

Maximizing Spectral Efficiency for High Mobility Systems with Imperfect Channel State Information

Ning Sun and Jingxian Wu

Abstract—This paper studies the optimum system design that can maximize the spectral efficiency of high mobility wireless communication systems with imperfect channel state information (CSI). The fast time-varying fading in high mobility systems can be tracked with pilot-assisted channel estimation. The percentage of pilot symbols in the transmitted symbols plays a critical role on the system performance: a higher pilot percentage yields a more accurate channel estimation, but also more overhead. The effects of pilot percentage are quantified through the derivation of the channel estimation mean squared error (MSE), which is expressed as a closed-form expression of various system parameters through asymptotic analysis. It is discovered that, if the pilots sample the channel above its Nyquist rate, then the estimation of the channel coefficients of data symbols through temporal interpolation yields the same asymptotic MSE as the direct estimation of the channel coefficients of the pilot symbols. Based on the statistical properties of the channel estimation error, we quantify the impacts of the imperfect CSI on the system performance by developing the analytical symbol error rate (SER) and a spectral efficiency lower bound of the communication system. The optimum pilot percentage that can maximize the spectral efficiency lower bound is identified through both analytical and simulation results.

Index Terms—High mobility communications, imperfect channel state information, channel estimation error, spectral efficiency.

I. INTRODUCTION

HIGH mobility wireless communications have received increasing attentions recently with the growing demands for applications such as high speed railways and aircraft communications. One of the main challenges faced by high mobility communications is the fast time-varying fading caused by the Doppler shift, which could be as high as 1,000 Hz for a 2.4 GHz system operating at a speed of 450 km/hr. In a high mobility system, the accurate estimation and tracking of the fast time-varying fading are critical to reliable system operations. Channel estimation can be performed either through the direct estimation of the fading coefficients [1]-[10], or through basis expansion models (BEMs) that transform fading coefficients to low-dimensional transform domains [11], [12].

Many channel estimation related works focus on the design of optimum pilot patterns that can minimize the channel estimation mean squared error (MSE) [1]-[6], [12]. In [1]-[4], the optimum pilot design for orthogonal frequency division multiplexing (OFDM) systems employing the minimum mean squared error (MMSE) channel estimation is discussed,

and it is shown that the MSE can be minimized by using identical equally-spaced pilot tones or clusters. The MMSE estimator requires the priori knowledge of channel statistics, yet such information is not needed by a least squares (LS) channel estimator. In [5], the optimum pilot pattern for the LS estimation of quasi-static channel in OFDM systems is obtained through numerical convex optimizations. The LS estimation of doubly selective fading channels are discussed in [6] for a multiple-input multiple-output (MIMO) OFDM system, where the pilot matrix is designed as a unitary matrix to avoid matrix inversions during the LS channel estimation. A windowed LS (WLS) channel estimation with a BEM channel model is proposed in [12], and it is shown that the estimation accuracy of WLS can approach that of MMSE-based estimators. All above methods are designed by using the MSE as a metric under the constraint of fixed pilot power and/or pilot numbers. They do not consider how the pilot patterns or imperfect channel state information (CSI) impacts the overall communication performance, such as the bit error rate (BER), spectral efficiency, or energy efficiency.

In high mobility systems, channel estimation errors are non-negligible and they might have significant impacts on the system performance and designs [7]-[9]. In [7], it is discovered that systems employing LS or MMSE channel estimations can achieve the same symbol error rate (SER) performance if the optimum receivers are designed by considering the statistics of the channel estimation errors. In [8], the impacts of channel estimation error on the BER of an ultra-wide band (UWB) system are studied. Both [7] and [8] use system error probability as the design metric. An information theoretic metric, a sum-rate lower bound of a two-way relay network, is used in [9] to evaluate the system performance in the presence of imperfect CSI. The sum-rate lower bound is numerically maximized by considering parameters such as training vector structures and the number of training symbols. A quasi-static block fading model is assumed in [7]-[9], thus the results are not applicable to high mobility systems. In [10], the tracking of a time-varying channel is achieved by using polynomial interpolations, and the results are used to quantify the BER of a two-way relay system. It is demonstrated that polynomial interpolations might not be sufficient to track the channel variation in high mobility systems.

In this paper, the optimum pilot design that maximizes the spectral efficiency of high mobility wireless communication systems is studied. The fast time-varying fading coefficients are estimated and tracked through the MMSE estimation and interpolation. The MSE of channel coefficients of pilot symbols and data symbols are studied through the asymptotic analysis, and the results are expressed as closed-form expres-

Manuscript received May 1, 2013; revised September 12, 2013; accepted November 13, 2013. The associate editor coordinating the review of this paper and approving it for publication was W. Zhang.

The authors are with the Department of Electrical Engineering, University of Arkansas, Fayetteville, AR, 72701, U.S.A. (e-mail: {nsun, wuj}@uark.edu). Digital Object Identifier 10.1109/TWC.2014.012314.130772



Fig. 1. The transmitted slot structure.

sions of parameters such as the maximum Doppler spread, the signal-to-noise ratio (SNR) of received pilot symbols, and the percentage of pilot symbols in the transmitted symbols. It is discovered that, if the pilots sample the channel at or above the Nyquist rate of the time-varying fading, then the MMSE interpolation of the channel coefficients of data symbols yields the same asymptotic MSE as the estimation of the channel coefficients of pilot symbols. The statistical properties of the estimated channel coefficients are studied, and the results are used to develop an analytical SER and a spectral efficiency lower bound for systems operating with imperfect CSI. A higher pilot percentage yields a better SER. However, a lower SER does not necessarily mean a better overall performance, considering the fact that the excessive use of pilot symbols means more overhead. Such a tradeoff relationship is revealed in the system spectral efficiency. The optimum pilot percentage that can maximize the spectral efficiency lower bound is analytically identified. The impacts of imperfect CSI on system performance are studied through both analytical and simulation results.

The remainder of this paper is organized as follows. The system model is presented in Section II. Section III studies the analytical asymptotic MSE for channel estimation and interpolation. The impacts of the imperfect CSI on the SER are studied in Section IV by analyzing the statistical properties of the estimated channel coefficients. In Section V, a spectral efficiency lower bound is developed for systems with imperfect CSI, and the optimum pilot percentage that maximizes the spectral efficiency is identified. Numerical results are given in Section VI, and Section VII concludes the paper.

II. SYSTEM MODEL

Consider a system that employs pilot-assisted channel estimation and experiences fast time-varying fading. At the transmitter, the data to be transmitted are divided into slots, and each slot has N_s modulated data symbols and $N_p \leq N_s$ pilot symbols. The values of N_s and N_p can be chosen such that $K = \frac{N_s}{N_p} + 1$ is an integer. The pilot symbols are equally spaced such that each pair of adjacent pilot symbols are separated by $K - 1$ data symbols. The slot structure is shown in Fig. 1, where P and D denote the pilot and data symbols, respectively. Define the symbol vector as $\mathbf{x} = [x_1, \dots, x_N]^T \in \mathcal{S}^{N \times 1}$, where $N = N_s + N_p$ is the total number of symbols per slot, \mathcal{S} is the modulation alphabet set, and \mathbf{A}^T represents the matrix transpose. Denote the k -th pilot symbol as $x_{i_k} = p_k$, where $i_k = kK$ is the index of the k -th pilot symbol, for $k = 1, \dots, N_p$. The average energy of the symbols is normalized to 1, $\mathbb{E}[|x_n|^2] = 1$, where \mathbb{E} is the mathematical expectation operator. Define the percentage of the pilot symbols as $\delta = \frac{N_p}{N} = \frac{1}{K}$.

