Optimizing Training Lengths and Training Intervals in Time-Varying Fading Channels

Stefano Savazzi, *Student Member, IEEE*, and Umberto Spagnolini, *Senior Member, IEEE*

***Abstract*—In time-varying faded channels the transmissions are organized into frames where the channel estimation is mainly training-based. The optimal design of the training structure is formulated here by finding the training length (the optimal number of contiguous pilots) and the training interval (the interval among two successive training phases) to maximize system throughput. The optimal balance of training and payload depends on the combination of Doppler frequency and frame length. The level of the signal to noise ratio and the fading dynamics constrain the quality of the estimate from training. It is shown that the length of the training can be conveniently traded for lower training intervals to reduce the estimate out-dating. For fast-varying fading and 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. Extensive simulations corroborate the design criteria. System throughput is maximized either for noisy binary transmission and for Gaussian input symbol distribution (i.e., by using information theoretic analysis).**

***Index Terms*—Gauss–Markov fading, MMSE channel estimation, throughput optimization, time-varying fading channels, training based channel estimation.**

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 receiver acquires the channel dynamics from known sequences of symbols (pilot symbols) transmitted during the training phase, information symbols are then estimated based on the acquired channel [1]. Pilot symbols are multiplexed with the information symbols by clustering and placing them at fixed (periodic) intervals so that the channel variations due to propagation environment and terminal mobility can be conveniently estimated. In single carrier systems these pilots are spread in time and they are inserted into the data stream according to a frame structure. Similarly, in multicarrier systems (e.g., OFDM) pilots are spread either in time and in frequency according to a specific time-frequency pattern.

Manuscript received February 08, 2008; revised October 10, 2008. First published November 11, 2008; current version published February 11, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Zhengyuan (Daniel) Xu. The material in this paper was presented in part at the Wireless Communication and Networking Conference (WCNC), Las Vegas, NV, April 2008.

The authors are with the Dipartimento di Elettronica e Informazione, Politecnico di Milano, I-20133 Milano, Italy (e-mail: savazzi@elet.polimi.it, spagnoli@elet.polimi.it).

Digital Object Identifier 10.1109/TSP.2008.2009270

Performance of communication systems is known to be severely limited by channel estimation errors due to noise as well as to the temporal variations of the channel thus making the estimates to be outdated. In this paper we tackle the general problem of designing the optimal pilot density for wireless transmissions over time-varying flat fading channels. Given that pilot placement can be considered as sampling of fading variations, the quality of the channel estimate is ruled by the density and the accuracy of the channel sampling. The optimal design is a tradeoff between maximizing the channel estimation quality and minimizing the training overhead to optimize the fraction of time spent in payload transmission.

Let the transmission be organized into frames of L symbols as shown in Fig. 1, a fraction of M symbols are reserved for channel estimation (pilot symbols) and $N = L - M$ symbols contain the information. The training efficiency $\eta = N/L$ is the fraction of time spent in transmission of information symbols. For the same efficiency η, the frame can be subdivided into an (arbitrary) number (Q) of sub-blocks consisting of a number of training symbols (M/Q) followed by information symbols (N/Q). The number of pilots (M/Q) within each sub-block and the training interval (L/Q), that is the time interval among two successive training phases, are here *jointly* optimized. The system throughput used as metric depends on training efficiency (η) and the probability of successful transmission of the data stream (in case of BPSK modulation). Alternatively, it can be used the average mutual information provided by the channel (in case of Gaussian input symbol distribution). This optimum criterion balances the use of training (M) and information payload (N) according to the combination of Doppler frequency and frame length. Minimum mean squared error (MMSE) channel estimation is exploited to further optimize performances and minimize the effects of channel outdating. The comparison in term of optimized throughput proves that as soon as the communication signal-to-noise ratio (SNR) is high enough and the Doppler (or the channel out-dating) becomes the limiting factor, there is a definite advantage in fragmenting the block of M training symbols within the frame $(Q > 1)$ into training sub-blocks of smaller length (as in pilot assisted modulation PSAM [1]) rather than having an highly reliable channel estimate by concentrating all the M training symbols as for $Q = 1$.

A. Related Works

Optimal training design in fading channels is a largely investigated topic. Early studies on the effects of training on multi-antenna channel capacity are in [2]. Therein it is shown that,

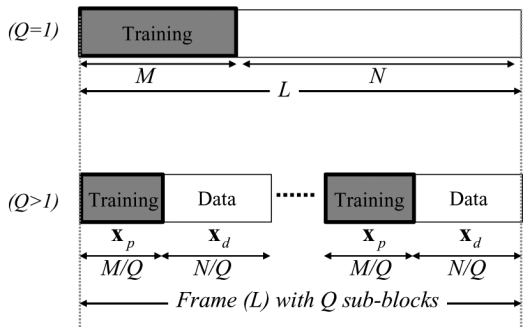

Fig. 1. Design parameters definitions and framing/sub-block partition structure.

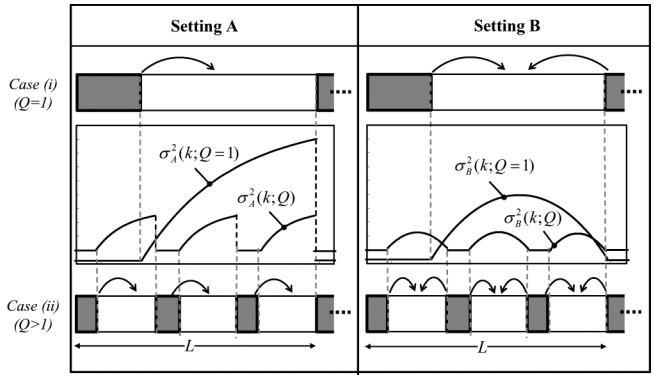

Fig. 2. Sample and hold channel estimate (Setting A) and interpolation of the channel estimates among two neighboring training blocks (Setting B). The same number of pilots can be either clustered at the beginning of each frame (Case i) or uniformly spread among the frame according to a given sub-block partition (Q) (Case ii). For $Q \geq 1$ the MSE for the settings A and B are superimposed to the frame structure.

under the block fading assumption,[1] it is convenient to spend no more than half the coherence interval in training. For block fading channels, the problem of designing the training interval is also tackled in [3], [4] (and [5]) for the multiple antenna transmission case by using a lowerbound to the information theoretic capacity as performance metric. The mutual information of the channel is expressed as a function of the channel estimation error. The mutual information can be maximized with respect to the length of the training as the estimation error is governed by the number of contiguous pilots in the training sequence. It is found that the optimal length of the training sequence still converges to half of the block length in the limit of SNR approaching to zero [5]. Moreover, if transmit powers for pilot and data symbols can be optimized, then the optimal number of training symbols is always equal to the number of transmit antennas. When training and data have the same power, the training efficiency η is lower [3].

The problem of joint optimizing the interval among subsequent training phases and the number of pilots for each phase arises when continuous time-varying fading is assumed (or equivalently when the channel estimate out-dates due to time-varying fading). By limiting the analysis to the case of one pilot symbol per training $(M/Q = 1)$, the problem of designing the interval among pilots to optimize a given criteria was tackled for various optimization metrics and channel estimation techniques, see [6]–[8]. Using the channel capacity as performance criteria, in [6] the training interval is optimized by assuming time-varying fading channels (for varying Doppler spreads) and MMSE channel estimation based on all past pilots symbols. Results are extended in [7] for multiple antennas scenario. Still by constraining one pilot per training sequence, the joint design of the training interval and of the pilot symbol transmit power is considered in [8] by optimizing the system throughput and in [9] by maximizing the cut-off rate for binary coded modulation and Gauss-Markov fading modeling.

A different approach is tackled in [10] where the design of the optimal pilot placement for time-varying fading is considered by constraining the fraction of pilot symbols in the data stream (e.g., training efficiency η) with MMSE channel estimator employed by Kalman filtering (channel variations are modeled by using a Gauss-Markov model [15]). It is shown that regular pilot placements can minimize the channel MSE.

By using a simple maximum likelihood (ML) estimator for block fading channels with orthogonal training sequences, in [11] and [12] the joint design of training length and training interval that maximize the percentage of information symbols within the frame is given numerically to satisfy a (given) target bit error-rate. Optimal designs are found as functions of the Doppler frequency (thus by using Clarke model for the fading dynamics) and of the number of antennas as for multiple input multiple output MIMO systems. Impact of estimators (other than ML) on the pilot placement design was not discussed.

B. Paper Organization and Contributions

In what follows we outline the original contributions of the paper as compared to the existing works, paper organization is given as well.

i) The dynamics of the channel state are introduced in Section II with system model. These are modeled by a first order Gauss-Markov process (similarly to [10]) as it is a simple and fairly general model to characterize the fading process [15], [25]. As a practical constraint, the channel over data block is estimated from training either by interpolation of the channel estimates from the two (closest in time) training sequences (referred as setting B, see Fig. 2) or by maintaining the channel estimate from the previous training sequence as unaltered (except for a scaling factor as for MMSE estimation) for the whole payload symbols [6]–[12] (referred as setting A, see Fig. 2). Both channel estimation techniques are described in Appendix C where we derive the MMSE performances in closed form as a function of the number of training symbols, the training interval and the AR-1 parameters (or equivalently the Doppler frequency).

ii) The problem of optimizing regular pilot deployment to maximize the throughput metrics is formulated in Section IV by using a successive optimization method. We *first* optimize the training interval (that is the optimal time distance between two successive training phases)

[1]The channel is assumed to be time-invariant within the frame or equivalently the length of the frame is small compared to the coherence time of the channel.

as in [10] for a given number of training symbols or equivalently for a given training efficiency $\eta = 1 - M/L$ (Problem 1), *next* we come out with the general problem of finding the optimal training efficiency $\hat{\eta}$ (e.g., the optimal number of pilots and training interval) that maximize the overall throughput (Problem 2). As performance metric, we compare two different choices: throughput for uncoded BPSK modulation (Section V) and system capacity for Gaussian input symbol distribution (Section VI). Differently from [11] and [12], the proposed throughput metric better highlights the fundamental tradeoff between training length and interval as it considers the impact of the SNR (through the BER) and of the fading dynamics (through the MSE of the channel estimate) jointly, without treating the BER as an external parameter in the optimization.

iii) By constraining the input symbol distribution to be binary valued (using BPSK modulation) and by approximating the performance criteria for high SNR and small enough Doppler (i.e., the channel variations during one frame period are small) throughput optimization can be cast into a convex optimization. Closed form conditions on the optimal training interval (solution to Problem 1) and on the training efficiency $\hat{\eta}$ (solution to Problem 2) are derived in Section V-B to maximize the probability of successful binary transmission for the MMSE channel estimation in settings A and B. An analytical framework is given herein to provide insights to the interplay between the optimal length (M/Q) and interval (L/Q) of the trainings as functions of the SNR and of the frame decorrelation. Moreover, we quantify for various settings how much gain is provided by the combination of the channel estimates from the previous and the upcoming training sequence (setting B) compared to the simpler channel sampling and hold strategy (setting A).

iv) Without limiting the analysis to BPSK modulation, in Section VI optimal training length and interval are derived to maximize the average mutual information similarly to [12]. Problem formulation is the same as in Section IV while input symbol distribution is now constrained to be Gaussian. The throughput [6] is approximated for high SNR and small enough Doppler to still have a convex optimization as for uncoded BPSK modulation (Section V).

To corroborate the analytical results, numerical simulations for BPSK transmissions are given in Sections V-A and V-B-I for each considered setting, while numerical results for capacity maximization are in Section VI-B.

II. System and Channel Model

Let the transmitter send N information symbols $\mathbf{x} = [x(1), \ldots, x(N)]^T$ over a frequency flat channel (see Fig. 1). Training design parameters for the problem at hand are: i) the training length (M) or training efficiency $(\eta = N/(N+M))$, notice that $M = (1-\eta)L$ and $N = \eta L$; ii) the number of sub-blocks (Q) that reduce the interval (L/Q) among two successive trainings.

