LOWER BOUNDS ON THE CHANNEL ESTIMATION ERROR FOR FAST-VARYING FREQUENCY-SELECTIVE RAYLEIGH MIMO CHANNELS

O. Simeone and U. Spagnolini

Dipartimento di Elettronica e Informazione, Politecnico di Milano, Milano (Italy)

ABSTRACT

In this paper, we derive a lower bound on the MSE matrix of training-based channel estimators for MIMO systems over fast-varying fading channels. To this end, we consider an ideal estimator that is able to estimate the long-term features of the channel (e.g., second order statistics, delays,..) with high accuracy while tracking the fast-varying fading fluctuations in optimum (MMSE) way. The bound on the MSE matrix is a valuable tool as a reference to assess the performance of any proposed estimator and it is proved to reduce to known results for simplified settings.

1. INTRODUCTION

Most of the algorithms as well as theoretical analyses devoted to MIMO systems have relied on the assumption of perfect channel estimate (also referred to as perfect channel state information). However, a practical approach has to take into account the effect of estimation errors. A setting of recognized effectiveness in the pursuit of Shannon's capacity sees the transmission organized in bursts, each divided into a training period and (one or more) payload section(s). Training sequences are used to sound the radio channel, that in general can be modelled as frequency-selective. Within this framework, it is of practical and theoretical relevance to derive a lower bound on the mean square error (MSE) matrix of the channel estimate.

The lower bound will be derived for two different multipath MIMO channel models corresponding to different assumptions about the geometry of the antenna arrays and the scatterers. 1) *Beamforming model* [1]: the elements of both the transmitting and receiving antenna arrays are co-located and the scatterers can be considered as point sources as it is reasonable to assume in outdoor environments. Each path of the multipath channel is fully characterized by directions of departure (DOD) and arrival (DOA), a delay and a complex amplitude (fading). The latter is in turn modelled as a temporally correlated Gaussian stationary process. 2) *Diversity model* [2]: the elements of both the transmitting and receiving antenna arrays are not co-located and/or the different scatterers have to be modelled as distributed sources. These assumptions are generally well suited for indoor environments. The amplitudes for each delay of the multipath can be modelled as spatially and temporally correlated jointly Gaussian random variables with zero mean (Rayleigh fading).

To compute a lower bound on the performance of any channel estimator for a frequency-selective MIMO system, we consider some simplifying assumptions. The multipath channel has some long-term characteristics (such as DOA's, DOD's, delays, power-delay profile and temporal correlation function for the beamforming model and delays, spatio-temporal correlation function and power-delay profile for the diversity model) and some fast-varying feature (fading). An ideal estimator should be able to estimate the long-term features of the channel with any accuracy and to track in an optimal (MMSE) way the variations of fast-varying features. By deriving the MSE matrix of the estimate for this ideal method, we therefore set a lower bound on the achievable performance of any estimation algorithm.

2. PROBLEM FORMULATION

We consider a MIMO link with M receiving and N transmitting antennas over a frequency-selective fading channel. After sampling at symbol-rate $1/T$, the baseband discrete-time channel between each transmitting and receiving antenna is assumed to have a temporal support smaller than, or equal to, W samples. The MIMO-FIR channel is estimated by the periodic insertion into the data stream of N training sequences, simultaneously transmitted by each of the N transmitting antennas. The fading variations are sufficiently slow to guarantee that the assumption of a static channel within each training period is reasonably satisfied.

Let $\mathbf{x}[\ell] = [x^{(1)}[\ell], \cdots, x^{(N)}[\ell]]^T$ be the $N \times 1$ vector that stacks the training sequences $\{x^{(n)}[\ell]\}_{\ell=1}^{L}$ for the $n = 1, 2, ..., N$ transmitting antennas and $\mathbf{H}_k[\ell]$ ($\ell = 1, 2, .., W$) the $M \times N$ channel matrix impulse response within the kth training period, the received signal is

$$\mathbf{y}_k[\ell] = \sum_{i=1}^{W} \mathbf{H}_k[i]\mathbf{x}[\ell - i] + \mathbf{n}_k[\ell] \qquad (1)$$

