FLEXIBLE MMSE DFE USING LENGTH AND LAG ADAPTATION

Yu Gong and Colin F. N. Cowan

Institute of Electronics, Communications and Information Technology
Queen's University of Belfast
Northern Ireland Science Park, Queen's Road
Belfast BT3 9DT, United Kingdom
email: y.gong, c.f.n.cowan@ee.qub.ac.uk

ABSTRACT

This paper investigates how to choose the optimum tap-length and decision delay for the decision feedback equalizer (DFE). Although the feedback filter length can be set as the channel memory, there is no closed-form expression for the feedforward filter length and decision delay. In this paper, first we analytically show that the two dimensional search for the optimum feedforward filter length and decision delay can be simplified to a one dimensional search, and then describe a new adaptive DFE where the optimum structural parameters can be self-adapted.

1. INTRODUCTION

The adaptive minimum mean square error (MMSE) equalizer has been widely used in communication systems. Although an ideal MMSE equalizer generally has infinite tap length [1], most designs use finite length filters because of simplicity and robustness. This brings up the problem of how to choose the tap length and decision delay which significantly affect the performance: on the one hand, though the MMSE is a monotonic non-increasing function of the tap-length, "too" long an equalizer not only unnecessarily increases the complexity but also introduces more adaptation noise; on the other hand, for a given filter length, the MMSE is generally a concave function of the decision delay [2]. Therefore, there exist *optimum* values of the tap-length and decision delay that best balance the steady-state performance and complexity. Most designs, however, choose the tap length and decision delay either based on pre-assumption of the channel or simply based on average experience, making it necessary to design flexible equalizers that can automatically find the "optimum" values of these structure parameters.

This research was sponsored by the UK Engineering and Physical Sciences Research Council under grant number GR/S00217/01.

The difficulties of designing such flexible equalizers come from not only the non existent closed-form relation between the MMSE and the tap-length or the decision delay, but also the constraint that either the tap-length or decision delay must take integer values. In [3], we proposed a novel length adaptation algorithm for the linear equalizer using a new concept of the pseudo "fractional tap-length" which can take fractional values. In the proposed approach, the relation between the MMSE and the tap-length was estimated on-line, and a gradient search algorithm was derived to adapt the *fractional tap-length* whose integer part gives the true tap-length. The proposed algorithm can converge to the optimum tap-length in the mean, but suffers from slow convergence under some scenarios. Fortunately, this problem has been successfully solved in our recent paper [4]. With similar idea, we also proposed a delay adaptation algorithm in [5] to search for the optimum decision delay for the linear equalizer.

Another widely used equalizer is the decision feedback equalizer (DFE) which consists of a feed-forward filter (FFF) $\{w_k\}$, a feedback filter (FBF) $\{b_k\}$ and a decision device. In a finite length DFE, there are three important structural parameters, namely the FFF tap-length N_f, the decision delay Δ and the FBF tap-length N_b. It has been proved in [6] that the FBF with length of $N_h - 1$ performs as well as any longer filters, where N_h is the channel length. In contrast, N_f is generally related to the inverse of the channel but has no closed-form expression, implying that even with the knowledge of N_h, it is still not straightforward to set an appropriate N_f. Moreover, the computation load for a DFE mainly comes from the FFF because the inputs to the FBF are the output from the hard decision device which are, for example, binary signals in a BPSK system. Thereby in practice, we can fix $N_b = N_h - 1$ where N_h can be roughly estimated by either order selection criteria (e.g [7]) or variable length adaptive algorithms (e.g. [4]), and search for the optimum values of N_f and Δ, which, in general, is a two dimensional searching problem.

