Training structure design optimization for continuous time-varying fading channels

S. Savazzi, U. Spagnolini
Dip. di Elettronica e Informazione, Politecnico di Milano, Piazza L. da Vinci 32, I-20133 Milano, Italy
e-mail: savazzi@elet.polimi.it, spagnoli@elet.polimi.it

***Abstract*— In fast-varying faded channels the transmission can be organized into frames where the channel estimation is mainly training-based. For fast-varying fading channels the training length and training interval can be optimized jointly to maximize the throughput. The optimal balance of training and payload depends on the combination of Doppler frequency and frame length. Here we show that, depending on the degree of mobility for high enough signal to noise ratio, there is a definite advantage in fragmenting the frame with dispersed segments of training symbols of smaller length rather than having a highly reliable channel estimate by concentrating all the training symbols at the beginning of the frame.**

I. INTRODUCTION

In wireless communications, reliable coherent reception is guaranteed as long as the channel estimation exhibits a sufficiently high quality. In most practical systems, the channel gain is estimated by using a fixed number of pilot symbols known to the receiver and multiplexed with the data symbols. Pilot symbols are usually clustered and placed at fixed (periodic) intervals so that the channel variations due to the propagation environment and terminal mobility can be conveniently estimated.

Performance of communication systems is severely limited by channel estimation errors that are typically caused by additive white Gaussian (AWGN) estimation noise as well as by the temporal variations of the channel that make the estimates outdated. A tradeoff between maximizing the channel estimation quality and minimizing the training overhead to maximize the fraction of time spent in payload transmission has to be accounted for.

Optimal training design in fading channels has been considered in [1] for the block-fading case by using information theoretic capacity as a performance metric. In [2] and [3] training is optimized by assuming continuous time-varying fading channels. In [3] MMSE (minimum mean square error) channel estimation is considered based on the current and all past pilot symbols, channel variations are modelled with a Gauss-Markov model [4]. By constraining the percentage of pilot symbols in the data stream it is shown that regular pilot placements minimize the MMSE. The optimal training length and training interval that maximize the throughput for a given target bit error-rate are numerically found in [5] and [6] for ML estimation and block-fading channels as functions of the Doppler frequency (by using Clarke's model of fading).

In this paper we tackle the problem of designing the optimal pilot placement (that is the training length and interval) to maximize system throughput for wireless transmissions over time-varying flat fading channels. Throughput is evaluated in terms of the fraction of time spent in transmission of the information symbols multiplied by the probability of successful transmission of the data stream (e.g., by assuming uncoded BPSK). As a practical constraint and in addition to previous works [2]-[6], the channel is estimated by MMSE interpolation of the channel estimates from the two (closest in time) training sequences at the boundary of an information block.

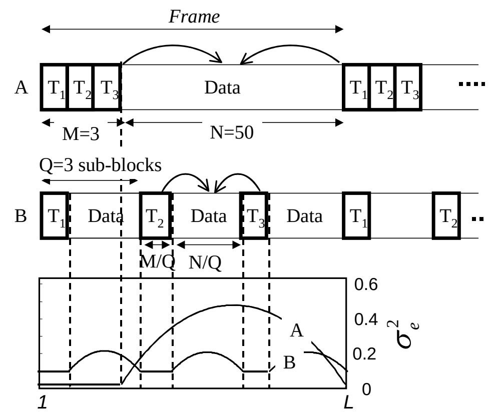

Fig. 1. Training structures and channel estimation mean squared error (MSE): (A) with interpolation among two neighboring training blocks and (B) with interpolaton among pilots that are uniformly spread among the frame

This paper is organized as follows, the system and the channel model are introduced in Sect. II where, for analytical convenience (as done also in [3]), the dynamics of the channel state are modeled by a first order Gauss-Markov process [4]. In Sect. III we evaluate the MSE performances of the MMSE channel estimator in closed form. In Sect. IV we formulate the general problem of optimizing the pilot deployment to maximize the throughput performances. In Sect. V closed form conditions on the optimal training interval and on the maximum achievable training efficiency $\hat{\eta}$ (that is the optimal fraction of symbols that should be reserved for data transmission) are derived in the asymptotically high SNR regime.

II. SYSTEM AND CHANNEL MODEL

We assume a single antenna transmitter sends N information symbols $\mathbf{x} = [x(1), ..., x(N)]$ over a narrowband channel to a single antenna receiver. As shown in figure 1, transmission is organized in frames of L symbols, where M are reserved for channel estimation (pilot symbols) and $N = L - M$ symbols contain information.