where $\mathbf{n}_k[\ell]$ denotes the additive Gaussian noise (AGN) that is assumed temporally uncorrelated but spatially correlated with covariance matrix $\mathbf{Q}$: $E[\mathbf{n}_k[\ell]\mathbf{n}_k[\ell-i]^H] = \mathbf{Q}\delta[i]$. By stacking the received L samples of the training sequences into a matrix $\mathbf{Y}_k = [\mathbf{y}_k[1], ..., \mathbf{y}_k[L]]$ the model (1) leads to

$$\mathbf{Y}_k = \mathbf{H}_k\mathbf{X} + \mathbf{N}_k, \quad (2)$$

where $\mathbf{H}_k = [\mathbf{H}_k[1], ..., \mathbf{H}_k[W]]$ is the $M \times NW$ MIMO-FIR channel matrix, $\mathbf{X}$ is the $NW \times L$ convolution matrix obtained from the N training sequences so that the ith column of $\mathbf{X}$ is $[\mathbf{x}[i-1]^T, \mathbf{x}[i-2]^T, ..., \mathbf{x}[i-W]^T]^T$, $\mathbf{N}_k$ has the same structure of $\mathbf{Y}_k$ and $1/L \cdot E[\mathbf{N}_k\mathbf{N}_k^H] = \mathbf{Q}$.

For future use, we stack the columns of $\mathbf{H}_k$ into a vector $\mathbf{h}_k = \mathrm{vec}\{\mathbf{H}_k\}$ and the L columns of the matrix $\mathbf{Y}_k$ into a $ML \times 1$ vector $\mathbf{y}_k = \mathrm{vec}\{\mathbf{Y}_k\}$, obtaining from (2)

$$\mathbf{y}_k = (\mathbf{X}^T \otimes \mathbf{I}_M)\mathbf{h}_k + \mathbf{n}_k = \bar{\mathbf{X}}\mathbf{h}_k + \mathbf{n}_k. \quad (3)$$

The AGN is $\mathbf{n}_k = \mathrm{vec}\{\mathbf{N}_k\}$ and the covariance matrix for temporally uncorrelated noise is $E[\mathbf{n}_k\mathbf{n}_k^H] = \mathbf{I}_L \otimes \mathbf{Q}$. Let $\hat{\mathbf{h}}_k$ be the channel estimate, our purpose is to evaluate a bound on the MSE matrix (i.e., correlation matrix of the channel estimation error) $MSE(\hat{\mathbf{h}}_k) = E[(\hat{\mathbf{h}}_k - \mathbf{h}_k)(\hat{\mathbf{h}}_k - \mathbf{h}_k)^H]$. The corresponding MSE of the estimate of $\mathbf{h}_k$ is $MSE_{\hat{\mathbf{h}}} = E[(\hat{\mathbf{h}}_k - \mathbf{h}_k)^H(\hat{\mathbf{h}}_k - \mathbf{h}_k)] = tr\{MSE(\hat{\mathbf{h}}_k)\}$. We recall that the (unconstrained) maximum likelihood (ML) estimate of the channel vector is known to be $\hat{\mathbf{h}}_{ML,k} = \bar{\mathbf{X}}^\dagger\mathbf{y}_k$ and the corresponding MSE matrix $MSE(\hat{\mathbf{h}}_{ML,k}) = ((\mathbf{X}^*\mathbf{X}^T)^{-1} \otimes \mathbf{Q})$ depends on the covariance of AGN and on the correlation properties of the training sequences.

3. CHANNEL MODELS

The $M \times N$ matrix $\mathbf{H}_k[\ell] = \mathbf{H}_k(\ell T)$ denotes the ℓth tap of the MIMO-FIR channel obtained by sampling at symbol-rate ($1/T$) the MIMO channel impulse response. In a multipath environment, the link between the nth transmitting antenna and the mth receiving antenna can be described by a combination of d paths, each characterized by a delay $\tau_{i,k}$, a power $\Omega_{i,k}$ and a normalized amplitude $[\mathbf{A}_{i,k}]_{m,n}$ (i denotes the dependence on the path, $i = 1, 2, ..., d$, and k runs across the training periods). The MIMO channel impulse response is thus a combination of d delayed replica of the known waveform $g(t)$, given by the convolution of the transmitted baseband pulse and the receiving filter: $\mathbf{H}_k(t) = \sum_{i=1}^{d} \sqrt{\Omega_{i,k}} \cdot g(t - \tau_{i,k})\mathbf{A}_{i,k}$. After sampling and ordering the W channel matrixes into the $M \times NW$ MIMO-FIR channel matrix, we get

