved in the simulations.**

***Index Terms*—Closed-loop design, deterministic annealing, open-loop design, predictive vector quantizer design.**

I. Introduction

A VECTOR quantizer (VQ) is a structure that implements a many-to-one (lossy) mapping of data from one domain to another. A main use of VQ has been in the compression of data for storage or transmission over communication channels. VQ has successfully found its way into several speech coding standards [1]–[3] and has also found application in image and video compression [4]–[7]. Such sources usually exhibit considerable temporal dependencies. Quantizers with memory offer a natural means to remove temporal redundancies and achieve better efficiency. Examples for such VQ structures include finite-state VQ, classified VQ, and predictive VQ. The latter is the quantizer structure of interest in this paper. While the complexity of predictive VQ may not be significantly higher than that of conventional memoryless VQ, the performance gains are often considerable.

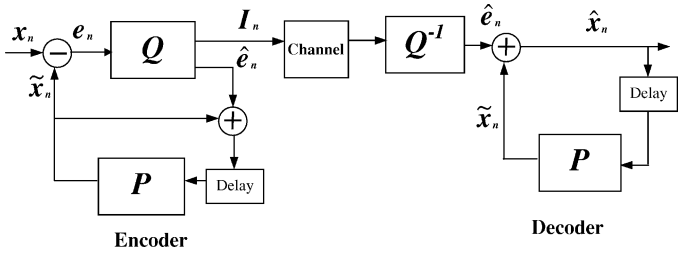


Fig. 1. Basic predictive vector quantizer system.

A predictive vector quantizer (PVQ) is basically a predictor in tandem with a memoryless vector quantizer. The system, hence, predicts an incoming vector and then quantizes the prediction error. The quantization is performed by finding in the codebook the best codevector with respect to a predefined distortion measure. Only the codevector index needs to be transmitted, as the decoder has an exact copy of the codebook. In order for the decoder to reconstruct the vectors correctly, it must be able to duplicate the encoder prediction. Therefore, the prediction must only depend on past reconstructed vectors that are known to both encoder and decoder. In other words, PVQ is a feedback system.

The PVQ structure is depicted in Fig. 1. Although the PVQ structure is simple and well-understood, its design is problematic, and standard methods often fail to produce optimal (or even good) predictors and quantizers. The feedback loop creates a complex relationship between predictor and quantizer. To design the quantizer, a representative training set of prediction errors is needed, but to obtain the prediction errors, we must run the system in closed-loop, which implies dependence on both the predictor and the quantizer to be designed. Further, there are open questions concerning the design of the predictor. Clearly, the choice of the best predictor is not independent of the quantizer in use. In addition, due to the piecewise constant nature of the quantizer function, it is difficult to optimize the predictor with respect to the reconstruction error.

Two simple approaches were introduced by Cuperman and Gersho [8]. The first approach is called "open-loop" design. It solves the quantizer design problem by assuming no feedback, and operating directly on original source vectors. An improvement can be achieved using their second approach called "closed-loop" design. In this case, an iterative design is used for

Manuscript received July 17, 2001; revised August 21, 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 associate editor coordinating the review of this paper and approving it for publication was Prof. Sheila S. Hemami.

H. Khalil is with the Digital Media Division, Microsoft Corporation, Redmond, WA USA.

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

Digital Object Identifier 10.1109/TSP.2002.806582

updating the training set and quantizer given a fixed predictor. An alternative (closed loop) design algorithm was presented by Chang and Gray [9], where both predictor and quantizer are jointly optimized, but in general, such design approaches exhibit significant stability problems especially at low bit rates. We recently proposed an approach in [10]–[12] called the asymptotic closed-loop (ACL) algorithm, which solves the stability problems over all bit rates. The method is open-loop in nature and thus inherently stable. However, the designed system asymptotically approaches closed-loop operation and effectively a feedback system is designed to produce a realistic PVQ system. A brief review of the traditional approaches and their strengths and weaknesses will be presented in Section II.

In this paper, we propose a PVQ design method that significantly outperforms all of the above methods. The new approach (DA-ACL) embeds the ACL method within a deterministic annealing (DA) framework. The overall DA-ACL method benefits from its two main components in a complementary way. ACL offers the means for stable quantizer design as it is performed in open-loop but asymptotically converges to the closed-loop system. DA offers two benefits: Its probabilistic framework replaces hard quantization with a differentiable cost function that can be jointly optimized for the predictor and quantizer parameters; and its annealing schedule helps in avoiding many poor local optima.

In contrast with traditional methods, the DA-ACL algorithm we propose does not require initialization of codebooks or prediction parameters as its outcome is independent of initialization. Various DA-based methods have been successfully applied in several applications including pattern recognition and signal compression [13], [14]. To implement DA within a procedure for PVQ design, we randomize the encoding rule of the predictive quantizer system and seek to minimize the expected distortion cost subject to a specified level of randomness measured by the Shannon entropy. The degree of randomness is parameterized by the "temperature" of the configuration (The temperature is the physical analogue of the Lagrange multiplier relating distortion/energy and entropy). The temperature is gradually lowered, and the system is reoptimized at each temperature. This is an annealing process consisting of maintaining the system at isothermal equilibrium while gradually reducing the temperature. At the limit of zero randomness, the algorithm directly minimizes the average distortion cost, and a deterministic encoder is obtained.

The paper is organized as follows. In Section II, we state the PVQ problem and outline known design algorithms as well as ACL. In Section III, we derive the new DA-ACL design algorithm. In Sections IV and V, we present simulation results and conclusions, respectively.

II. Problem Formulation and Known Approaches

In this section, we briefly state the problem, review existing techniques, and point out their shortcomings as motivation for the proposed approaches.

A. Problem

A typical PVQ system is shown in Fig. 1. Let $\boldsymbol{X} = \{\boldsymbol{x}_n\}_{n=0}^{N}$ be a vector-valued source over the k-dimensional Euclidean space $\boldsymbol{R}^k$. The predictor approximates the next input vector $\boldsymbol{x}_{n+1}$ given past reconstructed vectors as $\tilde{\boldsymbol{x}}_{n+1} = \boldsymbol{f}(\hat{\boldsymbol{x}}_n, \hat{\boldsymbol{x}}_{n-1}, \ldots)$. A first-order linear vector predictor is used for simplicity: $\tilde{\boldsymbol{x}}_{n+1} = \boldsymbol{A}\hat{\boldsymbol{x}}_n$, where $\boldsymbol{A}$ is a $k \times k$ matrix. The encoder $\boldsymbol{\gamma}$ assigns to the error vector $\boldsymbol{e}_n = \boldsymbol{x}_n - \tilde{\boldsymbol{x}}_n$ an index $i_n = \boldsymbol{\gamma}(\boldsymbol{e}_n)$. The decoder $\boldsymbol{\beta}$ assigns to index i_n a reproduction value $\hat{\boldsymbol{e}}_n = \boldsymbol{\beta}(i_n)$. Given a sequence of input vectors, the index sequence i_n, reconstruction sequence $\hat{\boldsymbol{x}}_n$, and corresponding prediction sequence $\tilde{\boldsymbol{x}}_n$ are defined recursively in n as

$$
\begin{aligned}
i_n &= \boldsymbol{\gamma}(\boldsymbol{e}_n) = \boldsymbol{\gamma}(\boldsymbol{x}_n - \tilde{\boldsymbol{x}}_n) \\
\hat{\boldsymbol{x}}_n &= \tilde{\boldsymbol{x}}_n + \boldsymbol{\beta}(i_n) \\
\tilde{\boldsymbol{x}}_{n+1} &= \boldsymbol{A}\hat{\boldsymbol{x}}_n.
\end{aligned}
$$

Although $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are important for system implementation, it is convenient for compression performance analysis to make abstraction of the index and define instead the quantizer function $\boldsymbol{Q}(\cdot)$, which takes in the prediction error and produces its reconstructed value: $\hat{\boldsymbol{e}}_n = \boldsymbol{Q}(\boldsymbol{e}_n) = \boldsymbol{\beta}(\boldsymbol{\gamma}(\boldsymbol{e}_n))$.