The statistical properties of the channel (e.g., the fading distribution) are known and fixed during the frame. Fading is modelled as Rayleigh distributed stationary process. At sampling times $t_k = kT_b$ with $k = 1, ..., L$ (T_b is the symbol period, perfect synchronization is herein assumed), the baseband received signal at the output of the matched filter is collected into an L-length vector $\mathbf{y} = [\mathbf{y}_p^T, \mathbf{y}^T]^T$ where vector $\mathbf{y}_p$ collects the M received pilot symbols $\mathbf{x}_p$ and $\mathbf{y}$ contains the N received data symbols $\mathbf{x}$:

$$\mathbf{y} = \mathbf{H} \cdot [\mathbf{x}_p^T, \mathbf{x}^T]^T + \mathbf{w}. \quad (1)$$

The AWGN noise samples $\mathbf{w} = [w(1), ..., w(L)]$, $\mathbf{w} \sim CN(\mathbf{0}, \sigma^2 \mathbf{I})$ are uncorrelated, $\sigma^2 = N_0/T_b$ and N_0 is the single sided noise power spectral density. Diagonal matrix $\mathbf{H} = diag(\mathbf{h}_p, \mathbf{h})$, with $\mathbf{h} = [h(1), ..., h(N)] \sim CN(\mathbf{0}, \Gamma \mathbf{R}_{N\times N})$ and $\mathbf{h}_p = [h_p(1), ..., h_p(M)] \sim CN(\mathbf{0}, \mathbf{R}_{M\times M})$ collects the baseband equivalent channel gains during the data transmission and the training phase, respectively. Covariance matrices $\mathbf{R}_{N\times N} = E[\mathbf{h}\mathbf{h}^H]$ and $\mathbf{R}_{M\times M} = E[\mathbf{h}_p\mathbf{h}_p^H]$ describe the channel correlation (as for Clarke model) $E[h(k)h^*(k+\ell)] = \Gamma J_0(2\pi f_D T_b \ell)$ [7] among different time samples at time distance ℓT_b, $J_0(\cdot)$ is the zeroth order Bessel function, $f_D = v f_c / c$ (f_c is the carrier frequency and v is the mobile speed) is the maximum Doppler frequency and $\Gamma = E[|h(k)|^2]$ is the average channel power. For notational convenience we define a normalized Doppler frequency as $f_d = f_D T_b$ (that is the maximum Doppler frequency multiplied by the symbol time).

Assuming constant modulus and unitary power signals, the SNR is defined as $\mu = \Gamma/\sigma^2$. Moreover, the training efficiency is the number of information symbols N divided by the total number of symbols in the frame $L = N + M$:

$$\eta = N/L, \quad (2)$$

in particular $M = (1 - \eta)L$ and $N = \eta L$.

AR-1 model of fading. Let the innovation process of the channel state $h(k)$ at time k be modelled by a first-order Gauss-Markov process (first order Auto Regressive or AR-1 [4]) $h(k) = \rho h(k-1) + u(k)$ where $u(k) \sim CN(0, (1-\rho^2)\Gamma)$ is the additive Gaussian driving noise. For bandwidths of $1/T_b \simeq 10kHz$ and Doppler frequency of the order of $f_D = 100\ Hz$, typical values for ρ are between 0.9 and 0.99 [8]. The channel gain at time k is constrained to follow a sequence from a known initial state, say $h(0)$:

$$h(k) = \rho^k h(0) + \sum_{n=0}^{k-1} \rho^n u(k-n). \quad (3)$$

The average channel gain at sampling time k given the initial state $h(0)$ is $E[h(k)|h(0)] = \rho^k h(0)$; the conditional variance is $var[h(k)|h(0)] = \Gamma(1 - \rho^{2k})$. The correlation over ℓ signalling intervals is $E[h(k)h^*(k+\ell)] = \rho^{|\ell|}$ and it depends on the mobility environment (and on the symbol time T_b) at hand.

We define for AR-1 the degree of fading decorrelation between the channel at the beginning and at the end of the frame (after L signalling intervals) as

$$\varepsilon = 1 - E[h(k)h^*(k+\ell)] = 1 - \rho^{|L|}. \quad (4)$$

For f_d small enough (so that channel estimation is still helpful for data decoding [6]) the AR-1 model can be also adapted to fit (from first order Taylor expansion) the Clarke model by letting $\rho = J_0(2\pi f_d) \approx 1 - \frac{1}{4}(\pi f_d)^2$ as in [5].

III. MMSE CHANNEL ESTIMATION

Consider the settings A and B in figure 1 where we assume that the channel gain $h(k)$ that impairs a given information symbol at time k is estimated by interpolating the channel estimates from two different training sequences at the boundaries of an information block. While in setting A the training symbols are clustered into a single block of M symbols, in setting B the same M training symbols are spread over the frame so that the fading is oversampled.

For setting B the frame can be partitioned into $1 < Q \leq \min(M, N)$ sub-blocks (notice that for $Q = 1$ setting B reduces to setting A) each containing N/Q information symbols and M/Q pilot symbols (so that the training efficiency η is the same as for setting A). A high number of sub-blocks causes the fading process to be highly oversampled: this reduces the degradation due to channel outdating at the price of an increased channel estimation error due to the lower number of training samples M/Q. Notice that for a given training efficiency η the sub-blocks are such that each partition contains at least one pilot symbol and one information symbol.

