OPTIMAL TRAINING FOR FREQUENCY-SELECTIVE FADING CHANNELS

H. Vikalo†, *B. Hassibi*‡, *B. Hochwald*‡, *and T. Kailath*†

† Information System Lab, Stanford University, CA
‡ Bell Labs, Lucent Technologies, Murray Hill, NJ

ABSTRACT

Many communications systems employ training, i.e., the transmission of known signals, so that the channel parameters may be learned at the receiver. This has a dual effect: too little training and the channel is improperly learned, too much training and there is no time left for data transmission before the channel changes. In this paper we use an information-theoretic approach to find the optimal amount of training for frequency selective channels described by a block-fading model. When the training and data powers are allowed to vary, we show that the optimal number of training symbols is equal to the length of the channel impulse response. When the training and data powers are instead required to be equal, the optimal number of symbols may be larger. We further show that at high SNR training-based schemes are capable of capturing most of the channel capacity, whereas at low SNR they are highly suboptimal.

1. INTRODUCTION

Frequency selective fading multipath channels are often encountered in wireless communication systems (see [1] and the references therein). To combat intersymbol interference (ISI) on such channels, receivers use various equalization techniques. Most practical communication systems learn the channel impulse response by means of training—they devote a portion of the transmission time to training symbols known to the receiver. Based on its received signals and the known training data, the receiver can estimate the channel parameters.

In this paper, we take an information-theoretic approach for finding the optimal parameters of a training-based transmission scheme. In particular, we find a lower bound on the capacity of training-based schemes assuming a frequency selective channel with block fading. The optimal training parameters are obtained via maximizing this lower bound. When the training and data powers are allowed to vary, we find the optimal power allocation and show that the optimal length of the training interval is equal to the length of the channel. Our results further show that at high SNR training-based schemes can achieve (most of the) capacity, whereas at low SNR they are highly suboptimal.

This work was supported in part by AFOSR under grant F49620-95-1-0525 and NSF under contract ECS-9529325.

2. CHANNEL MODEL

We assume a block-fading frequency-selective channel model, where the channel coefficients are constant for some discrete interval T, referred to as the *coherence interval*, after which they change to independent values held for another T channel uses, and so on. The block-fading model is a piecewise constant approximation of a time varying channel.

We further assume that the distribution of the coefficients of the channel response is known to both the transmitter and receiver. To obtain the realization of the channel at the receiver, part of each coherence interval is devoted to transmitting known training symbols. Hence training-based schemes comprise the following two phases:

1. **Training Phase**
During the training phase we model the transmission as
$$\mathbf{y}_\tau = \sigma_\tau \Theta_\tau \mathbf{h} + \mathbf{v}_\tau, \tag{1}$$
where $\mathbf{h} \in \mathcal{C}^{L\times 1}$ is the vector of the channel coefficients, $\mathbf{v}_\tau \in \mathcal{C}^{T_\tau \times 1}$ is a vector of independent additive Gaussian noise, and σ_τ^2 is the expected transmit power during the training phase. [In our scheme, the transmit powers during the training and data transmission phases may differ.] For simplicity of the presentation, we shall assume $R_{\mathbf{h}} = Ehh^* = I$. Further, $\Theta_\tau \in \mathcal{C}^{T_\tau \times L}$ is a matrix made up of training symbols known to the receiver,
$$\Theta_\tau = \begin{bmatrix} \theta_1 & 0 & 0 & \dots & 0 \\ \theta_2 & \theta_1 & 0 & \dots & 0 \\ \theta_3 & \theta_2 & \theta_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_L & \theta_{L-1} & \theta_{L-2} & \dots & \theta_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{T_\tau} & \theta_{T_\tau-1} & \theta_{T_\tau-2} & \dots & \theta_{T_\tau-L+1} \end{bmatrix}.$$

Since σ_τ^2 is the expected transmit power, the training symbols vector $\theta_\tau = [\theta_1\ \theta_2\ \dots\ \theta_{T_\tau}]^T$ satisfies $\operatorname{tr} \theta_\tau \theta_\tau^* = T_\tau$. An estimate of the channel is formed from the observed signals, $\mathbf{y}_\tau$, and Θ_τ,

$$\hat{\mathbf{h}} = f(\mathbf{y}_\tau, \Theta_\tau).$$

For a well-determined system of equations (and meaningful estimate) we need $T_\tau \geq L$, that is, at least as many equations as unknowns in (1).