The objective of a PVQ design algorithm is to obtain a predictor and quantizer which are matched to the source to be compressed, i.e., which minimize the reconstruction distortion.

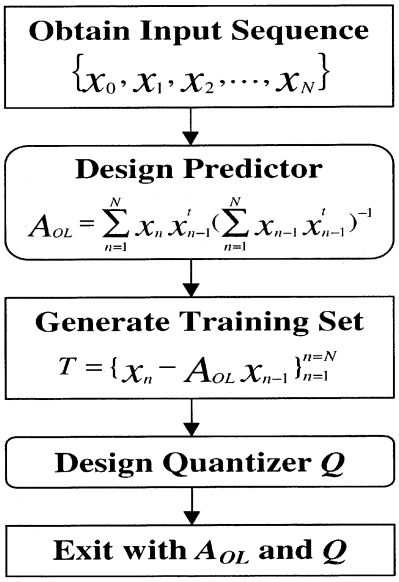


Fig. 2. Flow diagram for an implementation of the OL training procedure.

B. Open- and Closed-Loop Approaches

A flow diagram for the open-loop (OL) method of [8] is shown in Fig. 2. Both predictor and quantizer are designed based on the original *unquantized* source vectors. The auto-regressive predictor is obtained from the autocorrelation of the input source (^{t} denotes transposition):

$$
\boldsymbol{A}_{OL} = \boldsymbol{R}_1 \boldsymbol{R}_0^{-1} = \left(\sum_{n=1}^{N} \boldsymbol{x}_n \boldsymbol{x}_{n-1}^t \right) \left(\sum_{n=1}^{N} \boldsymbol{x}_{n-1} \boldsymbol{x}_{n-1}^t \right)^{-1}. \tag{1}
$$

A training set of prediction error vectors $\boldsymbol{T} = \{\boldsymbol{e}_n\}_{n=1}^{N}$ is generated from the *original* source vectors: $\boldsymbol{e}_n = \boldsymbol{x}_n - \tilde{\boldsymbol{x}}_n =$

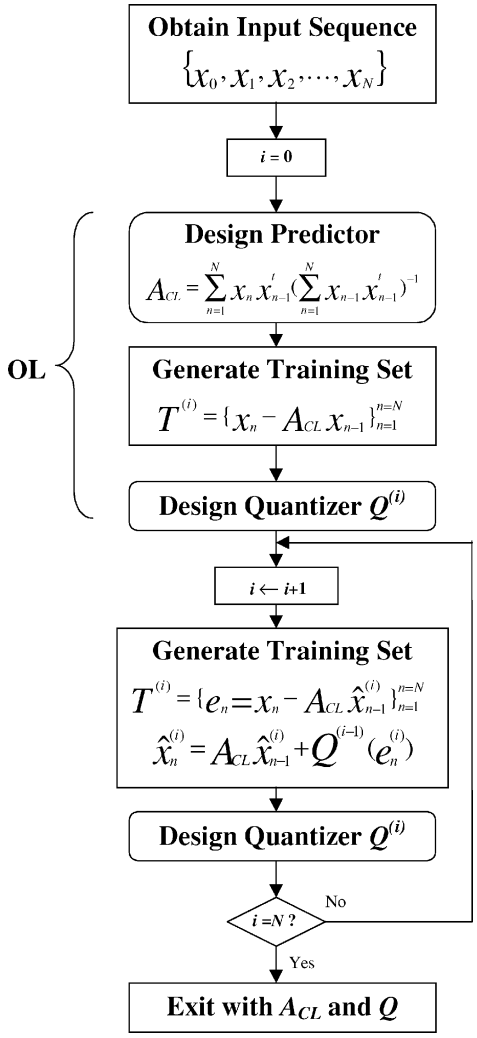


Fig. 3. Flow diagram for an implementation of the CL training procedure.

$\boldsymbol{x}_n - \boldsymbol{A}_{OL}\boldsymbol{x}_{n-1}$. The design of $\boldsymbol{Q}_{OL}$ consists of a straightforward VQ design procedure given the training set $\boldsymbol{T}$.

In the closed-loop (CL) approach of [8] (shown in Fig. 3), a closed-loop (real) system is used to generate the prediction errors for the *quantizer design* in an iterative fashion. The predictor, however, is designed using the open-loop method, i.e., $\boldsymbol{A}_{CL} = \boldsymbol{A}_{OL}$. Given a quantizer at iteration $i-1$, which we denote by $\boldsymbol{Q}^{(i-1)}$, a training set of prediction errors is generated for iteration i, $\boldsymbol{T}^{(i)} = \{\boldsymbol{e}_n^{(i)}\}_{n=1}^{N}$, where $\boldsymbol{e}_n^{(i)} = \boldsymbol{x}_n - \boldsymbol{A}_{CL}\hat{\boldsymbol{x}}_{n-1}^{(i)}$, and $\hat{\boldsymbol{x}}_n^{(i)} = \boldsymbol{A}_{CL}\hat{\boldsymbol{x}}_{n-1}^{(i)} + \boldsymbol{Q}^{(i-1)}(\boldsymbol{e}_n^{(i)})$. We use OL as the initialization: $\boldsymbol{Q}^{(0)} = \boldsymbol{Q}_{OL}$.

The main advantage of OL over CL is that its training set $\boldsymbol{T}$ is fixed. Therefore, we can design the PVQ by applying a standard optimization technique such as GLA [15]. Since the training set remains unchanged, the design algorithm is expected to converge to a locally optimal solution. However, the decoder does not have access to the original source vector for prediction. Therefore, during the actual operation of the compression system, prediction must be performed using *reconstructed* source vectors. Thus, the training set of prediction errors is statistically different from the prediction errors the quantizer will eventually encounter. This statistical mismatch, which is exacerbated by feedback through the prediction loop, results in poor performance. In the case of CL, since the training residuals were generated by a closed-loop coder, their statistics may be expected to be similar to those that will be observed in practice. However, convergence of the algorithm is not guaranteed, as the training set changes from iteration to iteration in an unpredictable fashion, and the predictor is not re-optimized for a new quantizer. In fact, the CL system sometimes exhibits catastrophically unstable behavior.

C. Steepest Descent (SD) Approach

The previous two methods implicitly assume that the bit rate is sufficiently high so that open-loop design of the predictor is adequate. Two gradient algorithms for designing predictive vector quantizers were developed by Chang and Gray [9]: the steepest descent algorithm and the stochastic gradient algorithm. Both of these methods, which are based on known adaptive filtering techniques, attempt to jointly optimize quantizer and predictor. We briefly summarize the steepest descent (SD) algorithm since it is less complex than the stochastic gradient method, and yields similar performance [9]. This method proposes to improve over the CL method by including optimization of the predictor. Since updates to the predictor affect the training residual and thereby the quantizer, which in turn impacts the training set and the predictor, a joint optimization is needed. Such joint optimization is achieved through an iterative procedure (see Fig. 4 for an implementation example).

The predictor in this case varies from iteration to iteration $\boldsymbol{A}_{SD} = \boldsymbol{A}^{(i)}$. Given a quantizer at iteration $i-1$, which is denoted by $\boldsymbol{Q}^{(i-1)}$, a training set of prediction errors $\boldsymbol{T}^{(i)} = \{\boldsymbol{e}_n^{(i)}\}_{n=1}^{N}$ is generated for iteration i, where

$$\boldsymbol{e}_n^{(i)} = \boldsymbol{x}_n - \boldsymbol{A}^{(i)}\hat{\boldsymbol{x}}_{n-1}^{(i)} \tag{2}$$

and

$$\hat{\boldsymbol{x}}_n^{(i)} = \boldsymbol{A}^{(i)}\hat{\boldsymbol{x}}_{n-1}^{(i)} + \boldsymbol{Q}^{(i-1)}(\boldsymbol{e}_n^{(i)}). \tag{3}$$

A necessary condition for predictor optimality is satisfied by requiring the gradient of the average distortion to vanish:

$$\nabla_{\boldsymbol{A}} D(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \nabla_{\boldsymbol{A}} \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \|\boldsymbol{x}_n - \hat{\boldsymbol{x}}_n\|^2 = 0. \tag{4}$$

