

A Self-Structured Adaptive Decision Feedback Equalizer

Yu Gong and Colin F. N. Cowan, *Senior Member, IEEE*

Abstract—In a decision feedback equalizer (DFE), the structural parameters, including the decision delay, the feedforward filter (FFF), and feedback filter (FBF) lengths, must be carefully chosen, as they greatly influence the performance. Although the FBF length can be set as the channel memory, there is no closed-form expression for the FFF length and decision delay. In this letter, first we analytically show that the two-dimensional search for the optimum FFF 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.

Index Terms—Adaptive algorithm, decision delay, decision feedback equalizer (DFE), tap-length.

I. PROBLEM STATEMENTS

THE typical structure of a decision feedback equalizer (DFE) consists of a feedforward filter (FFF), a feedback filter (FBF), and a decision device. Although an ideal DFE generally has infinite length [1], most designs use finite length filters because of simplicity and robustness. This brings up the problem of how to choose the structural parameters, including the FFF tap-length N_f , the decision delay Δ , and the FBF tap-length N_b , as they have great influence on the performance. Note that the decision delay determines to which transmitted symbol the detected symbol corresponds. To be specific, on the one hand, when N_f and N_b are fixed, there exists an optimum Δ that minimizes the minimum mean-square error (MMSE); on the other hand, for a given Δ , the MMSE is always a nonincreasing function of N_f or N_b , but too long a N_f or N_b not only unnecessarily increases the complexity with little MMSE improvement but also increases the adaptation noise when the adaptive algorithm is applied. Therefore, for all possible choices of N_f , N_b , and Δ , there must exist a group of optimum values of them that correspond to the best MMSE performance. The purpose of this letter is, therefore, to search for the optimum structural parameters that normally vary with different channels.

Some designs, by comparison, intend to reach a preset target MMSE with the simplest complexity, such as the smallest tap-length. However, how to choose the target MMSE, thereafter, becomes another issue that is different from that of this letter,

where we want the MMSE as small as possible. In some other approaches, the algorithm fixes the tap-length based on a tolerable complexity and finds the corresponding optimum Δ . For example, in [2], fixing the total number of taps " $N_{tot} = N_f + N_b$," Al-Dhahir *et al.* described a method to search for the optimum values of N_b and Δ that minimize the MMSE. Such an approach, however, has two disadvantages: First, it did not point out how to choose an appropriate value of N_{tot} , which, similar to that for the linear equalizer [3], should balance the complexity and performance and thus generally varies with the channel. Thus, the resulting structural parameters are only "locally" optimum for one particular choice of N_{tot} . Second, although the assumption of fixing N_{tot} attempts to place a constraint on the complexity, two DFEs with the same N_{tot} may vary significantly in complexity, depending on how the taps are assigned between the FFF and FBF. This is because the computation load for a DFE mainly comes from the FFF as the inputs to the FBF are the output from the hard decision device that are, for example, binary signals in a binary phase-shift keying (BPSK) system. Therefore, fixing N_{tot} is not equivalent to "fixing" the complexity. In fact, as will be shown later in this letter, the constraint on N_{tot} or N_f may result in performance loss. Thus, from the performance point of view, it is more appropriate to choose the tap-length based on specific channels rather than complexity constraint. It is true that, if the channel has very long memory, the DFE has high complexity. However, under such a scenario, it is difficult for a single DFE to do the equalization at all, and other approaches such as orthogonal frequency-division multiplexing (OFDM) should be applied. In general, how to trade off the MMSE performance and the complexity constraint is an issue beyond the scope of this letter.

It has been proved in [4] 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. Chen *et al.* pointed out in [5] that the DFE should have $N_f \geq N_h$ and $\Delta = N_f - 1$. In a more recent paper [6], Hillery *et al.* further argued that, when N_f is at least as long as the channel span, the DFE is insensitive to the decision delay for a fairly wide band. What the exact value of an appropriate N_f should be, however, still remains a question. In fact, in many cases, setting $N_f = N_h$ is not enough, as will be shown in an example later. Similarly, although Tidestav *et al.* and Lopez-Valcarce have shown in [7] and [8], respectively, that there is no need to increase N_f beyond a given value of Δ , they did not point out how to choose an appropriate Δ so that the best MMSE performance can be reached.

