Boosting MMSE receivers using AdaBoost

Onkar Dabeer
School of Technology and Computer Science
Tata Institute of Fundamental Research
Mumbai, India
Email: onkar@tcs.tifr.res.in

Srinidhi Nagaraja, A. Chockalingam
Department of ECE
Indian Institute of Science
Bangalore 560012, India
Email: {srinidhi,achockal}@ece.iisc.ernet.in

***Abstract*— MIMO systems with several antennas at the transmitter and receiver have the potential to enable high throughputs. One of the challenges in realizing this potential is the design of receivers that scale in terms of computation and performance. Towards this end, in this paper, we propose a receiver that uses a committee of linear receivers, whose parameters are estimated from training data using a variant of the *AdaBoost* algorithm, a celebrated supervised classification algorithm in machine learning. We call our receiver *boosted MMSE (B-MMSE) receiver* and we study its performance via simulations. The channel matrix is estimated from the pilot using the MMSE technique. We find that for a 4×4 system with a Rayleigh fading channel, using BPSK at a bit error rate (BER) of 10^{-2}, our receiver using the estimated channel matrix needs 2 dB less power than the plain MMSE receiver based on the estimated channel matrix. On an average, excluding the training phase, the receiver complexity equals 1.25 linear receivers. Thus the B-MMSE receiver is slightly more complex than the MMSE receiver and substantially less complex than the maximum-likelihood receiver. We also show that the coded BER performance with an outer rate-3/4 Turbo-code has a gain of 5 dB at 10^{-4} coded BER when the demodulated output from the proposed method is used in place of plain MMSE detection with MMSE estimated channel.**

Keywords: *AdaBoost, statistical boosting, MIMO detection, MMSE receiver, boosted-MMSE, machine learning.*

I. Introduction

Multiple-input multiple-output (MIMO) systems have received wide attention over the last decade owing to the exciting opportunities they provide in terms of increased spectral efficiency and/or diversity gain. In particular, the capacity of point-to-point MIMO channel increases linearly with the minimum of the number of transmit and receive antennas [1], [2]. This has led to the use of MIMO systems in wireless communications standards such as the IEEE 802.11n/802.11ac [3] and LTE [4]. One of the main impediments to successful deployment of MIMO systems is the increased complexity in decoding the received signals. Though the MIMO architecture provides a way around deep fades in a single path by transmitting data along multiple paths, this means that the signal from each antenna interferes with those from other antennas. Recovering the data from such interfering signals in the presence of additive noise using a low complexity receiver is a challenge and many authors have addressed it. Many algorithms have been proposed, which vary in the degree of performance-complexity trade-off. Linear receivers such as the zero-forcing (ZF) and minimum mean squared error (MMSE) detectors are valued for their low complexity, but their performance is far from optimal [5, Chapter 6]. At the other end, the sphere decoder and its variants attain optimal performance but have high complexity [6]. To bridge the gap between these extremes, a number of techniques have been proposed. In this paper, we propose a technique that yields performance and complexity between linear receivers and maximum likelihood decoding. We use a committee of linear receivers, which decodes the transmitted symbol based on the decisions of several linear receivers and their relative confidence.

Each linear receiver in the committee has its own matrix to be applied to the received data as well as a weight (signifying the confidence in the receiver), which is used while forming the committee decision. To learn these parameters, we propose a variant of a statistical boosting technique called AdaBoost. AdaBoost ([10], [11], [12]) is a machine learning algorithm, wherein a set of *weak* classifiers (whose classification errors are large but better than random guessing) are combined into a *strong* classifier. This is referred to as boosting and AdaBoost is popular in several applications. These include text categorization [13], text filtering [14], 'ranking' problems [15] and natural language processing [16]. In signal processing, it has also be used in the context of minimizing multimodal non-smooth functions [17].

