Robust Vector Quantization for Wireless Channels

Wen-Whei Chang, *Member, IEEE*, Tan-Hsu Tan, and De-Yu Wang

Abstract—This study focuses on two issues: parametric modeling of the channel and index assignment of codevectors, to design a vector quantizer that achieves high robustness against channel errors. We first formulate the design of a robust zero-redundancy vector quantizer as a combinatorial optimization problem leading to a genetic search for a minimum-distortion index assignment. Performance is further enhanced by the use of the Fritchman channel model that more closely characterizes the statistical dependencies between error sequences. This study also presents an index assignment algorithm based on the Fritchman model with parameter values estimated using a real-coded genetic algorithm. Simulation results indicate that the global explorative properties of genetic algorithms make them very effective in estimating Fritchman model parameters, and use of this model can match index assignment to expected channel conditions.

Index Terms—Genetic algorithm, index assignment, vector quantization.

I. INTRODUCTION

VECTOR QUANTIZERS (VQs) are used in many applications and allow optimum mapping of a large set of input vectors into a finite set of representative codevectors [1]. Transmitting VQ data over noisy channels changes the encoded information and consequently leads to severe distortions in the reconstructed output. Forward error control could be used to protect VQ data but it is more efficient to mitigate the effects of channel errors without adding redundant bits. This has motivated investigation into trying to design a nonredundant VQ system with increased robustness to channel errors. Conventional design approaches to channel robustness are available in two general categories: robust VQ (RVQ) and channel-optimized VQ (COVQ). The encoder–decoder pair of a COVQ system is jointly trained for the given channel, whereas an RVQ encoder is trained for a noiseless channel and then made robust by assigning suitable indices to the codevectors. Of the two schemes, the RVQ is popular where codebook training time is of primary concern, and the COVQ is more appropriate when higher levels of robustness are needed. A more comprehensive discussion of transmitting VQ data over noisy channels can be found in [2] and [3]. No matter which design scheme is used, special care must be taken to ensure that actual error characteristics are incorporated into the computation of channel transition probabilities. While this paper addresses the index assignment problem for use in RVQ, the design techniques used to refine the channel transition probabilities can be applied to COVQ design as well.

Finding the best index assignment requires searching every possible codebook permutation for the one that yields the minimum distortion under noisy channel conditions. However, because the nature of this solution is NP-complete, it requires enormous computational complexity which, for even small size codebooks, may be prohibitive. In [4]–[8], several practical procedures for solving the index assignment problem have been proposed. Although effectively reducing a system's sensitivity to channel errors, these approaches concentrate on mathematically simple memoryless binary symmetric channels. Unfortunately, however, transmission errors encountered in most real communication channels exhibit various degrees of statistical dependencies that are contingent on the transmission medium and on the particular modulation technique used. A typical example occurs in digital mobile radio channels, where bit-rate reduced speech signals suffer severe degradation from error bursts due to the combined effects of fading and multipath propagation. Further improvement can be realized through a more precise characterization of the channel on which index assignment design is based. Toward this end, recent work on VQs for channels with memory has considered binary Markov channels [9], [10] and Gaussian channels [11], [12]. For this investigation, we focused on the simple partitioned Markov chain model proposed by Fritchman [13]. This model has several practical advantages over the Gilbert model [14] and other models [15]. First, the Fritchman model is relatively simple and can characterize a wide range of digital channels, as evidenced by its applicability to performance analysis of various error control schemes [16], [17]. Second, as we shall see later, the channel transition probabilities of the Fritchman model have closed-form expressions that can be represented in terms of model parameters.

This study investigates the following optimization problems: 1) developing a model representative of real channel behavior and 2) assigning codevector indices for a given noisy channel. Traditionally, optimization problems have been dealt with using the gradient-descent algorithm [18]. However, gradient-descent approaches perform local searches and may fail to provide reliable results when used on complex optimization problems with multiple local optima. By contrast, the global explorative properties of genetic algorithms have made them very effective at solving constrained as well as combinatorial optimization problems [19]. The first part of this paper focuses on estimating the parameters of the Gilbert and Fritchman channel models from an experimental error-gap distribution. Exponential curve fitting can formulate the channel characterization as a constrained optimization problem that is amenable to the application of real-
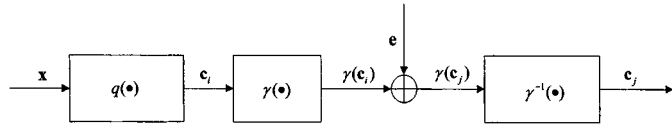
Fig. 1.   Block diagram of the VQ transmission system.

coded genetic algorithms. The second part of this study develops mathematical tools for use with the Fritchman model in optimizing the index assignment under noisy channel conditions. Simulation results indicate that use of Fritchman channel characterization enables the implementation of an index assignment that better matches the intrinsic natures of error statistics.

The rest of this paper is organized as follows. Section II addresses the index assignment problem with respect to the design of a robust vector quantizer. In Section III, we present the Gilbert and Fritchman channel models. Details of the genetic algorithms required for solving the channel modeling problem, along with some numerical results for comparison with experimental measurements, are provided in Section IV. Section V suggests a way to associate Fritchman model parameters with optimal index assignment of VQ codevectors. Section VI presents the index assignment performance comparison and investigates the robustness of the vector quantizer under channel mismatch conditions. Finally, Section VII gives our conclusions.

II. ROBUST VQ AND THE INDEX ASSIGNMENT PROBLEM

The vector quantizer provides better performance than the scalar quantizer by employing a multipath search that pursues a set of alternatives and chooses among them the best possible output sequence. A block diagram depicting the VQ transmission system is shown in Fig. 1. The VQ encoder searches through the codebook for the codevector $\mathbf{c}_i$ that best matches the input vector $\mathbf{x}$, and then transmits the corresponding codevector index $\gamma(\mathbf{c}_i)$ to the decoder in binary format. Here, the codebook consisting of $M = 2^m$ codevectors, $C = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_M\}$, is designed for a noiseless channel using the generalized Lloyd algorithm [20]. The resulting codebook, with an arbitrary permutation, is used as the input to the index assignment function, $\{\gamma(\mathbf{c}_i), 1 \leq i \leq M\}$, which yields the same codebook but with its codevectors in different locations.

