

Genetic Algorithm Optimization for Blind Channel Identification with Higher Order Cumulant Fitting

S. Chen, Y. Wu, and S. McLaughlin

Abstract—An important family of blind equalization algorithms identify a communication channel model based on fitting higher order cumulants, which poses a nonlinear optimization problem. Since higher order cumulant-based criteria are multimodal, conventional gradient search techniques require a good initial estimate to avoid converging to local minima. We present a novel scheme which uses genetic algorithms to optimize the cumulant fitting cost function. A microgenetic algorithm implementation is adopted to further enhance computational efficiency. As is demonstrated in computer simulation, this scheme is robust and accurate and has a fast convergence performance.

Index Terms—Blind channel identification, genetic algorithm, higher order cumulant fitting.

I. INTRODUCTION

MANY digital communication channels are subject to intersymbol interference (ISI) [1]–[3]. The ISI distortion of a channel is due to the restricted bandwidth allocated to the channel or the presence of multipath effects in the transmission medium. At the receiver, the ISI must be compensated to reconstruct the transmitted data symbols, and this is referred to as channel equalization. Traditionally, an equalizer learns the channel characteristics through an initial training period. During training, the transmitter sends out a data sequence that is known and available in proper synchronism at the receiver. The receiver uses this sequence as the desired reference signal to adjust the equalizer parameters. Once the channel characteristics have been learned sufficiently well, the equalizer can then switch to a decision-directed mode with the detected symbol sequence serving as the reference signal. This decision-directed adaptation allows an equalizer to track slow variations in the channel characteristics during actual data transmission.

Training consumes valuable bandwidth resource and should be avoided whenever possible. Furthermore, in multipoint broadcast communication systems, it is impossible to have a training period, and an equalizer must adjust itself based only on noisy channel observations without access to the desired reference. This is known as blind equalization. Blind equalization techniques developed since the pioneering work of Sato [4] can be categorized into three classes that either rely on: 1) Busgang-type techniques [4]–[9]; 2) higher order cumulants (HOC's) or equivalent higher order spectra [10]–[15]; or 3) a joint channel and data estimation approach [16]–[21].

Manuscript received November 4, 1996; revised May 16, 1997, November 26, 1997, and December 15, 1997.

S. Chen and Y. Wu are with the Department of Electrical and Electronic Engineering, University of Portsmouth, Portsmouth PO1 3DJ U.K.

S. McLaughlin is with the Department of Electrical Engineering, University of Edinburgh, Edinburgh EH9 3JL U.K.

Publisher Item Identifier S 1089-778X(97)09409-5.

Blind equalizers based on HOC techniques are very general and often achieve excellent practical results, provided that sufficient signal samples are available to accurately estimate higher order statistics.

For the second family of blind equalizers, a two-stage strategy is usually adopted, which first identifies a channel model using HOC fitting algorithms and then employs the estimated channel model to design an equalizer [14], [15]. The key step of this approach is its ability to obtain an accurate channel model. Once this model is available, a variety of existing equalizer design methods can be employed, ranging from a simple linear inverse filter to a sophisticated maximum likelihood sequence estimator [1], [22]–[24], depending on a tradeoff between performance and complexity. HOC fitting algorithms for blind channel identification are widely used in practice due to their simplicity and computational efficiency.

Cost functions of HOC fitting, however, are multimodal, and conventional gradient optimization techniques [14], [15] may converge to “wrong” solutions unless a good initial value for the channel parameters is provided, which is not always possible. To overcome the problem of local minima, simulated annealing (SA) has been implemented to optimize the HOC fitting cost function [25]. Here we propose genetic algorithms (GA's) [26]–[31] for blind channel identification based on HOC fitting. For this particular application, we believe that the GA approach is more efficient than SA. Our simulation study demonstrates that the GA-based scheme is robust to errors in estimated HOC's and can achieve a global optimal solution regardless of initial value of channel estimate. Furthermore, the number of parameters to be optimized in the problem of blind channel estimation is usually small, and GA's are particularly effective for this kind of optimization problems. The micro-GA (μ GA) implementation [30] is adopted in our scheme to further improve the convergence rate.

II. BACKGROUND

Typically, a digital communication channel can be modeled as a finite impulse response filter (also known as a moving average model) with an additive noise source [1]–[3]. Specifically, the received signal at sample k is given by

$$r(k) = \hat{r}(k) + e(k) = \sum_{i=0}^{n_a} a_i s(k-i) + e(k) \quad (1)$$

where $\hat{r}(k)$ denotes the noiseless channel output, n_a is the channel order, a_i are the channel taps, the symbol sequence $\{s(k)\}$ is independent and identically distributed, and $e(k)$ is a Gaussian white noise sequence with zero mean and variance $E[e^2(k)] = \sigma_e^2$. We assume that the channel and symbols

are real valued. This corresponds to the use of a multilevel pulse amplitude modulation (M -PAM) scheme with a symbol constellation defined by the set

$$s_i = 2i - M - 1, 1 \leq i \leq M. \quad (2)$$

For example, an 8-PAM scheme has a data symbol set $\{-7, -5, -3, -1, 1, 3, 5, 7\}$. Extension to complex-valued channels and modulation schemes is straightforward. The reason for concentrating on the simpler real-valued case is to avoid complicated communication terminologies, which some readers might not be familiar with.