To solve (4), Chang and Gray make the assumption that modifying $\boldsymbol{A}^{(i)}$ will only affect $\hat{\boldsymbol{x}}_n^{(i)}$ via the first term of (3), i.e., effects propagating through the quantizer in the second term of (3) may be neglected. This simplifying approximation is justifiable when changes are so small that no quantized value gets modified, in other words, when the system is virtually converged. It is not easily justified otherwise. Subject to this approximation, the optimal predictor is

$$\boldsymbol{A}^{(i)} = \boldsymbol{P}^{(i)}\boldsymbol{R}^{(i)^{-1}} \tag{5}$$

where $\boldsymbol{P}^{(i)}$ and $\boldsymbol{R}^{(i)}$ are the cross and auto correlation matrices

$$\boldsymbol{P}^{(i)} = \frac{1}{N} \sum_{n=1}^{N} \left(\boldsymbol{x}_n - \boldsymbol{Q}^{(i-1)}(\boldsymbol{e}_n^{(i)})\right) \left(\hat{\boldsymbol{x}}_{n-1}^{(i)}\right)^t \tag{6}$$

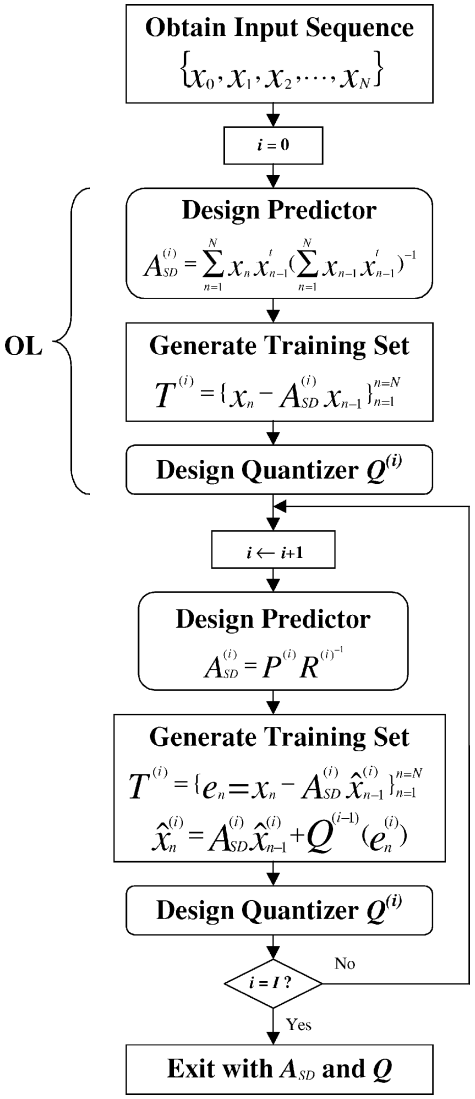


Fig. 4. Flow-diagram for an implementation of the SD training procedure.

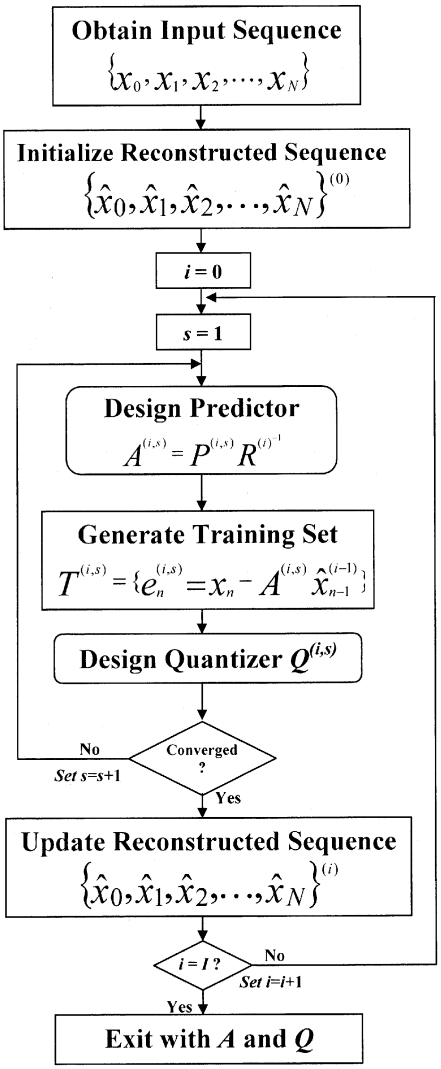


Fig. 5. Flow diagram of the ACL training procedure.

and

$$\boldsymbol{R}^{(i)} = \frac{1}{N} \sum_{n=1}^{N} \hat{\boldsymbol{x}}_{n-1}^{(i)} \left(\hat{\boldsymbol{x}}_{n-1}^{(i)} \right)^t \tag{7}$$

where, by assumption, $e_n^{(i)}$ in (6) may be computed using the available $A^{(i-1)}$.

Using $\boldsymbol{T}^{(i)}$ and the new $\boldsymbol{A}^{(i)}$, an optimized $\boldsymbol{Q}^{(i)}$ is obtained. Again, OL is used for the initialization, such that $\boldsymbol{Q}^{(0)} = \boldsymbol{Q}_{OL}$. For better stability, after calculating a new predictor, an updated training set is generated using the new predictor (and latest quantizer) before the new quantizer is designed. Similarly, an updated reconstructed sequence is generated using the new quantizer (and latest predictor) before the predictor is updated. In spite of these precautions, there is a fundamental stability problem, as there is complex interaction between predictor and quantizer. While reducing the update step size does improve the stability, it does not significantly improve the final results. In fact, in several PVQ applications, the improvement obtained by SD over the simpler CL method was reported as insignificant [4], [9].

D. Asymptotic Closed-Loop (ACL) Design

Motivated by the shortcomings of existing methods, we recently proposed the asymptotic closed-loop (ACL) approach [10], [11], which offers improved design stability. ACL design enjoys the best of both worlds, namely, it inherits the design stability of open-loop techniques while ultimately optimizing the system for closed-loop operation. In [10] and [11], our objective was quantizer design only and did not include optimization of the predictor, as the predictor in video coding applications usually takes the fixed form of motion compensation. In [12], we proposed an improvement to the basic ACL algorithm where the quantizer and predictor are jointly optimized, making the algorithm applicable to a wider array of applications. In this paper, we employ the term ACL to refer to the latter, more general version. The ACL procedure for PVQ design is illustrated in Fig. 5 and is explained later in more detail as it is an important component of the proposed deterministic annealing design approach.

The main objective of the design procedure is to avoid accumulation, through the prediction loop, of errors due to mismatched quantization. We therefore base our prediction on the

reconstructed vectors of the *previous iteration*. By basing prediction on an "older" version of reconstructed vectors, the prediction residuals are in effect calculated in open loop, and we can thus circumvent the destabilizing effects of the feedback of the closed-loop system. As a direct consequence of this "effectively open-loop" approach, a monotonic optimization procedure is possible for jointly optimizing both quantizer and predictor (for the given set of fixed vectors on which prediction is based). Once such quantizer and predictor are optimized, an improved set of reconstructed vectors is produced (also in open loop). The new set of reconstructed vectors is fixed for the next design iteration, and the steps are repeated until convergence.

To start the algorithm, a set of reconstructed vectors at iteration $i = 0$ $\{\hat{\boldsymbol{x}}_n^{(0)}\}_{n=0}^N$ is needed. This can be obtained using, for example, an OL-generated sequence. As with the SD algorithm, iteration counter i is incremented with each update of the reconstructed set. For each reconstructed set, the predictor and quantizer are iteratively optimized. Denote the prediction residuals and total distortion at iteration i achievable by the current predictor and quantizer as follows:

$$\boldsymbol{e}_n^{(i)} = \boldsymbol{x}_n - \boldsymbol{A}^{(i)}\hat{\boldsymbol{x}}_{n-1}^{(i-1)} \tag{8}$$

and

$$D^{(i)} = \sum_{n=1}^{N} \left\|\boldsymbol{e}_n^{(i)} - \boldsymbol{Q}^{(i)}(\boldsymbol{e}_n^{(i)})\right\|^2. \tag{9}$$

From (8) and (9), the distortion depends on two sets of system variables: The prediction parameters and quantizer parameters, but note that we have no closed-loop dependencies, as the prediction is based on the fixed reconstructed set from the previous iteration. An alternate minimization algorithm is proposed to minimize (9) as follows: For the current quantizer, the best predictor to minimize the distortion is obtained by descending along the gradient $\nabla_{\boldsymbol{A}^{(i)}} D^{(i)}$. This predictor is used to generate a new residual set. A new quantizer is optimized by GLA for the new training set. These two distortion-lowering steps can be repeated by iterating over the appropriately introduced subiteration counter s to convergence.

