Tracking Analysis of an LMS Decision Feedback Equalizer for a Wireless Channel

Veeraruna Kavitha and Vinod Sharma
Department of Electrical Communication Engineering,
Indian Institute of Science, Bangalore, 560012, India.
kavitha@pal.ece.iisc.ernet.in vinod@ece.iisc.ernet.in

***Abstract*— We consider a time varying wireless fading channel, equalized by an LMS Decision Feedback equalizer (DFE). We study how well this equalizer tracks the optimal MMSE-DFE (Wiener) equalizer. We model the channel by an Autoregressive (AR) process. Then the LMS equalizer and the AR process are jointly approximated by the solution of a system of ODEs (ordinary differential equations). Using these ODEs, we show via some examples that the LMS equalizer moves close to the instantaneous Wiener filter after initial transience. We also compare the LMS equalizer with the instantaneous optimal DFE (the commonly used Wiener filter) designed assuming perfect previous decisions and computed using perfect channel estimate (we will call it as IDFE). We show that the LMS equalizer outperforms the IDFE almost all the time after initial transience.**

Keywords : LMS-DFE, ODE approximation, Wiener filter, Convergence analysis, Tracking performance.

I. Introduction

A channel equalizer is an important component of a communication system and is used to mitigate the ISI (inter symbol interference) introduced by the channel. The equalizer depends upon the channel characteristics. In a wireless channel, due to multipath fading, the channel characteristics change with time. Thus it may be necessary for the channel equalizer to track the time varying channel in order to provide reasonable performance.

Least Mean Square linear equalizer (LMS-LE) is a simple equalizer and is extensively used ([2], [5]). For a fixed channel its convergence to the Wiener filter has been studied in [2], [14] (see also the references therein). Its performance on a wireless (time varying) channel has been studied theoretically in [7] (see also [5], [10] where the performance has been studied via simulations, approximations and upper bounds on probability of error etc).

Decision feedback equalizers (DFE) are nonlinear equalizers, which can provide significantly better performance than LE ([1], [15], [16]). A DFE feeds back the previous decisions of the transmitted symbols, to nullify the ISI due to them and makes a better decision about the current symbol. Although these equalizers have also been used for quite sometime, due to feedback their behavior is much more complex than that of the LEs. Hence their performance is not well understood. Existence of a hard decoder inside the feedback loop, due to its nonlinearity, makes the study all the more difficult. A DFE mainly exploits the finite alphabet structure of the hard decoder output ([4], [12]) and hence the hard decoder cannot be ignored (i.e., its performance is better than a system with a soft decoder).

For a DFE, statistics of the previous decisions are not known. Hence till recently, there was no known technique which provides an MMSE DFE (we will call is as DFE-WF for the rest of the paper) even for a fixed channel ([3], [12], [15]). We addressed this issue in [8] (see details below). Prior to [8], one usually designed an MMSE DFE by assuming perfect decisions (see, e.g., [12], [17]). For convenience, for the rest of the paper, we will call such a DFE as IDFE (Ideal DFE). In this paper IDFE is also computed *using perfect channel estimates*. The IDFE often outperforms the Linear Wiener filter significantly ([1], [15], [16]). But it is generally believed that DFE-WF, the true MSE optimal DFE (designed considering the decision errors), can outperform even this.

Another way to obtain an optimal filter is to replace the feedback filter at the receiver by a precoder at the transmitter ([3], [15]). This way one can indeed obtain the optimal filter but this requires the knowledge of the channel at the transmitter. For wireless channels, which are time varying, this is often not an attractive solution ([12], [15]).

Some research has been done to deal with the decision errors. Either the distribution of the decisions errors were approximated in designing an MSE optimal filter (IDFE being one such example) or some other appropriate criterion was used to get the optimal filter considering the errors in decisions. For example, in [19], authors approximated the errors in decisions with an AWGN (Additive White Gaussian Noise) independent of the input sequence and obtained a DFE Wiener filter. But as is stated in the paper this approximation is not realistic. In [4], the authors obtain an H^{∞} optimal DFE taking into account the decision errors. However no comparison to DFE-WF was provided.

One possibility of obtaining DFE-WF is via LMS-DFE. However, convergence of LMS-DFE is not well understood even for a fixed channel. Trajectory of the LMS-DFE algorithm, on a fixed channel, with a *soft decoder* in the feedback loop has been approximated by an ODE in [11]. But this ODE does not approximate the LMS-DFE with a hard decoder. Beneveniste et al. ([2]) have shown the ODE approximation of an LMS-DFE with a hard decoder. But the

This research is partially supported by DRDO-IISc program on advanced research in Mathematical Engineering.

ODE obtained by them is not explicit enough. Furthermore, they do not relate the attractors of this ODE to the DFE-WF.

In [8] we showed that the LMS-DFE asymptotically comes close to the DFE-WF at high SNRs. We also showed that, it can outperform the commonly used Wiener filter IDFE, at all practical SNRs. We thus concluded that the LMS-DFE can be used to obtain an equalizer close to the DFE-WF.

In this paper we study the behavior of an LMS-DFE while tracking a wireless channel. To study the tracking behavior theoretically, one needs to have a theoretical model of the fading channel. Auto Regressive (AR) processes have been shown to model such channels quite satisfactorily ([10], [13], [18]). We will model our channel by an AR(2) process as in [7]. The class of AR(2) processes includes the Random Walk model and the Filtered Random Walk model ([13]). Thus it is a very useful model for many wireless channels.

In this paper, we approximate the trajectory of the LMS-DFE tracking an AR(2) process by a set of ODEs. Using these ODEs we demonstrate via examples that an LMS-DFE in fact comes close to the instantaneous DFE-WF after some initial transience. We also show that it significantly outperforms the commonly used IDFE (even when designed using perfect channel estimates) at all practical SNRs. An interesting observation is that, the improvement is significant even at high SNRs where an IDFE does not suffer from error propagation.

The paper is organized as follows. Our system model, notations and assumptions are discussed in Section II. In Section III, we obtain an ODE approximation for the tracking trajectory of an LMS-DFE. Section IV provides some examples verifying our claims, while Section V concludes the section. Some of the proofs are provided in the Appendices.

II. System model and Notations

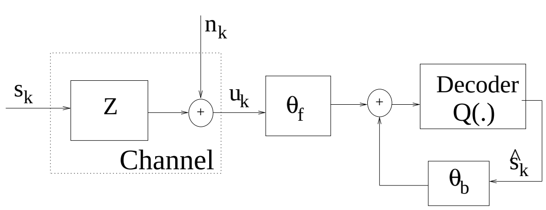

Fig. 1. Block diagram of a Wireless channel followed by a DFE.