One of the principal concerns in transmitting VQ data over noisy channels is that channel errors affect the bits that convey information about codevector indices. Assume that a channel's input and output are defined by the bit strings $\gamma(\mathbf{c}_i) = (\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{im})$ and $\gamma(\mathbf{c}_j) = (\gamma_{j1}, \gamma_{j2}, \ldots, \gamma_{jm})$. They differ when the error pattern $\mathbf{e} = (e_1, e_2, \ldots, e_m)$ corrupts the input, so that the channel output bit $\gamma_{jk} = \gamma_{ik} \oplus e_k$, where the symbol $\oplus$ denotes the bitwise logical exclusive-OR operation. Consequently, the decoder looks up the corresponding codevector in its codebook and releases J samples of the codevector $\mathbf{c}_j$, instead of $\mathbf{c}_i$, as the output. The resultant distortion between the codevectors $\mathbf{c}_i$ and $\mathbf{c}_j$ is denoted by $d(\mathbf{c}_i, \mathbf{c}_j)$, which will be assumed in the sequel to be a squared Euclidean distance, i.e., $d(\mathbf{c}_i, \mathbf{c}_j) = \|\mathbf{c}_i - \mathbf{c}_j\|^2$. In this framework, the overall distortion $E[\|\mathbf{x} - \mathbf{c}_j\|^2]$ can be divided into three terms: quantizer distortion $D_q = E[\|\mathbf{x} - \mathbf{c}_i\|^2]$, channel

distortion $D_c = E[\|\mathbf{c}_i - \mathbf{c}_j\|^2]$, and mixed-term distortion $D_m = 2E[(\mathbf{x} - \mathbf{c}_i)^T(\mathbf{c}_i - \mathbf{c}_j)]$. Among them, D_q does not depend on the channel condition and D_m is zero for optimum codebooks designed for noiseless channels [6]. Viewing from this perspective, we focus our attention on the index assignment problem for the purpose of minimizing channel distortion D_c. To begin, let $P(\gamma(\mathbf{c}_j)|\gamma(\mathbf{c}_i))$ denote the probability of receiving the index $\gamma(\mathbf{c}_j)$ given that the transmitted index is $\gamma(\mathbf{c}_i)$. For the specific index assignment function $\gamma(\mathbf{c}_i)$, $i = 1, 2, \ldots, M$, the average distortion due to channel errors is expressed as

$$D_c(\gamma) = \sum_{i=1}^{M} \sum_{j=1}^{M} P(\mathbf{c}_i) P(\gamma(\mathbf{c}_j)|\gamma(\mathbf{c}_i)) \|\mathbf{c}_i - \mathbf{c}_j\|^2 \quad (1)$$

where $P(\mathbf{c}_i)$ is the *a priori* probability of the codevector $\mathbf{c}_i$. From a mathematical standpoint, the index assignment problem is equivalent to a combinatorial optimization problem in which channel distortion $D_c(\gamma)$ is the objective function to be minimized.

The effect of channel errors on the quality of a vector quantizer is introduced through the values of the channel transition probabilities $P(\gamma(\mathbf{c}_j)|\gamma(\mathbf{c}_i))$. For a stationary and symmetric channel, they can be represented in terms of the probability distribution of the error pattern, i.e.,

$$P(\gamma(\mathbf{c}_j)|\gamma(\mathbf{c}_i)) = P(e_1, e_2, \ldots, e_m). \quad (2)$$

To permit theoretical analysis, most previous works on index assignment assumed that codevector indices were transmitted over memoryless binary symmetric channels [4]–[8]. This reduces channel transition probabilities to the following:

$$P(\gamma(\mathbf{c}_j)|\gamma(\mathbf{c}_i)) = \varepsilon^l (1 - \varepsilon)^{m-l} \quad (3)$$

where ε is the channel bit error rate (BER) and l is the number of ones occurring in the error pattern $\mathbf{e}$. However, the errors encountered over most wireless channels are not independent; rather, they tend to occur in bursts. It is, therefore, believed that further improvement can be obtained through intelligent exploitation of the statistical dependencies between error occurrences.

III. GILBERT AND FRITCHMAN CHANNEL MODELS

The design of a more sophisticated error protection scheme requires that parameterized probabilistic models be used to summarize some of the most relevant aspects of error statistics. Most studies have emphasized the use of a Markov chain consisting of a finite number of states with defined transition probabilities [15]. Among them, Gilbert [14] and Fritchman [13] models have been shown to adequately describe the observed error bursts occurring in digital mobile radio channels [18]. The Gilbert model consists of a Markov chain having an error-free state G and a bad state B, in which errors occur with the probability $1 - h$. The state transition probabilities are P and p for the G to B and B to G transitions, respectively. Notice that in the particular case of a Gilbert model with parameter values $P = 1$, $p = 0$, $h = 1 - \varepsilon$, the channel model reduces to a memoryless binary symmetric channel with the BER ε.
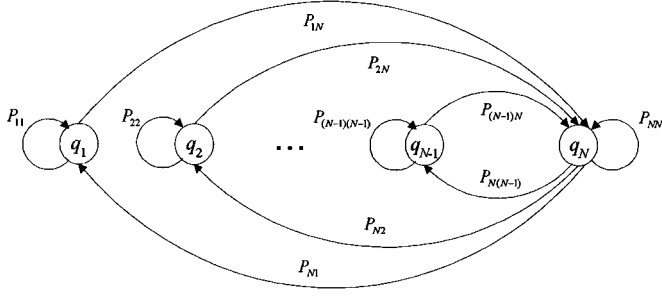
Fig. 2. Fritchman channel model with $N-1$ error-free states and a single error state.