Specifically, for the design of a predictor at iteration i and subiteration s, the optimal predictor is $\boldsymbol{A}^{(i,s)} = \boldsymbol{P}^{(i,s)}{\boldsymbol{R}^{(i)}}^{-1}$, where

$$\boldsymbol{P}^{(i,s)} = \sum_{n=1}^{N} \left(\boldsymbol{x}_n - \boldsymbol{Q}^{(i,s-1)}\left(\boldsymbol{e}^{(i,s-1)}\right)\right)\left(\hat{\boldsymbol{x}}_{n-1}^{(i-1)}\right)^t \tag{10}$$

$$\boldsymbol{R}^{(i)} = \sum_{n=1}^{N} \hat{\boldsymbol{x}}_{n-1}^{(i-1)}\left(\hat{\boldsymbol{x}}_{n-1}^{(i-1)}\right)^t. \tag{11}$$

Note that the assumption/approximation of Chang and Gray (see Section II-C) is employed here and the effects of the predictor on the *quantized* prediction error have been neglected.

The quantizer is optimized based on the new training set $\boldsymbol{T}^{(i,s)} = \{\boldsymbol{e}_n^{(i,s)}\}_{n=1}^N$, where $\boldsymbol{e}_n^{(i,s)} = \boldsymbol{x}_n - \boldsymbol{A}^{(i,s)}\hat{\boldsymbol{x}}_{n-1}^{(i-1)}$.

When the predictor and quantizer iterations reach convergence, we obtain the final quantizer and predictor of this iteration: $\boldsymbol{Q}^{(i)} = \boldsymbol{Q}^{(i,s_{\text{final}})}$ and $\boldsymbol{A}^{(i)} = \boldsymbol{A}^{(i,s_{\text{final}})}$, and correspondingly, $e^{(i)} = e^{(i,s_{\text{final}})}$.

We next calculate the new set of reconstructed vectors by

$$\hat{\boldsymbol{x}}_n^{(i)} = \boldsymbol{A}^{(i)}\hat{\boldsymbol{x}}_{n-1}^{(i-1)} + \boldsymbol{Q}^{(i)}(\boldsymbol{e}_n^{(i)}). \tag{12}$$

Fix the reconstructed set and prepare the initialization of the next iteration. Assign $\boldsymbol{Q}^{(i+1,0)} = \boldsymbol{Q}^{(i)}$, $\boldsymbol{A}^{(i+1,0)} = \boldsymbol{A}^{(i)}$, increment i, and reset $s = 1$. Iterate algorithm until a convergence test is satisfied, and exit the routine with $\boldsymbol{A}_{\text{ACL}} = \boldsymbol{A}^{(i_{\text{final}})}$, and $\boldsymbol{Q} = \boldsymbol{Q}^{(i_{\text{final}})}$.

Note that in (12), the predictor $\boldsymbol{A}^{(i)}$ and quantizer $\boldsymbol{Q}^{(i)}$ are used to encode based on *exactly* the same prediction error and reconstructed vectors used for their design. Those are the best predictor and quantizer for this setting, and thus ensure that the new reconstructed set better approximates the original. Under reasonable assumptions, this results in a better prediction base for the next iteration. Consequently, we expect nearly monotonic improvement throughout the process.

Note that the entire design is in open-loop mode since we compute prediction errors for the entire sequence before quantization. As the distortion is generally decreasing, we expect the process to approach convergence. Let us next consider the implications of convergence. At convergence, further iterations do not modify the quantizer and predictor, i.e., $\boldsymbol{Q}^{(i+1)} = \boldsymbol{Q}^{(i)}$, and $\boldsymbol{A}^{(i+1)} = \boldsymbol{A}^{(i)}$, respectively. This immediately implies that the reconstruction sequence remains unchanged, i.e., $\hat{\boldsymbol{x}}_n^{(i+1)} = \hat{\boldsymbol{x}}_n^{(i)}$, and so does the next-vector prediction sequence

$$\boldsymbol{A}_{\text{ACL}}\hat{\boldsymbol{x}}_{n-1}^{(i)} = \boldsymbol{A}_{\text{ACL}}\hat{\boldsymbol{x}}_{n-1}^{(i-1)}. \tag{13}$$

This, in turn, implies that the actual prediction is the same as if it were based on the reconstruction of the current iteration (instead of on the reconstruction from the previous iteration.) In other words, the procedure is asymptotically equivalent to closed-loop design, but the algorithm is running at all times in open-loop. The procedure is thus "open-loop" in nature, yet it converges to optimization of the closed-loop performance.

The ACL algorithm offers means to overcome the central difficulties that plague earlier PVQ design techniques, and indeed, simulation results in [10]–[12], as well as the results in this paper, provide ample experimental evidence of substantial performance gains.

E. Remaining Shortcomings

ACL still suffers from three significant shortcomings involving issues of convergence and suboptimality. First, we note that the quantizer employed in (12) has only been locally optimized, which naturally implies potential suboptimality of the new reconstructed set. Second, although the predictor and quantizer are jointly optimized, the predictor is optimized under a simplifying assumption while neglecting its effects on the quantized prediction error. This is due to the piecewise constant nature of the quantizer function which poses difficulties on optimizing signals to which it is applied (in particular, gradient-based approaches are precluded). This implies suboptimality and a consequent negative impact on convergence. Finally, even if the predictor and quantizer were truly optimized within an iteration, they would simply ensure minimum distortion of the new reconstructed sequence. While it is highly likely that an improved reconstructed sequence that

better approximates the original provides a better prediction basis for the next iteration, it is possible that occasionally it does not. In other words, full convergence is only guaranteed under the reasonable (yet not completely fail-proof) assumption that better reconstruction yields better prediction. One should note in passing that this is mainly an impediment for sample-wise convergence. In terms of the average over the entire sequence, under most reasonable statistical models it can be shown by arguments, based on the law of large numbers, that the probability of increase in average distortion approaches zero for long sequences. Indeed, in simulations, we experienced small limit cycles rather than full convergence, which we mainly attribute to the first two shortcomings, and only mildly to the last (for reasons that will become clearer in the sequel). We note, however, that no serious stability problems have ever been observed. The above suboptimality and convergence problems motivate the deterministic annealing method we propose in the next section.

III. Deterministic Annealing Approach to Predictive Vector Quantizer Design

In this paper, we propose a PVQ design method that significantly outperforms all of the methods described in the previous section. The new approach (DA-ACL) embeds the ACL method within a deterministic annealing framework. The DA-ACL algorithm we propose does not require initialization of codebooks or prediction parameters and can avoid many local minima of the cost surface. Further, due to the probabilistic formulation of the system, it is easy and natural to directly optimize all system parameters with respect to the same cost objective, which is the overall distortion, without recourse to simplifying assumptions. In particular, effects of the predictor on the reconstruction distortion via the quantizer module are directly taken into account as the quantizer is no longer a piecewise constant function.

Our approach is inspired by, and builds on, the DA approach for vector quantizer design [13], [14]. It is motivated by the observation of annealing processes in physical chemistry. 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 predictive quantizer 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 optimal configuration requires that all codevectors be coincident at the centroid of the source distribution; no initialization of quantizer or predictor is necessary.

We then track the Lagrangian minimum at successively lower levels of entropy by recalculating the optimum locations of the reproduction points and the encoding probabilities at each stage. At the limit of zero randomness, the algorithm directly minimizes the average distortion cost, and a deterministic encoder is obtained.