2. **Data Transmission Phase**
For this phase we have

$$\mathbf{y}_d = \sigma_d H_d \mathbf{s}_d + \sigma_\tau H_\tau \theta_\tau + \mathbf{v}_d, \quad E\mathbf{s}_d\mathbf{s}_d^* = R_\mathbf{s},$$

where $\mathbf{s}_d = [s_1\ s_2\ \dots\ s_{T_d}]^T$ is the vector of the transmitted data sequence, and $\mathbf{v}_d \in \mathcal{C}^{(T_d+L-1)\times 1}$ is the vector of additive white complex Gaussian noise with covariance $E\mathbf{v}_d\mathbf{v}_d^* = I$. Furthermore, the matrices $H_d \in \mathcal{C}^{(T_d+L-1)\times T_d}$ and $H_\tau \in \mathcal{C}^{(T_d+L-1)\times T_\tau}$ are defined as

$$H_d = \begin{bmatrix} h_1 & & & & & \\ h_2 & h_1 & & & & \\ \vdots & \vdots & \ddots & & & \\ h_L & h_{L-1} & \dots & h_1 & & \\ & \ddots & & & \ddots & \\ & & h_L & h_{L-1} & \dots & h_1 \\ & & & \ddots & \vdots & \vdots \\ & & & & h_L & h_{L-1} \\ & & & & & h_L \end{bmatrix}$$

$$H_\tau = \begin{bmatrix} h_L & h_{L-1} & \dots & h_2 \\ & h_L & \dots & h_3 \\ & & \ddots & \vdots \\ & & & h_L \\ & & & \end{bmatrix}.$$

The expected transmission power is σ_d^2. The estimate of the channel, $\hat{\mathbf{h}}$, is used to detect $\mathbf{s}_d$ from

$$\underbrace{\mathbf{y}_d - \sigma_\tau \hat{H}_\tau \theta_\tau}_{\mathbf{y}_d'} = \sigma_d \hat{H}_d \mathbf{s}_d + \underbrace{\sigma_d \tilde{H}_d \mathbf{s}_d + \sigma_\tau \tilde{H}_\tau \theta_\tau + \mathbf{v}_d}_{\mathbf{v}_d'}, \qquad (2)$$

where $\mathbf{v}_d'$ is the effective noise comprised of the additive noise and residual channel estimation error, and $\mathbf{y}_d'$ denotes the combination of the measured and known signals during the data transmission phase.

We note that the following relations hold due to conservation of time and energy,

$$T = T_\tau + T_d, \quad \sigma^2 T = \sigma_\tau^2 T_\tau + \sigma_d^2 T_d.$$

Clearly, increasing T_τ improves the channel estimate but that is achieved at the expense of the length of the data transmission interval T_d. Similar trade-off holds for σ_τ^2 and σ_d^2. We are interested in finding optimal (from the capacity point of view) parameters $(T_\tau, T_d, \sigma_\tau^2, \sigma_d^2)$ along with the optimal training sequence θ_τ. We should also mention that the channel estimate $\hat{\mathbf{h}}$ may, for instance, be the maximum-likelihood or linear minimum-mean-square-error estimate. Note that MMSE estimate $\hat{\mathbf{h}}$ is the conditional mean of $\mathbf{h}$ given θ_τ and $\mathbf{y}_\tau$. Hence $\hat{\mathbf{h}}$ and $\tilde{\mathbf{h}}$ are uncorrelated and so are $\hat{H}$ and $\tilde{H}$. Thus, when the channel estimate is MMSE, the effective noise $\mathbf{v}_d'$ in (2) is uncorrelated with the signal $\mathbf{s}_d$.

3. CAPACITY BOUNDS AND OPTIMAL PARAMETERS OF THE TRAINING-BASED TRANSMISSION SCHEME

The capacity in bits per channel use in the training-based scheme can be expressed as

$$\begin{aligned} C_\tau &= \sup_{p_{\mathbf{s}_d}, trE\mathbf{s}_d\mathbf{s}_d^* \leq T_d} \frac{1}{T} I(\mathbf{y}_\tau, \theta_\tau, \mathbf{y}_d; \mathbf{s}_d) \\ &= \sup_{p_{\mathbf{s}_d}, trE\mathbf{s}_d\mathbf{s}_d^* \leq T_d} \frac{1}{T}(I(\mathbf{y}_d, \mathbf{s}_d | \mathbf{y}_\tau, \theta_\tau) + \underbrace{I(\mathbf{y}_\tau, \theta_\tau; \mathbf{s}_d)}_{=0}) \end{aligned}$$

