

# OPTIMAL EMBEDDING OF KNOWN SYMBOLS FOR OFDM

*Srihari Adireddy and Lang Tong*

School of Electrical Engineering  
Cornell University  
Ithaca, NY 14853  
{srihari,ltong}@ee.cornell.edu

*Harish Viswanathan*

Bell Laboratories  
791 Holmdel-Keyport Road  
Holmdel, NJ 07733-0400  
harishv@lucent.com

## ABSTRACT

The problem of placing known symbols optimally for OFDM is considered. The channel is assumed to be quasi-static with a finite impulse response. Under the assumption that neither the transmitter and receiver know the channel, we optimize the training by maximizing a lower bound of the mutual information. It is shown that the lower bound is maximized by placing the known symbols periodically. Optimum energy trade-off between the training and the data is also obtained and illustrated through simulation.

## 1. INTRODUCTION

Orthogonal Frequency Division Multiplexing (OFDM) has emerged as an attractive modulation scheme for high data rate communication systems. It is being presently used in standards like Digital Video Broadcast (DVB) and Digital Audio Broadcast (DAB). Proposals for fourth generation systems include those that use OFDM as the modulation scheme. All of these standards mandate that the data stream contain known symbols. These known symbols serve various purposes such as training of receivers, synchronization. A significant fraction of the available bandwidth might be used for sending these known symbols especially in the mobile communications scenario where the channel changes rapidly.

While the insertion of known symbols leads to a reduction in maximum achievable data rate, it is mandatory to simplify the task of the receiver. The demand for higher bit rate motivates us to investigate into optimal design of training. This involves issues like amount of training, choice of training, and placement of training. The focus of this paper is designing optimal placement schemes of training symbols for OFDM systems. Hassibi and Hochwald [1] have considered the optimization of training and energy and data trade-off for multiple antenna systems for the flat fading scenario. In this case the performance is independent on the placement of known symbols. The problem of placement becomes important for the case of frequency selective fading channels.

We have previously considered the problem of joint optimization of symbol placement and equalizer for a symbol by symbol decision feedback receiver [3]. The performance criterion used

was average MSE. Optimal known symbol placement schemes that maximize the i.i.d mutual information between the input and the output were considered in [4] for single carrier and multi-carrier transmission systems. Since the purpose was to determine the effect of known symbol placement on detection, the channel was assumed to be known. It was shown that in the single carrier case, mutual information is maximized by breaking the known symbols into small blocks and placing them periodically. Optimal known symbol placement schemes for single carrier transmission case when the frame length goes to infinity with the percentage of known symbols is fixed was dealt with in [5]. In this scenario, since the number of known symbols goes to infinity, the assumption that the receiver knows the channel is reasonable.

Known symbol placement schemes for OFDM have been considered in [2]. The metric used for optimizing the placement is the mean square error of the channel estimate. However, channel estimation is just one facet of the problem. The placement of known symbols affects not only the channel estimate but also the detection. In this paper, we take the holistic view and try to optimize the placement of known symbols by maximizing the mutual information. We assume that the receiver and the transmitter do not know the channel. It is difficult to obtain the mutual information expressions for this scenario. Hence we obtain a tight lower bound as in [1] that is analytically tractable. Optimal placement of training for single carrier systems under the same frame work has been considered in [6].

Under the assumption that all the training symbols have equal energy, we show that the mutual information is maximized by a simple placement scheme. The known symbols are placed periodically in the OFDM symbol. That is, we pick equally spaced tones for training. This is the placement scheme that was also obtained in [2]. It is quite remarkable that this placement not only gives the best channel estimate but also maximizes the mutual information. This is not true in general as is evident from our results for single carrier systems [4].

This paper is organized as follows. In Section 2, we introduce the channel model and the transmission system. In Section 3, we state the general problem of optimizing training to maximize the training based i.i.d capacity of the system. In Section 4, we introduce a tight lower bound for the i.i.d capacity of the system and then we state the optimization problem in terms of this lower bound. In Section 5 we state the theorem about optimal training placement and optimal energy trade-off. In Section 6 we illustrate the ideas through simulations and finally conclude in Section 7.

