COMPARISON OF TRAINING, BLIND AND SEMI BLIND EQUALIZERS IN MIMO FADING SYSTEMS USING CAPACITY AS MEASURE

Veeraruna Kavitha *Vinod Sharma*

Department of Electrical Communication Engg Indian Institute of Science, Bangalore, India
email: kavitha@pal.ece.iisc.ernet.in, vinod@ece.iisc.ernet.in

ABSTRACT

Semiblind/blind equalizers are believed to work unsatisfactorily in fading MIMO channels compared to training based methods, due to slow convergence or high computational complexity. In this paper we revisit this issue. Defining a 'composite' channel for each equalizer, we compare the three algorithms based on the capacity of this channel. We show that in Ricean (with Line of sight, LOS) environment semiblind/blind algorithms outperform training equalizers. But in Rayleigh channels it is better to use training based methods. We also find the optimum training size in training and semiblind methods.

1. INTRODUCTION

An important component in any communication system is efficient, accurate channel estimation and equalization. Due to time varying nature of the wireless channel, for a good channel estimate one generally uses the training based methods and needs to send the training sequence frequently. Therefore, a significant ($\sim 18\%$ in GSM) fraction of the channel capacity is consumed by the training sequence. The usual blind equalization techniques have been found to be inadequate ([4]) due to their slow convergence and/or high computational complexity. Therefore, semi-blind algorithms, which use some training sequence along with blind techniques have been proposed (Chapter 7 of [4] and references therein). In this paper, we provide a systematic comparison of the training, blind and semi-blind algorithms.

In comparing training based methods with blind algorithms one encounters the problem of comparing the loss in BW (Bandwidth) in training based methods (due to training symbols) with the gain in BER (due to better channel estimation/equalization accuracies) as compared to the blind algorithms. We overcome this problem by comparing these methods via the channel capacity they provide. Towards this goal, we combine the channel, the equalizer and the decoder to form a composite channel. The capacity of this composite channel will be a good measure for comparison. We assume block fading channel, with independent fades between frames and the capacity is computed for a frame.

Similar problem for the training based methods has also been studied in [1] and [5]. They obtain a lower bound on the channel capacity and find the optimal training sequence length (and also placement in case of [1]). We not only study the problem of optimal training sequence length, but also compare it to blind and semi blind methods. Obtaining channel capacity of the composite channel for training based methods, in our model is not difficult. But for the blind and semiblind methods, one needs to know the equalizer value at the end of the frame (or till the point the algorithm is updated). At that point the equalizer will often be away from the equilibrium point. Thus, we need the value of the equalizer at a specific time under transience and it depends upon the initial conditions, the value of the fading channel and the receiver noise realizations. Thus it is not practically feasible to obtain the capacity of such a composite channel. We circumvent this problem by using the result in [9], where given the initial conditions, the equalizer value at any time (even under transience) can be approximated by the solution of an ODE. Based on this analysis we find that in LOS conditions even though the (semi) blind algorithm might not have converged, it can perform better than training based methods.

The paper is organized as follows. Section 2 describes our model and approach . Section 3, 4, and 5 consider training, blind and semiblind algorithms respectively. Section 6 compares the 3 algorithms using few examples and Section 7 concludes the paper.

Notation

The following notation is used throughout the paper. All Bold capital letters represent complex matrices, capital letters represent real matrices and small letters represent the complex scalars. Bold small letters with bar on the top represent complex column vectors, while the same bold small letters without bar represent their real counter parts given by (for an n dimensional complex vector $\bar{\mathbf{y}}$), $\mathbf{y} \triangleq Real(\bar{\mathbf{y}}) := [\ \bar{\mathbf{y}}_{1,r} \quad \bar{\mathbf{y}}_{2,r} \quad \cdots \quad \bar{\mathbf{y}}_{n,r} \quad \bar{\mathbf{y}}_{1,i} \quad \cdots \quad \bar{\mathbf{y}}_{n,i}\ ]^T$. Here $\bar{\mathbf{y}}_{k,r} + i\bar{\mathbf{y}}_{k,r}$ is the k^{th} element of $\bar{\mathbf{y}}$. $\mathbf{A}^H$, $\mathbf{A}^T$, $\mathbf{A}^{HT}$ represent Hermitian, transpose and conjugate of matrix $\mathbf{A}$ respectively. Two more real vectors corresponding to a complex vector are defined as, $\tilde{\mathbf{y}} = Real(\bar{\mathbf{y}}^{HT})$ and $\check{\mathbf{y}} = Real(i\bar{\mathbf{y}}^{HT})$. Note that $\mathbf{e}^T\tilde{\mathbf{y}}, \mathbf{e}^T\check{\mathbf{y}}$ equal the real and imaginary parts of the complex inner product $\bar{\mathbf{e}}^H\bar{\mathbf{y}}$ respectively. For any complex matrix $\mathbf{H}$, $vect(\mathbf{H})$ represents $\mathbf{H}$ in vector form by concatenating elements of all row vectors one after the other. We will represent $vect(\mathbf{H})$ by $\bar{\mathbf{h}}$. $vect(\{\mathbf{H}_l\}_{l=0}^{L-1})$ denotes $vect([\mathbf{H}_0\ \mathbf{H}_1\ \cdots \mathbf{H}_{L-1}])$. $\mathbb{E}(.)$ represents expectation.