Unfortunately, most current approaches mainly consider searching for the optimum Δ for a given N_f. For examples, in [8], Voois (et al.) expressed the decision delay as an explicit parameter for the DFE and described a method to obtain the optimum Δ for a given N_f. Further in [9], Al-Dhahir (et al.) proposed an efficient algorithm to calculate that optimum Δ. Obviously, non of these approaches guarantees to give the "global" optimum structural parameters where all possible N_f and Δ must be taken into consideration. In a recent paper [2], Hillery (et al.) suggested that, if $N_f = N_h$, the DFE may be insensitive to the decision delay. This conclusion, however, was mainly based on heuristic explanation, and we can easily give anti-examples as that will be shown later in this paper.

In this paper, we will analytically show that the two dimensional search for the optimum N_f and Δ can be simplified to one dimensional search by fixing $\Delta = N_f - 1$, base on which a novel DFE with adaptive N_f will be described. To highlight the key points, in this paper, we will assume the channel length N_h is known and $N_b = N_h - 1$, though in practice N_h must be estimated first.

2. OPTIMUM TAP-LENGTH

In general, an optimum FFF is anticausal and has infinite length [10], implying that the minimum MMSE with respect to all possible tap-length and decision delay is given by:

$$\xi_{\min} = \lim_{N_f \to \infty} \xi_{N_f,\,\Delta = N_f - 1}, \tag{1}$$

where $\xi_{N_f,\,\Delta}$ represents the MMSE with the FFF tap-length and decision delay being at N_f and Δ respectively. For simple expression, the right hand side of (1) will be denoted as $\xi_{N_f \to \infty,\,\Delta = N_f - 1}$ later. Then for a finite length DFE, the "global" optimum FFF tap-length $N_{f,\,\text{opt}}$ and decision delay Δ_{opt} can be defined as the minimum N_f and Δ that satisfy:

$$\xi_{N_f,\,\Delta} - \xi_{\min} < \mathcal{E}, \tag{2}$$

where $\mathcal{E}$ is a small positive constant which is preset according to the system requirements. This definition is based on the assumption that any MMSE satisfies (2) need not to be discriminated with $\xi_{\min}$.

In general, finding $N_{f,\,\text{opt}}$ and Δ_{opt} requires *brute force* search in two dimensions of N_f and Δ. For every possible N_f, only $\xi_{N_f,\,\Delta}$ for $N_f - 1 \leqslant \Delta \leqslant N_f + N_h - 2$ needs to be computed [11], though Δ is generally within the range of $[0,\ N_f + N_h - 2]$ outside which the received signal and the desired signal are uncorrelated[1]. However, even with this simplification and the efficient method to calculate $\xi_{N_f,\,\Delta}$ shown in [9], the total complexity in finding $N_{f,\,\text{opt}}$ and Δ_{opt} is still too high. Below we will show that the search in two dimension can be constrained to one dimension, greatly reducing the complexity.

[1]Without losing generality, we assume the transmission signals are mutual independent.

It has been proved in the Theorem 1 in [6] that, if $N_b = N_h - 1$ and $\Delta < N_f - 1$, then $w_k = 0$ for $k = \Delta + 1, \cdots, N_f - 1$, where w_k is the kth tap of the FFF. Thus for an infinite long FFF, we have:

$$\xi_{N_f \to \infty,\,\Delta} = \xi_{N_f = \Delta + 1,\,\Delta}, \qquad \text{for } \Delta = 0, 1, \cdots, \infty. \tag{3}$$

For a finite length FFF with $N_f = N$, the decision delay satisfies $0 \leqslant \Delta \leqslant N + N_h - 2$. Again from the Theorem 1 in [6], and noting (3), we have:

$$\xi_{N_f = N,\,\Delta} = \xi_{N_f = \Delta + 1,\,\Delta} = \xi_{N_f \to \infty,\,\Delta}, \text{ for } 0 \leqslant \Delta \leqslant N - 1. \tag{4}$$