---

This work was supported in part by the National Science Foundation under Contract CCR-9804019 and the Multidisciplinary University Research Initiative (MURI) under the Office of Naval Research Contract N00014-00-1-0564.

## 2. SYSTEM MODEL

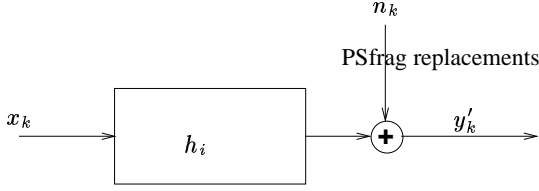


Fig. 1: System Model

In this section we give the channel model and define the transmission model for OFDM. The system model is shown in Fig 1. The channel  $\mathbf{h} = [h_0, h_1, \dots, h_L]^t$  has a finite impulse response of length  $(L + 1)$ .

We assume that taps of the channel  $\mathbf{h}$  are i.i.d complex Gaussian with zero mean and variance equal to  $\frac{1}{L+1}$ . The fading coefficients remain constant for  $T$  symbol periods and change to a new independent value. We assume that the receiver and the transmitter do not know the fading coefficients. The received signal is corrupted by additive white noise that is complex Gaussian with zero mean and variance  $\sigma_w^2$ . This model described above is an extension of the quasi-static flat fading model to quasi-static frequency selective fading.



Fig. 2: Transmitter side processing

Figure 2 shows the processing performed at the transmitter. The symbol stream is parsed into blocks of length  $(T - L)$  by the serial to parallel converters (S/P). We call these blocks OFDM blocks. This block is passed through a Inverse Discrete Fourier Transform (IDFT) matrix. The cyclic prefix (CP) of length  $L$  is appended to this block to form a super block. We then perform a parallel to serial (P/S) conversion of the super blocks and transmit them. We assume that the channel stays constant over the duration of a super block. Each OFDM block is of length  $(N + P)$  where  $N$  is number of number of unknown symbols and  $P$  is the number of known symbols in the OFDM block. This implies that total  $(N + P)$  tones are used for transmission. Known symbols are introduced in frequency as is the norm for most OFDM standards. The vector  $\mathbf{s} = [s_1, s_2, \dots, s_{(N+P)}]^t$  is formed by collecting the symbols in each OFDM block.



Fig. 3: Receiver Side Processing

Figure 3 illustrates the processing performed at the receiver.

At the receiver, the output that is due to symbols from two different OFDM blocks is removed. The remaining data is parsed into blocks of length  $(T - L)$  by the serial to parallel (S/P) converters and passed through a Discrete Fourier Transform (DFT) matrix. The vector  $\mathbf{y}$  is formed by collecting the output due to the block  $\mathbf{s}$ .

The channel is completely specified by the relation between the input  $\mathbf{s}$  and the output  $\mathbf{y}$ . The channel law is given by

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N+P} \end{bmatrix} = \underbrace{\begin{bmatrix} d_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & d_2 & & \mathbf{0} \\ & & \ddots & \\ \mathbf{0} & & & d_{N+P} \end{bmatrix}}_{\mathbf{D}} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{N+P} \end{bmatrix} + \mathbf{w} \quad (1)$$

where  $d_i$  is the  $i^{\text{th}}$  Fourier coefficient of the  $(N + P)$  point DFT of the channel  $\mathbf{h}$ . That is,

$$\begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_{N+P} \end{bmatrix} = \sqrt{(N + P)} \mathbf{W}_L \mathbf{h} \quad (2)$$

where  $\mathbf{W}_L$  is the truncated unit norm DFT matrix of size  $(N + P) \times (L + 1)$ . The  $(k, l)^{\text{th}}$  entry of this matrix is given by  $\frac{1}{\sqrt{N+P}} \exp \frac{-j2\pi(k-1)(l-1)}{N+P}$ . Intuitively speaking, the OFDM transmission scheme converts frequency selective fading in time to flat fading on each tone. The vector  $\mathbf{w}$  is additive, white Gaussian noise of variance  $\sigma_w^2$ .