Fading is stationary over frames so that correlation properties can be acquired from channel estimates over multiple sub-blocks. At sampling times $t_k = kT_b$ with $k = 1, \ldots, L$ (T_b is the symbol period, perfect synchronization is assumed herein), the baseband received signal at the output of the matched filter is collected into vector $\left[\mathbf{y}_p^T, \mathbf{y}^T\right]^T$ that contains the L/Q symbols received from any sub-block. The M/Q-length vector $\mathbf{y}_p$ collects the M/Q received pilot symbols $\mathbf{x}_p$ and the N/Q-length vector $\mathbf{y}$ contains the N/Q received data symbols $\mathbf{x}$:

$$\left[\mathbf{y}_p^T, \mathbf{y}^T\right]^T = \mathbf{H} \cdot \left[\mathbf{x}_p^T, \mathbf{x}^T\right]^T + \mathbf{w}. \tag{1}$$

The uncorrelated AWGN samples are $\mathbf{w} = [w(1), \ldots, w(L/Q)] = \left[\mathbf{w}_p^T, \mathbf{w}^T\right]$ with $\left[\mathbf{w}_p^T, \mathbf{w}^T\right] \sim CN\left(\mathbf{0}, \sigma_w^2 \mathbf{I}\right)$ and where $\sigma_w^2 = N_0/T_b$ with single-sided noise power spectral density N_0. Diagonal matrix $\mathbf{H} = \mathrm{diag}(\mathbf{h}_p, \mathbf{h})$ collects the baseband equivalent Rayleigh fading channel gains during the data transmission, $\mathbf{h} \sim CN(\mathbf{0}, \Gamma\mathbf{R})$, and the training phase, $\mathbf{h}_p \sim CN(\mathbf{0}, \Gamma\mathbf{R}_p)$, within one sub-block. Channel correlation embedded in covariance matrices $\mathbf{R} = \mathbb{E}[\mathbf{h}\mathbf{h}^H]/\Gamma$ and $\mathbf{R}_p = \mathbb{E}\left[\mathbf{h}_p\mathbf{h}_p^H\right]/\Gamma$ depends upon the propagation environment and the terminal mobility.

Symbols are independent random variables with unitary power ($\mathbb{E}[|x|^2] = 1$), the (average) SNR can be defined as

$$\mu = \frac{\Gamma}{\sigma_w^2} \tag{2}$$

scaling $\Gamma = E[|h|^2]$ is the channel power.

A. Auto Regressive Model of Fading

To develop an analytic framework that can help in providing insights to training optimization problem, let the innovation process of the channel state at sampling time k be modeled by first-order Gauss-Markov process (Auto Regressive AR-1 [20]). Channel samples are

$$h(k) = \rho h(k-1) + u(k) \tag{3}$$

where $u(k) \sim CN(0, (1-\rho^2)\Gamma)$ is the white Gaussian driving noise. For bandwidths in the 10 kHz range and Doppler spreads in the order of $f_D = vf_c/c \leq 100$ Hz (f_c is the carrier frequency, v is the mobile speed), AR-1 modeling holds true with typical values for ρ between $\rho = 1$ (block fading) and $\rho = 0.9$[14]. The fading correlation over d signaling interval is $E\left[h(k)h^*(k+d)\right]/\Gamma = \rho^{|d|}$ and it depends on the mobility environment (and on the symbol time T_b) at hand. Notice that when fading dynamics are small, AR-1 can be adapted to fit with other fading correlation models (e.g., Clarke model) by simply adjusting the parameter ρ.

The degree of fading variations within the frame are described by the frame correlation. As a design parameter for training optimization, in this paper we refer to the fading decorrelation over the frame length (after L signaling intervals)

$$\varepsilon = 1 - \rho^L. \tag{4}$$

Notice that training sequences are usually densely placed so that decorrelation is $\varepsilon \ll 1$.

III. Channel Estimation

Depending on the computational complexity of the channel estimation and on the delay constraints, there are different training structures and estimation methodologies [10]. The most simple (and suboptimal) one is referred as the setting A (case i) depicted in Fig. 2: a single training block ($Q = 1$) of M symbols is used to perform the channel estimation at one end of the training sequence, while the same channel estimate is maintained (except for a scaling factor to have an unbiased estimator, see, e.g., [9] and [13]) as if the fading would be fixed for the remaining N symbols of the data block. By using the same channel estimation technique, the frame might be also partitioned (case i) into $1 < Q \leq \min(M, N)$ sub-blocks of length L/Q each containing N/Q information symbols and M/Q pilot symbols. This is done so that each sub-block can preserve training efficiency η with the same number of pilot and information symbols. Dispersed trainings of smaller length are spread over the frame so that the fading is extremely oversampled for large Q. This reduces the degradation due to channel outdating at the price of an increased channel estimation error for the lower number of training samples M/Q. Notice that for a given training efficiency η the maximum number of sub-blocks Q is $Q_{\max} = \min(M, N) \leq L/2$ and it is such that each partition contains at least one pilot symbol and one information symbol.

More sophisticated interpolation techniques can be used to cope with the large MSE that is caused by channel outdating at the end of data block. Setting B assumes 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 boundary of an information block. Notice that acquiring the channel estimate through interpolation from two sequences of pilots improves the performance at the cost of additional delay experienced in detecting the data symbols. As for setting A, the frame might be now composed by one ($Q = 1$) block of N information symbols (case i) or by $Q > 1$ sub-blocks (case ii). If $Q = 1$ the channel estimate at time k, say $\hat{h}(k)$, results from a combination of the channel estimates obtained from the previous and the upcoming (belonging to the next frame) training sequences. Instead, in case the frame can be partitioned into Q sub-blocks each containing N/Q information symbols and M/Q pilot symbols (still preserving the same η), the channel estimation during data transmission is obtained by combining the two estimates from neighboring training sequences.

A. Channel Estimation From Training (Overview)

In what follows the fading will be assumed to be block-fading *only* within the training sequence so that for each sub-block $\mathbf{h}_p = [h_p(1), \ldots, h_p(M/Q)] = h_p\mathbf{1}$, but it is still varying within the N/Q information symbol (data) block $\mathbf{h} = [h(1), \ldots, h(N/Q)]^T$. The block-fading assumption within the training interval simplifies the analysis at the price of an optimistic estimation error, the corresponding performance degradation is quantified in Appendix A. Motivated by practical assumptions (small end-frame decorrelation, $\varepsilon \ll 1$), the AR-1 fading model is considered in the following for analytical derivation.

According to the block-fading assumption, let the channel gain be constant over the training sequence of M/Q symbols (still randomly varying among each sub-block) so that from (1) the received signal within one training sequence is

$$\mathbf{y}_p = h_p\mathbf{x}_p + \mathbf{w}_p. \tag{5}$$

To simplify the reasoning, the training sequence is considered as constant modulus so that $\mathbf{x}_p^H\mathbf{x}_p = M/Q$. The reader might refer to [16], [17] where the issue of optimal training sequence design is investigated. The ML estimator of the channel value is [13]

$$\hat{h}_p = \left(\mathbf{x}_p^H\mathbf{x}_p\right)^{-1}\mathbf{x}_p^H\mathbf{y}_p = h_p + \delta h_p \tag{6}$$

where $\mathbb{E}[\hat{h}_p] = h_p$ (as for unbiased estimator), $\delta h_p = \left(\mathbf{x}_p^H\mathbf{x}_p\right)^{-1}\mathbf{x}_p^H\mathbf{w}_p \sim CN(0, Q/(M\mu_p))$, with $\mu_p = 1/\sigma_w^2$ being the transmit SNR referred to the pilot symbols. The assumption of block-fading (5) yields to a better estimate (i.e., with lower variance) compared to the estimate that accounts for the channel variation within the training sequence.

As proved in Appendix A, if fading is time-varying over the pilots, the channel estimator (6) is biased at the edges of the training block and there is an increased error so that $\delta h_p \sim CN(0, Q/(\tilde{M}\mu_p))$, with $\tilde{M}/Q < M/Q$ and

$$\frac{\tilde{M}}{Q} \approx \max\left[1, \frac{\frac{M}{Q}}{1 + \frac{(1-\rho^2)}{6}\Gamma\mu_p\left(\frac{2M}{Q} - 1\right)\left(\frac{M}{Q} - 1\right)}\right]. \tag{7}$$

Here $\tilde{M}/Q$ can be regarded as the equivalent number of pilots in time-varying fading to have the same performances of block-fading. In the following, notation $\approx$ will used to indicate that the equality holds for small enough ε (say $\varepsilon \ll 1$ or equivalently $\rho \simeq 1$). Approximation (7) is corroborated by numerical analysis in Fig. 3: the equivalent number of pilots $\tilde{M}/Q$ for time-varying fading over the training are shown versus the fading correlation ρ and the training block length (M/Q). Approximation (7) is in dashed lines. Performance loss in terms of $\tilde{M}/Q < M/Q$ is negligible as far as the channel is fairly stable within the M/Q samples of the training sequence ($\rho < 0.99$) and for a low number of pilots.

B. MMSE Channel Estimation Performances in Time-Varying Fading

The estimated channel at the kth information symbol of a given sub-block is modeled as

$$\hat{h}(k) = h(k) + \delta h(k) \tag{8}$$

where for unbiased estimators it is

$$\sigma^2(k; Q) = \mathbb{E}\left[\left|h(k) - \hat{h}(k)\right|^2\right] = \mathbb{E}\left[\left|\delta h(k)\right|^2\right]. \tag{9}$$

Therein we evaluate the MSE (9) separately for the settings A and B with varying number of sub-blocks Q, while

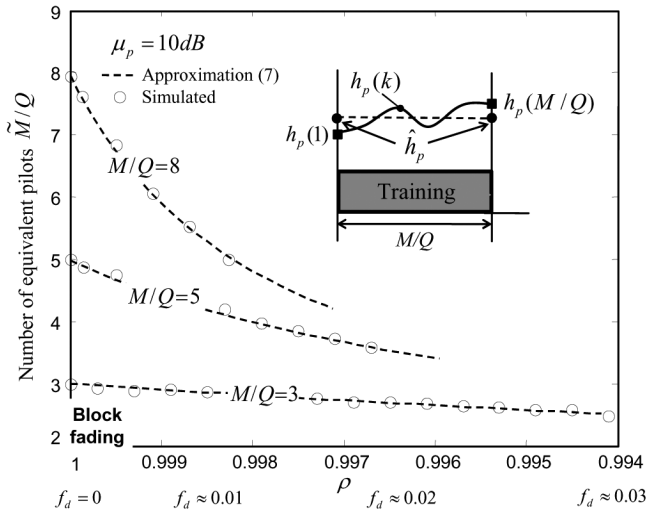

Fig. 3. Equivalent number of pilots $\tilde{M}/Q$ for time-varying fading over the training sequence versus the fading correlation ρ and the training block length (M/Q). Dashed lines refer to the approximation (7).

an illustrative comparison among different strategies is superimposed in Fig. 2 (for $\rho \approx 0.985$ (or $f_d \approx 0.04$) $\mu = \Gamma/\sigma^2 = \mu_p = 10$ dB ($\Gamma = 1$), $M = 3$ pilot symbols and $N = 50$ data symbols). As shown in Fig. 4, we refer to $\hat{h}_p^{(L)} \sim CN(h_p(M/Q), Q/\tilde{M}\mu_p)$ as the estimate of the channel drawn from the training sequence *before* the data block ($h_p(M/Q)$ is the fading term at the edge of the training), while $\hat{h}_p^{(R)} \sim CN(h_p(1), Q/\tilde{M}\mu_p)$ is the estimate (now of $h_p(1)$) obtained from the *upcoming* training block. In both cases estimators are in (6).

i) **Setting A.** The channel estimator $\hat{h}_A(k)$ at signaling interval k, with $k = 1, \ldots, N/Q$ (notice that index k refers to the information symbol position within one sub-block) is

$$\hat{h}_A(k) = \rho^k \times \hat{h}_p^{(L)}, \tag{10}$$

the channel estimate $\hat{h}_p^{(L)}$ from previous training sequence is practically maintained (except for a scaling factor) for the whole sub-block until a new training sequence is received. The MSE profile depends both on the channel outdating that is ruled by the interval N/Q among two trainings and on the channel estimation error (6) that is constrained by the training length. Let $k = \alpha N/Q$ with $\alpha \in [0, 1]$, the MSE (9) is function of the number of sub-blocks as

$$\sigma_A^2\left(\frac{\alpha N}{Q}; Q\right) = \Gamma\left[\left(1 - \gamma^{\alpha/Q}\right) + \frac{Q \cdot \gamma^{\alpha/Q}}{\mu\tilde{M}}\right] \tag{11}$$

where $\gamma \equiv \rho^{2N} = \mathbb{E}[h^*(0)h(N)]^2$ is the (square of the) frame correlation (or equivalently $\gamma = \left[(1-\varepsilon)^2/\rho^{2M}\right]$ being ε the frame decorrelation. The MSE profile (11) is illustrated on the left-hand side of Fig. 2 for the case of one single sub-block $Q = 1$ [$\sigma_A^2(\alpha N/Q; Q = 1)$, (case i)] and multiple sub-blocks $Q > 1$ [$\sigma_A^2(\alpha N/Q; Q)$, (case ii)]. For large decorrelation (that implies high Doppler) the largest MSE is at the end of each sub-block (or for $\alpha = 1$), for $Q > 1$ the amount of degradation due to channel out-dating is lower compared to $Q = 1$ as the fading is now oversampled with a larger factor Q. However, the reduced number of training samples M/Q cause a degradation of the channel estimation quality and thus higher estimation MSE.