To match the observed error sequences more closely, it is necessary to use a Markov model with more than one error-free state. Fritchman [13] investigated the more general case of a Markov chain with N states, partitioned into a group of k error-free states, $Q_A = \{q_1, q_2, \ldots, q_k\}$, and another group of $(N-k)$ error states, $Q_B = \{q_{k+1}, q_{k+2}, \ldots, q_N\}$. Let $\mathbf{e} = \{e_1, e_2, \ldots\}$ be an error sequence generated by the model and let $\mathbf{s} = \{s_1, s_2, \ldots\}$ be the sequence of states that evolves along with the transition probability $P_{ij} = \Pr(q_j \text{ at } n+1 | q_i \text{ at } n)$. At each time instant n, the occurrence of 0 or 1 in e_n will correspond respectively to the occurrence of s_n in group Q_A or in group Q_B. For this investigation, we considered a simplified version of the Fritchman model in which there is a single error state (i.e., $k = N - 1$) and no transitions are allowed between any error-free states. The model state transition diagram is shown in Fig. 2. This model has certain advantages toward channel characterization. One advantage is that efficient methods are available for estimating the model parameters by observing that the error-gap distribution is sufficient to uniquely specify a single-error-state Markov model. Another attractive feature is that a large variety of digital channels can be represented by appropriate definitions of the model's state transition probabilities [18]. For example, the Gilbert model can be considered a special case of the Fritchman model with two error-free states ($k = 2$) and an error state [13].

The principal difficulty in using Fritchman's channel characterization is that the state transition probabilities are not directly observable, so methods of deducing them from easily measured error statistics must be derived. The measurement data considered here is the error-gap distribution, denoted by $P(0^l|1)$, that gives the probability that at least l successive error-free bits will be encountered next on the condition that an error bit has just occurred. Assuming a stationary Markov chain, Fritchman [13] showed that the error-gap distribution can be expressed as the sum of $(N-1)$ exponentials

$$P(0^l|1) = \sum_{i=1}^{N-1} \alpha_i \beta_i^l \qquad (4)$$

where the values of α_i and β_i are referred to as the Fritchman model parameters. The relationship between these parameters and the state transition probabilities can be found in [13]. Proceeding in this way, the original descriptive modeling issue can be formulated as a constrained optimization problem in which

$(N-1)$ pairs of parameters $\{\alpha_i, \beta_i\}$ are the optimization variables to be identified. For this investigation, the sum of the squared errors between the measured error-gap distribution and its modeled fit was considered a suitable cost function. The least square approximation method leads to a constrained nonlinear optimization problem that can be stated as follows:

$$\min_{\alpha, \beta} E_g = \min_{\alpha, \beta} \sum_{l=1}^{L} \frac{1}{l} \left[\log_{10} P(0^l|1) - \log_{10} \sum_{i=1}^{N-1} \alpha_i \beta_i^l \right]^2 \qquad (5)$$

where $0 \leq \alpha_i, \beta_i \leq 1$, for $i = 1, 2, \ldots, N-1$, and L gives the longest interval between two consecutive errors.

IV. PARAMETER ESTIMATION IN GILBERT AND FRITCHMAN MODELS

The gradient-descent algorithm [18] is often used to identify optimal values of Gilbert and Fritchman model parameters for which the cost function E_g in (5) is minimal. Starting with a single point in the search space, the gradient-descent method tries to find the optimum solution by performing successive corrections on parameter estimates in the direction opposite to the gradient of the cost function. While this approach converges rapidly, its simple downhill search transitions can easily become trapped in local minima and, thus, miss finding the globally optimal solution. Thus, it is common practice to run the gradient search starting from a number of initial configurations and then to choose the best outcome from all those obtained. However, the computational burden can be intolerable and there would still be no guarantee that an optimal solution will be found. In an earlier work [21], we presented preliminary experimental results showing that the simulated annealing technique [22] is preferable to the gradient method for use in estimating Gilbert model parameters. However, for Fritchman channel modeling problems, the error-performance surface tends to exhibit many different convex regions and has been found difficult to optimize by means of the simulated annealing algorithm.

An alternative approach to function optimization is based on genetic algorithms (GAs) [19]. The main attraction of GAs arises from the fact that the given search space is explored in parallel by means of iterative modifications of a population of individuals. Each individual, called a chromosome, represents a potential solution to a given problem. A block diagram of a typical genetic evolution is shown in Fig. 3. Choosing an appropriate representation of chromosomes is the first step in applying GAs to solving optimization problems and one that conditions all subsequent steps of the implementation. Here, $(N-1)$ pairs of parameters $\{\alpha_i, \beta_i\}$ define the solution and hence, can be encoded into a chromosome as a list of real numbers, that is: $\mathbf{S} = [\alpha_1, \beta_1, \ldots, \alpha_{N-1}, \beta_{N-1}]$. This real-coded representation is more accurate for continuous optimization problems [23] and allows us to hybridize with problem-dependent heuristics to make an efficient implementation. The fitness values of all chromosomes were ranked with respect to the objective function $F(\mathbf{S}) = 1/E_g$ using (5). As a result of this evaluation, a particular group of chromosomes were selected from the population to generate offspring by subsequent recombination. To prevent premature convergence
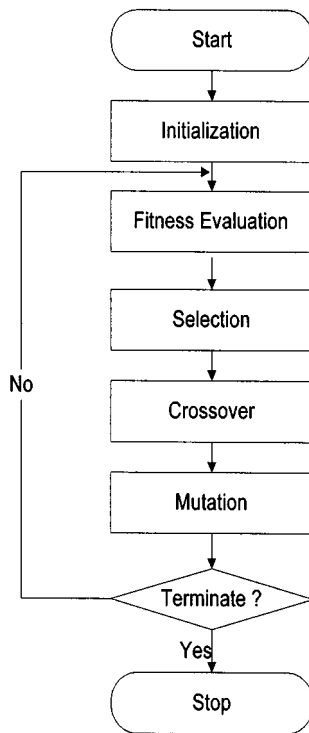
Fig. 3.   Flow diagram for the GA.