The parameter learning algorithm needs training data, which is *generated* at the receiver using an estimate of the channel matrix. Thus the *pilot or training* overhead actually sent over the channel remains the same as a classical system using a MMSE receiver. Specifically, as is common, for a $N_t \times N_r$ system, we send N_t pilot vectors of size $N_t \times 1$. At the receiver, the channel matrix is estimated using the pilot vectors. Then using the estimated channel matrix and knowledge of the SNR, at the receiver we simulate T channel input-output pairs. These T channel input-output pairs serve as the *training data* for our parameter learning algorithm. The channel estimation and the parameter learning based on the simulated training data needs to be carried out for every coherence interval.

It is fair to say that in recent years, the most successful classifiers in applications are based on a committee of simple classifiers and frequently use AdaBoost or some related algorithm (see [10]). On the other hand, linear receivers are popular in wireless communications due to their low computational complexity. It is natural to seek to improve over

978-1-4673-6187-3/13/$31.00 ©2013 IEEE

linear receivers by considering a committee of linear receivers, and given its success story, AdaBoost is a natural candidate to learn the parameters of such a committee. We note that in recent years, ideas from machine learning have been used in communications [18], [19], [20], [21], [22]. In particular, [22] proposes joint channel estimation and multiuser detection based on a "total least squares" approach. It uses boosting to search for the best solution, an idea similar to that proposed in [17] for optimization. This usage of boosting is different from our proposed receiver structure, which uses a combination of a committee of linear receivers and AdaBoost. To the best of our knowledge, this combination is new and has not been reported before. We believe that a committee of linear receivers is a basic and natural receiver structure that merits systematic study, and this paper makes a first effort in this direction.

We call our receiver boosted MMSE (B-MMSE), since at each stage of the algorithm we use the weighted empirical MMSE over the training data to find good linear receivers to incorporate in the committee. BER analysis of the B-MMSE receiver is hard due to its non-linear and iterative nature. In this first paper, our goal is to motivate the receiver, explain the learning algorithm in detail, and evaluate receiver performance (uncoded as well as coded BER) with simulations. We find that the B-MMSE receiver significantly outperforms the conventional MMSE receiver with an average complexity than is less than that of two linear receivers. In particular, our simulation results show a gain of about 2 dB at an uncoded BER of 10^{-2} for a 4×4 V-BLAST MIMO system with BPSK and $T = 64$ training vectors. Moreover, this gain in uncoded BER carries over and is amplified in the presence of an outer code. We find that the coded BER performance with an rate-3/4 turbo code has a gain of 5 dB at 10^{-4} coded BER when the demodulated output from the proposed method is used in place of plain MMSE detection with MMSE estimated channel, which is significant.

The rest of the paper is organized as follows. The system model is presented in Sec. II. The proposed AdaBoost algorithm applied to MIMO detection is presented in Sec. III. BER performance results are presented in Sec. IV. The conclusions are given in Sec. V.

II. System Model

Consider a V-BLAST MIMO system with N_t transmit and N_r receive antennas. Let $\boldsymbol{x} \in \mathsf{Q}^{N_t}$ be the transmitted vector where $\mathsf{Q} \subset \mathbb{C}$ is the modulation alphabet of size Q. Let $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ denote the channel gain matrix, whose entries are i.i.d. and $\mathbf{H} \sim \mathbb{C}\mathcal{N}(0, 1)$. The received vector $\boldsymbol{y}$ is given by

$$\boldsymbol{y} \quad = \quad \mathbf{H}\boldsymbol{x} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{n}$ is the noise vector whose entries are modeled as i.i.d. $\mathbb{C}\mathcal{N}(0, \sigma^2)$.