On another front, since the optimum FFF generally has infinite length, and with the presence of noise, $\xi_{N_f,\,\Delta}$ is normally a strict decreasing function of N_f for a given Δ. Thus we have $\xi_{N_f = N,\,\Delta} > \xi_{N_f = \Delta + 1,\,\Delta}$ for $N \leqslant \Delta \leqslant N + N_h - 2$. Further noting (3) gives:

$$\begin{aligned} \xi_{N_f = N,\,\Delta} &> \xi_{N_f = \Delta + 1,\,\Delta} \\ &= \xi_{N_f \to \infty,\,\Delta}, \text{ for } N \leqslant \Delta \leqslant N + N_h - 2. \end{aligned} \tag{5}$$

The above analysis provides important information in understanding the optimum FFF tap-length and decision delay. For clarity of exposition, we illustrate this problem through a numerical example by plotting the curves for "MMSE vs. Δ" in Fig. 1, where N_f is fixed at $4,\ 7,\ 12,\ 20$ and ∞ respectively, the channel vector is $[0.1\ 0.2\ 0.4\ 0.6\ 0.8\ 1\ 0.7]^{\mathrm{T}}$ with $N_h = 7$, SNR = 20dB and the transmission signal $x(n)$ are either $+1$ or -1 (BPSK) and mutual independent.

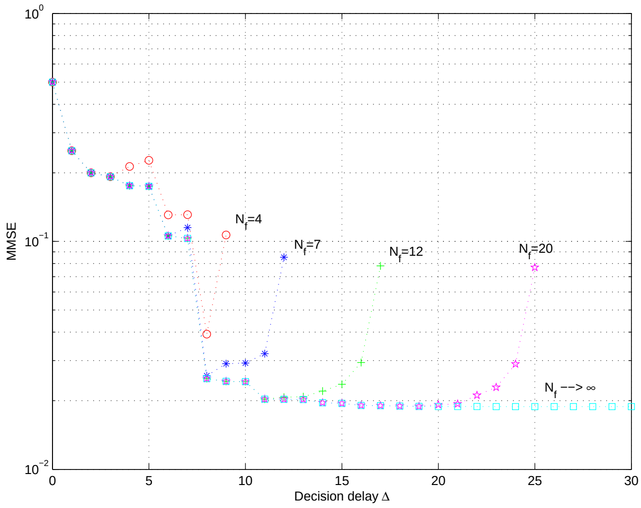

Fig. 1. MMSE vs Δ with N_f fixed at $4, 7, 12$ and ∞ respectively.

As shown in Fig. 1, the curves for different N_f form a "tree" structure, where the "trunk" of the tree corresponds to the curve for $N_f \to \infty$. For a finite N_f, the "MMSE vs. Δ" curve can be divided into two segments, corresponding to $0 \leqslant \Delta \leqslant N_f - 1$ and $N_f \leqslant \Delta \leqslant N_f + N_h - 2$ respectively. According to (4), the first segment of the curve overlaps with the "trunk", and according to (5), the second segment forms a "branch" of the tree which lies above the "trunk". Hence the "trunk" provides a lower MMSE bound for all possible N_f and Δ. Moreover, (3) shows that the "trunk" can also be formed by plotting the MMSE vs. $\{N_f = \Delta + 1, \Delta\}$ for $\Delta = 0, 1, \cdots, \infty$, implying that the "global" optimum FFF tap-length $N_{f,\,\mathrm{opt}}$ and decision delay Δ_{opt} must be on the "trunk" and satisfy $\Delta_{\mathrm{opt}} = N_{f,\,\mathrm{opt}} - 1$. Therefore, the search for $N_{f,\,\mathrm{opt}}$ needs only to be along the "trunk" by fixing $\Delta = N_f - 1$, which is in one dimension!