To simplify the analysis at the price of a moderate overestimate of the performance, in what follows the fading will be assumed to be block-fading *only* within the training sequence of M symbols so that $\mathbf{h}_p = h\mathbf{1}^T$ but still varying within the N information symbol (data) block $\mathbf{h} = [h(1), ..., h(N)]$. In practice, this neglects the outdating (or the increased error) that happens within the training block. The ML estimator of the block fading channel is $\hat{h} = \left(\mathbf{x}_p^H \mathbf{x}_p\right)^{-1} \mathbf{x}_p^H \mathbf{y}_p$with $\mathbf{y}_p = h\mathbf{x}_p + \mathbf{w}_p$ and

$$\hat{h} = h + \delta h \sim CN(h, 1/M\mu_p), \quad (5)$$

$\mu_p = \mu/\Gamma$ is the transmit SNR for the pilot symbols. To simplify, the training sequence is assumed as constant modulus so that $\mathbf{x}_p^H \mathbf{x}_p = M$.

A. MSE performances in time-varying fading

The estimated channel at the k-th symbol is modelled as

$$\hat{h}(k) = h(k) + \delta h(k) \quad (6)$$

where for unbiased estimators the MSE is

$$\sigma_e^2(k) = var\left[\hat{h}(k)\right] = E\left[|\delta h(k)|^2\right]. \quad (7)$$

In what follows we analyze the MSE (7) separately for the settings A and B while an illustrative comparison is in Fig. 1 (for $\rho = 0.98$ or equivalently $f_d \approx 0.04$, $\mu = \Gamma/\sigma^2 =$

$\mu_p = 10dB$ ($\Gamma = 1$), $M = 3$ pilot symbols and $N = 50$ data symbols).

Setting A: the channel interpolation $\hat{h}(k)$ is based on the estimated channel values $\hat{h}(0)$ and $\hat{h}(N+1)$ from the training (notice that $\hat{h}(N+1)$ refers to the channel estimate for the upcoming training sequence). The MMSE estimator is

$$\hat{h}(k) = w(\gamma, \alpha)\hat{h}(0) + w(\gamma, 1-\alpha)\hat{h}(N+1) \quad (8)$$

where $k = \alpha N$ and α lies within the support: $\alpha \in [0,1]$, weighting $w(\gamma, \alpha)$ with $\gamma = \rho^{2N}$ is derived in Appendix VIII. Recall that from (5) it is: $\hat{h}(0) = h(0) + \delta h(0)$, $\hat{h}(N+1) = h(N+1) + \delta h(N+1)$. It can be proved (not shown here) that (8) reduces to linear interpolation for all practical values of the Doppler frequency (say $\gamma > 0.5$) with $w(\gamma, \alpha) \approx \alpha$. The conditional mean of $\hat{h}(k)$, $E[\hat{h}(\alpha N)|h(0), h(N+1)]$ coincides with the conditional mean estimator

$$E[\hat{h}(\alpha N)|h(0), h(N+1)] = w(\gamma, \alpha)h(0) + w(\gamma, 1-\alpha)h(N+1) \quad (9)$$

thus showing that the noisy estimates from training do not add bias to the overall interpolator.

The MSE (7) includes the degradation due to channel outdating $var[h(\alpha N)|h(0), h(N+1)]$ (that is derived in the Appendix VIII by assuming perfect channel estimation from training) and due to the channel estimation errors $E[|\delta h(0)|^2] = E[|\delta h(N)|^2] = \left(M\mu_p\right)^{-1}$:

$$\sigma_e^2(\alpha N) = var[\hat{h}(\alpha N)|h(0), h(N+1)] + \left(M\mu_p\right)^{-1}\Psi(\gamma, \alpha) \quad (10)$$

where $\Psi(\gamma, \alpha) = \left[w(\gamma, \alpha)^2 + w(\gamma, 1-\alpha)^2\right]$. The MSE profile (10) for setting A is shown in figure 1. The most critical case (larger variance) for channel interpolation happens in the center of the payload interval ($\alpha = 1/2$) where, if the channel decorrelation is large, the degradation reaches its maximum level. However, in case the fading is almost static between the trainings ($\gamma \simeq 1$) it can be shown that the interpolation improves the channel estimate within the data interval as the interpolator can exploit both the estimates at the boundary.

Setting B: the frame is now partitioned into Q sub-blocks and the data block is of N/Q symbols. As for setting A the channel interpolation $\hat{h}(k)$ at time $k = \alpha N/Q$ (notice that index k now refers to the information symbol position within one sub-block) is a combination of the estimated channel values $\hat{h}(0)$ and $\hat{h}(N/Q+1)$ obtained from M/Q training samples each. The MMSE estimator has the form $\hat{h}(k) = w(\gamma, \alpha)\hat{h}(0) + w(\gamma, 1-\alpha)\hat{h}(N/Q+1)$ with conditional variance (similar to (10))

$$\begin{aligned}\sigma_e^2(\alpha K, Q) = var[\hat{h}(\alpha K)|h(0), h(K)] + \\ + Q\left(M\mu_p\right)^{-1}\Psi(\gamma^{1/Q}, \alpha)\end{aligned} \quad (11)$$

where $K = N/Q + 1$ and

$$var[\hat{h}(\alpha K)|h(0), h(K)] = \Gamma\frac{1 + \gamma^{1/Q} - \gamma^{(1-\alpha)/Q} - \gamma^{\alpha/Q}}{1 - \gamma^{1/Q}}. \quad (12)$$

Notice that the MSE profile (11) is the same as in (10) except for the reduced interval due to fragmentation ($\gamma^{1/Q} \equiv \rho^{2N/Q}$). The MSE profile (11) for setting B is in the example in figure 1. For high Doppler the most critical case still happens in the center of each sub-block $\alpha = 1/2$, however the amount of degradation due to channel outdating is lower compared to setting A as the fading is oversampled with factor $Q > 1$ at the price of a reduced number of training samples M/Q that increase the channel estimation error.