An important objective is to avoid accumulation of errors due to mismatched quantization through feedback. Adopting the ACL approach [10], we therefore base our prediction on the reconstructed vectors of the previous iteration. By basing prediction on an older set of reconstructed vectors, we can, in effect, remove the feedback of the closed-loop system. As a direct consequence, a monotonic optimization procedure is possible for jointly optimizing both quantizer and predictor. The details of the iterative algorithm are described in what follows.

At each iteration, we optimize the predictor and quantizer. The plain notation $(\boldsymbol{A}, \boldsymbol{Q})$ is used to refer to the latest available values for the parameters of the predictor and quantizer, whereas their values at the end of the iteration are denoted $(\boldsymbol{A}^{(i)}, \boldsymbol{Q}^{(i)})$. Assume that quantizer $\boldsymbol{Q}$, at iteration i, employs a codebook of M codevectors $\{\boldsymbol{y}_j\}_{j=1}^{M}$. For each fixed reconstruction sequence $\{\hat{\boldsymbol{x}}_n^{(i-1)}\}_{n=0}^{N}$, we can calculate the total distortion achievable by the current predictor and quantizer as follows. For the prediction residual $\boldsymbol{e}_n^{(i)} = \boldsymbol{x}_n - \boldsymbol{A}\hat{\boldsymbol{x}}_{n-1}^{(i-1)}$, the squared error distance from $\boldsymbol{y}_j$ is defined as

$$d(\boldsymbol{e}_n^{(i)}, \boldsymbol{y}_j) = \|\boldsymbol{e}_n^{(i)} - \boldsymbol{y}_j\|^2. \tag{14}$$

The random quantizer's encoding rule is represented by conditional probability distribution $P_{j|n}$, which measures the probability that prediction residual vector $\boldsymbol{e}_n^{(i)}$ is quantized to codebook entry $\boldsymbol{y}_j$. The random quantizer is hence specified by the codebook and the encoding probabilities. The corresponding joint distribution is denoted $P_{nj} = P_n P_{j|n}$.

Once the "hard" quantizer of traditional design approaches is replaced with the random quantizer, the total (expected) distortion at iteration i becomes

$$D = \sum_n \sum_j P_{nj} d(\boldsymbol{e}_n^{(i)}, \boldsymbol{y}_j). \tag{15}$$

Note that due to this probabilistic framework, the overall distortion is a differentiable function of all predictor and quantizer parameters. We minimize the distortion under a constraint on the entropy which measures the level of randomness of the system. The entropy of the system is

$$H = -\sum_n \sum_j P_{nj} \log P_{nj}. \tag{16}$$

In Lagrangian formulation, we minimize

$$F = D - TH \tag{17}$$

where the Lagrange multiplier T is referred to as "temperature." Note that this is exactly the Helmholtz free energy of statistical physics if we define the system energy by D and its entropy by H. Assuming uniform distribution over the training set $P_n = 1/N$, an expression for the optimal encoding probability at iteration i, $P_{j|n}^{(i)}$ can be obtained by setting

$$\frac{\partial F}{\partial P_{j|n}} = d(\boldsymbol{e}_n^{(i)}, \boldsymbol{y}_j) + T \log P_{j|n} + T = 0 \tag{18}$$

which yields (after normalization) the Gibbs distribution

$$P_{j|n}^{(i)} = \frac{e^{-d(e_n^{(i)}, y_j)/T}}{\sum_k e^{-d(e_n^{(i)}, y_k)/T}}. \quad (19)$$

As with VQ design using DA, the association probabilities are effectively controlled by the temperature. From (19), it can be seen that high temperature assigns uniform probability associations. As the temperature is lowered, those association probabilities become more discriminatory and assign higher probability to nearer codevectors. Hence, (19) is a probabilistic generalization of the standard nearest neighbor rule.

The optimal predictor satisfies

$$\nabla_{\boldsymbol{A}} F = \nabla_{\boldsymbol{A}} D = -2 \sum_n \sum_j P_n^{(i)} P_{j|n}^{(i)} \times \left(\boldsymbol{x}_n - \boldsymbol{A}\hat{\boldsymbol{x}}_{n-1}^{(i-1)} - \boldsymbol{y}_j \right) \left(\hat{\boldsymbol{x}}_{n-1}^{(i-1)} \right)^t = \boldsymbol{0} \quad (20)$$

and hence

$$\boldsymbol{A}^{(i)} = \left(\sum_n \boldsymbol{x}_n \left(\hat{\boldsymbol{x}}_{n-1}^{(i-1)} \right)^t - \sum_n \sum_j P_{j|n}^{(i)} \boldsymbol{y}_j \left(\hat{\boldsymbol{x}}_{n-1}^{(i-1)} \right)^t \right) \times \left(\sum_n \hat{\boldsymbol{x}}_{n-1}^{(i-1)} \left(\hat{\boldsymbol{x}}_{n-1}^{(i-1)} \right)^t \right)^{-1} \quad (21)$$

where (21) is the probabilistic counterpart of (10) and (11).

Finally, the optimal codevectors must satisfy

$$\nabla_{\boldsymbol{y}_j} F = \nabla_{\boldsymbol{y}_j} D = -2 \sum_n P_n^{(i)} P_{j|n}^{(i)} \left(\boldsymbol{e}_n^{(i)} - \boldsymbol{y}_j \right) = \boldsymbol{0} \quad (22)$$

which yields

$$\boldsymbol{y}_j^{(i)} = \frac{\sum_n P_{j|n}^{(i)} \boldsymbol{e}_n^{(i)}}{\sum_n P_{j|n}^{(i)}}. \quad (23)$$

This is a generalization of the standard centroid rule. Note that $\boldsymbol{e}_n^{(i)}$ here utilizes the updated version $\boldsymbol{A} = \boldsymbol{A}^{(i)}$.

The reconstructed set is updated by running a probabilistic counterpart to an iteration of ACL. Recall the ACL update rule:

$$\hat{\boldsymbol{x}}_n^{(i)} = \boldsymbol{A}^{(i)} \hat{\boldsymbol{x}}_{n-1}^{(i-1)} + \boldsymbol{Q}^{(i)}(\boldsymbol{e}_n^{(i)}). \quad (24)$$

Thus, the new reconstructed set is an open-loop update of the previous reconstructed set using the optimized predictor and quantizer. For the probabilistic quantizer of DA, we can rewrite (24) as

$$\hat{\boldsymbol{x}}_n^{(i)} = \boldsymbol{A}^{(i)} \hat{\boldsymbol{x}}_{n-1}^{(i-1)} + \sum_j P_{j|n}^{(i)} \boldsymbol{y}_j^{(i)} \quad (25)$$

where the quantization procedure in DA is equivalent to a weighted average of all the codevectors in the codebook based on the conditional probabilities.

In summary, the DA-ACL algorithm (see Fig. 6) revolves around four important steps.

1) Calculate conditional probabilities.
2) Calculate the optimal predictor.
3) Calculate the optimal quantizer.
4) Update the reconstructed set.