The maximum-likelihood receiver is given by:

$$\widehat{\boldsymbol{x}}_{ML} \quad = \quad \arg \min_{\boldsymbol{x} \in \mathsf{Q}^{N_t}} \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2. \tag{2}$$

The computational complexity in (2) is exponential in N_t since the set Q^{N_t} has size Q^{N_t}. Thus even for moderately large values of Q and N_t, the receiver is computationally intractable. To reduce complexity, it is common to consider a linear receiver: the estimate of $\boldsymbol{x}$ is given by $\mathcal{Q}(\mathbf{G}\boldsymbol{y})$, where $\mathbf{G} \in \mathbb{C}^{N_t \times N_r}$ is a suitably chosen matrix and $\mathcal{Q}(\boldsymbol{z})$ is the nearest neighbor quantizer for the input constellation Q applied to each component of the vector $\boldsymbol{z}$. The complexity of this receiver is dominated by the matrix multiplication $\mathbf{G}\boldsymbol{y}$ and is $O(N_t N_r)$. For example, for the MMSE receiver,

$$\mathbf{G}_{MMSE} = \left(\mathbf{H}^H \mathbf{H} + \frac{\sigma^2}{E_s} \mathbf{I} \right)^{-1} \mathbf{H}^H, \tag{3}$$

where we have assumed that $E[\boldsymbol{x}\boldsymbol{x}^H] = E_s \mathbf{I}$.

The MMSE receiver is widely used owing to its low complexity and improved performance compared to the ZF receiver. However, the performance of the MMSE receiver is far away from that of ML, even in the case of perfect channel gain information. In this paper, we aim to improve on MMSE receiver by using a *committee* of linear receivers instead of just a single MMSE receiver. The structure of the receiver we propose is described in the next section.

III. Boosting MMSE receiver

In this section, we describe a collection of linear receivers, whose decisions are combined to form a final decision. We call this a committee of linear receivers. Below, we describe the structure of the receivers and propose an algorithm to learn the parameters based on AdaBoost.

Committee of Linear Receivers: Let $\mathbf{G}_1, \cdots, \mathbf{G}_R$ be $N_t \times N_r$ complex matrices and let $\alpha_1, \cdots, \alpha_R$ be non-negative weights. Given observation $\boldsymbol{y}$, let $\widehat{\boldsymbol{x}}^{(r)} = \mathcal{Q}(\mathbf{G}_r \boldsymbol{y})$. Then our estimate of x_n (the nth component of $\boldsymbol{x}$) is formed as follows:

$$\widehat{x}_n = \arg \max_{q \in \mathsf{Q}} \sum_{r=1}^{R} \alpha_r 1(\widehat{x}_n^{(r)} = q), \tag{4}$$

where $1(\cdot)$ is the indicator function which is 1 if the argument is true, and is 0 otherwise. In words, this receiver assigns confidence weight α_r to receiver $\mathbf{G}_r$, and picks the input symbol which results in highest net confidence. Thus we are using a committee of linear receivers. The case $R = 1$ corresponds to usual linear receivers. We see that the complexity of the receiver is dominated by the matrix multiplication in the R receivers, that is, it is of order $O(RN_t N_r)$, which for a fixed R is much smaller than that of the ML receiver. To reduce complexity, we would like to choose R as small as possible.

To implement the committee of linear receivers, we wish to find good choices of R, $\mathbf{G}_1, \cdots, \mathbf{G}_R$ and $\alpha_1, \cdots, \alpha_R$ to get much better performance than MMSE. Towards this end, we use ideas from statistical learning. We describe our approach below starting with the description of how the training data is generated.

Generating Training Data: Let $\{\boldsymbol{x}_t\}_{t=1}^T$ be pre-generated with an i.i.d. uniform distribution from the input modulation constellation and stored in memory. Similarly we pre-generate i.i.d. $N(0, \sigma^2)$ noise samples $\{n_t\}_{t=1}^T$ and store them in memory. Then we generate the channel output samples using

$$\boldsymbol{y}_t = \mathbf{H}\boldsymbol{x}_t + \boldsymbol{n}_t.$$

Since $\{\boldsymbol{x}_t, \boldsymbol{n}_t\}_{t=1}^T$ are stored in the memory, only $\boldsymbol{y}_t$ needs to be generated for each coherence period. Hence the generation of training data has complexity $O(TN_tN_r)$. The data $\{\boldsymbol{x}_t, \boldsymbol{y}_t\}_{t=1}^T$ is used for learning the receiver parameters.