In what follows we tackle the problem of training structure design by evaluating the optimal efficiency η and the number of sub-blocks Q for setting B to maximize a given performance criteria.

IV. Optimization for maximum throughput

The throughput is the cost function to optimize and it is evaluated in terms of the average number of information symbols that are forwarded to the receiver:

$$G_{pkt}(\mu, \eta, \varepsilon, Q, \sigma_e^2) = \eta \times \Pr[N \text{ symbols correctly decoded}], \quad (13)$$

averaging is made with respect to the fading process. The performance criteria to quantify the reliability of communication is the average Frame Error Rate (FER) by assuming an uncoded BPSK system. For a generic number of sub-blocks Q, the bit error probability for a given information symbol depends on the position of the bit (k) within the data-block (for a given sub-block of length N/Q), the SNR (μ), the training sequence length (M/Q) and the fading (de)correlation (ε). Since BPSK is used then $\{\mathbf{x}\}_k = x(k) \in \{-1, +1\}$. The estimated channel (in any possible way) at k-th bit is modelled as in (6) and $\delta h(k) \sim CN(0, \sigma_e^2(k, Q))$.

The received baseband signal at position k is $y(k) = h(k)x(k) + w(k)$, detection of $x(k)$ can be carried out from decision variable [10]

$$z(k) = \mathrm{Re}\left\{y(k)\hat{h}(k)^*\right\} = |h(k)|^2 x(k) + \mathrm{Re}\left[\tilde{w}(k)\right] \quad (14)$$

and $\tilde{w}(k) = h(k)\delta h^*(k)x(k) + w(k)\left[h(k)^* + \delta h(k)^*\right]$. The corresponding conditional BER is $P(E|h(k)) = \Pr[z(k) < 0]$ [10]. Throughput (13) for a given number of sub-blocks Q follows as

$$G_{pkt} = \eta \times \mathbb{E}_{\mathbf{h}}[\prod_{k \in \xi(Q,L,\eta)} (1 - \Pr[z(k) < 0])]. \quad (15)$$

$\xi(Q, L, \eta)$ is the set of bit positions over the frame that contain an information symbol and depends on the frame structure (e.g., the number of sub-blocks Q, L and η), $\prod_{k \in \xi(Q,L,\eta)} [1 - \Pr[z(k) < 0]]$ is the conditional FER. Notice that for BPSK the throughput is so that $G_{pkt} \leq \eta \leq 1$.

The general optimization problem can be stated as maximizing the throughput metric (15) over the training efficiency η and over the number of sub-blocks Q jointly,

$$\begin{aligned}(\hat{\eta}, \hat{Q}) = \arg\max_{\eta, 1 \leq Q \leq L/2} G_{pkt}(\eta, Q|\mu, \varepsilon, \sigma_e^2) \\ \text{s.t. } Q/L \leq \eta \leq 1 - Q/L,\ Q/L \in \mathbb{N}\end{aligned} \quad (16)$$

Rather than solving (16), the optimization can be restated by first tackling the problem of optimizing the number of sub-blocks $\hat{Q}$ for a given training efficiency η, SNR (μ) and fading decorrelation (ε) (or Doppler frequency) so that

$$\text{(Problem 1) } \hat{Q}(\eta) = \arg \max_{1 \leq Q \leq Q_{\max}(\eta)} G_{pkt}(Q|\eta, \mu, \varepsilon, \sigma_e^2)$$
$$\text{s.t. } Q/L,\ Q/\eta L,\ Q/(1-\eta)L \in \mathbb{N} \quad (17)$$

and $Q_{\max}(\eta) = \min((1-\eta), \eta) \cdot L$ is the largest number of frame partitions (sub-blocks) for a given training efficiency and provided that each sub-block has at least one pilot symbol and one information symbol. Notice that the solution $\hat{Q}(\eta)$ shall be a divisor of M, N and L.

Next, using the solution to Problem 1 in (17) the optimal training efficiency $\hat{\eta}$ solution to (16) is

$$\text{(Problem 2) } \hat{\eta} = \arg \max_{1/L \leq \eta \leq 1-1/L} G_{pkt}(\eta, \hat{Q}(\eta)|\mu, \varepsilon, \sigma_e^2) \quad (18)$$

where the optimal fragmentation (number of sub-blocks) is $\hat{Q} = \hat{Q}(\hat{\eta})$.