$$\mathbf{H}_k = \sum_{i=1}^{d} \sqrt{\Omega_{i,k}} \cdot \mathbf{g}(\tau_{i,k})^T \otimes \mathbf{A}_{i,k}, \quad (4)$$

where $\mathbf{g}(\tau)$ is the $W \times 1$ vector that gathers the samples of the waveform delayed by τ. For convenience, we rearrange the entries of the MIMO-FIR matrix $\mathbf{H}_k$ by stacking the columns of the $M \times N$ channel taps $\mathbf{H}_k[\ell]$ (i.e., $\mathcal{H}_k = [\mathbf{h}_k[1], ..., \mathbf{h}_k[W]]$, where $\mathbf{h}_k[\ell] = \mathrm{vec}\{\mathbf{H}_k[\ell]\}$):

$$\mathcal{H}_k = \mathcal{A}_k \cdot \Omega_k \cdot \mathbf{G}(\boldsymbol{\tau}_k)^T \quad (5)$$

where $\mathcal{A}_k = [\mathrm{vec}\{\mathbf{A}_{1,k}\}, \cdots, \mathrm{vec}\{\mathbf{A}_{d,k}\}]$ is $MN \times d$, $\Omega_k = \mathrm{diag}\{\sqrt{\Omega_{1,k}}, ..., \sqrt{\Omega_{d,k}}\}$ and the $W \times d$ matrix $\mathbf{G}(\boldsymbol{\tau}_k)$ $= [\mathbf{g}(\tau_{1,k}), ..., \mathbf{g}(\tau_{d,k})]$ collects all the delayed waveforms. The channel vector can now be obtained as $\mathbf{h}_k = \mathrm{vec}\{\mathcal{H}_k\}$.

3.1. Beamforming model

The ith path is characterized by a DOD $\alpha_i^{(T)}$, a DOA $\alpha_i^{(R)}$ and a complex amplitude (fading) $\beta_{i,k}$ so that

$$\mathbf{A}_{i,k} = \beta_{i,k}\mathbf{a}_R(\alpha_i^{(R)})\mathbf{a}_T(\alpha_i^{(T)})^T \quad (6)$$

where $\mathbf{a}_T(\alpha)$ (or $\mathbf{a}_R(\alpha)$) is the $N \times 1$ (or $M \times 1$) vector containing the array response to a plane wave transmitted (or received) with the angle α. Notice that the two arrays do not need to have the same sensors' arrangement. The normalized faded amplitudes $\beta_{i,k}$ are uncorrelated zero mean circularly symmetric normal (Rayleigh fading) with unit power (i.e., $\boldsymbol{\beta}_k \sim \mathcal{CN}(0, \mathbf{I}_d)$); their temporal correlation across different bursts is the same for all the paths as it accounts for the mobility of the terminal: $E[\beta_{i,k}^*\beta_{i,k+n}] = \rho_n \; \forall i = 1, 2, ..., d$ with $\rho_o = 1$. According to (6), the matrix $\mathcal{A}_k$ can be factorized into the burst-independent term $\mathcal{A}'(\boldsymbol{\alpha}^{(T)}, \boldsymbol{\alpha}^{(R)}) = [\mathbf{a}_T(\alpha_1^{(T)}) \otimes \mathbf{a}_R(\alpha_1^{(R)}), ..., \mathbf{a}_T(\alpha_d^{(T)}) \otimes \mathbf{a}_R(\alpha_d^{(R)})]$ that contains the spatial signatures of the d stationary angles $\boldsymbol{\alpha}^{(T)} = [\alpha_1^{(T)}, \cdots, \alpha_d^{(T)}]^T$ and $\boldsymbol{\alpha}^{(R)} = [\alpha_1^{(R)}, \cdots, \alpha_d^{(R)}]^T$, and the $d \times 1$ vector $\boldsymbol{\beta}_k = [\beta_{1,k}, \cdots, \beta_{d,k}]^T$:

$$\mathcal{A}_k = \mathcal{A}'(\boldsymbol{\alpha}^{(T)}, \boldsymbol{\alpha}^{(R)})) \cdot \mathrm{diag}(\boldsymbol{\beta}_k). \quad (7)$$