ii) **Setting B.** The channel interpolation $\hat{h}_B(k)$ is now based on the estimated channel values $\hat{h}_p^{(L)}$ and $\hat{h}_p^{(R)}$ from the previous and upcoming training sequences, respectively (see Fig. 4). Correlation γ is known (or frame decorrelation ε) as it can be obtained by averaging over multiple channel measurements over multiple frames. The MMSE estimator is

$$\hat{h}_B(k) = q(\gamma, \alpha) \times \hat{h}_p^{(L)} + q(\gamma, 1-\alpha) \times \hat{h}_p^{(R)} \tag{12}$$

where weighting $q(\gamma, \alpha)$ depends on γ as derived in Appendix B. It can be shown that (12) reduces to linear interpolation for practical values of correlation ρ (say $\gamma > 0.5$) with $q(\gamma, \alpha) \simeq \alpha$. Since from (6) $\hat{h}_p^{(L)} = h_p(M/Q) + \delta h_p^{(L)}$, $\hat{h}_p^{(R)} = h_p(1) + \delta h_p^{(R)}$ and $\delta h_p^{(L)}, \delta h_p^{(R)} \sim CN(0, Q/\tilde{M}\mu_p)$, the MSE (for $\mathbb{E}\left[\delta h_p^{(L)} \delta h_p^{(R)}\right] = 0$) depends on the variance due to channel outdating, $var[\hat{h}_B(\alpha N/Q)|h_p(1),\ h_p(M/Q)]$ (see Appendix) and due to the channel estimation errors

$$\sigma_B^2\left(\frac{\alpha N}{Q}; Q\right) = var\left[\hat{h}_B\left(\frac{\alpha N}{Q}\right)|h_p(1), h_p\left(\frac{M}{Q}\right)\right] + \frac{Q}{\tilde{M}\mu_p}\Psi(\gamma^{1/Q}, \alpha) \tag{13}$$

with $\Psi(\gamma, \alpha) = q(\gamma, \alpha)^2 + q(\gamma, 1-\alpha)^2$, [see (14) at the bottom of the page]. The MSE profile (14) for setting B is shown on the right-hand side of Fig. 2 for same parameters as for setting A, for both cases i) and ii). The largest MSE for channel interpolation is in the center of the payload interval ($\alpha = 1/2$) where, if the channel decorrelation is large, the degradation reaches its maximum level. The amount of degradation for case ii) is lower compared to case i) as the channel out-dating is reduced even if channel estimates are less accurate as based on a reduced number of training samples M/Q. For the same number of sub-blocks, channel interpolation can further reduce

$$\sigma_B^2\left(\frac{\alpha N}{Q}; Q\right) = \Gamma\left[\frac{1 + \gamma^{1/Q} - \gamma^{(1-\alpha)/Q} - \gamma^{\alpha/Q}}{1 - \gamma^{1/Q}} + \frac{Q \times \Psi(\gamma^{1/Q}, \alpha)}{\tilde{M}\mu}\right]. \tag{14}$$

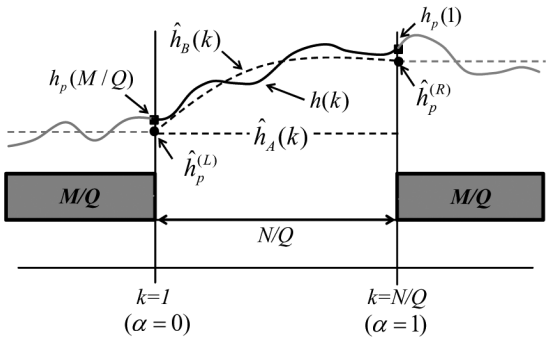

Fig. 4. Channel interpolation for setting A $\hat{h}_A(k)$ and B $\hat{h}_B(k)$ within a given sub-block of training length M/Q and payload N/Q.

the channel outdating and MSE with respect to setting A. Notice that in case the fading is almost static between the trainings ($\gamma \simeq 1$, or $\varepsilon \simeq 0$) then the quality of the channel estimate improves within the data interval (not shown here).

In what follows we tackle the problem of evaluating the optimal training efficiency $\hat{\eta}$ and the number of sub-blocks $\hat{Q}$ for settings A and B to maximize a given performance criteria.

IV. Optimization for Maximum Throughput

In this section, we formulate the problem of optimal design of the training interval and the training length for a given frame decorrelation ε, frame length L and SNR μ. The throughput is the cost function to be maximized and it is defined as the average number of *information* bits per channel use that are successfully forwarded to the receiver. For a frame of length L symbols with a payload of N information symbols, throughput is:

$$G_t(\mu,\eta,L,\rho,\sigma^2(k;Q)) = \eta \times \varsigma \times \Pr[\text{all } N \text{ symbols are correctly decoded}], \quad (15)$$

where ς is the number of bits per channel use according to the specific coding and modulation format employed for transmission. Notice that throughput (15) depends on the average performance with respect to the fading process. The throughput and optimal training structure are evaluated in this paper first by assuming a transmission that uses an uncoded BPSK modulation (Section V). Next the optimal design is evaluated by maximizing the average mutual information[2] of the channel by assuming Gaussian input symbol distribution (Section VI). We refer to corresponding sections for a more detailed analysis of each throughput function.

The optimization of training solved here can be stated by maximizing the throughput metric (15) over the training efficiency η and the number of sub-blocks Q:

$$(\hat{\eta},\hat{Q}) = \arg\max_{\eta,Q} G_t(\eta,\sigma^2(k;Q)|\mu,L,\rho)$$

$$\text{s.t. } 1 \le Q \le \frac{L}{2}, \quad \frac{Q}{L} \le \eta \le \frac{1-Q}{L}$$
$$Q, \frac{Q}{L} \in \mathbb{N}. \quad (16)$$

Without affecting the optimality of the results, optimization (16) can be conveniently restated by first tackling the problem of optimizing the number of sub-blocks $\hat{Q}$ for a given training efficiency η (and therefore given the number of training samples $M = (1-\eta)L$ and data symbols $N = \eta L$), SNR (μ) and fading correlation (ρ) so that

$$\begin{aligned}&(\text{Training interval optimization} - \text{Problem 1})\\ &\hat{Q}(\eta) = \arg\max_{Q} G_t\left(\sigma^2(k;Q)|\mu,\eta,L,\rho\right)\\ &\text{s.t. } 1 \le Q \le Q_{\max}\\ &Q, \frac{Q}{L}, \frac{Q}{\eta L}, \frac{Q}{(1-\eta)L} \in \mathbb{N}\end{aligned} \quad (17)$$

recall that $\hat{Q}(\eta) \le Q_{\max} = \min((1-\eta),\eta)\cdot L$ and $\hat{Q}(\eta)$ has to be a divisor of M, N and L.

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

$$\begin{aligned}&(\text{Training efficiency optimization} - \text{Problem 2})\\ &\hat{\eta} = \arg\max_{\eta} G_t(\eta,\sigma^2(k;\hat{Q}(\eta))|\mu,L,\rho)\\ &\text{s.t. } 1/L \le \eta \le 1 - 1/L\end{aligned} \quad (18)$$

where the optimal number of sub-blocks solution to (16) is nested into (18) as $\hat{Q}(\hat{\eta})$.

V. Throughput Analysis for Uncoded BPSK Modulation

By assuming the input stream as uncoded BPSK (with $\varsigma = 1$ bit/symbols), throughput (15) depends on the average Frame Error Rate (FER). For an arbitrary number of sub-blocks Q, the bit error probability is affected by: i) the fading correlation (ρ); ii) the position of the bit (k) within the data block for varying estimation accuracy $\sigma^2(k;Q)$ and training sequence lengths (M/Q); *iii)* the SNR (μ). For BPSK it is $\{\mathbf{x}\}_k = x(k) \in \{-1,+1\}$.

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 ([18, p. 61])

$$z(k) = \text{Re}\left(\frac{y(k)\hat{h}(k)^*}{|\hat{h}(k)|^2}\right) = x(k) + \frac{\text{Re}\left(\hat{h}(k)^*\bar{w}(k)\right)}{|\hat{h}(k)|^2} \quad (19)$$

the equivalent noise term $\bar{w}(k) = \delta h(k)x(k) + w(k)$ accounts for channel mismatch $\delta h(k) \sim CN(0,\sigma^2(k;Q))$ for any given channel estimation technique, as shown in Section III. The conditional BER is [25]

$$P(E|h(k)) = \Pr[z(k) < 0] \quad (20)$$

that modify throughput (15) as

[2]Here it is used the bound [6] to the information theoretic average mutual information that assumes Gaussian channel estimation noise. Notice that maximum average mutual information is expressed in bits per channel use and it can be interpreted as the limiting maximum achievable system throughput in the information theoretic sense.

$$G_t(\mu,\eta,L,\rho,\sigma^2(k;Q)) = \eta \times \left(\mathbb{E}_{\mathbf{h}} \left[\prod_{k\in\xi(Q,L,\eta)} [1-\Pr[z(k)<0]] \right] \right). \quad (21)$$

$\xi(Q,L,\eta)$ accounts for the set of bit positions over the frame that contain an information symbol and it depends on the frame structure (e.g., the number of sub-blocks Q, L and η). Notice that $\prod_{k\in\xi(Q,L,\eta)}[1-\Pr[z(k)<0]]$ is the conditional FER and for BPSK the throughput (21) is bounded so that $G_t(\mu,\eta,L,\rho,\sigma^2) \le \eta < 1$.

A. Numerical Optimization

Simulation results are given herein by assuming a Clarke model of fading so that the channel correlation among different time samples at time distance $d \times T_b$ is [23]

$$\frac{\mathbb{E}[h(k)h^*(k+d)]}{\Gamma} = J_0(2\pi f_d d) \quad (22)$$

where $J_0(\cdot)$ is the zero*th* order Bessel function, $f_d = vT_b/\lambda$ (λ is the signal wavelength, v is the mobile speed) is the maximum normalized Doppler frequency (that is defined as the maximum Doppler frequency normalized by the symbol rate $1/T_b$). Notice that, depending on the degree of terminal mobility and on the channel environment other correlation models might apply as well [22]. For a given channel environment the AR-1 model can be adapted to fit Clarke model (22) for small enough f_d [11]

$$\rho \approx 1 - f_d^2 \times \pi^2 \quad (23)$$

provided that it is $2\pi f_d L \ll 1$.

Fig. 5 show the throughput metric (21) versus the number of sub-blocks Q for a frame of $M = 30$ training symbols and $N = 30$ information symbols (with training efficiency $\eta = 0.5$). The optimal number of sub-blocks $\hat{Q}(0.5)$ to maximize the throughput (Problem 1 (17)) for setting A and B can be inferred on top and at bottom of Fig. 5, respectively. The fading realizations are obtained by simulating the Doppler as in [24], the channel values are estimated from the training symbols (6). Channel estimates are maintained as in (10) for the whole frame (setting A on top of Fig. 5) or they result as in (12) from linear interpolation (by placing $q(\gamma,\alpha) \approx \alpha$) between the two neighboring channel estimates (setting B at the bottom of Fig. 5). Conditional FER is simulated from the conditional BER (20) and then averaged over a large enough number of fading realizations. Optimum frame structures are indicated by markers for different values of the normalized maximum Doppler frequency f_d, and for SNR $\mu = 20\,\mathrm{dB}$ (dashed lines) and $\mu = 10\,\mathrm{dB}$ (solid lines).