In the ideal situation where the ISI is absent, the received signal is given by

$$r(k) = s(k - d) + e(k). \quad (3)$$

The optimal decision process in this case is trivial and can readily be shown to be

$$\hat{s}(k - d) = \begin{cases} s_1, & r(k) < -M + 2 \\ s_i, & 2i - M - 2 \leq r(k) < 2i - M, \\ & 1 < i < M \\ s_M, & r(k) \geq M - 2 \end{cases} \quad (4)$$

where $\hat{s}(k - d)$ denotes the estimate of $s(k - d)$. The ISI distortion refers to the fact that, in reality, the received signal is a mixture of several transmitted data symbols. Even without noise, the threshold decision rule (4) is no longer reliable because data symbols transmitted before and after $k - d$ will interfere with the decision regarding the transmitted symbol $s(k - d)$. If the channel model

$$\mathbf{a} = [a_0 a_1 \cdots a_{n_a}]^T \quad (5)$$

is known, this ISI distortion can be removed or minimized by employing an equalizer. Equalizer design given a known channel model is a well-developed field, and a variety of techniques are available [1], [22]–[24]. The channel model is generally unknown, however, and has to be identified first.

Blind channel identification refers to the determination of the channel model \mathbf{a} using only the noisy received signal $\{r(k)\}$ and some prior knowledge of statistical properties of $s(k)$. The problem is particularly difficult because the transfer function of the channel

$$A(z) = \sum_{i=0}^{n_a} a_i z^{-i} \quad (6)$$

is generally nonminimum phase. $A(z)$ is said to be minimum phase if all the zeros of $A(z)$ are inside the unit circle of the z -plane. If $A(z)$ has zeros outside the unit circle, it is nonminimum phase. For nonminimum phase channels, methods based on second-order statistics fail to work completely, and higher order statistics have to be employed. The second-order cumulant or autocorrelation function $C_{2,r}(\tau)$ of the received channel output sequence $\{r(k)\}$ is well known to be

$$\begin{aligned} C_{2,r}(\tau) &= \mathbb{E}[r(k)r(k+\tau)] \\ &= \sigma_s^2 \left(\sum_{i=\max\{0, -\tau\}}^{\min\{n_a, n_a - \tau\}} a_i a_{i+\tau} \right) + \delta(\tau) \sigma_e^2 \end{aligned} \quad (7)$$

where $\sigma_s^2 = \mathbb{E}[s^2(k)]$ is the symbol variance and $\delta(0) = 1$ and $\delta(\tau) = 0$ for $\tau \neq 0$. The fourth-order cumulant sequence $C_{4,r}(\tau_1, \tau_2, \tau_3)$ of $\{r(k)\}$ is defined as

$$\begin{aligned} C_{4,r}(\tau_1, \tau_2, \tau_3) &= \mathbb{E}[r(k)r(k+\tau_1)r(k+\tau_2)r(k+\tau_3)] \\ &\quad - \mathbb{E}[r(k)r(k+\tau_1)]\mathbb{E}[r(k+\tau_2)r(k+\tau_3)] \\ &\quad - \mathbb{E}[r(k)r(k+\tau_2)]\mathbb{E}[r(k+\tau_1)r(k+\tau_3)] \\ &\quad - \mathbb{E}[r(k)r(k+\tau_3)]\mathbb{E}[r(k+\tau_1)r(k+\tau_2)]. \end{aligned} \quad (8)$$

It can be shown that [32]

$$C_{4,r}(\tau_1, \tau_2, \tau_3) = \gamma_{4,s} \sum_{i=L}^U a_i a_{i+\tau_1} a_{i+\tau_2} a_{i+\tau_3} \quad (9)$$

where

$$\begin{aligned} L &= \max\{0, -\tau_1, -\tau_2, -\tau_3\} \\ U &= \min\{n_a, n_a - \tau_1, n_a - \tau_2, n_a - \tau_3\} \end{aligned} \quad (10)$$

and

$$\gamma_{4,s} = C_{4,s}(\tau_1, \tau_2, \tau_3)|_{\tau_1=\tau_2=\tau_3=0} \quad (11)$$

is the kurtosis of the transmitted symbol sequence $\{s(k)\}$. The four-order cumulant is considered because third-order cumulants of symmetric sequences are zero. The variance σ_s^2 and the kurtosis $\gamma_{4,s}$ are known in our application.

As is well known, the autocorrelation functions do not carry any phase information but the HOC's are very sensitive to phase properties. The frequency response of the transfer function $A(z)$ is defined by

$$\begin{aligned} A(\omega) &= A(z)|_{z=\exp(j\omega)} = |A(\omega)| \exp(j\angle A(\omega)), \\ \omega &\in [0, 2\pi) \end{aligned} \quad (12)$$

where $|A(\omega)|$ is the amplitude response and $\angle A(\omega)$ is the phase response. Consider the simplest case with $A(z) = a_0 + a_1 z^{-1}$. Assume that we have two transfer functions defined by

$$A_1(z) = 1.0 + 0.5z^{-1} \text{ and } A_2(z) = 0.5 + 1.0z^{-1} \quad (13)$$

where $A_1(z)$ is minimum phase since its zero $z_0 = -0.5$ is inside the unit circle, and $A_2(z)$ is nonminimum phase as its zero $z_0 = -2.0$ is outside the unit circle. Both have the same amplitude response $\sqrt{1.25 + \cos(\omega)}$ but very different phase responses, as illustrated in Fig. 1. In general, among all the $A(z)$'s with the same amplitude response $|A(\omega)|$, the one with minimum phase has the smallest phase deviation. The autocorrelation functions (assuming noise-free) and the diagonal slice of the fourth-order cumulants for the signals produced by the two systems defined in (13) can be calculated, and the results are summarized in Table I. It can be seen that autocorrelations totally lose the phase information of the signal but HOC's can resolve the signal phase. Another advantage of HOC-based techniques is that, unlike second-order statistics,

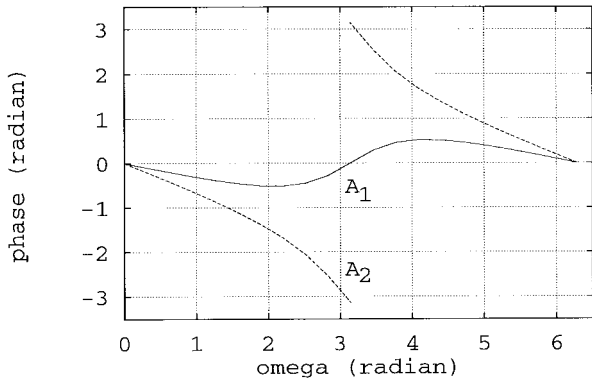


Fig. 1. Phase responses of $A_1(z) = 1.0 + 0.5z^{-1}$ and $A_2(z) = 0.5 + 1.0z^{-1}$.

