

ADAPTIVE COMMUNICATIONS AND SIGNAL PROCESSING LABORATORY  
CORNELL UNIVERSITY, ITHACA, NY 14853

---

## **Channel Estimation under Asynchronous Packet Interference**

**Cristian Budianu and Lang Tong**

**Technical Report No. ACSP-TR-03-03-01**

**March 2003**



## Abstract

This paper investigates the placement of training symbols within data packets of a wireless system in which transmissions are subject to asynchronous interference. The minimum mean square error of the training-based channel estimator is expressed as a function of the Fisher information of the received signal. It is shown that the placement that minimizes the minimum mean square error should be searched for within a set containing half as many elements as the number of training symbols in the packet. Furthermore, a lower bound on the minimum mean square error is derived and analyzed. It is shown that this bound is tight when the power of the interference is high. The placement of the training symbols in two clusters of equal or quasi-equal length at the two edges of the data packet minimizes the lower bound for all values of the parameters, and thus gives the solution of the problem for high values of the interference power. The influence of the training symbols placement on the data transmission performance is also investigated.

## Index Terms

Fisher information, FIM, MMSE, Pilot Symbols, Placement Schemes.

**EDICS:** 3-CEQU (Channel modeling, estimation, and equalization), 3-PERF (Performance Analysis, Optimization, and Limits).

## I. INTRODUCTION

This paper considers the channel estimation problem in the presence of asynchronous packet interference. Asynchronous interference arises in ad-hoc networks, wireless LANs, and even in cellular networks where packet collisions can not be avoided and packet synchronization is not perfect. Moreover, the interference packets can have a wide range of power levels due to the near-far effect.

Channel estimation is crucial in coherent symbol detection, optimal scheduling and power allocation [1], [2], and in the design of the medium access control protocol in random access networks [3], [4]. Typically, channel estimation is performed by including a certain number of training symbols in the data packet. When the channel is memoryless, the placement of these training symbols does not affect the performance and is designed to simplify the receiver implementation. When the channel has memory, however, the placement of training symbols can affect the performance significantly. Optimal training placement for intersymbol interference channels has been considered in [5], [6], [7], [8], [9].

Asynchronous packet interference introduces a different kind of channel memory. The event that a symbol of a data packet is hit by an interfering packet affects the chance that its adjacent symbols are

also hit. The effect of packet interference on the training symbols, however, is somewhat subtle. If we assume that an interference packet arrives randomly, and its position relative to the packet of interest is uniformly distributed, then the *average number* of training symbols hit by the interference is the same regardless how training symbols are placed in the packet. However, the *distribution* of the number of training symbols that are hit by the interference is a function of the placement. It is the distribution of the number of training symbols survived the interference—not the average number—that determines the performance of the channel estimator.

We assume that the receiver uses the training-based minimum mean square error (MMSE) channel estimator, *i.e.*, only those observations corresponding to the training symbols are used in the estimation. If there are  $N$  training symbols in a packet of size  $B$ , the brute-force approach to finding optimal placement requires comparing  $\binom{B}{N}$  possible placements. The lack of a simple expression for the MMSE coupled with the enormous number of possibilities makes the brute-force approach unappealing. Also, it is unlikely that such an approach will lead to useful insights.

In searching for the optimal placement, we first obtain an expression of MMSE as a function of the Fisher information matrix (FIM) of the received signal. This crucial step allows us to exploit the convexity of the Fisher information functional and therefore reduce the number of searches from  $\binom{B}{N}$  to  $\lfloor N/2 \rfloor + 1$ , which depends only on the number of training symbols and not on the size of the packet. Furthermore, independent of the system parameters such as signal-to-noise ratio, the optimal placement belongs to a fixed set of placements with either one or two clusters.

The main difficulty involved in obtaining the optimal placement in closed-form comes from the nonlinearity of the MMSE estimator. One way to overcome this problem is to consider a lower bound given by the MSE of a linear estimator that knows the position of the interference with respect to the data packet. This estimator and its MSE are called the genie estimator and the genie bound respectively. The genie estimator can only be approximated by a detect-then-estimate scheme where the receiver first detects the presence of the interference. What we gain in considering the genie estimator is that the relation between its MSE and the training placement can be obtained explicitly. We show that the placement that minimizes the genie bound has two clusters of equal or quasi-equal length at the two edges of the data packet, which is in contrast to widely accepted single cluster placement (such as that in GSM) and the uniformly distributed periodic placement. We further show that the genie bound is tight when the interference power is high, which implies that if the interference level is high, the two equal sized clusters placed at the two ends of the packets is optimal. In general, we can only conjecture that this placement is optimal for all values of the parameters involved; this conjecture is supported by

simulations.

Existing work on optimal placement focuses mainly on channels where self interference is introduced by channel memory [6], [7]. In such cases and under different metrics, the optimal placements tend to be scattered. In particular, for ISI channels, the optimal placement of training symbols is the *quasi-periodic placement* [6] where the pilot symbols are placed periodically with the minimum cluster size. When the interference comes from asynchronous packets, our result points to a different placement strategy where the training clusters are placed at the two ends of the packet.

The paper is organized as follows: Section II contains the model and is followed by Section III with the description and analysis of the MMSE channel estimation. In III-B the behavior of the MMSE is analyzed using the FIM and in Section III-C the genie lower bound is introduced and analyzed. Section IV studies the effect of the placement on the data communication performance. In Section V the simulations and numerical results provide additional insight into the problem. We conclude the paper in Section VI. Some derivations are given in the Appendix.

Notations : the vectors are in bold fonts,  $\mathbb{E}_X$  is the expectation with respect to the random variable  $X$ ,  $\mathbb{P}\{A\}$  the probability of the event  $A$ ,  $\nabla_{\mathbf{a}}f(\mathbf{a})$  the gradient operator with respect to vector  $\mathbf{a}$ , and  $\text{diag}(\mathbf{A})$  is a column vector formed by the diagonal elements of the square matrix  $\mathbf{A}$ . We use  $\mathbf{1}_u$  to denote an  $u \times 1$  vector with all elements equal to 1 and  $\mathbf{1} \triangleq \mathbf{1}_N$ . Whenever necessary the function  $f(\cdot)$  is written as  $f(\cdot; a)$  to emphasize the dependence on the parameter  $a$ . Given the vector  $\mathbf{a}$ ,  $a_k$  is its  $k$ -th element and  $(f(x))_k$  is the  $k$ -th element of the vectorial function  $f(x)$ . If  $A, B$  are square matrices,  $A \geq B$  means that  $A - B$  is positive semidefinite. The complex Gaussian distribution with mean  $\boldsymbol{\mu}$  and covariance matrix  $\mathbf{C}$  is denoted by  $\mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$ , while  $\mathcal{CN}(\mathbf{x}; \boldsymbol{\mu}, \mathbf{C})$  is its probability density function. The set of strictly positive natural numbers is denoted by  $\mathbb{N}^*$ .

## II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider a point-to-point one-way communication link. The channel is assumed Rayleigh block-flat-fading, *i.e.*, the channel is constant during the transmission of one packet, and has independent realizations in different packets. The symbols of the data packets can be either training or data. The receiver obtains an estimate of the channel based only on the training symbols from the current packet received.

The communication is subject to the usual i.i.d. complex additive white Gaussian noise (CAWGN) with known variance  $\sigma^2$ . A data packet is also affected by interference, modeled as a packet of i.i.d. CAWGN with known variance  $\sigma_q^2$ . The interference affects a contiguous subset of the symbols of the packet depending on its position as described using Fig. 1.

The following assumptions about the interference are made: only one interference packet can hit the data packet at a time and the data and interference packets have the same length. The relative position of the data and interference packets is not known either by the transmitter or the receiver and is distributed uniformly.

Denote by  $B$  the total number of symbols of one packet, out of which  $N \geq 2$  are allocated to training. The symbols of one received packet are given by the  $B \times 1$  vector

$$\mathbf{y} = \mathbf{s}A + \mathbf{z}, \quad (1)$$

where  $A \sim \mathcal{CN}(0, 1)$  is the complex scalar channel parameter,  $\mathbf{s}$  a  $B \times 1$  vector representing one block of transmitted symbols, and  $\mathbf{z}$  the total noise vector that includes the CAWGN and the interference. The probability density function (pdf) of  $\mathbf{z}$  is a mixture of Gaussians. We assumed the variance of the channel equal to 1 without loss of generality.

From the assumption that the interference packets are of the same length  $B$  as the data packets, it follows that the relative position between the two packets can be described by a discrete random variable  $V$  distributed uniformly on  $\{1, \dots, 2B - 1\}$ . As shown in Fig. 1, if  $V \in \{1, \dots, B\}$  then the first  $V$  symbols of the data packet are hit. Similarly, if  $V \in \{B + 1, \dots, 2B - 1\}$  then the last  $2B - V$  symbols of the data packet are hit. The distribution of the total noise vector  $\mathbf{z}$  is obtained noting that, conditioned on  $\{V = v\}$ ,

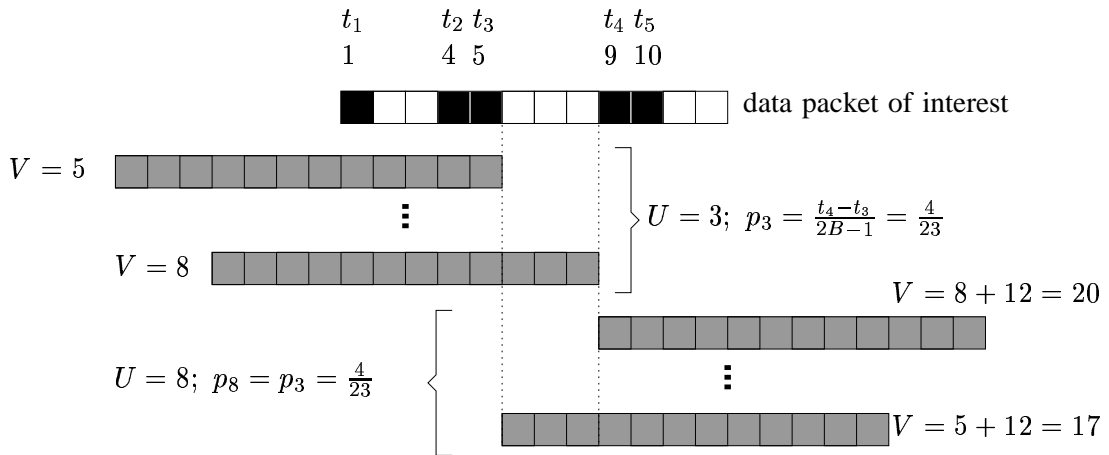


Fig. 1. Data packet and different possible positions of the interference packet. The training symbols are shaded in black. The interference packets are colored in gray.  $B = 12$ ,  $N = 5$ ,  $\mathcal{J} = \{1, 4, 5, 9, 10\}$

$\mathbf{z}|V = v$  is CAWGN with independent components

$$\begin{aligned} f(\mathbf{z}) &= \sum_{v=1}^{2B-1} \frac{1}{2B-1} f(\mathbf{z}|V = v), \\ f(\mathbf{z}|V = v) &= \frac{1}{\pi^N |\mathbf{G}_v|} \exp(-\mathbf{z}^H \mathbf{G}_v^{-1} \mathbf{z}). \end{aligned}$$

For any  $v \in \{1, \dots, 2B-1\}$  the elements of the diagonal matrices  $\mathbf{G}_v$  may take only two values,  $\sigma^2$  when the  $v$ -th symbol is interference-free, and  $\sigma_h^2 \triangleq \sigma^2 + \sigma_q^2$  when the  $v$ -th symbol is hit by the interference

$$\text{diag}(\mathbf{G}_v) = \begin{cases} [\sigma_h^2 \mathbf{1}_v^T & \sigma^2 \mathbf{1}_{B-v}^T]^T & \text{if } v \in \{1, \dots, B-1\} \\ \sigma_h^2 \mathbf{1}_B & \text{if } v = B \\ [\sigma^2 \mathbf{1}_{v-B}^T & \sigma_h^2 \mathbf{1}_{2B-v}^T]^T & \text{if } v \in \{B+1, \dots, 2B-1\} \end{cases}. \quad (2)$$

A channel estimate for each packet is obtained only from the received training symbols. Denote by  $\mathcal{J} = \{t_1, \dots, t_N\} \subset \{1, \dots, B\}$ ,  $1 \leq t_1 < t_2 < \dots < t_N \leq B$ , the ordered set of indexes of the training symbols within the packet. Using an  $N \times B$  selection matrix  $\mathbf{P}$  we extract these symbols into an  $N \times 1$  vector  $\mathbf{x}$

$$\mathbf{x} = \mathbf{P}\mathbf{y} = \mathbf{1}A + \mathbf{P}\mathbf{z} = \mathbf{1}A + \mathbf{n}, \quad (3)$$

where, the  $k$ -th row of  $\mathbf{P}$  has only one non-zero element in the column  $t_k$  that gives the position of the  $k$ -th training symbol within the data packet. All the training symbols are chosen equal to 1.