## 3. PROBLEM STATEMENT

In this section we formulate the problem of designing optimal training. Training symbols are introduced in  $\mathbf{s}$  to estimate the channel  $\mathbf{h}$ . We define as set  $\mathcal{P}$ , the indices of the tones used for training and as set  $\mathcal{P}_e$ , the indices of the tones used for transmitting data. The placement scheme is completely specified by the set  $\mathcal{P}$ . We denote as  $\mathbf{s}_t = (s_{1t}, \dots, s_{Pt})^t$  the vector of symbols used for training. We use the subscript  $1t$  to represent the smallest element of the set  $\mathcal{P}$  and so on. Let  $\mathbf{s}_d$  be the vector of data symbols namely  $\mathbf{s}_d = (s_{1d}, \dots, s_{Nd})^t$ .

The power constraint on the system is formulated as

$$\frac{1}{(N + P)} (\mathbb{E} \text{tr} \mathbf{s}_d \mathbf{s}_d^H + \text{tr} \mathbf{s}_t \mathbf{s}_t^H) = 1. \quad (3)$$

We do not constrain the data and training powers to be same. If  $\rho_d = \frac{1}{N} \mathbb{E} \text{tr} \mathbf{s}_d \mathbf{s}_d^H$  and  $\rho_t = \frac{1}{P} \text{tr} \mathbf{s}_t \mathbf{s}_t^H$ , then the above equation can be written as

$$\frac{N\rho_d + P\rho_t}{N + P} = 1. \quad (4)$$

In most OFDM systems, all the training symbols are constrained to be of the same energy. Hence we assume that  $|s_{it}|^2 = \rho_t$ ,  $i = 1, \dots, P$ .

We restrict ourselves to receivers of the structure given in Figure 4. We assume that the receiver forms the MMSE estimate of the channel using only training. The receiver then uses  $\mathbf{y}_d$  and the MMSE estimate  $\hat{\mathbf{D}}$  to perform the decoding. Such receivers are computationally simple, analytically tractable and are used

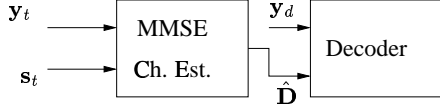


Fig. 4: Receiver Structure

in practical systems. The i.i.d capacity of the system [7] is then equal to

$$C(\mathcal{P}, \rho_d, \rho_t, \mathbf{s}_t) \triangleq \max_{f_{i.i.d}(\mathbf{s}_d)} I(\mathbf{y}_d, \hat{\mathbf{D}}; \mathbf{s}_d) \quad (5)$$

where the probability distribution  $f_{i.i.d}(\mathbf{s}_d)$  and the training  $\mathbf{s}_t$  are such that the input power constraint is satisfied.

Our objective then is to obtain optimal placement scheme  $\mathcal{P}^*$ , optimal energy trade-off  $(\rho_d^*, \rho_t^*)$  and optimal training symbols  $\mathbf{s}_t^*$  as

$$(\mathcal{P}^*, \rho_d^*, \rho_t^*, \mathbf{s}_t^*) = \arg \max_{\mathcal{P}, \rho_d, \rho_t, \mathbf{s}_t} C(\mathcal{P}, \rho_d, \rho_t, \mathbf{s}_t). \quad (6)$$