TABLE I

A COMPARISON OF CUMULANTS AND AUTOCORRELATIONS FOR $A(z) = a_0 + a_1z^{-1}$ WITH $a_0 = 1.0, a_1 = 0.5$ (MINIMUM PHASE) AND $a_0 = 0.5, a_1 = 1.0$ (NONMINIMUM PHASE). THE RECEIVED CHANNEL OUTPUT IS $r(k)$. AUTOCORRELATION FUNCTIONS LOSE THE PHASE INFORMATION OF THE SIGNAL

	$r(k)$	Minimum phase	Nonminimum phase
4th-order cumulants	$C_{4,r}(0, 0, 0)/\gamma_{4,s}$	$(a_0^4 + a_1^4) = 1.0625$	$(a_0^4 + a_1^4) = 1.0625$
	$C_{4,r}(1, 1, 1)/\gamma_{4,s}$	$a_0a_1^3 = 0.125$	$a_0a_1^3 = 0.5$
autocorrelation functions	$C_{2,r}(0)/\sigma_s^2$	$(a_0^2 + a_1^2) = 1.25$	$(a_0^2 + a_1^2) = 1.25$
	$C_{2,r}(1)/\sigma_s^2$	$a_0a_1 = 0.5$	$a_0a_1 = 0.5$

HOC's are insensitive to Gaussian noise. This can readily be verified by comparing (9) with (7).

Assume that N received signal samples $\{r(k)\}_{k=1}^N$ are used to compute the time estimate $\hat{C}_{4,r}(\tau_1, \tau_2, \tau_3)$ of the fourth-order cumulant. $\hat{C}_{4,r}(\tau_1, \tau_2, \tau_3)$ is obtained by replacing the moments in (8) with their respective sample averages based on the N samples of $r(k)$. The optimal channel estimate $\hat{\mathbf{a}}$ can be obtained by minimizing the following cumulant-fitting cost function [15], [25]:

$$J_4(\mathbf{a}) = \sum_{\tau=-n_a}^{n_a} \left(\hat{C}_{4,r}(\tau, \tau, \tau) - \gamma_{4,s} \sum_{i=\max\{0, -\tau\}}^{\min\{n_a, n_a - \tau\}} a_i a_{i+\tau}^3 \right)^2. \quad (14)$$

In the construction of the fourth-order cumulant-based cost function (14), only the diagonal slice of $\hat{C}_{4,r}(\tau_1, \tau_2, \tau_3)$ is used as this results in much reduced computational complexity. Moreover, it has been found in practice that fitting cumulants over their entire region of support does not give any significant improvement in channel model estimation. It is obvious that the accuracy of the estimated cumulant $\hat{C}_{4,r}(\tau_1, \tau_2, \tau_3)$ will influence the result of this approach. To reduce the estimation variance of $\hat{C}_{4,r}(\tau_1, \tau_2, \tau_3)$, the segment-average scheme [11] is often used to compute $\hat{C}_{4,r}(\tau_1, \tau_2, \tau_3)$. In this approach, blind channel estimation is formulated as a standard optimization problem with the cost function $J_4(\mathbf{a})$. This is attractive since the concepts and principles of the related optimization are widely understood.

With the exception of [25], existing algorithms for HOC fitting employ gradient search techniques, yet the associated cost functions are well known to be multimodal. Even with

measures of providing good initial channel estimates, it has been observed that gradient algorithms sometimes converge to local minima [14]. Using GA's to optimize the cost function (14) has the advantage of ensuring a global optimal channel estimate in the limit. Moreover, a communication channel model (5) typically contains a few tap coefficients. Thus the number of parameters to be optimized in (14) is small, and GA's are often very efficient in solving this kind of "small-dimensional" optimization problems. We also discovered in simulation that GA's are less sensitive to noisy errors in the time estimate of $\hat{C}_{4,r}(\tau_1, \tau_2, \tau_3)$.

In reality, the channel order n_a is unknown and needs to be identified. Several model-order selection criteria can be applied to determine the correct order n_a [32]–[34], and they will not be repeated here. This model-order selection process will add considerably more computational complexity. A much simpler method is to overfit with an upper bound $\hat{n}_a \geq n_a$. Thus the cost function used in the optimization process is modified as

$$\hat{J}_4(\mathbf{a}) = \sum_{\tau=-\hat{n}_a}^{\hat{n}_a} \left(\hat{C}_{4,r}(\tau, \tau, \tau) - \gamma_{4,s} \sum_{i=\max\{0, -\tau\}}^{\min\{\hat{n}_a, \hat{n}_a - \tau\}} a_i a_{i+\tau}^3 \right)^2. \quad (15)$$

Function (15) is more complicated than (14) because it contains more local minima, and this will generally cause more problems for gradient-based methods. The GA-based method, however, should in principle be capable of identifying those nonexistent taps with (near) zero values, since the true global optimal values for nonexistent taps are zeros. This has been confirmed in simulation. An inspection of the obtained channel estimate will allow deleting those insignificant taps. Thus, the proposed method has an additional advantage of much simpler model-order selection.