We consider a system with a wireless channel and a DFE (see Figure 1). Inputs $\{s_k\}$ enter a time varying finite impulse response channel $\{z_{k,l}\}_{l=0}^{L-1}$, and are corrupted by additive white Gaussian noise $\{n_k\}$ with variance σ^2. The channel output, u_k, at any time k, is given by,

$$u_k = \sum_{l=0}^{L-1} s_{k-l} z_{k,l} + n_k.$$

The time variations of the channel are modeled by an AR(2) process,

$$Z_k = d_1 Z_{k-1} + d_2 Z_{k-2} + \mu W_k \qquad (1)$$

where W_k is an IID vector sequence of Gaussian random variables (Gaussian assumption is not really needed) and $Z_k = \begin{bmatrix} z_{k,0}, & z_{k,1} & \cdots & z_{k,L-1} \end{bmatrix}$. The channel outputs u_k pass through a DFE with a hard decoder. The details about the equalizer are given below. We use the following notations and assumptions.

- We assume BPSK modulation, i.e., $s_k \in \{+1, -1\}$.
- Sequences $\{s_k\}$ and $\{n_k\}$ are IID (independent, identically distributed) and independent of each other.
- The equalizer forward filter is given by $\{\theta_{f_l}\}_{l=0}^{N_f-1}$, while the feedback filter is given by $\{\theta_{b_l}\}_{l=1}^{N_b}$.
- $N_L \triangleq N_f + L - 1$.
- The decisions are obtained after hard decoding. Hence decision $\hat{s}_k$ is given by,

$$\begin{aligned} \hat{s}_k &= Q\left(\sum_{l=0}^{N_f-1} \theta_{f_l} u_{k-l} + \sum_{l=1}^{N_b} \theta_{b_l} \hat{s}_{k-l}\right) \text{ where} \\ Q(x) &:= \begin{cases} +1 & if \quad x \geq 0, \\ -1 & if \quad x < 0. \end{cases} \end{aligned} \qquad (2)$$

- For any vector, x, we use x_l to represent its l^{th} component. x_l^k, $l \leq k$, represents the vector $\begin{bmatrix} x_k & x_{k-1} & \cdots & x_l \end{bmatrix}^T$.
- The following vector notations are used throughout.

$$\begin{array}{llll} S_k \triangleq s_{k-N_L+1}^k, & N_k \triangleq n_{k-N_f+1}^k \\ U_k \triangleq u_{k-N_f+1}^k, & \hat{S}_k \triangleq \hat{s}_{k-N_b+1}^k, \\ X_k \triangleq [\, U_k^T \quad \hat{S}_{k-1}^T \,]^T, & G_k \triangleq [\, S_k^T \quad X_k^T \,]^T, \\ \theta_f \triangleq {\theta_f}_0^{N_f-1}, & \theta_b \triangleq {\theta_b}_1^{N_b}, \\ \theta \triangleq [\, {\theta_f}^T \quad {\theta_b}^T \,]^T. \end{array}$$

- θ_k represent the time varying equalizer at time k.
- Let $\mathcal{S} := \{+1, -1\}$. For a fixed (θ, Z), G_k is a Markov chain made of the channel input S_k, channel output U_k and the decoder output $\hat{S}_{k-1}$ when the channel is the fixed vector Z and the equalizer is fixed at θ. G_k takes values in $\mathcal{S}^{N_L} \times \mathcal{S}^{N_b} \times \mathcal{R}^{N_f}$, where $\mathcal{R}$ is the set of real numbers. We represent throughout this paper the current and previous state values of this Markov chain by the ordered pairs (i, y), (j, y') respectively. Here i, j take values from the discrete part of the state space, $\mathcal{S}^{N_L} \times \mathcal{S}^{N_b}$, while y, y' take values in $\mathcal{R}^{N_f}$.
- $Z\theta = \{z\theta_l\}_{l=0}^{N_L-1}$ represents the convolution of the channel $\{z_l\}$and forward filter θ_f.
- $B(\theta, \delta)$, $\bar{B}(\theta, \delta)$ are the open and closed balls respectively with center θ and radius δ.
- $K(\epsilon, M) := \{(\theta_1, \theta_2) \in \mathcal{R}^{N_f} \times \mathcal{R}^{N_b} :$
$\epsilon < |\theta_1| \leq M, |\theta_2| \leq M\}$.
- $E_{j,y'}^{\theta,Z}$ represents the expectation of the Markov chain G_k for fixed channel, equalizer pair (θ, Z) when the initial condition is j, y'.
- $E_{j,y';\theta,Z}$ represents the expectation of the Markov chain $\{G_k, \theta_k, Z_k, Z_{k-1}\}$ with initial condition (j, y', θ, Z).
- $E_{(j,y');(i,y)}^{\theta,Z;\theta',Z'}$ represents the expectation of the Markov chain pair $\{(G_k, G_k')\}$ under the initial condition j, y', i, y. Here G_k is the Markov chain for fixed channel, equalizer pair (θ, Z) with initial condition j, y'

while G'_k is the one for channel, equalizer pair (θ', Z') with initial condition i, y. When both the initial conditions are same it is simply represented by, $E^{\theta,Z;\theta',Z'}_{i,y}$.

- $P^{\theta,Z^k}(.|.)$, $\Pi^{\theta,Z}$, $E^{\theta,Z}$ respectively represent the $k-$step transition function, the stationary distribution, and the expectation wrt to the stationary measure (existence will be shown) for a fixed channel, equalizer pair (θ, Z).
- We use a DFE, θ, to track the wireless channel modeled by an AR(2) process, $\{Z_k\}$. The LMS algorithm is used to continuously update the equalizer θ to cater to the time varying channel.

$$\begin{aligned} \theta_{k+1} &= \theta_k - \mu H_{1\theta_k}(G_k) \text{ where} \\ H_{1\theta}(G) &\stackrel{\triangle}{=} X\left(X^t\theta - s\right). \end{aligned} \quad (3)$$

III. ODE Approximation

We can rewrite the channel AR(2) process (1) as,

$$Z_{k+1} = (1-d_2)Z_k + d_2 Z_{k-1} + \mu(W_k + \eta Z_k). \quad (4)$$

We will show below that the trajectory (θ_k, Z_k) given by equations (4), (3) can be approximated by the solution of the following system of ODEs,

$$\begin{aligned} (1+d_2)\dot{Z}(t) &= h(Z(t)), \qquad \text{if } d_2 \in (-1,1], \\ \frac{d^2Z(t)}{dt^2} &= h(Z(t)), \qquad \text{if } d_2 = -1, \\ \frac{d^2Z(t)}{dt^2} + \eta_1 \dot{Z}(t) &= h(Z(t)), \\ & \qquad \text{if } d_2 \text{ is close to } -1, \end{aligned} \quad (5)$$

$$\dot{\theta}(t) = h_1(\theta(t), Z(t)), \quad (6)$$

where

$$\begin{aligned} h(Z) &\stackrel{\triangle}{=} E(W_k + \eta Z) = E(W_1) + \eta Z, \\ h_1(\theta, Z) &\stackrel{\triangle}{=} E^{\theta,Z}\left[X_k\left(X_k^t\theta - s_k\right)\right] \\ &= -R_{xx}(\theta,Z)\theta + R_{xs}(\theta,Z), \\ \eta &= \frac{d_1+d_2-1}{\mu}, \\ R_{xx}(\theta,Z) &= E^{\theta,Z}\left(XX^t\right), \\ R_{xs}(\theta,Z) &= E^{\theta,Z}(Xs). \end{aligned}$$