Next, we characterize the placement of training symbols through the probability mass function of the relative position of the interference and training symbols. Consider the random variable  $U \in \{0, 1, \dots, 2N-1\}$  that gives the position of the interference packet with respect to the *training symbols*. Some possible values are shown in Fig. 1. Similar to  $V$ , if  $U \in \{1, \dots, N\}$  the first  $U$  training symbols are hit. If  $U \in \{N+1, \dots, 2N-1\}$  the last  $2N-U$  training symbols are hit. If  $U = 0$  no training symbol is hit by the interference. The random variable  $U$  is not distributed uniformly anymore. Specifically,  $U$  is determined by  $V$ , and its distribution can be obtained from the placement of the training symbols by taking into account the uniform distribution of  $V$ . Defining  $\varepsilon \triangleq \frac{1}{2B-1}$ ,  $\mathcal{S}_{N-1} \triangleq \{1, \dots, N-1\}$  and  $\mathcal{S}_{N+1} \triangleq \{N+1, \dots, 2N-1\}$ , we obtain the distribution  $p_u \triangleq \mathbb{P}\{U = u\}$  of  $U$  as

$$p_u = \begin{cases} \varepsilon(t_{u+1} - t_u) & \text{if } u \in \mathcal{S}_{N-1} \\ \varepsilon(B - (t_N - t_1)) & \text{if } u = N \\ p_{u-N} & \text{if } u \in \mathcal{S}_{N+1} \\ p_N - \varepsilon & \text{if } u = 0 \end{cases}. \quad (4)$$

We use  $\mathbf{p} \triangleq [p_0, p_1, \dots, p_{2N-1}]^T$  for the distribution of  $U$ . Denote by  $\mathcal{P}$  the set of all distributions  $\mathbf{p}$  that satisfy the conditions (4), *i.e.*,

$$\mathcal{P} \triangleq \left\{ \mathbf{p} \in \mathbb{R}^{2N} \mid \sum_{u=0}^{2N-1} p_u = 1; p_u = p_{u+N}, u = 1, \dots, N-1; p_0 = p_N - \varepsilon; \frac{p_u}{\varepsilon} \in \mathbb{N}^* \forall u \in \{1, \dots, N\} \right\}. \quad (5)$$

From equation (4) it follows that the placement of the training symbols (the set  $\mathcal{J}$ ) is determined uniquely by  $\mathbf{p}$  if and only if  $p_0 = 0$ . In this case  $t_1 = 1$  and  $t_N = B$ , *i.e.*, the packet starts and ends with a training symbol. However, it will be shown that the MMSE channel estimator and the MMSE are the same for all placements with the same  $\mathbf{p}$ . This is why in the channel estimation part of the paper we'll refer to placements through their corresponding  $\mathbf{p}$ .

For the placement given in Fig. 1, we have  $\mathcal{J} = \{1, 4, 5, 9, 10\}$  and  $\mathbf{p} = \frac{1}{23}[2, 3, 1, 4, 1, 3, 3, 1, 4, 1]^T$ . Frequently referred in the paper is the placement of training symbols in two clusters of equal (or nearly equal) length at the edges of the packet. This placement is given by

$$\bar{\mathcal{J}} = \left\{ 1, \dots, \left\lfloor \frac{N}{2} \right\rfloor, B - \left\lfloor \frac{N}{2} \right\rfloor + 1, \dots, B \right\}, \quad (6)$$

and the corresponding distribution is

$$\bar{p}_u = \begin{cases} \varepsilon & \text{if } u \notin \left\{ \left\lfloor \frac{N}{2} \right\rfloor, N + \left\lfloor \frac{N}{2} \right\rfloor, 0 \right\} \\ (B - N + 1)\varepsilon & \text{if } u \in \left\{ \left\lfloor \frac{N}{2} \right\rfloor, N + \left\lfloor \frac{N}{2} \right\rfloor \right\} \\ 0 & \text{if } u = 0 \end{cases}. \quad (7)$$

We now complete the channel model given in (3) by deriving the pdf of the received signal. Given the placement, or equivalently  $\mathbf{p}$ , the pdf of  $\mathbf{n}$  is

$$f(\mathbf{n}) = \sum_{u=0}^{2N-1} p_u f(\mathbf{n} | U = u),$$

$$f(\mathbf{n} | U = u) = \frac{1}{\pi^N |\mathbf{D}_u|} \exp(-\mathbf{n}^H \mathbf{D}_u^{-1} \mathbf{n}),$$

where the matrices  $\mathbf{D}_u$  are defined as

$$\text{diag}(\mathbf{D}_u) = \begin{cases} [\sigma_h^2 \mathbf{1}_u^T & \sigma^2 \mathbf{1}_{N-u}^T]^T & \text{if } u \in \mathcal{S}_{N-1} \\ \sigma_h^2 \mathbf{1}_N & \text{if } u = N \\ [\sigma^2 \mathbf{1}_{u-N}^T & \sigma_h^2 \mathbf{1}_{2N-u}^T]^T & \text{if } u \in \mathcal{S}_{N+1} \\ \sigma^2 \mathbf{1}_N & \text{if } u = 0 \end{cases}. \quad (8)$$

Finally, the pdf of  $\mathbf{x}$ , the vector of received training symbols, is

$$f(\mathbf{x}) = \sum_{u=0}^{2N-1} p_u f(\mathbf{x}|U=u) \quad (9)$$

$$f(\mathbf{x}|U=u) = \frac{1}{\pi^N |\mathbf{C}_u|} \exp(-\mathbf{x}^H \mathbf{C}_u^{-1} \mathbf{x}), \quad (10)$$

where

$$\mathbf{C}_u \triangleq \mathbf{1}\mathbf{1}^T + \mathbf{D}_u. \quad (11)$$

The parameters of the system are  $\{\mathcal{J}, B, N, \sigma^2, \sigma_q^2\}$ . It was specified the use of  $\mathbf{p}$  instead of  $\mathcal{J}$  for convenience. Channel estimation depends on all the system parameters, but the dependence will not be always expressed explicitly.

### III. CHANNEL ESTIMATION

#### A. The Bayesian MMSE Estimator

The Bayesian MMSE estimator of the channel parameter  $A$  is given by

$$\begin{aligned} \hat{A}(\mathbf{x}; \mathbf{p}) &= \mathbb{E}\{A|\mathbf{x}\} = \mathbb{E}_U \mathbb{E}_A \{A|\mathbf{x}, U\} \\ &= \mathbb{E}_U \{\mathbf{1}^T \mathbf{C}_U^{-1} \mathbf{x} | \mathbf{x}, U\} \\ &= \mathbf{1}^T \mathbb{E}_{U|\mathbf{x}} \{\mathbf{C}_U^{-1}\} \mathbf{x}. \end{aligned} \quad (12)$$

Conditioned on the position of the interference  $U$ , we have the well-known Gaussian model; relation (12) follows from the conditional expectation in Gaussian models, see *e.g.*, [10], p.326. Writing the expectation explicitly, we have

$$\hat{A}(\mathbf{x}; \mathbf{p}) = \mathbf{1}^T \left( \sum_{u=0}^{2N-1} \frac{p_u f(\mathbf{x}|U=u)}{f(\mathbf{x})} \mathbf{C}_u^{-1} \right) \mathbf{x}. \quad (13)$$

The performance of the Bayesian MMSE estimator is given by the MMSE

$$\mathcal{E}(\mathbf{p}) = \mathbb{E}[|A|^2] - \mathbb{E}[|\hat{A}(\mathbf{x}; \mathbf{p})|^2]. \quad (14)$$

The MMSE estimator and the MMSE depend on the distribution  $\mathbf{p}$  of the random variable  $U$ . Our goals are to characterize this dependence and to find the placement(s)  $\mathbf{p}_0$  that minimizes the MMSE (14) under the conditions (5) imposed by the physical model, *i.e.*, to find

$$\arg \min_{\mathbf{p} \in \mathcal{P}} \{\mathcal{E}(\mathbf{p})\}. \quad (15)$$

In general, the set of solutions of the problem above depends on the choice of the system parameters  $\{B, N, \sigma^2, \sigma_q^2\}$ .



Given a placement  $\mathcal{J}$  and its distribution  $\mathbf{p}$ , define its mirror reflection by  $\mathcal{J}^{\leftarrow} \triangleq \{N+1-t_N, \dots, N+1-t_1\}$  and the corresponding  $\mathbf{p}^{\leftarrow}$ . Note that if the placement is symmetric we have  $\mathbf{p} = \mathbf{p}^{\leftarrow}$ . Because of the left-right symmetry of the model, the mirror reflection  $\mathbf{p}^{\leftarrow}$  has the same MMSE as  $\mathbf{p}$ . Thus if  $\mathbf{p}_0$  is a solution of (15), so is  $\mathbf{p}_0^{\leftarrow}$ .

In our case the MMSE estimator and its performance are non-linear functions of  $\mathbf{p}$ . This makes their analysis a hard problem.

### B. MMSE and Optimal Placement

In this subsection we search for the training placement that minimizes the MMSE. Towards this goal, we first establish the connection between the MMSE (14) and the Fisher information matrix of the received signal and then show that the optimal placement has either two clusters placed at the two ends of the packet, or one cluster; thus it should be searched for within a certain subset with  $\lfloor \frac{N}{2} \rfloor + 1$  elements.

For a random complex vector  $\mathbf{X}$  with probability density function (pdf)  $f$ , the FIM is defined as

$$\mathbf{J}(f) = \mathbb{E} \left\{ \nabla_{\mathbf{x}^*} \{\log f(\mathbf{x})\} (\nabla_{\mathbf{x}^*} \{\log f(\mathbf{x})\})^H \right\}. \quad (16)$$

Some regularity conditions on  $f$  are necessary for the FIM to exist, see [10] for details. These conditions are satisfied by the distributions considered in our problem.