TABLE I
RESULTS FROM THE GRADIENT-DESCENT AND GENETIC ALGORITHM
APPROACHES FOR ESTIMATING GILBERT MODEL PARAMETERS

	Gradient-descent	Genetic Algorithm	
		Uniform Mutation	Annealing Mutation
Run #1	0.154968	0.043520	0.036368
Run #2	0.112683	0.038849	0.036368
Run #3	0.115964	0.036393	0.036368
Run #4	0.054852	0.040656	0.036368
Run #5	0.034670	0.036971	0.036368
Average	0.100627	0.039278	0.036368

of the population, this investigation employed the linear ranking selection scheme [24] that sorted individuals by increasing order of fitness, and then assigned the expected number of offspring according to their relative ranking. Crossover among the selected chromosomes then proceeded by exchanging substrings of two chromosomes between two randomly selected crossover points. The crossover probability, denoted by p_c, is defined as the ratio of the number of offspring produced in each generation to the population size. After crossover, with a probability of p_m, the mutation operator was used to introduce random variations into the genetic structure of the chromosomes. When the maximum number of generations was reached, the best chromosome in the final population was taken as GA's solution for functional optimization.

Having a proper mutation operator is critical for both the convergence rate and the final performance of real-coded GAs. A usual way to achieve mutation is to generate a random number and then replace one of the gene values of an existing chromosome with a probability of p_m. However, GAs with such a uniform mutation find it difficult to perform a fine local search when applied to multidimensional parameter optimization problems. Initiated by the merits of simulated annealing [25], we propose a new mutation operator that uses the probabilistic acceptance test technique internally to determine whether to accept the mutated solution, or to stay with the previous solution. More specifically, mutated chromosomes with higher fitness values are always included in the next generation, whereas those with inferior values are only accepted with a certain probability that decreases as the effective temperature is decreased. This implies that more diversity among chromosomes can be introduced to allow a random search of the solution space in the early stage of the optimization process and, in later stages, the chances

of fitter chromosomes being replaced become less. Thus, the search will move toward some feasibility regions likely to contain the global optimum as the temperature gradually decreases, but escapes from local optima are still possible since inferior solutions are allowed. Here, we refer to $\mathbf{S}$ as the chromosome to be mutated and to the T_n as the temperature at the nth generation. The temperature is decreased according to the cooling schedule $T_n = T_0/(n+1)$ [26], where T_0 is the initial temperature. Relevant aspects of the annealing mutation scheme are summarized as below.

1) Apply mutation to chromosome $\mathbf{S}$, if the value of a random number generated between 0 and 1 is less than the specified mutation probability p_m. Otherwise, leave the chromosome intact.

2) Perturb chromosome $\mathbf{S}$ to obtain $\mathbf{S}'$ by adding an element of zero-mean noise r to the randomly selected gene s_i. The only difference between $\mathbf{S}$ and $\mathbf{S}'$ is that the ith gene value $s_i' = s_i + r$, where r is normally distributed with a standard deviation of $\sigma = 0.05 T_n$.

3) Calculate the fitness difference between new and old chromosomes, $\Delta F = F(\mathbf{S}') - F(\mathbf{S})$. Accept the new chromosome $\mathbf{S}'$ if it results in a net increase in fitness, i.e., $\Delta F > 0$. If $\Delta F < 0$, the probability of accepting an inferior solution $\mathbf{S}'$ over $\mathbf{S}$ is given by $e^{\Delta F/T_n}$.

To test the validity of various parameter estimation algorithms, a series of sample error sequences were generated using a narrowband mobile radio channel simulator [27]. Each sample error sequence was 100 000 bits long. The channel condition was defined by the Jakes model [28] with the following parameters: carrier frequency= 900 MHz, vehicle speed= 100 km/h, number of low-frequency oscillators= 8, and average SNR= 10 dB. The Jakes model was designed to simulate Rayleigh fading channels and has been adopted by standard groups for use in testing candidate speech coding schemes as well as radio-link error-control protocols [29]. Using this model, we simulated error sequences for optimum differential phase-shift keying (DPSK) modulation at a data rate of 20 kb/s. For each of simulated error sequences, we first evaluated the measured values of $P(0^l|1)$ by computing the ratio of consecutive series of error-free bits with lengths equal to or greater than l to the total number of error bursts. The optimal identification of an N-state Fritchman model was then formulated as the least square approximation of the measured error-gap distribution by summing the $(N-1)$ exponentials.

Experimental results obtained from the gradient-descent and genetic approaches for estimating Gilbert model parameters are presented in Table I. The performances were evaluated in terms

TABLE II
RESULTS FROM THE GRADIENT-DESCENT AND GENETIC ALGORITHM
APPROACHES FOR ESTIMATING FRITCHMAN MODEL PARAMETERS

	Gradient-descent	Genetic Algorithm	
		Uniform Mutation	Annealing Mutation
Run #1	0.039161	0.026896	0.022008
Run #2	0.168571	0.040753	0.022191
Run #3	0.117410	0.031968	0.021973
Run #4	0.062529	0.045399	0.021973
Run #5	0.050357	0.036682	0.021973
Average	0.087606	0.036340	0.022046

Fig. 4. Experimental error-gap distribution $P(0^l|1)$ and its Fritchman modeled fit for an error sequence typical of the Jakes model.

of the cost function E_g in (5), which indicates the the sum of the squared errors between the measured error-gap distribution and the corresponding modeled fit. To compare the algorithms with respect to their consistency, we ran each algorithm five times, each time beginning with a randomly selected parameter setting. In the GA implementation, the parameter values used for the maximum number of generations, the population size, the crossover probability, and the mutation probability were empirically determined to be 2000, 50, 0.6, and 0.1, respectively. As shown in Table I, the estimation process using the gradient-descent method often terminates in an unsatisfactory local minimum. This shortcoming is inherent in optimization algorithms that perform downhill searches along the function E_g consisting not of a unique global minimum but of many local minima. In contrast, the GA has a global searching capability and, therefore, leads to a greater accuracy in the estimation of Gilbert model parameters. It is also important to note that, relative to uniform mutation, the use of annealing mutation further reduces the distortion E_g and the sensitivity to initial parameter setting. This better result is due to the fact that the annealing process allows random searching of the solution space initially and only very local searching in later stages.