In (5), when d_2 is close but not equal to -1, two ODEs approximate the same AR(2) process. This is an important case and results when a second order AR process approximates a fading channel with a U-shaped band limited spectrum. It is obtained for small values of f_dT where f_d is the maximum Doppler frequency shift and T is the symbol transmission time. For example if f_dT equals 0.04, 0.01 or 0.005 the channel is approximated by an AR(2) process with (d_1, d_2, μ) equal to $(1.9707, -0.9916, 0.00035)$, $(1.9982, -0.9995, 1.38e^{-6})$ and $(1.9995, -0.9999, 8.66e^{-8})$ respectively (see e.g., [10])). One could approximate such an AR(2) process with the first order ODE of (5). However this approximation will not be very accurate and will require μ to be very small. In this case, the second order ODE approximates the channel trajectory better. We will plot these approximations in Section IV.

One can easily see that the above system of ODEs have unique global solutions that are bounded for any finite time (more details are in the technical report [9]).

Let $Z(t,t_0,Z), \theta(t,t_0,\theta)$ represent the solutions of the ODEs (5), (6) with initial conditions $Z(t_0) = Z$, $\theta(t_0) = \theta$, and $\dot{Z}(t_0) = 0$ whenever the channel is approximated by a second order ODE.

Let $V_k \stackrel{\triangle}{=} (Z_k, \theta_k)$ and $V(k) \stackrel{\triangle}{=} (Z(\mu^\alpha k, 0, Z), \theta(\mu k, 0, \theta))$, where $\alpha = 1$ if $Z(.,.,.)$ is the solution of a first order ODE and $= 1/2$ otherwise. We prove Theorem 1, as in [7].

Theorem 1: For any finite $T > 0$, for all $\delta > 0$ and for any initial condition (G, θ, Z), with $d_2 Z_{-1} + d_1 Z_0 = Z$, $\dot{Z}(t_0) = 0$ whenever the channel is approximated by a second order ODE and $\theta_0 = \theta$,

$$P_{G,Z,\theta}\left\{\sup_{\left\{1 \le k \le \frac{T}{\mu^\alpha}\right\}} |V_k - V(k)| \ge \delta\right\} \to 0,$$

as $\mu \to 0$, uniformly for all $(Z,\theta) \in Q$, if Q is contained in the bounded set containing the solution of the ODEs (5), (6) till time T.

Proof : Please see the Appendix A.

Thus we obtain the ODE approximation for the LMS-DFE tracking an AR(2) process. The approximating ODE (6) suggests that, its instantaneous attractors will be same as the LMS-DFE attractors when the channel is fixed at the instantaneous value of the channel ODE (5) (as in [8]). We have shown in [8] that these LMS-DFE attractors are close to the DFE-WF at high SNRs. Hence the ODE suggests that the LMS-DFE may move close to the instantaneous DFE-WFs. We will in fact see that this is true for the examples we study in the next section.

One can easily see that the solution of the channel (AR(2) process) ODE is,

$$Z(t) := \begin{cases} C_1 e^{\frac{\eta}{1+d_2}t} - \frac{E(W)}{\eta}, \\ \qquad \eta \neq 0, d_2 \in (-1,1], \\ \frac{E(W)}{1+d_2}t + C_1, \\ \qquad \eta = 0, d_2 \in (-1,1], \\ C_1 cosh(\sqrt{\eta}\ \ t) - \frac{E(W)}{\eta}, \\ \qquad \eta > 0, d_2 = -1, \\ C_1 cos(\sqrt{|\eta|}\ \ t) - \frac{E(W)}{\eta}, \\ \qquad \eta < 0, d_2 = -1, \\ \frac{E(W)}{2}\ \ t^2 + C_1, \\ \qquad \eta = 0, d_2 = -1, \\ C_1 e^{-2at} + \frac{E(W)}{2a}\ \ t, \\ \qquad \eta = 0, d_2 \text{ close to } -1, \\ C_1 e^{-at} cos(\sqrt{|\eta| - a^2}\ \ t) - \frac{E(W)}{\eta}, \\ \qquad \eta < -a^2, d_2 \text{ close to } -1, \\ C_1 e^{-at} cosh(\sqrt{\eta + a^2}\ \ t) - \frac{E(W)}{\eta}, \\ \qquad otherwise, \end{cases} \quad (7)$$

where the constant C_1 is chosen appropriately to match the initial condition of the approximated AR(2) process.

One of the uses of the above ODE approximation is that we can approximately obtain the performance (e.g., BER, MSE) of our LMS-DFE at any time by using the trajectory of this ODE. Of course, obtaining BER theoretically is still a problem because the BER of a system with a fixed known channel and a fixed DFE is still not available. But our ODE approximation is still useful because one can obtain the performance (transient as well as stationary) of the LMS-DFE with only one simulation, which would not be possible otherwise.

IV. EXAMPLES

We use the ODE approximation of the previous section to obtain some interesting conclusions. The ODE approximation gives accurate deterministic approximation of the LMS-DFE and the channel trajectory for practical values of step sizes. Hence as commented above, using these ODEs one can get good estimates of instantaneous performance measures like, Bit Error Rate (BER) and Mean Square Error (MSE) for almost all realizations of the LMS-DFE and the channel trajectory. Using the channel ODE, one can also obtain the same performance measures for instantaneous IDFE. Then one can compare the IDFE and the LMS-DFE along the entire time axis.

In all the examples below, we estimate the DFE-WF as in [8] (directly using steepest descent algorithm by approximating the gradient of the MSE with difference of estimated MSEs at two close points divided by the distance between the same two points). In the figures solid line, dash dot line, dash dash line respectively represent the true coefficient trajectory, the ODE approximation and the IDFE trajectory respectively while the stars represent the DFE-WFs.

