PEP-BASED OPTIMAL TRAINING FOR MIMO SYSTEMS IN WIRELESS CHANNELS

Khawza I. Ahmed, Cihan Tepedelenlioglu and Andreas Spanias

Department of Electrical Engineering
Arizona State University, Tempe, AZ 85287-5706
{khawza.ahmed, cihan, spanias}@asu.edu

ABSTRACT

In this paper, PEP (pairwise error probability) is derived for MIMO systems in the presence of channel estimation error, where the channel is estimated using training insertion and linear channel estimation. This training-based PEP expression is utilized to design the training sequence and for optimal power allocation between the training and data symbols. The loss of coding gain due to the training is quantified at high SNR. It is shown that both the proposed PEP-based and the existing capacity-based optimal training approach of [1] reveal the same power allocation for constant modulus (CM) constellations. The performance of our optimal power allocation scheme falls between the nonoptimized equal power allocation and perfectly known channel cases, and converges to the latter for larger data transmission intervals. Numerical and simulation results are presented to justify the usefulness of our analytical findings.

1. INTRODUCTION

With the increasing demand of high data rates, future wireless communication systems need to combat multipath fading in a bandwidth efficient manner. A widely used technique for this purpose is multiple antennas in the receiver side with some kind of combining of the received signal. However, with the recent development triggered by Telatar and Foschini [2, 3], placement of multiple antennas at both the transmitter and the receiver side has become a subject of intensive research due to the much higher capacity manifested by spatial multiplexing and diversity gains.

To achieve the capacity in a MIMO-setup, Tarokh *et al.* [4] introduced space-time trellis codes (STTC). In [5], the effect of channel estimation error (CEE) on the performance of space-time codes is investigated, where a generic channel estimate is modelled as the actual fading coefficients perturbed by additive Gaussian noise and an optimal detection rule that maximizes the *a posteriori* probability in the presence of CEE is obtained. A PEP expression as a function of the correlation coefficient μ in the presence of channel estimation error is obtained only for constant modulus constellations, e.g., PSK. The key observation is that the code design criteria, namely, the rank and determinant criteria remain unchanged. However, there exists a loss of coding gain attributed to the imperfect knowledge of channel and furthermore, an error floor results if μ does not approach to unity with the increase of SNR. In [6], it is shown that there is no loss of diversity gain even when the detection rule that is optimal with perfectly known channel coefficients is used just by replacing the channel coefficients with their estimates. The variance of the channel estimate is assumed to be proportional to the noise variance in [6] to make analysis tractable. Since no explicit training scheme is considered in [5] and [6], the issue of optimal training design is not raised and accordingly the derived PEP expressions are not explored for the design of training symbols or optimal power allocation between the training and data symbols. Hassibi *et al.* [1] resorts to the maximization of the lower bound of average training-based channel capacity to design the training scheme and determine the optimal allocation. Although the mutual information between the transmitted input and processed output as an optimality criterion provides important design guidelines, PEP (pairwise error probability) is a more practical measure of communication system performance. Thus, the objective and the main contribution of this paper is the proposed design of a training scheme for a MIMO system that minimizes the PEP. In particular, a novel PEP expression is obtained when the MMSE-based estimates are used in the decoding rule as if they are the true channel. Our training-based PEP expression provides the optimal choice of training symbols and more importantly optimal power allocation between the training and the data symbols. Furthermore, the loss of coding gain at high SNR is also quantified.

The notation used throughout is as follows. Bold upper and lower case letters denote matrices and vectors respectively. $(\cdot)^T$ $(\cdot)^{\mathcal{H}}$ and $(\cdot)^*$ denote transpose, Hermitian transpose and conjugate respectively, $\mathbf{I}_L$ denotes identity matrices of size L, $\mathcal{CN}(0, \sigma^2)$ is the complex-normal zero-mean where the real and imaginary components of each random entry are independent with variance $\sigma^2/2$, (q, r) element of matrix $\mathbf{A}$ is denoted by $[\mathbf{A}]_{q,r}$; $E[\cdot]$ and tr$(\cdot)$ respectively denote statistical average, trace of a matrix; $(\cdot)_o$ is used to distinguish the optimal parameters. vec$(\mathbf{A})$ denotes the vector formed by stacking the columns of $\mathbf{A}$ on top of each other.