We next investigated the performance dependence of channel characterization on the number of exponentials used to approximate the error-gap distribution. Table II presents the results obtained from the curve fitting for the Fritchman model with three error-free states and one error state. A comparison between Tables I and II reveals that the Fritchman model is preferable to the Gilbert model for use in Markov characterization of sample error sequences. This can be explained by (5), which indicates that the number $(N-1)$ of exponentials should be the number of distinct line segments embedded in the measured value, expressed logarithmically, of the error-gap distribution. To elaborate further, we show in Fig. 4 the experimentally measured error-gap distribution $P(0^l|1)$ for an error sequence typical of the Jakes model. Also shown in the figure is the Fritchman modeled fit with parameter values estimated using the GA with annealing mutation. It can be noted that within the error-gap distribution curve, three line segments with different slopes can be distinguished, making it particularly suited for a Fritchman model with three error-free states ($N-1=3$). It is also clear from this figure that the Fritchman modeled fit provides an approximation of the experimental error-gap distribution to a reasonable degree of accuracy.

V. ALGORITHMS FOR DETERMINING THE INDEX ASSIGNMENT

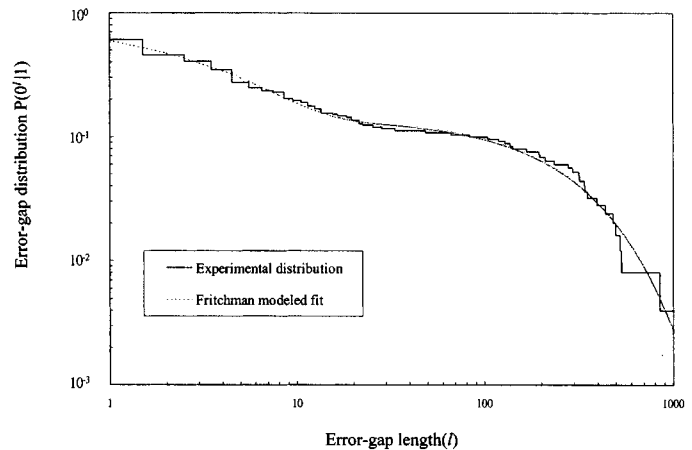For transmission of VQ data over noisy channels, distortions due to channel impairments can be greatly reduced by assigning suitable indices to the codevectors. For a fixed codebook and a given channel, every possible index assignment function γ must be examined to find the one that minimizes the channel distortion $D_c(\gamma)$. This task belongs to the class of NP-complete problems, since there are $M!$ possible combinations of indices for codebooks of size M. With this constraint, an exact search for the optimal index assignment is not feasible and many practical index assignment algorithms are suboptimal. The case of memoryless binary symmetric channel has been considered in conjunction with binary switching algorithm in [5], with simulated annealing algorithm in [6], and with Hadamard transform in [8]. Extension of these results to wireless channels requires that channel transition probabilities be carefully derived to account for the statistical dependencies between error occurrences. This motivates us to seek to incorporate the Fritchman channel modeling into the index assignment design.

A distinctive feature of the Fritchman model lies in its channel transition probabilities that have closed-form expressions represented in terms of model parameters. Let $\{\gamma(\mathbf{c}_i), \ 1 \leq i \leq M\}$ denote the set of codevector indices, with $\gamma(\mathbf{c}_i) = (\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{im})$ being the binary expansion of the index associated with the codevector $\mathbf{c}_i$. For the Fritchman model with $(N-1)$ error-free states and one error state, the effective BER ε is given by [31]

$$\varepsilon = \frac{1}{1 + \sum_{i=1}^{N-1}(\alpha_i\beta_i/1 - \beta_i)} \qquad (6)$$

and the channel transition probabilities are expressed as

$$P(\gamma(\mathbf{c}_j)|\gamma(\mathbf{c}_i)) = \pi \prod_{k=1}^{m} \mathbf{P}_e(\gamma_{ik} \oplus \gamma_{jk})\mathbf{1} \qquad (7)$$

where $\mathbf{1}$ is a vector of ones and the initial state probabilities π in the form

$$\pi = \left[\varepsilon\frac{\alpha_1\beta_1}{1-\beta_1}, \varepsilon\frac{\alpha_2\beta_2}{1-\beta_2}, \ldots, \varepsilon\frac{\alpha_{N-1}\beta_{N-1}}{1-\beta_{N-1}}, \varepsilon\right]. \qquad (8)$$

In (7), the term $(\gamma_{ik} \oplus \gamma_{jk})$ only assumes the values of 0 or 1, and the corresponding matrix $\mathbf{P}_e$ is given by

$$
\mathbf{P}_e(0) = \begin{bmatrix} \beta_1 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & \cdots & \beta_{N-1} & \vdots \\ \alpha_1\beta_1 & \cdots & \alpha_{N-1}\beta_{N-1} & 0 \end{bmatrix} \quad (9)
$$

or

$$
\mathbf{P}_e(1) = \begin{bmatrix} & & 1-\beta_1 \\ \mathbf{0}_{(N-1)\times(N-1)} & & \vdots \\ & & 1-\beta_{N-1} \\ 0\cdots 0 & & 1-\sum_{i=1}^{N-1}\alpha_i\beta_i \end{bmatrix}. \quad (10)
$$

We next developed an algorithm to use with the Fritchman model for optimizing the index assignment over channels with memory. The optimization approach taken here is the GA which performs global exploration among a population of chromosomes. A similar approach to index assignment that uses a parallel GA was introduced in [7] but with the major difference of considering memoryless binary symmetric channels. The parallel GA operates by dividing the total population into a number of smaller subpopulations and then executing the main loop of the GA on each subpopulation separately. A few reasons can be listed concerning our choice of the GA over the parallel GA for functional optimization. First, the issue of whether it is better to use a single large population or multiple small subpopulations in the GA implementation does not appear to be settled and may be problem dependent. Second, no matter which approach is used, incorporating the Fritchman model into the index assignment can serve to better track the intrinsic natures of channel errors. The input to the proposed algorithm consists of a domain-specific codebook and Fritchman model parameters for the given channel that will carry the codevector indices. The algorithm eventually halts and gives as output the index assignment that minimizes the channel distortion $D_c(\gamma)$.