The data and pilot symbols are transmitted over the fast time-varying fading channel with additive white Gaussian noise (AWGN). The signals observed at the receiver is

$$\mathbf{y} = \sqrt{E_0} \cdot \mathbf{X} \cdot \mathbf{h} + \mathbf{z}, \quad (1)$$

where $\mathbf{y} = [y_1, \dots, y_N]^T \in \mathcal{C}^{N \times 1}$ is the received signal, $\mathbf{z} = [z_1, \dots, z_N]^T \in \mathcal{C}^{N \times 1}$ is the AWGN with covariance matrix $\mathbf{R}_z = \sigma_z^2 \mathbf{I}_N$, \mathbf{I}_N is a size- N identity matrix, E_0 is the average transmission energy of a symbol, $\mathbf{h} = [h_1, \dots, h_N]^T \in \mathcal{C}^{N \times 1}$ is the channel fading coefficient vector, and $\mathbf{X} = \text{diag}(\mathbf{x})$ is a size- N diagonal matrix with the transmitted signal vector \mathbf{x} on its diagonal. The time-varying fading coefficients are correlated with the cross-correlation being

$$\rho(m - n) = \mathbb{E}[h_m h_n^*] = J_0(2\pi f_D |m - n| T_s), \quad (2)$$

where $\rho(k) \in (-1, 1]$, f_D is the maximum Doppler spread of the fading channel, T_s is the symbol period, and $J_0(x)$ is the zero-order Bessel function of the first kind.

It is assumed that the energy per symbol E_0 is fixed for both data and pilot symbols. Therefore, the average transmission power is $P_0 = \frac{E_0}{T_s}$.

III. IMPACTS OF PILOT PERCENTAGE ON CHANNEL ESTIMATION

In this section, the impacts of the pilot percentage on the channel estimation MSE are analytically studied through the asymptotic analysis. The channel estimation is performed in two steps: the receiver first obtains an estimate of the channel coefficients at pilot locations, then the channel coefficients at non-pilot locations are obtained by performing MMSE interpolations over the estimated CSI at pilot locations.

A. MMSE Channel Estimation at Pilot Locations

At the receiver, the distorted observations of the pilot symbols can be expressed as

$$\mathbf{y}_p = \sqrt{E_0} \mathbf{P} \mathbf{h}_p + \mathbf{z}_p \quad (3)$$

where $\mathbf{y}_p = [y_{i_1}, y_{i_2}, \dots, y_{i_{N_p}}]^T \in \mathcal{C}^{N_p \times 1}$, $\mathbf{h}_p = [h_{i_1}, h_{i_2}, \dots, h_{i_{N_p}}]^T \in \mathcal{C}^{N_p \times 1}$, and $\mathbf{z}_p = [z_{i_1}, z_{i_2}, \dots, z_{i_{N_p}}]^T \in \mathcal{C}^{N_p \times 1}$ are the received signal vector, fading vector, and noise vector at pilot locations, respectively, and \mathbf{P} is a size $N_p \times N_p$ diagonal matrix with the diagonal elements being $[p_1, \dots, p_{N_p}]^T \in \mathcal{S}^{N_p \times 1}$.

The estimate of the channel fading at the pilot locations \mathbf{h}_p can be obtained by minimizing the average MSE, $\sigma_{p, N_p}^2 = \frac{1}{N_p} \mathbb{E}(\|\hat{\mathbf{h}}_p - \mathbf{h}_p\|^2)$, as

$$\hat{\mathbf{h}}_p = \mathbf{W}_p^H \mathbf{y}_p, \quad (4)$$

where $\hat{\mathbf{h}}_p$ is an estimate of \mathbf{h}_p . The MMSE estimation matrix \mathbf{W}_p can be calculated as

$$\mathbf{W}_p = \sqrt{E_0} (E_0 \mathbf{P} \mathbf{R}_{hh} \mathbf{P}^H + \sigma_z^2 \mathbf{I}_{N_p})^{-1} \mathbf{P} \mathbf{R}_{hh} \quad (5)$$

where $\mathbf{R}_{hh} = \mathbb{E}[\mathbf{h}_p \mathbf{h}_p^H] \in \mathcal{C}^{N_p \times N_p}$ with its elements defined in (2), \mathbf{A}^H denotes the matrix Hermitian operation, and \mathbf{I}_{N_p} is a size- N_p identity matrix. The channel auto-correlation matrix \mathbf{R}_{hh} is a Toeplitz matrix with the (m, n) -th element being $J_0(2\pi f_D |m - n| \frac{T_s}{\delta})$ as defined in (2), where δ is the percentage of pilot symbols.

The error correlation matrix, $\mathbf{R}_{ee} = \mathbb{E}[\mathbf{e}_p \mathbf{e}_p^T]$, with $\mathbf{e}_p = \hat{\mathbf{h}}_p - \mathbf{h}_p$, can be calculated as

$$\mathbf{R}_{ee} = \mathbf{R}_{hh} - \mathbf{R}_{hh} \left(\mathbf{R}_{hh} + \frac{1}{\gamma_0} \mathbf{I}_{N_p} \right)^{-1} \mathbf{R}_{hh} \quad (6)$$

where $\gamma_0 = \frac{E_0}{\sigma_z^2}$ is the signal-to-noise ratio (SNR) without fading, and the assumption $|p_n|^2 = 1$ is used in the above equation. This assumption can be easily met by choosing only constant amplitude symbols, such as phase shift keying symbols, as the pilot symbols. It should be noted that the data symbols do not need to be of constant amplitude.

The average MSE can then be calculated as

$$\sigma_{p,N_p}^2 = \frac{1}{N_p} \text{trace}(\mathbf{R}_{ee}). \quad (7)$$

From (6) and (7), the calculation of the MSE involves matrix inversion and the trace operation. In order to explicitly identify the impacts of pilot percentage on the MSE, we resort to the asymptotic analysis by letting $N_p \rightarrow \infty$ and $N_s \rightarrow \infty$ while keeping a finite pilot percentage δ and data rate $R_s = \frac{1}{T_s}$. The results are presented as follows.

Proposition 1: When $N_p \rightarrow \infty$ while keeping a finite δ and R_s , the asymptotic MSE, $\sigma_p^2 = \lim_{N_p \rightarrow \infty} \sigma_{p,N_p}^2$, of the estimated channel coefficient at the pilot locations is

$$\sigma_p^2 = 1 - \frac{8\gamma_0 \arctan\left(\sqrt{\frac{2\gamma_0 - \frac{\alpha}{\pi}}{2\gamma_0 + \frac{\alpha}{\pi}}}\right)}{\pi \sqrt{(2\gamma_0)^2 - \left(\frac{\alpha}{\delta}\right)^2}}, \quad \text{for } \delta \geq \frac{\alpha}{\pi}, \quad (8)$$

where $\alpha = 2\pi f_d T_s$, γ_0 is the SNR without fading, and δ is the pilot percentage.