that is, the capacity in a training-based scheme is the supremum of the mutual information between the transmitted and received signals during the data transmission phase, given the transmitted and received signals during the training phase. In general, finding this capacity is a hard problem. Therefore, we find a lower bound on the capacity for a particular choice of the channel estimate. From (2),

$$\mathbf{y}_d' = \sigma_d \hat{H}_d \mathbf{s}_d + \mathbf{v}_d'. \qquad (3)$$

We assume that $\hat{H}$ in (3) is obtained from the mean-square error (MMSE) estimate of the channel $\mathbf{h}$. The choice of the estimator is driven by its property that the additive noise and signal in (3) are then uncorrelated; thus, $\mathbf{v}_d'$ in (3) is additive noise uncorrelated with $\mathbf{s}_d$. The training-based scheme assumes that the channel estimate $\hat{\mathbf{h}}$ (and, consecutively, $\hat{H}$ in (3)) is correct, an assumption often made in practical transmission schemes. Hence, the channel capacity of the training-based scheme is as same as the capacity of a known channel system, subject to the additive noise with the covariance matrix $R_{\mathbf{v}_d'} =$

$$E\mathbf{v}_d'\mathbf{v}_d'^* = E\left[\sigma_d^2 \tilde{H}_d \mathbf{s}_d \mathbf{s}_d^* \tilde{H}_d^* + \sigma_\tau^2 \tilde{H}_\tau \theta_\tau \theta_\tau^* \tilde{H}_\tau^*\right] + I. \quad (4)$$

Choosing the signal covariance $R_\mathbf{s} = I$ leads to a lower bound on $C_\tau =$

$$\begin{aligned} &\max_{R_\mathbf{s}, trR_\mathbf{s} = T_d} E\frac{T - T_\tau}{T + L - 1} \log\det\left(I + \sigma_d^2 R_\mathbf{v}^{-1} \hat{H}_d^* R_\mathbf{s} \hat{H}_d\right) \\ &\geq E\frac{T - T_\tau}{T + L - 1} \log\det\left(I + \sigma_d^2 R_\mathbf{v}^{-1} \hat{H}_d^* \hat{H}_d\right) \end{aligned}$$

Define the normalized channel, $\bar{H}_d$,as

$$\bar{H}_d = \sqrt{\frac{LT_d}{\text{trE}(\hat{H}_d^* \hat{H}_d)}}\hat{H}_d.$$

Then we can write capacity bound as

$$C_\tau \geq E\frac{T-T_\tau}{T+L-1}\log\det\left(I+\sigma_d^2\frac{\sum_{i=1}^{L}\sigma_{\hat{h}_i}^2}{L}R_{\mathbf{v}_d'}^{-1}\bar{H}_d\bar{H}_d^*\right) \quad (5)$$

We are interested in finding parameters of the transmission scheme that maximize the capacity lower bound in (5). In particular, we maximize the lower bound on capacity in (5) with respect to the training data sequence θ_τ, training power σ_d^2, and length of the training interval T_τ. The result is given below and the proof is omitted for brevity.

Theorem 1 (Optimizing the training-based scheme) *The optimal length of the training interval for the training-based transmission scheme over a frequency-selective channel is equal to the length of the channel, $T_\tau = L$, and the lower bound on the capacity is given by*

$$C_\tau \geq \frac{T-L}{T+L-1}E\log\det(I+\rho_{eff}\bar{H}\bar{H}^*),$$

where

$$\rho_{eff} = \begin{cases} \frac{\sigma^2 T}{T-2L}(\sqrt{\gamma}-\sqrt{\gamma-1})^2 & \text{for } T > 2L \\ \frac{(\sigma^2 T)^2}{4L(1+\sigma^2 T)} & \text{for } T = 2L \\ \frac{\sigma^2 T}{2L-T}(\sqrt{-\gamma}-\sqrt{-\gamma+1})^2 & \text{for } T < 2L \end{cases}$$

and $\gamma = \frac{(T-L)(1+\sigma^2 T)}{\sigma^2 T(T-2L)}$.
The optimal power allocation is given by

$$\sigma_d^2 = \begin{cases} (\gamma-\sqrt{\gamma(\gamma-1)})\sigma^2\frac{T}{T-L} & \text{for } T > 2L \\ \frac{1}{2}\sigma^2\frac{T}{T-L} & \text{for } T = 2L \\ (\gamma+\sqrt{\gamma(\gamma-1)})\sigma^2\frac{T}{T-L} & \text{for } T < 2L \end{cases},$$

$$\sigma_\tau^2 = \sigma_d^2 + (\sigma^2-\sigma_d^2)\frac{T}{L}.$$

For high and low SNR the results of Theorem 1 specialize as follows.