#### 4. LOWER BOUND ON TRAINING-BASED CAPACITY

In this section we obtain a tight lower bound for  $C(\mathcal{P}, \rho_d, \rho_t, \mathbf{s}_t)$  and optimize training with respect to this bound. We have

$$\begin{aligned} C(\mathcal{P}, \mathbf{s}_t, \rho_d, \rho_t) &= \max_{f_{i.i.d}(\mathbf{s}_d)} I(\mathbf{y}_d; \mathbf{s}_d | \hat{\mathbf{D}}) + \underbrace{I(\hat{\mathbf{D}}; \mathbf{s}_d)}_{=0} \\ &= \max_{f_{i.i.d}(\mathbf{s}_d)} I(\mathbf{y}_d; \mathbf{s}_d | \hat{\mathbf{D}}), \end{aligned} \quad (7)$$

because the MMSE estimate  $\hat{\mathbf{D}}$  is independent of  $\mathbf{s}_d$ . The relationship between  $\mathbf{y}_d$  and  $\mathbf{s}_d$  is given by

$$\mathbf{y}_d = \underbrace{\begin{bmatrix} d_{1d} & & \\ & \ddots & \\ & & d_{Nd} \end{bmatrix}}_{\mathbf{D}_d} \underbrace{\begin{bmatrix} s_{1d} \\ \vdots \\ s_{Nd} \end{bmatrix}}_{\mathbf{s}_d} + \underbrace{\begin{bmatrix} w_1 \\ \vdots \\ w_{Nd} \end{bmatrix}}_{\mathbf{w}_d}. \quad (8)$$

This can be rewritten as

$$\mathbf{y}_d = \hat{\mathbf{D}}_d \mathbf{s}_d + \underbrace{\tilde{\mathbf{D}}_d \mathbf{s}_d + \mathbf{w}_d}_{\boldsymbol{\nu}_d} \quad (9)$$

where  $\hat{\mathbf{D}}_d$  is the estimate of  $\mathbf{D}_d$  and  $\tilde{\mathbf{D}}_d$  is the error in the estimate. It is difficult to evaluate the i.i.d capacity because the distribution of  $\boldsymbol{\nu}_d$  is difficult to characterize. Therefore we obtain a lower bound on the i.i.d channel capacity then reformulate the problem of optimization in terms of this lower bound. We obtain the lower bound by letting the probability distribution of  $\boldsymbol{\nu}_d$  be that of the worst possible noise [1] among the class of the distributions that have the same first and second order statistics as  $\boldsymbol{\nu}_d$ .

It is easy to see that  $\boldsymbol{\nu}_d$  is zero mean. The auto-correlation of  $\boldsymbol{\nu}_d$  is given by

$$\mathbf{R}_{\boldsymbol{\nu}} = \mathbb{E} \tilde{\mathbf{D}}_d \mathbf{s}_d \mathbf{s}_d^H \tilde{\mathbf{D}}_d^H + \mathbb{E} \mathbf{w}_d \mathbf{w}_d^H \quad (10)$$

$$= \rho_d \mathbf{R}_e + \sigma_w^2 \mathbf{I} \quad (11)$$

$$(12)$$

The matrix  $\mathbf{R}_e$  is diagonal since the symbols  $s_{kd}$  and  $s_{ld}$  are independent for  $k \neq l$ . We also note that the noise  $\boldsymbol{\nu}_d$  is uncorrelated to the signal  $\hat{\mathbf{D}}_d \mathbf{s}_d$ . This is because the error  $\tilde{\mathbf{D}}_d$  is uncorrelated with  $\hat{\mathbf{D}}_d$  (property of the MMSE estimate).

It can be shown, as in [1], that the worst case noise then has a Gaussian probability distribution that is independent of the signal. Therefore we have

$$\begin{aligned} C(\mathcal{P}, \rho_d, \rho_t, \mathbf{s}_t) &\geq C_{lb}(\mathcal{P}, \rho_d, \rho_t, \mathbf{s}_t) \\ C_{lb}(\mathcal{P}, \rho_d, \rho_t, \mathbf{s}_t) &\triangleq \mathbb{E} \log \det \left( \mathbf{I} + \rho_d \mathbf{R}_{\boldsymbol{\nu}}^{-1} \hat{\mathbf{D}}_d \hat{\mathbf{D}}_d^H \right) \end{aligned} \quad (13)$$

where the expectation is with respect to the random variable  $\hat{\mathbf{D}}_d$ . At low SNR, the probability distribution of  $\boldsymbol{\nu}_d$  is in fact Gaussian. This expects us to believe that the bound is tight at low SNR. We conjecture that using the same arguments as in [1] we can show that the bound is tight at high SNR.