*Lemma 1:* The MMSE (14) can be written as

$$\mathcal{E}(\mathbf{p}) = 1 - \mathbf{1}^T \mathbf{J}(f(\cdot; \mathbf{p})) \mathbf{1}. \quad (17)$$

The pdf  $f(\cdot; \mathbf{p})$  is given by (9); here we indicated the dependence on  $\mathbf{p}$  explicitly.

**Proof** From the properties of Gaussian densities we have

$$\nabla_{\mathbf{x}^*} f(\mathbf{x}|U = u) = (-1) f(\mathbf{x}|U = u) \mathbf{C}_u^{-1} \mathbf{x}. \quad (18)$$

Using this, the expression (13) of  $\hat{A}(\mathbf{x}; \mathbf{p})$  becomes

$$\hat{A}(\mathbf{x}; \mathbf{p}) = -\frac{1}{f(\mathbf{x}; \mathbf{p})} \mathbf{1}^T \sum_{u=0}^{2N-1} p_u \nabla_{\mathbf{x}^*} f(\mathbf{x}|U = u). \quad (19)$$

Further,

$$|\hat{A}(\mathbf{x}; \mathbf{p})|^2 = \frac{1}{(f(\mathbf{x}; \mathbf{p}))^2} \mathbf{1}^T \left( \sum_{u=0}^{2N-1} p_u \nabla_{\mathbf{x}^*} f(\mathbf{x}|U = u) \sum_{u=0}^{2N-1} p_u (\nabla_{\mathbf{x}^*} f(\mathbf{x}|U = u))^H \right) \mathbf{1}, \quad (20)$$

and  $\mathbb{E} |\hat{A}(\mathbf{x}; \mathbf{p})|^2 = \mathbf{1}^T \mathbf{J}(f(\cdot; \mathbf{p})) \mathbf{1}$ . Using this in (14) and taking into account that  $A \sim \mathcal{CN}(0, 1)$ , the lemma follows. ■

This lemma allows us to use the convexity property of the FIM functional. This property is stated in the next lemma, which is an extension to complex vectors of a weaker form found in [11],[12].

*Lemma 2:* The Fisher information matrix of a random complex vector  $\mathbf{X}$  with pdf  $f$  is a convex functional of  $f$ . Specifically, for  $\mathbf{X}_1$  and  $\mathbf{X}_2$  random vectors with densities  $f_1$  and  $f_2$  respectively, and  $a \in [0, 1]$  an arbitrary number, the following inequality holds :

$$a\mathbf{J}(f_1) + (1 - a)\mathbf{J}(f_2) \geq \mathbf{J}(af_1 + (1 - a)f_2). \quad (21)$$

**Proof** See [13]. ■

The convexity property given above allows us to reduce the number of possible solutions of the optimization problem (15). To use this property we rewrite the pdf of the received signal  $f(\cdot, \mathbf{p})$  given in (9) as a mixture of densities

$$f(\mathbf{x}; \mathbf{p}) = \sum_{u=0}^{N-1} 2p_u \left( \frac{1}{2}f(\mathbf{x}|U = u) + \frac{1}{2}f(\mathbf{x}|U = N + u) \right) + \varepsilon f(\mathbf{x}|U = N). \quad (22)$$

The coefficients of this mixture are  $\{2p_0, 2p_1, \dots, 2p_{N-1}, \varepsilon\}$ . One can observe that the maximum of  $\mathbf{1}^T \mathbf{J}(f(\cdot; \mathbf{p})) \mathbf{1}$  is realized when one of the coefficients from the set given above takes its maximum possible value, and all the others take their minimum value. This result is stated in the next theorem.

*Theorem 1:* The placement that minimizes the MMSE (14) belongs to the set  $\mathcal{P}_0$  that contains the placements with either two clusters placed at the two ends of the packet, or a single cluster. Formally, from (4),  $\mathcal{P}_0$  is defined as

$$\mathcal{P}_0 \triangleq \{ \mathbf{p} \in \mathcal{P} \mid \exists m \in \{1, \dots, N\} : p_m = (B - N + 1)\varepsilon; p_u = \varepsilon, \forall u = 1, \dots, N, u \neq m \}, \quad (23)$$

and we have

$$\arg \min_{\mathbf{p} \in \mathcal{P}} \{ \mathcal{E}(\mathbf{p}) \} \subset \mathcal{P}_0.$$

**Proof** See the appendix. ■

The size of  $\mathcal{P}_0$  depends only on  $N$  and increases linearly with it. Even if the distributions vectors  $\mathbf{p}$  of  $\mathcal{P}_0$  depend on  $B$  and  $N$ , the physical structure of the corresponding placements does not. Although the set  $\mathcal{P}_0$  has  $N$  elements, since the mirror pairs have the same MSE, the optimal solution should be searched for among only  $\lfloor \frac{N}{2} \rfloor + 1$  placements.

### C. A Genie Lower Bound on the MMSE

A lower bound on the MMSE can be obtained by considering the performance of a receiver helped by a genie who provides the current value of  $U$ , *i.e.*, the position of the interference packet with respect to the training symbols. For each value  $U = u$  we have a Gaussian model, for which the MMSE estimator and its MSE are well known. Consider the following estimator, that assumes the random variable  $U$  known:

$$\tilde{A}(\mathbf{x}, U = u) \triangleq \mathbb{E}\{A|\mathbf{x}, U = u\} = \mathbf{1}^T \mathbf{C}_u^{-1} \mathbf{x}. \quad (24)$$

Its MSE for each  $U = u$  is given by  $m(u)$  and the averaged MSE by  $\xi(\mathbf{p})$

$$\begin{aligned} m(u) &\triangleq \mathbb{E}\{|\tilde{A}(\mathbf{x}, U) - A|^2|U = u\} = 1 - \mathbf{1}^T \mathbf{C}_u^{-1} \mathbf{1} \\ \xi(\mathbf{p}) &\triangleq \mathbb{E}\{|\tilde{A}(\mathbf{x}, U) - A|^2\} = \sum_{u=0}^{2N-1} p_u m(u). \end{aligned} \quad (25)$$

Given the state  $U = u$ ,  $\tilde{A}(\mathbf{x}, U = u)$  is the MMSE estimator. Thus, for any other estimator  $\hat{A}_0(\mathbf{x}, U)$  we have

$$m(u) \leq \mathbb{E}\{|\hat{A}_0(\mathbf{x}, U) - A|^2|U = u\}.$$

The estimators  $\hat{A}_0(\mathbf{x}, U)$  include those ones that do not use  $U$ , so the relation above applies to the true MMSE estimator  $\hat{A}(\mathbf{x}; \mathbf{p})$  (13) for each  $u$ . Thus  $\xi(\mathbf{p})$  is a lower bound (the genie lower bound) for the MMSE

$$\xi(\mathbf{p}) \leq \mathbb{E}\{|\hat{A}(\mathbf{x}; \mathbf{p}) - A|^2\}.$$

One can observe that the genie bound is the bound obtained by applying the convexity of the FIM, *i.e.*, Lemma 2, to the pdf (9) of the received signal expressed as a mixture of Gaussians.

The next theorem shows that the genie lower bound is tight when the power of the interference is high for any placement  $\mathbf{p}$ .

*Theorem 2:* Let  $\xi(\mathbf{p}; \sigma_q^2)$  and  $\mathcal{E}(\mathbf{p}; \sigma_q^2)$  be the genie bound and the MMSE respectively, where the dependence on the power of the interference  $\sigma_q^2$  has been shown explicitly. For any choice of  $(B, N, \sigma^2)$  we have

$$\lim_{\sigma_q^2 \rightarrow \infty} (\mathcal{E}(\mathbf{p}; \sigma_q^2) - \xi(\mathbf{p}; \sigma_q^2)) = 0. \quad (26)$$

**Proof** See the appendix. ■

The genie lower bound can be optimized with respect to  $\mathbf{p}$ . The result is given below.

*Theorem 3:* Placing the training symbols in two equal or quasi-equal clusters at the two ends of the data packet minimizes the genie lower bound for any set of values of the parameters involved. Specifically, using the probability distribution  $\bar{\mathbf{p}}$  given by

$$\bar{p}_u = \begin{cases} \varepsilon & \text{if } u \notin \{\lfloor \frac{N}{2} \rfloor, N + \lfloor \frac{N}{2} \rfloor, 0\} \\ (B - N + 1)\varepsilon & \text{if } u \in \{\lfloor \frac{N}{2} \rfloor, N + \lfloor \frac{N}{2} \rfloor\} \\ 0 & \text{if } u = 0 \end{cases}, \quad (27)$$

$\bar{\mathbf{p}}$  and  $\bar{\mathbf{p}}^{\leftarrow}$  are the only distributions that minimize the genie lower bound  $\xi(\mathbf{p})$  subject to the conditions (5):

$$\{\bar{\mathbf{p}}, \bar{\mathbf{p}}^{\leftarrow}\} = \arg \min_{\mathbf{p} \in \mathcal{P}} \{\xi(\mathbf{p})\}. \quad (28)$$

Note that  $\bar{\mathbf{p}} = \bar{\mathbf{p}}^{\leftarrow}$  if  $N$  is even.

**Proof** Taking into account the definitions (11, 8) of  $\mathbf{C}_u$  and  $\mathbf{D}_u$ , for  $u = 0, \dots, N$ , we have

$$m(u) = \frac{1}{1 + \sum_{k=1}^N \frac{1}{D_u(k,k)}} = \frac{1}{1 + \frac{u}{\sigma^2 + \sigma_q^2} + \frac{N-u}{\sigma^2}}.$$

Replacing  $u$  with a continuous variable  $x$ , for  $x \in [0, N]$ ,  $m(x)$  is a strictly convex function of  $x$ . Thus the function  $m(x) + m(N - x)$  is strictly decreasing on  $[0, \frac{N}{2}]$ . Since the MSE of the genie estimator is the same if the interference hits the first  $u$  symbols or the last  $u$ , for  $u = 1, \dots, 2N - 1$  we have  $m(u) = m(2N - u)$ . Using this and the symmetry property (4) of  $\mathbf{p}$ ,  $\xi(\mathbf{p})$  can be written

$$\xi(\mathbf{p}) = \sum_{u=1}^N p_u (m(u) + m(N - u)) - \frac{1}{2B - 1} m(0).$$

It follows that under the conditions (4),  $\bar{\mathbf{p}}$  given by (27) minimizes  $\xi(\mathbf{p})$ . ■

From Theorem 2 it follows that increasing  $\sigma_q^2$  we can make the MMSE be as close to the genie bound as wanted. Because the function  $m(x)$  preserves its strict convexity when  $\sigma_q^2 \rightarrow \infty$  and there is a finite number of possible placements, it follows that even in this case the placements  $\bar{\mathbf{p}}$  given by (27) and  $\bar{\mathbf{p}}^{\leftarrow}$  are the only solutions of the optimization problem stated in Theorem 3. We have the following corollary.