2. SYSTEM MODEL

We consider a MIMO system equipped with N_t transmit and N_r receive antennas in a flat fading environment. The fading coefficients from transmit antenna k to receive antenna i is denoted as $h^{(ki)}$. We assume antennas are separated far enough to ensure independent fading at each of the receiver antenna and therefore, fading coefficients can be modelled as independent complex Gaussian random variables of variance σ_h^2, i.e., $E\{|h^{(ki)}|^2\} = \sigma_h^2$. We also assume that the coefficients remain constant for time period T, and change to a new realization in the next time period. The symbol transmitted from antenna k at time l is denoted as $y^{(lk)}/\sqrt{N_t}$ where $y^{(lk)}$ chosen from a finite constellation points is normalized so that its average energy is unity and thus $1/N_t$ denotes the average energy of each of the symbols transmitted from each antenna. The received value $x^{(li)}$ at receive antenna i and at time l is given by

The work in this paper was supported in part by NSF Career Grant No. CCR-0133841 and by Arizona State University.

$$x^{(li)} = \frac{1}{\sqrt{N_t}} \sum_{k=1}^{N_t} y^{(lk)} h^{(ki)} + w^{(li)}, \quad (1)$$

where the additive noise $w^{(li)}$ at receive antenna i and time l is i.i.d complex Gaussian noise with mean zero and variance σ_w^2. Equation (1) can be written compactly in a matrix form as

$$\mathbf{X} = \frac{1}{\sqrt{N_t}} \mathbf{YH} + \mathbf{W}, \quad (2)$$

where $[\mathbf{Y} \in \mathcal{C}^{T \times N_t}]_{l,i} = y^{(li)}$, $[\mathbf{X} \in \mathcal{C}^{T \times N_r}]_{l,k} = x^{(lk)}$ and $[\mathbf{H} \in \mathcal{C}^{N_t \times N_r}]_{k,i} = h^{(ki)}$ for $k = 1, \cdots, N_t$, $i = 1, \cdots, N_r$ and $l = 1, \cdots, T$. Note that SNR at each of received antennas is $\eta = \sigma_h^2 / \sigma_w^2$.

3. PROBLEM FORMULATION

To decode the data symbols in a frame, we assume that the receiver first estimates the channel and then uses this estimated channel as if it were the true channel. To facilitate the channel estimation for a particular frame, the training matrix $\mathbf{Y}_\tau$ known to the receiver is appended before the data matrix. Consequently, the received matrix $\mathbf{X}_\tau$ in the training phase is given by

$$\mathbf{X}_\tau = \frac{\sigma_\tau}{\sqrt{N_t}} \mathbf{Y}_\tau \mathbf{H} + \mathbf{W}_\tau, \quad (3)$$

where σ_τ^2 is the total transmitted energy in each time slot of the training phase from all the N_t transmit antennas with $\text{tr}\left(\mathbf{Y}_\tau \mathbf{Y}_\tau^{\mathcal{H}}\right) = N_t T_\tau$. In the data transmission phase a data matrix $\mathbf{Y}_\delta$ is transmitted over the rest $T_\delta := T - T_\tau$ time samples and received matrix $\mathbf{X}_\delta$ is given by

$$\mathbf{X}_\delta = \frac{\sigma_\delta}{\sqrt{N_t}} \mathbf{Y}_\delta \mathbf{H} + \mathbf{W}_\delta, \quad (4)$$

where σ_δ^2 is the total transmitted energy in each time slot of the data transmission phase with $\text{tr}\left(\text{E}\left[\mathbf{Y}_\delta \mathbf{Y}_\delta^{\mathcal{H}}\right]\right) = N_t T_\delta$. The estimate $\hat{\mathbf{H}}$ is obtained from (3) and used in (4) to decode $\mathbf{Y}_\delta$. If we assume that on an average, considering both the training and the data transmission phases, the total transmitted energy at each time slot is unity, the energy conservation gives us the following constraint:

$$\sigma_\tau^2 T_\tau + \sigma_\delta^2 T_\delta = T. \quad (5)$$

A power allocation ratio $\gamma := \sigma_\tau^2 T_\tau / T \in (0, 1)$ indicates the percentage of the total power employed for training. Clearly, the performance of the receiver will depend on the quality of the channel estimate $\hat{\mathbf{H}}$ and the power allocation ratio γ. With $\tilde{\mathbf{H}} = \mathbf{H} - \hat{\mathbf{H}}$ the channel estimator MSE denoted as $\sigma_{\tilde{h}}^2 := 1/(N_r N_t) \times E[\text{vec}(\tilde{\mathbf{H}})\text{vec}(\tilde{\mathbf{H}})^{\mathcal{H}}]$ is a measure of the quality of the channel estimate and will depend on the training symbols in $\{\mathbf{Y}_\tau\}$ and the training interval T_τ. Furthermore, due to the power budget constraint in (5), if we allocate too much power for training (i.e., $\gamma \approx 1$), the channel estimates will be better but the detectability of the data will be susceptible to noise due to weak data SNR. On the other hand, too little power for training (i.e., $\gamma \approx 0$) deteriorates the estimated channel quality, and results in poor detection despite the high data SNR. Therefore, our objective is to derive the PEP expressions in the presence of training and minimize the PEP to obtain optimal training symbols and optimal power allocation. Moreover, we will analyze the performance of training with equal power allocation, i.e., $\sigma_\tau^2 = \sigma_\delta^2 = 1$, where $\gamma = T_\tau / T$ and also training with optimal power allocation, i.e., $\gamma = \gamma_o$ and compare these with the performance of the perfect CSI case.

4. CHANNEL ESTIMATION AND DECODING RULES

LMMSE estimate of the channel $\mathbf{H}$ from (3) is given by

$$\hat{\mathbf{H}} = \frac{\sqrt{N_t}}{\sigma_\tau} \left(\frac{\sigma_w^2 N_t}{\sigma_h^2 \sigma_\tau^2} \mathbf{I}_{N_t} + \mathbf{Y}_\tau^{\mathcal{H}} \mathbf{Y}_\tau \right)^{-1} \mathbf{Y}_\tau^{\mathcal{H}} \mathbf{X}_\tau, \quad (6)$$

and the estimation error variance is

$$\begin{aligned} \sigma_{\tilde{h}}^2 &:= \frac{1}{N_r N_t} E[\text{vec}(\tilde{\mathbf{H}})\text{vec}(\tilde{\mathbf{H}})^{\mathcal{H}}] \\ &= \frac{\sigma_h^2}{N_r N_t} \text{tr}\left(\left[\mathbf{I}_{N_t} + \frac{\sigma_\tau^2 \sigma_h^2}{N_t \sigma_w^2} \mathbf{Y}_\tau^{\mathcal{H}} \mathbf{Y}_\tau \right]^{-1} \bigotimes \mathbf{I}_{N_r} \right). \quad (7) \end{aligned}$$

To express the decoding rules, let us rewrite (4) as

$$\mathbf{X}_\delta = \frac{\sigma_\delta}{\sqrt{N_t}} \mathbf{Y}_\delta \hat{\mathbf{H}} + \underbrace{\frac{\sigma_\delta}{\sqrt{N_t}} \mathbf{Y}_\delta \tilde{\mathbf{H}} + \mathbf{W}_\delta}_{\mathbf{V}_\delta}, \quad (8)$$

where we have used $\tilde{\mathbf{H}} := \mathbf{H} - \hat{\mathbf{H}}$ and the term associated with the estimation error due to noise at the training phase, and the noise at the data phase are lumped together and denoted as the signal dependent noise $\mathbf{V}_\delta$. Due to the orthogonality property of LMMSE estimator, $\hat{\mathbf{H}}$ and $\tilde{\mathbf{H}}$ are uncorrelated and it can be shown that for a given $\mathbf{Y}_\delta$, $\mathbf{V}_\delta$ is Gaussian and independent to $\mathbf{Y}_\delta \hat{\mathbf{H}}$. $[\mathbf{X}_\delta]_{l,i}$ can be written as

$$x_\delta^{(li)} = \frac{\sigma_\delta}{\sqrt{N_t}} \sum_{k=1}^{N_t} y_\delta^{(lk)} \hat{h}^{(ki)} + \underbrace{\frac{\sigma_\delta}{\sqrt{N_t}} \sum_{k=1}^{N_t} y_\delta^{(lk)} \tilde{h}^{(ki)} + w_\delta^{(li)}}_{v_\delta^{(li)}},$$