Finding R, $\mathbf{G}_1, \cdots, \mathbf{G}_R$ and $\alpha_1, \cdots, \alpha_R$: We adapt the Stagewise Additive Modeling using Multiclass Exponential Loss (SAMME) algorithm proposed by [12]. The SAMME algorithm itself is a multiclass generalization of the celebrated supervised classification algorithm AdaBoost [10, Chapter 10]. At a high level, the algorithm proceeds as follows. We pick $\mathbf{G}_1$ to minimize the empirical mean square error (MSE) on the training data. Then we look at all input symbols in the training data where the receiver $\mathcal{Q}(\mathbf{G}_1\boldsymbol{y})$ makes an error. We next assign weights to the training data so that the erroneous symbols have higher weight and then we find $\mathbf{G}_2$ to minimize the weighted empirical MSE over the training data. The process continues in this stagewise manner to find $\mathbf{G}_r, \alpha_r$ one-by-one. To keep R (and thus the complexity) low, we adapt it to the channel condition.

Let $\mathbf{X} = [\boldsymbol{x}_1, ..., \boldsymbol{x}_T]$ and let $x_{n,t}$ denote the (n,t) entry of the matrix. The algorithm we propose is the following.

1) We initialize the training set weights with

$$w_{n,t} = \frac{1}{N_tT} \quad \text{for all} \quad 1 \le t \le T, \quad 1 \le n \le N_t.$$

2) For rounds $r = 1, \cdots, R$:
   a) Find $\mathbf{G}_r$ by minimizing the weighted MSE:

   $$\sum_{t=1}^{T}\sum_{n=1}^{N_t} w_{n,t}|x_{n,t} - \tilde{x}_{n,t}^{(r)}|^2$$

   where

   $$\tilde{\boldsymbol{x}}_t^{(r)} = \mathbf{G}_r\boldsymbol{y}_t.$$

   Assuming $T \ge N_r$ and $\sum_{t=1}^T w_{t,n}\boldsymbol{y}_t\boldsymbol{y}_t^H$ to be non-singular, the nth row of $\mathbf{G}_r$ is given by

   $$\sum_{t=1}^{T} w_{t,n}x_{n,t}\boldsymbol{y}_t^H \left[\sum_{t=1}^{T} w_{t,n}\boldsymbol{y}_t\boldsymbol{y}_t^H\right]^{-1}.$$

   b) Let $\widehat{x}_{n,t}^{(r)} = \mathcal{Q}(\tilde{x}_{n,t}^{(r)})$ and compute the error rate:

   $$\text{ERR}_r = \sum_{n=1}^{N_t}\sum_{t=1}^{T} w_{n,t} 1\left(x_{n,t} \neq \widehat{x}_{n,t}^{(r)}\right),$$

   where $1(\cdot)$ is an indicator function, which is 1 if the argument is true and is 0 otherwise.

   c) Set

   $$\alpha_r = \ln\left(\frac{1 - \text{ERR}_r}{\text{ERR}_r}\right) + \ln(Q - 1).$$

   d) Set

   $$w_{n,t} = w_{n,t}\exp\left(\alpha_r 1\left(x_{n,t} \neq \widehat{x}_{n,t}^{(r)}\right)\right)$$

   and normalize the $\{w_{n,t}\}$ so they sum to 1.

3) Output $\mathbf{G}_1, \cdots, \mathbf{G}_R$ and $\alpha_1, \cdots, \alpha_R$.

Adaptation to channel condition: If the channel condition is good and $\text{ERR}_r = 0$ for some r, then continuing the iterations will not change the weights, so the iteration can be stopped at this point. Further if the channel is bad and $\text{ERR}_r > (1 - \frac{1}{Q})$, the filter is doing worse than random guessing, so the iteration can be stopped and the current round ignored (except when $r = 1$).