For any vector (matrix) $\bar{\mathbf{y}}$ ($\mathbf{H}$), $\hat{\bar{\mathbf{y}}}$ ($\hat{\mathbf{H}}$) represents its estimate. Complex sequence $\{\bar{\mathbf{a}}_l :\ n_1 \le l \le n_2\}$ is represented by $\bar{\mathbf{a}}_{n_1}^{n_2}$. Same notation is used for a real sequence. $\mathcal{E}_l$ represents l length convolutional matrix of dimension $lm \times (l+M-1)n$, formed from $m \times n$ dimensional complex matrices $\{\mathbf{E}_p\}_{p=0}^{M-1}$. L, M represents channel and equalizer length respectively and $M_L \triangleq M+L-1$.

2. THE MODEL AND OUR APPROACH

Consider a wireless channel with m transmit antennas and n receive antennas with $m \le n$. The time axis is divided into frames; each frame consisting of N channel uses. The transmitted symbols are chosen from a finite alphabet, $\mathcal{S} = \{s_1, s_2, \cdots, s_{n_S}\}$. At time k, vector $\bar{\mathbf{a}}(k) \varepsilon \mathcal{S}^m$ is transmitted from the m transmit antennas. We use $\mathcal{S}^{(N_t)}$ to represent the set $\mathcal{S}^{m(N-N_t)}$. We represent the elements of $\mathcal{S}^{(N_t)}$ by $\{\bar{\mathbf{s}}_i; 1 \le i \le (n_S)^{m(N-N_t)}\}$ i.e., each $\bar{\mathbf{s}}_i$ is a $m(N-N_t)$ length complex vector formed from elements of $\mathcal{S}$.

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

The channel $\{\mathbf{H}_l\}_{l=0}^{L-1}$is assumed constant during a frame (quasi static channel). We assume it to vary independently from frame to frame. We further assume that the training length $N_t \geq M_L$. This assumption is required for designing an equalizer directly (like in CMA algorithm). Thus, the channel estimate and/or equalizer, for a frame depends upon the received symbols during that frame only. Therefore, we will consider a single frame in this paper.

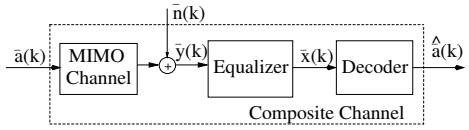

Fig. 1. Block Diagram of the System

The vectors received at the n receive antennas in any frame are, (we assume the first N_t correspond to training symbols)

$\bar{\mathbf{y}}(k) = \sum_{l=0}^{L-1} \mathbf{H}_l \bar{\mathbf{a}}(k-l) + \bar{\mathbf{n}}(k), \quad \forall 1 \leq k \leq N$

where $\bar{\mathbf{n}}(k)$ is an iid sequence of complex Gaussian vectors with mean 0 and co-variance $\sigma_n^2 I$ (denoted by $C\mathcal{N}(0, \sigma_n^2 I)$) and $\bar{\mathbf{h}} \sim C\mathcal{N}(\mu_{\bar{\mathbf{h}}}, C_h)$ (Note $\bar{\mathbf{h}} = vect(\{\mathbf{H}_l\}_{l=0}^{L-1})$). This is a Rayleigh or Rician channel. At the receiver one wants to estimate $\bar{\mathbf{a}}(k)$ (This is hard decoding. Quantized soft decoding can be taken care of in the same way). One common way is to use an equalizer at the receiver to nullify $\{\mathbf{H}_l\}_{l=0}^{L-1}$and then detect the transmitted symbol.

We assume that the channel statistics is available at the transmitter and the receiver but the actual channel $\{\mathbf{H}_l\}_{l=0}^{L-1}$is not known to the receiver and the transmitter. The receiver tries to estimate $\{\mathbf{H}_l\}_{l=0}^{L-1}$, or directly obtain an equalizer to estimate/detect the information symbols transmitted. For this the most common method used in wireless channels is to send a known training sequence in the frame. This is used by the receiver to estimate $\{\mathbf{H}_l\}_{l=0}^{L-1}$(say via Minimum Mean Square Error (MMSE) estimator) and then obtain an equalizer. In the rest of the frame, information symbols are transmitted and are decoded at the receiver using the equalizer. If a longer training sequence is used, we obtain a better channel estimate resulting in lower BER. However, we loss in channel BW (capacity) because information symbols are sent for a shorter duration. Thus one needs to find the 'optimal' training sequence length for a given channel. Alternatively, one can estimate an equalizer using only the statistics of the received and transmitted signal. These are blind methods and do not require training sequences, but may not be accurate. One can expect that by combining some of the blind methods with the training based methods we can obtain the same performance with a shorter training sequence and hence more capacity.