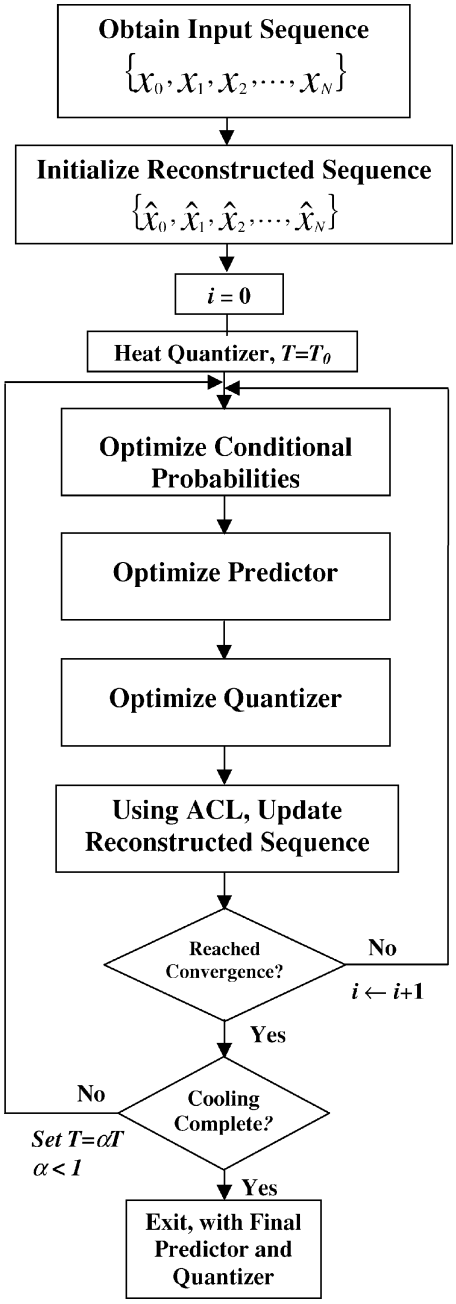
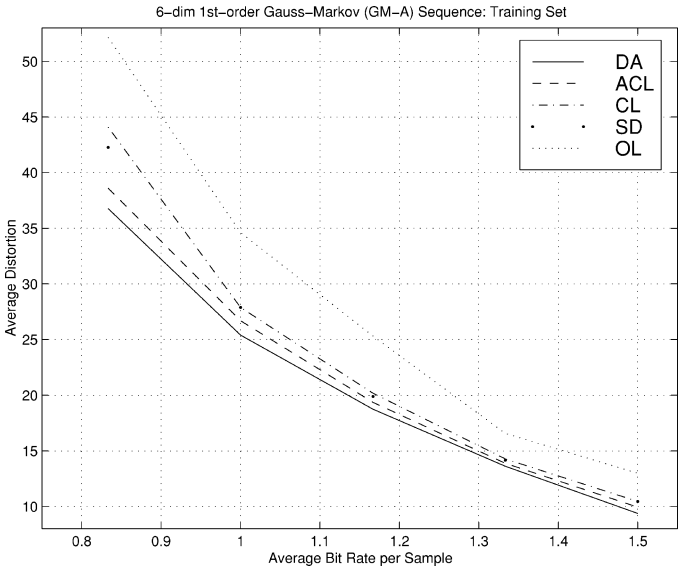
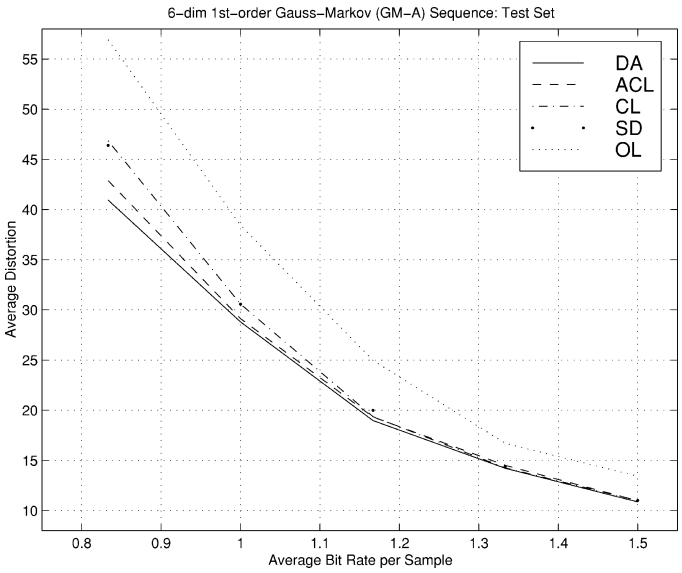


Fig. 6. Flow diagram for the DA-ACL training procedure.

The four steps are explicitly given by (19), (21), (23), and (25) and are repeated until a convergence test is satisfied. At convergence, the temperature T is lowered and the whole process is repeated. To start the algorithm, a reconstructed set is initialized (as shown in Fig. 6). In our experiments, the original sequence was used as initialization. However, this choice is not critical to the quality of the design. We have tested arbitrary initialization (with the all zero sequence) as well as initialization with the OL-reconstruction sequence and observed that although initialization had an effect on the execution time of the design, it did not have a significant impact on the quality of the resulting system.

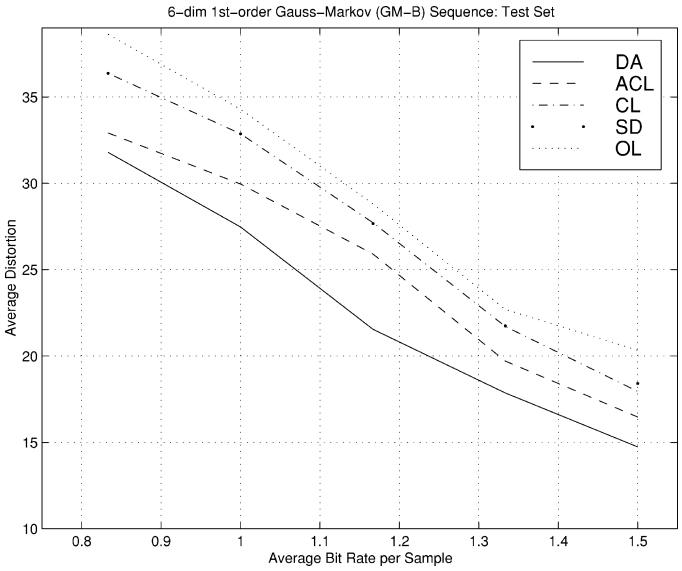


(a)

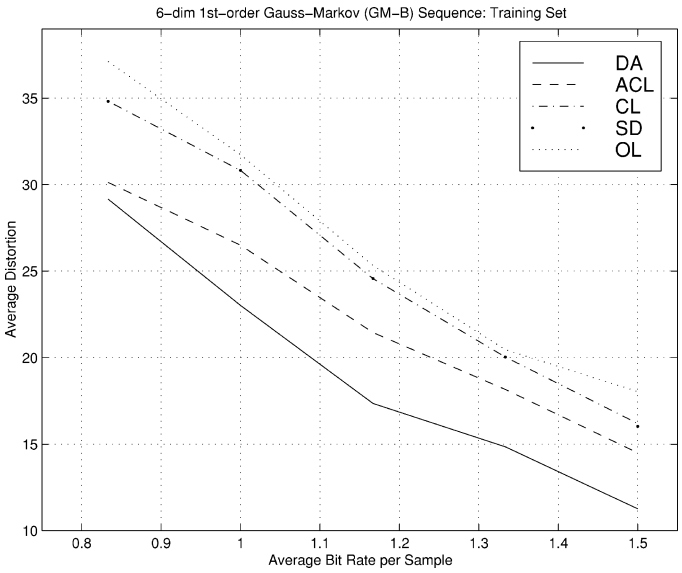


(b)

Fig. 7. Synthetic source GM-A: Average distortion of OL, CL, SD, and ACL design and the proposed DA-ACL approach to predictive quantizer design at various bit rates for (a) training set and (b) test set.



(a)



(b)

Fig. 8. Synthetic source GM-B: Average distortion of OL, CL, SD, and ACL design and the proposed DA-ACL approach to predictive quantizer design at various bit rates for (a) training set and (b) test set.

Recall the optimality and convergence shortcomings of ACL enumerated in Section II-E. With respect to the first shortcoming, the proposed DA approach offers substantial means to avoid poor local optima. Further, it optimizes the predictor and quantizer jointly with respect to the reconstruction error, without neglecting the predictor's impact on the quantized prediction error (thanks to the differentiability of the probabilistic quantizer). It hence eliminates the second shortcomings. The only remaining gap in ensuring full convergence is due to step (4). Equation (25) merely utilizes the current best quantizer and predictor to generate a new reconstructed set which more closely resembles the original set subject to the entropy constraint. Clearly, it is not very probable yet possible that better reconstruction of a given vector would result in degraded prediction of the next vector and increased contribution to the Lagrangian. However, the Lagrangian is averaged over many samples. In fact, the probability of increase in the overall Lagrangian approaches zero for long sequences. In our simulations of DA-ACL we have always witnessed monotone nonincreasing Lagrangian at a given temperature. In particular, the algorithm always reached complete convergence.

IV. Simulation Results

A number of synthetic sources, with different characteristics, have been used in the experiments. For natural sources, we use line spectral frequency (LSF) parameters which were extracted from a speech source. In this section, we give detailed simulation results to demonstrate the power of the proposed DA-ACL algorithm.