Manuscript received September 22, 2005; revised November 9, 2005. This work was supported by the U.K. Engineering and Physical Sciences Research Council under Grant GR/S00217/01. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dominic K. C. Ho.

The authors are with the Institute of Electronics, Communications, and Information Technology, Queen's University of Belfast, Belfast BT3 9DT, U.K. (e-mail: y.gong@ee.qub.ac.uk; c.f.n.cowan@ee.qub.ac.uk).

Digital Object Identifier 10.1109/LSP.2005.862601

Since, as was just mentioned, the complexity of a DFE mainly comes from the FFF part, the DFE is more tolerant to an overestimated N_b than an overestimated N_f . So in practice, we may let $N_b = N_h - 1$, where N_h can be roughly estimated by either order selection criteria (e.g., [9]) or variable length adaptive algorithms (e.g., [3]), and find the optimum values of N_f and Δ . This is clearly a two-dimensional search problem. Most current approaches, unfortunately, mainly consider searching for the optimum Δ for a given N_f or N_{tot} . For example, in [10], 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 [11], Al-Dhahir *et al.* proposed an efficient algorithm to calculate that optimum Δ . None of these approaches, including that in [2], guarantees to give the *true* optimum structural parameters, where all possible N_f and Δ must be taken into consideration. Besides, all these approaches are based on the “brute-force” search by directly manipulating the input correlation matrix, resulting in high complexity.

It has been shown in [4] that, when $N_b = N_h - 1$ and N_f is large enough, the optimum decision delay satisfies $\Delta = N_f - 1$. This result itself, however, does not directly lead to the conclusion that the optimum N_f must be “large enough” to satisfy the above equation. In fact, one of the contributions of this letter is to prove in Section II that the optimum values of N_f and Δ do satisfy $\Delta = N_f - 1$, and the two-dimensional search for the optimum N_f and Δ can then be simplified to a one-dimensional search. The proof itself, moreover, provides a deep insight in understanding the DFE. In Section III, a novel DFE with adaptive N_f is proposed, and numerical results are provided in Section IV to verify the analysis and the proposed algorithm. To highlight the key points, in the rest of this letter, we will assume the channel length N_h is known and $N_b = N_h - 1$, though in practice, N_h must be estimated, as was mentioned previously.

II. OPTIMUM FFF TAP-LENGTH AND DECISION DELAY

A. Two-Dimensional Definition

In general, an optimum FFF is acausal and has infinite length [1], implying that the minimum MMSE with respect to all possible tap-length and decision delay is given by

$$\xi_{\min} = \lim_{N_f \rightarrow \infty} \xi(N_f, \Delta = N_f - 1) \quad (1)$$

where $\xi(N_f, \Delta)$ represents the MMSE with the FFF tap-length and decision delay being at N_f and Δ , respectively. For simplicity, the right-hand side of (1) will be denoted as $\xi(N_f \rightarrow \infty, \Delta = N_f - 1)$ later. Then the optimum FFF tap-length $N_{f,\text{opt}}$ and decision delay Δ_{opt} are defined as the minimum N_f and Δ that satisfy

$$\xi(N_f, \Delta) - \xi_{\min} < \mathcal{E} \quad (2)$$

where \mathcal{E} is a small positive constant that is preset according to the system requirements. This definition is based on the assumption that any MMSE that satisfies (2) need not be discriminated with ξ_{\min} . It is clear that the larger the value of \mathcal{E} , the more sacrifice to the MMSE in order to achieve length reduction.

In general, finding $N_{f,\text{opt}}$ and Δ_{opt} requires *brute-force* search in two dimensions of N_f and Δ for a reasonably

large number of N_f . For every N_f , only $\xi(N_f, \Delta)$ for $N_f - 1 \leq \Delta \leq N_f + N_h - 2$ need to be computed [12], 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 [11], 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 reduced to one dimension.