To address the issue of fair comparison of various equalizers (training, blind and semiblind), we form a 'composite' channel, made of the channel, equalizer and the decoder, with input as the transmitted information data vector corresponding to one complete block $\bar{\mathbf{a}}_{N_t+1}^N$. Corresponding decisions $\hat{\bar{\mathbf{a}}}_{N_t+1}^N$form output for this composite channel. It forms a finite input - output alphabet time invariant channel. This is because the channel state is not known to the transmitter and hence the transmitter would experience average behavior in every frame. We will show that one can compute the composite channel's transition matrix and hence it's capacity $C(N_t)$, for a given N_t, once the statistics of the original channel are known at the transmitter and receiver. We then find the optimal N_t which maximizes $C(N_t)$.

Obtaining the capacity of the composite channel for training based methods is easy at least conceptually, and is carried out in section 3. Next we will consider blind algorithms. We consider, CMA as it has been one of the most used and successful algorithms ([4]). Here using the results in [9], we obtain a lower bound on capacity by limiting the set of input distributions to stationary and ergodic processes (stationarity can be relaxed to asymptotic stationarity). Comparing the capacity of the training based equalizer with that of a lower bound in case of blind/semi-blind algorithms ensures that the later is better by at least the amount obtained.

3. TRAINING BASED CHANNEL EQUALIZER

The MMSE estimator (Wiener filter)([8]) of the channel is obtained using mN_t training symbols. A MMSE equalizer is then designed using the channel estimate. The symbols obtained from the equalizer are decoded using minimum distance criterion (equivalent to ML decoding in Gaussian channels).

Let $\bar{\mathbf{y}}_{TS} \triangleq \bar{\mathbf{y}}_L^{N_t} = \mathbf{A}_{TS}\bar{\mathbf{h}} + \bar{\mathbf{n}}_L^{N_t}$. This represents received samples corresponding to the last $m(N_t - L + 1)$ training symbols. Here $\mathbf{A}_{TS}$ is an $n(N_t - L + 1) \times Lnm$ matrix formed appropriately using training symbols $\bar{\mathbf{a}}_1^{N_t}$. Note that only the last $N_t - L + 1$ samples of the output corresponding to training symbols can be used for channel estimation. The MMSE channel estimator ([8]) is given by $\hat{\bar{\mathbf{h}}} = \mu_{\bar{\mathbf{h}}} + C_h \mathbf{A}_{TS}^H (\mathbf{A}_{TS} C_h \mathbf{A}_{TS}^H + \sigma_n^2 I)^{-1} (\bar{\mathbf{y}}_{TS} - \mathbf{A}_{TS}\mu_{\bar{\mathbf{h}}})$, and $(\bar{\mathbf{h}}, \hat{\bar{\mathbf{h}}})$ are jointly Gaussian with mean $(\mu_{\bar{\mathbf{h}}}, \mu_{\bar{\mathbf{h}}})$.

The corresponding M length left MMSE equalizer (of dimension $m \times Mn$) equals, $\mathbf{E}(\hat{\bar{\mathbf{h}}}) = \left[\left(\hat{\mathcal{H}}_M \hat{\mathcal{H}}_M^H + \sigma_n^2 I \right)^{-1} \hat{\mathcal{H}}_M \mathcal{I}_{M_L n, m} \right]^H$, where $\mathcal{I}_{M_L n, m}$ is the matrix formed by first m columns of the identity matrix of dimension $M_L n \times M_L n$, and $\hat{\mathcal{H}}_M$ represents the $Mn \times M_L m$ dimensional convolutional matrix formed using $\hat{\bar{\mathbf{h}}}$. Using $\bar{\mathbf{h}}$ and $\mathbf{E}(\hat{\bar{\mathbf{h}}})$, define convolutional matrices, $\mathcal{H}_{N - N_t + M - 1}$ and $\mathcal{E}_{N - N_t}$, each of dimensions $(N - N_t + M - 1)n \times (N - N_t + M_L - 1)m$, and $(N - N_t)m \times (N - N_t + M - 1)n$ respectively. The equalizer output corresponding to transmitted input vector $\bar{\mathbf{a}}_{N_t+1}^N$will be,

$$\bar{\mathbf{x}}_{N_t+1}^N = \mathcal{E}_{N-N_t} \left(\mathcal{H}_{N - N_t + M - 1} \bar{\mathbf{a}}_{N_t - M_L + 2}^{N} + \bar{\mathbf{n}}_{N_t - M + 2}^{N} \right). \quad (1)$$