When the Doppler and the SNR are high, channel outdating becomes the main limiting factor while the impact of channel estimation error is almost negligible. For these settings, it is preferable to have dispersed trainings of smaller length, spread over the frame so that the channel is extremely oversampled and the number of sub-block $\hat{Q}$ is the largest possible: $\hat{Q}(0.5) = Q_{\max} = L/2 = 30$, for $\mu = 20\,\mathrm{dB}$ and $f_d > 0.02$. Each

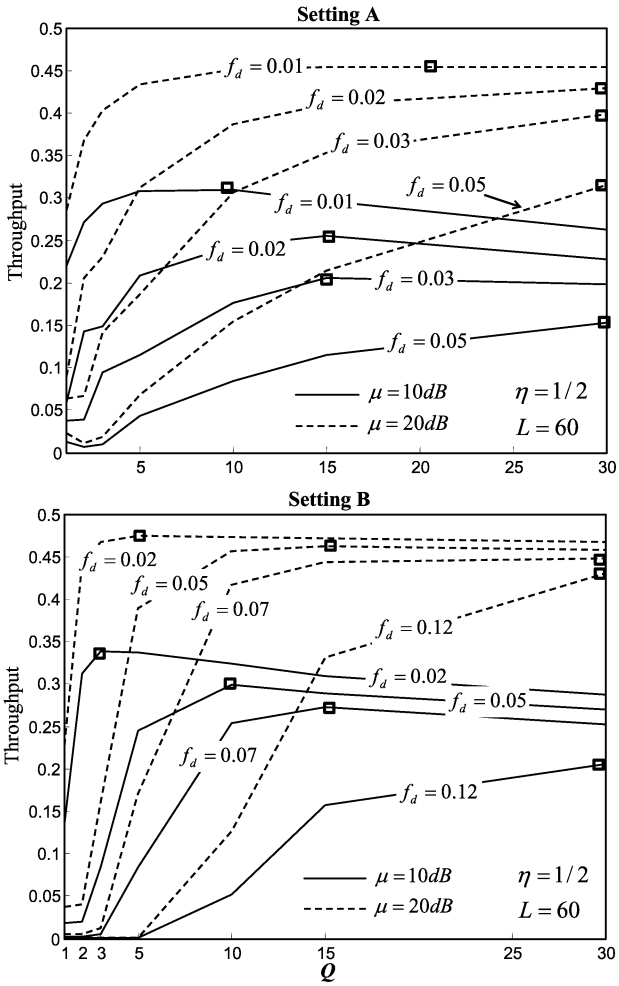

Fig. 5. Throughput G_t versus the number of sub-blocks Q, for a given training efficiency $\eta = 1/2$, $L = 60$ and for varying Doppler f_d and SNR $\mu = 10\,\mathrm{dB}$ (dashed lines), $\mu = 20\,\mathrm{dB}$ (solid lines). Setting A is on the top, while setting B is at the bottom. Square markers refer to the optimal number of sub-blocks $\hat{Q}$ (solution to Problem 1 (17)).

sub-block thus consists of one pilot symbol followed by one information symbol (recall that $\eta = 0.5$).

For low Doppler environments so that channel outdating is negligible and low SNR ($\mu = 10\,\mathrm{dB}$), the channel estimation error is now the main limiting factor. For $f_d = 0.02$ the optimal number of sub-blocks is $\hat{Q}(0.5) = 15 < Q_{\max}$ for setting A and $\hat{Q}(0.5) = 3 < Q_{\max}$ for setting B: each sub-block consists of $30/\hat{Q}(0.5)$ training symbols followed by $30/\hat{Q}(0.5)$ information symbols.

As an intuitive perspective, notice that at high SNR the limiting constraint on $\hat{Q}$ is the coherence time $T_c \simeq (1/2\pi f_d) \cdot T_b$ of the fading process so that the length of each sub-block $L/\hat{Q}$ should be upper-bounded as $L/\hat{Q} < T_c/T_b$ for setting A. When interpolation from neighboring channel estimates is used (setting B), the sub-block length $L/\hat{Q}$ can be made even larger than the coherence time of the channel until at least one channel estimate (either from the previous or the upcoming training phase) is of any help in decoding, therefore sub-block length is upper-bounded as $L/\hat{Q} < 2T_c/T_b$. This observation is confirmed in Fig. 5 for both settings

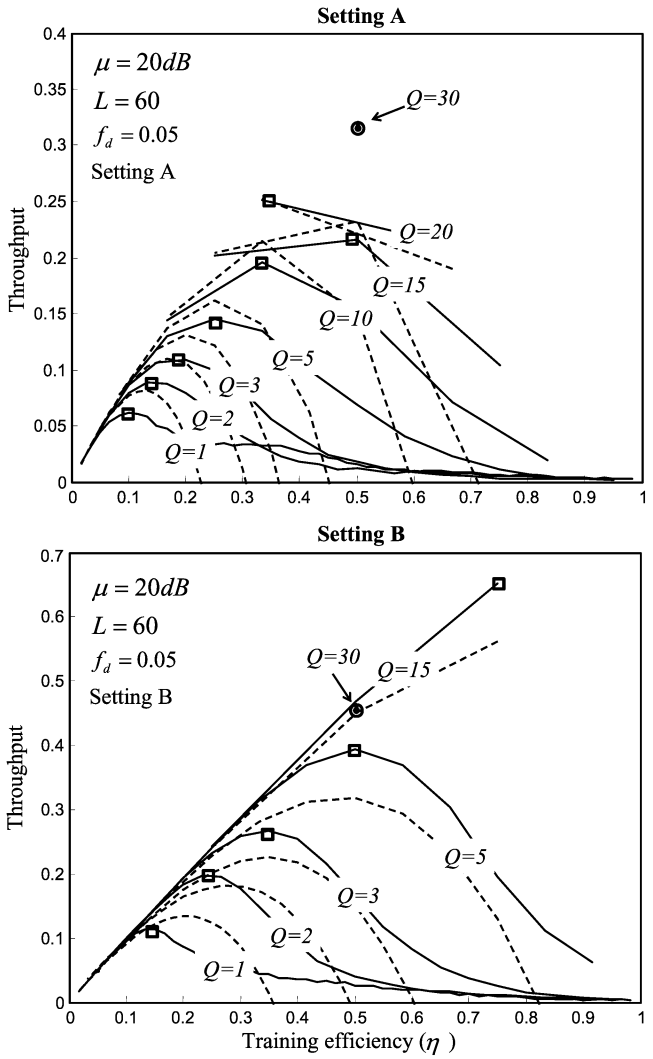

Fig. 6. Throughput G_t versus the training efficiency η for a given number of sub-blocks Q, $L = 60, \mu = 20\,\mathrm{dB}$ and $f_d = 0.05$. Optimal training efficiency is shown for a given number of sub-blocks Q, $f_d = 0.05$, $L = 60$, and $\mu = 20\,\mathrm{dB}$. Setting A is on the top, while setting B is at the bottom. Square markers refer to the maximum achievable throughput and optimal training efficiency $\hat{\eta}$ given the number of sub-blocks Q [e.g., for $Q = \hat{Q}$ this is the optimal solution to Problem 2 (18)]. Dashed lines refer to asymptotic ($\mu \to \infty$) approximation of throughput (29).

Channel interpolation (setting B) achieves the best throughput performances, as expected. The optimal training interval (e.g., $L/\hat{Q}$) for setting B is larger than that for setting A. Moreover, since the number of pilots is fixed, the training sequences are more clustered for setting B to gain from an improved channel estimation.

In Fig. 6 we show now the throughput metric G_t versus the training efficiency η (solid lines) for varying sub-blocks Q ($L = 60$ and $\mu = 20\,\mathrm{dB}$), and for setting A (on the top) and setting B (at the bottom). The maximum normalized Doppler frequency is $f_d = 0.05$ in both cases. Notice that for a given Q the training efficiencies that can be explored are $Q/L < \eta < 1 - Q/L$. When interpolation is used (setting B) meaningful performance benefits are observed with respect to setting A for any combination of Q and η. At $\mu = 20\,\mathrm{dB}$ the optimal pilot deployment for setting B (solution to Problem 2, (16)) requires $\hat{\eta} = 3/4$ 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. Instead, for setting A it is $\hat{\eta} = 1/2$ and $\hat{Q} = 30$, so that each burst consists of 1 training symbol followed by 1 information symbol.

B. Asymptotic Analysis ($\mu \to \infty$)

In order to gain insight into the optimization problem, here we derive a closed form that approximates the throughput measure (21) for large SNR and small Doppler ($\varepsilon \ll 1$) modeled as AR-1 (Appendix C). The purpose of this section is to cast the optimization problem (16) into the framework of convex optimization. Notice that the main reasonings behind the approximations are: i) the fading is stable within the training block of M/Q pilots, any performance loss can be accounted for separately by replacing M/Q with $\tilde{M}/Q$ (see Fig. 3); ii) all the impairments $\bar{w}$ due to channel estimation error and outdating in (19) are approximated as white Gaussian noise so that the conditional error probabilities can be taken from conventional $\mathcal{Q}()$ function as in [26]; iii) the estimate $\left|\hat{h}(k)\right|^2$ is unbiased so that $\mathbb{E}[|\hat{h}(k)|^2] = \left(1 - \sigma^2(k;Q)\right)\mathbb{E}[|h(k)|^2] \simeq \mathbb{E}[|h(k)|^2]$[3]; iv) the conditional FER is approximated as

$$\mathbb{E}_{\mathbf{h}} \prod_{k\in\xi(Q,L,\eta)} [1 - P(E|h(k))] \simeq \left[1 - \sum_{k\in\xi(Q,L,\eta)} \mathbb{E}_{h(k)}[P(E|h(k))]\right] \quad (24)$$

so that

$$G_t(\mu,\eta,L,\rho,\sigma^2(k,Q)) \simeq \eta \times \left[1 - \sum_{k\in\xi(Q,L,\eta)} \mathbb{E}_{h(k)}[P(E|h(k))]\right]. \quad (25)$$

By using the Gaussian approximation of equivalent noise $\tilde{w}$ and by noticing that $\mathrm{Re}(x) \le |x|$, the conditional BER can be bounded as in [25] by $P(E|h(k)) \ge \mathcal{Q}(\sqrt{2 \cdot SNR_k})$. From (19) the (effective) SNR at the decision variable is [26]

$$SNR_k = \frac{|\hat{h}(k)|^2}{\sigma_w^2} \cdot \frac{1}{1 + \sigma^2(k;Q)/\sigma_w^2} \approx \frac{|h(k)|^2}{\sigma_w^2} \cdot \frac{1}{1 + \sigma^2(k;Q)/\sigma_w^2}. \quad (26)$$

The average BER (evaluated w.r.t. fading over different frames but still in the same location k within the frame) scales as [25]

$$\mathbb{E}_{h(k)}[P(E|h(k))] \ge \mathbb{E}_{h(k)}\left[\mathcal{Q}\left(\sqrt{SNR_k}\right)\right] \simeq \frac{1}{4\mu}\left(1 + \mu\frac{\sigma^2(k;Q)}{\Gamma}\right). \quad (27)$$

[3]This assumption holds when the Doppler is small. Note that for high SNR the most significant source of error is the MSE of channel estimation that adds to AWG noise term [7].

Notice that in the limit of $\mu \to \infty$ the average BER performances are limited by the channel outdating as

$$\lim_{\mu\to\infty} \mathbb{E}_{h(k)}[P(E|h(k))] > \lim_{\mu\to\infty} \frac{1}{4\mu}\left(1+\mu\frac{\sigma^2(k;Q)}{\Gamma}\right) = \frac{1}{4SIR} \tag{28}$$

where $SIR = \Gamma/\sigma^2(k;Q)$ can be regarded as an equivalent signal to interference ratio [12].

By substituting (27) into (25) the throughput scales as

$$G_t(\mu,\eta,L,\rho,\sigma^2(k;Q)) \simeq \tilde{G}_t$$
$$\tilde{G}_t = \eta \times \left(1 - \frac{1}{4\mu}\left[\eta L + Q\cdot\frac{\mu}{\Gamma}\sum_{k=1}^{\eta L/Q}\sigma^2(k;Q)\right]\right) \tag{29}$$

the last equality follows from the stationarity of the fading (i.e., the MSE profile is the same for each sub-block, as training sequences are equally split among each frame partition). For AR-1 model of fading the MSE $\sigma^2(k;Q)$ is defined in (11) and (14) for setting A and B, respectively.

In Fig. 6 we assess the tightness of the proposed throughput approximation (29) (dashed lines) as compared to the simulated throughput (solid lines). The dynamics of the channel are modeled by an AR-1 model where the correlation parameter ρ is a function of the Doppler f_d as in (23). We used the MSE profiles $\sigma^2(k;Q)$ developed in Section III.

Closed form solutions for both Problem 1 (17) and 2 (18) are now derived by using the throughput approximation (29).