Proof: The proof is in Appendix A. ■

B. MMSE Channel Interpolation

Once the estimates of the channel information at the pilot locations are obtained, they can be interpolated to obtain the channel estimations of the entire slot.

Consider the estimation of the fading coefficients with symbol indices $\{i'_k = (k-1)K + u\}_{k=1}^{N_p}$, where $u = 1, \dots, K-1$ correspond to the indices of the non-pilot data symbols. Define the fading vector to be estimated through interpolation as $\mathbf{h}_d = [h(i'_1), \dots, h(i'_{N_p})]^T \in \mathcal{C}^{N_p \times 1}$.

The receiver obtains the estimate of the channel fading at non-pilot locations, $\hat{\mathbf{h}}_d$, as

$$\hat{\mathbf{h}}_d = \mathbf{W}_d^H \hat{\mathbf{h}}_p, \quad (9)$$

where $\hat{\mathbf{h}}_d$ is an estimate of \mathbf{h}_d and $\mathbf{W}_d \in \mathcal{R}^{N_p \times N_p}$ is the MMSE channel estimation matrix to minimize $\frac{1}{N_p} \mathbb{E}(\|\hat{\mathbf{h}}_d - \mathbf{h}_d\|^2)$.

With the orthogonal principal, $\mathbb{E}[(\hat{\mathbf{h}}_d - \mathbf{h}_d)\hat{\mathbf{h}}_p^T] = \mathbf{0}$, we have $\mathbf{W}_d^H = \mathbf{R}_{d\hat{h}} \mathbf{R}_{\hat{h}\hat{h}}^{-1}$, where

$$\mathbf{R}_{d\hat{h}} \triangleq \mathbb{E}(\mathbf{h}_d \hat{\mathbf{h}}_p^H) = \sqrt{E_0} \mathbf{R}_{dh} \mathbf{P}^H \mathbf{W}_p, \quad (10a)$$

$$\mathbf{R}_{\hat{h}\hat{h}} \triangleq \mathbb{E}(\hat{\mathbf{h}}_p \hat{\mathbf{h}}_p^H) = \mathbf{W}_p^H (E_0 \mathbf{P} \mathbf{R}_{hh} \mathbf{P}^H + \sigma_z^2 \mathbf{I}_{N_p}) \mathbf{W}_p \quad (10b)$$

Thus the MMSE spatial interpolation can be expressed by

$$\hat{\mathbf{h}}_d = \mathbf{R}_{d\hat{h}} \mathbf{R}_{\hat{h}\hat{h}}^{-1} \hat{\mathbf{h}}_p. \quad (11)$$

The cross-correlation matrix, $\mathbf{R}_{dh} = \mathbb{E}(\mathbf{h}_d \hat{\mathbf{h}}_p^H) \in \mathcal{R}^{N_p \times N_p}$, is a Toeplitz matrix with its first row being $[\rho(-K+u), \rho(-2K+u), \dots, \rho(-N_p K+u)]$ and the first column $[\rho(-K+u), \rho(u), \dots, \rho((N_p-2)K+u)]^T$.

Combining (4), (10) and (11) yields

$$\hat{\mathbf{h}}_d = \sqrt{E_0} \mathbf{R}_{dh} \mathbf{P}^H (E_0 \mathbf{P} \mathbf{R}_{hh} \mathbf{P}^H + \sigma_z^2 \mathbf{I}_{N_p})^{-1} \mathbf{y}_p. \quad (12)$$

The corresponding error correlation matrix, $\mathbf{\Psi}_{ee} \triangleq \mathbb{E}[(\hat{\mathbf{h}}_d - \mathbf{h}_d)(\hat{\mathbf{h}}_d - \mathbf{h}_d)^T]$, can then be calculated by

$$\mathbf{\Psi}_{ee} = \mathbf{R}_{hh} - \mathbf{R}_{dh} \left(\mathbf{R}_{hh} + \frac{1}{\gamma_0} \mathbf{I}_{N_p} \right)^{-1} \mathbf{R}_{hd}, \quad (13)$$

where $\mathbf{R}_{dd} = \mathbb{E}(\mathbf{h}_d \mathbf{h}_d^H) = \mathbf{R}_{hh}$ is used in the above equation, and $\mathbf{R}_{hd} = \mathbf{R}_{dh}^H$. The average MSE for spatial interpolation is $\sigma_{e,N_p}^2 = \frac{1}{N_p} \text{trace}(\mathbf{\Psi}_{ee})$. The asymptotic average MSE for channel estimation through temporal interpolation is given in the following proposition.

Proposition 2: When $N_p \rightarrow \infty$ while keeping a finite δ , if $\delta \geq \frac{\alpha}{\pi}$, then channel estimations through temporal interpolation yields the same asymptotic MSE as channel estimations at pilot locations, i.e., $\sigma_e^2 = \lim_{N_p \rightarrow \infty} \sigma_{e,N_p}^2 = \sigma_p^2$, with σ_p^2 defined in (8).

Proof: The proof is in Appendix B. ■

The results in Proposition 2 state that the temporal interpolation will not degrade the channel estimation performance, as long as the channel coefficients are sampled by the pilots at a rate no less than the Nyquist rate, i.e., $\frac{1}{T_p} \geq 2f_d$, or equivalently, $\delta \geq \frac{\alpha}{\pi}$. The temporal interpolation introduces a time shift in the correlation between \mathbf{h}_d and \mathbf{h}_p . A shift in the time domain corresponds to a phase shift in the frequency domain as shown in (33) in Appendix C. The asymptotic MSE is only related to the squared amplitude of the frequency domain representation of the channel correlation as in (34). If there is no spectrum aliasing, then the phase shift does not have any impact on the asymptotic MSE.

IV. IMPACTS OF PILOT PERCENTAGE ON SYMBOL ERROR PROBABILITY

In this section, the statistical properties of the estimated channel are studied, and the results are used to derive the SER in the presence of imperfect CSI.

A. Statistical Properties of the Estimated Channel

To build an explicit relationship between the channel estimation MSE and the SER, we first study the statistical properties of the estimated channel. To simplify notation, the data symbol index is dropped in the subsequent analysis.

Proposition 3: For a system operating in a Rayleigh fading channel, the estimated channel coefficient, \hat{h} , is a complex Gaussian random variable (CGRV) with zero mean and variance $\sigma_{\hat{h}}^2 = 1 - \sigma_e^2$, i.e., $\hat{h} \sim \mathcal{N}(0, 1 - \sigma_e^2)$, where σ_e^2 is the channel estimation MSE.

Proof: The proof is in Appendix C. ■

Corollary 1: Consider a system operating in a Rayleigh fading channel. Conditioned on the estimated channel coefficient \hat{h} , the true channel coefficient h is Gaussian distributed with mean \hat{h} and variance σ_e^2 , i.e., $h|\hat{h} \sim \mathcal{N}(\hat{h}, \sigma_e^2)$.