It is clear that the vector $\bar{\mathbf{a}}_{N_t - M_L + 2}^{N_t}$, corresponding to the tail of the training sequence is common to all the frames and is known. The output, $\hat{\bar{\mathbf{a}}}_{N_t+1}^N$, of the decoder is obtained from symbol by symbol decoding of $\bar{\mathbf{x}}_{N_t+1}^N$. We compute the overall transition probabilities, $\{P(\hat{\bar{\mathbf{a}}}_{N_t+1}^N / \bar{\mathbf{a}}_{N_t+1}^N)\}$, of the composite channel by first computing transition probabilities, $\{P(\hat{\bar{\mathbf{a}}}_{N_t+1}^N / \bar{\mathbf{a}}_{N_t+1}^N, \bar{\mathbf{h}}, \hat{\bar{\mathbf{h}}})\}$, given $\bar{\mathbf{h}}, \hat{\bar{\mathbf{h}}}$ and then averaging over all values of $\bar{\mathbf{h}}, \hat{\bar{\mathbf{h}}}$. ($\bar{\mathbf{h}}, \hat{\bar{\mathbf{h}}}$ are Gaussian with known joint distribution). By defining B_i as $B_i := \{\mathbf{x}_{N_t+1}^N \in R^{2m(N - N_t)} : \cap_{l \neq i} \|\bar{\mathbf{s}}_l - \bar{\mathbf{x}}_{N_t+1}^N\|^2 \geq \|\bar{\mathbf{s}}_i - \bar{\mathbf{x}}_{N_t+1}^N\|^2\}$, the transition probabilities given $(\bar{\mathbf{h}}, \hat{\bar{\mathbf{h}}})$, from (1), are given by $P(\hat{\bar{\mathbf{a}}}_{N_t+1}^N = \bar{\mathbf{s}}_i / \bar{\mathbf{a}}_{N_t+1}^N = \bar{\mathbf{s}}_j, \bar{\mathbf{h}}, \hat{\bar{\mathbf{h}}}) = P_{rob}(\mathbf{x}_{N_t+1}^N \in B_i / \bar{\mathbf{a}}_{N_t+1}^N = \bar{\mathbf{s}}_j, \bar{\mathbf{h}}, \hat{\bar{\mathbf{h}}})$.

It is easy to see that the overall transition probabilities of the composite channel, $\{P(\hat{\bar{\mathbf{a}}}_{N_t+1}^N / \bar{\mathbf{a}}_{N_t+1}^N)\}$, can be computed at its receiver and transmitter once the statistics of the original channel is known. Given N_t, the composite channel becomes a time invariant channel with $C(N_t) = \sup_{P(\bar{\mathbf{a}}_{N_t+1}^N) \in \mathcal{P}(\mathcal{S}^{(N_t)})} I(\hat{\bar{\mathbf{a}}}_{N_t+1}^N, \bar{\mathbf{a}}_{N_t+1}^N)$, where $I(\hat{\bar{\mathbf{a}}}_{N_t+1}^N, \bar{\mathbf{a}}_{N_t+1}^N)$ represents the mutual information with input pmf (probability mass function) $P(\bar{\mathbf{a}}_{N_t+1}^N)$ and transition probabilities $\{P(\hat{\bar{\mathbf{a}}}_{N_t+1}^N / \bar{\mathbf{a}}_{N_t+1}^N)\}$. $\mathcal{P}(\mathcal{S}^{(N_t)})$ is the set of probability mass functions on $\mathcal{S}^{(N_t)}$ and is compact (Note $\bar{\mathbf{a}}_{N_t+1}^N \in \mathcal{S}^{(N_t)}$). Since $P(\hat{\bar{\mathbf{a}}}_{N_t+1}^N / \bar{\mathbf{a}}_{N_t+1}^N)$ is independent of the input pmf $P(\bar{\mathbf{a}}_{N_t+1}^N)$, the mutual information $I(\hat{\bar{\mathbf{a}}}_{N_t+1}^N, \bar{\mathbf{a}}_{N_t+1}^N)$ is a strictly concave function of $P(\bar{\mathbf{a}}_{N_t+1}^N)$ ([2], p.31) and hence optimization over the convex set results in a global maximum. Capacity for training equalizer equals $C(N_t^*)$, where N_t^* is the optimum training length.

3.1. Computation of $C(N_t)$

As in GSM, the block length N can be as large as 140 or even more. In such cases direct computation of $C(N_t)$ may be difficult. Hence we calculate a lower bound to the capacity by restricting the input vector $\bar{\mathbf{a}}_{N_t+1}^N$ to be Markov chain. Given a specific Markov chain, $I(\hat{\bar{\mathbf{a}}}_{N_t+1}^N; \bar{\mathbf{a}}_{N_t+1}^N)$, is computed recursively (note that output $\hat{\bar{\mathbf{a}}}(i)$ is Hidden Markov) similar to the way in [10]. We show below that this provides a tight lower bound.

