# Optimum Training for CSI Acquisition in Cognitive Radio Channels

Alberto Rico-Alvariño and Carlos Mosquera
Signal Theory and Communications Department, Universidade de Vigo 36310 Vigo, Spain
email: {alberto,mosquera}@gts.uvigo.es

***Abstract*—One of the main issues in the Overlay Cognitive Radio Paradigm (a framework where the secondary transmitter partially cooperates with the primary user) is the acquisition of channel state information at the secondary transmitter. In [1] a simple interaction framework that could allow the estimation of the channel values at the secondary transmitter was presented, although its performance was not characterized. In this paper, we extend the framework in [1] to the MIMO and SISO Time-Varying channels, present closed form expressions for the mean square error of these channel estimates, and derive optimum training sequences to minimize the estimation variance.**

## I. Introduction

In the last few years the interest of knowing the potential of those cognitive radio systems where the Secondary Transmitter (ST) is aware of the message to be conveyed by the Primary Transmitter (PT) has been found to be of special interest, in which is known as the Overlay Cognitive Radio paradigm [2]. Although the study of a general cooperation scheme between primary and secondary transmitters was introduced in [3], the most realistic model for Overlay Cognitive Radio might be the one introduced by Jovicić and Viswanath [1], the so-called *Cognitive Radio Channel*, that introduces additional constraints to a general cooperative channel: the ST is aware of the primary message, which conforms a partially cooperative interference channel; the Primary Receiver (PR) uses a single user decoder, i.e., is completely unaware of the presence of the ST; and the rate of the primary system is not compromised.

In [1] the problem of obtaining Channel State Information (CSI) at the ST was shown to be critical: if the ST does not know the PT to PR and ST to PR (complex) channel values, the primary signal contributions coming from the ST and the PT could result in destructive interference, thus causing a severe degradation in the primary link (see e.g. [4] [5] for some scenarios where CSI is critical). In the same work, the authors described a simple interaction scheme between the PR and ST that would allow to obtain CSI. If we assume that the PR estimates the *equivalent channel* (the channel resulting from the addition of the contributions coming from the ST and the PT) and broadcasts its value, the ST might gain access to CSI in the following way: let $\alpha$ be the channel from the ST to the PR, and $\eta$ the channel from the PT to the PR. In the first time slot, when the ST joins the network, it does not transmit the primary message, so the received waveform at the PR is

$$r_i = \eta x_i + z_i, \tag{1}$$

with $x_i$ the primary codeword, and $z_i$ a sample of Gaussian noise. At that time slot, the PR broadcasts the estimated channel $h_i \approx \eta$. In the second time slot, the ST allocates $\|\gamma\|^2$ units of power to the primary message, so the received waveform is

$$r_{i+1} = (\eta + \gamma\alpha)\, x_{i+1} + z_{i+1}. \tag{2}$$

At this time, the PR broadcasts the estimated channel $h_{i+1} \approx \eta + \gamma\alpha$, so the channel values $\alpha$ and $\eta$ can be approximately obtained as $\alpha \approx \frac{h_{i+1}-h_i}{\gamma}$, $\eta \approx h_i$. Note that the one-tap precoding sequence $\boldsymbol{\gamma} = [0, \gamma]$ can be considered as a **training sequence** for the channel estimation problem.

Although this simple interaction framework has been cited in other works like [6], and even for a multiple antenna channel in [7], to the best of the authors knowledge this CSI acquisition technique has not been sufficiently studied in the literature. In this paper, we derive closed-form expressions for the mean squared error (MSE) of the channel estimates in a Single Input Single Output (SISO) channel, in a Multiple Input Multiple Output (MIMO) channel with transmit beamforming and in a SISO time-varying channel.

The remaining of the paper is structured as follows: in Section II the SISO time-invariant channel is studied; Section III extends the framework in [1] to transmit beamforming MIMO channels; a similar approach is taken to study the SISO time-varying channel in Section IV; Section V presents the results; finally, Section VI concludes the paper.

## II. SISO Channel

We will assume $\alpha$, $\eta \in \mathbb{C}$ are the time-invariant ST to PR and PT to PR channels, respectively, and model the $i$-th feedback message as

$$y_i = (\alpha\gamma_i + \eta) + n_i,\, n_i \sim \mathcal{CN}\left(0, \sigma^2\right) \tag{3}$$

where $n_i$ is a zero-mean Gaussian RV that accounts for the estimation error and $\gamma_i$ is the one-tap pre-equalizer previously introduced. We are assuming through the paper that feedback is error-free.

Research supported by the European Regional Development Fund (ERDF) and the Spanish Government under projects DYNACS (TEC2010-21245-C02-02/TCM) and COMONSENS (CONSOLIDER-INGENIO 2010 CSD2008-00010), and the Galician Regional Government under project CN 2012/260 "Consolidation of Research Units: AtlantTIC".