where $\sigma_v^{2(l)} := E[|v_\delta^{(li)}|^2] = (\sigma_\delta^2 / N_t) \sum_{k=1}^{N_t} |y_\delta^{(lk)}|^2 \sigma_{\tilde{h}}^2 + \sigma_w^2$, $\hat{h}^{(ki)} := [\hat{\mathbf{H}}]_{k,i}$ and $\tilde{h}^{(ki)} := [\tilde{\mathbf{H}}]_{k,i}$ for $k = 1, \cdots, N_t$, $i = 1, \cdots, N_r$ and $l = 1, \cdots, T_\delta$. In decoding $y_\delta^{(lk)}$, we have at least two options. One option is to minimize

$$\sum_{l=1}^{T_\delta} \sum_{i=1}^{N_r} \left| x_\delta^{(li)} - \frac{\sigma_\delta}{\sqrt{N_t}} \sum_{k=1}^{N_t} y_\delta^{(lk)} \hat{h}^{(ki)} \right|^2, \quad (9)$$

with respect to $y_\delta^{(lk)}$. Note that (9) is the ML receiver when $\hat{h}^{(ki)} = h^{(ki)}$. However, in the presence of CEE, (9) is not optimal and the ML estimator of $y_\delta^{(lk)}$ minimizes

$$\sum_{l=1}^{T_\delta} \sum_{i=1}^{N_r} \left[\frac{\left| x_\delta^{(li)} - \frac{\sigma_\delta}{\sqrt{N_t}} \sum_{k=1}^{N_t} y_\delta^{(lk)} \hat{h}^{(ki)} \right|^2}{\sigma_v^{2(l)}} + \ln \sigma_v^{2(l)} \right]. \quad (10)$$

Notice that decision rule in (10) will yield the same result as in (9), if $y_\delta^{(lk)}$ has constant modulus entries $\forall k, l$. Therefore, in general, receivers that assume the channel estimate after training to be perfect are suboptimal. However, these are employed in practice due to the simplicity of the decoding rule. Hence, we will focus on (9) throughout our analysis.

5. PEP WITH CHANNEL ESTIMATION ERROR

We will evaluate the PEP when the decision rule in (9) is used. The PEP is the probability of decoding the code matrix $\mathbf{Y}'_\delta$ erroneously while $\mathbf{Y}_\delta$ was transmitted and can be expressed as

$$\begin{aligned} &P(\mathbf{Y}_\delta \to \mathbf{Y}'_\delta | \hat{\mathbf{H}}) \\ &= Q\left(\frac{\sum_{l=1}^{T_\delta} \sum_{i=1}^{N_r} \left| \frac{\sigma_\delta}{\sqrt{N_t}} \sum_{k=1}^{N_t} (y_\delta^{(lk)} - {y'}_\delta^{(lk)}) \hat{h}^{(ki)} \right|^2}{\sqrt{\sum_{l=1}^{T_\delta} \sum_{i=1}^{N_r} 2\sigma_v^{2(l)} \left| \frac{\sigma_\delta}{\sqrt{N_t}} \sum_{k=1}^{N_t} (y_\delta^{(lk)} - {y'}_\delta^{(lk)}) \hat{h}^{(ki)} \right|^2}} \right) \\ &\leq Q\left(\sqrt{\frac{\sum_{l=1}^{T_\delta} \sum_{i=1}^{N_r} \frac{\sigma_\delta^2}{N_t} \left| \sum_{k=1}^{N_t} (y_\delta^{(lk)} - {y'}_\delta^{(lk)}) \hat{h}^{(ki)} \right|^2}{2\sigma_v^2}} \right), \quad (11) \end{aligned}$$

where the inequality in (11) is obtained by using

$$\sigma_v^2 := (\sigma_\delta^2/N_t)\sum_{k=1}^{N_t} \nu\sigma_{\tilde{h}}^2 + \sigma_w^2 = \sigma_\delta^2\nu\sigma_{\tilde{h}}^2 + \sigma_w^2 \geq \sigma_v^{2(l)},$$