Let $\theta_h := [\ \mu_{\bar{\mathbf{h}}}^T \quad vect(C_h)^T\ ]^T$. This parametric vector characterizes complex Gaussian random variable $\bar{\mathbf{h}}$ completely. Let $g(\theta_h)$ represent the conditional probabilities of the composite channel for a given θ_h and let $f(P(.|.))$ be a capacity achieving input distribution for given conditional probabilities $P(.|.)$. The proof of the following lemma is provided in [7].

Lemma 1 *g and f are continuous functions. Hence $f \circ g$ is continuous in θ_h.*

For channels with zero ISI (say with θ_h^0 as parametric vector), it is easy to see that $g(\theta_h^0)$ is memoryless and $f \circ g(\theta_h^0)$ will be IID. Therefore, by the above lemma, for channels with small ISI or for systems with equalizers compensating the ISI to a good extent, a tight lower bound on the capacity can be achieved by restricting the input distributions to l-step Markov chains. Recently, we found that the lower bound obtained by using the uniform IID input is itself tight in most cases.

4. BLIND CMA EQUALIZER

The CMA cost function for a single user MIMO channel with same source alphabet for all m transmit antennae can be written as ([3]),
$\mathbf{E}_{CMA} = \arg\min_{\mathbf{E}=[\bar{\mathbf{e}}_1,\ldots,\bar{\mathbf{e}}_m]^T} \Sigma_{l=1}^m \mathbb{E}\left(|\bar{\mathbf{e}}_l^H \bar{\mathbf{y}}_{k-M+1}^k|^2 - R_2^2\right)^2$
or equivalently (terms in the summation are positive)

$$\bar{\mathbf{e}}_{cma\,l} = \arg\min_{\bar{\mathbf{e}}_l} \mathbb{E}\left(|\bar{\mathbf{e}}_l^H \bar{\mathbf{y}}_{k-M+1}^k|^2 - R_2^2\right)^2,\ l=1,2,\ldots m$$

where $\bar{\mathbf{e}}_l^T$, nM length vector, represents the l^{th} row and
$R_2 = \mathbb{E}|\bar{\mathbf{a}}_{N-M_L+1}^N|^4 / \mathbb{E}|\bar{\mathbf{a}}_{N-M_L+1}^N|^2$ (we assume input is stationary).

To obtain the above optimum, the corresponding m update equations for a given value of $\bar{\mathbf{h}}$ are ($1 \le l \le m$, $M_L \le k \le N$), (Note that $\bar{\mathbf{y}}_{k-M+1}^k = \bar{\mathbf{z}}(k) + \bar{\mathbf{n}}_{k-M+1}^k$, with $\bar{\mathbf{z}}(k) \triangleq \mathcal{H}_M \bar{\mathbf{a}}_{k-M_L+1}^k$)

$$\mathbf{e}_l(k+1) = \mathbf{e}_l(k) + \mu H_{CMA}\left(\mathbf{e}_l(k), \mathbf{z}(k), \mathbf{n}_{k-M+1}^k\right) \qquad (2)$$

where with $\bar{\mathbf{y}} = \bar{\mathbf{z}} + \bar{\mathbf{n}}$ ($\tilde{\mathbf{y}}$ and $\check{\mathbf{y}}$ are defined in notations)
$H_{CMA}(\mathbf{e}, \mathbf{z}, \mathbf{n}) \triangleq ((\mathbf{e}^T\mathbf{y})^2 + (\mathbf{e}^T\check{\mathbf{y}})^2 - R_2^2)\,((\mathbf{e}^T\tilde{\mathbf{y}})\tilde{\mathbf{y}} + (\mathbf{e}^T\check{\mathbf{y}})\check{\mathbf{y}})$.

A close look at (2) shows that all m sub cost functions are same and the different equalizers should be initialized appropriately to extract the desired source symbols. In [3] a new joint CMA algorithm is proposed that ensures that the MIMO CMA separates all the sources successfully irrespective of the initial conditions. In this work, we choose the initial condition $\mathbf{E}_0^*$ (which will be used in all frames) such that the channel capacity is optimized. This solves initialization problem to a great extent in blind case with the original CMA itself. The problem will be solved to a greater extent in the semiblind algorithm, as here a rough estimate of the training based equalizer forms the initializer.

The equalizer adaptation(2) can be started only after leaving out the first $M_L - 1$ received samples. This ensures that the equalizer adaptation is independent of previous frame symbols. One may also update the equalizer tap only for a fraction of the symbols in the frame. This reduces delay in processing.