Note that if we remove the term $\gamma_i$, we have that the parameters $\alpha$ and $\eta$ are not identifiable. We define the vector observation $\mathbf{y} = [y_1, ..., y_M]^T$ resulting from stacking $M$ scalar observations as

$$\mathbf{y} = \boldsymbol{\gamma}\alpha + \mathbf{1}_M \eta + \mathbf{n} \tag{4}$$

with $\boldsymbol{\gamma} = [\gamma_1, ... \gamma_M]^T$, $\mathbf{1}_M$ a column vector with its $M$ entries equal to one and $\mathbf{n} \sim \mathcal{N}\left(0, \sigma^2 \mathbf{I}_M\right)$. It can be easily seen that (4) follows a Gaussian Linear Model [8], so efficient estimators $\hat{\alpha}$ and $\hat{\eta}$ exist such that they attain the Cramér-Rao Bound (CRB) variance, which is given by

$$Var_{\hat{\alpha}} = \frac{M}{\sigma^2 \det \boldsymbol{\mathcal{I}}}, \quad Var_{\hat{\eta}} = \frac{\|\boldsymbol{\gamma}\|^2}{\sigma^2 \det \boldsymbol{\mathcal{I}}} \tag{5}$$

with the determinant of the Fisher Information Matrix (FIM) $\boldsymbol{\mathcal{I}}$

$$\det \boldsymbol{\mathcal{I}} = \sigma^{-4} \left( M \|\boldsymbol{\gamma}\|^2 - \|\boldsymbol{\gamma}^H \mathbf{1}_M\|^2 \right). \tag{6}$$

For a given total power $\|\boldsymbol{\gamma}\|^2 \leq P$, it is clear that the values of $\boldsymbol{\gamma}$ that maximize the determinant of $\boldsymbol{\mathcal{I}}$, and, therefore, minimize the CRB, are those with $\boldsymbol{\gamma}^H \mathbf{1}_M = 0$ and $\|\boldsymbol{\gamma}\|^2 = P$. Just by taking any vector of this family, we arrive to

$$Var_{\hat{\alpha}} = \frac{\sigma^2}{P}, \quad Var_{\hat{\eta}} = \frac{\sigma^2}{M}. \tag{7}$$

At the view of this results, we conclude that the *training sequence* $\boldsymbol{\gamma} = [0, \gamma]$ introduced in [1] is not optimum in the sense of minimum estimation variance for a given total power, as $\boldsymbol{\gamma}^H \mathbf{1}_2 = \gamma \neq 0$.

## III. MIMO CHANNEL

In this section we extend the results from the previous one and the framework presented in [1] to the case of transmit beamforming MIMO channels.

Let us denote by $\mathbf{H}_s \in \mathbb{C}_{N_r \times N_s}$ the (assumed to be flat fading) MIMO channel from the ST to the PR, and as $\mathbf{H}_p \in \mathbb{C}_{N_r \times N_p}$ the MIMO channel from the PT to the PR. We will assume that the PT is transmitting only one data layer at a time by applying a fixed beamforming vector $\mathbf{w}_p \in \mathbb{C}_p^N$, and the ST is also conveying the primary information by using a beamforming vector $\mathbf{w}_{s,i} \in \mathbb{C}^{N_s}$, which can be time-varying. We will also assume that the PR is estimating the SIMO channel (combination of the beamforming + MIMO channel) such that the SIMO channel estimate $\mathbf{y}_i \in \mathbb{C}_{N_r}$ can be written as

$$\mathbf{y}_i = \mathbf{H}_p \mathbf{w}_p + \mathbf{H}_s \mathbf{w}_{s,i} + \mathbf{n}_i, \, \mathbf{n}_i \sim \mathcal{N}\left(\mathbf{0}, \mathbf{C}_n\right) \tag{8}$$

where $\mathbf{n}_i$ accounts for the estimation error.

If the PR broadcast these channel estimates the PT would be able to acquire some CSI. Note that the assumption of fixed beamforming $\mathbf{w}_p$ makes the separate estimation of $\mathbf{H}_p$ and $\mathbf{w}_p$ impossible, so we will denote by $\mathbf{g} \doteq \mathbf{H}_p \mathbf{w}_p$ the SIMO channel consisting on the combination of beamforming and MIMO channel from the PT to the PR. The objective of the ST is to estimate both $\mathbf{g}$ and $\mathbf{H}_s$ from the observations $\mathbf{y}_i$, by treating the sequence of beamforming vectors $\mathbf{w}_{s.i}$ as a training sequence.