Corollary 1 (High and low SNR)

1. *At high SNR, lower bound on capacity is given by*

$$C_\tau \geq E\frac{T-L}{T+L-1}\log\det(I+\frac{\sigma^2 T}{(\sqrt{T_d}+\sqrt{L})^2}\bar{H}_d\bar{H}_d^*),$$

while the optimal power allocation is

$$\sigma_d^2 = \frac{\sqrt{T_d}}{\sqrt{T_d}+\sqrt{L}}\cdot\sigma^2\frac{T}{T_d}.$$

2. *At low SNR, lower bound on capacity is given by*

$$C_\tau \geq E\frac{T-L}{T+L-1}\log\det(I+\frac{(\sigma^2 T)^2}{4T_d}\bar{H}_d\bar{H}_d^*),$$

while the optimal power allocation is given by

$$\sigma_d^2 = \frac{1}{2}\cdot\sigma^2\frac{T}{T_d}.$$

Some comments regarding Theorem 1 are appropriate. Intuitively, longer training intervals provide better estimates of the channel, thus decreasing the power of the effective noise. However, longer training intervals mean less time for data transmission. Theorem 1 implies that spending time sending data is more important than spending time training; the optimum training interval is set to its minimum meaningful length. Note that increasing the training interval increases the capacity logarithmically (in lower noise power), but decreases it linearly (in time).

3.1. Equal powers

The assumption made throughout the paper is that the communication system can provide two different transmission power levels, one for the training and one for the data transmission phase. However, if practical constraints impose equal power, i.e., $\sigma_\tau^2 = \sigma_d^2 = \sigma^2$, the capacity lower bound can be written as

$$C_\tau \geq E\frac{T-T_\tau}{T+L-1}\log\det\left(I+\frac{\sigma^4 T_\tau}{1+\sigma^2(T_\tau+L)}\bar{H}_d\bar{H}_d^*\right).$$

Further simplifications of this capacity lower bound expressions are possible for the special cases of high and low SNR.

1. At high SNR, we can write the capacity lower bound as

$$C_\tau \geq E\frac{T-T_\tau}{T+L-1}\log\det\left(I+\frac{\sigma^2 T_\tau}{T_\tau+L}\bar{H}_d\bar{H}_d^*\right). \quad (6)$$

Optimum length of the training interval can be obtained by evaluating (6) for various T_τ, $L \leq T_\tau < T$.

2. At low SNR, using $\log(I+A) = \log e(A - A^2/2 + A^3/3\ldots)$, we obtain following expression for the capacity lower bound

$$C_\tau \geq \sigma^4 L\log(e)\frac{T-T_\tau}{T+L-1}T_\tau(T-T_\tau).$$

Upon taking the derivative with respect to T_τ, one can notice that the capacity bound is maximized for T_τ found as a solution of the quadratic equation

$$T_\tau^2 - \frac{4}{3}TT_\tau + \frac{1}{3}T^2 = 0.$$

Solving for T_τ, we find that $T_\tau = \frac{1}{3}T$, and a third of a coherence interval should be devoted to training.

4. SIMULATION RESULTS AND CONCLUSION

Figure 1 shows the training-based lower bounds on capacity as a function of the block length T for $\sigma^2 = 6dB$ and the channel length $L = 4$. By allowing the training and data transmission powers to vary, we achieve approximately $5 - 10\%$ increase in capacity. At $T = 50$, achieved capacity is approximately 20% below the (unrealistic) capacity achieved when the receiver knows the channel perfectly.

In Figure 2, the optimal transmit power allocation σ_d^2 and σ_τ^2 is plotted as a function of the block length. The dashed line in Figure 2 denotes the case of equal training and data transmission powers $\sigma_d^2 = \sigma_\tau^2$. Figure 2 illustrates what is implied by Theorem 1 — we need to spend more power for training than for transmission when $T > 2L$, more power for transmission than for training when $T < 2L$, and the same power for both when $T = 2L$.