While the basics of the operation of the GA were illustrated in Fig. 3, it requires some elaboration. The search space of interest is a set of permutations of the indices; encoding each chromosome as a list of indices that number the representative codevectors can solve the representation problem. For GAs with such a permutation representation, special care must be taken to ensure that genetic operators will not yield illegal offspring in which some indices are missed while other indices are duplicated. The proposed GA for solving the index assignment problem consists of the following steps.

1) Randomly generate an initial population of chromosomes, each of which corresponds to one particular index assignment function, i.e., $\mathbf{S} = [\gamma(\mathbf{c}_1), \gamma(\mathbf{c}_2), \dots, \gamma(\mathbf{c}_M)]$.
2) Evaluate the fitness of each chromosome $\mathbf{S}$ by the objective function $F(\mathbf{S}) = 1/D_c(\gamma)$, where $D_c(\gamma)$ is calculated by substituting (7) in (1).
3) Assign the expected number of offspring to each chromosome according to the linear ranking selection scheme [24].

TABLE III
CHANNEL DISTORTION USING VARIOUS VECTOR QUANTIZERS IN FIVE DIFFERENT RUNS OF THE BINARY SWITCHING ALGORITHM, GENETIC ALGORITHM AND FOR THE ENSEMBLE AVERAGE OF $M!$ CODES

| | $\rho = 0.0$ | | | | $\rho = 0.5$ | | | |
| | J=4 | | J=6 | | J=4 | | J=6 | |
	BSA	GA	BSA	GA	BSA	GA	BSA	GA
Run #1	0.1245	0.1245	0.2329	0.2208	0.1486	0.1442	0.2311	0.2247
Run #2	0.1260	0.1244	0.2294	0.2212	0.1506	0.1441	0.2346	0.2242
Run #3	0.1305	0.1245	0.2258	0.2213	0.1462	0.1442	0.2289	0.2251
Run #4	0.1298	0.1245	0.2359	0.2214	0.1458	0.1441	0.2294	0.2254
Run #5	0.1253	0.1245	0.2288	0.2204	0.1464	0.1441	0.2310	0.2243
Average	0.1272	0.1245	0.2306	0.2210	0.1475	0.1441	0.2310	0.2247
Random	0.2378		0.4812		0.2811		0.5361	

4) Apply the crossover to the selected chromosomes when the value of a random number generated between 0 and 1 is less than the crossover probability p_c. The partially-matched crossover operator [30] is adopted here. This crossover is performed by exchanging the substrings between two randomly selected positions and then resolving conflicting assignments with a repairing procedure.
5) With a probability of p_m, the mutation is applied to each chromosome by inverting its substring between two randomly selected positions.
6) Repeat steps 2 to 5 until the maximum number of generations has been reached. After termination, the best chromosome in the final population is taken as the optimal index assignment.

VI. INDEX ASSIGNMENT PERFORMANCE COMPARISONS AND RESULTS

Experiments were carried out to investigate the potential advantages of using genetic algorithms to improve the robustness of nonredundant VQ coding systems. The input signals considered here include first order Gauss–Markov sources described by $x(n) = \rho x(n-1) + w(n)$, where $w(n)$ is zero-mean, unit-variance white Gaussian noise, with correlation coefficients of $\rho = 0$ and $\rho = 0.5$. Each of these was tested with rate $R = 1$ bit/sample vector quantizers having the following codebook sizes and dimension values $(M, J): (16, 4), (64, 6), (256, 8)$. The codebooks were designed for a noiseless channel using the standard generalized Lloyd algorithm [20]. Distortions between the source sequence and each possible codevector were calculated, and the codevector with the smallest distortion was selected as the best fit. Table III presents the vector quantization results associated with various index assignment algorithms for the case where the bits in the codevector indices are subjected to the sample error sequences described in Section IV. The performances of the binary switching algorithm (BSA) [5] and the GA were measured in terms of channel distortion $D_c(\gamma)$ and compared with the ensemble-averaged distortion yielded by random index assignment. The parameter values used in the GA implementation for the maximum number of generations, the population size, the crossover probability, and the mutation probability were 500, 400, 0.6, and 0.2, respectively. Numerical results obtained from the BSA and GA are presented for five runs of each algorithm, each run beginning with a randomized index assignment. The results of these experiments clearly demonstrate the improved performance achievable using the BSA and GA
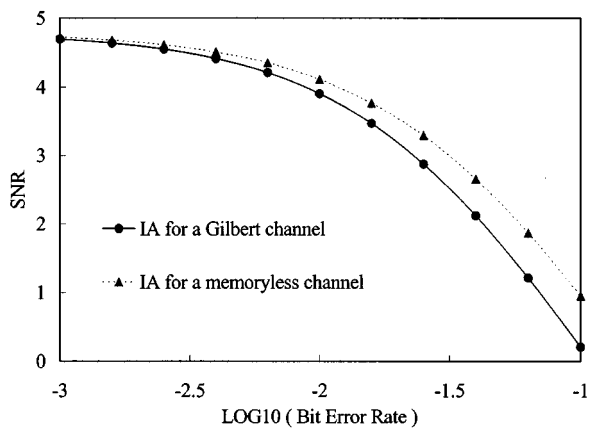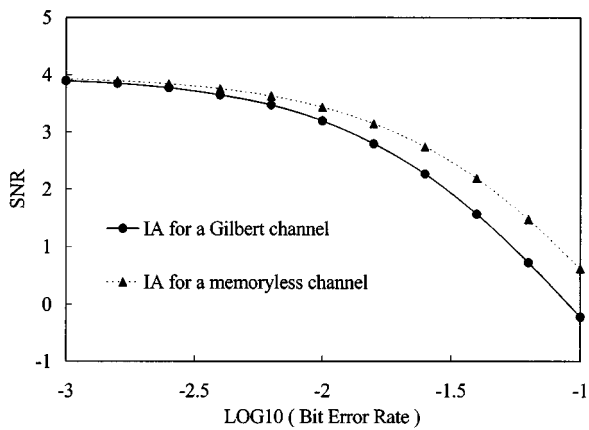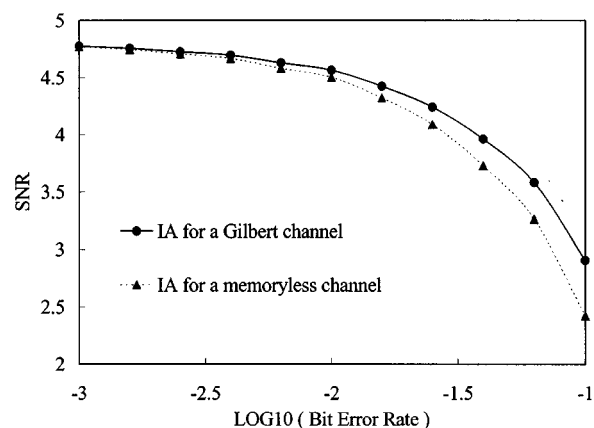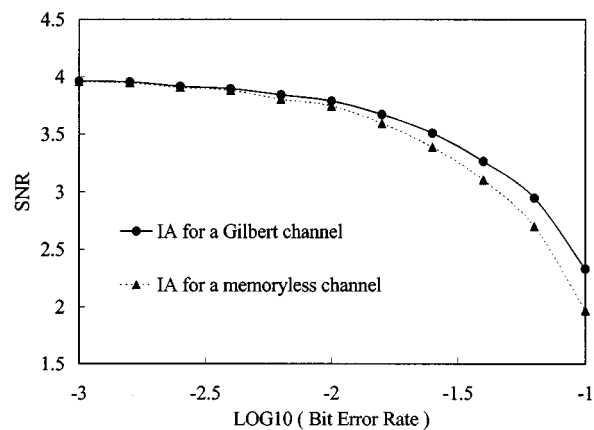
Fig. 5. SNR(dB) performance of a vector quantizer with index assignment (IA) on a memoryless channel. (a) Codebook size = 256, vector dimension = 8, source = Gaussian i.i.d. (b) Codebook size = 256, vector dimension = 8, source = 1st-order Gauss-Markov ($\rho = 0.5$).