Proof: The proof is in Appendix D. ■

The receiver performs detection based on the knowledge of the received sample y and the estimated channel coefficient \hat{h} . We have the following corollary regarding the likelihood function, $p(y|\hat{h}, x)$, in the presence of imperfect CSI.

Corollary 2: Consider a system operating in a Rayleigh fading channel. If the channel coefficient is obtained through

MMSE channel estimation, then the likelihood function, $p(y|\hat{h}, x)$, is a Gaussian probability density function (pdf), with the conditional mean, $u_{y|x,\hat{h}}$, and conditional variance, $\sigma_{y|x,\hat{h}}^2$, given by

$$u_{y|x,\hat{h}} = \sqrt{E_0} \hat{h} x, \quad (14)$$

$$\sigma_{y|x,\hat{h}}^2 = E_0 \sigma_e^2 |x|^2 + \sigma_z^2. \quad (15)$$

where σ_e^2 is the channel estimation MSE.

Proof: The proof is in Appendix E. ■

B. SER in the Presence of Imperfect CSI

The SER performance of systems with M -ary phase shift keying (MPSK) modulations and imperfect CSI is studied in this subsection by utilizing the statistical properties of the channel estimation error.

For systems with equiprobable transmitted symbols and imperfect CSI, the SER can be minimized by maximizing the likelihood function, $p(y|\hat{h}, x)$, which is a Gaussian pdf with the conditional mean and variance given in Corollary 2. From Corollary 2, the maximum likelihood decision rule for system with MPSK can be expressed as

$$\hat{x} = \underset{x \in \mathcal{S}}{\operatorname{argmin}} \left\{ \frac{|y - u_{y|x,\hat{h}}|^2}{\sigma_{y|x,\hat{h}}^2} \right\} = \underset{x \in \mathcal{S}}{\operatorname{argmin}} \left\{ |\mu - x|^2 \right\} \quad (16)$$

where \mathcal{S} is the MPSK modulation alphabet set, and $\mu = \frac{1}{\sqrt{E_0}} \hat{h}^* y$ is the decision variable for MMSE channel estimation, with a^* being the complex conjugate operator.

From eqn. (16), the SER can be calculated by finding the probability that the decision variable μ is outside the decision region of the transmitted symbol, thus the SER depends on the statistical properties of μ . Given \hat{h} and the transmitted symbol x , the decision variable μ is also Gaussian distributed with the conditional mean and variance given by

$$u_{\mu|x,\hat{h}} = |\hat{h}|^2 x, \quad (17a)$$

$$\sigma_{\mu|x,\hat{h}}^2 = |\hat{h}|^2 \left(\sigma_e^2 + \frac{1}{\gamma_0} \right). \quad (17b)$$

Note that the identity $|x|^2 = 1$ is used in the above derivation for the MPSK modulated system.

With the statistical properties of the decision variable given in (17), the SER of MPSK modulated systems with the imperfect CSI is given in the following proposition.

Proposition 4: For an MPSK modulated system operating in fast time-varying Rayleigh fading channels, if the channel is estimated with an MMSE estimator, then the SER is

$$P(E) = \frac{1}{\pi} \int_0^{\pi - \frac{\pi}{M}} \left[1 + \zeta \cdot \frac{\sin^2(\frac{\pi}{M})}{\sin^2(\phi)} \right]^{-1} d\phi, \quad (18)$$

where $\zeta = \frac{1 - \sigma_e^2}{\sigma_e^2 + \frac{1}{\gamma_0}}$, and σ_e^2 is the asymptotic MSE of the channel estimation given in Proposition 2.

Proof: The proof is in Appendix F. ■

In Proposition 4, the SER is expressed as a function of the channel estimation MSE σ_e^2 and the SNR γ_0 . Since the asymptotic MSE σ_e^2 is a function of δ and f_d , the SER can be expressed as an explicit function in δ , f_d , and γ_0 .

V. MAXIMIZING SPECTRAL EFFICIENCY WITH IMPERFECT CHANNEL INFORMATION

In this section, we study the optimum pilot design by maximizing a lower bound of the spectral efficiency in the presence of imperfect CSI. A higher pilot percentage yields a better channel estimation, thus less detection errors at the receiver. On the other hand, increasing pilot percentage will decrease the number of data bits transmitted per unit time per unit bandwidth. The optimum pilot percentage that can balance this tradeoff is studied in this section.

Considering the presence of both pilot symbols and channel estimation error, we can calculate the effective system spectral efficiency as

$$\eta = \mathbb{E}_{\hat{h}} \left[\frac{N_s}{N} C(\hat{h}) \right] = (1 - \delta) \mathbb{E}_{\hat{h}} \left[C(\hat{h}) \right], \quad (19)$$

where the expectation is performed with respect to \hat{h} , $C(\hat{h}) = \max_{p(x)} \mathbb{I}(y; x | \hat{h})$ is the maximum mutual information between y and x given the knowledge of the estimated channel coefficient \hat{h} , with $p(x)$ being the pdf of the input x . $C(\hat{h})$ can be considered as the channel capacity in the presence of imperfect CSI, and it quantifies the impact of channel estimation error on the channel capacity.

It is difficult to obtain the exact expression of the conditional channel capacity $C(\hat{h})$. A lower bound on $C(\hat{h})$ is given as follows.

Lemma 1: For a system operating in a Rayleigh fading channel with pilot-assisted MMSE channel estimation, the channel capacity conditioned on the estimated channel coefficient is lower bounded by

$$C_{\text{low}}(\hat{h}) = \log \left(1 + |\hat{h}|^2 \frac{1}{\sigma_e^2 + \frac{1}{\gamma_0}} \right) \quad (20)$$

Proof: The proof is in Appendix G. ■

Based on the results in Lemma 1, a lower bound on the effective spectral efficiency is given by the following proposition.

Proposition 5: For a system that employs MMSE channel estimation and experiences Rayleigh fading, the average spectral efficiency is lower bounded by

$$\eta_{\text{low}} = (1 - \delta) \exp \left(\frac{\sigma_e^2 + \frac{1}{\gamma_0}}{1 - \sigma_e^2} \right) \Gamma \left(0, \frac{\sigma_e^2 + \frac{1}{\gamma_0}}{1 - \sigma_e^2} \right), \quad (21)$$

where $\Gamma(s, x) = \int_x^\infty t^{s-1} e^{-t} dt$ is the incomplete Gamma function.

Proof: The proof is in Appendix H. ■

It is straightforward to show that σ_e^2 is a decreasing function in δ and an increasing function in f_d . Thus $\xi(\delta) = \frac{1 - \sigma_e^2}{\sigma_e^2 + \frac{1}{\gamma_0}}$ is an increasing function in δ . The spectral efficiency lower bound in (21) can be alternatively represented as (46) in Appendix I, which can be decomposed into two components, $g_1(\delta) = (1 - \delta)$ and $g_2(\delta) = \int_0^\infty \exp(-v) \log \left(1 + v \frac{1 - \sigma_e^2}{\sigma_e^2 + \frac{1}{\gamma_0}} \right) dv$. The linear function $g_1(\delta)$ is strictly decreasing in δ , and it contributes to the spectral efficiency loss due to a higher pilot percentage. On the other hand, $g_2(\delta)$ is an increasing function in δ , and it contributes to the spectral efficiency

gain due to a more accurate channel estimation at a higher δ . Therefore, $g_1(\delta)$ and $g_2(\delta)$ reveal two opposite effects of the pilot percentage δ on the spectral efficiency.