III. METHOD

With the goal being to find a global optimum solution as quickly as possible, we adopt the so-called μ GA [30], which appears to offer certain advantages. This version of GA uses a population that is much smaller than typically employed, which can make it less computationally burdensome. In [30], it was reported that the μ GA can find optimal regions faster than standard GA's for selected optimization problems. Allowing a single sequence of a μ GA to converge, however, may not be very useful apart from quickly locating local optima. Therefore, after such convergence, the population is reinitialized randomly while the best individual found up to that point is copied into the newly generated population. This iterative reinitialization is repeated until no further improvement is evidenced.

Each parameter has a search range of $(-1, 1)$ and is coded into a 16-bit string. Although other choices of encoding are clearly available, this is deemed sufficient for the task. Reference [30] utilized a population size of five, but we choose a value roughly two times the number of parameters, in anticipation of facing more difficult problems that may require a larger collection of individuals in the population. The crossover rate is set to 1.0, with the number of crossover

points typically set to four. No mutation is employed, as the reinitialization routine serves to introduce diversity. Tournament selection [30] is employed to determine parents for reproduction.

More specifically, the procedure is summarized as follows.

- Step 1) Given a set of received signal samples $\{r(k)\}_{k=1}^N$, assume an overlength \hat{n}_a and compute the required time estimate of the fourth-order cumulant, $\hat{C}_{4,r}(\tau, \tau, \tau)$ for $-\hat{n}_a \leq \tau \leq \hat{n}_a$, to form the cumulant fitting cost function (15).
- Step 2) With a set of randomly initialized channel parameter vectors $\{\hat{\mathbf{a}}_i\}_{i=1}^{n_p}$, where n_p is the population size, use the μ GA to optimize the cost function (15). The fitness function value f_i corresponding to $\hat{\mathbf{a}}_i$ is defined as

$$f_i = \frac{1}{\hat{J}_4(\hat{\mathbf{a}}_i)}. \quad (16)$$

In the case that $\hat{J}_4(\hat{\mathbf{a}}_i)$ is (near) zero, a very large value is assigned to f_i . The operation of the μ GA involves two loops.

- Step 2.1) The inner loop evaluates and evolves a population. Let f_{best} be the best fitness value of the current population. The population is assumed to have converged if

$$\sum_{i=1}^{n_p} (f_{\text{best}} - f_i) < \alpha f_{\text{best}} \quad (17)$$

where α is a predefined small positive constant.

- Step 2.2) After the convergence of the inner loop, the convergence of the outer loop is tested. Let $\hat{\mathbf{a}}_{\text{best}}$ be the best channel estimate in the current population and $\hat{\mathbf{a}}_{\text{Rec}}$ be the best channel estimate recorded in the previous population. The overall process has converged if

$$\|\hat{\mathbf{a}}_{\text{best}} - \hat{\mathbf{a}}_{\text{Rec}}\| < \delta \sqrt{4(n_a + 1)} \quad (18)$$

where the small positive scalar δ defines the final search accuracy, and $\sqrt{4(n_a + 1)}$ is the Euclidean norm of the search space (as the search range for each parameter is $[-1, 1]$). Otherwise, the population is reinitialized, $\hat{\mathbf{a}}_{\text{Rec}}$ is reset to $\hat{\mathbf{a}}_{\text{Rec}} = \hat{\mathbf{a}}_{\text{best}}$, and the inner loop restarts.

To increase the chance of converging to a true global optimum solution, a more robust outer loop test can be employed which only terminates the overall process after the test (18) has been satisfied for several consecutive times.

A ‘‘normalization’’ measure is developed from (9) to improve the rate of convergence. We make sure that every candidate $\hat{\mathbf{a}}_i$ satisfies

$$\sum_{j=0}^{\hat{n}_a} \hat{a}_{i,j}^4 = \frac{\hat{C}_{4,r}(0, 0, 0)}{\gamma_{4,s}} \quad (19)$$

where $\hat{a}_{i,j}$ is the j th element of $\hat{\mathbf{a}}_i$. In the population initialization, the taps of each channel vector first take values randomly from the interval $(-1, 1)$, and each chosen model is then normalized according to (19). Whenever a new generation is produced, each member of the population is also normalized. This ensures that each member of the population is inside the feasible set of channel models. Without this normalization measure, mating will inevitably produce some population members that are far outside the possible set of candidate channel models, and function evaluations for these members will merely waste computation.

It is well known that sign and time-shift ambiguities exist in blind channel identification based on HOC fitting. Sign ambiguity is manifested by the fact that both the true channel \mathbf{a} and $-\mathbf{a}$ are global optimal solutions of (15) [or (14)]. Time-shift ambiguity can be illustrated in the following example. Let the true channel model be $\mathbf{a} = [a_0 \ a_1 \ \dots \ a_{n_a}]^T$ with $a_0 \neq 0$ and $a_{n_a} \neq 0$. Suppose that $\hat{n}_a = n_a + 2$ is used. Then $[a_0 \ a_1 \ \dots \ a_{n_a} \ 0 \ 0]^T$, $[0 \ a_0 \ a_1 \ \dots \ a_{n_a}]^T$, and $[0 \ 0 \ a_0 \ a_1 \ \dots \ a_{n_a}]^T$ are all global optimal solutions of (15). These ‘‘ambiguity’’ problems are common to blind equalization techniques. A solution to time-shift ambiguity is to fix one of the channel coefficients. We do not fix a tap value but instead check if the first tap of a population member is zero (absolute value smaller than a threshold), and if so, a shifting is performed to ensure that the first tap is always nonzero.