*Corollary 1:* For any values of the system parameters  $(B, N, \sigma^2)$ , there is a level of interference  $\bar{\sigma}_q^2(B, N, \sigma^2)$  such that the placements  $\bar{\mathbf{p}}$  and  $\bar{\mathbf{p}}^{\leftarrow}$  are the only placements that minimize the MMSE (14) (i.e., solutions of the general problem (15)) for all  $\sigma_q^2 > \bar{\sigma}_q^2(B, N, \sigma^2)$ . ■

*Remark 1:* Using the properties of the genie bound given in the proof of Theorem 3, the placement that maximizes the genie bound  $\xi(\mathbf{p})$  (i.e., the worst) can be found as well, and it corresponds to placing

all the training symbols into one cluster. Using Theorem 2 we can conclude that placing the training symbols in one cluster provides the worst performance at high values of the interference power. ■

Notice that the solution  $\bar{\mathbf{p}}$  given by (27) obtained using the genie bound corresponds to the placement in  $\mathcal{P}_0$  of theorem (1) for which  $p_{\lfloor \frac{N}{2} \rfloor} = (B - N + 1)\varepsilon$ . Unfortunately, we were not able to show that  $\bar{\mathbf{p}}$  is the solution of the problem for any choice of the parameters. Besides the genie bound solution and the asymptotic solution, the simulations suggest the following conjecture.

*Conjecture 1:* The placement that minimizes the MMSE (14) is in two equal or quasi-equal clusters at the edges of the packet, *i.e.*,

$$\{\bar{\mathbf{p}}, \bar{\mathbf{p}}^{\leftarrow}\} = \arg \min_{\mathbf{p} \in \mathcal{P}} \{\mathcal{E}(\mathbf{p})\}.$$

■

#### IV. DATA COMMUNICATION UNDER ASYNCHRONOUS INTERFERENCE

In this section we investigate the influence of the placement of training symbols on the data transmission performance. Using the setup of Section II, we consider specific data transmission and reception procedures and evaluate the achievable rate using the mutual information between the input and the output.

We consider that the transmitter uses BPSK signaling with coding and interleaving. The data stream is partitioned in  $B - N$  substreams at the transmitter and each substream is coded independently of the others. The binary symbols of the codewords of one substream are inserted in successive data packets in the same position ( $k$ -th). Thus each data packet contains one binary symbol from a codeword of each substream. The receiver decodes each substream independently of the others. The communication channel under these constraints is equivalent to  $B - N$  parallel independent channels. We name “channel  $k$ ” the channel corresponding to the data symbols in the  $k$ -th position ( $s_k$ ).

Before decoding, the receiver recovers each data symbol coherently using the channel estimate. To make the model tractable, we assume that the receiver uses the genie estimator instead of the true MMSE estimator. In general, the results obtained will be upper bounds to the performance of the system, but if the power of the interference is high, the use of genie and MMSE estimates provides identical results.

Two different receiver scenarios are considered. First, we consider a receiver that performs individual hard reception, and then decoding. The second case is a receiver that decodes coherently the received symbols without any intermediate processing.

### A. Achievable Rate with Hard Reception

In this subsection we calculate the achievable rate of the system under the constraints given and assuming a receiver that first obtains a hard estimate of each symbol and then does decoding. When the interference power level is high, the data symbols that are hit are received with a very low SNR. If the receiver tries to detect them (binary), the probability of error will be close to one half. Also, if all training symbols of one packet are hit by the interference, then the probability of binary detection error is close to one half for all data symbols of that packet. These two situations will be defined as erasure events.

From the transmission model (1), the  $k$ -th received symbol (assume  $k \notin \mathcal{J}$ , i.e., data symbol) is given by

$$y_k = s_k A + z_k. \quad (29)$$

Introduce the binary random variable  $H_k$ ;  $\{H_k = 1\}$  is the event that the  $k$ -th received symbol is hit by interference and  $\{H_k = 0\}$  its complement. Conditioned on  $H_k$ ,  $z_k$  is Gaussian;  $z_k | H_k \sim \mathcal{CN}(0, \sigma^2 + H_k \sigma_q^2)$ . When the power of the interference is high, the erasure events can be detected accurately. Equivalently, we assume that besides  $U$ , the detector knows  $H_k$  for each  $k$ . Assuming equiprobable input symbols, the ML detector is

$$\tilde{s}_k \triangleq \begin{cases} \text{erasure} & \text{if } H_k = 1 \text{ or } U = N \\ \text{sgn}\{\mathcal{R}e\{\tilde{A}^* y_k\}\} & \text{otherwise} \end{cases}. \quad (30)$$

If the decoder does not use the information provided by the channel estimation part anymore, the achievable rate for channel  $k$  is given by the capacity of the binary channel with erasures

$$\underline{C}_k = (1 - P_{k,r})(1 + P_{k,e} \log_2(P_{k,e}) + (1 - P_{k,e}) \log_2(1 - P_{k,e})),$$

where the probability of erasure  $P_{k,r}$  and the conditional probability of error  $P_{k,e}$  are respectively given by

$$\begin{aligned} P_{k,r} &= \mathbb{P}\{\{H_k = 1\} \cup \{U = N\}\} \\ &= \begin{cases} \frac{B+(t_1-k)}{2B-1} & \text{if } t_1 > 1 \text{ and } 1 \leq k < t_1 \\ \frac{B+(k-t_N)}{2B-1} & \text{if } t_N < B \text{ and } t_N < k \leq B \\ \frac{B}{2B-1} & \text{otherwise} \end{cases} \\ P_{k,e} &= \mathbb{P}\{\tilde{s}_k \neq s_k | (\{H_k = 1\} \cup \{U = N\})^C\}. \end{aligned}$$

The probabilities  $P_{k,e}$  are derived in the appendix. The achievable rate (lower bound on capacity) of the system considered is given by the average

$$\underline{C} = \frac{1}{B} \sum_{k \notin \mathcal{J}} \underline{C}_k. \quad (31)$$

The placement that optimizes this lower bound cannot be found in closed form. Moreover, the numerical evaluations indicate that the problem has different solutions for different choices of the parameters  $(B, N, \sigma^2)$ . Note that if the noise has moderately low variance, at high interference power, the variation of the probabilities of error  $\mathbb{P}\{\tilde{s}_k \neq s_k | H_k = 0, U = u\}$  with  $u$  is small and the placement influence on the achievable rate is mainly through the probabilities of erasure  $P_{k,r}$ . For example, if the training symbols are placed in one cluster, then the data symbols from the first and last data clusters have high probability of erasure. This can lower significantly the performance of the system, as shown and discussed further in Section V.

### B. Reliable Rate with Soft Reception

The lower bound on capacity given before is valid only when the interference power is high. In order to obtain a performance measure for any value of the interference power, we consider a receiver that decodes coherently the received signal using the genie estimate but without performing hard detection first. An upper bound on the achievable rate of the transmission system under the constraints imposed in this section is calculated as the average of the capacities of the  $B - N$  parallel independent channels.

Denote by  $q_0$  the distribution of the data symbols  $s_k$ . The upper bound on the capacity of channel  $k$  is given by

$$\begin{aligned} \bar{C}_k &= \max_{q_0} \left\{ I(s_k; y_k, U, \tilde{A}(\mathbf{x}, U)) \right\} \\ &= \max_{q_0} \left\{ I(s_k; U, \tilde{A}(\mathbf{x}, U)) + I(s_k; y_k | U, \tilde{A}(\mathbf{x}, U)) \right\} \\ &= \max_{q_0} \left\{ I(s_k; y_k | U, \tilde{A}(\mathbf{x}, U)) \right\} \end{aligned} \quad (32)$$

$$\begin{aligned} &= \max_{q_0} \mathbb{E}_{U, \tilde{A}(\mathbf{x}, U)} \left\{ I(q_0(s_k); f(y_k | s_k, U, \tilde{A}(\mathbf{x}, U))) \right\} \\ &= \mathbb{E}_{U, \tilde{A}(\mathbf{x}, U)} \left\{ I(q_0(s_k); f(y_k | s_k, U, \tilde{A}(\mathbf{x}, U))) \Big|_{q_0 = \{\frac{1}{2}, \frac{1}{2}\}} \right\}. \end{aligned} \quad (33)$$

Relation (32) follows from the independence between  $s_k$  and the random variables  $U$  and  $\tilde{A}(\mathbf{x}, U)$ . Using the symmetry properties of the conditional distribution of the interference  $z_k$  and of the transmitted signal constellation, one can prove that the input distribution that achieves the maximum of  $I(q_0(s_k); f(y_k | s_k, U, \tilde{A}(\mathbf{x}, U)))$  is the uniform one. Relation (33) holds because this distribution is the

same for all  $U$  and  $\tilde{A}(\mathbf{x}, U)$ . The distribution  $f(y_k|s_k, U, \tilde{A}(\mathbf{x}, U))$  is given in the appendix, (56). So the upper bound on the capacity of channel  $k$  is obtained by averaging the capacities of the channels given by each  $U$  and  $\tilde{A}(\mathbf{x}, U)$ . The upper bound on the achievable rate is

$$\bar{C} = \frac{1}{B} \sum_{k \notin \mathcal{J}} \bar{C}_k. \quad (34)$$

Because of the mixture of distributions, this upper bound can be calculated only numerically; some results and further comments are provided in the next section.

## V. SIMULATIONS AND NUMERICAL RESULTS

The simulations were done for the following parameters :  $B = 80$ ,  $N = 6$ ,  $\sigma^2 = -16.9dB$ ,  $\sigma_A^2 = 1$ . The MMSE is evaluated from equation (14).

For the numerical values considered, taking into account Theorem 1 and the remark that follows, the optimal scheme should be searched for within a set with 4 elements. The corresponding placements are the first four placements in Fig. 2.

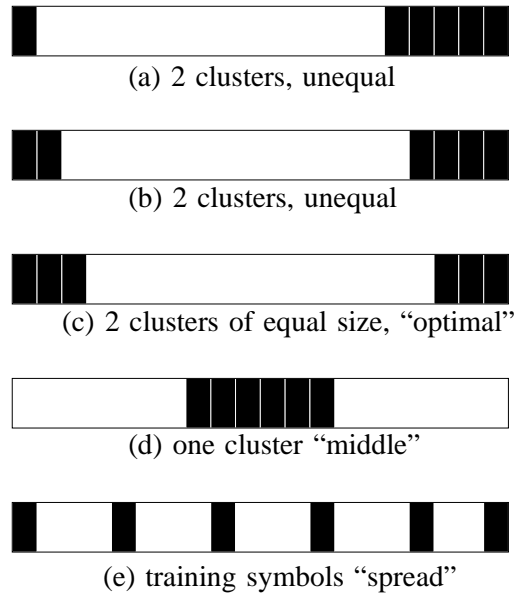


Fig. 2. The training schemes compared in Figs. 3 and 4

Fig. 3 shows that we can gain more than  $10dB$  by using the "optimal" placement over the placement that uses one cluster (in the middle of the packet). This is the maximum gain that can be obtained; according to the genie bound the "middle" placement offers the worst performance. Once we have two clusters placed at the edges of the packet, the gain obtained by using the "optimal" placement is smaller,



up to 2dB. The figure illustrates well the behavior of the MMSE in three regimes. When the power of the interference is small, all training schemes have the same MMSE, as expected. When  $\sigma_q^2$  is large, then the MMSE of the three schemes considered is as predicted by the corollary. In this regime the large gap between the “middle” placement and the others can be easily explained by observing that for reasonable small values of  $\sigma^2$ , the performance of the estimator is significantly degraded if all the training symbols are hit, while if only one of them is not hit the estimation can be done reasonably well. One may observe as well that the variation of the MMSE with the power of the interference is not monotonic. This effect is somehow reduced for the situation of this figure, and will be explained next.