The spectral efficiency lower bound is shown as a function of δ in Fig. 2 under various values of the normalized Doppler spread $f_D T_s$, where T_s is the symbol period. For a system with symbol rate 100 ksym/s and operating at 1.9 GHz, the range of the Doppler spread considered in the figure is between 100 Hz ($f_D T_s = 10^{-3}$) to 1 KHz ($f_D T_s = 10^{-2}$), which correspond to a mobile speed in the range between 56.8 km/hr and 568.4 km/hr. For all the system configurations in Fig. 2, it is observed from the numerical results that η_{low} is empirically concave in δ ¹. When δ is small, *e.g.*, $\delta < 0.095$ for $f_D T_s = 0.01$, η_{low} increases in δ . This indicates that, when δ is small enough, the impact of channel estimation error dominates the effective spectral efficiency. On the other hand, when δ becomes large enough, *e.g.*, $\delta > 0.095$ for $f_D T_s = 0.01$, increasing δ further will degrade the spectral efficiency because of the excessive overhead caused by the high percentage of the pilot symbols.

For those system configurations that empirically show a concave behavior of the spectral efficiency lower bound, we can obtain the optimum pilot percentage that maximizes the spectral efficiency lower bound by solving the equation $\frac{\partial \eta_{\text{low}}}{\partial \delta} = 0$, which can be expressed as

$$\exp\left(\frac{\sigma_e^2 + \frac{1}{\gamma_0}}{1 - \sigma_e^2}\right) \Gamma\left(0, \frac{\sigma_e^2 + \frac{1}{\gamma_0}}{1 - \sigma_e^2}\right) \left[\frac{(1 - \delta)(1 + \frac{1}{\gamma_0})}{(1 - \sigma_e^2)^2} (\sigma_e^2)' - 1 \right] = \frac{(1 - \delta)(1 + \frac{1}{\gamma_0})}{(\sigma_e^2 + \frac{1}{\gamma_0})(1 - \sigma_e^2)} (\sigma_e^2)', \quad (22)$$

where $(\sigma_e^2)'$ is the first derivative of σ_e^2 given as

$$(\sigma_e^2)' = \frac{-4\gamma_0\alpha \left[\delta \sqrt{(2\gamma_0)^2 - (\frac{\alpha}{\delta})^2} - 2\alpha \arctan\left(\sqrt{\frac{2\gamma_0 - \frac{\alpha}{\delta}}{2\gamma_0 + \frac{\alpha}{\delta}}}\right) \right]}{\pi\delta^3 \left[(2\gamma_0)^2 - (\frac{\alpha}{\delta})^2 \right]^{\frac{3}{2}}} \quad (23)$$

The eqn. (22) in δ can be solved numerically by using standard software packages, such as `fsolve` in Matlab. It should be noted that (22) should be used only if the spectral efficiency lower bound for a given system configuration is concave in the pilot percentage, and this can be empirically verified through numerical results before the application of (22).

VI. NUMERICAL RESULTS

Numerical and simulation results are provided in this section to verify the results obtained in this paper, and to demonstrate the impacts of imperfect CSI on the system performance.

Fig. 3 plots the numerical and simulated MSE of the channel estimation at pilot and data locations as a function of the pilot percentage, δ , under various values of slot length N . The normalized Doppler spread is $f_D T_s = 0.005$ and the SNR is $\gamma_0 = 10$ dB. The analytical results with finite N are evaluated by using (6) and (7), and the asymptotic analytical results with $N \rightarrow \infty$ are from Proposition 2. In the simulations, a

¹We were not able to analytically show the concavity of η_{low} in δ due to the complex relationship involved between σ_e^2 and δ . However, the concavity is observed in all of our numerical results through extensive simulations and numerical studies, under a wide range of system parameters.

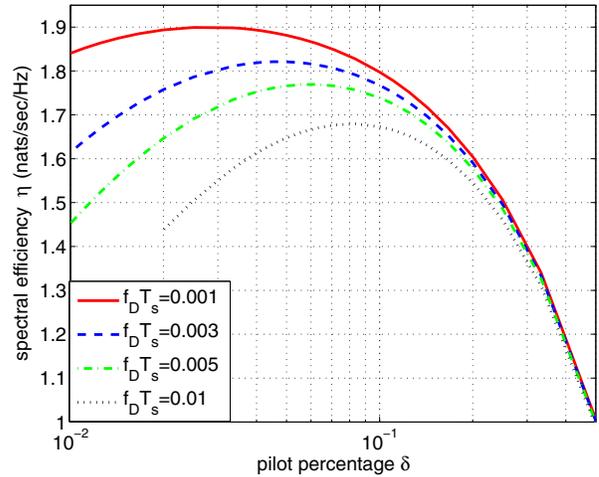


Fig. 2. The spectral efficiency as a function of the pilot percentage.

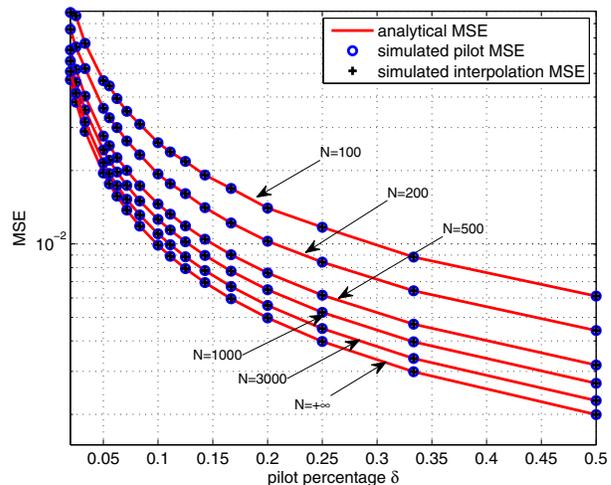


Fig. 3. The MSE of the channel estimation under various symbol number N .

slot length of $N = 75,000$ is used to approximate the infinite slot length. As expected, the MSE is a decreasing function of the pilot percentage. At a given pilot percentage, a larger N leads to a more accurate channel estimation due to the fact that more pilots are available for estimation. When N is large, *e.g.*, $N > 3000$, increasing N further only leads to a slight improvement in the MSE.

Fig. 4 shows the asymptotic MSE in Proposition 2 as a function of pilot percentage, under various values of the normalized Doppler spread $f_D T_s$. The SNR is $\gamma_0 = 10$ dB. The MSE obtained from simulations is also shown in the figure. Both the MSE for channel estimation at pilot positions and the MSE for channel interpolation at non-pilot locations are shown in the figure, and they are the same as predicted by Proposition 2. Excellent agreement is observed between the analytical MSE obtained with infinite frame length and the simulation results with finite frame length. As expected, the MSE is a decreasing function in δ , and an increasing function in $f_D T_s$.