To begin with we consider a stable channel (for which all the poles are inside the unit circle),

$$Z_k \quad = \quad .4995 Z_{k-1} + 0.5 Z_{k-2} + 0.0001 W_k.$$

Here W_k is a Gaussian IID random vector with independent components of unit variance and its mean given by,

$$\begin{bmatrix} 0.26 & 0.34 & 0.25 & 0.064 & -0.13 & -0.19 & -0.16 & 0 & 0.064 & 0.064 \end{bmatrix}.$$

We consider a five tap feed-forward filter and a five tap feed-back filter. The LMS step-size equals $\mu = 0.001$. (In theory it is assumed that the step-size μ of LMS is also equal to the channel step-size. However one can absorb the difference into one of the $H_1()$, $H()$ functions.) The noise variance $\sigma^2 = 0.05$. We plot some of the channel and the equalizer filter coefficient trajectories along with their ODE approximations in Figure 2 (the trajectories for the other co-efficients behave in the same way but we do not present due to lack of space. These are provided in [9], which also contains additional examples). We start the LMS and the ODEs at time $t = 0$ with the instantaneous IDFE. We also plot the instantaneous DFE-WF and the IDFEs in the same figure. We can see that the ODE approximation is quite accurate for all the co-efficients. The approximation for the feed-forward coefficients is better than for the feed-back coefficients. We also see that the LMS-DFE is very close to the instantaneous DFE-WF after some initial transience. Furthermore, the IDFE trajectory is away from the DFE-WF in most of the cases. We also plot the instantaneous BER and MSE of the IDFE and the LMS-DFE (both calculated from the corresponding ODE approximations) in last two sub figures of Figure 2. One can see a huge improvement (upto 35%) of LMS-DFE (also of DFE-WF) over the IDFE both in terms of BER, MSE after the initial transience (Figure 2). On the other hand, performance of the LMS-DFE is quite close to that of the DFE-WF.

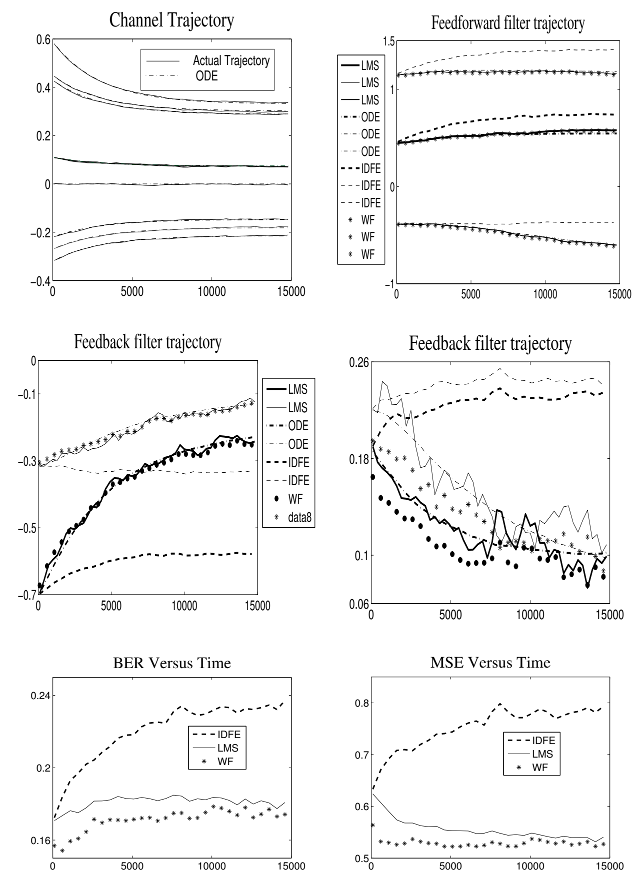

Fig. 2. A Stable channel with $d_1 = 0.4995$, $d_2 = 0.5$, $\mu = 1e^{-4}$ and mean a constant multiple of $[0.26,\ 0.34,\ 0.25,\ 0.064,\ -0.13,\ -0.19,\ -0.16,\ 0,\ 0.064,\ 0.064]$.

Next, we consider a marginally stable channel in Figure 3. $Z_k = 1.9999998 Z_{k-1} - Z_{k-2} + 0.0000001 W_k$ (One can see from (7) that $\sqrt{\frac{d_1+d_2-1}{\mu}}$ gives the period of oscillations). W_k is generated as before. Again there are five taps in the feed-forward filter and five in the feed-back filter. The step size of the LMS equals, $\mu = 0.001$. Here the channel trajectory is approximated by a cosine waveform. From Figure 3 we can make the same observations as in the stable case. In particular we see that the LMS-DFE has BER and MSE upto

50% less than for the IDFE. We also see that the LMS-DFE is always (after initial transience) very close to the DFE-WF while the IDFE stays quite away. This again explains the poor performance of the IDFE (in terms of BER, MSE) over the LMS-DFE after initial transience.

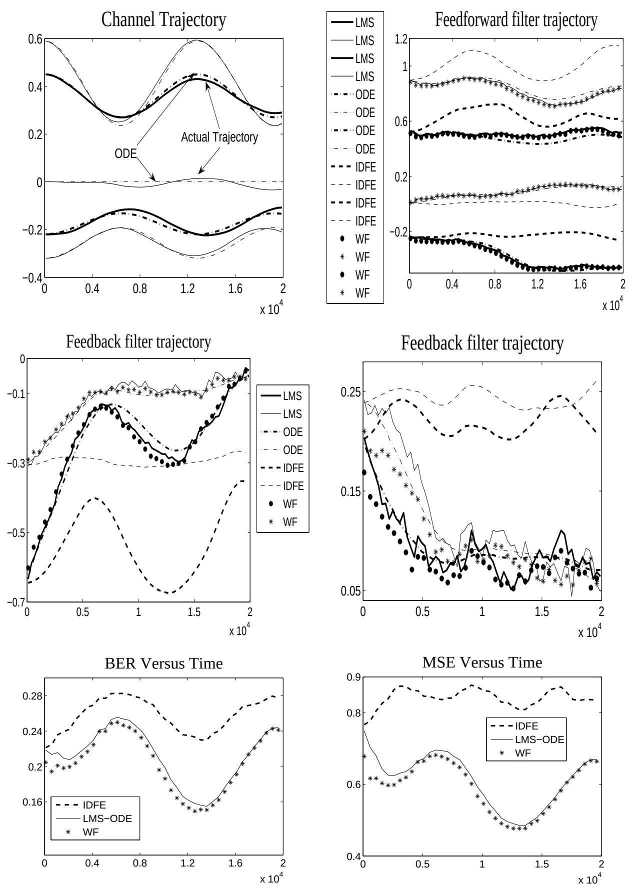

Fig. 3. A Marginally stable channel with $\mu = 1e^{-7}$, $d_1 = 1.9999998$, $d_2 = -1$ and mean a constant multiple of $[0.26,\ 0.34,\ 0.25,\ 0.064,\ -0.13,\ -0.19,\ -0.16,\ 0,\ 0.064,\ 0.064]$.