For real constellations like BPSK, a more suitable CMA cost function would be $((\mathbf{e}^T\mathbf{y})^2 - R_2^2)^2 + (\mathbf{e}^T\check{\mathbf{y}})^4$. We conducted simulations with this cost function for BPSK over complex channels.

In the next subsection, we show how analytically we can obtain the value of CMA equalizer approximately at any time t and then proceed with obtaining the channel capacity.

4.1. CMA Equalizer approximated by ODE

Now we assume that the input is stationary and ergodic. Each of the m update equations in (2) is similar to the CMA update equation for SISO. Therefore it is easy to see that all the proofs in [9] for convergence of the CMA trajectory to the solution of an ODE (Ordinary differential equation) hold. Thus the update equation(2) for any given $\bar{\mathbf{h}}$, can be approximated by the trajectory of the ODE,

$$\dot{\mathbf{e}}_l(t) = \hat{H}_{CMA}(\mathbf{e}_l(t)) \triangleq \mathbb{E}_{\mathbf{z}}[\mathbb{E}_{\mathbf{n}}(H_{CMA}(\mathbf{e}_l(t), \mathbf{z}, \mathbf{n}))] \qquad (3)$$

where $\bar{\mathbf{z}} \triangleq \mathcal{H}_M \bar{\mathbf{a}}_{N-M_L+1}^N$ (under stationarity). The approximation can be made accurate with high probability by taking μ small.

We can solve (3) numerically for each l and obtain the equalizer $\mathbf{E}(T)$ at time $T = \mu(N - M_L + 1)$ (smaller T if equalizer is updated only for a fraction of the frame) which approximates the CMA equalizer at the end of the frame. These co-efficients are used for decoding of the entire frame. It is clear that $\mathbf{E}(T)$ is a function of $\bar{\mathbf{h}}$, $\mathbf{E}_0$ and $P(\bar{\mathbf{a}}_{N-M_L+1}^N)$. Define $\mathcal{E}(T) := \mathcal{E}_{N-N_t}(T)$, the $N - N_t$ length convolutional matrix constructed using $\mathbf{E}(T)$.
Let $\mathcal{E}_{\mathcal{R}} \triangleq \begin{bmatrix} Re(\mathcal{E}(T)) & -Im(\mathcal{E}(T)) \\ Im(\mathcal{E}(T)) & Re(\mathcal{E}(T)) \end{bmatrix}$. Here $Re(\mathcal{E}(T))$, $Im(\mathcal{E}(T))$ represent matrices formed by keeping only the real or imaginary part, respectively of, each component of the matrix $\mathcal{E}(T)$.

As in the previous section, conditioned on $\bar{\mathbf{h}}$, $P(\bar{\mathbf{a}}_{N_t+1}^N)$ and $\mathbf{E}_0$, the transitional probabilities of the approximate composite channel obtained by solving the ODE are (the probabilities of the initial $L-1$ symbols will be different, but the number is very small with respect to the block length and hence it can be neglected),

$$P(\hat{\bar{\mathbf{a}}}_{N_t+1}^N = \bar{\mathbf{s}}_i / \bar{\mathbf{a}}_{N_t+1}^N = \bar{\mathbf{s}}_j; \mathbf{E}_0, P(\bar{\mathbf{a}}_{N_t+1}^N), \bar{\mathbf{h}}) = Prob(\mathcal{E}_{\mathcal{R}} \mathbf{y}_{N_t-M+2}^N \in B_i / \bar{\mathbf{a}}_{N_t+1}^N = \bar{\mathbf{s}}_j; \mathbf{E}_0, P(\bar{\mathbf{a}}_{N_t+1}^N), \bar{\mathbf{h}})$$

The overall transition probabilities for a given $P(\bar{\mathbf{a}}_{N_t+1}^N)$ and $\mathbf{E}_0$ are now obtained by averaging over $\bar{\mathbf{h}}$. The capacity(lower bound as input is restricted to be stationary and ergodic)
$C(\mathbf{E}_0) := \sup_{P(\bar{\mathbf{a}}_{N_t+1}^N)} I(\hat{\bar{\mathbf{a}}}_{N_t+1}^N; \bar{\mathbf{a}}_{N_t+1}^N / \mathbf{E}_0, P(\bar{\mathbf{a}}_{N_t+1}^N))$ of the approximate channel for a given $\mathbf{E}_0$, is finite. The overall capacity of this discrete memoryless channel is $C_{CMA} \approx \sup_{\mathbf{E}_0} C(\mathbf{E}_0)$, which is also finite. When the receiver and the transmitter have channel statistics, one can choose ${\mathbf{E}_0}^*$ and $P^*(\bar{\mathbf{a}}_{N_t+1}^N)$, such that $I(\hat{\bar{\mathbf{a}}}_{N_t+1}^N; \bar{\mathbf{a}}_{N_t+1}^N / \mathbf{E}_0^*, P^*(\bar{\mathbf{a}}_{N_t+1}^N))$ is as close as possible to C_{CMA} in principle at least. In simulations we obtained a lower bound by estimating the mutual information for uniform IID input distribution and $\mathbf{E}_0^* = \mathbf{E}(\mu_{\bar{\mathbf{h}}})$ ($E(\hat{\bar{\mathbf{h}}})$ is defined in previous section) and found it to be tight.