In Algorithm 1 below, we give the pseudo-code of the parameter learning algorithm.

Discussion of the algorithm: Step 2a) of the above algorithm is computationally most expensive and determines the complexity of the algorithm. The complexity of forming the matrix $\mathbf{G}_r$ is $O(TN_tN_r) + O(N_r^3)$, and since we have R rounds, the complexity is of order $O(RN_r(TN_t + N_r^2))$, which is substantially smaller than the exponential complexity of the ML receiver. The training complexity is incurred only once per coherence time of the channel. Thus in applications such as WLANs or low-mobility cellular users, the training complexity is not large. Also we let R depend on the channel realization. As a consequence, R is usually small, and is large only occasionally. Consequently, the average complexity of the receiver is low.

The specific algorithm we have used to compute $\mathbf{G}_1, \cdots, \mathbf{G}_R$ and $\alpha_1, \cdots, \alpha_R$ is an adaptation of the SAMME algorithm [12], which itself is motivated by AdaBoost (see for example [10, Chapter 10]). We note that in AdaBoost, usually the classifier in each round is chosen to minimize a bound on the weighted classification error. However, this optimization problem does not have an analytical solution in our case. Instead of resorting to numerical optimization, we have taken the approach of using weighted MSE as the error metric in our algorithm, which leads to a closed-form solution in Step 2a) of the algorithm. Thus our learning algorithm differs from AdaBoost/SAMME algorithm in Step 2a) described above.

Since we start with equal weights initially, and we use the weighted empirical MSE for choosing the linear receivers in the committee, we call our receiver Boosted MMSE (B-MMSE). It is natural to ask if this committee always has better performance than say the MMSE receiver. However, BER analysis of the B-MMSE receiver is analytically intractable due to the nonlinear nature of the receiver. In the next section, we use simulations to study its performance.

IV. BER Performance of the B-MMSE Receiver

In this section we present the simulated BER performance of the proposed B-MMSE receiver and compare it with the MMSE receiver. Throughout we consider a 4×4 V-BLAST MIMO system. While the description of the receiver and the parameter learning algorithm so far is applicable to any modulation, in this section we focus on BPSK.

Training Phase:

$$\mathbf{Y}_T \leftarrow [\boldsymbol{y}_1 \boldsymbol{y}_2 \cdots \boldsymbol{y}_t \cdots \boldsymbol{y}_T]$$
$$\mathbf{X} \leftarrow [\boldsymbol{x}_1 \boldsymbol{x}_2 \cdots \boldsymbol{x}_t \cdots \boldsymbol{x}_T]$$

Initializations:

$$stop\_flag \leftarrow 0 \qquad r \leftarrow 1$$
$$\boldsymbol{\alpha} \leftarrow \phi \qquad \mathcal{G} \leftarrow \phi$$

while $stop\_flag = 0$ **do**

Find $\mathbf{G}_r$:
The nth row of $\mathbf{G}_r$ is given by

$$\sum_{t=1}^{T} w_{t,n} x_{n,t} \boldsymbol{y}_t^H \left[\sum_{t=1}^{T} w_{t,n} \boldsymbol{y}_t \boldsymbol{y}_t^H \right]^{-1}.$$

where $x_{n,t}$ is the (n,t)th entry of $\mathbf{X}$

$$\widehat{\mathbf{X}}_T^{(r)} \leftarrow \mathcal{Q}(\mathbf{G}_r \mathbf{Y}_T)$$

$$\text{ERR}_r = \sum_{n=1}^{N_t} \sum_{t=1}^{T} w_{n,t} 1 \left(x_{n,t} \neq \widehat{x}_{n,t}^{(r)} \right)$$

if $ERR_r > (1 - \frac{1}{Q})$ **then**

if $r = 1$ **then**

$$\boldsymbol{\alpha} \leftarrow \{1\}, \mathcal{G} \leftarrow \{\mathbf{G}_1\}$$

end

$stop\_flag \leftarrow 1$

else if $ERR_r = 0$ **then**

$$\alpha_r = \ln \left(\frac{1 - \epsilon}{\epsilon} \right) + \ln(Q - 1)$$

where $\epsilon > 0$ has a very small value.