The Proposition 1 below provides a closed form solution to Problem 1 (17): the optimal number of sub-blocks (thus the optimal training interval) for a given number of training symbols (e.g., for a given training efficiency η) is found to be a function of the SNR (μ), the frame decorrelation (4) (ε), and the frame length (L).

Proposition 1: For asymptotically high SNR and $\varepsilon \ll 1$ the optimal solution to

$$\text{(Problem 1) } \hat{Q}(\eta) = \arg\max_{Q} \tilde{G}_t(\sigma^2(k,Q)|\mu,\eta,L,\rho) \quad \text{s.t. } 1 \le Q \le Q_{\max} \tag{30}$$

is

$$\hat{Q}(\eta) \approx \begin{cases} Q_{\max} \text{ when } \mu\varepsilon \ge \varphi L \times \Omega(\eta) \\ \sqrt{\frac{L(1-\eta)\eta\mu\varepsilon}{\varphi}} \text{ when } \frac{\varphi}{(1-\eta)\eta L} \le \mu\varepsilon < \varphi L \times \Omega(\eta) \\ 1 \text{ when } \mu\varepsilon < \frac{\varphi}{(1-\eta)\eta L} \end{cases} \tag{31}$$

constant φ is $\varphi = 1$ for setting A, and $\varphi = 3$ for setting B, $\Omega(\eta) = \min[\eta/1-\eta, 1-\eta/\eta]$. Recall that $Q_{\max} = L \times \min(\eta, 1-\eta) \le L/2$.

Proof: See Appendix D ∎

Remark 2: It follows from (31) that $\lim_{\mu\to\infty}\hat{Q}(\eta) = Q_{\max}$.

Remark 2 states that if $\varepsilon > 0$ (or $f_d > 0$) and SNR μ is large enough so that $\mu\varepsilon \ge \varphi L \times \Omega(\eta)$ then the frame should be partitioned such that the number of sub-blocks is the largest (or the training interval is the smallest). This result was also observed numerically in [12] by optimizing the efficiency η to satisfy a given BER requirement. For a given training efficiency η, the maximum number of sub-blocks is $\hat{Q}(\eta) = Q_{\max}$ such that each sub-block has at least one pilot symbol and one information symbol as for PSAM modulation [1]. Same result was also obtained in Fig. 5 showing that for $\mu = 20$ dB there is clear advantage in partitioning the frame by using the maximum number of sub-blocks $\hat{Q}(0.5) = Q_{\max} = L/2$. Notice that this condition is still valid in case fading is time-varying over the training block.[4].

On the contrary, Proposition 1 shows that for low Doppler with $\varepsilon \simeq 0$ (or $f_d \simeq 0$) and for a SNR μ such that $\mu\varepsilon < \varphi L$ then there is no advantage in fragmenting the frame, therefore, $\hat{Q}(\eta) \simeq 1$. In this case, a tradeoff there exist between the number of training symbols that should be clustered together to gain from improved channel estimation and the length of the interval between two adjacent training phases (training interval).

Optimal settings that can be derived using Proposition 1 are still applicable as far as performance degradation of channel estimation due to time-varying fading over the training block can be neglected. Given an optimal number of sub-blocks $\hat{Q}(\eta)$ and frame length L that define the training interval $L/\hat{Q}(\eta)$, the number of (clustered) pilots in each fragment is $(1-\eta)L/\hat{Q}(\eta)$ while the number of information symbols is $\eta L/\hat{Q}(\eta)$. Notice that for setting A, the optimal number of sub-blocks should be larger as the impact of channel outdating is predominant in any case.

Using the optimal solution $\hat{Q}(\eta)$ in (31) we now maximize the throughput measure (29) (as for Problem 2 (18)) by finding the optimal training efficiency $\hat{\eta}$ such that

$$\text{(Problem 2) } \hat{\eta} = \arg\max_{\eta} \tilde{G}_t(\eta, \sigma^2(k;\hat{Q}(\eta))|\mu,L,\rho) \quad \text{s.t. } \frac{1}{L} \le \eta \le 1 - \frac{1}{L}. \tag{32}$$

for the optimal number of sub-blocks is $\hat{Q}(\eta)$ from (31).

Proposition 3: The optimal training efficiency solution to (32) scales, for small enough ε and $\varepsilon\mu \ge \varphi L$, as

$$\lim_{\mu\to\infty} \hat{\eta} \approx 1 - \max\left(\frac{1}{L}, \frac{1}{1+\varphi}\sqrt{\varepsilon}\right) = \hat{\eta}_\infty. \tag{33}$$

Proof: See Appendix E ∎

Proposition 3 shows that when the SNR is high so that the channel outdating is the limiting factor (as for $\mu \to \infty$), the required fraction of pilot symbols for each setting is

$$\text{(Setting A, } \varphi = 1) \lim_{\mu\to\infty}[1-\hat{\eta}_A] \approx \frac{1}{2}\sqrt{\varepsilon} \tag{34}$$

$$\text{(Setting B, } \varphi = 3) \lim_{\mu\to\infty}[1-\hat{\eta}_B] \approx \frac{1}{4}\sqrt{\varepsilon} \tag{35}$$

and should increase with the squared root of the frame decorrelation $\varepsilon < 1$. Finally, since

$$\frac{\text{Fraction of Pilots (Setting B)}}{\text{Fraction of Pilots (Setting A)}} = \lim_{\mu\to\infty}\frac{1-\hat{\eta}_B}{1-\hat{\eta}_A} = \frac{1}{2}$$

[4]In this case the large SNR condition is sufficient (but not necessary) as performance degradation due to channel variations within the training intervals have been neglected (i.e., $M/Q \le M/Q$).

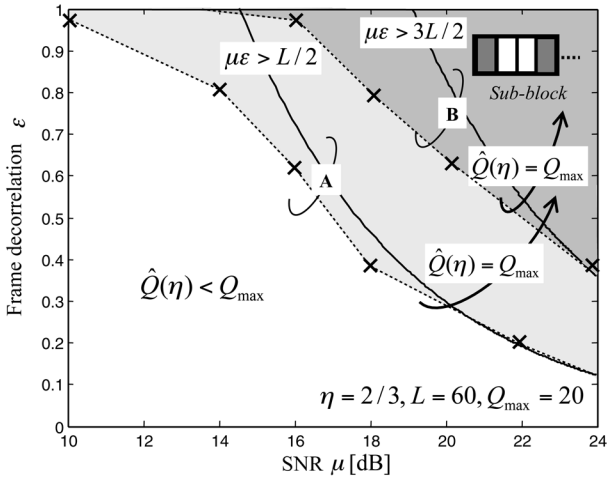

Fig. 7. Regions of SNR μ and frame decorrelation ε pair to support maximum fragmentation $Q_{\max}$ for setting A (light gray shaded area) and setting B (dark gray shaded area) for $L = 60$, $\eta = 2/3$ so that $Q_{\max} = 20$ sub-blocks and $L/Q_{\max} = 3$ symbols each. Empty markers refer to the region boundary for setting A, while filled markers refer to the boundary for setting B.

it proves that channel estimation by MMSE interpolation (setting B) requires only half the number of pilots compared to the case where the channel is sampled (estimated) and hold for the whole payload (setting A).

Remark 4: The optimal number of sub-blocks scales as $\lim_{\mu\to\infty} \hat{Q}(\hat{\eta}) \approx \max(1, (L/1+\varphi)\sqrt{\varepsilon})$.

Proof: From Remark 2 since $\mu\varepsilon \geq \varphi L$ then $\lim_{\mu\to\infty} \hat{Q}(\hat{\eta}) \approx L \cdot \lim_{\mu\to\infty} \min((1-\hat{\eta}), \hat{\eta})$. As $\lim_{\mu\to\infty} \hat{\eta} > 1/2$ from (34) and (35) then $\lim_{\mu\to\infty} \hat{Q}(\hat{\eta}) \approx L(1 - \lim_{\mu\to\infty} \hat{\eta})$. ■

The interplay between the limiting optimal efficiency $\lim_{\mu\to\infty} \hat{\eta}$ solution to (32) and the optimal training design $\hat{\eta}$ that maximize the exact throughput (18) will be assessed in what follows by numerical analysis.

1) Numerical Examples for Training Design ($\mu \to \infty$): Fig. 7 evaluates the condition for maximum fragmentation (in solid lines) from Remark 2 $\mu\varepsilon \geq \varphi L \times \Omega(\eta)$ for setting A and B (with $\eta = 2/3$, $L = 60$) as compared with simulated performances (markers). Shaded areas refer to regions in terms of SNR μ and frame decorrelation ε pair where optimal frame design prescribes trainings of one pilots each with interval among two successive pilots of $L/Q_{\max} = 3$ symbols (as $Q_{\max} = 20$). Condition ($\mu\varepsilon \geq \varphi L/2$ from Remark 2 for $\eta = 2/3$) is tight enough in predicting the exact regions for small enough ε and large μ. Notice that region for setting A is larger compared to that for channel interpolation (setting B) as simpler estimation requires a denser pilot placement to limit the large channel out-dating.

Fig. 8 compares the limiting training efficiency $\hat{\eta}_\infty$ derived in closed form from Proposition 3 with the optimal solution to Problem 2 (18) $\hat{\eta}$ for varying SNR μ and Doppler environments. The optimal limiting training efficiency $\hat{\eta}_\infty$ (33) is shown on top of Fig. 8 in solid lines for both setting A and B versus the fading decorrelation ε. Optimal number of sub-blocks is shown at the bottom for the same settings. Notice that ε can be mapped onto an equivalent normalized Doppler frequency by using the mapping in subfigure where we set $\rho = J_0(2\pi f_d)$). Dots refer to

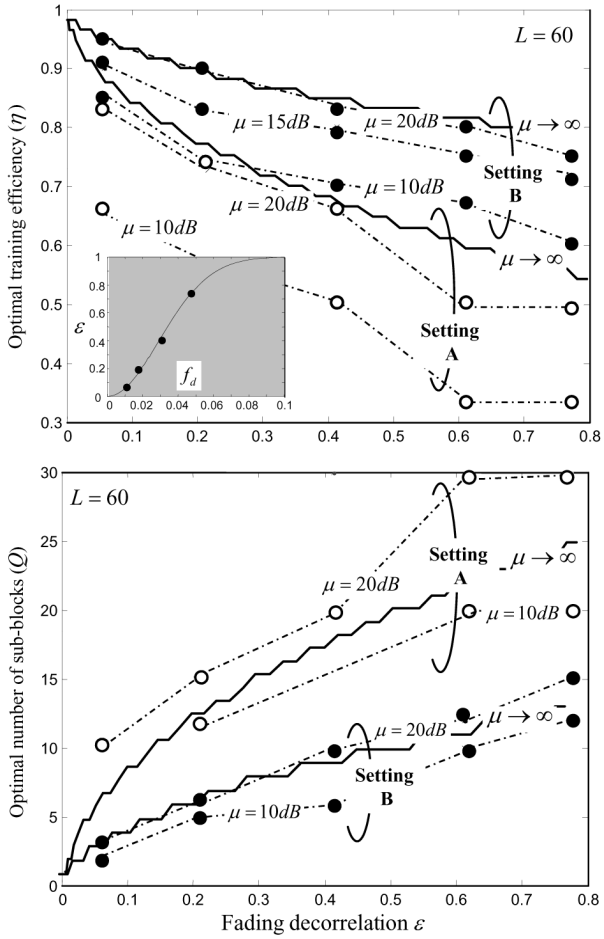

Fig. 8. Training efficiency $\hat{\eta}_\infty$ (33) (on the top) and optimal number of sub-blocks $\hat{Q}(\hat{\eta}_\infty)$ (at the bottom) in solid lines for setting A and B versus the fading decorrelation ε (mapping between ε and. f_d by using (4) and (23) is included in shaded subfigure in down-left corner), markers refer to optimized settings (for throughput in (21)) for setting A (empty markers) and B (filled markers) for varying SNR: μ =10, 15, and 20 dB.

optimized efficiency $\hat{\eta}$ solution to Problem 2 (18) for setting B with finite SNR, μ =10, 15, 20 dB, instead empty dots refer to setting A with finite SNR μ =10, 20 dB. In both cases exact throughput (21) is maximized. We show that the proposed bound is tight enough in predicting the exact performances still from SNR $\mu = 20$ dB. Notice that throughput is significantly degraded for SNRs below 10 dB as $\hat{\eta} < 0.7$. For fading decorrelation $\varepsilon \simeq 0.05$ (or $f_d \simeq 0.01$) the maximum limiting training efficiency (33) for setting B is $\hat{\eta}_\infty = 0.95$ while for setting A is $\hat{\eta}_\infty = 0.9$. Notice that these results are confirmed by simulating the optimal efficiency $\hat{\eta}$ still for $\mu = 20$ dB. The true optimal number of sub-blocks is the divisor of $L = 60$ that is nearest to $\hat{Q}(\hat{\eta}) \approx \max(1, (1 - \hat{\eta}_\infty)L)$: it is found from simulations $\hat{Q} = 3$ ($\hat{Q}(\hat{\eta}) \simeq 3.35$) for setting B and $\hat{Q} = 5$ ($\hat{Q}(\hat{\eta}) \simeq 6.7$) for setting A. When the fading decorrelation is as high as $\varepsilon \simeq 0.75$ (or $f_d \simeq 0.05$), the maximum achievable training efficiency for setting B is $\hat{\eta}_\infty \approx 0.78$, while for setting A is $\hat{\eta}_\infty \approx 0.58$. For $L = 60$ and $\mu = 20$ dB the optimal training efficiencies (see also Fig. 6) are found to be $\hat{\eta} = 0.75$ for setting B and $\hat{\eta} = 0.5$ for setting A. The optimal number of sub-blocks is $\hat{Q} = 15$ ($\hat{Q}(\hat{\eta}) \approx 12.6$) and $\hat{Q} = 30$ ($\hat{Q}(\hat{\eta}) \approx 25.2$).