### A. Estimation problem

For the sake of clarity we will denote $\mathbf{H} \doteq \mathbf{H}_s$ and $\mathbf{w}_i \doteq \mathbf{w}_{s,i}$, as the primary MIMO channel and beamforming vectors are included in the SIMO channel $\mathbf{g}$. At a given time instant $i$, our observation will be

$$\mathbf{y}_i = \mathbf{g} + \mathbf{H}\mathbf{w}_i + \mathbf{n}_i. \tag{9}$$

If we stack $M$ observations into a column vector we obtain the $MN_r \times 1$ vector $\mathbf{y}$ as follows:

$$\mathbf{y} = \begin{bmatrix} \mathbf{g} \\ ... \\ \mathbf{g} \end{bmatrix} + \begin{bmatrix} \mathbf{H}\mathbf{w}_1 \\ ... \\ \mathbf{H}\mathbf{w}_M \end{bmatrix} + \begin{bmatrix} \mathbf{n}_1 \\ ... \\ \mathbf{n}_M \end{bmatrix}. \tag{10}$$

As we are interested in estimating both the vector $\mathbf{g}$ and the matrix $\mathbf{H}$, we will rewrite (10) as

$$\mathbf{y} = \left(\mathbf{1}_M \otimes \mathbf{I}_{N_r}\right)\mathbf{g} + \left(\mathbf{W} \otimes \mathbf{I}_{N_r}\right)\mathbf{h} + \mathbf{n} \tag{11}$$

with $\mathbf{I}_M$ the $M \times M$ identity matrix, $\otimes$ the Kronecker product operator, and $\mathbf{h} = \text{vec}\mathbf{H}$ is the result of stacking the columns of $\mathbf{H}$ into a vector, so $\mathbf{h} \in \mathbb{C}^{N_s N_r}$. The training sequence matrix is the result of stacking into a matrix the training sequence $\mathbf{w}_i^T$: $\mathbf{W} = [\mathbf{w}_1, ..., \mathbf{w}_M]^T \in \mathbb{C}_{M \times N_s}$. The vector $\mathbf{n} \sim \mathcal{CN}(0, \mathbf{C})$, with $\mathbf{C} = \mathbf{C}_n \otimes \mathbf{I}_M$, is the result of stacking the $M$ noise vectors $\mathbf{n}_i$. With this, we can rewrite (11) as

$$\mathbf{y} = \mathbf{K}\mathbf{g} + \mathbf{R}\mathbf{h} + \mathbf{n} \tag{12}$$

with $\mathbf{K} \doteq \left(\mathbf{1}_M \otimes \mathbf{I}_{N_r}\right) \in \mathbb{C}_{N_r M \times N_r}$ and $\mathbf{R} \doteq \left(\mathbf{W} \otimes \mathbf{I}_{N_r}\right) \in \mathbb{C}_{N_r M \times N_r N_s}$. It can be easily seen that (12) is a Gaussian Linear Model [8], so if we define $\mathbf{A} = [\mathbf{K}\,\mathbf{R}]$ and $\mathbf{b} = \left[\mathbf{g}^T \, \mathbf{h}^T\right]^T$ the Minimum Variance Unbiased (MVU) Estimator (which is efficient) is given by

$$\hat{\mathbf{b}} = \left(\mathbf{A}^H \mathbf{C}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^H \mathbf{C}^{-1} \mathbf{y} \tag{13}$$

which is distributed according to

$$\hat{\mathbf{b}} \sim \mathcal{CN}\left(\mathbf{b}, \left(\mathbf{A}^H \mathbf{C}^{-1} \mathbf{A}\right)^{-1}\right). \tag{14}$$

Note that in the previous equations we have assumed $M \geq N_s + 1$ so the provided inverse matrices exist.

### B. Training sequence design

We will design our training sequence $\mathbf{W}$ in order to minimize the total estimation variance, subject to a total power constraint $P$:

$$\begin{aligned} &\text{minimize} \quad \text{tr}\left(\mathbf{A}^H \mathbf{C}^{-1} \mathbf{A}\right)^{-1} \\ &\text{subject to} \quad \text{tr}\left(\mathbf{W}^H \mathbf{W}\right) \leq P. \end{aligned} \tag{15}$$

In the following, we will assume that $\mathbf{C} = \sigma^2 \mathbf{I}_{MN_r}$, so the matrix in the objective function reads as

$$\mathbf{A}^H \mathbf{C}^{-1} \mathbf{A} \;\; = \;\; \frac{1}{\sigma^2} \mathbf{M} \otimes \mathbf{I}_{N_r} \tag{16}$$

with

$$\mathbf{M} \doteq \begin{bmatrix} \mathbf{1}_M^T \\ \mathbf{W}^H \end{bmatrix} [\mathbf{1}_M, \mathbf{W}] . \tag{17}$$

Therefore, the objective function in (15) reads as

$$\text{tr } \left(\mathbf{A}^H \mathbf{C}^{-1} \mathbf{A}\right)^{-1} = \sigma^2 N_r \text{tr } \mathbf{M}^{-1}. \tag{18}$$

As $N_r$ and $\sigma^2$ do not depend on $\mathbf{W}$, we can rewrite (15) as

$$\begin{array}{ll} \text{minimize} & \text{tr } \mathbf{M}^{-1} \\ \text{subject to} & \text{tr } \left(\mathbf{W}^H \mathbf{W}\right) \leq P. \end{array} \tag{19}$$