Fig. 6. SNR(dB) performance of a vector quantizer with IA on a Gilbert channel. (a) Codebook size = 256, vector dimension = 8, source = Gaussian i.i.d. (b) Codebook size = 256, vector dimension = 8, source = 1st-order Gauss-Markov ($\rho = 0.5$).

in comparison to the performance obtained from random index assignment. Furthermore, the improvement has a tendency to increase for larger codebook sizes and for more heavily correlated Gaussian sources. This observation is in accord with the results produced by the memoryless binary symmetric channel, reported earlier [5]–[7]. Compared with the BSA, the effectiveness of the GA for use in solving the minimum distortion index assignment problem is clearly demonstrated. The main reason for this is that the BSA suffers from the problem of convergence toward locally optimal solutions that are critically dependent upon the initial index assignment [5]. By contrast, the global explorative properties of GA can help to find a solution that shows greater accuracy and lower sensitivity to the choice of initial index assignment. This explains why the genetic search strategy converges to a point where the corresponding channel distortion is more consistent from run to run. Indeed, for cases in which $J = 4$, the genetic-type index assignment yielded the same result in all the tests.

The next step in the present investigation concerned the performance degradation that may result from using genetic-type index assignment under channel mismatch conditions. To illustrate some of the results, Figs. 5 and 6 compare the overall SNR of the vector quantizer for the memoryless and the Gilbert channels, respectively. These results are presented for

a vector quantizer with a codebook size and vector dimension of $(M, J) = (256, 8)$. The SNRs were measured for the codebooks before and after applying genetic-type index assignment for the channel BERs ranging from 10^{-3} to 10^{-1}. Each graph shows results that compare the case where the index assignment was designed for a memoryless channel against the case that was designed for the Gilbert channel. Simulation results indicate that the accuracy of the channel model used in developing the index assignment algorithm is extremely important to the performance of the vector quantizer. The investigation further showed that for higher channel BERs, the increase in SNR is more noticeable when an index assignment matches real channel behavior.

VII. CONCLUSION

This study presents a novel means of exploiting Markov characterization of error sequences in the design of a robust vector quantizer for noisy channels with memory. We first emphasized the importance of matching the real channel behavior to the channel model on which the index assignment design is based. This task was accomplished by using finite-state Markov chain models to characterize the statistical dependencies in relative occurrences of errors. Simulation results indicate that a hybrid

strategy incorporating an annealing mutation operator into the real-coded genetic algorithm leads to greater accuracy in estimating Gilbert and Fritchman model parameters. Our method coincides with the optimization process of fitting mixtures of exponential functions to experimental error-gap distributions. Finally, the merits of using genetic algorithms for assigning binary indices to VQ codevectors were explored. We concluded that with the aid of Fritchman channel characterization the index assignment algorithm can be developed to better track the intrinsic natures of channel errors. While we only addressed the index assignment problem associated with designing a robust VQ, the design techniques used to refine the channel transition probabilities can be applied to the channel-optimized VQ design as well.

ACKNOWLEDGMENT

The authors are very grateful to the unknown reviewers and the associate editor, L. Hanzo, for their careful readings of this paper and their constructive suggestions.