From the Theorem 3 in [6], we have $\xi_{N_f \to \infty,\, \Delta=0} \geqslant \cdots \geqslant \xi_{N_f \to \infty,\, \Delta = N_f - 1}$, which is, according to (3), equivalent to $\xi_{N_f = 1,\, \Delta = 0} \geqslant \cdots \geqslant \xi_{N_f = N,\, \Delta = N-1}$ for $N = 1, \cdots, \infty$. Thus $\xi_{N_f,\, \Delta = N_f - 1}$ (or the "trunk") is a strict non-increasing function of N_f, and then $N_{f,\,\mathrm{opt}}$ can be redefined as the minimum N_f that satisfies:

$$\xi_{N_f,\, \Delta = N_f - 1} - \xi_{\min} < \mathcal{E}. \tag{6}$$

It is clear from Fig. 1 that, in this example, we have $N_{f,\,\mathrm{opt}} = 12$ and $\Delta_{\mathrm{opt}} = 11$ since the "trunk" curve almost *flats out* after $\Delta = 11$.

As shown in [6], when N_f is large enough ($N_f = 12$ in this example), the optimum decision delay that minimizes the MMSE for a given N_f must satisfy $\Delta_o = N_f - 1$; otherwise ($N_f = 4$ and 7 in this example), we may have $N_f \leqslant \Delta_o \leqslant N_f + N_h - 2$, but then from (5), we have $\xi_{N_f,\, \Delta_o} > \xi_{N_f = \Delta_o + 1,\, \Delta_o}$, which implies that, rather than searching for Δ_o for a given N_f, it is more appropriate to increase N_f with $\Delta = N_f - 1$. Alternatively, "$\Delta_o = N_f - 1$" may be used as an index to indicate whether the optimum N_f has been reached or not, e.g., to increase N_f until $\Delta_o = N_f - 1$ is satisfied. In this paper, we are more interested in finding the optimum N_f adaptively as will be shown in the next section.

3. THE FFF TAP-LENGTH ADAPTATION

The previous section shows that the two dimensional search for $N_{f,\,\mathrm{opt}}$ and Δ_{opt} can be simplified to one dimensional search by fixing $\Delta = N_f - 1$, making it possible to adapt N_f instantaneously as in [4].

3.1. Algorithm

First, we observe that, if N_f is large enough and $\Delta = N_f - 1$, the first several coefficients of the FFF will be very small, so is the difference between the MMSE and the so-called "segment MMSE" which is defined as $\xi_{\{N_f - K\}} \triangleq \mathrm{E}|e^2_{\{N_f(n)-K\}}(n)|$, where:

$$e_{\{N_f - K\}}(n) \triangleq x(n - \Delta) - \tilde{x}_{\{N_f - K\}}(n - \Delta), \tag{7}$$

and

$$\tilde{x}_{\{N_f - K\}}(n - \Delta) \triangleq \mathbf{w}^{\mathrm{T}}_{K+1:N_f}(n) \cdot \mathbf{y}_{K+1:N_f}(n) + z_b(n), \tag{8}$$

where $x(n)$ is the transmitted signal, $z_b(n)$ is the FBF output, K is a positive integer much smaller than N_f, and $\mathbf{w}_{K+1:N_f}(n)$ and $\mathbf{y}_{K+1:N_f}(n)$ consist of the last $N_f - K$ coefficients of the FFF tap-vector $\mathbf{w}(n)$ and input-vector $\mathbf{y}(n)$ respectively. Note that $\tilde{x}_{N_f}(n)$ and $e_{\{N_f\}}(n)$ are the equalization output and error signal of the DFE.

Therefore, similar to [4], it is possible to define the optimum N_f by comparing $\xi_{\{N_f - K\}}$ and $\xi_{\{N_f\}}$. To be specific, the optimum FFF tap-length, $N'_{f,\,\mathrm{opt}}$, can be defined as the minimum N_f that satisfies:

$$\xi_{\{N_f - K\}} - \xi_{\{N_f\}} < \mathcal{E}', \tag{9}$$

where $\mathcal{E}'$ is also a smaller positive constant. It is clear that, when $\mathcal{E}$ and $\mathcal{E}'$ are small, the optimum FFF tap-length based on (6) and (9) are either same or very close, i.e. $N_{f,\,\mathrm{opt}} \approxeq N'_{f,\,\mathrm{opt}}$. To make sense of (9), we must ensure $N_f \geqslant K + 1$, below which no $N'_{f,\,\mathrm{opt}}$ can be differentiated.