The $MNW \times 1$ channel vector $\mathbf{h}_k$ thus reduces to:

$$\begin{aligned}\mathbf{h}_k &= (\mathbf{G}(\boldsymbol{\tau}) \diamond \mathcal{A}'(\boldsymbol{\alpha}^{(T)}, \boldsymbol{\alpha}^{(R)})) \cdot \Omega \cdot \boldsymbol{\beta}_k = \\ &= \mathbf{T}(\boldsymbol{\tau}, \boldsymbol{\alpha}^{(T)}, \boldsymbol{\alpha}^{(R)}, \Omega) \cdot \boldsymbol{\beta}_k, \end{aligned} \quad (8)$$

where $\diamond$ denotes the Khatri-Rao (or columnwise Kronecker product). It is worthwhile to point out that the $MNW \times d$ matrix $\mathbf{T}(\boldsymbol{\tau}, \boldsymbol{\alpha}^{(T)}, \boldsymbol{\alpha}^{(R)}, \Omega) = (\mathbf{G}(\boldsymbol{\tau}) \diamond \mathcal{A}'(\boldsymbol{\alpha}^{(T)}, \boldsymbol{\alpha}^{(R)})) \cdot \Omega$ is independent on the burst and the faded amplitudes $\boldsymbol{\beta}_k$ are correlated across time: $\mathbf{R}_\beta(n) = E[\boldsymbol{\beta}_k\boldsymbol{\beta}_{k-n}^H] = \mathbf{R}_\beta\rho_n = \rho_n\mathbf{I}_d$.

The matrix $\mathbf{T}$ could be rank-deficient as the space-time signatures of different paths having the same DOD's and DOA's and the same delays do not contribute to the matrix $\mathbf{T}$ with linearly independent columns. In practice, the angles (or delays) have to be compared with the array resolution (or the signal resolution) in order to assess the independence of the columns. Let $\mathrm{rank}(\mathbf{T}) = r \leq d$ and let $\mathbf{U}$ be the $MNW \times r$ orthonormal matrix such that $\text{span}\{\mathbf{T}\} = \text{span}\{\mathbf{U}\}$ (i.e., in terms of the SVD of $\mathbf{T}$: $\mathbf{T} = \mathbf{U}\Lambda^{1/2}\mathbf{V}^H$), the channel (8) can be equivalently restated as

$$\mathbf{h}_k = \mathbf{U} \cdot \mathbf{d}_k. \tag{9}$$

The newly defined amplitudes $\mathbf{d}_k$ are mutually uncorrelated

$$\mathbf{R}_{\mathbf{d}}(n) = E[\mathbf{d}_k \mathbf{d}_{k-n}^H] = \Lambda \rho_n \tag{10}$$

with $\Lambda = \text{diag}\{\lambda_1, \lambda_2, ..., \lambda_r\}$ and $\lambda_i = \mu_i[\mathbf{T}\mathbf{T}^H]$, the notation $\mu_i[\cdot]$ denoting the ith eigenvalue of its argument. The matrix Λ represents the power-delay/angle profile that accounts for the power on each vector of the basis $\mathbf{U}$.

3.2. Diversity model

In a rich scattering environment (e.g., indoor), according to the central limit, we can assume $\text{vec}\{\mathbf{A}_{i,k}\} \sim \mathcal{CN}(0, \mathbf{R}_i)$, where the $MN \times MN$ matrix $\mathbf{R}_i$ accounts for the correlation among the transmitting and receiving antennas and it is normalized so that $[\mathbf{R}_i]_{kk} = 1$. The correlation between transmitting and receiving antennas can be considered separable: $\mathbf{R}_i = \mathbf{R}_{T,i} \otimes \mathbf{R}_{R,i}$, where $\mathbf{R}_{T,i}$ and $\mathbf{R}_{R,i}$ are the correlation matrix of the fading at the transmitting and receiving antennas, respectively [2].

According to (5), the channel vector $\mathbf{h}_k$ becomes

$$\mathbf{h}_k = (\mathbf{G}(\boldsymbol{\tau})\Omega \otimes \mathbf{I}_{MN}) \cdot \text{vec}\{\mathcal{A}_k\} = \mathbf{T}(\boldsymbol{\tau}, \Omega) \cdot \boldsymbol{\beta}_k, \tag{11}$$