$$\boldsymbol{\alpha} \leftarrow \boldsymbol{\alpha} \bigcup \{\alpha_r\}, \mathcal{G} \leftarrow \mathcal{G} \bigcup \{\mathbf{G}_r\}$$
$$stop\_flag \leftarrow 1$$

else

$$\alpha_r \leftarrow \ln \left(\frac{1 - \text{ERR}_r}{\text{ERR}_r} \right) + \ln(Q - 1)$$

$$w_{n,t} \leftarrow w_{n,t} \exp \left(\alpha_r 1 \left(x_{n,t} \neq \widehat{x}_{n,t}^{(r)} \right) \right)$$

$$w_{n,t} \leftarrow \frac{w_{n,t}}{\sum_{n=1}^{N_t} \sum_{t=1}^{T} w_{n,t}}$$

$$\boldsymbol{\alpha} \leftarrow \boldsymbol{\alpha} \bigcup \{\alpha_r\}, \; \mathcal{G} \leftarrow \mathcal{G} \bigcup \{\mathbf{G}_r\}$$

if $r = R$ **then** $stop\_flag \leftarrow 1$ **else** $r \leftarrow r + 1$

end

end

Detection Phase:

for $r \leftarrow 1$ ***to*** $|\mathcal{G}|$ **do**

$$\widehat{\boldsymbol{x}}^{(r)} \leftarrow \mathcal{Q}(\mathbf{G}_r \boldsymbol{y})$$

end

$$\widehat{x}_n \leftarrow \arg\max_{q \in \mathsf{Q}} \sum_{r=1}^{R} \alpha_r 1(\widehat{x}_n^{(r)} = q)$$

Algorithm 1: Boosting MMSE Receiver

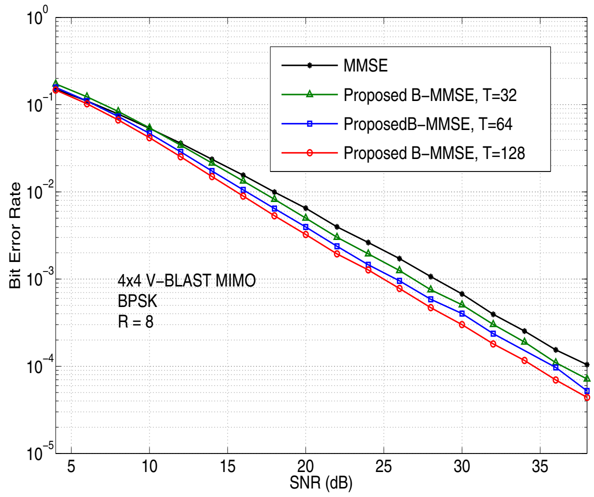

Fig. 1. BER performance of the B-MMSE receiver in 4×4 V-BLAST MIMO for different values of T. MMSE channel estimate is used.

In the simulations, we use the MMSE channel estimate $\widehat{\mathbf{H}}$ which can be obtained by sending an $N_t \times N_t$ *orthonormal* pilot matrix $\mathbf{X}_{pilot}$ over the channel with power P and observing the corresponding output

$$\mathbf{Y}_{pilot} = \sqrt{P}\mathbf{H}\mathbf{X}_{pilot} + \mathbf{W}$$

where the columns of $\mathbf{W}$ are i.i.d. noise vectors. The corresponding MMSE channel estimate is

$$\widehat{\mathbf{H}} = \frac{P}{P + \sigma^2} \mathbf{Y}_{pilot} \mathbf{X}_{pilot}^H. \tag{5}$$

In Figure 1, we compare the B-MMSE receiver for different values of training length T with the MMSE receiver. We can see that the proposed method, with 64 training vectors, has a gain of about 2 dB when compared to MMSE detector at a BER of 10^{-2}. The performance improves with increasing number of training vectors, but this increases the complexity of training.