#### 5. OPTIMAL TRAINING

In this section we optimize the placement by maximizing the lower bound on capacity. We find that  $C_{lb}(\mathcal{P}, \rho_d, \rho_t, \mathbf{s}_t)$  is independent of  $\mathbf{s}_t$  as long as we use equal energy training schemes. We can therefore exclude  $\mathbf{s}_t$  as an argument of  $C_{lb}$  and obtain optimal training  $(\mathcal{P}^*, \rho_d^*, \rho_t^*)$  as

$$(\mathcal{P}^*, \rho_d^*, \rho_t^*) = \arg \max_{\mathcal{P}, \rho_d, \rho_t} C_{lb}(\mathcal{P}, \rho_d, \rho_t) \quad (14)$$

We then have the following theorem.

**Theorem 1** Under the assumption that  $N = mP$ , and  $P \geq (L + 1)$ , all the following placements are optimal

$$\mathcal{P}^* = \{i, i+m+1, i+2(m+1), \dots, i+(P-1)(m+1)\} \quad (15)$$

where  $i$  can take values from 1 to  $(m+1)$ . For any of these placements the lower bound is given by

$$C_{lb}(\mathcal{P}^*, \rho_d, \rho_t) = NE \log \left( 1 + \frac{\rho_d}{\sigma_w^2} \frac{P \rho_t d d^*}{P \rho_t + (L+1)(\rho_d + \sigma_w^2)} \right) \quad (16)$$

where  $d$  is a complex Gaussian random variable with zero mean and unit variance. The optimal energy distribution is given by

$$\begin{aligned} \rho_d^* &= (\sqrt{\gamma} - \sqrt{\gamma-1}) \frac{\sqrt{\gamma}}{g} \\ \rho_t^* &= (\sqrt{\gamma} - \sqrt{\gamma-1}) \frac{1}{\sqrt{hk}}. \end{aligned} \quad (17)$$

$$(18)$$

where  $h = \frac{P}{(N+P)}$ ,  $g = \frac{N}{(N+P)}$ ,  $k = \frac{P(N-L-1)}{(L+1)((N+P)+N\sigma_w^2)}$  and  $\gamma = \frac{h}{k} + 1$ .

*Proof:* Refer to [6].

It was shown in [2] that the same set of placements minimizes the mean square error in the estimate of  $\mathbf{h}$ . It should be noted that this does not imply that these placements should minimize  $C_{lb}(\mathcal{P}, \rho_d, \rho_t)$ . Indeed there exist placements such that some data tones have a lower MSE than that with  $\mathcal{P}^*$ .

## 6. SIMULATION

Figure 5 shows the variation of  $C_{lb}$  with the percentage of known symbols at SNR=20dB. Figure 6 show the same at an SNR of 0dB. In either case we plot the curves for both  $\rho_d = \rho_t = 1$  and also  $\rho_d, \rho_t$  optimized case. We assume that the optimal known symbol placement scheme is used when  $P$  is a multiple of  $N$ . It should be noted for some values of  $P$ , it might not be possible to place the known symbols periodically. We find that when we optimize  $\rho_d, \rho_t$  the lower bound is maximized by making the number of known symbols as small as possible which in this case is equal to  $(L + 1)$ . For the equal training and data energy case, the lower bound first rises fast, reaches a maximum and then falls at a slower rate. We find that, as expected, the optimum percentage of known symbols decreases with increasing SNR. At high SNR, the gain in optimizing  $\rho_d, \rho_t$  is minimal.