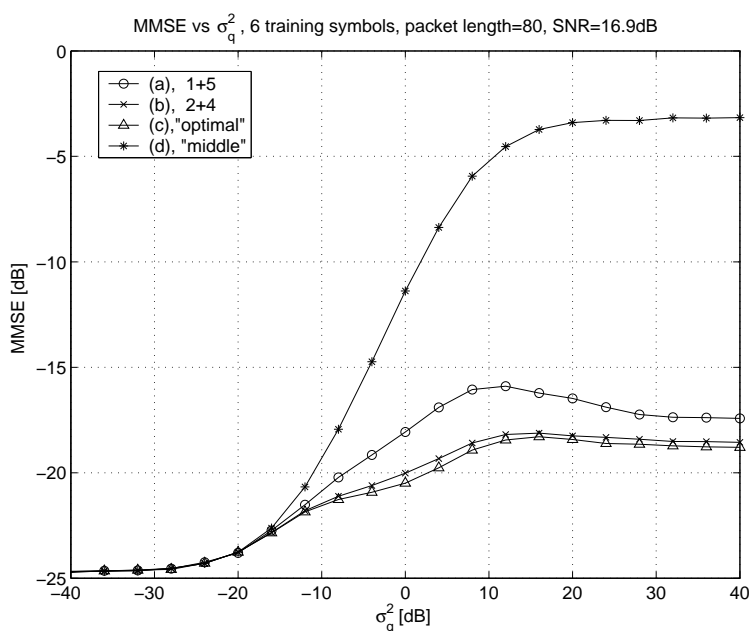


Fig. 3. The MMSE of the first four training schemes from Fig. 2. The legend shows the number of training symbols in each cluster.

In Fig. 4 we compared the MMSE with the genie bound for three of the placements represented in Fig. 2, *i.e.*, “optimal”, “middle” and “spread” schemes. As predicted by Theorem 2, the MMSE converges to the genie bound for  $\sigma_q^2$  large enough. For the “optimal” and the “middle” placements, the genie bounds are relatively tight for all  $\sigma_q^2$ . The interesting fact is that the MMSE of the “spread” placement scheme has a bell shape and the genie bound is not tight. The gap to the MMSE of the “optimal” placement can be predicted using the convexity property of the FIM. Alternatively, it can be explained by thinking that the coefficients  $f(U = u|\mathbf{x})$  in the expression (13) of the MMSE estimator act like an embedded maximum *a posteriori* (MAP) soft detector. The detection can be done better if there are fewer events with high

*a priori* probabilities. This happens if the symbols are grouped into two clusters placed at the edges or in one cluster; in these cases the MMSE is close to the genie bound. The previous argument works very well when the interference power has moderately high values. However, when the interference is weak or absent its position can't be detected, but the detection is unnecessary and all the placements provide the same performance, as one would expect, while at high values of the interference power the detection can be done accurately for any placement.

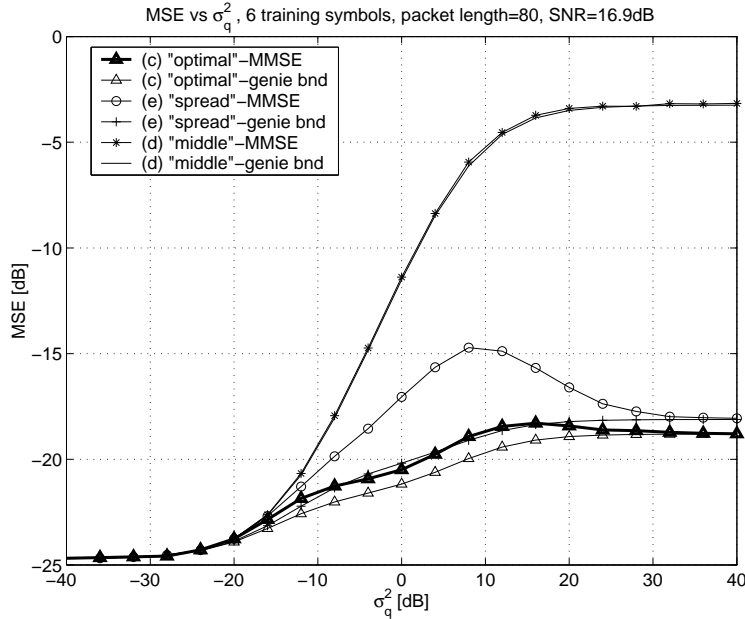


Fig. 4. The MMSE of the “optimal”, “middle” and “spread” placements and their genie bounds

In Fig. 5 the MMSE of three placement schemes function of the length of the packet length  $B$  is shown. The interference power  $\sigma_q^2$  was chosen close to the value that provides the worst performance. The figure also contains the variation of the genie bounds associated. If the packet has only training symbols, then the placement does not influence the MMSE. The variation of the MMSE and the genie bound can be explained by observing that when the packet length  $B$  is large, the probability that all the training symbols are hit simultaneously becomes smaller and smaller. Also, it can be noted that the genie bound can't be achieved by increasing  $B$ . This is expected, since for large values of  $B$  the distribution of the received signal is close to a mixture of only two distributions corresponding to the two dominant events, *i.e.*, the interference hits the left or the right cluster. The gap to the genie bound is explained by the convexity of the FIM functional.

Some numerical results for the lower bound  $\underline{C}$  given by (31) on the achievable rate of the binary

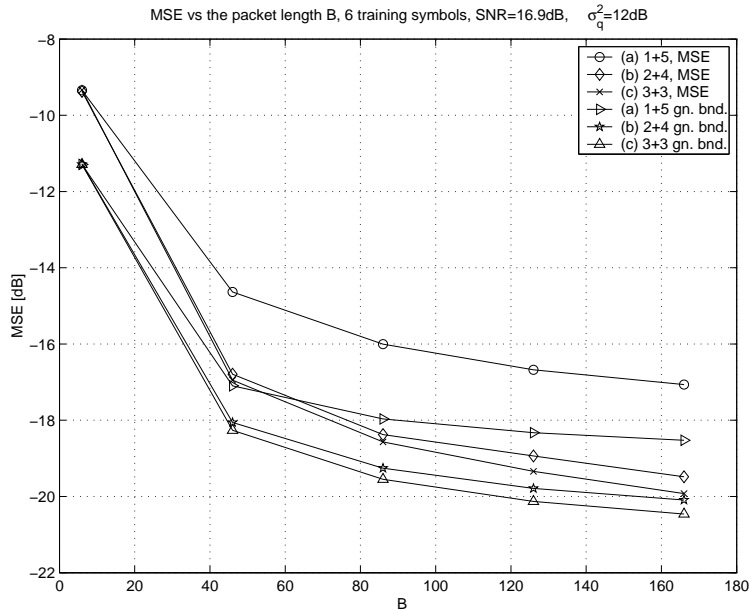


Fig. 5. The performance of 3 placement schemes and their genie bounds when the packet length varies.  $N = 6$ ,  $\sigma^2 = -16.9dB$ ,  $\sigma_A^2 = 1$ ,  $\sigma_q^2 = 12dB$ .

channel with errors and erasures when  $\sigma_q^2 \rightarrow \infty$  are given in Table I. First, it can be observed that  $\underline{C}$  can

Placement - $\mathcal{J}$	$\underline{C}[bps]$
$\bar{\mathcal{J}} = \{1, 2, 3, 78, 79, 80\}$	0.4329
$\mathcal{J}_{middle} = \{38, \dots, 43\}$	0.3313
$\mathcal{J}_{one\ edge} = \{1, \dots, 6\}$	0.2282
$\mathcal{J}_{max} = \{1, 9, 28, 53, 72, 80\}$	0.4334

TABLE I

THE LOWER BOUND ON THE CHANNEL CAPACITY OF THE BINARY CHANNEL WITH ERRORS AND ERASURES, ( $B = 80$ ,

$$N = 6, \sigma^2 = -16.9dB, \sigma_A^2 = 1, \sigma_q^2 \rightarrow \infty)$$

be almost doubled by selecting the right placement. The big difference is given by the erasures that occur when all the training symbols are hit. However, once the packet begins and ends with a training symbol, the influence of the training placement is relatively low. The placement  $\mathcal{J}_{max}$  that maximizes the bound on the achievable rate was found by numerical evaluations. Except that it is symmetric, this placement does not correspond to a regular scheme (*e.g.*, is not uniform). Moreover, if the system parameters are

changed, the placement that maximizes  $\underline{C}$  is different. For example, if we change only the noise power to  $\sigma^2 = 0dB$ , then we obtained  $\mathcal{J}_{max} = \{5, 14, 30, 51, 67, 76\}$ .

The upper bound  $\bar{C}$  given by (34) on the achievable rate of the channel with soft outputs for the “optimal” and “middle” placements is represented in Fig. 6. It can be seen that at high values of the interference power the improvement is around 0.1 bps, or 30%. The numerical values of the capacity obtained for high  $\sigma_q^2$  are very close to those given in table I; in this situation (low noise power, high interference power) the loss induced by considering the channel as a binary channel with erasures is very small.

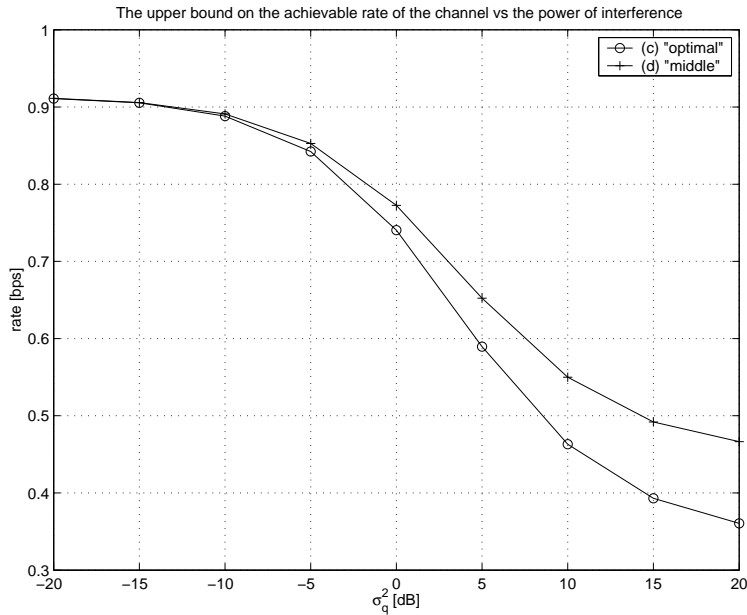


Fig. 6. The upper bound on the capacity of the communication system;  $B = 80$ ,  $N = 6$ ,  $\sigma^2 = -16.9dB$ ,  $\sigma_A^2 = 1$ .

## VI. CONCLUSIONS

In this paper we considered the channel estimation problem in the presence of an asynchronous packet interference. Using a connection between the Bayesian MMSE and the FIM of the received signal, it was shown that the optimal placement should be searched for within a small set whose structure does not depend on the parameters of the system. It was shown that placing the training symbols in two clusters of equal or nearly equal sizes optimizes the MMSE for high values of the interference power. The behavior of the genie bound and the simulations suggest that this is the solution of the problem for any set of

system parameters. For data transmission the numerical results show that the influence of the placement on the achievable data rate can be significant.

## APPENDIX

### *Proof of Theorem 1*

Taking into account expression (22) of  $f(\mathbf{x})$ , that is

$$f(\mathbf{x}; \mathbf{p}) = \sum_{u=0}^{N-1} 2p_u \left( \frac{1}{2}f(\mathbf{x}|U=u) + \frac{1}{2}f(\mathbf{x}|U=N+u) \right) + \varepsilon f(\mathbf{x}|U=N),$$

we introduce a new set of coefficients to remove the dependence among the elements of vector  $\mathbf{p}$ .