Note that

$$\mathbf{M} = \begin{bmatrix} M & \mathbf{1}_M^T \mathbf{W} \\ \mathbf{W}^H \mathbf{1}_M & \mathbf{W}^H \mathbf{W} \end{bmatrix}, \tag{20}$$

so the trace of the inverse matrix of $\mathbf{M}$ can be written as a function of the Schur complements of the submatrices in $\mathbf{M}$ as tr $\mathbf{M}^{-1} =$ tr $\mathbf{S}_M^{-1} +$ tr $S_{\mathbf{W}^M\mathbf{W}}^{-1}$, with $\mathbf{S}_M$ and $S_{\mathbf{W}^H\mathbf{W}}$ the Schur complements of $M$ and $\mathbf{W}^H\mathbf{W}$ in $\mathbf{M}$. The inverse of the latter can be expanded by using the Sherman-Morrison formula:

$$\mathbf{S}_{\mathbf{W}^H\mathbf{W}}^{-1} = \left(\mathbf{W}^H\mathbf{W}\right)^{-1} + \frac{1}{M} \frac{\left(\mathbf{W}^H\mathbf{W}\right)^{-1} \mathbf{W}^H \mathbf{1}_M \mathbf{1}_M^T \mathbf{W} \left(\mathbf{W}^H\mathbf{W}\right)^{-1}}{M - \mathbf{1}_M^T \mathbf{W} \left(\mathbf{W}^H\mathbf{W}\right)^{-1} \mathbf{W}^H \mathbf{1}_M} \tag{21}$$

so the objective function can be seen to be

$$\text{tr } \mathbf{M}^{-1} = c \left(1 + \frac{\text{tr } \mathbf{Q}}{M}\right) + \text{tr } \left(\mathbf{W}^H\mathbf{W}\right)^{-1} \tag{22}$$

where

$$\mathbf{Q} \doteq \left(\mathbf{W}^H\mathbf{W}\right)^{-1} \mathbf{W}^H \mathbf{1}_M \mathbf{1}_M^T \mathbf{W} \left(\mathbf{W}^H\mathbf{W}\right)^{-1} \tag{23}$$

and

$$c \doteq S_{\mathbf{W}^H\mathbf{W}}^{-1} = \left(M - \mathbf{1}_M^T \mathbf{W} \left(\mathbf{W}^H\mathbf{W}\right)^{-1} \mathbf{W}^H \mathbf{1}_M\right)^{-1} \tag{24}$$

In the following, we will minimize separately the two terms in the sum (22).

*1) Minimization of* $c\left(1 + \frac{tr\ \mathbf{Q}}{M}\right)$*:* If we define $\mathbf{j} = \left(\mathbf{W}^H\mathbf{W}\right)^{-1} \mathbf{W}^H \mathbf{1}_M$ then we have that tr $\mathbf{Q} =$ tr $\mathbf{j}\mathbf{j}^H = \|\mathbf{j}\|^2$ so we can write

$$c\left(1 + \frac{\text{tr } \mathbf{Q}}{M}\right) = \frac{1 + \frac{1}{M}\|\mathbf{j}\|^2}{M - \mathbf{1}_M^T \mathbf{W}\mathbf{j}}. \tag{25}$$

The denominator is always positive since

$$\mathbf{1}_M^T \mathbf{W}\mathbf{j} = \mathbf{1}_M^T \mathbf{P}_{\mathbf{W}} \mathbf{1}_M \leq M \tag{26}$$

with $\mathbf{P}_{\mathbf{W}} = \mathbf{W}\left(\mathbf{W}^H\mathbf{W}\right)^{-1}\mathbf{W}^H$ the projection matrix into the subspace spanned by the columns of $\mathbf{W}$, so (25) is clearly minimized when $\mathbf{j} = \mathbf{0}$ or, equivalently $\mathbf{W}^H \mathbf{1}_M = 0$.

Note that this minimization is not affected by the power constraint.

*2) Minimization of tr* $\left(\mathbf{W}^H\mathbf{W}\right)^{-1}$*:* This minimization is going to be affected by the power constraint tr $\mathbf{W}^H\mathbf{W} \leq P$. As tr $\mathbf{A} = \sum_{i=1}^N \lambda_i(\mathbf{A})$ and tr $\mathbf{A}^{-1} = \sum_{i=1}^N \lambda_i\left(\mathbf{A}^{-1}\right) = \sum_{i=1}^N \lambda_i^{-1}(\mathbf{A})$ for $\mathbf{A} \in \mathbb{C}_{N\times N}$ we can state our optimization problem as

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{i=1}^{N_s} \frac{1}{\lambda_i} \\ \text{subject to} & \displaystyle\sum_{i=1}^{N_s} \lambda_i \leq P, \quad -\lambda_i \leq 0 \end{array} \tag{27}$$

which is convex. If we define $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_{N_s}]$, it can be easily proved by using the Karush-Kuhn-Tucker (KKT) conditions, for example, that the optimum value is given by $\boldsymbol{\lambda} = \frac{P}{N_s}\mathbf{1}_{N_s}$, leading to an objective function value of tr $\left(\mathbf{W}^H\mathbf{W}\right)^{-1} = \frac{N_s^2}{P}$.