As a final remark, it should be noticed that for asymptotically large SNR μ the number of training samples (M) should never exceed the number of information symbols (N) as $\lim_{\mu\to\infty}\hat{\eta} > 1/2\,\forall\,\varepsilon < 1$, thus, extending to continuous time-varying fading the results for block fading of [3].

VI. Information Theoretic Analysis

In this section, we derive the optimal training length and training interval so that the channel capacity is maximized. By maximizing the mutual information $I(\mathbf{y};\mathbf{x}|\mathbf{y}_p,\mathbf{x}_p)$[3] with respect to the input symbol distribution, we define the throughput function as the maximum achievable capacity

$$G_t(\mu,\eta,L,\sigma^2) = \frac{1}{L}\max_{\Pr(\mathbf{x})}[I(\mathbf{y};\mathbf{x}|\mathbf{y}_p,\mathbf{x}_p)]. \tag{36}$$

A. Asymptotic Analysis ($\mu\to\infty$)

A lowerbound to throughput (capacity bound) in the presence of channel estimation error and out-dating can be found in [6] by computing the capacity of the following channel

$$y(k) = \hat{h}(k)x(k) + \bar{w}(k) \tag{37}$$

where the estimates $\hat{h}(k)$ are assumed to be independent (and identically distributed, i.i.d.) and $\bar{w}(k)$ is approximated as white Gaussian noise process (as in Section V). Independent fading states can be obtained by using an ideal interleaver with a sufficiently deep memory (namely larger than the coherence time of the fading process) that randomly spreads the symbols among multiple sub-blocks to combat deep fades. Based on these assumptions, throughput can be bounded as

$$G_t(\mu,\eta,L,\sigma_e^2) \geq \eta\times\mathbb{E}_{\mathbf{h}}\left[\sum_{k\in\xi(Q,L,\eta)}\frac{\log(1+SNR_k)}{N}\right] \tag{38}$$

where $\mathbb{E}_{\mathbf{h}}\left[\sum_{k\in\xi(Q,L,\eta)}(1/N)\log(1+SNR_k)\right]$ is the maximum average mutual information for channel model (37) and Gaussian noise.

We now approximate the bound (38) for large SNR and small Doppler as in Section V-B. Using the Jensen inequality $\mathbb{E}_{\mathbf{h}}[\sum_{k\in\xi(Q,L,\eta)}\log(1+SNR_k)] \leq \sum_{k\in\xi(Q,L,\eta)}\log\left(1+\mathbb{E}_{h(k)}[SNR_k]\right)$ and effective SNR approximation (26), throughput (38) can be approximated as

$$G_t \simeq \tilde{G}_t(Q,\eta) = \eta\times\int_0^1\log\left(1+\frac{\mu}{1+\mu\frac{\sigma^2\left(\frac{\alpha\eta L}{Q};Q\right)}{\Gamma}}\right)d\alpha. \tag{39}$$

Notice that for block fading (or $\rho = 1$) it is $\sigma^2(k;Q) \to \mathbb{E}\left[|\Delta h_p|^2\right]$ because for $\lim_{\sigma^2\to\mathbb{E}[|\Delta h_p|^2]} G_t = \lim_{\sigma^2\to\mathbb{E}[|\Delta h_p|^2]}\tilde{G}_t = \eta\times\log\left(1+\Gamma/\sigma_w^2+\mathbb{E}[|\Delta h_p|^2]\right)$ as [3]. The optimal number of sub-blocks, thus the optimal training interval [Problem 1 (17)] for a given number of training symbols (e.g., for a given training efficiency η), can be approximated by solving the following (convex) optimization problem:

$$\begin{aligned}&\text{(Problem 1) } \hat{Q}(\eta) = \arg\max_{1\leq Q\leq Q_{\max}}\tilde{G}_t(Q,\eta)\\ &Q,\frac{Q}{L},\frac{Q}{\eta L},\frac{Q}{(1-\eta)L}\in\mathbb{N}.\end{aligned} \tag{40}$$

Remark 4: It follows from (40) that $\lim_{\mu\to\infty}\hat{Q}(\eta) = Q_{\max}$.

Using the optimal solution $\hat{Q}(\eta)$ in (40) we maximize now the throughput measure (39) [as for Problem 2 in (18)] by finding the optimal training efficiency $\hat{\eta}$ such that

$$\text{(Problem 2) } \hat{\eta} = \arg\max_{1/L\leq\eta\leq 1-1/L.}\tilde{G}_t(\hat{Q}(\eta),\eta) \tag{41}$$

where the optimal number of sub-blocks is $\hat{Q}(\hat{\eta})$ from Problem 1 (40).

Remark 5: In the limiting case of asymptotically high SNR and $\varepsilon \ll 1$ the optimal training efficiency [for Problem 2 in (41)] $\hat{\eta}_\infty = \lim_{\mu\to\infty}\hat{\eta}$ is the solution to the following:

$$\begin{aligned}&[1+\omega(\varsigma,\hat{\eta}_\infty,\rho)]\log_2(1+\omega(\varsigma,\hat{\eta}_\infty,\rho))\\ &\qquad = \frac{\varsigma}{(1-\rho)} + \omega(\varsigma,\hat{\eta}_\infty,\rho)\end{aligned} \tag{42}$$

with $\omega(\varsigma,\eta,\rho) = \varsigma(1-\eta)/[\eta(1-\rho)]$ and constant ς is $\varsigma = 1/2$ (setting A) and $\varsigma = 2$ (setting B).

Proof: For small enough ε the optimization criteria (39) is approximated by using the linear piecewise approximation of $\sigma^2(\alpha N/\hat{Q}(\eta);\hat{Q}) \simeq \tilde{\sigma}^2(\alpha N/\hat{Q}(\eta);\hat{Q})$ (Appendix C) for both setting A and B. After straightforward algebra, the optimal training efficiency becomes

$$\begin{aligned}\hat{\eta}_\infty &= \lim_{\mu\to\infty}\hat{\eta}\\ &\approx \arg\max_{1/L\leq\eta\leq 1-1/L.}\eta\\ &\quad\times[\log_2(1+\omega(\varsigma,\eta,\rho))\\ &\qquad+\omega(\varsigma,\eta,\rho)\log_2(1+\omega^{-1}(\varsigma,\eta,\rho)1)]\\ &\approx \arg\max_{1/L\leq\eta\leq 1-1/L.}\eta\times[\log_2(1+\omega(\varsigma,\eta,\rho))].\end{aligned}$$

for $\omega(\varsigma,\eta,\rho)\log_2\left(1+\omega(\varsigma,\eta,\rho)^{-1}\right)\approx 1$. It follows the (42) by nulling the derivative of $\eta[\log_2(1+\omega(\varsigma,\eta,\rho))]$ with respect to η. ∎

B. Numerical Optimization of Average Mutual Information

In this section, ergodic capacity is maximized for optimal training designs. Fading realizations and channel estimates are obtained as in Section V-A, while channel capacity is obtained by averaging over a large enough number of fading realizations. Fig. 9 corresponds (for the same settings) to Fig. 5 where the throughput metric is now replaced by the capacity bound (38). Throughput is shown versus the number of sub-blocks Q for a frame of $M = 30$ training symbols and $N = 30$ information symbols ($L = 60$, $\eta = 0.5$) so that the optimal number of sub-blocks $\hat{Q}$ (Problem 1 (17)) for setting A and B (Fig. 9) can be inferred. Optimum frame structure is shown for

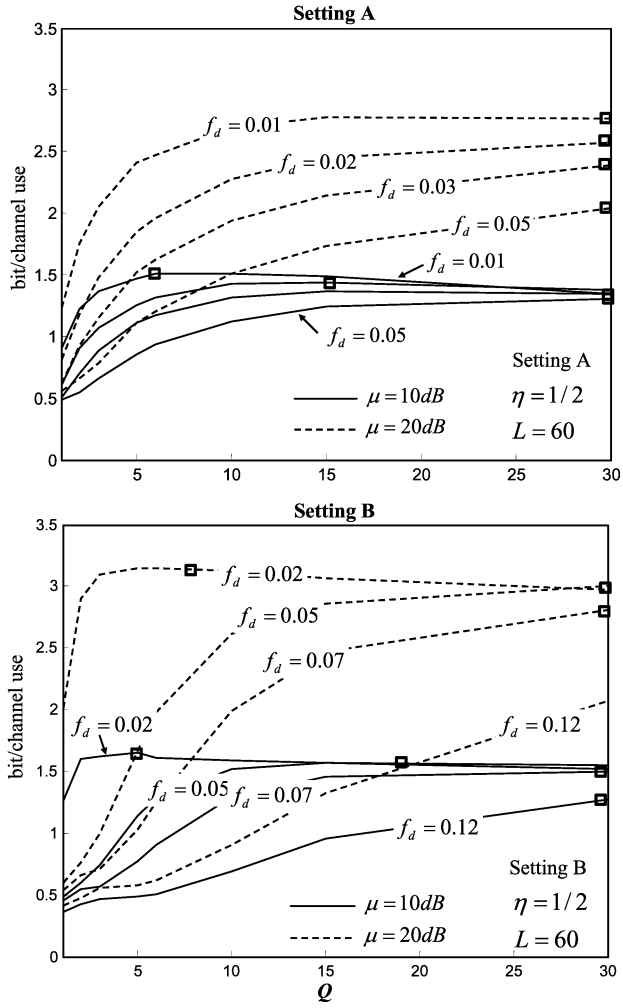

Fig. 9. Throughput function G_t (38) for Gaussian input symbol distribution versus the number of sub-blocks Q, setting A is shown on top, while setting B is at the bottom (for simulation settings in Fig. 5).

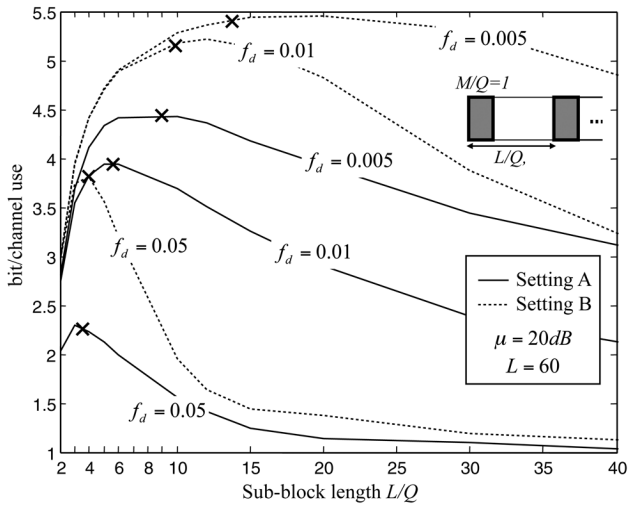

Fig. 10. Throughput performances versus the sub-block length (L/Q) where each sub-block has only one pilot $(M/Q = 1)$. Results are shown for both settings A and B and for varying Doppler environments. Asymptotical optimal training intervals $\lim_{\mu\to\infty} L/Q(\hat{\eta}) = (1-\hat{\eta}_\infty)^{-1}$ with $\hat{\eta}_\infty$ solution to (42) are also superimposed by cross markers.

different values of the normalized maximum Doppler frequency f_d and for SNR $\mu = 20$ dB (dashed lines) and $\mu = 10$ dB (solid lines). The results confirm some conclusions as for the uncoded BPSK transmission (Section V): dispersed trainings with smaller length have to be preferred when the product $\mu \times \varepsilon$ is large enough (see Proposition 1). Given the same SNR μ, frame decorrelaion ε and frame length L, the optimal values for the training periods are similar to those obtained by minimizing the packet error rate in Fig. 5. In all cases, the achievable maximum throughput (in bit/channel use) is larger compared to that obtained in Fig. 5 as modulation is not limited to BPSK (Section V). In this case, larger degrees of freedom are available to further maximize the spectral efficiency.