B. One-Dimensional Definition

It has been proved in [4, Theorem 1] that, if $N_b = N_h - 1$ and $\Delta < N_f - 1$, then $w_k = 0$ for $k = \Delta + 1, \dots, N_f - 1$, where w_k is the k th tap of the FFF. Thus, for an infinitely long FFF, we have

$$\xi(N_f \rightarrow \infty, \Delta) = \xi(N_f = \Delta + 1, \Delta), \quad \text{for } \Delta = 0, 1, \dots, \infty. \quad (3)$$

For a finite length FFF with $N_f = N$, the decision delay satisfies $0 \leq \Delta \leq N + N_h - 2$. Again from [4, Theorem 1], and noting (3), we have

$$\begin{aligned} \xi(N_f = N, \Delta) &= \xi(N_f = \Delta + 1, \Delta) \\ &= \xi(N_f \rightarrow \infty, \Delta), \quad \text{for } 0 \leq \Delta \leq N - 1. \end{aligned} \quad (4)$$

On the other hand, since the optimum FFF generally has infinite length, with the presence of noise, $\xi(N_f, \Delta)$ is normally a strictly decreasing function of N_f for a given Δ . Thus, we have $\xi(N_f = N, \Delta) > \xi(N_f = \Delta + 1, \Delta)$ for $N \leq \Delta \leq N + N_h - 2$. Further noting (3) gives

$$\begin{aligned} \xi(N_f = N, \Delta) &> \xi(N_f = \Delta + 1, \Delta) \\ &= \xi(N_f \rightarrow \infty, \Delta) \\ &\quad \text{for } N \leq \Delta \leq N + N_h - 2. \end{aligned} \quad (5)$$

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 MMSE curves with respect to Δ in Fig. 1, where N_f is fixed at 4, 7, 12, 20, and ∞ , respectively, the channel vector is $[0.1 \ 0.2 \ 0.4 \ 0.6 \ 0.8 \ 1 \ 0.7]^T$ with $N_h = 7$, SNR = 20 dB, and the transmission signal $x(n)$ are either +1 or -1 (BPSK) and mutually independent.

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 $N_f \rightarrow \infty$. For a finite N_f , the “MMSE versus Δ ” curve can be divided into two segments, corresponding to $0 \leq \Delta \leq N_f - 1$ and $N_f \leq \Delta \leq N_f + N_h - 2$, 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 that 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 versus $\{N_f = \Delta + 1, \Delta\}$ for $\Delta = 0, 1, \dots, \infty$, implying that the optimum FFF tap-length $N_{f,\text{opt}}$ and decision delay Δ_{opt} must be on the “trunk” and satisfy $\Delta_{\text{opt}} = N_{f,\text{opt}} - 1$. Therefore, the

¹Without losing generality, we assume the transmission signals are mutually independent.

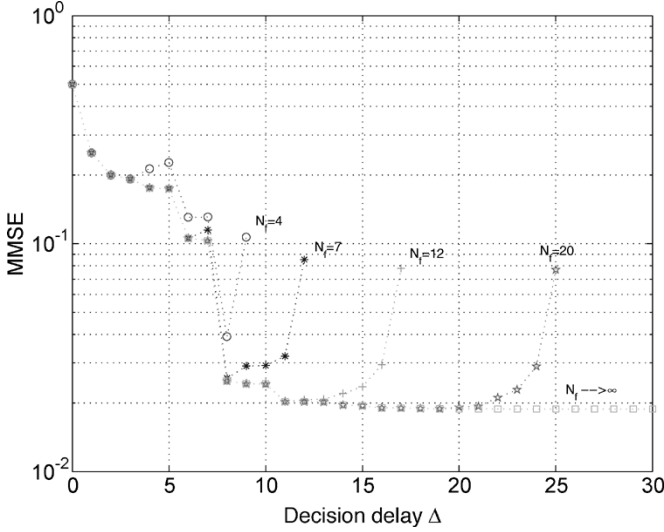


Fig. 1. MMSE versus Δ with N_f fixed at 4, 7, 12, and ∞ , respectively.

search for $N_{f,\text{opt}}$ needs only to be along the “trunk” by fixing $\Delta = N_f - 1$, which is in one dimension.