Finally, in Figure 4, we consider a stable channel with d_2 close to -1 (from the Figure, it actually looks like a marginally stable channel but its magnitude is reducing at a very small rate, $\frac{1+d_2}{\sqrt{\mu}}$, as d_2 is very close to -1). In this case, as is shown theoretically, a better ODE approximation is obtained by a second order ODE. Here, the channel trajectory is approximated by an exponentially reducing cosine waveform. We considered the AR(2) process, which approximates the fading channel with band limited and U-shaped spectrum and received with $f_dT = 0.001$. One can see that, the LMS-DFE once again tracks the instantaneous DFE-WF after initial transience (in this case more than half of the first cycle, as this is a fast varying channel) and that the IDFE performs poorly in comparison with the LMS-DFE and the DFE-WF.

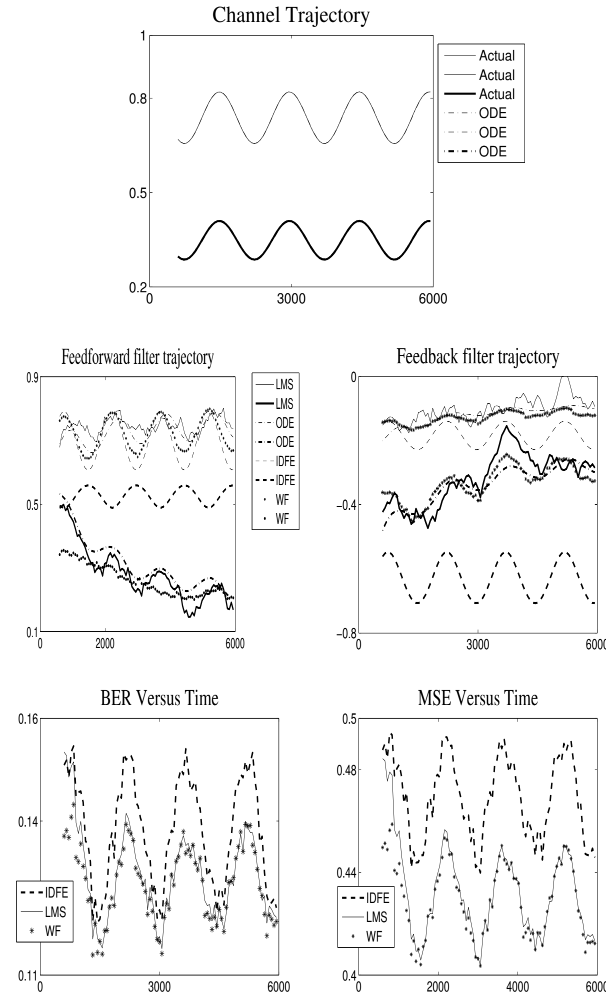

Fig. 4. A Stable channel with mean a constant multiple of $[0.41, .82, .41]$. It is obtained for $f_dT = 0.001$, i.e., $d_1 = 1.999982$, $d_2 = -.9999947$ and $\mu = 1.3997e-010$.

V. Conclusions

We study an LMS-DFE tracking a wireless channel, approximated by an AR(2) process. We considered a long-standing problem of tracking the true MSE optimal DFE. We approximated the LMS-DFE trajectory with the solution of a system of ODEs. Using this ODE approximation, we showed that the LMS-DFE comes close to the instantaneous DFE-WF after the initial transience. We also saw that the performance measures BER and MSE of the LMS-DFE are quite close to that of the DFE-WF after the transient period. We thus, conclude that the LMS-DFE can be used to track the DFE-WF.

Furthermore, we also compared the LMS-DFE with IDFE, the popular WF designed assuming perfect past decisions (also designed from perfect channel estimate). IDFE is shown to be far away from the DFE-WF (also from the LMS-DFE) trajectory throughout the entire time axis. Its performance (BER, MSE) is substantially worse than that of the DFE-WF and the LMS-DFE.

Appendix A

Proof of Theorem 1 : We consider a general system (13)-(14) in Appendix B and prove the ODE approximation for this in Theorem 2. The channel, equalizer pair, (θ_k, Z_k) given by equations (4), (3), is a specific example of the general system (13), (14). Thus Theorem 1 is proved if we show that (θ_k, Z_k) given by (4), (3) satisfies the assumptions **A**.1-**A**.3 and **B**.1-**B**.4 of Theorem 2 in Appendix B.

The AR(2) process $\{Z_k\}$ in (4) clearly satisfies the assumptions **A.**1 - **A.**3 as is shown in [7]. If (θ_k, Z_k) stay constant and equal (θ, Z), then $\{G_k\}$ is a Markov chain whose transition probabilities $P^{\theta,Z}(G, \mathbf{A})$ are a function of (θ, Z) alone. Thus condition **B**.1 is satisfied. One can easily see that conditions **B.**2, **B.**4 are satisfied by the LMS-DFE (the details are in [9]). Next we verify **B**.3 (More details of this proof are in [9]).

Verification of Assumptions B.3(b) and B.3(c)i :

One can clearly see that for all initial conditions j, y', i, y and all equalizers θ,

$$E^{\theta,Z}_{j,y'}\left(|U_n|^p\right) \leq \left\{ \begin{array}{ccc} C & if & p > N_f \\ C'\,|y'|^p & if & p \leq N_f, \end{array} \right. \quad (8)$$

$$E^{\theta,Z}_{(j,y');(i,y)}\left(|U_n - U'_n|^p\right) \leq \left\{ \begin{array}{ccc} C & if & p > N_f \\ C'\,|y - y'|^p & if & p \leq N_f, \end{array} \right. \quad (9)$$

whenever the channel $Z \in B(0, \epsilon')$ for any ϵ'. Using (8), (9) and Lemma 1, 2 of the Appendix C we show that all the hypothesis required for the Proposition 2, p.253, [2] are satisfied in [9]. By this Proposition (which is reproduced in Appendix D of [9]),

$$h_1(\theta, Z) = \lim_{k\to\infty} P^{\theta,Z^k} H_{1\theta}(j, y'), \quad (10)$$

exists for every channel, equalizer pair (θ, Z), and for any initial condition j, y'. Please note here that P^{θ,Z^k} as mentioned in Section II represents the k-step transition function of the Markov chain $\{G_k\}$, with the channel and the equalizer fixed at θ, Z. Also, by the same Proposition, for some constants $C < \infty$, $q > 0$ and $\rho < 1$,

$$\left| h_1(\theta, Z) - P^{\theta,Z^k} H_{1\theta}(i, y) \right| \leq C\rho^k(1 + |y|^q). \quad (11)$$

We also get the existence of,

$$\nu_{\theta,Z}(G) \;\; \triangleq \;\; \sum_{k\geq 0} P^{\theta,Z^k}(H_1(\theta, G) - h_1(\theta, Z)),$$

for all channel, equalizer pairs (θ, Z), which satisfy the assumption **B**.3b. Finally, by uniformity of all the inequalities for any $\theta \in K(\epsilon, M)$, $Z \in B(0, \epsilon')$, assumption **B**.3(c)i is satisfied, i.e.,

$$|\nu_{\theta,Z}(i, y)| \leq C_6(1 + |y|^q), \text{ with } C_6 < \infty.$$

In [9], we showed that for every fixed (θ, Z), a unique stationary measure of the Markov chain G_k exists and is continuous in (θ, Z). Hence $h_1(\theta, Z) = E^{\theta,Z} H_{1\theta}$.