The interplay between the limiting optimal efficiency $\hat{\eta}_\infty$ solution to (42) and the optimal design $\hat{\eta}$ that maximize the bound (38) is assessed in Fig. 10 for both setting A and B. Throughput performances are shown for varying length of the sub-blocks (L/Q) and by constraining (for $\mu = 20$ dB) each sub-block to have only one pilot. This setting was also considered in [6]: training is maximally dispersed and the training length is minimal, $M/Q = 1$. Optimal training periods[5] $L/\hat{Q}(\hat{\eta}) = (1-\hat{\eta})^{-1}$ can be inferred for varying Doppler environments, moreover analytical results (cross markers) are superimposed to numerical analysis (solid and dashed lines) and they refer to optimal training periods $(1-\hat{\eta}_\infty)^{-1}$ solution to (42) for asymptotically high SNR. As confirmed by [12], the optimal sub-block length reduces for high Doppler to allow for larger fading oversampling. Similarly as what observed for BPSK case, the optimal training period relates closely with the coherence time interval $1/2\pi f_d$ of the fading process, as expected.

As far as optimal pilot placement is concerned, similar conclusions on training length and interval tradeoff can be made for both BPSK and information theoretic cases. Although specific results are not provided in this paper, we expect that the main reasonings of the analysis would be marginally affected by the particular modulation and coding employed for transmission.

VII. Concluding Remarks

In this paper, we optimized the training in wireless communication systems where pilot symbols are multiplexed within the information symbols to provide an estimate of the Rayleigh faded channel. Estimate is obtained either by linear combination of the estimates from the previous and the upcoming training sequences (setting B) or by holding the estimate from previous training until new pilots are received (setting A). Given that pilots can be considered as sampling of fading variations, the quality of the channel estimate is related to how many contiguous pilots are used in sampling the fading and how frequently the fading estimate is acquired. The training optimization is tackled by maximizing the system throughput. As optimization strategy, we *first* maximize the time interval between two successive training phases, *next* we optimize the number of pilot symbols (the training length) to be deployed during each phase (that constrains the *quality* of the sampling). The fundamental tradeoff between maximizing the channel estimation quality (for large training lengths and small training intervals)

[5]Recall that for sub-block structures with one pilot $(M/Q = 1)$ it is: $\hat{Q}(\hat{\eta}) = \min(1-\eta, \eta) = 1-\eta$

and maximizing the fraction of time spent in payload transmission is accounted for. By deriving closed form solutions to optimization problems that are valid for small frame decorrelation ε and large (but finite) SNR μ, we provide analytical insights to optimal training designs.

When the product of ε and μ is large (i.e., $\mu\varepsilon > \varphi L$, see Proposition 1), channel out-dating becomes the main limiting factor. In this case, it is preferable to have dispersed trainings of smaller length (case ii) spread over the frame so that the fading is extremely oversampled and the training interval is the smallest possible (largest Q). Instead, when the product $\mu\varepsilon$ is low (i.e., $\mu\varepsilon < \varphi L$) the channel estimation error becomes a limiting factor. Optimal training design results now as a tradeoff between concentrating all the training symbols at the beginning of the frame or spreading the pilots over the frame.

We quantify how much gain is provided by a channel estimation strategy that combines the channel estimates from neighboring training sequences (setting B) compared to the simpler processing based on (channel) sampling and holding (setting A). As a general result for large SNR, interpolating the channel estimates from two trainings requires half the number of pilots compared to simpler methodology (setting A). In both cases the optimal fraction of pilot symbols scales (at large SNR) as the square root of the frame decorrelation $\sqrt{\varepsilon}$ (Proposition 2). Results and observations are further corroborated by optimizing the training for maximum mutual information over Gaussian input symbol distribution. Although specific analysis is not dealt with in this paper, it is expected the main reasoning of the analysis to be still valid regardless of the particular modulation and coding (other than BPSK) employed.

APPENDIX

Channel Estimation Error for Time Varying Fading Over Pilots: To simplify the notation, let the training sequence consist of M symbols (e.g., for $Q = 1$). The received signal during one training sequence can be written in general as

$$\mathbf{y}_p = \mathrm{diag}(\mathbf{h}_p) \times \mathbf{x}_p + \mathbf{w}_p \tag{43}$$

where vector $\mathbf{h}_p = [h_p(1), \ldots, h_p(M)]$ now collects the *time-varying* fading channels over the M pilots (see Fig. 3). Assume that the channel to be estimated at one edge of the training is $h_p(1)$, the fading dynamics are modeled by AR-1 with parameter ρ so that

$$h_p(k) = \rho^{k-1} h_p(1) + \sum_{n=0}^{k-2} \rho^n u(k-1-n) \tag{44}$$

for $k = 2, \ldots, M$ and $u \sim CN(0, (1-\rho^2)\Gamma)$. Notice that a similar reasoning holds true for estimation of the channel $h_p(M)$ (or $h_p(M/Q)$) at the other edge of the training sequence. Using (44) the ML estimator $\hat{h}_p$ of the channel at the edge of the frame

$$\hat{h}_p = \left(\mathbf{x}_p^H \mathbf{x}_p\right)^{-1} \mathbf{x}_p^H \mathbf{y}_p = \psi(\rho) h_p(1) + \delta h_p \tag{45}$$

is now biased with respect to $h_p(1)$[6] as $\psi(\rho) = M^{-1} \sum_{k=0}^{M-1} \rho^k \leq 1$. Moreover, time-varying fading within the training block causes the estimation error δh_p to increase as

$$\mathbb{E}\left[|\delta h_p|^2\right] = \frac{1}{M\mu_p} + \frac{1}{M^2} \times \mathbb{E}\left[\left|\sum_{k=1}^{M-1} u(M-k) \sum_{n=0}^{k-1} \rho^n\right|^2\right] \approx \frac{1}{M\mu_p} + \frac{(1-\rho^2)\,\Gamma}{6} \frac{(2M-1)(M-1)}{M}.$$

Assuming that the bias at the edge of the training block is negligible for small enough Doppler ($\psi(\rho) \approx 1$ for $\rho \approx 1$) and training block length ($M \ll L$), then the estimator performance $\hat{h}_p \approx h_p + \delta h_p$ is limited by error $\delta h_p \sim CN(0, 1/(\tilde{M}\mu_p))$ with

$$\tilde{M} \approx \max\left[1, \frac{M}{1 + \frac{(1-\rho^2)\Gamma\mu_p(2M-1)(M-1)}{6}}\right] \tag{46}$$

and being $\tilde{M}$ the equivalent number of pilots to have the same performance of block-fading. Notice that $\tilde{M}$ ranges between $\tilde{M} = 1$ and $\tilde{M} = M$, with $\tilde{M} \approx M$ for $\rho \approx 1$ as expected.

MMSE Interpolation (From Known Neighboring Channel Values): 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 $\alpha \in [0, 1]$. We first evaluate the conditional probability density function (notice that all the components are Gaussian):

$$\begin{aligned} &p(h(k)|h(0), h(K)) \\ &= \frac{p(h(K)|h(0), h(k)) \cdot p(h(0), h(k))}{p(h(K)|h(0)) \cdot p(h(0))} \\ &= \frac{p(h(k)|h(K)) \cdot p(h(K))}{p(h(k))} \cdot \frac{p(h(k)|h(0))}{p(h(K)|h(0))} \\ &= \frac{p(h(k)|h(K)) \cdot p(h(k)|h(0))}{p(h(k))} \\ &\quad \cdot \underbrace{\frac{p(h(K))}{p(h(K)|h(0))}}_{\text{not depend on } h(k)} \end{aligned} \tag{47}$$

where in the second equality we used the Markov properties of AR and thus $p(h(K)|h(0), h(k)) = p(h(K)|h(k))$[13]. Since a combination of Gaussian random variables is still Gaussian, we can focus the attention to the exponent of the pdf $p(h(k)|h(0), h(K))$ so that, by neglecting the last term $p(h(K))/p(h(K)|h(0))$ as it acts as a scaling factor, we can evaluate the mean and variance of the equivalent Gaussian model

$$p(h(k)|h(0), h(K)) \propto \exp\left(-\frac{\phi(h(k)|h(0), h(K))}{\Gamma}\right) \tag{48}$$

as $p(h(k)|h(K))$, $p(h(k)|h(0))$ and $p(h(k))$ in (47) are all Gaussian

$$p(h(k)|h(K)) \propto \exp\left(-\frac{|h - \mathbb{E}[h(k)|h(K)]|^2}{var[h(k)|h(K)]}\right) \tag{49}$$

$$p(h(k)|h(0)) \propto \exp\left(-\frac{|h - \mathbb{E}[h(k)|h(0)]|^2}{var[h(k)|h(0)]}\right) \tag{50}$$

[6]Bias is the same with respect to $h_p(M)$ as fading is stationary over the training sequence.

and $p(h(k)) \propto \exp(-|h|^2/\Gamma)$. Dynamics of the channel are modeled as AR-1 process, it is $E[h(k)|h(0)] = \rho^{K\alpha}h(0)$, $E[h(k)|h(K)] = \rho^{K(1-\alpha)}h(K)$, $var[h(k)|h(K)] = \Gamma(1-\rho^{2K(1-\alpha)})$ and $var[h(k)|h(0)] = \Gamma(1-\rho^{2K\alpha})$ for $k = \alpha K$, and $\alpha = 0, \ldots, 1$. By using (47), (49) and (50), the exponent term in (48) is

$$\begin{aligned}\phi(h(\alpha K)|h(0),h(K)) &= \frac{\left|h(\alpha K)-\rho^{K(1-\alpha)}h(K)\right|^2}{(1-\rho^{2K(1-\alpha)})} \\ &\quad + \frac{\left|h(\alpha K)-\rho^{K\alpha}h(0)\right|^2}{(1-\rho^{2K\alpha})} - |h|^2 . \end{aligned} \tag{51}$$

After straightforward algebra

$$\phi = \frac{\sigma_1^2+\sigma_2^2-\sigma_1^2\sigma_2^2}{\sigma_1^2\sigma_2^2} \times \left| h - \frac{\sigma_2^2 h_K + \sigma_1^2 h_0}{\sigma_1^2+\sigma_2^2-\sigma_1^2\sigma_2^2} \right|^2 \tag{52}$$

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 $h_K = \rho^{K(1-\alpha)}h(K)$, moreover $\gamma \equiv \rho^{2K}$ is the square of the correlation coefficient between $h(0)$ and $h(K)$. From (52) the conditional variance is

$$var[h(k)|h(0),h(K)] = \Gamma\frac{1+\gamma-\gamma^{1-\alpha}-\gamma^{\alpha}}{1-\gamma} \tag{53}$$

and the conditional mean (i.e., the MMSE estimator) is the combination

$$\begin{aligned}\hat{h}(\alpha K) &= \mathbb{E}[h(\alpha K)|h(0),h(K)] \\ &= q(\gamma,\alpha)h(0) + q(\gamma,1-\alpha)h(K)\end{aligned} \tag{54}$$

with

$$q(\gamma,\alpha) = \frac{1-\gamma^{1-\alpha}}{1-\gamma}\gamma^{\alpha/2}. \tag{55}$$

Weighting term $q(\gamma,\alpha)$ (or $q(\gamma,1-\alpha)$) for varying values of intertraining correlation γ shows that as far as $\gamma \geq 0.1$ the MMSE estimator (54) can be approximated by the *linear* interpolation $\mathbb{E}[h(\alpha K)|h(0),h(K)] \approx \alpha h(0) + (1-\alpha)\,h(K)$ (or equivalently $q(\gamma,\alpha) \approx \alpha$). For smaller (and somewhat unrealistic values of γ, here $\gamma = 10^{-6}$), the MMSE estimator privileges only the neighboring values when $\alpha \sim 1/2$, thus, giving up with any possible estimation from neighboring channel values (as $q(\gamma,\alpha)$ gets small quite easily versus α).