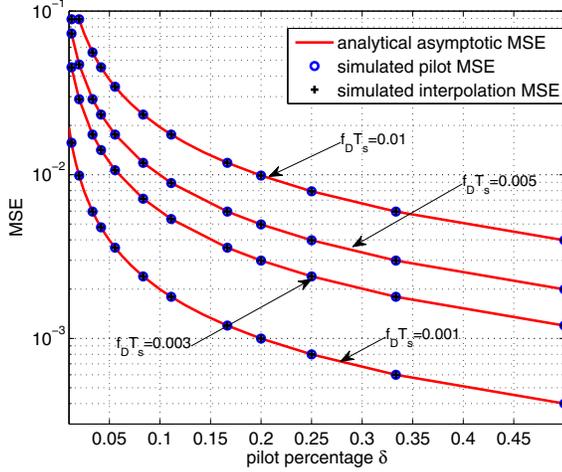


Fig. 4. The asymptotic MSE as a function of the pilot percentage.

The analytical and simulated SER of systems employing binary phase shift keying (BPSK) modulation are shown in Fig. 5. The SNR is $\gamma_0 = 10$ dB. The analytical results can accurately predict the actual SER in the presence of imperfect CSI. Similar to the MSE, the SER is a decreasing function in δ and an increasing function in $f_D T_s$. As δ increases, the SER approaches a lower bound. Since the SER is decreasing in δ , the SER lower bound is achieved at $\delta = 1$, which corresponds to the best possible channel estimation. From (8), the MSE at $\delta = 1$ is

$$\bar{\sigma}_e^2 = 1 - \frac{8\gamma_0 \arctan\left(\sqrt{\frac{2\gamma_0 - \alpha}{2\gamma_0 + \alpha}}\right)}{\pi\sqrt{(2\gamma_0)^2 - \alpha^2}}, \quad (24)$$

Replacing σ_e^2 in (18) with $\bar{\sigma}_e^2$ in the above equation yields the lower bound on the SER. The SER lower bounds are 0.0237, 0.0235, and 0.0233 for systems with $f_D T_s = 0.01$, 0.005, and 0.001, respectively. The SER of systems with perfect CSI is also calculated, and the result is 0.0233, which is independent of δ and coincides with the lower bound for the system with $f_D T_s = 0.001$ and imperfect CSI.

In addition, it is also observed from Fig. 5 that, when δ is small, the SER decreases dramatically as δ increases. When δ reaches a certain threshold, increasing δ further only yields very small performance gains, i.e., the slope of the SER- δ curve approaches zero as δ increases. Therefore, the desired pilot percentage can be chosen as the point such that $\left|\frac{\partial P(E)}{\partial \delta}\right| = \epsilon$, with ϵ being a small number. The slope of the SER- δ curve can be calculated as

$$\frac{\partial P(E)}{\partial \delta} = \frac{1}{\pi} \int_0^{\pi - \frac{\pi}{M}} \left\{ \frac{(1 + \frac{1}{\gamma_0})(\sigma_e^2)' \frac{\sin^2(\frac{\pi}{M})}{\sin^2(\phi)}}{\left[\sigma_e^2 + \frac{1}{\gamma_0} + (1 - \sigma_e^2) \frac{\sin^2(\frac{\pi}{M})}{\sin^2(\phi)}\right]^2} \right\} d\phi \quad (25)$$

where $(\sigma_e^2)'$ is the first derivative of σ_e^2 with respect to δ given in (23). In Fig. 6, the desired pilot percentage is solved by choosing $\epsilon = 10^{-5}$ and shown as functions of the normalized Doppler spread $f_D T_s$ under different SNRs. The desired pilot percentage increases as $f_D T_s$ increases. This is intuitive because a channel that changes faster needs a higher pilot percentage.

Fig. 7 shows the spectral efficiency maximizing pilot percentage as a function of the normalized Doppler spread $f_D T_s$,

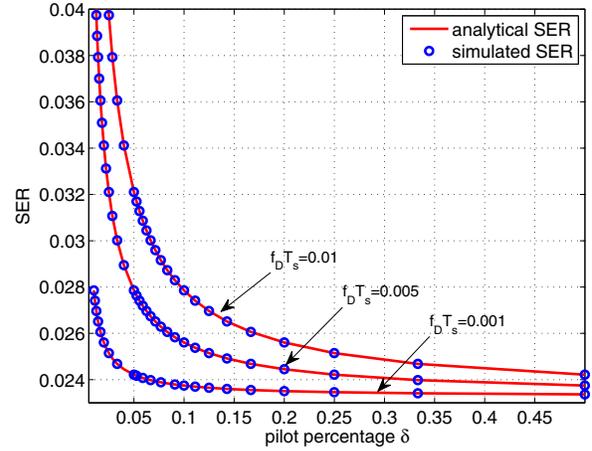


Fig. 5. The SER as a function of the pilot percentage δ .

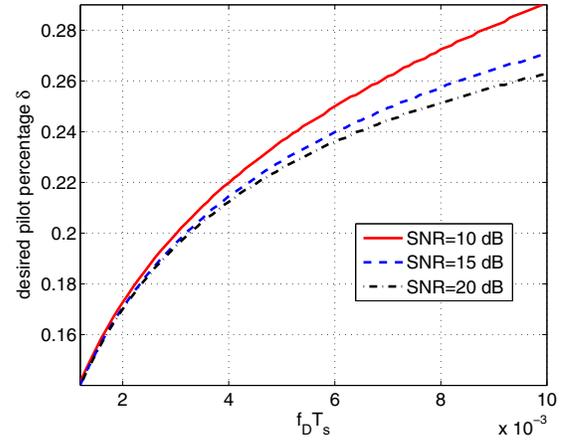


Fig. 6. The desired pilot percentage obtained by solving (25) with $\epsilon = 10^{-5}$.

under various values of SNR. The optimum pilot percentage is calculated by solving (22). The optimum pilot percentage is a monotonically increasing function in $f_D T_s$, because more pilots per unit time are required to compensate the faster channel variation at higher Doppler spreads. At SNR = 10 dB, increasing $f_D T_s$ from 2×10^{-3} to 10^{-2} will double the optimum pilot percentage from 4% to 8%. In addition, a lower pilot percentage is required for systems with higher SNR due to the better channel estimation quality when the SNR is high.

VII. CONCLUSIONS

The optimum system designs for high mobility wireless communication systems with imperfect CSI have been studied in this paper. The asymptotic channel estimation MSE has been quantified as a closed-form expression of the percentage of pilots used for MMSE channel estimation. By analyzing the statistical properties of the estimated channel coefficients, we derived the explicit SER and a spectral efficiency lower bound of communication systems operating with imperfect CSI. It has been shown through theoretical study that, if the pilot samples the channel at a rate no less than the Nyquist rate of the time-varying channel, MMSE channel estimation at pilot locations or MMSE channel interpolation at non-pilot locations yield the same MSE. Numerical results indicated that the imperfect CSI could have significant impacts on

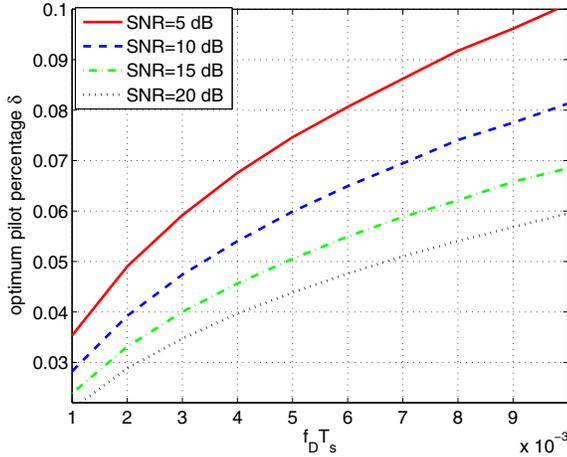


Fig. 7. The optimum pilot percentage as a function of the normalized Doppler spread.

system performance, especially at high Doppler spread. The negative impacts of channel estimation error could be reduced by carefully choosing the pilot percentage.