REFERENCES

[1] R. M. Gray, "Vector quantization," *IEEE Acoust. Speech, Signal Processing Mag.*, vol. 1, pp. 4–29, Apr. 1984.
[2] P. Hedelin, P. Knagenhjelm, and M. Skoglund, "Vector quantization for speech transmission," in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds. Amsterdam, The Netherlands: Elsevier, 1995.
[3] ——, "Theory for transmission of vector quantization data," in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds. Amsterdam, The Netherlands: Elsevier, 1995.
[4] J. R. B. De Marca and N. S. Jayant, "An algorithm for assigning binary indices to the codevectors of a multidimensional quantizer," in *Proc. IEEE Int. Conf. Communications*, Seattle, WA, June 1987, pp. 1128–1132.
[5] K. Zeger and A. Gersho, "Pseudo-Gray coding," *IEEE Trans. Commun.*, vol. 38, pp. 2147–2158, Dec. 1990.
[6] N. Farvardin, "A study of vector quantization for noisy channels," *IEEE Trans. Inform. Theory*, vol. 36, pp. 799–809, July 1990.
[7] J. S. Pan, F. R. McInnes, and M. A. Jack, "Application of parallel genetic algorithm and property of multiple global optima to VQ codevector index assignment for noisy channels," *Electron. Lett.*, vol. 32, pp. 296–297, Feb. 1996.
[8] P. Knagenhjelm and E. Agrell, "The Hadamard transform — A tool for index assignment," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1139–1151, July 1996.
[9] N. Phamdo, F. Alajaji, and N. Farvardin, "Quantization of memoryless and Gaussian sources over binary Markov channels," *IEEE Trans. Commun.*, vol. 45, pp. 668–675, June 1997.
[10] H. S. Wang and N. Moayeri, "Modeling, capacity, and joint source/channel coding for Rayleigh fading channels," in *Proc. IEEE Vehicular Technology Conf.*, Secaucus, NJ, May 1993, pp. 473–479.
[11] V. Kafedziski and D. Morell, "Vector quantization over Gaussian channels with memory," in *Proc. IEEE Int. Conf. Communications*, Seattle, WA, June 1995, pp. 1433–1437.
[12] M. Skoglund and T. Ottosson, "Joint equalization and soft decoding for vector quantizer over channels with intersymbol interference," in *Proc. IEEE Int. Conf. Communications*, Dallas, TX, June 1996, pp. 1025–1029.
[13] B. D. Fritchman, "A binary channel characterization using partitioned Markov chains," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 221–227, Apr. 1967.
[14] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell Syst. Tech. J.*, vol. 39, pp. 1253–1265, Sept. 1960.
[15] L. N. Kanal and A. R. K. Sastry, "Models for channels with memory and their applications to error control," *Proc. IEEE*, vol. 66, pp. 724–744, July 1978.
[16] T. Sato, K. Tokuda, M. Kawabe, and T. Kato, "Simulation of burst error models and an adaptive error control scheme for high speed data transmission over analog cellular systems," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 443–452, May 1991.
[17] A. I. Drukarev and K. P. Yiu, "Performance of error-correcting codes on channels with memory," *IEEE Trans. Commun.*, vol. COM-34, pp. 513–521, June 1986.
[18] J. Y. Chouinard, M. Lecours, and G. Y. Delisle, "Estimation of Gilbert's and Fritchman's models parameters using the gradient method for digital mobil radio channels," *IEEE Trans. Veh. Technol.*, vol. 37, pp. 158–166, Aug. 1988.
[19] D. E. Goldberg, *Genetic Algorithm in Search, Optimization and Machine Learning*. New York: Addison-Wesley, 1989.
[20] 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.
[21] T. H. Tan and W. W. Chang, "Estimation of the Gilbert model parameters using the simulated annealing method," *Electron. Lett.*, vol. 32, pp. 1256–1258, July 1996.
[22] E. Aats and J. Korst, *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*. New York: Wiley, 1989.
[23] L. J. Eshelman and J. D. Schaffer, "Real-coded genetic algorithms and interval-schemata," in *Foundation of Genetic Algorithms*, G. J. E. Rawlins, Ed. San Mateo, CA: Morgan Kaufmann, 1991, pp. 205–218.
[24] J. E. Baker, "Adaptive selection methods for genetic algorithms," in *Proc. 1st Int. Conf. Genetic Algorithms and Their Applications*, Pittsburgh, PA, July 1985, pp. 101–111.
[25] S. Kirkpatrick, C. D. Gelartt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.
[26] H. Szu and R. Hartley, "Fast simulated annealing," *Phys. Lett. A*, vol. 122, pp. 157–162, 1987.
[27] K. Pahlavan and A. H. Levesque, *Wireless Information Networks*. New York: Wiley, 1995.
[28] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.
[29] A. H. Levesque and P. J. Kush, "Evaluation and correction of the fading model used in the error mask generator computer program," Data Services Task Group, Contribution no. TR45.3.2.5/93.02.08.06, TIA/TR45.3, 1993.
[30] D. Goldberg and R. Lingle, "Alleles, loci and the traveling salesman problem," in *Proc. 1st Int. Conf. Genetic Algorithms*, J. Grefenstette, Ed., Hillsdale, NJ, July 1985.
[31] W. Turin, *Performance Analysis of Digital Transmission Systems*. Rockville, MD: Computer Science, 1990.

Wen-Whei Chang (S'86-M'89) received the B.S. degree in communication engineering from National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 1980, and the M.Eng. and Ph.D. degrees in electrical engineering from Texas A & M University, College Station, TX, in 1985 and 1989, respectively.

Since August 1989, he has been an associate professor with the Department of Communication Engineering, National Chiao Tung University, Hsinchu, Taiwan, R.O.C. His current research interests include speech processing, language identification, and secure communication.

Tan-Hsu Tan received the B.S. degree from National Taiwan Institute of Technology and M.S. degree from National Tsing Hua University, in 1983 and 1988, respectively, both in electrical engineering, and the Ph.D. degree in electronics engineering from National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 1998.

Since 1988, he has been with the Department of Electrical Engineering, National Taipei University of Technology, Taipei, Taiwan, R.O.C., where he is currently an associate professor. His research interests include mobile radio communications, speech signal processing, and optimization algorithms.

De-Yu Wang received the B.S. and M.S. degrees in electrical engineering from Chung Cheng Institute of Technology in 1986 and 1991, respectively, and the Ph.D. degree in communication engineering from National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 1999.

From 1991 to 2000, he was an assistant researcher with Chung Shan Institute of Science and Technology, Taoyuan, Taiwan. In August 2000, he joined the Department of Electrical Engineering, Nankai College of Technology and Commerce, Nantou, Taiwan, R.O.C., as an assistant professor. His research interests include speech coding, speech processing, and voice security.