Consider the mixture coefficients  $\mathbf{c} = [c_1, \dots, c_N]^T$ , that satisfy  $c_k \geq 2\varepsilon, \forall k = 1, \dots, N-1, c_N \geq 0$  and  $\sum_{k=1}^N c_k = 1 - \varepsilon$ , and the following pdf

$$f_c(\mathbf{x}; \mathbf{c}) \triangleq \sum_{u=1}^{N-1} c_u \left( \frac{1}{2}f(\mathbf{x}|U=u) + \frac{1}{2}f(\mathbf{x}|U=N+u) \right) + \varepsilon f(\mathbf{x}|U=N). \quad (35)$$

Define the function  $g : \mathcal{P} \rightarrow \mathbb{R}^N$

$$(g(\mathbf{p}))_u = 2p_u, \quad u = 1, \dots, N-1 \quad (36)$$

$$(g(\mathbf{p}))_N = 2p_0 = 2(p_N - \varepsilon), \quad (37)$$

Observe that from (35) and (22) we have  $f_c(\cdot; g(\mathbf{p})) = f(\cdot; \mathbf{p})$ . Defining  $\mathcal{C}_2 \triangleq g(\mathcal{P})$ , and  $g_1 : \mathcal{P} \rightarrow \mathcal{C}_2$ ,  $g_1(\mathbf{p}) \triangleq g(\mathbf{p})$ , then  $g_1$  is a bijective function.  $\mathcal{C}_2$  is given by

$$\mathcal{C}_2 \triangleq \{ \mathbf{c} \in \mathbb{R}^N \mid c_k \geq 2\varepsilon, k = 1, \dots, N-1; \sum_{k=1}^N c_k = 1 - \varepsilon; c_k/(2\varepsilon) \in \mathbb{N}, k = 1, \dots, N \}.$$

$\mathcal{C}_2$  and  $\mathcal{P}$  are both discrete sets with finite number of elements. Define  $\mathcal{C}_1 \subset \mathbb{R}^N$  the smallest closed set that contains  $\mathcal{C}_2$ . Clearly,

$$\mathcal{C}_1 \triangleq \{ \mathbf{c} \in \mathbb{R}^N \mid c_k \geq 2\varepsilon, k = 1, \dots, N-1; c_N \geq 0; \sum_{k=1}^N c_k = 1 - \varepsilon \}. \quad (38)$$

Observe that the set  $\mathcal{C}_1$  is a bounded set with bounds given by linear constraints. A point  $\mathbf{c} \in \mathcal{C}_1$  is a vertex if  $\forall \mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C}_1$  and  $\forall a, b \geq 0, a + b = 1$ , we have

$$a\mathbf{c}_1 + b\mathbf{c}_2 = \mathbf{c} \implies a \in \{0, 1\}.$$

The set of vertices of  $\mathcal{C}_1$  is given by

$$\mathcal{V} \triangleq \{\mathbf{c} \mid \exists m \in \{1, \dots, N-1\} : c_m = 1 - (2N-3)\varepsilon; c_N = 0; c_k = 2\varepsilon \forall k \neq m, k \neq N\} \\ \cup \{\mathbf{c} \mid c_N = 1 - (2N-1)\varepsilon, c_k = 2\varepsilon \forall k \neq N\}.$$

There is the following relation among the regions above  $\mathcal{V} \subset \mathcal{C}_2 \subset \mathcal{C}_1$ . Note that the  $\mathbf{c}$  vectors in region  $\mathcal{C}_1$  take continuous values. Consider the problem of finding  $\mathbf{c} \in \mathcal{C}_1$  that maximize  $\mu(\mathbf{c}) = \mathbf{1}^T \mathbf{J}(f(\cdot; \mathbf{c})) \mathbf{1}$

$$\mathbf{c}_0 \triangleq \arg \max_{\mathbf{c} \in \mathcal{C}_1} \{\mu(\mathbf{c})\}. \quad (39)$$

From the convexity of the FIM functional and the convexity of the region  $\mathcal{C}_1$  it is straightforward that the solution  $\mathbf{c}_0 \in \mathcal{V}$ . Moreover, since  $\mathcal{V} \subset \mathcal{C}_2 \subset \mathcal{C}_1$ , we have

$$\mathbf{c}_0 = \arg \max_{\mathbf{c} \in \mathcal{C}_2} \{\mu(\mathbf{c})\} \quad (40)$$

It follows that any solution  $\mathbf{p}_0$  of problem (15) is given by  $g_1^{-1}(\mathbf{c}_0)$  which means that the set of solutions of (15) is the set  $\mathcal{P}_0 \triangleq g_1^{-1}(\mathcal{V})$ . The set  $\mathcal{P}_0$  contains all those vectors  $\mathbf{p} \in \mathcal{P}$  for which one of the vector elements has its maximum possible value. ■

### *Proof of Theorem 2*

We show the theorem by considering a specific suboptimal estimator denoted  $\hat{A}_d(\mathbf{x})$  and showing that

$$\lim_{\sigma_q^2 \rightarrow \infty} (\mathbb{E}\{|\hat{A}_d(\mathbf{x}) - A|^2\} - \xi(\mathbf{p}; \sigma_q^2)) = 0. \quad (41)$$

Using the optimality of the MMSE estimator, *i.e.*,  $\mathbb{E}\{|\hat{A}_d(\mathbf{x}) - A|^2\} \geq \mathcal{E}(\mathbf{p}; \sigma_q^2)$ , the lemma follows.

The estimator  $\hat{A}_d(\mathbf{x})$  considered first detects the position  $U$  of the interference packet with respect to the training symbols, and then does linear estimation based on the detected position. The estimator is defined such that if the detection is correct, then the estimate obtained is the same as the one given by the genie estimator. The proof shows that the probability of error can be made arbitrarily small by increasing  $\sigma_q^2$  and the estimation error is bounded for the cases when detection is wrong.

The detector is described by its decision regions  $\mathcal{D}_k$ ,  $k = 0, \dots, 2N-1$ ,  $\bigcup_{k=0}^{2N-1} \mathcal{D}_k = \mathbb{C}^N$  :

$$\hat{U}(\mathbf{x}) = k \text{ if } \mathbf{x} \in \mathcal{D}_k, \forall k \in \{0, \dots, 2N-1\}. \quad (42)$$

The decisions regions are given by

$$\begin{aligned}\mathcal{D}_k &= \{\mathbf{x} : s(|\mathbf{x}| - x_T \mathbf{1}) = s(\text{diag}(\mathbf{D}_k) - \sigma^2 \mathbf{1})\} \quad k = 1, \dots, 2N - 1 \\ \mathcal{D}_0 &= \mathbb{C}^N \setminus \bigcup_{k=1}^{2N-1} \mathcal{D}_k.\end{aligned}$$

$|\mathbf{x}|$  is the vector of absolute values and the threshold  $x_T$  is a positive number that can be chosen function of  $\sigma_q^2$ .  $s(y)$  is the step function for  $y \in \mathbb{R}$  and  $s(\mathbf{y})$ ,  $\mathbf{y} \in \mathbb{R}^N$  is the  $N \times 1$  vector with components  $s(y_k)$ . The threshold  $x_T$  will be chosen function of the interference power  $\sigma_q^2$ . The diagonal matrices  $\mathbf{D}_k$  have been defined in (8).

Using the detected value  $\hat{U}(\mathbf{x})$  the estimator is

$$\hat{A}_d(\mathbf{x}) = \mathbf{1}^T \mathbf{C}_{\hat{U}(\mathbf{x})}^{-1} \mathbf{x}, \quad (43)$$

and its MSE is given by

$$\mathbb{E}\{|\hat{A}_d(\mathbf{x}) - A|^2\} = \sum_{u=0}^{2N-1} \mathbb{P}\{U = u\} \mathbb{E}\{|\hat{A}_d(\mathbf{x}) - A|^2 | U = u\}.$$

To show the theorem is enough to prove for each  $u$  that

$$\lim_{\sigma_q^2 \rightarrow \infty} (\mathbb{E}\{|\hat{A}_d(\mathbf{x}) - A|^2 | U = u\} - \mathbb{E}\{|\tilde{A}(\mathbf{x}, u) - A|^2 | U = u\}) = 0. \quad (44)$$

In the previous equation  $\tilde{A}(\mathbf{x}, u)$  is the genie-aided estimator that was already defined.

$$\begin{aligned}\mathbb{E}\{|\hat{A}_d(\mathbf{x}) - A|^2 | U = u\} &= \sum_{v=0}^{2N-1} \int_{\mathbf{x} \in \mathcal{D}_v} \int_A |\hat{A}_d(\mathbf{x}) - A|^2 f(\mathbf{x}, A | U = u) dA d\mathbf{x} \\ &= \int_{\mathbf{x} \in \mathcal{D}_u} \int_A |\hat{A}_d(\mathbf{x}) - A|^2 f(\mathbf{x}, A | U = u) dA d\mathbf{x} \\ &\quad + \sum_{v \neq u} \int_{\mathbf{x} \in \mathcal{D}_v} \int_A |\hat{A}_d(\mathbf{x}) - A|^2 f(\mathbf{x}, A | U = u) dA d\mathbf{x} \\ &\leq \int_{\mathbf{x}} \int_A |\tilde{A}(\mathbf{x}, u) - A|^2 f(\mathbf{x}, A | U = u) dA d\mathbf{x} \\ &\quad + \sum_{v \neq u} \int_{\mathbf{x} \in \mathcal{D}_v} \int_A |\hat{A}_d(\mathbf{x}) - A|^2 f(\mathbf{x}, A | U = u) dA d\mathbf{x} \quad (45) \\ &= \mathbb{E}\{|\tilde{A}(\mathbf{x}, U) - A|^2 | U = u\} \\ &\quad + \sum_{v \neq u} \int_{\mathbf{x} \in \mathcal{D}_v} \int_A |\hat{A}_d(\mathbf{x}) - A|^2 f(\mathbf{x}, A | U = u) dA d\mathbf{x}.\end{aligned}$$

Relation (45) holds because if  $\hat{U}(\mathbf{x}) = U$ , then  $\hat{A}_d(\mathbf{x}) = \tilde{A}(\mathbf{x}, U)$ , as it was mentioned before.