APPENDIX

A. Proof of Proposition 1

Performing eigenvalue decomposition of \mathbf{R}_{hh} in (6), we can rewrite the MSE as

$$\begin{aligned} \sigma_{p, N_p}^2 &= \frac{1}{N_p} \sum_{n=1}^{N_p} \left[\lambda_n - \left(\lambda_n + \frac{1}{\gamma_0} \right)^{-1} \lambda_n^2 \right] \\ &= \frac{1}{N_p} \sum_{n=1}^{N_p} \left(\frac{\lambda_n}{\lambda_n \gamma_0 + 1} \right), \end{aligned} \quad (26)$$

where λ_n is the n -th eigenvalue of \mathbf{R}_{hh} . Based on Szego's Theorem [15], when $N_p \rightarrow \infty$, (26) can be rewritten as

$$\sigma_p^2 = \lim_{N_p \rightarrow \infty} \sigma_{p, N_p}^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\frac{\Lambda(\Omega)}{\Lambda(\Omega)\gamma_0 + 1} \right] d\Omega, \quad (27)$$

where $\Lambda(\Omega) = \sum_{k=-\infty}^{\infty} J_0(2\pi f_D |k| T_p) e^{-jk\Omega}$ is the discrete-time Fourier transform (DTFT) of $\{J_0(2\pi f_D |k| T_p)\}_k$, with $T_p = \frac{T_s}{2}$ being the space between two pilot symbols.

The Fourier transform (FT) of the continuous-time function $J_0(2\pi f_D t)$ is [17]

$$\Pi(\omega) = \frac{2\text{rect}\left(\frac{\omega}{4\pi f_D}\right)}{\sqrt{(2\pi f_D)^2 - \omega^2}}, \quad (28)$$

where $\omega = \frac{\Omega}{T_p}$ is the analog angular frequency with unit radians per second. From (28), the DTFT $\Lambda(\Omega)$ can then be written as

$$\Lambda(\Omega) = \frac{1}{T_p} \sum_{n=-\infty}^{\infty} \Pi\left(\omega - n \frac{2\pi}{T_p}\right) = \sum_{n=-\infty}^{\infty} \frac{2\text{rect}\left(\frac{\Omega - 2\pi n}{2\beta}\right)}{\sqrt{\beta^2 - (\Omega - 2\pi n)^2}}, \quad (29)$$

where $\beta = 2\pi f_D T_p$.

Based on the sampling theorem, when $\delta \geq 2f_D T_s$, there is no spectrum aliasing in (29), the DTFT in (29) can be simplified to

$$\Lambda(\Omega) = \frac{2\text{rect}\left(\frac{\Omega}{2\beta}\right)}{\sqrt{\beta^2 - \Omega^2}}, \quad -\pi \leq \Omega \leq \pi. \quad (30)$$

The MSE in (27) can then be simplified to

$$\sigma_p^2 = \frac{1}{2\pi} \int_{-\beta}^{\beta} \left[\frac{2}{2\gamma_0 + \sqrt{\beta^2 - \Omega^2}} \right] d\Omega \quad (31)$$

It should be noted that $\beta \leq \pi$ when $\delta \geq 2f_D T_s$.

By changing the integration variable $\Omega = \beta \sin(x)$, we can solve the above integral with the following integration

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} [a + b \cos(x)]^{-1} dx = \frac{4 \arctan\left(\sqrt{\frac{a-b}{a+b}}\right)}{\sqrt{a^2 - b^2}}, \quad (32)$$

where the equation is derived by combining [14, eqn. (2.553.3)] with the identity $\arctan(jx) = j \arctanh(x)$ for $x \in \mathcal{R}$ and $j^2 = -1$. The results in (8) can then be obtained by applying (32) to (31).

B. Proof of Proposition 2

The Toeplitz matrix, \mathbf{R}_{dh} , is uniquely determined by the sequence $\mathbf{t}_{dh} = [t_{-N_p}, \dots, t_0, \dots, t_{N_p-2}]^T$, where $t_k = \rho(kK + u) = J_0(2\pi f_D T_p |k + \frac{u}{K}|)$. The temporal interpolation introduces a time shift, $\frac{u}{K} T_p$, between the sequences \mathbf{t}_{dh} and $\{J_0(2\pi f_D |k| T_p)\}$. A shift in the time domain corresponds to a phase shift in the frequency domain. Therefore, when $N_p \rightarrow \infty$ and $\delta \geq \frac{\alpha}{\pi}$, the DTFT of the sequence \mathbf{t}_{dh} can be calculated as

$$\Lambda_{dh}(\Omega) = \Lambda(\Omega) \times \exp\left(j \frac{u}{K} \Omega\right), \quad -\pi \leq \Omega \leq \pi, \quad (33)$$

where $\Lambda(\Omega)$ is the DTFT of $\{J_0(2\pi f_D |k| T_p)\}$ and it is given in (30).

Based on [15, Lemma 2], \mathbf{R}_{dh} is asymptotically equivalent to a circulant matrix, $\mathbf{C}_{dh} = \mathbf{U}_N^H \mathbf{D}_{dh} \mathbf{U}_N$, where \mathbf{U}_N^H is the unitary discrete Fourier transform (DFT) matrix with the $(m+1, n+1)$ -th element being $(\mathbf{U}_N)_{m+1, n+1} = \frac{1}{\sqrt{N}} \exp[-j2\pi \frac{mn}{N}]$, and \mathbf{D}_{dh} is a diagonal matrix with its k -th diagonal element being $(\mathbf{D}_{dh})_{k,k} = \Lambda_{dh}\left(\frac{k-1}{N}\right)$.

Similarly, the Toeplitz matrix, \mathbf{R}_{hh} , is asymptotically equivalent to a circulant matrix, $\mathbf{C}_{hh} = \mathbf{U}_N^H \mathbf{D}_{hh} \mathbf{U}_N$, where \mathbf{D}_{hh} is a diagonal matrix with its k -th diagonal element being $(\mathbf{D}_{hh})_{k,k} = \Lambda\left(\frac{k-1}{N}\right)$, with $\Lambda(\Omega)$ defined in (29).