MSE Linear Approximation: The linear piecewise approximation $\tilde{\sigma}^2(\alpha N/Q,;Q)$ of the MSE profile $\sigma^2(k;Q)$ is herein developed for settings A and B. Approximations assume small frame decorrelation ε and are validated by numerical analysis (see Section V-A). The following approximation valid for $\varepsilon \ll 1$ will be used in the following:

$$1-\gamma^{\alpha/Q} = 1-\rho^{2\alpha N/Q} = \frac{2\alpha\eta\varepsilon}{Q} + o(\varepsilon) \tag{56}$$

with $\varepsilon = 1-\rho^L \approx L(1-\rho)$.

Setting A: When ε is small enough, the MSE profile within a given sub-block of data length N/Q can be approximated by a linear function so that, recalling that $k = \alpha N/Q$, $\sigma_A^2(\alpha N/Q;Q) \simeq \tilde{\sigma}_A^2(\alpha N/Q;Q) = \Gamma[a+b/2\cdot\alpha]$ for $\alpha = 0,\ldots,1$. Being $M = (1-\eta)L$, from the MSE profile (11) it is

$$a = \frac{\sigma_A^2(\alpha=0;Q)}{\Gamma} = \frac{Q}{\mu(1-\eta)L} \tag{57}$$

notice that $\lim_{\mu\to\infty} a = 0$. For ε small enough, by substituting (56) into the MSE profile (11) it is

$$b = 2\left(\frac{\sigma_A^2(\alpha=1;Q)}{\Gamma} - a\right) \approx \frac{4\eta\varepsilon}{Q}\left[1-\frac{Q}{\mu(1-\eta)L}\right]. \tag{58}$$

The following relationship will be useful in the main text (Proposition 1 and 3)

$$\int_0^1 \tilde{\sigma}_A^2\left(\frac{\alpha\eta L}{Q};Q\right)d\alpha = \frac{Q}{\mu(1-\eta)L} + \frac{\eta\varepsilon}{Q}\left[1-\frac{Q}{\mu(1-\eta)L}\right] \tag{59}$$

and $\lim_{\mu\to\infty}\int_0^1 \tilde{\sigma}_A^2(\alpha\eta L/Q;Q)d\alpha = \eta\varepsilon/Q$.

Setting B: When ε is small enough, the MSE profile within a given sub-block of data length N/Q can be fit by a triangular function ($k = \alpha N/Q$ for $\alpha = [0,1]$) $\sigma_B^2(\alpha N/Q,Q) \simeq \tilde{\sigma}_B^2(\alpha N/Q,Q) = \Gamma[a+b\cdot|\alpha-1/2|]$. From the MSE profile (14) it is

$$a+\frac{b}{2} = \frac{\sigma_B^2(\alpha=0;Q)}{\Gamma} = \frac{\sigma_B^2(\alpha=1;Q)}{\Gamma} = \frac{Q}{\mu(1-\eta)L} \tag{60}$$

and using (56)

$$a = \frac{\sigma_B^2\left(\alpha=\frac{1}{2};Q\right)}{\Gamma} \approx \frac{\eta\varepsilon}{2Q}\left[1+\frac{Q^2}{(1-\eta)L\mu\eta\varepsilon} - \frac{1}{Q(1-\eta)L\mu}\right]. \tag{61}$$

while $b = 2\left(\sigma_B^2(\alpha=0;Q)/\Gamma - a\right)$ is

$$b \approx \frac{Q}{\mu(1-\eta)L}\left(1-\frac{\eta\varepsilon}{Q^2}\left[(1-\eta)L\mu - \frac{1}{Q}\right]\right) + o\left(\frac{1}{\mu}\right). \tag{62}$$

The following relationship will be useful for Proposition 1 and 3:

$$\int_0^1 \tilde{\sigma}_B^2\left(\frac{\alpha\eta L}{Q};Q\right)d\alpha = \frac{3Q}{4(1-\eta)L\mu} + \frac{\eta\varepsilon}{4Q} \tag{63}$$

and $\lim_{\mu\to\infty}\int_0^1 \tilde{\sigma}_B^2(\alpha\eta L/Q;Q)d\alpha \approx \eta\varepsilon/(4Q)$.

Proposition 1:

Proof: After straightforward algebra, Problem 1 (30) can be restated as

$$\begin{aligned}\hat{Q}(\eta) &\approx \arg\max_{1\leq Q\leq Q_{\max}} \tilde{G}_t(\sigma^2(k;Q)|\mu,\eta,L,\rho) \\ &\approx \arg\min_{1\leq Q\leq Q_{\max}} \frac{\eta L}{4\Gamma}\int_0^1 \tilde{\sigma}^2\left(\frac{\alpha\eta L}{Q};Q\right)d\alpha\end{aligned} \tag{64}$$

recall that $Q_{\max} = \min(\eta, 1-\eta)L$. As shown in Appendix C the MSE profile $\sigma^2(\cdot)$ can be approximated, for small enough frame decorrelation ε in (4), by using a linear piecewise approximation so that $\sigma^2(\alpha\eta L/Q;Q) \approx \tilde{\sigma}^2(\alpha\eta L/Q;Q)$. The problem reduces to

$$\hat{Q}(\eta) \approx \arg\min_{1\leq Q\leq Q_{\max}}\left(\frac{\varphi\cdot Q}{\mu(1-\eta)L} + \frac{\eta\varepsilon}{Q}\right) \tag{65}$$

where we used the integrals $\int_0^1 \tilde{\sigma}^2(\alpha\eta L/Q;Q)d\alpha$ derived in (59) for setting A and in (63) for setting B that yields to $\varphi = 1$ and $\varphi = 3$, respectively. Solution is now straightforward. ■

Proposition 3:

Proof: In the limit for asymptotically large SNR, the problem can be restated as

$$\hat{\eta}_\infty = \lim_{\mu\to\infty}\hat{\eta} \approx \arg\max_{\eta}\lim_{\mu\to\infty}\tilde{G}_t(\eta,\tilde{\sigma}^2(k;\hat{Q}(\eta))|\mu,L,\rho) \tag{66}$$

where $\tilde{\sigma}^2(k;\hat{Q}(\eta))$ is the piecewise approximation in Appendix C. From Proposition 1 (Remark 2), having large SNR, $\varepsilon\mu \geq \varphi L$, it is $\hat{Q}(\eta) = Q_{\max}$. By substituting $Q = \hat{Q}(\eta) = Q_{\max}$ into (66), (66) becomes

$$\hat{\eta}_\infty \approx \arg \max_{1/L \leq \eta \leq 1-1/L} \eta \times \left(1 - \frac{\left(\frac{\eta}{(1+\varphi)}\right)^2}{\min((1-\eta),\eta)}\varepsilon \right). \quad (67)$$

Since for $1/L < \eta \leq 1/2$ and $\varepsilon \ll 1$, the limiting throughput $\eta\left(1-\eta\varepsilon/(1+\varphi)^2\right)$ is an increasing function of η, the feasible set for (67) is such that $1/2 < \eta \leq 1 - 1/L$, thus

$$\hat{\eta}_\infty \approx \arg \max_{1/2 < \eta < 1-1/L} \eta \times \left(1 - \frac{\eta^2\varepsilon}{(1-\eta)(1+\varphi)^2} \right) \quad (68)$$

and (66) now has solution $\hat{\eta}_\infty \approx 1 - \max(1/L, 1/(1+\varphi)\sqrt{\varepsilon})$. ■

ACKNOWLEDGMENT

The authors acknowledge L. B. Milstein and P. C. Cosman for focusing the attention to this timely topic during their visit at the University of California at San Diego (UCSD).

REFERENCES

[1] 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.

[2] T. L. Marzetta, "Blast training: Estimating channel characteristics for high-capacity space-time wireless," presented at the Proc. 37th Ann. Allerton Conf. Commun., Control, Comput., Sep. 22–24, 1999.

[3] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theroy*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[4] O. Simeone and U. Spagnolini, "Lower bound on training-based channel estimation error for frequency-selective block-fading Rayleigh MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 11, pp. 3265–3277, Nov. 2004.

[5] F. Chan, "Statistical methods on detecting superpositional signals in a wireless channel," Ph.D. dissertation, Univ. of South Wales, Sydney, Australia, Jul. 2006.

[6] J. Baltersee, G. Fock, and H. Meyr, "An information theoretic foundation of synchronized detection," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2115–2123, 2001.

[7] J. Baltersee, G. Fock, and H. Meyr, "Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 12, pp. 2358–2368, 2001.

[8] S. Akin and M. C. Gursoy, "Training optimization for Gaussian-Markov Rayleigh fading channels," in *Proc. IEEE ICC*, Glasgow, Scotland, Jun. 2007, pp. 5999–6004.

[9] S. Misra, A. Swami, and L. Tong, "Optimal training over the Gauss-Markov fading channel: a cutoff rate analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2004, vol. 3, pp. 809–812.

[10] M. Dong, L. Tong, and B. M. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1403–1418, May 2004.

[11] Q. Sun, D. C. Cox, H. C. Huang, and A. Lozano, "Estimation of continuous flat fading MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 549–553, Oct. 2002.

[12] V. Pohl, P. H. Nguyen, V. Jungnickel, and C. v. Helmolt, "Continuous flat-fading MIMO channels: Achievable rate and optimal length of the training and data phases," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1889–1900, Jul. 2005.

[13] M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 2005.

[14] 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, Jun. 2005.

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

[16] A. J. Grant, "Optimal sequences for efficient estimation of multipath channels," in *Proc. IEEE 5th Int. Symp. Spread Spectrum Technol. Appl.*, Sep. 2–4, 1998, vol. 1, pp. 56–58.

[17] E. d. Carvalho and D. Slock, "Cramer-rao bounds for semi blind, blind and training sequence based channel estimation," in *Proc. 1st IEEE Signal Proc. Workshop on Signal Process. Adv. Wireless Commun.*, Apr. 1997, vol. 1, pp. 129–132.

[18] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge , U.K.: Cambridge Univ. Press, 2005.

[19] M. Stojanovic, J. Proakis, and J. Catipovic, "Analysis of the impact of the channel estimation errors on the performance of a decision feedback equalizer in fading multipath channels," *IEEE Trans. Commun.*, vol. 43, pp. 877–886, Feb. 1995.

[20] M. K. Tsatsanis, G. B. Giannakis, and G. Zhou, "Estimation and equalization of fading channels with random coefficients," *IEEE Trans. Signal Process.*, vol. 53, no. 2–3, pp. 211–229, Sep. 1996.

[21] L. Lindbom, "A Wiener filtering approach to the design of tracking algorithm," Ph.D., Uppsala Univ., Uppsala, Sweden, 1995.

[22] K. E. Baddour and N. C. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1650–1650, Jul. 2005.

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

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

[25] J. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.

[26] M. A. R. Baissas and A. M. Sayeed, "Pilot-based estimation of time-varying multipath channels for coherent cdma receivers," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 2037–2037, Aug. 2002.

Stefano Savazzi (S'06) received the M.Sc. in telecommunication engineering and the Ph.D. degree in information technology (both with honors) from the Dipartimento di Elettronica e Informazione, Politecnico di Milano, Milan, Italy, in 2004 and 2008, respectively.

He was a Visiting Researcher with Uppsala University in 2005 and with the University of California San Diego (UCSD) in 2007. He is currently a Postdoctoral Researcher with Politecnico di Milano. His current research interests include signal processing aspects for digital wireless communications and, more specifically, channel estimation, wireless sensor networks for geophysical applications, and cooperative networking.

Dr. Savazzi won the Dimitris N. Chorafas Foundation Award for best Ph.D. dissertation.

Umberto Spagnolini (SM'03) received the Dott. Ing. Elettronica degree (*cum laude*) from the Politecnico di Milano, Italy, in 1988.

Since 1988, he has been with the Dipartimento di Elettronica e Informazione, Politecnico di Milano, where he is Full Professor in Telecommunications. He is the cofounder of Wireless System Technology (WiSyTech), a spinoff company of Politecnico di Milano on Software Defined Radio. His interests are in the area of statistical signal processing. Specific areas of interest include channel estimation and space-time processing for wireless communication systems; synchronization and cooperation in wireless sensor networks; parameter estimation/tracking and wavefield interpolation applied to UWB radar, oil exploration, and remote sensing.

Dr. Spagnolini served as Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (1999–2006) and on the Editorial Board and Technical Program Committees of several conferences in all the areas of interests.