A. Experiment Details

All nonprobabilistic methods, namely, OL, CL, SD, and ACL employ the particular codebook design variant of GLA, called

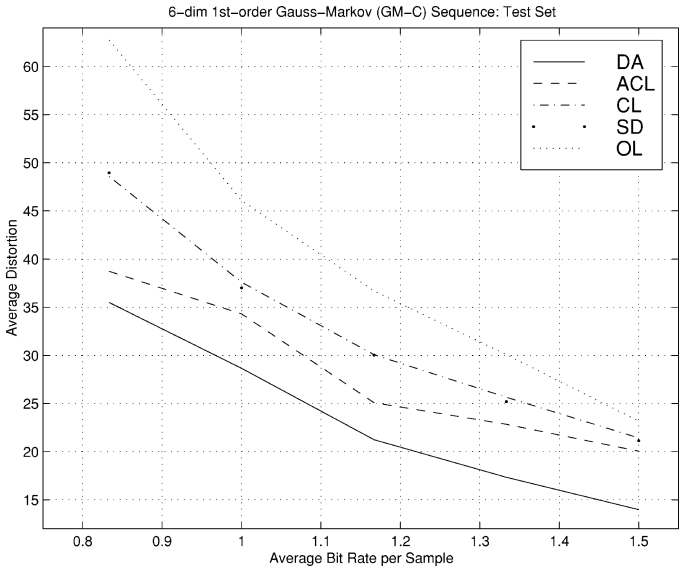


Fig. 9. Synthetic source GM-C: Average distortion of OL, CL, SD, and ACL design and the proposed DA-ACL approach to predictive quantizer design at various bit rates for (a) training set and b) test set.

selective splitting, which is described in [16]. The selective splitting codebook design algorithm initializes a codebook with a codevector at the centroid of the training set and then proceeds to selectively split individual codevectors based on criteria involving distortion and entropy. In the current work, fixed rate is assumed, and thus, the criterion used is only distortion based. Selective splitting has the advantage of not relying on initialization at the price of additional complexity (which is typically substantially lower than repeated design with different random codebook initializations). In addition, when the VQ codebook is used within a PVQ system, iterative updates to the training set often render codevectors unused, and selective splitting offers a natural way to fill the empty cells. The selective splitting design method was used for all relevant method implementations to ensure fair comparison.

The complexity of the design varies among the algorithms. The OL method is essentially the least complex since the training set is unchanged. The CL method is an order of magnitude more complex than OL since, with each new training set, complexity similar to OL is needed for the design. The SD and ACL methods are of similar complexity, but both are slightly more complex than CL since the predictor is also optimized. Finally, the DA-ACL algorithm is considerably more complex than the ACL method due to the slow annealing process dictated by DA. However, it is important to note that design complexity is handled offline. The end result of the design procedure is a predictor and a quantizer, and thus, the compression runtime complexity is the same regardless of the design method.

TABLE I
GAIN IN DECIBELS OF DA-ACL OVER ACL FOR THE VARIOUS SYNTHETIC SOURCE PROCESSES, AT THE DIFFERENT BIT RATES (INDICATED IN BITS PER VECTOR)

Source		5	6	7	8	9
GM-A	Train	0.21	0.22	0.14	0.09	0.26
	Test	0.20	0.05	0.82	0.09	0.05
GM-B	Train	0.14	0.61	0.92	0.87	1.10
	Test	0.15	0.38	0.80	0.43	0.48
GM-C	Train	0.43	0.77	0.70	1.26	1.58
	Test	0.38	0.78	0.72	1.20	1.57

TABLE II
GAIN IN DECIBELS OF DA-ACL OVER CL FOR THE VARIOUS SYNTHETIC SOURCE PROCESSES, AT THE DIFFERENT BIT RATES (INDICATED IN BITS PER VECTOR)

Source		5	6	7	8	9
GM-A	Train	0.79	0.40	0.32	0.21	0.47
	Test	0.59	0.26	0.09	0.01	0.06
GM-B	Train	0.77	1.27	1.51	1.30	1.58
	Test	0.58	0.78	1.09	0.85	0.86
GM-C	Train	1.26	1.29	1.55	1.72	1.99
	Test	1.36	1.18	1.51	1.70	1.86

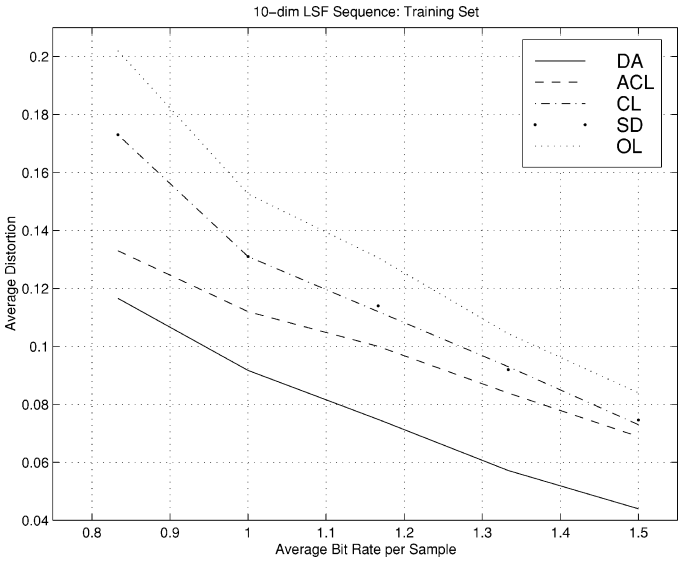
B. Experiments on Synthetic Sources

In the first set of experiments, the proposed DA-ACL approach is compared to the OL, CL, SD, and ACL methods on several synthetic sources. Specifically, sources with the following characteristics were synthesized:

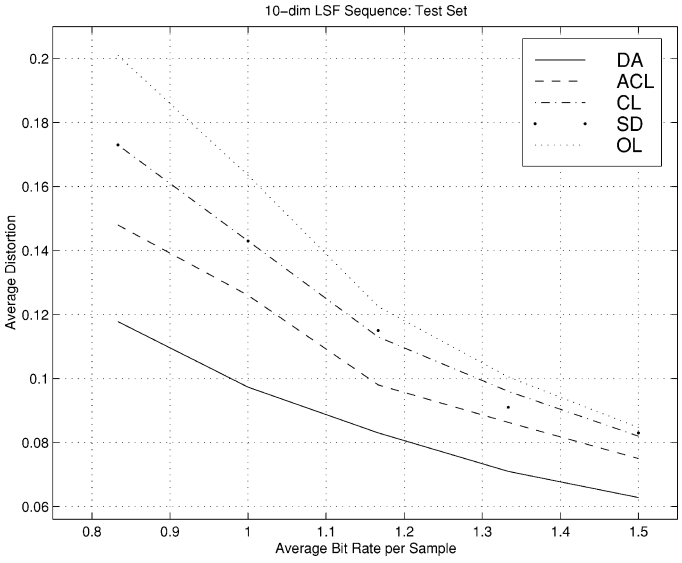
Source GM-A: a six-dimensional (6-D) first-order Gauss-Markov source with intra-vector and inter-vector correlation coefficients of 0.9;

Source GM-B: a 6-D Gauss-Markov source with intra-vector and inter-vector correlation of 0.8 and 0.95, respectively;

Source GM-C: a 6-D first-order Gauss-Markov source with inter-vector correlation of 0.8 and (in an effort to produce a more realistic source) intra-vector correlation varying along the components, in the range of 0.5 to 0.95;



(a)



(b)

Fig. 10. Line spectral frequency parameters: Average distortion of OL, CL, SD, and ACL design and the proposed DA-ACL approach to predictive quantizer design at various bit rates for (a) training set and b) test set.

Source LM: a ten-dimensional (10-D) Laplace-Markov source with intra-vector and inter-vector correlation of 0.95.

The training and test sets consisted of independent sequences of 10 000 and 1000 vectors, respectively. A first-order predictor is used in the PVQ design.