From the Theorem 3 in [4], we have $\xi(N_f \rightarrow \infty, \Delta = 0) \geq \dots \geq \xi(N_f \rightarrow \infty, \Delta = N_f - 1)$, which is, according to (3), equivalent to $\xi(N_f = 1, \Delta = 0) \geq \dots \geq \xi(N_f = N, \Delta = N - 1)$ for $N = 1, \dots, \infty$. Thus, $\xi(N_f, \Delta = N_f - 1)$ (or the “trunk”) is a strictly nonincreasing function of N_f , and then $N_{f,\text{opt}}$ can be redefined as the minimum N_f that satisfies

$$\xi(N_f, \Delta = N_f - 1) - \xi_{\min} < \mathcal{E}. \quad (6)$$

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

C. Some Discussions

As was shown in [4], 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 $N_f \leq \Delta_o \leq N_f + N_h - 2$, but then from (5), we have $\xi(N_f, \Delta_o) > \xi(N_f = \Delta_o + 1, \Delta_o)$, which implies that, from the performance point of view, rather than searching for Δ_o for a given N_f , it is more appropriate to increase N_f with $\Delta = N_f - 1$.

In [2], another approach of reducing the searching space for the structure parameters from two to one dimension was also described, which was achieved by assuming the FBF cancels all post-cursor ISI so that $\Delta = N_{\text{tot}} - 2N_b + N_h - 1$ and then finding the optimum N_b that minimizes the MMSE, where the total number of taps $N_{\text{tot}} = N_f + N_b$ is fixed and $1 \leq N_b \leq N_h - 1$.² It can be easily verified that such an approach, from the performance point of view, is equivalent to finding the optimum N_f that minimizes $\xi(N_f, \Delta)$ for $N_{\text{tot}} - N_h + 1 \leq N_f \leq N_{\text{tot}} - 1$, where $\Delta = 2N_f + N_h - N_{\text{tot}} - 1$ and $N_b = N_h - 1$.³ This corresponds to $N_f - 1$ number of searching

²Note that [2] and this letter use slightly different notations for the structure parameters.

³Under such a scenario, it is equivalent to having $N_b = N_{\text{tot}} - N_f$ and $N_b = N_h - 1$ in terms of the MMSE performance.

points on the “tree” structure described above. More specifically, because it can also be easily verified that $N_f - 1 \leq \Delta \leq N_f + N_h - 2$ for every N_f , there are $N_f - 2$ of the searching points on the “branches,” and one searching point on the “trunk,” corresponding to $N_f = N_{\text{tot}} - N_h$ as only then $\Delta = N_f - 1$. Since the “branches” always lie above the “trunk,” this indicates that the approach described in [2] cannot guarantee convergence to the optimum value, and the performance loss may happen when $N_{\text{tot}} < N_{f,\text{opt}} + N_h - 1$, where $N_{f,\text{opt}}$ is the optimum N_f defined in (6). The exact value of the performance loss depends on the specific channel. On the other hand, if $N_{\text{tot}} > N_{f,\text{opt}} + N_h - 1$, unnecessary complexity will arise.

III. FFF TAP-LENGTH ADAPTATION

A. Cost Function

The previous section shows that the two-dimensional search for $N_{f,\text{opt}}$ and Δ_{opt} can be simplified to a one-dimensional search by fixing $\Delta = N_f - 1$, making it possible to adapt N_f instantaneously as in [3]. First, we need a cost function on which N_f adaptation can be based. Unfortunately, we cannot directly apply (6) since it requires ξ_{\min} , which is normally *a priori* unknown. An alternative definition for the optimum FFF tap-length is thus introduced below.

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 $\xi_{\{N_f-K\}} \triangleq E|e_{\{N_f(n)-K\}}(n)|^2$, where we define

$$e_{\{N_f-K\}}(n) \triangleq x(n - \Delta) - \left(\mathbf{w}_{K+1:N_f}^T \cdot \mathbf{y}_{K+1:N_f}(n) + z_b \right) \quad (7)$$

where z_b is the FBF output, K is a positive integer much smaller than N_f , and $\mathbf{w}_{K+1:N_f}$ and $\mathbf{y}_{K+1:N_f}(n)$ consist of the last $N_f - K$ coefficients of the FFF tap-vector \mathbf{w} and input-vector $\mathbf{y}(n)$, respectively. K is set to avoid local minima corresponding to zero channel taps. Note that $e_{\{N_f\}}(n)$ and $\xi_{\{N_f\}}$ are just the error signal and MMSE of the DFE.