In Figure 2, we show the histogram of the number of rounds R over 50,000 channel realizations. We see that with probability 0.94 we need only $R \leq 2$ and the mean value of R is 1.25. We have restricted the maximum value to be $R_{max} = 8$, and this maximum value is used with probability 0.024. Since the complexity of B-MMSE is equal to R times that of MMSE, we see that the mean complexity is only 1.25 times the MMSE receiver.

It is of interest to see how much gain is achieved by the B-MMSE receiver in a coded system compared to plain MMSE receiver. To address this question, in Fig. 3, we compare the turbo coded performance of B-MMSE with that of plain MMSE detector with MMSE estimate of the channel. The rate of the code used is $3/4$. It is seen that a *significant gain* of 5 dB is achieved by the B-MMSE receiver at a coded BER of 10^{-4}.

V. Conclusions

We have proposed the B-MMSE receiver: it is a committee of linear receivers whose parameters are learnt using a variant

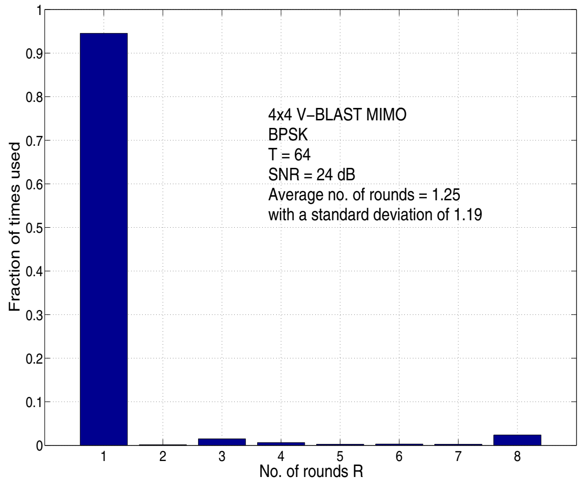

Fig. 2. Histogram of the number of rounds R needed over independent channel realizations. The average value of R is 1.25.

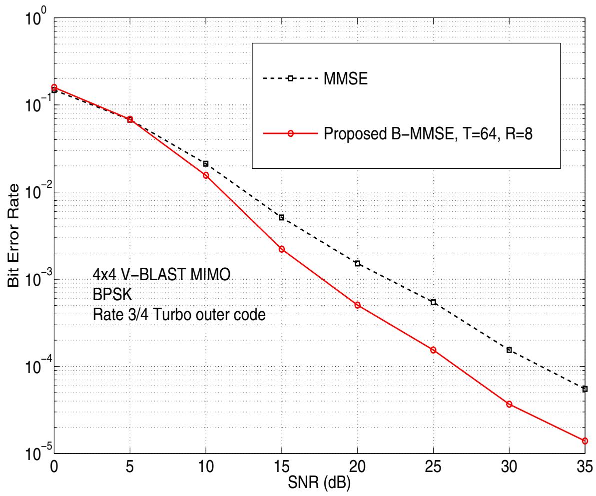

Fig. 3. Turbo coded performance of the B-MMSE receiver.

of AdaBoost, a popular boosting algorithm from machine learning. For BPSK modulation, the coded and uncoded BER of B-MMSE is superior to that of the MMSE receiver with only a small increase in average complexity. These results show that a popular theme in machine learning - a committee of simple weak classifiers boosted to form a strong classifier - is useful in the communications context and should be explored more. In particular, we believe a committee of linear receivers is a basic structure, and given the encouraging results in this paper, it needs to be studied further. From a practical standpoint, immediate directions for future work include performance evaluation for 4-QAM, 16-QAM, 64-QAM and techniques for improved simulation of training data. Any analytical insights into the performance would also be valuable to understand the performance gains/limitations of such committee based receivers.