If $\xi_{\{N_f - K\}}$ and $\xi_{\{N_f\}}$ are known for every N_f, the N_f adaptation rule can be simply obtained as:

$$N_f(n+1) = N_f(n) - M \cdot \mathrm{sign}\left(\xi_{\{N_f(n)\}} - \xi_{\{N_f(n) - K\}} + \mathcal{E}'\right), \tag{10}$$

where $N_f(n)$ is the FFF tap-length at time n and M is the step-size parameter which must be an integer. It can be easily verified that (10) converges to within a range of $(N'_{f,\,\mathrm{opt}} - M,\ N'_{f,\,\mathrm{opt}} + M)$.

In practice, unfortunately, $\xi_{\{N_f - K\}}$ and $\xi_{\{N_f\}}$ are not a priori known. Then the concept of the *pseudo-fractional tap-length* introduced in [3, 4] may be applied. Specifically, defining $n_f(n)$ as the *psedo-fractional FFF tap-length* which can be decimal, we have the adaptation rule as:

$$n_f(n+1) = (n_f(n) - \alpha) - \gamma \cdot \left[e^2_{\{N_f(n)\}}(n) - e^2_{\{N_f(n) - K\}}(n)\right], \tag{11}$$

where both α and γ are small positive numbers, and α is the leaky factor which is used to prevent the length adapting into necessarily large values. To ensure stability, we must have $\alpha \ll \gamma$. Initially we have $n_f(0) = N_f(0)$.

The "true" FFF tap-length $N_f(n)$ is determined according to:

$$N_f(n+1) = \begin{cases} \lfloor n_f(n) \rfloor, & |N(n) - n_f(n)| \geqslant M \\ N_f(n), & \text{otherwise} \end{cases} \tag{12}$$

where $\lfloor . \rfloor$ rounds the embraced value to the nearest integer. It is implied from (12) that $N_f(n)$ is modified by a value of

M each time. To be specific, if $N_f(n)$ is to be increased, then M number of zeros are padded at the *head* of the FFF. Otherwise if N_f is to be decreased, the *first* M coefficients of the FFF are taken out.

3.2. Performance Analysis

Since (10) and (11) are based on the same cost function, they should converge to the same tap-length. In fact, following the similar procedures to those in [4], we can easily verify that, if $\alpha/\gamma = \mathcal{E}'$, (11) converges to within a range of $(N'_{f,\text{ opt}} - M,\ N'_{f,\text{ opt}} + M)$ in the mean. The detail of the analysis, however, is not shown here due to the space constraint.

Since the proposed length adaptation is based on instantaneous gradient search, it imposes very low complexity. To be specific, the extra complexity brought by the length adaptation mainly comes from the adaption recursion (11) and the computation of $e^2_{\{N_f(n)\}}(n)$ and $e^2_{\{N_f(n)-K\}}(n)$. At each iteration, (11) requires 3 additions and 1 multiplications. But we note that $e_{\{N_f(n)\}}(n)$ and $\tilde{x}_{\{N_f\}}(n)$ are also required by the coefficient adaption itself, and

$$\tilde{x}_{\{N_f\}}(n-\Delta) = \tilde{x}_{\{N_f-K\}}(n-\Delta) + \mathbf{w}^{\mathrm{T}}_{1:K}(n)\cdot \mathbf{y}_{1:K}(n), \tag{13}$$

which implies we need only one more addition to obtain $\tilde{x}_{\{N_f-K\}}(n-\Delta)$, and, from (7), another one more addition to obtain $e_{\{N_f(n)-K\}}(n)$. Then considering the *square* manipulation, we need 2 more additions and 2 more multiplications to obtain $e^2_{\{N_f(n)\}}(n)$ and $e^2_{\{N_f(n)-K\}}(n)$ simultaneously. In total, at each iteration, we only require 5 extra additions and 3 extra multiplication for the proposed length adaption algorithm, regardless of the filter length.