The second term can be bounded above :

$$\begin{aligned} I_u &\stackrel{\Delta}{=} \sum_{v \neq u} \int_{\mathbf{x} \in \mathcal{D}_v} \int_A |\hat{A}_d(\mathbf{x}) - A|^2 f(\mathbf{x}, A|U = u) dA d\mathbf{x} \\ &\leq \mathbb{P}\{\mathbf{x} \notin \mathcal{D}_u | U = u\} \sup_{\mathbf{x} \in \mathbb{C}^N; v \neq u} \left\{ f(\mathbf{x}|U = u) \int_A |\hat{A}_d(\mathbf{x}) - A|^2 f(A|\mathbf{x}, U = u) dA \right\}. \end{aligned} \quad (46)$$

Look first at the second factor that is the  $\sup\{\cdot\}$ . We have:

$$|\hat{A}_d(\mathbf{x}) - A|^2 \leq 2 \left( |\hat{A}_d(\mathbf{x})|^2 + |A|^2 \right). \quad (47)$$

For any fixed  $\mathbf{x}$ , there is a  $v$  such that

$$\begin{aligned} \hat{A}_d(\mathbf{x}) &= \tilde{A}(\mathbf{x}, v) \\ |\hat{A}_d(\mathbf{x})|^2 &= |\mathbf{1}^T \mathbf{C}_v^{-1} \mathbf{x}|^2 \\ &= \left| \mathbf{1}^T \left( \mathbf{D}_v^{-1} - \frac{\mathbf{D}_v^{-1} \mathbf{1} \mathbf{1}^T \mathbf{D}_v^{-1}}{1 + \mathbf{1}^T \mathbf{D}_v^{-1} \mathbf{1}} \right) \mathbf{x} \right|^2. \end{aligned}$$

The elements of  $\mathbf{D}_v^{-1}$  are bounded for any values of the parameters involved and for any choice of  $v$ ; the same is true for the vector  $\mathbf{1}^T \left( \mathbf{D}_v^{-1} - \frac{\mathbf{D}_v^{-1} \mathbf{1} \mathbf{1}^T \mathbf{D}_v^{-1}}{1 + \mathbf{1}^T \mathbf{D}_v^{-1} \mathbf{1}} \right)$ . The bound can be computed explicitly. Defining  $C_1 \stackrel{\Delta}{=} (\max_{v,k} \{|\mathbf{1}^T \mathbf{C}_v^{-1} \mathbf{e}_k|\})^2$ , we get

$$|\hat{A}_d(\mathbf{x})|^2 \leq C_1 |\mathbf{x}|^2. \quad (48)$$

Now continue with the expression under the  $\sup\{\cdot\}$  in (46). A calculation using Gaussian pdf's shows that

$$\int_A |A|^2 f(A|U = u, \mathbf{x}) dA = \frac{|\mathbf{1}^T \mathbf{D}_u^{-1} \mathbf{x}|^2}{(1 + \mathbf{1}^T \mathbf{D}_u^{-1} \mathbf{1})^2} + (1 + \mathbf{1}^T \mathbf{D}_u^{-1} \mathbf{1})^{-1}.$$

From a similar argument as before, there exist constants  $C_3 > 0$ ,  $C_4 > 0$ , such that

$$\int_A |A|^2 f(A|U = u, \mathbf{x}) dA \leq C_3 |\mathbf{x}|^2 + C_4$$

Using this in the formula of  $I_u$ , we obtain

$$I_u \leq \mathbb{P}\{\mathbf{x} \notin \mathcal{D}_u | U = u\} \sup_{\mathbb{C}^N} \{(2(C_1 + C_3)|\mathbf{x}|^2 + 2C_4) f(\mathbf{x}|U = u)\} \quad (49)$$



For any square matrix  $\mathbf{A}$ , denote by  $\lambda_{\min}(\mathbf{A})$  and  $\lambda_{\max}(\mathbf{A})$  the minimum and maximum eigenvalues respectively. We have the inequalities

$$\begin{aligned} \mathbf{x}^H \mathbf{C}_u^{-1} \mathbf{x} &\geq \lambda_{\min}(\mathbf{C}_u^{-1}) |\mathbf{x}|^2 \\ &\geq \frac{1}{\lambda_{\max}(\mathbf{C}_u)} |\mathbf{x}|^2. \end{aligned}$$

Substituting in  $I_u$ ,

$$\begin{aligned} I_u &\leq \mathbb{P}\{\mathbf{x} \notin \mathcal{D}_u | U = u\} \sup_{\mathbb{C}^N} \left\{ (2(C_1 + C_3) |\mathbf{x}|^2 + 2C_4) \frac{1}{\pi^N |\mathbf{C}_u|} \exp \left\{ -\frac{|\mathbf{x}|^2}{\lambda_{\max}(\mathbf{C}_u)} \right\} \right\} \quad (50) \\ &\leq \mathbb{P}\{\mathbf{x} \notin \mathcal{D}_u | U = u\} \sup_{\mathbb{C}^N} \left\{ (2(C_1 + C_3) |\mathbf{x}|^2 + 2C_4) \frac{1}{\pi^N \lambda_{\max}(\mathbf{C}_u) (\lambda_{\min}(\mathbf{C}_u))^{N-1}} \right. \\ &\quad \left. \exp \left\{ -\frac{|\mathbf{x}|^2}{\lambda_{\max}(\mathbf{C}_u)} \right\} \right\}. \end{aligned}$$

$\lambda_{\min}(\mathbf{C}_u) \geq \sigma^2$  which is bounded away from zero. Observe that the expression that multiplies  $\mathbb{P}\{\mathbf{x} \notin \mathcal{D}_u | U = u\}$  is uniformly bounded above by a constant  $C_5 > 0$  that does not depend on  $\lambda_{\max}(\mathbf{C}_u)$ . So  $C_5$  does not depend on  $\sigma_q^2$ ,  $u$  or  $v$ .

We showed that

$$I_u \leq C_5 \mathbb{P}\{\mathbf{x} \notin \mathcal{D}_u | U = u\}. \quad (51)$$

To complete the proof of (44) we show that

$$\mathbb{P}\{\mathbf{x} \notin \mathcal{D}_u | U = u\} \xrightarrow{\sigma_q^2 \rightarrow \infty} 0.$$

Use the union bound to upper bound the error probability. The distribution of  $\mathbf{x}$  conditioned on  $U$  is given by (10). Consider  $u \in \{1, \dots, N-1\}$ . In this case the marginal conditional distribution of each element is given by

$$f(x_k | U = u) = \begin{cases} \frac{1}{\pi(1+\sigma^2+\sigma_q^2)} \exp\left(-\frac{|x_k|^2}{1+\sigma^2+\sigma_q^2}\right) & \text{if } k = 1, \dots, u \\ \frac{1}{\pi(1+\sigma^2)} \exp\left(-\frac{|x_k|^2}{1+\sigma^2}\right) & \text{if } k = u+1, \dots, N \end{cases}.$$

$$\mathbb{P}\{\mathbf{x} \notin \mathcal{D}_u | U = u\} \leq \sum_{k=1}^u \mathbb{P}\{|x_k| \leq x_T | U = u\} + \sum_{k=u+1}^N \mathbb{P}\{|x_k| > x_T | U = u\} \quad (52)$$

$$= \sum_{k=1}^u (1 - \exp(-x_T^2/(1+\sigma^2+\sigma_q^2))) + \sum_{k=u+1}^N \exp(-x_T^2/(1+\sigma^2)) \quad (53)$$

To show that  $\forall \epsilon > 0 \exists x_T, \sigma_q^2$  such that  $\mathbb{P}\{\mathbf{x} \notin \mathcal{D}_u | U = u\} < \epsilon$ , just choose  $x_T$  such that  $\exp(-x_T^2/(1+\sigma^2)) < \frac{\epsilon}{N}$  and then  $\sigma_q^2$  such that  $1 - \exp(-x_T^2/(1+\sigma^2+\sigma_q^2)) < \frac{\epsilon}{N}$ . The proof is similar for the rest of values  $u = 0, N, \dots, 2N-1$ . This completes the last step of the proof of (44) and the theorem follows. ■

Probability of detection error when  $\sigma_q^2 \rightarrow \infty$

In this section of the appendix we'll derive the detector  $\tilde{s}_k$  and the conditional probabilities of error  $P_{k,e}$ . Towards this, we first derive the binary ML detector without considering  $\sigma_q^2 \rightarrow \infty$ . Then we derive the probability of error for this binary detector. From the transmission model (1), the  $k$ -th received symbol (assume  $k \notin \mathcal{J}$ , *i.e.*, data symbol) is given by

$$y_k = s_k A + z_k. \quad (54)$$

The receiver uses the genie estimator, *i.e.*, it knows the realization of the random variable  $U$ . Consider  $U = u$  and write the channel parameter  $A = \tilde{A}(\mathbf{x}, u) + \Delta A(\mathbf{x}, u)$ , where  $\tilde{A}(\mathbf{x}, u)$  is the genie estimator (24). The received data symbol  $y_k$  becomes

$$y_k = s_k \tilde{A}(\mathbf{x}, u) + s_k \Delta A(\mathbf{x}, u) + z_k. \quad (55)$$

Conditioned on  $U = u$  the random variables  $\tilde{A}(\mathbf{x}, u)$  and  $\Delta A(\mathbf{x}, u)$  are independent complex Gaussian with distributions

$$\begin{aligned} \tilde{A}(\mathbf{x}, u) &\sim \mathcal{CN}(0, \mathbf{1}^T \mathbf{C}_u^{-1} \mathbf{1}) \\ \Delta A(\mathbf{x}, u) &\sim \mathcal{CN}(0, \mathbf{1} - \mathbf{1}^T \mathbf{C}_u^{-1} \mathbf{1}). \end{aligned}$$

The pdf of the observation  $y_k$  conditioned on  $U$ ,  $\tilde{A}$ , and the transmitted symbol  $s_k$  is

$$\begin{aligned} f(y_k | s_k, U, \tilde{A}) &= \mathbb{P}\{H_k = 0 | U\} f(y_k | s_k, H_k = 0, U, \tilde{A}) + \mathbb{P}\{H_k = 1 | U\} f(y_k | s_k, H_k = 1, U, \tilde{A}) \\ &= \mathbb{P}\{H_k = 0 | U\} \mathcal{CN}(y_k; s_k \tilde{A}, \mathbf{1} - \mathbf{1}^T \mathbf{C}_U^{-1} \mathbf{1} + \sigma^2) \\ &\quad + \mathbb{P}\{H_k = 1 | U\} \mathcal{CN}(y_k; s_k \tilde{A}, \mathbf{1} - \mathbf{1}^T \mathbf{C}_U^{-1} \mathbf{1} + \sigma^2 + \sigma_q^2). \end{aligned} \quad (56)$$

The arguments  $\mathbf{x}$  and  $U$  in the expressions of  $\tilde{A}$  and  $\Delta A$  were omitted for simplicity.

Assuming that the transmitted symbols are equiprobable, the symbol-by-symbol maximum likelihood (ML) detector for the  $k$ -th symbol (which is assumed to be data symbol) is found by solving

$$\hat{s}_k = \arg \max_{s_k \in \{\pm 1\}} f(y_k | s_k, U, \tilde{A}). \quad (57)$$

Taking into account the monotonicity of the exponential function, one can show that

$$\hat{s}_k = \text{sgn}\{\mathcal{R}e\{\tilde{A}^* y_k\}\}. \quad (58)$$

Conditioned on  $H_k$  and  $U$  the received signal  $y_k$  is Gaussian, thus  $\mathbb{P}\{\hat{s}_k \neq s_k | U, H_k\}$  can be computed using the well-known formula for the probability of error of BPSK signaling in Rayleigh fading [14]. If  $y = as + n$  is the received complex signal,  $s \in \{\pm 1\}$  the BPSK transmitted symbols with equal priors,

$a \sim \mathcal{CN}(0, \sigma_a^2)$  the known channel parameter, and  $n \sim \mathcal{CN}(0, \sigma_n^2)$  the noise, then the probability of detection error when  $\hat{s} = \text{sgn}\{\mathcal{R}e\{a^*y\}\}$  is given by

$$P_e = \frac{1}{2} \left( 1 - \frac{1}{\sqrt{\frac{\sigma_n^2}{\sigma_a^2} + 1}} \right). \quad (59)$$