In practice, the convergence performance of the algorithm can only be observed through the best value of the cost function $\hat{J}_4(\hat{\mathbf{a}}_{\text{best}})$. In simulation, the performance of the algorithm can also be assessed by the following mean tap error (MTE)

$$\text{MTE} = \|\pm \hat{\mathbf{a}}_{\text{best}} - \mathbf{a}\|^2 \quad (20)$$

where \mathbf{a} with $a_i = 0$ for $i > n_a$ is the true channel. In (20), $-\hat{\mathbf{a}}_{\text{best}}$ is used if $\hat{\mathbf{a}}_{\text{best}}$ converges to $-\mathbf{a}$, otherwise $\hat{\mathbf{a}}_{\text{best}}$ is used. This is necessary as both \mathbf{a} and $-\mathbf{a}$ are correct global optimal channel estimates for HOC fitting cost functions (see the above discussion on sign ambiguity).

IV. RESULTS

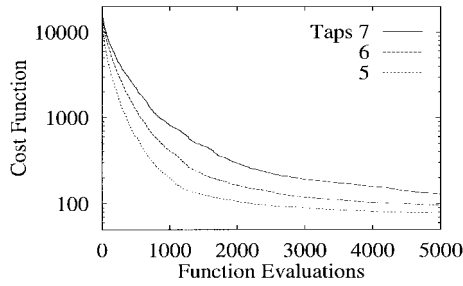
Simulation was conducted to test the proposed scheme using two channels taken from [2]. The impulse response of these two channels are given by

$$\begin{array}{l} \text{Channel 1} \quad \mathbf{a} = [-0.21 \quad -0.50 \quad 0.72 \quad 0.36 \quad 0.21]^T \\ \text{Channel 2} \quad \mathbf{a} = [0.227 \quad 0.460 \quad 0.688 \quad 0.460 \quad 0.227]^T \end{array} \quad (21)$$

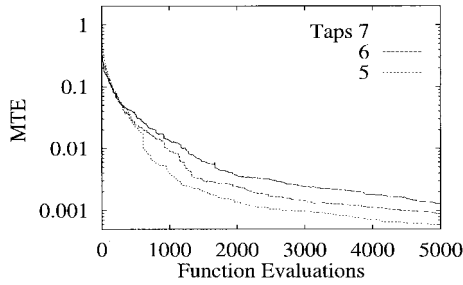
respectively. There were 8-PAM data symbols transmitted, and 50 000 noisy received data samples were used to compute the time estimate of the fourth-order cumulant. The signal-to-noise ratio (SNR) in the simulation was defined as

$$\text{SNR} = E[\hat{r}^2(k)]/E[e^2(k)] = \sigma_s^2 \left(\sum_{i=0}^{n_a} a_i^2 \right) / \sigma_e^2. \quad (22)$$

All the results were obtained by averaging over 100 different runs. Each run used a different sequence of noisy received data samples and a different randomly initialized population.

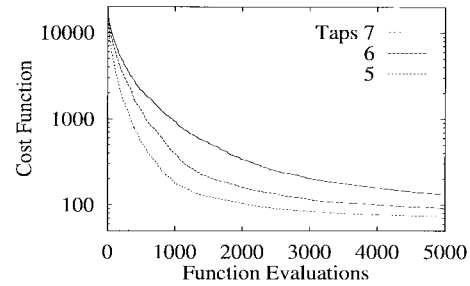


(a)

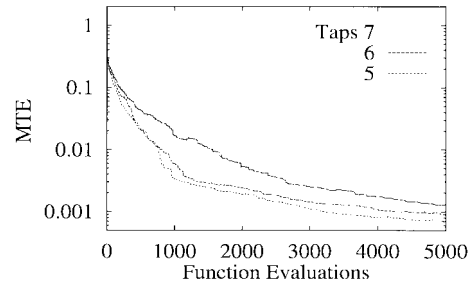


(b)

Fig. 2. (a) Cost function and (b) MTE versus number of function evaluations averaged over 100 different runs. Channel 1, 8-PAM, and SNR = 20 dB.

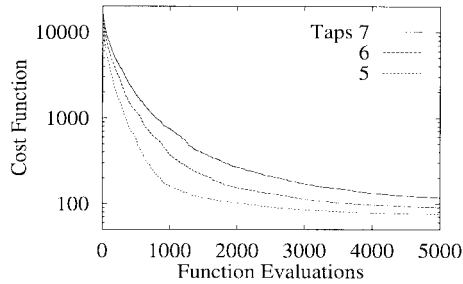


(a)

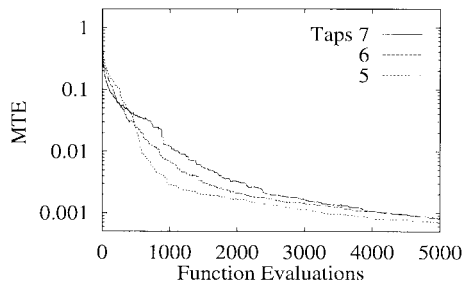


(b)

Fig. 4. (a) Cost function and (b) MTE versus number of function evaluations averaged over 100 different runs. Channel 1, 8-PAM, and SNR = 40 dB.



(a)



(b)

Fig. 3. (a) Cost function and (b) MTE versus number of function evaluations averaged over 100 different runs. Channel 1, 8-PAM, and SNR = 30 dB.

Figs. 2–4 depict evolutions of the cost function $\hat{J}_4(\hat{\mathbf{a}}_{\text{best}})$ and the MTE for channel 1 with different SNR conditions and assumed channel lengths \hat{n}_a , respectively. Tables II and III summarize the blind identification results (mean \pm standard deviation) for channel 1 with SNR's of 20 and 40 dB, respectively. Simulation results for channel 2 are similarly given in Figs. 5–7 and Tables IV and V. It is well known that channel 2 is much more difficult to equalize than channel 1, and this was reflected in terms of accuracy in our simulation results.