where $\nu := \max_{l,k} |y_\delta^{(lk)}|^2 \geq 1$. Notice that $\nu = 1$ for constant modulus constellations, e.g., PSK and the inequality in (11) reduces to equality. The Chernoff bound on (11) gives us

$$\begin{aligned} &P(\mathbf{Y}_\delta \to \mathbf{Y}'_\delta|\hat{\mathbf{H}}) \\ &\leq \exp\left(-\frac{\frac{\sigma_\delta^2}{N_t}\sum_{l=1}^{T_\delta}\sum_{i=1}^{N_r}\left|\sum_{k=1}^{N_t}(y_\delta^{(lk)} - {y'}_\delta^{(lk)})\hat{h}^{(ki)}\right|^2}{4(\sigma_\delta^2\sigma_{\tilde{h}}^2\nu + \sigma_w^2)}\right) \quad (12) \\ &= \exp\left(-\frac{\sigma_\delta^2}{4N_t}\frac{\sum_{i=1}^{N_r}\hat{\mathbf{h}}_i^{\mathcal{H}}\mathbf{A}\hat{\mathbf{h}}_i}{(\sigma_\delta^2\sigma_{\tilde{h}}^2\nu + \sigma_w^2)}\right) = \prod_{i=1}^{N_r}\exp\left(-\frac{\sigma_\delta^2\hat{\mathbf{h}}_i^{\mathcal{H}}\mathbf{A}\hat{\mathbf{h}}_i}{4N_t(\sigma_\delta^2\sigma_{\tilde{h}}^2\nu + \sigma_w^2)}\right) \quad (13) \end{aligned}$$

where $\hat{\mathbf{h}}_i$ is the i-th column of $\hat{\mathbf{H}}$ and $\mathbf{A} := (\mathbf{Y}_\delta - \mathbf{Y}'_\delta)^{\mathcal{H}}(\mathbf{Y}_\delta - \mathbf{Y}'_\delta)$. With $N_A \leq N_t$ number of non-zero eigenvalues $\{\lambda_m\}_{m=1}^{N_A}$ of $\mathbf{A}$, the expectation of (13) over the channel realization gives us

$$\begin{aligned} P(\mathbf{Y}_\delta \to \mathbf{Y}'_\delta) &\leq \left(\prod_m^{N_A}\frac{1}{1 + \frac{\sigma_{\hat{h}}^2\sigma_\delta^2\lambda_m}{4N_t(\sigma_{\tilde{h}}^2\sigma_\delta^2\nu + \sigma_w^2)}}\right)^{N_r} \\ &\leq \left(\prod_m^{N_A}\lambda_m\right)^{-N_r}\left(\underbrace{\frac{\sigma_{\hat{h}}^2\sigma_\delta^2}{4N_t(\sigma_{\tilde{h}}^2\sigma_\delta^2\nu + \sigma_w^2)}}_{f(\cdot)}\right)^{-N_rN_A}. \quad (14) \end{aligned}$$

When $\sigma_{\tilde{h}}^2 = 0$, $\sigma_{\hat{h}}^2 = \sigma_h^2$ and thus with $\sigma_\delta^2 = 1$ and average SNR per receive antenna $\eta := \sigma_h^2/\sigma_w^2$, (14) reduces to the perfect CSI case given by

$$P(\mathbf{Y}_\delta \to \mathbf{Y}'_\delta) \leq \left(\prod_m^{N_A}\lambda_m\right)^{-N_r}\left(\frac{\eta}{4N_t}\right)^{-N_rN_A}. \quad (15)$$

Comparing (14) and (15), we observe that when channel MSE $\sigma_{\tilde{h}}^2$ is proportional to the observed noise variance σ_w^2, there is no loss of diversity, but a loss of coding gain occurs due to the presence of CEE. This conclusion is also drawn in [5] and [6], however, no method is proposed to utilize the PEP expression to enhance the performance. In the next sections we discuss how this loss of coding gain can be minimized by the judicious choice to training symbols and the power allocation between the training and the data symbols.