Substituting  $\sigma_n^2 = 1 - (\mathbf{1}^T \mathbf{C}_u^{-1} \mathbf{1}) + \sigma^2 + h\sigma_q^2$  and  $\sigma_a^2 = \mathbf{1}^T \mathbf{C}_u^{-1} \mathbf{1}$  in (59) we obtain

$$\theta(u, h) \triangleq \mathbb{P}\{\hat{s}_k \neq s_k | U = u, H_k = h\} = \frac{1}{2} \left( 1 - \sqrt{\frac{\mathbf{1}^T \mathbf{C}_u^{-1} \mathbf{1}}{\sigma^2 + h\sigma_q^2 + 1}} \right) = \frac{1}{2} \left( 1 - \sqrt{\frac{1 - m(u)}{\sigma^2 + h\sigma_q^2 + 1}} \right). \quad (60)$$

When the interference power is high, the probability of detection error is close to one half if the current data symbol is hit or if all the training symbols are hit:

$$\begin{aligned} \lim_{\sigma_q^2 \rightarrow \infty} \theta(u, 1; \sigma_q^2) &= \frac{1}{2}, \\ \lim_{\sigma_q^2 \rightarrow \infty} \theta(N, h; \sigma_q^2) &= \frac{1}{2}. \end{aligned}$$

These two situations were defined as erasure events; when the power of the interference is high ( $\sigma_q^2 \rightarrow \infty$ ), the receiver can detect them accurately (*i.e.*, detect  $H_k$  with negligible error). We obtain the detector  $\tilde{s}$  (30)

$$\tilde{s}_k \triangleq \begin{cases} \text{erasure} & \text{if } H_k = 1 \text{ or } U = N \\ \text{sgn}\{\mathcal{R}e\{\tilde{A}^* y_k\}\} & \text{otherwise} \end{cases}. \quad (61)$$

For channel  $k$ , the probability of erasure  $P_{k,e}$  was defined as

$$\begin{aligned} P_{k,e} &= \mathbb{P}\{\tilde{s}_k \neq s_k | (\{H_k = 1\} \cup \{U = N\})^C\} \\ &= \mathbb{P}\{\hat{s}_k \neq s_k | (\{H_k = 1\} \cup \{U = N\})^C\}. \end{aligned}$$

To give the probabilities  $P_{k,e}$  in a convenient form, we introduce some new notations. The  $N$  training symbols divide the data symbols into maximum  $N + 1$  data blocks. For each  $u = 1, \dots, N + 1$  define  $\mathcal{B}_u$  as the set of indices of the  $u$ -th block of data symbols, and  $\Delta_u$  its size

$$\mathcal{B}_u \triangleq \begin{cases} \emptyset & \text{if } t_{u+1} - t_u = 1 \text{ or } t_{u-1} = 1 \text{ or } t_u = N \\ \{1, \dots, t_1 - 1\} & \text{if } u = 0 \text{ and } t_1 \geq 2 \\ \{t_u + 1, \dots, t_{u+1} - 1\} & \text{if } u = 1, \dots, N - 1 \text{ and } t_{u+1} - t_u \geq 2 \\ \{t_u + 1, \dots, N\} & \text{if } u = N \text{ and } t_u \leq N - 1 \end{cases}, \quad (62)$$

$$\Delta_u \triangleq |\mathcal{B}_u| = \begin{cases} t_1 - 1 & \text{if } u = 0 \\ t_{u+1} - t_u - 1 & \text{if } u = 1, \dots, N - 1 \\ N - t_N & \text{if } u = N \end{cases}. \quad (63)$$

The probability of error is given below for each of the cases  $k \in \mathcal{B}_0$ ,  $k \in \mathcal{B}_N$  and  $k \notin \mathcal{B}_0 \cup \mathcal{B}_N$ :

$$\begin{aligned}
P_{k,e;k \in \mathcal{B}_0} &= \frac{1}{B-1-(t_1-k)} \left( (k-1)\theta(0,0) + \sum_{u=1}^{N-1} (1+\Delta_u)\theta(N-u,0) + \Delta_N\theta(0,0) \right) \\
P_{k,e;k \in \mathcal{B}_N} &= \frac{1}{B-1-(k-t_N)} \left( \Delta_0\theta(0,0) + \sum_{u=1}^{N-1} (1+\Delta_u)\theta(u,0) + (B-k)\theta(0,0) \right) \\
P_{k,e;k \in \mathcal{B}_w, 0 < w < N} &= \frac{1}{B-1} \left( \Delta_0\theta(0,0) + \sum_{u=1}^{w-1} (1+\Delta_u)\theta(u,0) + (k-t_w)\theta(w,0) \right. \\
&\quad \left. + (t_{w+1}-k)\theta(N-w,0) + \sum_{u=w+1}^{N-1} (1+\Delta_u)\theta(N-u,0) + \Delta_N\theta(0,0) \right).
\end{aligned}$$

*Proof of Lemma 2*

**For the technical report only**

The proof follows the one for real random variables given in [11].

The FIM of a random vector  $\mathbf{X}$  with pdf  $p(\mathbf{x})$  is given by

$$\begin{aligned}
\mathbf{J}(p) &= \mathbb{E}[(\nabla \ln(p))(\nabla \ln(p))^H] \\
&= \int \frac{(\nabla p)(\nabla p)^H}{p} d\mathbf{x}.
\end{aligned}$$

where all the gradients are with respect to  $\mathbf{x}^*$ . In this proof most of the times we omit the argument  $\mathbf{x}$  for simplicity.

Let  $a, b$  be two positive scalars,  $a + b = 1$ ,  $p_1(\mathbf{x})$  and  $p_2(\mathbf{x})$  two probability distributions where  $\mathbf{x}$  is an  $N \times 1$  vector, and  $P(\mathbf{x}) \triangleq ap_1(\mathbf{x}) + bp_2(\mathbf{x})$  their mixture with coefficients  $a$  and  $b$ .

Considering the function

$$\psi(\mathbf{x}) \triangleq \sqrt{ap_1(\mathbf{x})} + j\sqrt{bp_2(\mathbf{x})},$$

we have

$$P = |\psi|^2 = ap_1 + bp_2.$$

The FIM of  $P$  can be written function of  $|\psi|$

$$\mathbf{J}(P) = \int \frac{(\nabla P)(\nabla P)^H}{P} d\mathbf{x} \tag{64}$$

$$= 4 \int (\nabla |\psi|)(\nabla |\psi|)^H d\mathbf{x} \tag{65}$$

The following identities can be obtained by direct calculation

$$\begin{aligned}\nabla\psi &= \frac{1}{2} \left( \sqrt{\frac{a}{p_1}} \nabla p_1 + j \sqrt{\frac{b}{p_2}} \nabla p_2 \right) \\ (\nabla\psi)(\nabla\psi)^H &= \frac{1}{4} \left( \frac{a}{p_1} \nabla p_1 (\nabla p_1)^H + \frac{b}{p_2} \nabla p_2 (\nabla p_2)^H + j \sqrt{\frac{ab}{p_1 p_2}} (\nabla p_2 (\nabla p_1)^H - \nabla p_1 (\nabla p_2)^H) \right).\end{aligned}$$

It follows shortly that :

$$\begin{aligned}\frac{1}{2} ((\nabla\psi)(\nabla\psi)^H + (\nabla\psi^*)(\nabla\psi^*)^H) &= \frac{1}{4} \left( \frac{a}{p_1} \nabla p_1 (\nabla p_1)^H + \frac{b}{p_2} \nabla p_2 (\nabla p_2)^H \right) \\ \frac{1}{2} \int ((\nabla\psi)(\nabla\psi)^H + (\nabla\psi^*)(\nabla\psi^*)^H) d\mathbf{x} &= \frac{1}{4} (a\mathbf{J}(p_1) + b\mathbf{J}(p_2)).\end{aligned}\quad (66)$$

On the other hand, we have

$$\begin{aligned}\psi(\mathbf{x}) &= R(\mathbf{x}) \exp(jS(\mathbf{x})), \\ \nabla\psi &= (\nabla R) \exp(jS) + j(\nabla S)R \exp(jS).\end{aligned}$$

Thus,

$$\begin{aligned}(\nabla\psi)(\nabla\psi)^H &= (\nabla R)(\nabla R)^H + R^2(\nabla S)(\nabla S)^H + j((\nabla S)R(\nabla R)^H - (\nabla R)(\nabla S)^H R), \\ \frac{1}{2} ((\nabla\psi)(\nabla\psi)^H + (\nabla\psi^*)(\nabla\psi^*)^H) &= (\nabla R)(\nabla R)^H + R^2(\nabla S)(\nabla S)^H\end{aligned}\quad (67)$$

Comparing (65), (66), and (67) we get

$$a\mathbf{J}(p_1) + b\mathbf{J}(p_2) = \mathbf{J}(P) + 4 \int R^2(\nabla S)(\nabla S)^H d\mathbf{x}.\quad (68)$$

Thus,

$$a\mathbf{J}(p_1) + b\mathbf{J}(p_2) \geq \mathbf{J}(P),\quad (69)$$

where the inequality means that the difference is positive definite.

## REFERENCES

- [1] R.Knopp and P.A.Humblet, "Information capacity and power control in single cell multi-user communications ," in *Proc. Intl Conf. Comm.*, (Seattle, WA), pp. 331–335, June 1995.
- [2] D.N.C.Tse and S.V.Hanly, "Multiaccess Fading Channels : I Polymatroid Structure, Optimal Resource Allocation and Throughput Capacities ," *IEEE Trans. Information Theory*, vol. 44, pp. 2796–2815, November 1998.
- [3] S. Adireddy and L. Tong, "Medium Access Control using Channel State Information for Large Sensor Networks," in *Proceedings of the IEEE Workshop on Multimedia Signal Processing 2002*, (St. Thomas, US Virgin Islands), December 2002. <http://acsp.ece.cornell.edu/pubC.html/>.

- [4] S. Adireddy and L. Tong, "Exploiting decentralized channel state information for random access," *Submitted to IEEE Trans. Info. Theory*, November 2002. <http://acsp.ece.cornell.edu/pubJ.html>.
- [5] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM System," *IEEE Trans. on Consumer Electronics*, vol. 44, pp. 1122–1128, Aug. 1998.
- [6] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of known symbols for frequency-selective block-fading channels," *IEEE Trans. Info. Theory*, vol. 48, pp. 2338–2353, August 2002. <http://www.ece.cornell.edu/~ltong>.
- [7] M. Dong and L. Tong, "Optimal Design and Placement of Pilot Symbols for Channel Estimation," *IEEE Trans. on Signal Processing*, vol. 50, pp. 3055–3069, December 2002.
- [8] S. Ohno and G. B. Giannakis, "Optimal Training and Redundant Precoding for Block Transmissions with Application to Wireless OFDM," *IEEE Trans. on Communications*, vol. 50, December 2002.
- [9] C. Budianu and L. Tong, "Channel Estimation for Space-Time Block Coding Systems," *IEEE Trans. Signal Processing*, vol. 50, pp. 2515–2528, Oct 2002.
- [10] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ 07632: Prentice Hall, 1993.
- [11] B. Frieden, A. Plastino, A. Plastino, and B. Soffer, "Fisher-based thermodynamics: Its Legendre transform and concavity properties," *Physical Review E*, vol. 60, pp. 48–53, July 1999.
- [12] A. Dembo, T. Cover, and J. Thomas, "Information Theoretic Inequalities," *IEEE Transactions on Information Theory*, vol. 37, pp. 1501–1518, June 1991.
- [13] C. Budianu and L. Tong, "Channel Estimation under Asynchronous Packet Interference," Tech. Rep. ACSP-TR-03-03, Cornell University, March 2003. <http://acsp.ece.cornell.edu/pubJ.html>.
- [14] J. Proakis, *Digital Communications*. McGraw Hill, 4 ed., 2001.