In Fig. 7, we show the average distortion as a function of bits per dimension for the source GM-A. It can be seen that the gains achievable for this source can sometimes be quite small, and this is due to simplicity of the source which does not pose an "optimization challenge." For the more complicated sources, such as source GM-B and GM-C, gains are much more substantial. The cases of GM-B and GM-C are shown in Figs. 8 and 9, respectively. The gains in dB for the comparison of DA-ACL with ACL, for all three synthetic sources, are shown in Table I. Table II compares our new design approach DA-ACL to CL, which is the commonly used method for training of PVQ. It is evident that gains achievable in those complicated sources can be quite significant, ranging from 0.48 to 1.99 dB for the highest bit rates. It is interesting to note that gains within the training set increase with increasing rate. This is mainly due to the fact that at higher rates, more codevectors need to be optimized, and the strength of the DA approach comes into play as the optimization avoids many local optima into which a traditional design approach would fall.

TABLE III
GAIN IN DECIBELS OF DA-ACL OVER ACL FOR THE LSF PARAMETER SOURCE, AT THE DIFFERENT BIT RATES (INDICATED IN BITS PER VECTOR)

	5	6	7	8	9
Train	0.57	0.87	1.26	1.66	1.95
Test	0.99	1.12	0.72	0.85	0.77

TABLE IV
GAIN IN DECIBELS OF DA-ACL OVER CL FOR THE LSF PARAMETER SOURCE, AT THE DIFFERENT BIT RATES (INDICATED IN BITS PER VECTOR)

	5	6	7	8	9
Train	1.71	1.55	1.75	2.11	2.20
Test	1.67	1.67	1.34	1.31	1.16

For the LM source, gains between 0.62 to 1.22 dB were achieved when comparing DA-ACL to ACL and between 1.42 to 2.19 dB when comparing DA-ACL to CL.

C. Experiments on Line Spectral Frequencies

Several speech coding standards are centered around the encoding of line spectral frequency (LSF) parameters. LSF parameters typically constitute 25 to 50% of the speech coding rate. In addition, LSF parameters have significant and direct impact on the intelligibility and quality of decoded speech. Thus, high-quality encoding of LSF parameters is an important objective, and careful design of PVQ is warranted. A frame of speech generally includes a 10-D LSF vector that usually exhibits significant correlation with previous frames. A common approach is to use a first- or second-order interframe predictor to remove such redundancies. In this experiment, we use a first-order predictor and quantizers that operate at a variety of bit rates.

In Fig. 10, we show the average distortion as a function of bits per dimension for the LSF source. This bit rate range is below the standard rates used to code LSF parameters, but we target this range for simplicity because it is realizable using a single stage VQ, and the test demonstrates the optimization capabilities of DA-ACL on a natural source. Again, the training and test sets consisted of independent sequences of 10 000 and 1000 vectors, respectively. The gains in decibels for the comparison of DA-ACL with ACL, and DA-ACL with CL, for the LSF source, are shown in Tables III and IV, respectively. It can be seen that DA-ACL may gain over ACL by more than 1 dB and over CL by more than 2 dB.

V. CONCLUSIONS

This paper describes a new approach to training predictive vector quantizers, which does not suffer from the statistical

mismatch typical of OL training algorithms, nor from the instability experienced in CL and SD approaches. The proposed iterative algorithm is open-loop in nature but asymptotically optimizes the closed-loop system. A deterministic annealing formulation of the design problem is presented that allows for the simultaneous optimization of both predictor and quantizer while minimizing the reconstruction distortion. Moreover, the approach avoids many suboptimal local minima. The new design approach is tested on a number of synthetic and natural sources and provides significant gains on most sources at various bit rates.

REFERENCES

[1] M. R. Schroeder and B. S. Atal, "Code-excited linear prediction (CELP): High quality at very low bit rates," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Mar. 1985, pp. 937–940.

[2] "Coding of speech at 8 Kbits/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)," ITU-T Study Group, 6.5 ed., 1995.

[3] "Dual rate speech coder for multimedia communication transmitting at 5.3 and 6.3 kbit/s," ITU-T, ITU-T G.723, 1995.

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

[5] P. C. Cosman, R. M. Gray, and M. Vetterli, "Vector quantization of image subbands: A survey," *IEEE Trans. Image Processing*, vol. 5, pp. 202–225, Feb. 1996.

[6] F. Kossentini, W. C. Chung, and M. J. T. Smith, "Conditional entropy-constrained residual VQ with application to image coding," *IEEE Trans. Image Processing*, vol. 5, pp. 311–320, Feb. 1996.

[7] L. Corte-Real and A. P. Alves, "A very low bit rate video coder based on vector quantization," *IEEE Trans. Image Processing*, vol. 5, pp. 263–273, Feb. 1996.

[8] V. Cuperman and A. Gersho, "Vector predictive coding of speech at 16 kbits/s," *IEEE Trans. Commun.*, vol. COM-33, pp. 685–696, July 1985.

[9] P. C. Chang and R. M. Gray, "Gradient algorithms for designing predictive vector quantizers," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 679–690, Aug. 1986.

[10] K. Rose, H. Khalil, and S. L. Regunathan, "Open-loop design of predictive vector quantizers for video coding," *Proc. IEEE Int. Conf. Image Process.*, vol. 3, pp. 953–957, Oct. 1998.

[11] H. Khalil, K. Rose, and S. L. Regunathan, "The asymptotic closed-loop approach to predictive vector quantizer design with application in video coding," *IEEE Trans. Image Processing*, vol. 10, pp. 15–23, Jan. 2001.

[12] H. Khalil and K. Rose, "Robust predictive vector quantizer design," in *Proc. Data Compression Conf.*, Snowbird, UT, Mar. 2001, pp. 33–42.

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

[14] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems," *Proc. IEEE*, vol. 86, pp. 2210–2239, Nov. 1998.

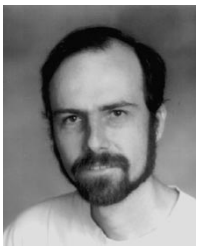
[15] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. COM-28, pp. 84–95, Jan. 1980.

[16] H. Khalil and K. Rose, "Selective splitting approach to entropy-constrained single/multi-stage vector quantizer design," in *Proc. SPIE Conf. Electron. Imaging*, San Jose, CA, Jan. 2000.

Hosam Khalil (S'98–M'01) was born in Washington, DC, in 1972. He received the B.S. and M.S. degrees in electrical engineering with honors from Cairo University, Cairo, Egypt, in 1993 and 1996, respectively. He received the Ph.D. degree from the University of California, Santa Barbara, in 2001.

From October 1993 to July 1996, he was an Assistant Teacher at Cairo University. He was also employed part-time during that period with IBM, Cairo. In the summer of 1998, he was an intern in the Multimedia Communications Research Laboratory, Bell Laboratories, Murray Hill, NJ. From March 1999 to March 2000, he was a Software Engineer at SignalCom, Inc., Goleta, CA. In April 2000, he joined Microsoft Corporation, Digital Media Division, Redmond, WA, where he is a Software Design Engineer. His research interests are in image, video, and speech compression and speech recognition.

Mr. Khalil was a recipient of the UC Regents Fellowship.



Kenneth Rose (S'85–M'91–SM'01–F'03) received the B.Sc. (summa cum laude) and M.Sc. (magna cum laude) degrees in electrical engineering from Tel-Aviv University, Tel-Aviv, Israel, in 1983 and 1987, respectively, and the Ph.D. degree from the California Institute of Technology (Caltech), Pasadena, in 1991.

From July 1983 to July 1988, he was with Tadiran Ltd., Holon, Israel, where he carried out research in the areas of image coding, image transmission through noisy channels, and general image processing. From September 1988 to December 1990, he was a graduate student at Caltech. In January 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 information theory, source and channel coding, image coding and processing, speech and general pattern recognition, and nonconvex optimization in general. He is also particularly interested in the relations between information theory and statistical physics and their potential impact on fundamental and practical problems in diverse disciplines.

Dr. Rose currently serves as Editor for Source-Channel Coding for the IEEE TRANSACTIONS ON COMMUNICATIONS. He co-chaired the technical program committee of the 2001 IEEE Workshop on Multimedia Signal Processing. In 1990, he received (with A. Heiman) the William R. Bennett Prize Paper Award from the IEEE Communications Society.