References

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311-335, March 1998.

[2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.,* vol. 10, no. 6, pp. 585-595, November 1999.

[3] G. Hiertz, D. Denteneer, L. Stibor, Y. Zang, X. P. Costa, and B. Walke, "The IEEE 802.11 universe," *IEEE Commun. Mag.,* vol. 48, no. 1, pp. 62-70, January 2010.

[4] J. Lee, J.-K. Han, and J. Zhang, "MIMO technologies in 3GPP LTE and LTE-advanced," *EURASIP J. Wirel. Commun. Netw.*, March 2009.

[5] A. Sibille, C. Oestges, and A. Zanella, *MIMO: From Theory to Implementation*, Academic Press, 2010.

[6] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inform. Theory,* vol. 45, no. 5, pp. 1639-1642, July 1999.

[7] G. Taricco and E. Biglieri, "Space-time decoding with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1874-1888, July 2005.

[8] Y. Chen and N. C. Beaulieu, "Optimum pilot symbol assisted modulation," *IEEE Trans. Commun.*, vol. 55, no. 8, pp. 1536-1546, August 2007.

[9] J. Zhang, *Optimal Detection with Imperfect Channel Estimation for Wireless Communications*, Ph.D Thesis, University of New York, September 2009.

[10] Hastie et al, *The Elements of Statistical Learning*, Second Edition, Springer, 2009.

[11] Y. Freund and R. Schapire, "A decision theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, pp. 119-139, 1997.

[12] J. Zhu, H. Zou, S. Rosset, T. Hastie, "Multi-class AdaBoost," *Statistics And Its Interface,* vol. 2, pp. 349-360, 2009.

[13] R. E. Schapire and Y. Singer, "BoosTexter: A boosting-based system for text categorization," *Machine Learning,* 39(2/3):135-168, 2000.

[14] R. E. Schapire, Y. Singer, and A. Singhal, "Boosting and Rocchio applied to text filtering," *SIGIR 98: Proceedings of the 21st Annual International Conference on Research and Development in Information Retrieval*, 1998.

[15] Y. Freund, R. Iyer, R. E. Schapire, and Y. Singer, "An efficient boosting algorithm for combining preferences," *Machine Learning: Proceedings of the Fifteenth International Conference*, 1998.

[16] M. Haruno, S. Shirai, and Y. Ooyama, "Using decision trees to construct a practical parser," *Machine Learning*, 34:131149, 1999.

[17] S. Chen, X. Wang, and C. J. Harris, "Experiments with repeating weighted boosting search for optimization signal processing applications," *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 35, no. 4, pp. 682-693, August 2005

[18] B. J. Frey, *Graphical Models for Machine Learning and Digital Communication*, Cambridge, MA: MIT Press, 1998.

[19] P. Som, T. Datta, N. Srinidhi, A. Chockalingam, and B. S. Rajan, "Low-complexity detection in large-dimension MIMO-ISI channels using graphical models," *IEEE Jl. Sel. Topics in Sig. Process.*, vol. 5, no. 8, pp. 1497-1511, December 2011.

[20] M. Hansen, B. Hassibi, A. G. Dimakis, and W. Xu, "Near-optimal detection in MIMO systems using Gibbs sampling," *Proc. IEEE ICC'2009*, Hawaii, December 2009.

[21] N. Srinidhi, T. Datta, A. Chockalingam, and B. S. Rajan, "Layered tabu search algorithm for large-MIMO detection and a lower bound on ML performance," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 2955-2963, November 2011.

[22] J. Zhang, S. Chen, X. Mu, and L. Hanzo, "Joint channel estimation and multiuser detection for SDMA/OFDM based on dual repeated weighted boosting search," *IEEE Trans. Veh. Tech.,* vol. 60, no. 7, pp. 3265-3275, September 2011.

[23] P. L. Bartlett and M. Traskin, "AdaBoost is consistent," *J. Mach. Learn. Res.*, 8 (December 2007), 2347-2368.