4. NUMERICAL RESULTS

We have done extensive numerical simulations with different channel specifications to verify the proposed algorithm. Due to the space constraint, however, only one example is shown below, where the same channel as that in Fig. 1 is applied, $N_b = N_h - 1$, $\Delta = N_f - 1$ and the normalized LMS with step-size 0.2 is used to adapt both the FFF and FBF coefficients. For the N_f adaptation, we have $\alpha = 0.01$, $\gamma = 0.5$ and $K = 3$ for (11), and $M = 1$ for (12).

Fig. 2 shows the N_f learning curves for different initialization at $N_f(0) = 4$ and 20 respectively. Both curves are for one typical simulation run. Recalling that the optimum FFF tap-length is 12 as shown in Fig. 1, we can clearly observe that the algorithm converges to the optimum tap-length for both initializations. In practice, if N_h is known, we may set $N_f(0) = N_h$ as it may be a "good guess" of the optimum N_f and the convergence time may be shortened.

Fig. 3 shows the MSE learning curves for $N_f = 7$, $N_f = 12$ and the adaptive N_f with $N_f(0) = 4$ respectively.

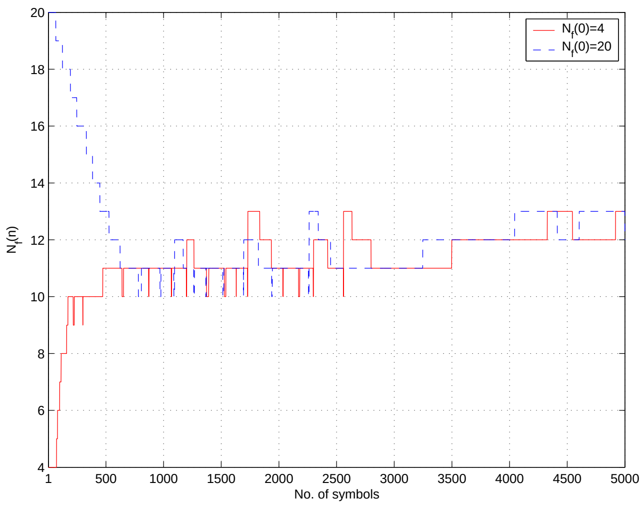

Fig. 2. N_f learning curves for different initializations.

For clarity of exposition, all MSE learning curves are obtained by averaging over 50 independent runs. It is clearly shown that the DFE with $N_f = N_{f,\text{ opt}} = 12$ performs significantly better than the that with $N_f = N_h = 7$. Thus in this example, setting N_f as N_h is obviously not large enough, unlike what has been implied in [2]. But it is clear that the DFE with adaptive N_f performs almost as well as that with $N_f = N_{f,\text{ opt}} = 12$.

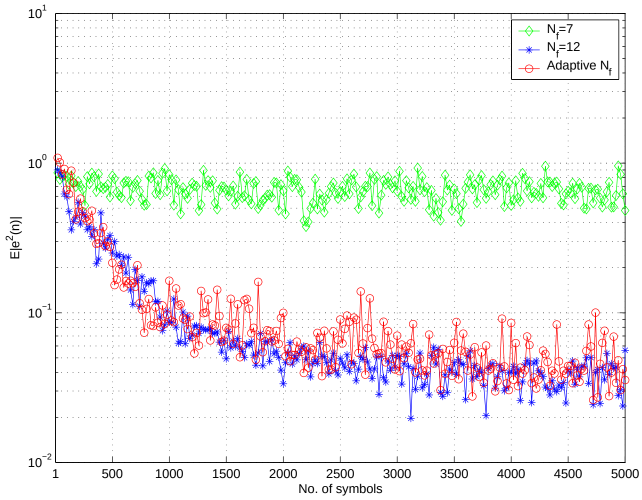

Fig. 3. MSE learning curves.