Verification of Assumption B.3a :

One can easily see that for (θ, Z), (θ', Z') from a compact set, Q, there exists a constant C depending upon Q such that, for all $k > N_f$,

$$\begin{aligned} &|H_{1\theta}(G_k(\theta, Z)) - H_{1\theta'}(G_k(\theta', Z'))| \quad (12) \\ &\qquad \leq \;\; C\,|(\theta, Z) - (\theta', Z')|\,(1 + |N_k|). \end{aligned}$$

Using limit (10) and the upper bound (12) we get,

$$\begin{aligned} &|h_1(\theta, Z) - h_1(\theta', Z')| \\ &= \left| \lim_{k\to\infty} \left(P^{\theta,Z^k} H_{1\theta}(j, y') - P^{\theta',Z'^k} H_{1\theta'}(j, y') \right) \right| \\ &= \left| \lim_{k\to\infty} E^{\theta,Z;\theta',Z'}_{(i,y);(j,y')} \{ H_{1\theta}(G_k(\theta, Z)) - H_{1\theta'}(G_k(\theta', Z')) \} \right| \\ &\leq \;\; C'\,|(\theta, Z) - (\theta', Z')|\,, \end{aligned}$$

whenever (θ, Z), (θ', Z') are in a compact set Q.

Verification of Assumption B.3(c)ii :

Note that,

$$P^{\theta,Z}\nu_{\theta,Z}(j, y') = \sum_{k\geq 1} \left\{ P^{\theta,Z^k} H_{1\theta}(j, y') - h(\theta, Z) \right\}.$$

Hence, for any (θ, Z), (θ', Z') pair and any j, y',

$$\begin{aligned} &\left| P^{\theta,Z}\nu_{\theta,Z}(j, y') - P^{\theta',Z'}\nu_{\theta',Z'}(j, y') \right| \\ &\leq \;\; \left| \sum_{k=1}^{n} \left(P^{\theta,Z^k} H_{1\theta}(j, y') - P^{\theta',Z'^k} H_{1\theta}(i, y) \right) \right| \\ &\qquad +(n-1)\,|h_1(\theta, Z) - h_1(\theta', Z')| \\ &\qquad + \left| \sum_{k=1}^{n} \left(P^{\theta',Z'^k} H_{1\theta}(j, y') - P^{\theta',Z'^k} H_{1\theta'}(j, y') \right) \right| \\ &\qquad + \left| \sum_{k\geq n} \left\{ P^{\theta,Z^k} H_{1\theta}(j, y') - h(\theta, Z) \right\} \right| \\ &\qquad + \left| \sum_{k\geq n} \left\{ P^{\theta',Z'^k} H_{1\theta'}(j, y') - h(\theta', Z') \right\} \right|. \end{aligned}$$

The second term is bounded by a constant multiple of the term, $n\,|(\theta, Z) - (\theta', Z')|$, as h_1 is locally Lipschitz (proved in the previous para). Using the upper bound (11), one can see that the fourth and fifth terms are bounded by a constant multiple of the term $\rho^n(1+|y'|^q)$. Without loss of generality, we can further choose $q \geq 2$. The third term can be bounded because $|H_{1\theta}(i, y) - H_{1\theta'}(i, y)| \leq C\,|y|^2\,|\theta - \theta'|$ whenever θ, θ' come from a compact set and because $P^{\theta',Z'^k}_{j,y'}\,|y|^4 \leq C'\left(1 + |y'|^4\right)$ for any k. Hence we get,

$$\begin{aligned} &\left| P^{\theta,Z}\nu_{\theta,Z}(j, y') - P^{\theta',Z'}\nu_{\theta',Z'}(j, y') \right|^2 \\ &\leq \;\; B_1 n \sum_{k=1}^{n} \left| P^{\theta,Z^k} H_{1\theta}(j, y') - P^{\theta',Z'^k} H_{1\theta}(j, y') \right|^2 \\ &\qquad + \left(B_2 n^2\,|(\theta, Z) - (\theta', Z')|^2 + B_3\rho^{2n} \right)(1 + |y'|^{2q}). \end{aligned}$$

Fix $\epsilon > 0$, $M > 0$, $\epsilon' > 0$. Define $\tau \stackrel{\triangle}{=} \inf_n \left\{(\theta_n, Z_n) \notin \bar{K}(\epsilon, M) \times \bar{B}(0,\epsilon')\right\}$. By Lemma 3 and 4 for any m,

$$\begin{aligned}
&E_{i,y,\theta_0,Z_0}\Big\{ I(m+1 \le \tau) \left| P^{\theta_{m+1},Z_{m+1}} \nu_{\theta_{m+1},Z_{m+1}}(G_{m+1}) \right. \\
&\qquad\qquad \left. - P^{\theta_m, Z_m} \nu_{\theta_m,Z_m}(G_{m+1}) \right|^2 \Big\} \\
&\le \quad B_1 n^2 C_5 \mu^{0.5} \left(1 + |y|^4\right) + B_2 n^2 C_6 \mu^{0.5} \left(1+|y|^{2q}\right) \\
&\qquad + B_3 \rho^{2n} \left(1+|y|^{2q}\right) \\
&\le \quad B\left(n^2 \mu^{0.5} + \rho^{2n}\right)\left(1+|y|^{2q}\right).
\end{aligned}$$

Now, we choose $n = \left\lceil log\mu^{0.5}. \left(log\rho^2\right)^{-1} \right\rceil$, where $\lceil x \rceil$ represents the smallest integer $\ge x$. Then,

$$log\rho^{2n} \quad \ge \quad log\mu^{0.5}.$$

Hence we have for some constant C depending upon ρ,

$$n^2\mu^{0.5} + \rho^{2n} \le C\left(1 + \left|log\mu^{0.5}\right|^2\right)\mu^{0.5} + \mu^{0.5}.$$

Then for any $\lambda < 0.5$ (as $\lim_{x\to 0} x^\alpha (log(x))^2 = 0$ whenever $\alpha > 0$, by applying L'Hospital's rule twice),

$$\begin{aligned}
&E_{i,y,\theta_0,Z_0}\Big\{ I(m+1 \le \tau) \left| P^{\theta_{m+1},Z_{m+1}} \nu_{\theta_{m+1},Z_{m+1}}(G_{m+1}) \right. \\
&\qquad\qquad \left. - P^{\theta_m, Z_m} \nu_{\theta_m,Z_m}(G_{m+1}) \right|^2 \Big\} \\
&\qquad\qquad \le \quad B'(\lambda)\mu^\lambda \left(1+|y|^{2q}\right).
\end{aligned}$$