Therefore, similar to that in [3], it is possible to define the optimum N_f by comparing $\xi_{\{N_f-K\}}$ and $\xi_{\{N_f\}}$. To be specific, we can define an alternative optimum FFF tap-length $N'_{f,\text{opt}}$ as the minimum N_f that satisfies

$$\xi_{\{N_f-K\}} - \xi_{\{N_f\}} < \mathcal{E}' \quad (8)$$

where \mathcal{E}' is also a smaller positive constant [3]. It is clear that, when \mathcal{E} and \mathcal{E}' are small, the optimum FFF tap-length based on (6) and (8) are either the same or very close, i.e., $N_{f,\text{opt}} \approx N'_{f,\text{opt}}$. We note that, to make sense of (8), we must ensure $N_f \geq K + 1$, below which no $N'_{f,\text{opt}}$ can be differentiated.

B. Adaptation Rule

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

$$N_f(n+1) = N_f(n) - M \cdot \text{sign}(\xi_{\{N_f(n)\}} - \xi_{\{N_f(n)-K\}} + \mathcal{E}') \quad (9)$$

where $N_f(n)$ is the FFF tap-length at time n , and M is the step-size parameter that must be an integer. It can be easily verified

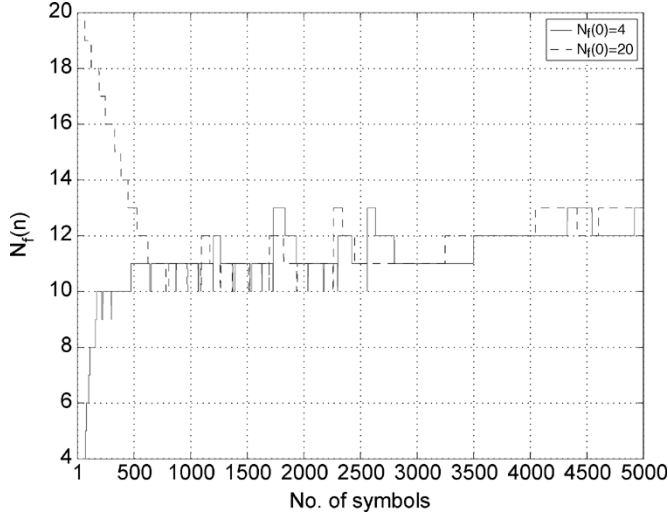


Fig. 2. N_f learning curves for different initializations.

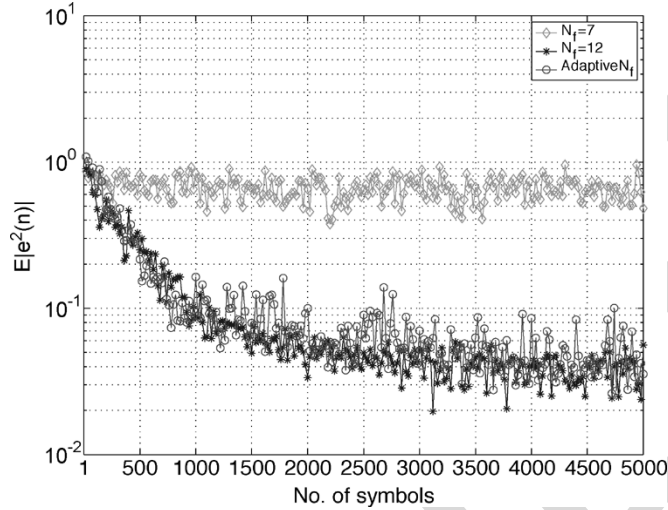


Fig. 3. MSE learning curves.

that (9) converges to within a range of $(N'_{f,\text{opt}} - M, N'_{f,\text{opt}} + M)$.