*3) Putting all together:* From the previous results, if we can find a matrix that meets the following properties

1) The $N_s$ eigenvalues of $\mathbf{W}^H\mathbf{W}$ are all equal to $\frac{P}{N_s}$, and
2) $\mathbf{W}^H\mathbf{1}_M = \mathbf{0}$,

then the optimum training sequence will be given by $\mathbf{W}$. We can write the Singular Value Decomposition (SVD) of the matrix $\mathbf{W}$ as

$$\mathbf{W} = \mathbf{U} \begin{bmatrix} \boldsymbol{\Sigma} \\ \mathbf{0}_{M-N_s\times N_s} \end{bmatrix} \mathbf{V}^H \tag{28}$$

with $\mathbf{U} \in \mathbb{C}_{M\times M}$ and $\mathbf{V} \in \mathbb{C}_{N_s\times N_s}$ unitary matrices, and $\boldsymbol{\Sigma} = \text{diag}\left(\sigma_1 ... \sigma_{N_s}\right)$ is a diagonal matrix containing the nonzero singular values of $\mathbf{W}$.

Now, we have that $\lambda_i\left(\mathbf{W}^H\mathbf{W}\right) = \sigma_i^2$, so property 1 does only depend on the values of the matrix $\boldsymbol{\Sigma}$.

In order to characterize the second property, we can rewrite (28) as the *thin SVD*

$$\mathbf{W} = [\mathbf{U}_1\, \mathbf{U}_2] \begin{bmatrix} \boldsymbol{\Sigma} \\ \mathbf{0}_{M-N_s\times N_s} \end{bmatrix} \mathbf{V}^H = \mathbf{U}_1 \boldsymbol{\Sigma} \mathbf{V}^H \tag{29}$$

with $\mathbf{U}_1 \in \mathbb{C}_{M\times N_s}$ the matrix containing the first $N_s$ columns of $\mathbf{U}$. Condition 2 can be rewritten as $\mathbf{V}^H\boldsymbol{\Sigma}\mathbf{U}_1^H\mathbf{1}_M = \mathbf{0}$. Note that $\mathbf{V}$ and $\boldsymbol{\Sigma}$ are invertible, so the previous condition is equivalent to $\mathbf{U}_1^H\mathbf{1}_M = \mathbf{0}$, that only depends on the submatrix $\mathbf{U}_1$. Therefore, it is possible to find a matrix $\mathbf{W}$ that meets the two conditions at the same time by means of the following procedure:

1) Let $\mathbf{U}$ be an orthonormal base of $\mathbb{C}^M$ with $\frac{1}{\sqrt{M}}\mathbf{1}_M$ as a vector.
2) Choose $N_s$ of the vectors in $\mathbf{U}$ except $\frac{1}{\sqrt{M}}\mathbf{1}_M$. Put them into the matrix $\mathbf{U}_1$.
3) Set $\boldsymbol{\Sigma} = \sqrt{P/N_s}\mathbf{I}_{N_s}$.
4) Let $\mathbf{V}$ be an orthonormal base of $\mathbb{C}^{N_s}$.
5) Obtain the matrix training sequence as $\mathbf{W} = \mathbf{U}_1\boldsymbol{\Sigma}\mathbf{V}^H$.

With this family of training sequences, the matrix $\mathbf{M}$ in (20) is block-diagonal, and the optimization problem in (15) is solved with a value of

$$\text{tr}\ \left(\mathbf{A}^H\mathbf{C}^{-1}\mathbf{A}\right)^{-1} = \sigma^2 N_r \left(\frac{1}{M} + \frac{N_s^2}{P}\right). \quad (30)$$

## IV. Time Varying Channel

In this section we will study a scenario with Time Varying (TV) channels. We will assume that for a given observation period of $M$ samples, the TV channels $\boldsymbol{\alpha} = [\alpha_1 \ldots, \alpha_M]^T$, $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_M]^T$ can be written following a Basis Expansion Model (BEM) as

$$\boldsymbol{\alpha} = \mathbf{F}_\alpha \mathbf{b}_\alpha,\ \boldsymbol{\eta} = \mathbf{F}_\eta \mathbf{b}_\eta, \quad (31)$$

with $\mathbf{b}_\alpha \in \mathbb{C}^{K_\alpha}$ and $\mathbf{b}_\eta \in \mathbb{C}^{K_\eta}$ the BEM coefficients, and $\mathbf{F}_\alpha$ and $\mathbf{F}_\eta$ the matrices containing the $K_\alpha$ and $K_\eta$ first columns of the $M \times M$ Discrete Fourier Transform (DFT) matrix $\mathbf{F}$:

$$\mathbf{F} = \frac{1}{\sqrt{M}} \left[\begin{array}{cccc} \mathbf{f}_0 & \mathbf{f}_1 & \ldots & \mathbf{f}_{M-1} \end{array}\right] \quad (32)$$

with $\mathbf{f}_i = \left[1,\ e^{-j\frac{2\pi i}{M}},\ e^{-j2\frac{2\pi i}{M}},\ \ldots e^{-j(M-1)\frac{2\pi i}{M}}\right]^T$.