Figure 2 shows the throughput metric (15) versus the number of sub-blocks Q for a frame of $M = 30$ training symbols and $N = 30$ information symbols (the frame length is $L = 60$ symbols with training efficiency $\eta = 1/2$). The fading realizations are obtained by simulating the Doppler spectrum as done in [11], the fading term is estimated from the training symbols, channel estimate results from linear interpolation between the two neighboring channel estimates. Conditional FER is simulated by evaluating the sign of (14), throughput measure is then averaged over a large enough number of fading realizations. Optimum frame structure analysis is carried out for different values of the normalized maximum Doppler frequency f_d and for a SNR $\mu = 20dB$ (dashed lines) and $\mu = 10dB$ (solid lines). When the Doppler and the SNR are high the channel outdating becomes the main limiting factor while the impact of channel estimation error is almost negligible. In this case it is preferable to have dispersed trainings of smaller length and spread over the frame so that the fading is extremely oversampled and the number of sub-blocks (Problem 1 (17)) is the largest (therefore $\hat{Q}(1/2) = L/2 = 30$ at $\mu = 20dB$ and $f_d > 0.02$). For low Doppler environments ($f_d = 0.01$, 0.02) and low SNR ($\mu = 10dB$) with negligible channel outdating, the channel estimation error is now the main limiting factor. For $f_d = 0.02$ the optimal number of sub-blocks is $\hat{Q}(1/2) = 3$, each burst consists of $30/\hat{Q}$ training symbols followed by $30/\hat{Q}$ information symbols.

In figure 3 we show now the throughput metric versus the training efficiency (solid lines) for a given number of sub-blocks Q ($L = 60$ and $\mu = 20dB$). The maximum normalized Doppler frequency is $f_d = 0.05$. Notice that for a given Q the training efficiencies that can be explored are $Q/L < \eta < 1 - Q/L$ as in (16). At $\mu = 20dB$ the optimal pilot deployment (solution to Problem 2, (18)) requires $\hat{\eta} = 0.75$ and $\hat{Q} = 15$, thus each sub-block consists of $M/\hat{Q} = (1-\hat{\eta})L/\hat{Q} = 1$ pilot symbols followed by $N/\hat{Q} = \hat{\eta}L/\hat{Q} = 3$ information symbols.

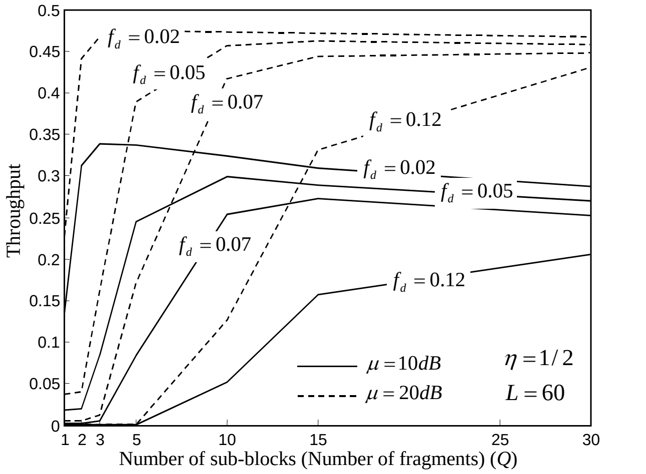

Fig. 2. Throughput metrci versus the number of sub-blocks Q, for a frame structure $L = 60$, training efficiency $\eta = 0.5$ and for different values of the normalized maximum Doppler frequency f_d. The SNR is fixed to $\mu = 20dB$ (dashed lines) and $\mu = 10dB$ (solid lines).

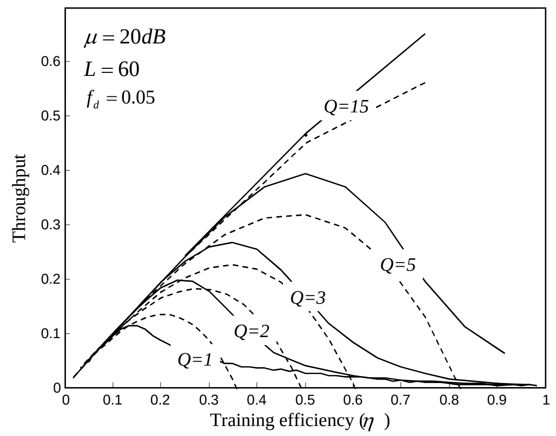

Fig. 3. Optimal training efficiency for a given number of sub-blocks Q, $f_d = 0.05$ $L = 60$ and $\mu = 20dB$ to maximize the throughput (Problem 2 (18)). Dashed lines refer to throughput approximation (22).

V. Asymptotic analysis (SNR$\rightarrow \infty$)

We now approximate the throughput measure (15) for large enough SNR and small Doppler. The dynamics of the channel are now modeled by using the AR-1 model defined in Sect II. The purpose of the section is to cast the optimization problem (16) into the framework of convex optimization. Notice that the main reasonings of the approximations are the follows *i)* the pdf of the impairments in (14) is approximated as a Gaussian random variable so that the conditional error probabilities can be expressed by using the $\mathcal{Q}()$ function as in [10]: in fact, this approximation holds for high SNR and low Doppler as shown in [12]; *ii)* the unconditional FER is approximated for large enough SNR and small Doppler as

$$E_{\mathbf{h}}[\prod_{k\in\xi(Q,\eta)}[1-P(E|h(k))]]\approx[1-\sum_{k\in\xi(Q,\eta)}\mathbb{E}_{h(k)}[P(E|h(k))] \quad (19)$$

In the following notation $\approx$ will used to indicate that the equality holds for fading decorrelation ε small enough.