6. PEP-BASED OPTIMAL $\mathbf{Y}_\tau$

From (7) it is evident that the design of training matrix $\mathbf{Y}_\tau$ affects $\sigma_{\tilde{h}}^2$. Furthermore, since for LMMSE estimator, $\sigma_{\hat{h}}^2 = \sigma_h^2 - \sigma_{\tilde{h}}^2$, it is evident from (14) that the decrease of estimator error variance $\sigma_{\tilde{h}}^2$ increases $f(\cdot)$ and thereby decreases PEP. Therefore, optimal $\mathbf{Y}_\tau = \mathbf{Y}_{\tau o}$ that minimizes PEP can be obtained from

$$\underset{\mathbf{Y}_\tau, \mathrm{tr}\left(\mathbf{Y}_\tau^{\mathcal{H}}\mathbf{Y}_\tau\right)=N_tT_\tau}{\arg\ \min} \frac{1}{N_t}\mathrm{tr}\left(\left[\mathbf{I}_{N_t} + \frac{\sigma_\tau^2\sigma_h^2}{N_t\sigma_w^2}\mathbf{Y}_\tau^{\mathcal{H}}\mathbf{Y}_\tau\right]^{-1}\right) \quad (16)$$

which results to $\mathbf{Y}_{\tau o}^{\mathcal{H}}\mathbf{Y}_{\tau o} = T_\tau\mathbf{I}_{N_t}$, i.e., the columns of $\mathbf{Y}_{\tau o}$ are orthonormal with a multiplicative constant. The same conclusion is drawn in [1] when the optimality criterion is based on the lower bound of the average capacity. With the optimal choice of $\mathbf{Y}_\tau$, (7) becomes

$$\sigma_{\tilde{h}}^2 = \frac{N_t\sigma_w^2\sigma_h^2}{N_t\sigma_w^2 + \sigma_h^2\sigma_\tau^2T_\tau}, \quad (17)$$

and $\hat{\mathbf{H}}$ in (6) becomes

$$\hat{\mathbf{H}} = c\mathbf{H} + \mathbf{U}, \quad (18)$$

where $c := \sigma_h^2\sigma_\tau^2T_\tau/(\sigma_w^2N_t + \sigma_h^2\sigma_\tau^2T_\tau)$ and $[\mathbf{U}]_{i,k} \sim \mathcal{CN}(0, \sigma_u^2)$ with $\sigma_u^2 = N_t\sigma_w^2\sigma_h^4T_\tau\sigma_\tau^2/(\sigma_w^2N_t + \sigma_h^2\sigma_\tau^2T_\tau)^2$. Notice that since the total training power is dictated by $\sigma_\tau^2T_\tau$, from (17) it is evident by increasing the training power $\sigma_{\tilde{h}}^2$ can be minimized. But due to the total power constraint, it will decrease the data SNR and affect the performance. The optimal power allocation derived in the next section addresses this issue. Since we are estimating $\mathbf{H}$, i.e., $N_t \times N_r$ unknowns, we need to have at least $N_t \times N_r$ measurements and thus $T_\tau \geq N_t$ would give us a meaningful estimate of $\mathbf{H}$. To maintain the maximum throughput, here we use $T_\tau = N_t$ and $\sigma_\tau^2 = \gamma T/N_t$ with optimal $\gamma = \gamma_o$. In fact it is shown in [1] that when $\gamma = \gamma_o$, the choice $T_\tau = N_t$ maximizes the capacity. From (18) we can write $\hat{h}^{(ki)} = ch^{(ki)} + u^{(ki)}$, and it can be shown that for a given estimate $\hat{h}^{(ki)}$, $h^{(ki)} = c\mu\sigma_h\hat{h}^{(ki)}/\sigma_{\hat{h}} + z^{(ki)}$, where

$$\mu := \frac{E[\mathrm{Real}\{h^{(ki)}\hat{h}^{(ki)*}\}]}{\sqrt{E[|h^{(ki)}|^2]E[|\hat{h}^{(ki)}|^2]}} = \frac{c\sigma_h}{\sigma_{\hat{h}}},$$

and $z^{(ki)} \sim \mathcal{CN}(0, \sigma_u^2\sigma_h^2/\sigma_{\hat{h}}^2)$. Therefore, PEP expression in (14) can be written in terms of μ, σ_z^2 and c and furthermore, with $c = 1$, which is true for high SNR only, (14) becomes same as (13) in [5].