here $\mathbf{T}(\boldsymbol{\tau}, \Omega)$ is $MNW \times MNd$ matrix composed of the $W \times d$ stationary matrix $\mathbf{G}(\boldsymbol{\tau})\Omega$, and $\boldsymbol{\beta}_k = \text{vec}\{\mathcal{A}_k\} = \text{vec}\{[\mathbf{A}_{1,k}, \cdots, \mathbf{A}_{d,k}]\}$ is a $MNd \times 1$ vector of the faded amplitudes that are spatially (i.e., among sensors) correlated. The correlation of the random amplitudes is $\mathbf{R}_\beta(n) = \mathbf{R}_\beta \rho_n$, where the spatial correlation matrix $\mathbf{R}_\beta$ is block-diagonal, and ρ_n accounts for the temporal correlation. The model simplifies when the fading correlation is independent on the path ($\mathbf{R}_i = \mathbf{R}$ for $\forall i$) as $\mathbf{R}_\beta = \mathbf{I}_d \otimes \mathbf{R}$, where $\mathbf{R}$ can be further assumed to be separable, i.e., $\mathbf{R} = \mathbf{R}_T \otimes \mathbf{R}_R$.

Similarly to the beamforming model, the channel vector can be rewritten by introducing the SVD $\mathbf{G}(\boldsymbol{\tau})\Omega = \mathbf{U}_g \Lambda_g^{1/2} \mathbf{V}_g^H$ obtaining

$$\mathbf{h}_k = (\mathbf{U}_g \otimes \mathbf{I}_{MN}) \cdot \mathbf{d}_k = \mathbf{U} \cdot \mathbf{d}_k, \tag{12}$$

where the correlation of the $r = MNr_g$ amplitudes $\mathbf{d}_k$ depends on the eigenvalues $\mu_i[\mathbf{G}(\boldsymbol{\tau})\Omega^2\mathbf{G}(\boldsymbol{\tau})^T] = [\Lambda_g]_{ii}$ $(i = 1, .., r_g)$ as in

$$\mathbf{R}_{\mathbf{d}}(n) = (\Lambda_g^{1/2}\mathbf{V}_g^H \otimes \mathbf{I}_{MN})\mathbf{R}_\beta(\mathbf{V}_g\Lambda_g^{1/2} \otimes \mathbf{I}_{MN})\rho_n. \tag{13}$$

However, when the correlation is independent on the path the correlation matrix (13) simplifies as

$$\mathbf{R}_{\mathbf{d}}(n) = (\Lambda_g \otimes \mathbf{R})\rho_n. \tag{14}$$

4. LOWER BOUND ON THE MSE MATRIX OF THE CHANNEL ESTIMATE

According to the previous Section, the channel vector for the kth training period can be written as

$$\mathbf{h}_k = \mathbf{T}\boldsymbol{\beta}_k = \mathbf{U} \cdot \mathbf{d}_k, \tag{15}$$

where $\mathbf{U}$ denotes the orthonormal $MNW \times r$ matrix that spans the same subspace as $range\{\mathbf{T}\}$, and $\mathbf{d}_k$ is the $r \times 1$ vector of the burst-varying amplitudes. The long term features of the channel are accounted for by the basis $\mathbf{U}$. This is assumed to be estimated with any degree of accuracy since it is possible to devise a consistent estimator $\hat{\mathbf{U}}$ such that $cov\{\hat{\mathbf{U}}\} \to \mathbf{0}$ for $K \to \infty$. For instance, since $E[\mathbf{h}_k\mathbf{h}_k^H] = \mathbf{T}\mathbf{T}^H$, the leading eigenvectors of the sample correlation matrix $(1/K)\sum_{k=1}^K \mathbf{h}_k\mathbf{h}_k^H$ can be used as an estimate $\hat{\mathbf{U}}$ that is consistent if the fading is asymptotically uncorrelated (i.e., $\rho_n \to 0$ for $n \to \infty$). Notice that this estimate does not require the knowledge of the array manifolds [5]. Alternatively, we could perform a structured estimate of the long-term parameters of the multipath such as angles and delays. A thorough discussion on the estimators is not covered here. In addition, the correlation properties of the faded amplitudes, i.e., $\mathbf{R}_{\mathbf{d}}(n) = \mathbf{R}_{\mathbf{d}}\rho_n$ are known as well.