5. CONCLUSIONS

In this paper, we investigated the structure adaptation for the DFE equalizer, i.e., how to choose the tap-length and the decision delay to give the best performance. First, we showed that, with $N_b = N_h - 1$, the general two dimensional search for the optimum FFF tap-length and decision delay can be simplified to a one dimensional search by fixing $\Delta = N_f - 1$, and then proposed a novel DFE where the optimum N_f can be self-adapted. Numerical simulations were also given to verify the analysis and proposed algorithm.

We note that, under the scenario where the channel memory is very large, the complexity may be high to set the feedback filter length as the channel memory. Fortunately, since the complexity of a DFE mainly comes from the feedforward filter, it may not be too bad to have a large feedback filter length. But after all, when the channel memory is too large, it is difficult for a single DFE to equalize the channel at all. Some other approaches, such as OFDM, must be considered.

Another import issue is the equalizer design for the channel with sparse impulse response and large delay spread. Under such scenario, only adjusting the tap-length is not enough, and the space between the taps must also be carefully chosen (see [12] and the references there in).

6. REFERENCES

[1] S. U. H. Qureshi, "Adaptive equalization," *Proceedings of the IEEE*, vol. 73, no. 9, pp. 1349 – 1387, Sept. 1985.

[2] W. J. Hillery, M. D. Zoltowski, and M. Fimoff, "Decision feedback equalizer design for insensivity to decision delay," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03)*, Hong Kong, April 2003, vol. 4, pp. IV –505–8.

[3] Y. Gong and C. F. N. Cowan, "Structure adaptation of linear MMSE adaptive filters," *IEE Proceedings - Vision, Image and Signal Processing*, vol. 151, no. 4, pp. 271 – 277, Aug. 2004.

[4] Y. Gong and C. F. N. Cowan, "An LMS style variable tap-length algorithm for structure adaptation," *IEEE Trans Signal Processing, To appear in July 2005*.

[5] Y. Gong and C. F. N. Cowan, "A variable time lag algorithm for the FIR equalizer," in *12th European Signal Processing Conference (EUSIPCO2004)*, Vienna, Austria, Sept. 2004.

[6] Y. Gong and C. F. N. Cowan, "Optimum decision delay of the finite-length DFE," *IEEE Signal Processing Letters*, vol. 11, no. 11, pp. 858 – 861, Nov. 2004.

[7] H. Akaike, "A new look at the statistical model identification," *IEEE Trans Automat. Contr*, vol. AC-19, pp. 716 – 723, Dec. 1974.

[8] P. A. Voois, I. Lee, and J. M. Cioffi, "The effect of decision delay in finite-length decision feedback equalization," *IEEE Trans. on Information Theory*, vol. 42, no. 2, pp. 618 – 621, March 1996.

[9] N. Al-Dhahir and J. M. Cioffi, "Efficient computation of the delay-optimized finite-length MMSE-DFE," *IEEE Trans. on Signal Processing*, vol. 44, no. 5, pp. 1288 – 1292, May 1996.

[10] J. Salz, "Optimum mean-square decision feedback equalization," *Bell System Techniqual Journal*, vol. 52, pp. 1341 – 1373, Oct. 1973.

[11] N. Al-Dhahir and J. M. Cioffi, "MMSE decision-feedback equalizers: finite-length results," *IEEE Trans. on Information Theory*, vol. 41, no. 4, pp. 961 – 975, July 1995.

[12] M. J. Lopez and A. C. Singer, "A DFE coefficient placement algorithm for sparse reverberant channels," *IEEE Trans. on Commun.*, vol. 49, no. 8, pp. 1334 – 1338, Aug. 2001.