7. PEP-BASED OPTIMAL γ

We now pursue optimal power allocation $\gamma = \gamma_o$ for given frame length T (i.e., total average transmitted power is fixed to T) with optimal $\mathbf{Y}_\tau = \mathbf{Y}_{\tau o}$. Replacing $\sigma_{\tilde{h}}^2$ in $f(\cdot)$ by (17) and using $\sigma_{\hat{h}}^2 = \sigma_h^2 - \sigma_{\tilde{h}}^2$, $\sigma_\tau^2 = \gamma T/T_\tau$ and $\sigma_\delta^2 = (1-\gamma)T/T_\delta$ we can write $f(\cdot)$ in (14)

$$f(\gamma) = \frac{\sigma_h^2T}{4N_t\sigma_w^2(T_\delta - N_t\nu)} \cdot \frac{\gamma(1-\gamma)}{\gamma + b}, \quad (19)$$

where

$$b := \frac{N_t(1 + \frac{\eta\nu T}{T_\delta})}{\eta T(1 - \frac{N_t\nu}{T_\delta})}.$$

Optimal $\gamma = \gamma_o$ can be obtained by differentiating $f(\gamma)$ with respect to γ. Considering the value of $N_t\nu$ compared to T_δ we have the following results.

Case-I: with $T_\delta = N_t\nu$, we have

$$\begin{aligned} \gamma_o &= \arg\max_\gamma \gamma(1-\gamma) = 1/2, \quad (20) \\ f(\gamma_o) &= \left(\frac{(\eta T)^2}{(4N_t)^2(T_\delta + \eta\nu T)}\right). \end{aligned}$$

Case-II: $T_\delta > N_t\nu$ makes $b > 0$ and we have

$$\begin{aligned} \gamma_o &= \arg\max_\gamma \frac{\gamma(1-\gamma)}{\gamma + b} = \sqrt{b(b+1)} - b, \quad (21) \\ f(\gamma_o) &= \frac{\eta T}{4N_t(T_\delta - N_t\nu)}(\sqrt{b+1} - \sqrt{b})^2. \end{aligned}$$

Case-III: $T_\delta < N_t\nu$ makes $b < -1$ and we have

$$\gamma_o = \arg\max_{\gamma} \frac{\gamma(1-\gamma)}{-\gamma - b} = -b - \sqrt{b(b+1)} \tag{22}$$

$$f(\gamma_o) = \frac{\eta T}{4N_t(N_t\nu - T_\delta)}(\sqrt{-b} - \sqrt{-(b+1)})^2.$$

High and Low SNR: At high SNR

$$\gamma_o = \frac{\sqrt{N_t\nu}}{\sqrt{T_\delta} + \sqrt{N_t\nu}} \tag{23}$$

and at low SNR $\gamma_o = 1/2$.
Note that the PEP-based γ_o equals to the capacity-based [1] γ_o with $\nu = 1$.

8. LOSS OF PERFORMANCE AT HIGH SNR

It is appealing to quantify the loss of coding gain due to training at high SNR. We have the following results.
With equal power training $\sigma_\tau^2 = \sigma_\delta^2 = 1$, i.e., $\gamma = T_\tau/T$ and $T_\tau = N_t$ we have

$$P(\mathbf{Y}_\delta \to \mathbf{Y}'_\delta) \le \left(\prod_{m=1}^{N_A} \lambda_m\right)^{-N_r} \left(\frac{\eta}{4N_t} \cdot \frac{N_t}{N_t + N_t\nu}\right)^{-N_r N_A}. \tag{24}$$

With optimal power training, i.e., $\gamma = \gamma_o$ given by (23), we have

$$P(\mathbf{Y}_\delta \to \mathbf{Y}'_\delta) \le \left(\prod_{m=1}^{N_A} \lambda_m\right)^{-N_r} \left(\frac{\eta}{4N_t} \cdot \frac{T_\delta + N_t}{(\sqrt{T_\delta} + \sqrt{N_t\nu})^2}\right)^{-N_r N_A}. \tag{25}$$