Recalling (15), the signal model (3) for $K \to \infty$ can be restated in terms of the known matrix $\mathbf{F} = \bar{\mathbf{X}}\mathbf{U}$ as

$$\mathbf{y}_k = \mathbf{F}\mathbf{d}_k + \mathbf{n}_k. \tag{16}$$

Therefore, the task of the ideal channel estimator boils down to the MMSE estimation of the amplitudes $\mathbf{d}_k$. In accordance with our framework, the number of training periods available to the estimator is taken to be $K \to \infty$ so that the frequency domain MMSE theory can be applied. Therefore, the MMSE estimate of the amplitudes in the frequency domain is $\mathcal{F}\{\hat{\mathbf{d}}_k\} = \mathbf{S}_{dy}(\omega)\mathbf{S}_{yy}(\omega)^{-1}\mathcal{F}\{\mathbf{y}_k\}$ where $\mathbf{S}_{dy}(\omega) = \mathcal{F}\{E[\mathbf{d}_k\mathbf{y}_{k-n}^H]\}$ denotes the discrete-time Fourier transform of the crosscorrelation matrix between $\{\mathbf{d}_k\}$ and $\{\mathbf{z}_k\}$ and $\mathbf{S}_{yy}(\omega)$ is similarly defined. The bound on the MSE matrix $MSE(\hat{\mathbf{h}}_k) = \mathbf{U}MSE(\hat{\mathbf{d}}_k)\mathbf{U}^H$ can then be evaluated in closed form for a uniform Doppler spectrum $S_\rho(\omega) = 1/2f_D$ over the support $\omega \in [-2\pi f_D, +2\pi f_D]$. Omitting the straightforward proof, we get the MSE matrix

$$MSE(\hat{\mathbf{h}}_k) = 2f_D \cdot \mathbf{U}(2f_D\mathbf{R}_d^{-1} + \mathbf{R}_w)^{-1}\mathbf{U}^H, \tag{17}$$

where $\mathbf{R}_w = \mathbf{U}^H MSE(\hat{\mathbf{h}}_{ML,k})^{-1}\mathbf{U}$. It follows that the mean square error $MSE_{\hat{\mathbf{h}}} = E[||\hat{\mathbf{h}}_k - \mathbf{h}_k||^2]$ reads

$$MSE_{\hat{\mathbf{h}}} = 2f_D \cdot \sum_{i=1}^r \frac{1}{\mu_i[2f_D\mathbf{R}_d^{-1} + \mathbf{R}_w]}. \tag{18}$$

5. CONCLUDING DISCUSSIONS

The bound (17-18) generalizes some known results related to the performance evaluation of MMSE or ML channel estimation presented in the literature. Some of these connections are discussed below.

1. Static channel ($f_D = 0$): if the channel is static the MSE matrix is $MSE(\hat{\mathbf{h}}_k) = \mathbf{0}$. Indeed, in this case the channel vector is a constant, $\mathbf{h}_k = \bar{\mathbf{u}}$, that can be consistently estimated with covariance $O(1/K)$ by just averaging the ML estimates $\hat{\mathbf{h}}_{ML,k}$ obtained from the different slots.

2. Ideal training and noise ($\mathbf{Q} = \sigma^2\mathbf{I}_M$ *and* $\mathbf{X}^*\mathbf{X}^T = L\sigma_x^2\mathbf{I}_{NW}$): let us consider spatially white noise ($\mathbf{Q} = \sigma^2\mathbf{I}_M$) and ideal training sequences (i.e., orthogonal between any two transmitting antennas and with an impulsive temporal correlation: $\mathbf{X}^*\mathbf{X}^T = L\sigma_x^2\mathbf{I}_{NW}$), in this case $\mathbf{R}_w = L\sigma_x^2/\sigma^2\mathbf{I}_r$.

2.1. Beamforming model: the eigenvalues for the beamforming model can be evaluated as $\mu_i[2f_D\mathbf{R}_d^{-1} + \mathbf{R}_w] = L\sigma_x^2/\sigma^2 + 2f_D/\lambda_i$ for $i = 1, 2, .., r$ so that the MSE becomes

$$MSE_{\hat{\mathbf{h}}_k} = 2f_D \cdot \sum_{i=1}^{r} \frac{\lambda_i}{2f_D + \lambda_i \frac{L\sigma_x^2}{\sigma^2}}. \quad (19)$$