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] may be applied. Specifically, defining $n_f(n)$ as the *pseudo-fractional FFF tap-length*, which can take fractional values, we have the length adaptation rule as

$$n_f(n+1) = (n_f(n) - \alpha) - \gamma \cdot [e^2_{\{N_f(n)\}}(n) - e^2_{\{N_f(n)-K\}}(n)] \quad (10)$$

where both α and γ are small positive numbers. Specifically, α is an additive leaky factor that is used to prevent the length adapting into unnecessarily large values, and γ is the step-size parameter for the above adaption rule. 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

$$N_f(n+1) = \begin{cases} \lfloor n_f(n) \rfloor, & |N_f(n) - n_f(n)| \geq M \\ N_f(n), & \text{otherwise} \end{cases} \quad (11)$$

where $\lfloor \cdot \rfloor$ rounds the embraced value to the nearest integer. It is implied from (11) 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.

Since (9) and (10) are based on the same cost function, they should converge to the same tap-length. In fact, following similar procedures to those in [3], we can easily verify that, if $\alpha/\gamma = \mathcal{E}'$, (10) 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 constraints. Finally, we note that, since the N_f adaptation of (10) is based on instantaneous gradient search, it imposes very low complexity on the DFE.

IV. NUMERICAL SIMULATIONS

This section gives numerical simulation to verify the proposed algorithm, 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 (10), and $M = 1$ for (11).

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 shorten the convergence time. We note that, similar to the famous LMS algorithm, the instantaneous nature of (10) makes the N_f learning curves keep “bouncing” between 12 and 13 after convergence.

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. 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 that with $N_f = N_h = 7$. Thus, in this example, having $N_f = N_h$ is obviously not large enough. However, it is clear that the DFE with adaptive N_f performs almost as well as that with $N_f = N_{f,\text{opt}} = 12$.

V. CONCLUSION

In this letterer, first we showed that, with $N_b = N_h - 1$, the generally 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.

REFERENCES

- [1] J. Salz, “Optimum mean-square decision feedback equalization,” *Bell Syst. Tech. J.*, vol. 52, pp. 1341–1373, Oct. 1973.
- [2] N. Al-Dhahir and C. Fragouli, “How to choose the number of taps in a DFE,” in *Proc. 36th Annu. Conf. Information Science Systems*, Princeton, NJ, Mar. 2002.

- [3] Y. Gong and C. F. N. Cowan, "An LMS style variable tap-length algorithm for structure adaptation," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2400–2407, Jul. 2005.
- [4] —, "Optimum decision delay of the finite-length DFE," *IEEE Signal Process. Lett.*, vol. 11, no. 11, pp. 858–861, Nov. 2004.
- [5] S. Chen, B. Mulgrew, and S. McLaughlin, "Adaptive Bayesian equalizer with decision feedback," *IEEE Trans. Signal Process.*, vol. 41, no. 9, pp. 2918–2927, Sep. 1993.
- [6] W. J. Hillery, M. D. Zoltowski, and M. Fimoff, "Decision feedback equalizer design for insensitivity to decision delay," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 4, Hong Kong, Apr. 2003, pp. IV-505–IV-508.
- [7] C. Tidestav, A. Ahlen, and M. Sternad, "Realizable MIMO decision feedback equalizers: Structure and design," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 121–133, Jan. 2001.
- [8] R. Lopez-Valcarce, "Realizable linear and decision feedback equalizer: Properties and connections," *IEEE Trans. Signal Process.*, vol. 52, no. 3, pp. 757–773, Mar. 2004.
- [9] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 12, pp. 716–723, Dec. 1974.
- [10] P. A. Voois, I. Lee, and J. M. Cioffi, "The effect of decision delay in finite-length decision feedback equalization," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 618–621, Mar. 1996.
- [11] N. Al-Dhahir and J. M. Cioffi, "Efficient computation of the delay-optimized finite-length MMSE-DFE," *IEEE Trans. Signal Process.*, vol. 44, no. 5, pp. 1288–1292, May 1996.
- [12] —, "MMSE decision-feedback equalizers: finite-length results," *IEEE Trans. Inf. Theory*, vol. 41, no. 4, pp. 961–975, Jul. 1995.