TABLE II

IDENTIFICATION RESULTS FOR CHANNEL 1 WITH 8-PAM AND SNR = 20 dB

	true	estimate (mean \pm standard deviation)		
		$\hat{n}_a = 4$	$\hat{n}_a = 5$	$\hat{n}_a = 6$
a_0	-0.21	-0.20975 \pm 0.01183	-0.21124 \pm 0.01483	-0.21098 \pm 0.01414
a_1	-0.50	-0.49957 \pm 0.00548	-0.49975 \pm 0.00632	-0.49835 \pm 0.00707
a_2	0.72	0.72092 \pm 0.00447	0.71965 \pm 0.00447	0.72087 \pm 0.00548
a_3	0.36	0.35788 \pm 0.00775	0.35859 \pm 0.01000	0.35823 \pm 0.01049
a_4	0.21	0.21073 \pm 0.01265	0.21058 \pm 0.01140	0.20870 \pm 0.01304
a_5	—	—	-0.00041 \pm 0.02236	-0.00184 \pm 0.02236
a_6	—	—	—	-0.00173 \pm 0.02345

TABLE III

IDENTIFICATION RESULTS FOR CHANNEL 1 WITH 8-PAM AND SNR = 40 dB

	true	estimate (mean \pm standard deviation)		
		$\hat{n}_a = 4$	$\hat{n}_a = 5$	$\hat{n}_a = 6$
a_0	-0.21	-0.21029 \pm 0.01000	-0.21384 \pm 0.01304	-0.20804 \pm 0.03435
a_1	-0.50	-0.49931 \pm 0.00548	-0.49769 \pm 0.00632	-0.49170 \pm 0.07092
a_2	0.72	0.71989 \pm 0.00447	0.72071 \pm 0.00447	0.71788 \pm 0.02302
a_3	0.36	0.36004 \pm 0.00837	0.35800 \pm 0.00837	0.35821 \pm 0.01095
a_4	0.21	0.21070 \pm 0.01378	0.20924 \pm 0.01304	0.20900 \pm 0.03493
a_5	—	—	-0.00103 \pm 0.03033	-0.00679 \pm 0.03795
a_6	—	—	—	-0.00265 \pm 0.02665

Computational complexity of the proposed scheme is summarized in Table VI, where N_g is the number of generations for the μ GA to achieve convergence. In the derivation of this theoretical complexity, we have used a population size of $n_p = 2(\hat{n}_a + 1)$ and only taken into account the computational requirements of cumulant calculation and function evaluation. Computational complexity of genetic operations are typically negligible, compared with the complexity of cost function evaluation. For the channels tested in the simulation, the GA-based scheme typically required a few hundreds of generations to obtain an adequate solution. Thus the required number

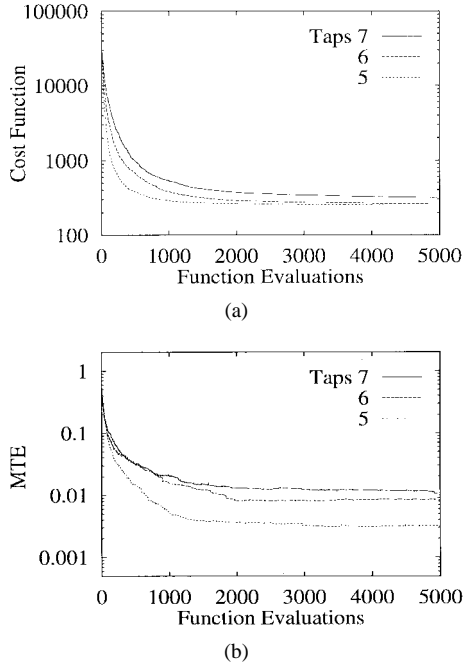


Fig. 5. (a) Cost function and (b) MTE versus number of function evaluations averaged over 100 different runs. Channel 2, 8-PAM, and SNR = 20 dB.

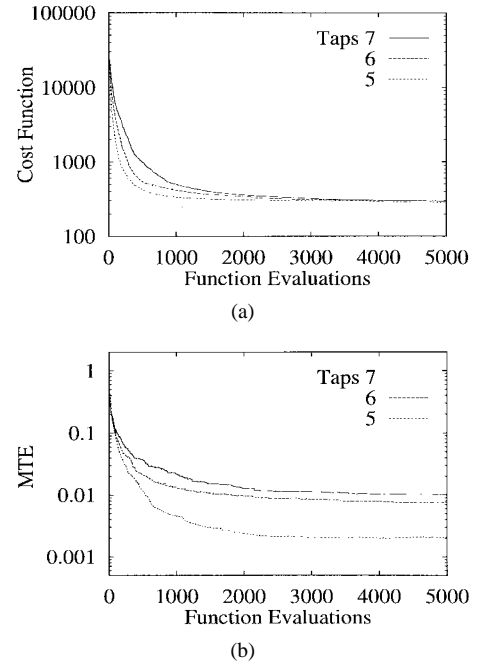


Fig. 7. (a) Cost function and (b) MTE versus number of function evaluations averaged over 100 different runs. Channel 2, 8-PAM, and SNR = 40 dB.