Note that $\mathbf{F}_x^H \mathbf{F}_x = \mathbf{I}_{K_x}$ but $\mathbf{F}_x \mathbf{F}_x^H \neq \mathbf{I}_M$ in general, for $x \in \{\eta, \alpha\}$.

Let $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_M]^T$ be the sequence of pre-equalizers applied at the ST, and $\boldsymbol{\Gamma} = \text{diag}(\boldsymbol{\gamma})$. We can write our observation vector $\mathbf{y} = [y_1, \ldots, y_M]^T$ as

$$\mathbf{y} = \boldsymbol{\Gamma}\mathbf{F}_\alpha \mathbf{b}_\alpha + \mathbf{F}_\eta \mathbf{b}_\eta + \mathbf{n} \quad (33)$$

with $\mathbf{n} \sim \mathcal{CN}\left(0, \frac{\sigma^2}{M}\mathbf{I}_M\right)$. The factor $1/M$ is introduced in the noise variance to have a constant *signal to noise ratio per observation*, independently from the value of $M$.

The objective of the ST is to estimate the parameters $\mathbf{b}_\alpha$ and $\mathbf{b}_\eta$ in order to predict the channel. Since (33) is a linear model, an efficient estimator $\hat{\mathbf{b}}$ of $\mathbf{b} \doteq \left[\mathbf{b}_\alpha^T, \mathbf{b}_\eta^T\right]^T$ can be found with a distribution $\hat{\mathbf{b}} \sim \mathcal{CN}(\mathbf{b}, \mathbf{C}_b)$ with

$$\mathbf{C}_b = \frac{\sigma^2}{M}\left(\mathbf{H}^H\mathbf{H}\right)^{-1} \quad (34)$$

and $\mathbf{H} = \left[\begin{array}{cc} \boldsymbol{\Gamma}\mathbf{F}_\alpha & \mathbf{F}_\eta \end{array}\right]$.

In the following we will assume that $M \geq K_\alpha + K_\eta$ so the inverse in (34) exists. The total estimation variance can be written as

$$\frac{\sigma^2}{M}\text{tr}\ \left(\mathbf{H}^H\mathbf{H}\right)^{-1} = \frac{\sigma^2}{M}\text{tr}\left[\begin{array}{cc} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^H & \mathbf{I}_{K_\alpha} \end{array}\right]^{-1} \quad (35)$$

with $\mathbf{A} = \mathbf{F}_\alpha^H \boldsymbol{\Gamma}^H \boldsymbol{\Gamma}\mathbf{F}_\alpha$ and $\mathbf{B} = \mathbf{F}_\alpha^H \boldsymbol{\Gamma}^H \mathbf{F}_\eta$. Note that the trace can be written as a function of the Schur complements of the block matrices as

$$\text{tr}\left[\begin{array}{cc} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^H & \mathbf{I}_{K_\alpha} \end{array}\right]^{-1} = \text{tr}\ \mathbf{S}_{\mathbf{I}_{K_\alpha}}^{-1} + \text{tr}\ \mathbf{S}_{\mathbf{A}}^{-1} \quad (36)$$

that can be writte, following the Woodbury matrix identity, as

$$\mathbf{S}_{\mathbf{I}_{K_\alpha}}^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}^H\left(\mathbf{I}_{K_\alpha} + \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^H\right)^{-1}\mathbf{B}\mathbf{A}^{-1} \quad (37)$$

and

$$\mathbf{S}_{\mathbf{A}}^{-1} = \mathbf{I}_{K_\alpha} + \mathbf{B}\left(\mathbf{A}^{-1} + \mathbf{B}^H\mathbf{B}\right)^{-1}\mathbf{B}^H. \quad (38)$$

Now, we proceed to separately minimize the traces of the inverses of the Schur complements

### A. Minimization of tr $\mathbf{S}_{\mathbf{A}}^{-1}$

Note that

$$\text{tr}\ \mathbf{S}_{\mathbf{A}}^{-1} = K_\alpha + \text{tr}\ \mathbf{B}\left(\mathbf{A}^{-1} + \mathbf{B}^H\mathbf{B}\right)^{-1}\mathbf{B}^H \geq K_\alpha \quad (39)$$

since the second term in the sum is a positive semidefinite matrix. Recall that $\mathbf{B} = \mathbf{F}_\eta^H \boldsymbol{\Gamma}\mathbf{F}_\alpha$, so (39) will be minimized when $\mathbf{B} = \mathbf{0}$. If we set

$$\gamma_i = \sqrt{r}e^{-j\frac{2\pi i K_\eta}{M}},\ i = 1, \ldots, M \quad (40)$$