By assuming that the pdf of the impairments $\tilde{w}(k)$ in (14) is Gaussian and by noticing that $\mathrm{Re}(x)\leq|x|$, the conditional BER can be bounded as done in [10] by $P(E|h(k))\geq\mathcal{Q}\left(\sqrt{2\cdot SNR_k}\right)$ with SNR at the decision variable

$$SNR_k\approx\frac{\mu\cdot|h(k)|^2/\Gamma}{1+\mu\cdot\sigma_e^2(k,Q)} \quad (20)$$

where $\sigma_e^2(k,Q)$ is in (11). The bound is shown in [12] to be tight enough for practical (small) Doppler environments and large enough SNR.

The average BER (wrt fading over different frames but still in the same location k within the frame) scales as [10]

$$\mathbb{E}_{h(k)}[P(E|h(k))]\approx\frac{1}{4\mu}\left(1+\frac{\mu}{\Gamma}\sigma_e^2(k,Q)\right). \quad (21)$$

Notice that in the limit for asymptotically high μ the average BER performances are limited by the channel outdating so that $\mathbb{E}_{h(k)}[P(E|h(k))]>1/4SIR$ where $SIR=\Gamma/\sigma_e^2(k,Q)$ can be regarded as an equivalent signal to interference ratio (SIR) as done in [6].

By substituting (21) into (19) the throughput (15) becomes

$$G_{pkt}\approx\tilde{G}_{pkt}=\eta\left(1-\frac{1}{4\mu}\left[\eta L+Q\cdot\frac{\mu}{\Gamma}\sum_{k=1}^{\eta L/Q}\sigma_e^2(k,Q)\right]\right) \quad (22)$$

the last equality follows from the stationarity of the fading (i.e., the MSE profile is the same for each sub-block, see figure 1, as training sequences are equally split among each frame partition). For AR-1 model of fading $\sigma_e^2(k,Q)$ is defined for setting A,B in (10) and (11), respectively.

Lemma 1: For high SNR such that $\mu\varepsilon>3L$ and ε (4) small enough, the optimal solution to Problem 1 (17) is $(\forall\eta)$

$$\hat{Q}(\eta)=\arg\max_{1\leq Q\leq Q_{\max}(\eta)}G_{pkt}(Q|\eta,\cdot)\approx Q_{\max}(\eta) \quad (23)$$

Therefore if $\varepsilon>0$ (or $f_d>0$) and SNR μ is large enough (depending on the fading decorrelation as $\mu\varepsilon>3L$) then the frame should be partitioned such that the number of sub-blocks is the largest (for a given training efficiency this is $Q_{\max}(\eta)=L\cdot\min((1-\eta),\eta)<L/2$).

Proof: Due to lack of space herein we give the main reasoning of the proof. By using the throughput approximation (22) the Problem 1 (17) reduces to

$$\hat{Q}(\eta)\approx\arg\min_{1\leq Q\leq Q_{\max}(\eta)}\int_0^1\sigma_e^2(\alpha N/Q,Q)d\alpha \quad (24)$$

where, for small enough ε (notice that $\gamma^{\alpha/Q}\approx1-2\alpha\eta\varepsilon/Q$), and recalling that $k=\alpha N/Q$ it is

$$\int_0^1\sigma_e^2(\alpha N/Q,Q)d\alpha\approx\frac{3Q}{4(1-\eta)\mu L}+\frac{\eta\varepsilon}{4Q} \quad (25)$$

minimization of (25) with respect to Q has the straightforward solution

$$\hat{Q}^{(c)}(\eta)\approx\sqrt{L(1-\eta)\eta\mu\varepsilon/3} \quad (26)$$

and $\forall\eta$ it is $\hat{Q}^{(c)}(\eta)>Q_{\max}(\eta)$ for $\mu\varepsilon>3L$. ■

Using the optimal solution $\hat{Q}(\eta)$ in Lemma 1 we maximize now the throughput measure (Problem 2 in (18)) by finding optimal training efficiency η.

Proposition 2: The optimal training efficiency scales with ε (small enough) and for $\varepsilon\mu\geq3L$ as

$$\lim_{\mu\to\infty}\hat{\eta}\approx1-\max(1/L,\sqrt{\varepsilon}/4)=\hat{\eta}_\infty. \quad (27)$$

Moreover, using Lemma 1 the optimal number of sub-blocks is $\lim_{\mu\to\infty}\hat{Q}(\hat{\eta})=Q_{\max}(\lim_{\mu\to\infty}\hat{\eta})\approx\max(1,\frac{L}{4}\sqrt{\varepsilon})$.