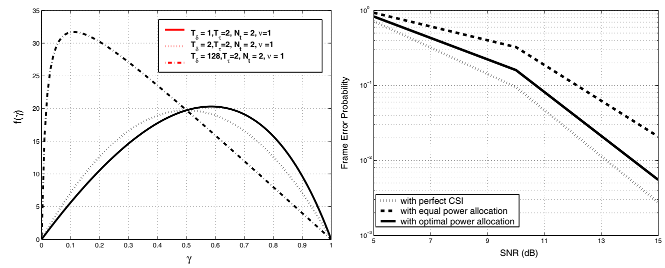
When $\nu = 1$, bounds in (24) and (25) are tight and it appears that equal power training suffers from around 3dB loss as compared to the perfect CSI case whereas the optimal power allocation performs around $10\log[(\sqrt{T_\delta} + \sqrt{N_t})^2/(T_\delta + N_t)]$dB worse than that of perfect CSI case. Therefore optimal power allocation performs same as equal power when $T_\delta = N_t$ and better at higher T_δ.

9. NUMERICAL AND SIMULATION EXAMPLES

We consider a MIMO system with $N_t = N_r = 2$ and $\sigma_h^2 = 1$, $\nu = 1$, i.e., PSK constellations and $T_\tau = N_t$. Optimal values of γ are calculated from the expressions derived for different cases depicted in Section 7 and shown in Table-I. It is observed that γ_0 calculated using (23) for high SNR well-approximates the actual values of γ_o even at low SNR, e.g., 5 dB. Therefore, lower values of T_δ results higher γ_o, i.e., allocation of more power for training. $f(\gamma)$ in (19) vs. γ is plotted in Fig. 1(a) at 25dB SNR. We observe that numerically $f(\gamma)$ is maximum at γ_o obtained from our respective derived expressions. Simulation results on training-based transmission of 4-PSK using space-time trellis codes are shown in Fig. 1(b). We have chosen the 4-state code for 2 transmitter and 2 receiver described in [4]. Training symbols are chosen so that $\mathbf{Y}_\tau^{\mathcal{H}}\mathbf{Y} = T_\tau \mathbf{I}_{N_t}$ where $T_\tau = N_t = 2$, data transmission interval $T_\delta = 128$ and for optimal power allocation γ_o is chosen according to (23). From Fig. 1(b), we observe about 1 dB loss for optimal power allocation, whereas about 2.9 dB loss for equal power allocation. From our derived expressions for PEP in Section 8 these losses are predicted as around 1dB and 3dB respectively verifying the accuracy of our derived expressions.

Table 1. Values of optimal γ in different scenarios

SNR (in dB)	γ_o: Case-I $T_\delta = 2$		γ_o: Case-II $T_\delta = 128$		γ_o:Case-III $T_\delta = 1$	
	Eqn. (20)	(23)	(21)	(23)	(22)	(23)
5	0.5	0.5	0.125	0.111	0.574	0.585
25	0.5	0.5	0.111	0.111	0.585	0.585



(a) (b)

Fig. 1. (a) $f(\gamma)$ defined in (19) vs. γ, the power allocation ratio, at 25dB SNR, (b) Performance of space-time trellis code (STTC) for 4-PSK with 2 b/s/Hz in the presence of training scheme with $N_t = 2, N_r = 2$.

10. CONCLUSION

We consider training design for a MIMO system that minimizes PEP. Optimal training symbols and expressions for optimal power allocation are obtained for different scenarios. It is observed that for constant modulus constellations equal power training performs 3dB worse than the perfect CSI case whereas optimal power training performs in between, depending on the interval of data transmission phase.

Acknowledgment: The authors would like to thank Tansal Gucluoglu for providing the program for the STTC.

11. REFERENCES

[1] B. Hassibi and B. Hochwald, “How much training is needed in multiple-antenna wireless links?” *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951– 963, Apr. 2003.

[2] E. Telatar, “Capacity of multi-antenna gaussian channels,” *Tech. memo., AT&T Bell Labs*, June 1995.

[3] G. J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,” *AT&T Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[4] V. Tarokh, N. Seshadri, and A. R. Calderbank, “Space-time codes for high data rate wireless communication: performance criterion and code construction,” *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[5] V. Tarokh, A. Naguib, and N. Seshadri, “Space-Time codes for high data rate wireless communication: performance criteria in the presence of channel estimation errors, mobility, and multiple paths,” *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 199–207, Feb. 1999.

[6] E. G. Larsson, “Diversity and channel estimation errors,” *IEEE Trans. Commun.*, vol. 52, no. 2, pp. 205–208, Feb. 2004.