Figure 7 shows the variation of the lower bound as the coherence interval of the channel increases for both  $\rho_d, \rho_t$  optimized and  $\rho_d = \rho_t = 1$  cases. We find that as  $T$  increases the lower bound converges to the value of the known channel capacity. This implies that at large  $T$ , capacity can be achieved by training.

## 7. CONCLUSIONS

Under the assumption that the receiver forms training-only based MMSE estimate of the channel, we obtained the optimal known symbol placement schemes that maximize a tight lower bound on mutual information for OFDM systems. We find that the mutual information is maximized by selecting equally spaced symbols as training. It is surprising that such a simple placement has the optimality property. We also obtain optimum energy trade-off between data and training. We find from simulations that when we optimize data and training energy  $\rho_d, \rho_t$  the lower bound is maximized by making  $P$  as small as possible. When we constrain ourselves to equal training and data energy case, the lower bound reaches maximum at a particular percentage of known symbols. This optimum percentage decreases with SNR. From simulations, we find that as the coherence interval of the channel goes to infinity training is optimal. That is we can achieve the capacity of the channel with training.

## 8. REFERENCES

- [1] B.Hassibi and B.Hochwald. "How much Training is Needed in Multiple-Antenna Wireless Links". *Submitted to IEEE Trans. Information Theory*, August 2000.
- [2] R.Negi and J.Cioffi. "Pilot Tone Selection for Channel Estimation in a Mobile OFDM System". *IEEE Tran. Consumer Electronics*, 44(3):1122–1128, August 1998.
- [3] S.Adireddy and L.Tong. "Detection with Embedded Known Symbols : Optimal Symbol Placement and Equalization". In *Proc. ICASSP*, Istanbul, Turkey, May 2000.
- [4] S.Adireddy and L.Tong. "Optimal Placement of Known Symbols". *Submitted to IEEE Tran. Information Theory*, July 2000.
- [5] S.Adireddy and L.Tong. "Repeatable Quasi Periodic Placement of Known Symbols". In *Submitted to ISIT*, Washington DC, June 2001.
- [6] S.Adireddy, L.Tong, and H.Viswanathan. "Optimal Placement of Training for OFDM and Single Carrier Systems". *Submitted to IEEE Tran. Information Theory*, February 2001.
- [7] Shlomo Shamai (Shitz) and Rajiv Laroia. "The Intersymbol Interference Channel: Lower bounds on Capacity and Channel Precoding Loss". *IEEE Trans. Inform. Theory*, 42(5):1388–1404, September 1996.

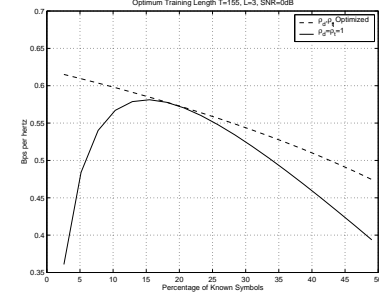


Fig. 5: Variation of lower bound with percentage of known symbols at SNR = 0dB,  $T = 155$  and  $L = 3$ .

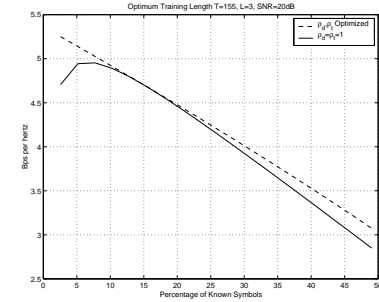


Fig. 6: Variation of lower bound with percentage of known symbols at SNR = 20dB,  $T = 155$  and  $L = 3$ .

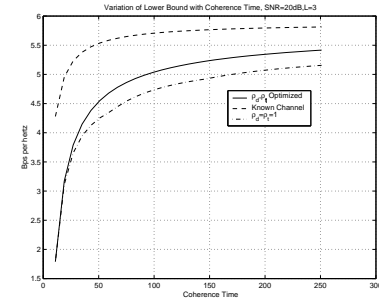


Fig. 7: Variation of lower bound with coherence interval at SNR = 20dB and  $L = 3$