Proof: From Lemma 1, $\lim_{\mu\to\infty}\hat{Q}(\eta)\approx Q_{\max}(\eta)$ then

$$\lim_{\mu\to\infty}\hat{\eta}\approx\arg\max_{1/L\leq\eta\leq1-1/L}\lim_{\mu\to\infty}\tilde{G}_{pkt}(\eta,Q_{\max}(\eta)|\cdot). \quad (28)$$

and using (25)

$$\lim_{\mu\to\infty}\hat{\eta}\approx\arg\max_{1/2<\eta<1-1/L}\eta\times\left(1-\frac{\eta^2}{16(1-\eta)}\varepsilon\right) \quad (29)$$

with limiting solution (for $\eta^2\varepsilon\approx\varepsilon$) that scales as in (27). ■

The training efficiency (27) is shown in solid lines in figure 4 versus the fading decorrelation ε. Notice that ε can be mapped onto an equivalent normalized Doppler frequency by using the mapping in sub-figure where we set $\rho=J_0(2\pi f_d)$). Filled circles refer to optimized settings (for exact throughput in (15)) with finite SNR, $\mu=20,10dB$ and $8dB$. We show that the proposed bound is tight enough in predicting the exact performances when the SNR is higher than $20dB$. For $f_d=0.01$ ($\varepsilon\approx0.05$) the maximum achievable training efficiency is $\hat{\eta}_\infty=1-\max(1/L,\sqrt{\varepsilon}/4)=0.95$. Notice that this result is confirmed by simulating the optimal efficiency $\hat{\eta}$ for $\mu=20dB$. Since $L=60$ the optimal number of sub-blocks is the divisor of L that is nearest to $\hat{Q}(\hat{\eta})\approx\max(1,L\sqrt{\varepsilon}/4)$, that is $\hat{Q}=3$ ($\max(1,L\sqrt{\varepsilon}/4)=3.35$). When the fading decorrelation is higher for $f_d=0.05$ ($\varepsilon\approx0.75$) the maximum achievable training efficiency is $\hat{\eta}_\infty=0.83$. For $L=60$ and $\mu=20dB$ the optimal training efficiency is found to be $\hat{\eta}=0.75$. The optimal number of sub-blocks is $\hat{Q}=15$ ($\max(1,L\sqrt{\varepsilon}/4)=12.9$).

VI. Concluding remarks

The problem of optimal pilot placement to maximize system throughput for wireless transmissions over time-varying (flat) fading channels has been tackled in this paper by maximizing the fraction of time spent in transmission of the information symbols multiplied by the probability of successful transmission of the data stream (assuming uncoded BPSK). Practical constraint on the channel estimation processing is included so that the fading term that impairs a given information symbol is estimated by interpolating the channel estimates from the (only) two (closest in time) training sequences at the boundary of an information block. We optimize the length (number of pilot symbols, M/Q) of the training sequence and the training interval (N/Q) with the frame partitioned into Q sub-blocks.

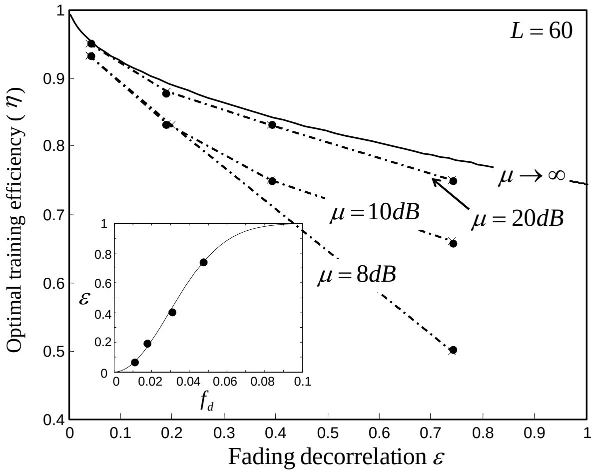

Fig. 4. Training efficiency bound $\hat{\eta}_\infty$ (solid line) versus the fading decorrelation ε (corresponding Doppler frequencies are shown in down left corner sub-figure). Numerical optimization of throughput (15) for varying $\mu = 8dB$, $10dB$ and $20dB$ is shown in black markers.

Optimal pilot deployment results as a tradeoff between minimizing the impact of channel outdating (by choosing an high number of sub-blocks to have the fading process highly oversampled as in setting B) and minimizing the channel estimation error by concentrating all the training symbols (setting A). By approximating the throughput measure for low Doppler, it has been proved (Lemma 1) that when the SNR is high enough (but still finite) there is a definite advantage in fragmenting the frame length with dispersed trainings of one symbol each, rather than having an highly reliable channel estimate by concentrating the pilots. For the same case, we also derived a scaling law (Proposition 2) for the optimal training efficiency and training interval in the limit for asymptotically high SNR. The optimal fraction of symbols to be reserved for training $1-\hat{\eta}_\infty$ depends on the degree of the fading variations (frame decorrelation ε). Extension of training optimization to OFDM and MIMO systems is only a matter of constraining the problem setting to the specific system requirements.

VII. Acknowledgements

The authors would like to acknowledge Prof. L. B. Milstein and Prof. P. C. Cosman for many fruitful discussions during authors' visit to University of California San Diego (UCSD).