For SIMO, $m=1$, we have, (see proof in [7])

Lemma 2 *$I(\hat{\bar{\mathbf{a}}}_{N_t+1}^N; \bar{\mathbf{a}}_{N_t+1}^N / \bar{\mathbf{e}}_0, P(\bar{\mathbf{a}}_{N_t+1}^N))$ is a continuous function of $\bar{\mathbf{e}}_0$ and $P(\bar{\mathbf{a}}_{N_t+1}^N)$. Also $C(\bar{\mathbf{e}}_0)$ is continuous in $\bar{\mathbf{e}}_0$.*

Thus in SIMO case, there exist $\bar{\mathbf{e}}_0^*$ and $P^*(\bar{\mathbf{a}}_{N_t+1}^N)$, such that the corresponding suprema are achieved (when $\bar{\mathbf{e}}_0$ is restricted to a compact domain). More recently, we have obtained this result for MIMO case also.

5. SEMI-BLIND CMA ALGORITHM

In this variant of the semi-blind algorithm, we use MMSE equalizer of the training based channel estimator $\mathbf{E}(\hat{\bar{\mathbf{h}}})$ obtained in section 3 as the initializer for the CMA algorithm. The equalizer co-efficients obtained from the CMA at the end of the frame (or

fraction of the frame) are used for decoding of data for the whole frame. Once again we use the ODE approximation of the CMA trajectory in the capacity analysis. The difference from the blind case, being that the initializer $\mathbf{E}_0$ for the CMA is given by the training based channel estimator. Now $T{=}\mu(N{-}N_t)$. Then the conditional probabilities are obtained by first conditioning on and then averaging with respect to $\bar{\mathbf{h}}, \hat{\bar{\mathbf{h}}}$ as in section 3.

As before, given N_t, the capacity (lower bound), $C_{SB}(N_t){=}\sup_{P(\bar{\mathbf{a}}^N_{N_t+1})} I(\hat{\bar{\mathbf{a}}}^N_{N_t+1}; \bar{\mathbf{a}}^N_{N_t+1}/P(\bar{\mathbf{a}}^N_{N_t+1}))$, exists and $C_{SB} \approx C_{SB}(N_t^*)$. Again for SIMO, we have (proof is in [7]),

Lemma 3 *$I(\hat{\bar{\mathbf{a}}}^N_{N_t+1}; \bar{\mathbf{a}}^N_{N_t+1}/P(\bar{\mathbf{a}}^N_{N_t+1}))$ is a continuous function of $P(\bar{\mathbf{a}}^N_{N_t+1})$ and hence supremum can be achieved.*

6. SIMULATIONS

Simulations have been carried out over complex Gaussian channels with BPSK modulations. We consider $2{\times}2$ MIMO channels for simulations. We set $C_h{=}\sigma_h^2 I$. We normalized the channel gain to one for both the receive antennas. We fixed N =64, L =M =2. We calculated the capacity of the composite channel for all the equalizers. We noticed that the equalizer value at the end of the frame was away from the equilibrium point in both blind and semiblind algorithms. In Figure 2 we plotted capacity versus transmitted power with $\sigma_n^2{=}1$. We varied K-factor (ratio of power in the mean component to that in the varying component) of the channel during our experiments. For large K, there is an improvement of up to 2 dB ($\approx 50\%$ improvement in TX power) in semiblind/blind (blind being the best) algorithms at around 12dB compared to the training method. At K =1, the improvement is $\approx 26\%$ and semiblind is the best. As the K-factor approaches 0 (Rayleigh channel), improvement in semiblind diminishes but the blind becomes much worse. Infact, in Rayleigh channel, the blind capacity is almost zero.

We have observed that for training and semiblind equalizers, the capacity increases with N_t, reaches a maximum and starts decreasing. From this, one can estimate the 'optimal' number of training symbols, N_t^* (see Table 1 for some examples).