with $r$ a positive number, we have that

$$\boldsymbol{\Gamma}\mathbf{F}_\alpha = \sqrt{\frac{r}{M}}\left[\begin{array}{cccc} \mathbf{f}_{K_\eta} & \mathbf{f}_{K_\eta+1} & \ldots & \mathbf{f}_{K_\eta+K_\alpha-1} \end{array}\right] \quad (41)$$

so $\mathbf{B} = \mathbf{F}_\eta^H \boldsymbol{\Gamma}\mathbf{F}_\alpha = \mathbf{0}$. In this case, the training sequence $\boldsymbol{\gamma}$ is *modulating* the channel, thus artificially introducing a faster time variation in the channel from the ST to the PR. In this case, the estimator is able to distinguish the lowpass variations of the channel, which correspond to $\boldsymbol{\eta}$, from those variations at higher frequencies, which correspond to $\boldsymbol{\alpha}$. With this training sequence we have that tr $\mathbf{S}_{\mathbf{A}}^{-1} = K_\alpha$.

### B. Minimization of tr $\mathbf{S}_{\mathbf{I}_{K_\alpha}}^{-1}$

Since

$$\text{tr}\ \mathbf{S}_{\mathbf{I}_{K_\alpha}}^{-1} = \text{tr}\ \mathbf{A}^{-1} + \text{tr}\ \mathbf{A}^{-1}\mathbf{B}^H\left(\mathbf{I}_{K_\alpha} + \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^H\right)^{-1}\mathbf{B}\mathbf{A}^{-1} \geq \text{tr}\ \mathbf{A}^{-1}, \quad (42)$$

the second term of the sum is also minimized when $\mathbf{B} = \mathbf{0}$.

We can find an optimum training sequence for the first term tr $\mathbf{A}^{-1} = \text{tr}\ \left(\mathbf{F}^H\boldsymbol{\Gamma}^H\boldsymbol{\Gamma}\mathbf{F}\right)^{-1}$ as follows. Let $\boldsymbol{\Sigma} \doteq \boldsymbol{\Gamma}^H\boldsymbol{\Gamma}$. Since $\mathbf{A}$ is positive definite we have that

$$\text{tr}\ \mathbf{A}^{-1} \geq \sum_{i=1}^{K_\alpha}\frac{1}{a_i} \quad (43)$$

with equality if and only if $\mathbf{A}$ is diagonal, and where

$$a_i = \mathbf{f}_{i-1}^H \boldsymbol{\Sigma}\mathbf{f}_{i-1} = \frac{1}{M}\text{tr}\ \boldsymbol{\Sigma} \quad (44)$$

is the $i$-th element of the diagonal of $\mathbf{A}$. Therefore, the optimum value of $\boldsymbol{\Sigma}$, subject to tr $\boldsymbol{\Sigma} \leq P$ is

$$\boldsymbol{\Sigma} = \frac{P}{M}\mathbf{I}_M \quad (45)$$

so that $\mathbf{A}$ is diagonal and tr $\mathbf{A}^{-1} = \frac{K_\alpha M}{P}$.

### C. Putting all together

It can be easily seen that the optimality conditions obtained in (IV-A) and (IV-B) are compatible, since a training sequence

$$\boldsymbol{\gamma} = \sqrt{\frac{P}{M}}\left[1,\ e^{-j\frac{2\pi K_\eta}{M}},\ \cdots e^{-j(M-1)\frac{2\pi K_\eta}{M}}\right] \quad (46)$$

meets (45) and (40). Therefore, the total estimation variance is $\sigma^2\left(\frac{K_\eta}{M} + \frac{K_\alpha}{P}\right)$.

## V. Results

In Figures 1-3 the estimation variance results for the SISO, MIMO and TV channels are shown. In all cases, dashed lines represent the analytical expression for the MSE with the *On-Off*[1] procedure described in [1], solid lines those of the optimum training sequence, and squares and circles the results of the Monte Carlo simulations in the *On-Off* and optimum cases, respectively. As expected, the optimum training clearly outperforms the *On-Off* one, being this difference specially noticeable in the TV channel.