VIII. Appendix

Let us assume that we know the channel at the boundary of the information data block of length K, say $h(0)$ and $h(K)$ and we choose $k = \alpha K$ to be within the support: $k \in [0, K]$ or $\alpha \in [0, 1]$. Our purpose is now to evaluate the degree of uncertainty, say $var[\hat{h}(k)|h(0), h(K)]$. By expanding the terms for the probability density function (notice that all the components are Gaussian) it is

$$p(\hat{h}(k)|h(0), h(K)) = \mathcal{K} \times \frac{p(h(k)|h(K)) \cdot p(h(k)|h(0))}{p(h(k))} \tag{30}$$

where $\mathcal{K} = p(h(K))/p(h(K)|h(0))$ do not depend on $h(k)$ and we used the Markov properties of AR and thus $p(h(K)|h(0), h(k)) = p(h(K)|h(k))$ [9]. Since a combination of Gaussian random variables is still Gaussian, we can focus the attention to the exponent of the pdf $p(h(k)|h(0), h(K))$ by neglecting the term $\mathcal{K}$ as it acts as a scaling factor. The probability density function $p(h(k)|h(0), h(K))$ can be written by expanding each product term from (30) as $p(h(k)|h(0), h(K)) \propto \exp(-\phi/2\Gamma)$. From AR-1 model it is clearly $E[h(k)|h(0)] = \rho^{K\alpha}h(0)$, $E[h(k)|h(K)] = \rho^{K(1-\alpha)}h(K)$, $var[h(k)|h(K)] = \Gamma(1 - \rho^{2K(1-\alpha)})$, $var[h(k)|h(0)] = \Gamma(1 - \rho^{2K\alpha})$ and

$$\phi = \frac{\sigma_1^2 + \sigma_2^2 - \sigma_1^2\sigma_2^2}{\sigma_1^2\sigma_2^2}\left[|h|^2 - 2\,\mathrm{Re}\left[\frac{\sigma_2^2 h_K + \sigma_1^2 h_0}{\sigma_1^2 + \sigma_2^2 - \sigma_1^2\sigma_2^2}h\right] + ..\right] \tag{31}$$

with $\sigma_1^2 = 1-\rho^{2K(1-\alpha)} = 1-\gamma^{1-\alpha}$, $\sigma_2^2 = 1-\rho^{2K\alpha} = 1-\gamma^{\alpha}$, $h_0 = \rho^{K\alpha}h(0)$ and $h_K = \rho^{K(1-\alpha)}h(K)$. $\gamma \equiv \rho^{2K}$ is the square of the correlation at the end of the data-frame. The conditional variance is found to be:

$$var[\hat{h}(k)|h(0), h(K)] = \Gamma\left(1 + \gamma - \gamma^{1-\alpha} - \gamma^{\alpha}\right)/(1-\gamma) \tag{32}$$

The conditional mean (i.e., the MMSE estimator[1]) is $E[\hat{h}(\alpha K)|h(0), h(K)] = w(\gamma, \alpha)h(0) + w(\gamma, 1-\alpha)h(K)$ with

$$w(\gamma, \alpha) = \gamma^{\alpha/2}\left(1 - \gamma^{1-\alpha}\right)/(1-\gamma) \tag{33}$$

References

[1] B. Hassibi and B. M. Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links?" *IEEE Trans. on Info. Theroy*, vol. 49, no. 4, pp. 951- April 2003.

[2] J. K. Cavers "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.* vol. 40, pp. 686-693, Nov. 1991.

[3] M. Dong, L. Tong, and B. M. Sadler, "Optimal Insertion of Pilot Symbols for Transmissions Over Time-Varying Flat Fading Channels," *IEEE Trans. on Signal Proc.*, Vol. 52, No. 5, pp. 1403, May 2004

[4] A. Gaston, W. Chriss, E Walker, "A multipath fading simulator for radio," *IEEE Trans. Veh. Technol.* vol. VT-22, pp. 241-244, 1973l.

[5] Q. Sun, Student, D. C. Cox, H. C. Huang, Member, A. Lozano, "Estimation of Continuous Flat Fading MIMO Channels," *IEEE Trans. on Wireless Comm*, Vol. 1, No. 4, pp. 549October 2002.

[6] V. Pohl, P. H. Nguyen, V. Jungnickel, C von Helmolt, "Continuous Flat-Fading MIMO Channels: Achievable Rate and Optimal Length of the Training and Data Phases," *IEEE Trans. on Wireless Comm.* Vol. 4, No 4, pp. 1889, July 2005.

[7] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, vol. CS-11, pp. 360–393, Dec. 1963.

[8] I. Abou-Faycal, M. Medard, and U. Madhow, "Binary adaptive coded pilot symbol assisted modulation over Rayleigh fading channels without feedback," *IEEE Trans. Commun.*, vol. 53, pp. 1036-1046, June 2005.

[9] M.Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall., 2005.

[10] J. Proakis, *Digital Communications*, 4th Ed., McGraw Hill, 2001.

[11] W. C. Jakes Jr., *Microwave Mobile Communications*. New York: Wiley, 1974.

[12] M. A. R. Baissas and A. M. Sayeed, "Pilot-Based Estimation of Time-Varying Multipath Channels for Coherent CDMA Receivers," *IEEE Trans. on Signal Proc*, Vol. 50, No. 8, pp 2037- August 2002.

[1] It can be easily proved that this is the MMSE estimate by the statistical orthogonality with the channel estimation error.