We have shown in the technical report [9] that the ODEs (5), (6) have unique bounded solution for any finite time interval. Hence the condition (17) given below is satisfied for any finite time T for some pair of compact sets Q_1, Q_2. Thus all the hypothesis of Theorem 1 are satisfied. ■

APPENDIX B : ODE APPROXIMATION OF A GENERAL SYSTEM

We consider the following general system,

$$\begin{aligned}
Z_{k+1} &= (1-d_2)Z_k + d_2 Z_{k-1} + \mu H(Z_k, W_k), && (13)\\
\theta_{k+1} &= \theta_k + \mu H_1(Z_k, \theta_k, G_{k+1}), && (14)
\end{aligned}$$

where equation (13) satisfies all the conditions in **A**.1–**A**.3 and the equation (14) satisfies the assumptions **B**.1–**B**.4, both given in the next para. We will show that the above equations can be approximated by the solution of the ODE's,

$$\begin{aligned}
(1+d_2)\,\dot{Z}(t) &= h(Z(t)), \quad \text{if } d_2 \in (-1,1], \\
\frac{d^2 Z(t)}{dt^2} &= h(Z(t)), \quad \text{if } d_2 = -1, \\
\frac{d^2 Z(t)}{dt^2} + \eta_1 \dot{Z}(t) &= h(Z(t)), \quad \text{if } d_2 \text{ is close to } -1,
\end{aligned} \qquad (15)$$

$$\dot{\theta}(t) = h_1(Z(t), \theta(t)), \qquad (16)$$

where the function h_1 is defined in the assumptions given below and $h(Z) = E[H(Z,W)]$, with $\eta_1 = \frac{1+d_2}{\sqrt{\mu}}$.

We make the following assumptions for the system (13) :

A.1 $\{W_k\}$ is an IID sequence.

A.2 $h(Z) = E\left[H\left(W_k, Z\right)\right]$ is a C^1 function.

A.3 For any compact set Q, there exists a constant $C(Q)$, such that $E|H(Z,W)|^2 \le C(Q)$ for all $Z \in Q$, where the expectation is taken wrt W.

We make the following assumptions for (14), which are similar to that in [2]. Let $D \subset \mathcal{R}^d$ be an open subset.

B.1 There exists a family $\{P_{Z,\theta}\}$ of transition probabilities $P_{Z,\theta}(G, \mathbf{A})$ such that, for any Borel subset $\mathbf{A}$,

$$P[G_{n+1} \in \mathbf{A} | \mathcal{F}_n] = P_{Z_n,\theta_n}(G_n, \mathbf{A})$$

where $\mathcal{F}_k \stackrel{\triangle}{=} \sigma(\theta_0, Z_0, Z_1, W_1, W_2, \cdots, W_k, G_0, G_1, \cdots, G_k)$. This in turn implies that the tuple $(G_k, \theta_k, Z_k, Z_{k-1})$ forms a Markov chain.

B.2 For any compact subset Q of D, there exist constants C_1, q_1 such that for all $(Z,\theta) \in D$ we have

$$|H_1(Z,\theta,G)| \quad \le \quad C_1(1+|G|^{q_1}).$$

B.3 There exists a function h_1 on D, and for each $Z, \theta \in D$ a function $\nu_{Z,\theta}(.)$ such that

a) h_1 is locally Lipschitz on D.

b) $(I - P_{Z,\theta})\nu_{Z,\theta}(G) = H_1(Z,\theta,G) - h_1(Z,\theta)$.

c) For all compact subsets Q of D, there exist constants C_3, C_4, q_3, q_4 and $\lambda \in [0.5, 1]$, such that for all $Z, \theta, Z^{'}, \theta^{'} \in Q$

i) $|\nu_{Z,\theta}(G)| \le C_3(1+|G|^{q_3})$,

ii) $E_{G,A}\{\left|P_{Z_k,\theta_k}\nu_{Z_k,\theta_k}(G_k) - P_{Z_k,\theta_k}\nu_{Z_{k-1},\theta_{k-1}}(G_k)\right|^2 I(k < \tau(Q))\} \le C_4\left(1+|G|^{q_4}\right)\mu^\lambda$.

B.4 For any compact set Q in D and for any $q > 0$, there exists a $\mu_q(Q) < \infty$, such that for all n, G, $A = (Z,\theta) \in \mathcal{R}^d$

$$\begin{aligned}
&E_{G,A}\left\{ I(Z_k, \theta_k \in Q, k \le n)\left(1+|G_{n+1}|^q\right)\right\} \\
&\qquad\qquad \le \quad \mu_q(Q)\left(1+|G|^q\right),
\end{aligned}$$

where $E_{G,A}$ represents the expectation taken with $G_0, Z_0, \theta_0 = G, Z, \theta$.

Let $Z(t, t_0, Z), \theta(t, t_0, \theta)$ represent the solutions of the ODEs (15), (16) with initial conditions $Z(t_0) = Z$, $\theta(t_0) = \theta$. For second order ODEs the additional initial condition is given by $\dot{Z}(t_0) = 0$. Let Q_1 and Q_2 be any two compact subsets of D, such that $Q_1 \subset Q_2$ and we can choose a $T > 0$ such that there exists an $\delta_0 > 0$ satisfying

$$d\left((Z(t,0,Z), \theta(t,0,\theta)), Q_2^c\right) \ge \delta_0, \qquad (17)$$

for all $(Z,\theta) \in Q_1$ and all t, $0 \le t \le T$. We prove Theorem 2, following the approach used in [2]. Parts of this theorem are presented in [7]. Let $V_k \stackrel{\triangle}{=} (Z_k, \theta_k)$ and $V(k) \stackrel{\triangle}{=} (Z(\mu^\alpha k, 0, Z), \theta(\mu k, 0, \theta))$, where $\alpha = 1$ if $Z(.,.,.)$ is solution of a first order ODE and $1/2$ otherwise.

Theorem 2: Assume, $E|H(Z,W)|^4 \le C_1(Q)$ for all Z in any given compact set Q of D. Also assume **A**.1–**A**.3 and **B**.1–**B**.4. Furthermore, pick compact sets Q_1, Q_2, and positive constants T, δ_0 satisfying (17). Then for all $\delta \le \delta_0$

and for any initial condition G, with $Z_{-1} = Z_0 = Z$, $\dot{Z}$ $(t_0) = 0$ (whenever $Z(.,.,.)$ is solution of a second order ODE), and $\theta_0 = \theta$,

$$P_{G,Z,\theta}\left\{ \sup_{1 \le k \le \lfloor \frac{T}{\mu^{\alpha}} \rfloor} |V_k, -V(k)| \ge \delta \right\} \to 0 \text{ as } \mu \to 0$$

uniformly for all $Z, \theta \in Q_1$.

Proof : The proof is given in the Technical Report [6].