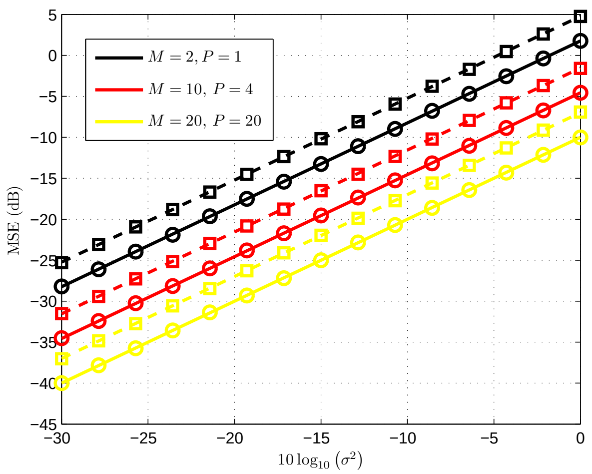


Fig. 1. SISO channel

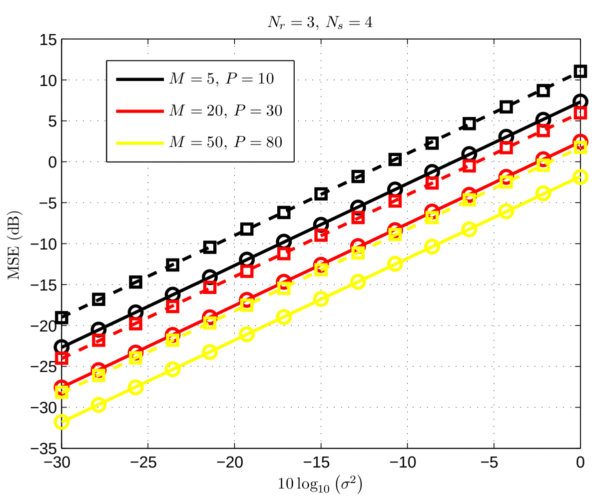


Fig. 2. MIMO channel with $N_r = 3$ antennas at the PR and $N_s = 4$ antennas at the ST.

## VI. Conclusions

In this paper we have studied the problem of acquiring CSI at the transmitter in the cognitive radio channel. We

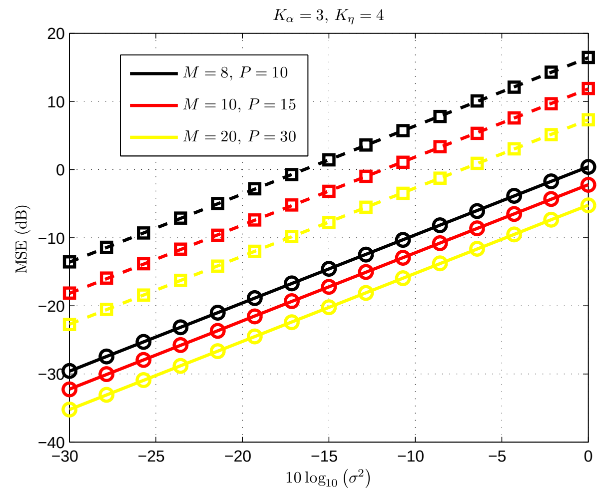


Fig. 3. TV channel with $K_\alpha = 3$ BEM coefficients for the channel from the ST and $K_\eta = 4$ BEM coefficients for that from the PT.

have analyzed from an estimation theory point of view the framework presented in [1] (extended to the MIMO and Time-Varying scenario), derived closed form expressions for the variance of the channel estimators and obtained optimum training sequences that minimize such a variance. The obtained results are consistent with classical pilot design schemes, based on orthogonalization and equal power distribution [9]. Future work includes studying more realistic feedback schemes (quantized SNR, for example), and taking into account the possible *outage* caused by this training procedure, as pointed out in [1].

## References

[1] A. Jovicic and P. Viswanath, "Cognitive radio: An information-theoretic perspective," *Information Theory, IEEE Transactions on*, vol. 55, no. 9, pp. 3945–3958, 2009.

[2] A. Goldsmith, S. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proceedings of the IEEE*, vol. 97, pp. 894–914, May 2009.

[3] N. Devroye, P. Mitran, and V. Tarokh, "Limits on communications in a cognitive radio channel," *Communications Magazine, IEEE*, vol. 44, no. 6, pp. 44–49, 2006.

[4] A. Rico-Alvarino, C. Mosquera, and F. Perez-Gonzalez, "Overlay cognitive transmission in a multicarrier broadcast network with dominant line of sight reception," *Wireless Communications, IEEE Transactions on*, 2012.

[5] A. Rico-Alvarino and C. Mosquera, "Overlay cognitive transmission in ofdm point to point systems exploiting primary feedback," in *Cognitive Information Processing (CIP), 2012 3rd International Workshop on*, pp. 1 –6, may 2012.

[6] S. Srinivasa and S. Jafar, "Cognitive radios for dynamic spectrum access-the throughput potential of cognitive radio: A theoretical perspective," *Communications Magazine, IEEE*, vol. 45, no. 5, pp. 73–79, 2007.

[7] R. Manna, R. Louie, Y. Li, and B. Vucetic, "Cooperative spectrum sharing in cognitive radio networks with multiple antennas," *Signal Processing, IEEE Transactions on*, no. 99, pp. 1–1, 2011.

[8] S. Kay, "Fundamentals of statistical signal processing: Estimation theory, 1993."

[9] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *Signal Processing, IEEE Transactions on*, vol. 51, pp. 1615 – 1624, june 2003.

[1]For the MIMO case, the *On-Off* approach uses only one transmit antenna in each time slot. For the SISO case, when more than two time slots are used, the *On-Off* sequence is of the form $0, \gamma, 0, \gamma \ldots$