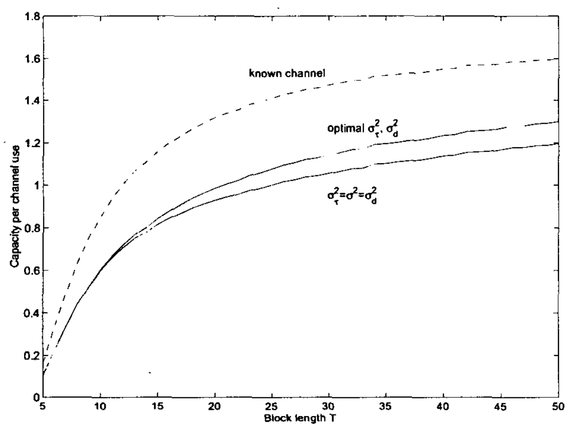

Fig. 1. Training-based lower capacity bound

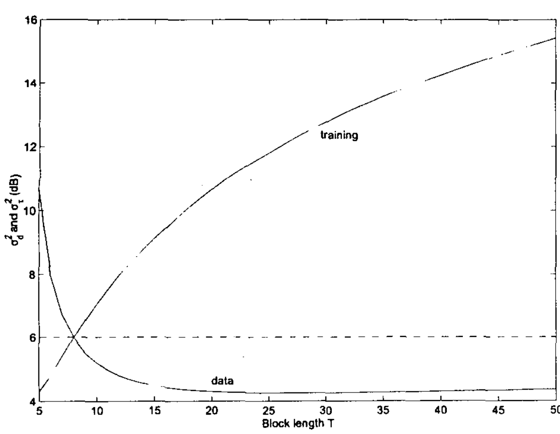

Fig. 2. Optimal power allocation

Now that we have determined the optimal amount of training for any training-based communication system, the question that remains is *how good are training-based schemes*? To answer this question one would need to compute the actual capacity of a block-fading frequency-selective channel and to compare it with the training-based capacity lower bounds we obtained. Unfortunately,com puting this capacity, in the general case, is an open problem. However, we have the following result whose proof we omit for brevity.

Theorem 2 *At high SNR* ($\sigma \to \infty$)*, the capacity of a block-fading frequency-selective channel with coherence interval T is given by*

$$C = (T - L)\log \sigma^2 + O(1). \quad (7)$$

Alternatively, at low SNR ($\sigma \to 0$)*, we have*

$$C = O(\sigma^2). \quad (8)$$

Now studying Theorem 1 at high SNR yields $\rho_{\text{eff}} = \frac{\sigma^2 T}{(\sqrt{T-L}+\sqrt{L})^2}$. Thus, since $\bar{H}\bar{H}^*$ is generically nonsingular:

$$\begin{aligned} C_\tau &\geq \frac{T-L}{T+L-1} E \log\det\left(I + \frac{\sigma^2 T}{(\sqrt{T-L}+\sqrt{L})^2}\bar{H}\bar{H}^*\right) \\ &\geq \frac{T-L}{T+L-1} E \log\det\left(\frac{\sigma^2 T}{(\sqrt{T-L}+\sqrt{L})^2}\bar{H}\bar{H}^*\right) \\ &= \frac{T-L}{T+L-1}\log\det\sigma^2 I_{T+L-1} + O(1) \\ &= (T-L)\log\sigma^2 + O(1). \end{aligned}$$

In other words, training-based schemes achieve capacity at high SNR!

At low SNR, on the other hand, examination of Theorem 1 yields

$$C_\tau \geq O(\sigma^4), \quad (9)$$

and, in fact, this bound is tight at low SNR because the additive noise $\mathbf{v}'_d$ is almost Gaussian. Comparing this to Theorem 2 shows that training-based schemes are highly suboptimal at low SNR.

5. REFERENCES

[1] E. Biglieri, J. Proakis, and S. Shamai, "Fading Channels: Information-Theoretic and Communications Aspects," *IEEE Trans. on Inf. Theory*, vol. 44, no. 6, Oct. 1998.

[2] B. Hassibi and B. M. Hochwald. "How Much Training is Needed in Multiple-Antenna Wireless Links?" submitted to *IEEE Trans. on Inf. Theory*. Download available at `http://mars.bell-labs.com`.

[3] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *Eur.Trans. Telecom.*, vol. 10, pp. 585-595, Nov. 1999.