For low SNR (in practice $L\sigma_x^2/\sigma^2 \ll 2f_D/\lambda_i \leq 1/\lambda_i$) the MSE reads

$$MSE_{\hat{\mathbf{h}}_k} \simeq \sum_{i=1}^{r} \lambda_i = MN \sum_{j=1}^{d} \Omega_j \left\|\mathbf{g}(\tau_i)\right\|^2, \quad (20)$$

while for high SNR or small Doppler frequency (i.e., $L\sigma_x^2/\sigma^2 \gg 2f_D/\lambda_i$) the MSE

$$MSE_{\hat{\mathbf{h}}_k} \simeq 2f_D \frac{\sigma^2}{L\sigma_x^2} r \quad (21)$$

is proportional to $r = \text{rank}(\mathbf{T})$, i.e., to the number of parameters to be estimated on each burst. If terminals are moving fast enough, the fading amplitudes are temporally uncorrelated and $f_D = 1/2$. In this case, the MSE (21) for $M = 1$ coincides with the MSE bound on the channel estimation error derived in [5] for a SIMO system.

2.2. Diversity model: the MSE depends on the $r = MNr_g$ eigenvalues of the matrix $2f_D\mathbf{R}_d^{-1} + \mathbf{R}_w$, that can be expressed in terms of the eigenvalues of $\mathbf{R}_T$ and $\mathbf{R}_R$:

$$MSE_{\hat{\mathbf{h}}_k} = 2f_D \sum_{i=1}^{r_g} \sum_{n=1}^{N} \sum_{m=1}^{M} \frac{\lambda_i \mu_n[\mathbf{R}_T]\mu_m[\mathbf{R}_R]}{2f_D + \frac{L\sigma_x^2}{\sigma^2}\lambda_i \mu_n[\mathbf{R}_T]\mu_m[\mathbf{R}_R]}. \quad (22)$$

For low SNR the MSE coincides with (20) derived for the beamforming model. For high SNR it is

$$MSE_{\hat{\mathbf{h}}_k} \simeq 2f_D \frac{\sigma^2}{L\sigma_x^2} MNr_g, \quad (23)$$

showing that, as in the beamforming case, for high SNR the MSE bound depends on the number of parameters to be estimated on each burst (i.e., $r = MNr_g$ parameters). Furthermore, the MSE bound (23) is upper bounded by the worst case of uncorrelated fading ($f_D = 1/2$).

For a frequency-flat fading channel the number of clusters with different delays is $d = 1$ and $W = 1$. In this case $\mathbf{U} = \mathbf{I}_{MN}$, and the MSE matrix is

$$MSE(\hat{\mathbf{h}}_k) = 2f_D \cdot \left(\frac{2f_D}{\Omega}\mathbf{R}^{-1} + \frac{L\sigma_x^2}{\sigma^2}\mathbf{I}_{MN}\right)^{-1}. \quad (24)$$

By further assuming spatially uncorrelated fading ($\mathbf{R} = \mathbf{I}_{MN}$) the MSE reduces to the result in [3]

$$MSE(\hat{\mathbf{h}}_k) = \frac{2f_D\Omega}{2f_D + \Omega\frac{L\sigma_x^2}{\sigma^2}}\mathbf{I}_{MN} \quad (25)$$

$$\Rightarrow \quad MSE_{\hat{\mathbf{h}}_k} = \frac{2f_D\Omega MN}{2f_D + \Omega\frac{L\sigma_x^2}{\sigma^2}}. \quad (26)$$

Moreover, if the faded amplitudes are uncorrelated across different bursts ($f_D = 1/2$) and for large SNR ($\Omega L\sigma_x^2/\sigma^2 \gg 1$) the MSE (25) becomes

$$MSE_{\hat{\mathbf{h}}_k} = \frac{\sigma^2 MN}{L\sigma_x^2}, \quad (27)$$

that coincides with the Cramer Rao Bound bound for the ML estimator carried out on a burst-by-burst basis (see, e.g., [4]).

6. REFERENCES

[1] G. G. Raleigh, J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, pp. 357-366, March 1998.

[2] D. Shiu, G. J. Foschini, M. J. Gans, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 502-513, March 2000.

[3] J. Baltersee, G. Flock, H. Meyr, "Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 2358-2368, Dec. 2001.

[4] T. L. Marzetta, "Blast training: estimating channel characteristics for high capacity space-time wireless," Proceedings of Allerton Conference, Monticello, 1999.

[5] M. Nicoli, O. Simeone, and U. Spagnolini, "Multi-slot estimation of frequency-selective fast-varying channels," to appear in *IEEE Trans. Signal Processing.*