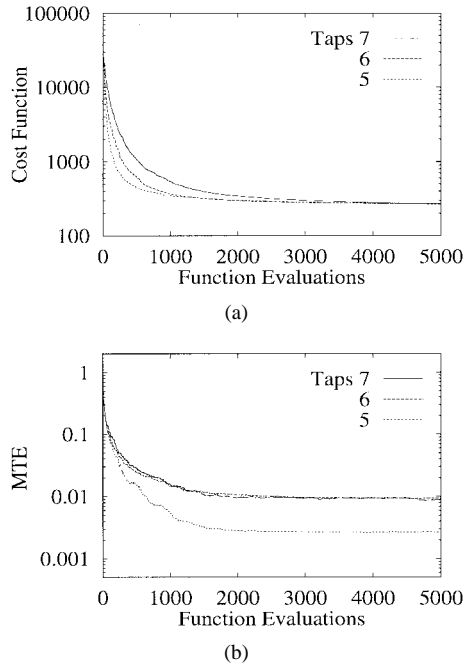


Fig. 6. (a) Cost function and (b) MTE versus number of function evaluations averaged over 100 different runs. Channel 2, 8-PAM, and SNR = 30 dB.

of function evaluations for the scheme was typically a few thousand, as confirmed in Figs. 2–7.

V. DISCUSSION

Some observations can readily be drawn from the simulation results. In the simulation, the μ GA-based scheme always converged close to a global optimal channel estimate and the optimization process converged quickly. Compared with other existing methods of HOC fitting, the method appears

TABLE IV
IDENTIFICATION RESULTS FOR CHANNEL 2 WITH 8-PAM AND SNR = 20 dB

	true	estimate (mean \pm standard deviation)		
	$n_a = 4$	$\hat{n}_a = 4$	$\hat{n}_a = 5$	$\hat{n}_a = 6$
a_0	0.227	0.22731 \pm 0.02828	0.22221 \pm 0.03146	0.21743 \pm 0.03450
a_1	0.460	0.45727 \pm 0.03114	0.45679 \pm 0.03564	0.45481 \pm 0.04919
a_2	0.688	0.68913 \pm 0.02646	0.68095 \pm 0.02569	0.67408 \pm 0.04254
a_3	0.460	0.45744 \pm 0.02280	0.46426 \pm 0.02775	0.46870 \pm 0.05119
a_4	0.227	0.22564 \pm 0.02811	0.22322 \pm 0.02720	0.22265 \pm 0.04658
a_5	—	—	0.01577 \pm 0.08837	0.01514 \pm 0.08081
a_6	—	—	—	0.00359 \pm 0.06964

TABLE V
IDENTIFICATION RESULTS FOR CHANNEL 2 WITH 8-PAM AND SNR = 40 dB

	true	estimate (mean \pm standard deviation)		
	$n_a = 4$	$\hat{n}_a = 4$	$\hat{n}_a = 5$	$\hat{n}_a = 6$
a_0	0.227	0.22698 \pm 0.01304	0.21768 \pm 0.03162	0.20608 \pm 0.05339
a_1	0.460	0.45168 \pm 0.01789	0.45670 \pm 0.03674	0.44640 \pm 0.05431
a_2	0.688	0.69342 \pm 0.01414	0.68378 \pm 0.02881	0.67003 \pm 0.04837
a_3	0.460	0.45715 \pm 0.01817	0.46087 \pm 0.03674	0.47454 \pm 0.05119
a_4	0.227	0.23020 \pm 0.01549	0.22442 \pm 0.03619	0.23998 \pm 0.05908
a_5	—	—	0.01842 \pm 0.08456	0.01185 \pm 0.08532
a_6	—	—	—	-0.00529 \pm 0.06671

TABLE VI
COMPUTATIONAL COMPLEXITY OF THE PROPOSED μ GA. N_g IS THE NUMBER OF GENERATIONS FOR THE ALGORITHM TO CONVERGE, N IS THE NUMBER OF SAMPLES, AND \hat{n}_a IS THE ASSUMED CHANNEL LENGTH

Complexity of cumulant computation	multiplications: $(8\hat{n}_a + 5)N - 4\hat{n}_a(\hat{n}_a - 1) + 5$ additions: $(4\hat{n}_a + 3)N - 2\hat{n}_a(\hat{n}_a - 2) - 2$
Complexity per function evaluation	multiplications: $3\hat{n}_a^2 + 10\hat{n}_a + 5$ additions: $\hat{n}_a^2 + 4\hat{n}_a + 1$
Number of function evaluations	$2(\hat{n}_a + 1)N_g$

more accurate and robust. This is demonstrated by very small standard deviations of estimated channel taps over different runs with randomly chosen channel populations. We also performed a range of simulations using four- and 16-PAM

data symbols and reducing the length of data samples for cumulant computation to 20 000. The results, not shown, are similar to those reported here. For other existing methods of HOC fitting, it is common that estimation accuracy will reduce and estimation variance will increase as SNR decreases. For the GA-based method, the simulation results suggest that, at least for the two channels tested, SNR has little effect on convergence rate and estimation accuracy. As was expected, our method is capable of identifying nonexistent channel taps with (near) zero values (at least an order smaller than values for existing taps). This is particularly important when an overlengthed channel model is used to avoid time-consuming model-order selection.

For this particular application, the convergence rate of the GA-based scheme is faster than that of the SA-based scheme reported in [25]. All the HOC-fitting algorithms have the same requirements of cumulant computation. Since the computational complexity per function evaluation is similar for GA and SA, we can compare their complexity by their required numbers of function evaluations. The number of function evaluations required for the GA scheme to converge was typically a few thousand in the simulation. This is compared favorably with the SA scheme. The number of function evaluations for the SA algorithm of [25] is $(\hat{n}_a + 1)N_T N_S T_{\text{red}}$, where N_S and N_T are the lengths of the two loops for step adjustment and T_{red} is the number of "temperature reductions." Quoting the figures from [25] with $N_T = 40$, $N_S = 20$ and $T_{\text{red}} = 5$, the required number of function evaluations for the channels used in our simulation study ($\hat{n}_a = 6$) would nearly be 30 000.