Appendix C

In this Appendix we state the Lemmas used in Appendix A. Their proofs are provided in [9].

Lemma 1: Let $A(n) = \left\{ \hat{S}_k \ne \hat{S}'_k ; k = 1, 2, \cdots, n \right\}$. Given ϵ, M, ϵ', there exist positive $C_2 < \infty$, and $\rho < 1$ such that, for all $Z \in \bar{B}(0, \epsilon')$, $\theta \in \bar{K}(\epsilon, M)$ and all n,

$$P^{\theta,Z}_{(i,y);(j,y')}(A(n)) \le C_2 \rho^n .$$

Lemma 2: For any θ, Z, for any pair of initial conditions (j, y'), (i, y) and for any $n > N_f + N_L + N_b$,

$$P^{\theta,Z}_{(j,y');(i,y)} \left(\left\{ \hat{S}_{n-1} = \hat{S}'_{n-1}, \hat{S}_n \ne \hat{S}'_n \right\} \right) = 0.$$

Lemma 3: There exists a constant C_5 such that for all n, for all initial conditions (i, y), $(\theta_0, Z_0) \in \bar{K}(\epsilon, M) \times B(0, \epsilon')$,

$$E_{i,y;\theta_0,Z_0} \left\{ I(m+1 \le \tau) \left| P^{\theta_{m+1}, Z_{m+1}}{}^{n} H_{1\theta_{m+1}}(G_{m+1}) - P^{\theta_m, Z_m}{}^{n} H_{1\theta_{m+1}}(G_{m+1}) \right|^2 \right\} \le C_5 \mu^{0.5} \left(1 + |y|^4 \right).$$

Lemma 4: For any given ϵ, ϵ', M, there exists a constant $C_6 > 0$ such that for all initial conditions (i, y), $(\theta_0, Z_0) \in (\bar{K}(\epsilon, M) \times \bar{B}(0, \epsilon'))$ and for any $q > 0$,

$$E_{i,y;\theta_0,Z_0} \left\{ I(m+1 \le \tau) \left| (\theta_{m+1}, Z_{m+1}) - (\theta_m, Z_m) \right|^2 (1 + |U_{m+1}|^q) \right\} \le C_6 \mu^{0.5} (1 + |y|^q)$$

References

[1] C. A. Belfiore and J. H. Park Jr, "Decision feedback equalization", Proc. IEEE, Vol. 67, No. 8, Aug 1979, 1143 - 1156.

[2] A. Benveniste, M. Mietivier and P. Priouret, "Adaptive Algorithms and Stochastic Approximation", Springer-Verlag, April 1990.

[3] J. M. Cioffi, C. P. Dudevoir, M. V. Eyuboglu and G. D. Forney Jr, "MMSE Decision-Feedback Equalizers and Coding-Part I: General Results", IEEE Trans. Comm., Vol 43, 1995, 2582-2594.

[4] A. T. Erdogan, B. Hassibi and T. Kailath, "MIMO Decision Feedback Equalization from an H^{∞} Perspective". IEEE Trans. on Signal Process., Vol. 52, No. 3, March 2004, pp 734-745.

[5] S. Haykin, "Adaptive Filter theory", Third Edition, Prentice-Hall International inc, 1996.

[6] V. Kavitha and V. Sharma, "DD/Training-based Linear Equalizers for MIMO Fading channels: Tracking Performance", Technical report no: TR-PME-2007-01, DRDO-IISc program on mathematical engineering, ECE Dept., IISc, Bangalore, January 2007. (downloadable from http://pal.ece.iisc.ernet.in/PAM/tech_rep07.html)

[7] V. Kavitha and V. Sharma, "Tracking performance of an LMS-Linear Equalizer for fading channels", Presented in Allerton Conference 2006. (See also, V. Kavitha and V. Sharma, " Tracking performance of an LMS-Linear Equalizer for fading channels using ODE approach", Technical report no:TR-PME-2005-11, DRDO-IISc programme on mathematical engineering, ECE Dept., IISc, Bangalore, October 2005. (downloadable from http://pal.ece.iisc.ernet.in/PAM/tech_rep05.html))

[8] V. Kavitha and V. Sharma, "LMS Versus Wiener filter for a Decision feedback equalizer", Presented in Allerton Conference 2006. (See also, V. Kavitha, V. Sharma, "LMS Versus Wiener filter for a Decision feedback equalizer", Technical report no: TR-PME-2006-06, DRDO-IISc program on mathematical engineering, ECE Dept., IISc, Bangalore, June 2006. (downloadable from http://www.pal.ece.iisc.ernet.in/PAM/tech_rep06.html)).

[9] V. Kavitha and V. Sharma, "Tracking Analysis of an LMS Decision Feedback Equalizer for a Wireless Channel", Technical report no: TR-PME-2006-19, DRDO-IISc program on mathematical engineering, ECE Dept., IISc, Bangalore, October 2006. (downloadable from http://www.pal.ece.iisc.ernet.in/PAM/tech_rep06.html).

[10] C. Komminakis, C. Fragouli, A. Sayed and R.Wesel, "Multiple-input Multiple-output Fading Channel Tracking and Equalization using Kalman Estimation", IEEE Trans. on Signal Process., Vol 50, No. 5, May 2002, pp 1064-1076.

[11] H. Kushner and G. Yin, "Stochastic Approximation Algorithms and Applications", Springer, 1997.

[12] E. A. Lee and D. G. Messerschmitt, "Digital communications", 2nd Edn., Kluwer Academic Publishers, 1994.

[13] L. Lindbom, M. Sternad and A. Ahlen, "Tracking of Time-Varying Mobile Radio Channels Part II: A Case Study", IEEE Trans. Communications, Vol. 50, No. 1, Jan 2002.

[14] O. Macchi and E. Eweda, "Convergence Analysis of Self-Adaptive Equalizers", IEEE Trans on Information Theory, Vol.30, No.2, March 1984, pp 161-176.

[15] J. G. Proakis, "Digital Communications", New York: McGraw-Hill, 2000.

[16] D. P. Taylor, G. M. Vitetta, B. D. Hart and A. Mammela, "Wireless channel equalization", Eur. Trans. Telecom., Vol. 9, 1998, 117-143.

[17] C. Tidestav, A. Ahlen and M. Sternad,"Realizable MIMO Decision Feedback Equalizers: Structure and Design", IEEE Trans. on Signal Processing, Vol. 49, No. 1, Jan 2001, pp 121 - 133.

[18] M. Tsatsanis, G. Giannakis and G. Zhou, "Estimation and Equalization of Fading Channels with Random Coefficients", Signal Process., Vol 53, no. 2-3, pp 211-229, Sept 1996.

[19] M. Sternad, A. Ahlen and E. Lindskog, "Robust decision feedback equalizers", in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process, vol. 3, Minneapolis, MN, Apr. 1993, pp. 555-558.