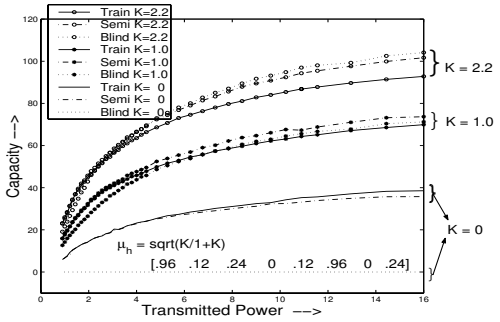

Fig. 2. Capacity versus transmitted power

Table 1 also shows comparison of various equalizers with respect to noise variance for two different K factors, for a fixed transmit power $E_A{=}12dB$. These channels have non zero mean for imaginary terms but ISI and ICI have zero mean. For high K systems, blind algorithm is the best at reasonable SNR's. This is because the other algorithms are loosing in capacity because of training symbols. But we have seen that at very low SNR's (near 0dB, not shown in the table, but this effect can be seen in Figure2) eventually the blind algorithm becomes worse. As commented in section 4, we observed in case of high K, the blind algorithm was successfully separating all the sources (which also explains comparatively good performance of blind algorithms at high K). But for low K, the blind algorithm was not separating the sources resulting in bad performance (K = 0.7 case in Table 1). Also from Table 1, we can see an SNR improvement of 66% (18%) at $\sigma_n^2{=}1$ for K=2.5 (K=0.7) in semiblind/blind over the training based method.

Table 1. $\mu_{\bar{\mathbf{h}}}{=}\sqrt{K/(1+K)}$ $[0.8{+}i0.6\ 0\ 0\ 0\ 0\ 0.8{+}i0.6\ 0\ 0]$, $E_A{=}12dB$

σ_n^2	K = 2.5			K = 0.7		
	Training (C, N_t^*)	Semi (C, N_t^*)	Blind C	Training (C, N_t^*)	Semi (C, N_t^*)	Blind C
1	(102.6, 3)	(110.8, 3)	114.4	(64.2, 9)	(66.4, 7)	62.0
1.2	(99.9, 4)	(108.5, 2)	112.3	(61.9, 9)	(64.4, 5)	60.2
2	(89.7, 3)	(99.0, 3)	102.8	(54.5, 8)	(57.8, 4)	53.7
4	(70.9, 3)	(76.6, 2)	80.3	(41.4, 7)	(43.4, 4)	41.3

7. CONCLUSIONS

We compared blind/semiblind equalizers with training based algorithms. Information capacity is the most appropriate measure for this comparison. We observed that the semiblind methods perform superior to training as well as blind methods in LOS conditions ($\approx$50 to 66% improvement in transmit power) even when they have not converged to the equilibrium point. But for Rayleigh fading, the semiblind methods are bad compared to training based and the blind methods become completely useless. Our approach could also be used to obtain the optimum number of training symbols. Experiments have been conducted for flat fading (no ISI) channels also. But the improvement was only about 20% [6]. More recently, we modified the semiblind algorithm, where we change the stepsize μ in (2) based on $\hat{\bar{\mathbf{h}}}$ and obtained a significant improvement.

8. REFERENCES

[1] S. Adireddy, L. Tong, H. Viswanathan, "Optimal Placement of Training for Frequency-Selective Block-Fading Channels," *IEEE Trans. on Inf. Theory*, Vol. 48, 2338-2353, 2002.

[2] T. Cover, J. A. Thomas, "Elements of Information Theory", John Wiley & Sons, 1991.

[3] Z. Ding, Y. Li, "Blind Equalization and Identification", Marcel Dekker Inc, NewYork 2000.

[4] G.B.Giannakis, Y.Hua, P.Stoica, L.Tong, "Signal Processing Advances in Wireless and Mobile Communications", Trends in Channel estimation and equalization. Vol.1, Prentice Hall, Upper Saddle River, NJ 2000.

[5] B. Hassibi, B.M. Hochwald, "How much training is needed in Multiple-Antenna wireless links?" *IEEE Trans. on Inf. Theory*, Vol. 49, 951-963, 2003.

[6] V.Kavitha, V.Sharma, "Information Theoretic Comparison of Training, Blind and Semi Blind Signal Separation Algorithms in MIMO Systems", SPCOM 2004.

[7] V.Kavitha, V.Sharma, "Comparison of Training, Blind and Semi Blind Equalizers in MIMO Fading Systems using Capacity as measure", Technical report no: TR-PME-2004-10, DRDO-IISc programme on mathematical engineering, ECE Dept., IISc, Bangalore, Sept 2004. (downloadable from http://pal.ece.iisc.ernet.in/PAM/tech_rep04.html)

[8] S. M. Kay, "Fundamentals of Statistical Signal Processing ESTIMATION THEORY", Prentice Hall, New Jersey, 1993.

[9] V.Sharma, Naveen Raj V, "Convergence and Performance Analysis of Godard family and Multi Modulus Algorithms for Blind Equalization", to appear in IEEE Trans. SP.

[10] S. Singh and Vinod Sharma, "Algorithms for computing the capacity of Markov channels with side information", IEEE Information Theory Workshop, 2002.