REFERENCES

- [1] S. U. H. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, pp. 1349–1387, Sept. 1985.
- [2] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1983.
- [3] C. F. N. Cowan and P. M. Grant, *Adaptive Filters*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
- [4] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation systems," *IEEE Trans. Commun.*, vol. COM-23, pp. 679–682, 1975.
- [5] D. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, pp. 1867–1875, 1980.
- [6] J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, no. 2, pp. 459–472, 1983.
- [7] G. Picchi and G. Prati, "Blind equalization and carrier recovering using a stop-and-go decision-directed algorithm," *IEEE Trans. Commun.*, vol. COM-35, pp. 877–887, 1987.
- [8] N. K. Jablon, "Joint blind equalization, carrier recovery, and timing recovery for high-order QAM signal constellations," *IEEE Trans. Signal Processing*, vol. 40, pp. 1383–1398, June 1992.
- [9] S. Chen, S. McLaughlin, P. M. Grant, and B. Mulgrew, "Multi-stage blind clustering equalizer," *IEEE Trans. Commun.*, vol. 43, no. 3, pp. 701–705, 1995.
- [10] K. S. Lii and M. Rosenblatt, "Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes," *Ann. Statist.*, vol. 10, pp. 1195–1208, 1982.
- [11] G. B. Giannakis and J. M. Mendel, "Identification of nonminimum phase system using higher order statistics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 360–377, 1989.
- [12] H.-H. Chiang and C. L. Nikias, "Adaptive deconvolution and identification of nonminimum phase FIR systems based on cumulants," *IEEE Trans. Automat. Contr.*, vol. 35, pp. 36–47, 1990.
- [13] D. Hatzinakos and C. L. Nikias, "Blind equalization using a tricepstrum-based algorithm," *IEEE Trans. Commun.*, vol. 39, pp. 669–682, May 1991.
- [14] F.-C. Zheng, S. McLaughlin, and B. Mulgrew, "Blind equalization of nonminimum phase channels: Higher order cumulant based algorithm," *IEEE Trans. Signal Processing*, vol. 41, pp. 681–691, Feb. 1993.
- [15] J. K. Tugnait, "Blind equalization and estimation of digital communication FIR channels using cumulant matching," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 1240–1245, 1995.
- [16] M. Ghosh and C. L. Weber, "Maximum-likelihood blind equalization," in *Proc. SPIE*, vol. 1565, pp. 188–195, San Diego, CA, 1991.
- [17] N. Seshadri, "Joint data and channel estimation using blind trellis search techniques," *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 1000–1011, 1994.
- [18] K. Giridhar, J. J. Shynk, and R. A. Iltis, "A modified Bayesian algorithm with decision feedback for blind adaptive equalization," in *Preprints 4th IFAC Int. Symposium Adaptive Systems in Control and Signal Processing*, France, 1992, pp. 737–742.
- [19] E. Zervas, J. Proakis, and V. Eyuboglu, "A quantized channel approach to blind equalization," in *Proc. ICC'92*, Chicago, 1992, vol. 3, pp. 351.8.1–351.8.5.
- [20] J. G. Proakis, "Adaptive algorithms for blind channel equalization," in *Proc. 3rd IMA Conf. Mathematics Signal Processing*, Univ. of Warwick, U.K., 1992.
- [21] S. Chen and Y. Wu, "Maximum likelihood joint channel and data estimation using genetic algorithms," *IEEE Trans. Signal Processing*, vol. 46, pp. 1469–1473, May 1998.
- [22] D. Williamson, R. A. Kennedy, and G. W. Pulford, "Block decision feedback equalization," *IEEE Trans. Commun.*, vol. 40, pp. 255–264, Feb. 1992.
- [23] S. Chen, B. Mulgrew, and S. McLaughlin, "Adaptive Bayesian equalizer with decision feedback," *IEEE Trans. Signal Processing*, vol. 41, pp. 2918–2927, Sept. 1993.
- [24] G. D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, no. 3, pp. 363–378, 1972.
- [25] J. Ilow, D. Hatzinakos, and A. N. Venetsanopoulos, "Blind equalizers with simulated annealing optimization for digital communication systems," *Int. J. Adapt. Contr. Signal Process.*, vol. 8, pp. 501–522, 1994.
- [26] J. H. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: Univ. of Michigan Press, 1975.
- [27] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA: Addison-Wesley, 1989.
- [28] J. J. Grefenstette, "Optimization of control parameters for genetic algorithms," *IEEE Trans. Syst. Man, Cybern.*, vol. SMC-16, no. 1, pp. 122–128, 1986.
- [29] L. Davis, Ed., *Handbook of Genetic Algorithms*. New York: Van Nostrand Reinhold, 1991.
- [30] K. Krishnakumar, "Micro-genetic algorithms for stationary and nonstationary function optimization," in *Proc. SPIE Intell. Cont. Adapt. Syst.*, vol. 1196, pp. 289–296, 1989.
- [31] L. Yao and W. A. Sethares, "Nonlinear parameter estimation via the genetic algorithm," *IEEE Trans. Signal Processing*, vol. 42, pp. 927–935, Apr. 1994.
- [32] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications," *Proc. IEEE*, vol. 79, pp. 278–305, Mar. 1991.
- [33] G. B. Giannakis and J. M. Mendel, "Cumulant-based order determination of non-Gaussian ARMA models," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1411–1423, Aug. 1990.
- [34] C. L. Nikias and A. P. Petropulu, *Higher-Order Spectra Analysis—A Nonlinear Signal Processing Framework*. Englewood Cliffs, NJ: Prentice-Hall, 1993.