Based on [16, Theorem 2.1], the error correlation matrix, Ψ_{ee} , is asymptotically equivalent to a circulant matrix, $\mathbf{C}_{ee} = \mathbf{C}_{hh} - \mathbf{C}_{dh} \left(\mathbf{C}_{hh} + \frac{1}{\gamma_0} \mathbf{I}_{N_p}\right)^{-1} \mathbf{C}_{dh}^H = \mathbf{U}_N^H \mathbf{D}_{ee} \mathbf{U}_N$, where $\mathbf{D}_{ee} = \mathbf{D}_{hh} - \mathbf{D}_{dh} \left(\mathbf{D}_{hh} + \frac{1}{\gamma_0} \mathbf{I}_{N_p}\right)^{-1} \mathbf{D}_{dh}^H$. It is apparent that \mathbf{D}_{ee} is diagonal given that \mathbf{D}_{hh} and \mathbf{D}_{dh} are diagonal.

Based on Szego's Theorem, we have

$$\sigma_e^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\Lambda(\Omega) - \frac{|\Lambda_{dh}(\Omega)|^2}{\Lambda(\Omega) + \frac{1}{\gamma_0}} \right] df. \quad (34)$$

Since $\delta \geq \frac{\alpha}{\pi}$, there is no spectrum aliasing for $\Lambda(\Omega)$ and $\Lambda_{dh}(\Omega)$ when $-\pi \leq \Omega \leq \pi$. As a result, it can be easily seen from (33) that $|\Lambda_{dh}(\Omega)|^2 = |\Lambda(\Omega)|^2$ when $-\pi \leq \Omega \leq \pi$. Therefore (34) can be simplified to (27), and this completes the proof.

C. Proof of Proposition 3

Since \mathbf{h}_p and \mathbf{z}_p are zero mean Gaussian distributed, the received vector corresponding to the pilot symbols, \mathbf{y}_p , is zero mean Gaussian distributed with auto-correlation matrix $\mathbf{R}_{yy} = E_0 \mathbf{P} \mathbf{R}_{hh} \mathbf{P}^H + \sigma_z^2 \mathbf{I}_{N_p}$. From (12), the estimated channel vector $\hat{\mathbf{h}}_d$ is a linear transformation of \mathbf{y}_p , thus $\hat{\mathbf{h}}_d$ is zero mean Gaussian distributed with auto-correlation matrix given by

$$\mathbf{R}_{\hat{d}\hat{d}} = \mathbf{R}_{dh} \left(\mathbf{R}_{hh} + \frac{1}{\gamma_0} \mathbf{I}_{N_p} \right)^{-1} \mathbf{R}_{dh}^H. \quad (35)$$

Combining (13) with (35) yields $\mathbf{R}_{\hat{d}\hat{d}} = \mathbf{R}_{hh} - \Psi_{ee}$. Therefore, $\sigma_{\hat{h}}^2 = \lim_{N \rightarrow \infty} \frac{1}{N} \text{trace}(\mathbf{R}_{\hat{d}\hat{d}}) = 1 - \sigma_e^2$.

D. Proof of Corollary 1

Denote the estimation error vector $\mathbf{e}_d = \hat{\mathbf{h}}_d - \mathbf{h}_d$. Since both $\hat{\mathbf{h}}_d$ and \mathbf{h}_d are zero-mean Gaussian distributed, \mathbf{e}_d is zero-mean Gaussian distributed. The cross-covariance matrix between \mathbf{e}_d and $\hat{\mathbf{h}}_d$ is $\mathbb{E}(\mathbf{e}_d \hat{\mathbf{h}}_d^H) = 0$ by following the orthogonal principal. Therefore, \mathbf{e}_d and $\hat{\mathbf{h}}_d$ are uncorrelated. The conditional mean can then be calculated as $\mathbf{u}_{h|\hat{h}} = \mathbb{E}(\mathbf{h}_d | \hat{\mathbf{h}}_d) = \hat{\mathbf{h}}_d - \mathbb{E}(\mathbf{e}_d | \hat{\mathbf{h}}_d) = \hat{\mathbf{h}}_d$. The auto-covariance matrix is, $\mathbb{E}[(\mathbf{h}_d - \mathbf{u}_{h|\hat{h}})(\mathbf{h}_d - \mathbf{u}_{h|\hat{h}})^H] = \mathbb{E}[(\mathbf{h}_d - \hat{\mathbf{h}}_d)(\mathbf{h}_d - \hat{\mathbf{h}}_d)^H] = \Psi_{ee}$. The conditional variance is thus $\sigma_{h|\hat{h}}^2 = \lim_{N \rightarrow \infty} \frac{1}{N} \text{trace}(\Psi_{ee}) = \sigma_e^2$.

E. Proof of Corollary 2

Since h conditioned on \hat{h} is Gaussian distributed, it is straightforward that $y = \sqrt{E_0} h x + z$ conditioned on \hat{h} and x is Gaussian distributed. The conditional mean and variance can be directly calculated by using the result from Corollary 1.

F. Proof of Proposition 4

Given the estimated CSI \hat{h} and the transmitted symbol x , the conditional SER equals to the probability that the decision variable μ is outside of the decision region of x . Since μ conditioned on \hat{h} and x is Gaussian distributed, the conditional error probability can be written as [7] and [18]

$$P(E|\hat{h}) = \frac{1}{\pi} \int_0^{\pi - \frac{\pi}{M}} \exp \left\{ - \frac{|u_{\mu|x,\hat{h}}|^2 \sin^2(\frac{\pi}{M})}{\sigma_{u|x,\hat{h}}^2 \sin^2(\phi)} \right\} d\phi \quad (36)$$

Substituting the values of $u_{\mu|x,\hat{h}}$ and $\sigma_{u|x,\hat{h}}^2$ from (17) into (36) yields

$$P(E|\hat{h}) = \frac{1}{\pi} \int_0^{\pi - \frac{\pi}{M}} \exp \left[- \frac{|\hat{h}|^2 \sin^2(\frac{\pi}{M})}{(\sigma_e^2 + \frac{1}{\gamma_0}) \sin^2(\phi)} \right] d\phi \quad (37)$$

The unconditional error probability $P(E) = \mathbb{E}[P(E|\hat{h})]$ can then be calculated by

$$P(E) = \frac{1}{\pi} \int_0^{\pi - \frac{\pi}{M}} \int_0^{+\infty} \exp \left[- \frac{\mu \sin^2(\frac{\pi}{M})}{(\sigma_e^2 + \frac{1}{\gamma_0}) \sin^2(\mu)} \right] p_{|\hat{h}|^2}(\mu) d\mu d\phi,$$

where $p_{|\hat{h}|^2}(\mu)$ is the pdf of $|\hat{h}|^2$. From Proposition 3, $\hat{h} \sim \mathcal{CN}(0, 1 - \sigma_e^2)$, thus $|\hat{h}|^2$ is an exponentially distributed random variable with mean $1 - \sigma_e^2$. Performing change of variable $v = \frac{\mu}{1 - \sigma_e^2}$ and integrating with respect to v results in (18).

G. Proof of Lemma 1

The conditional mutual information is defined as

$$\mathbf{I}(y; x | \hat{h}) = \mathbb{E}_{x,y} \left[\log p(y|x, \hat{h}) \right] - \mathbb{E}_{x,y} \left[\log p(y|\hat{h}) \right]. \quad (38)$$

From Corollary 2, $p(y|x, \hat{h})$ is a Gaussian pdf with the conditional mean and variance given in (14). Then

$$\mathbb{E}_{x,y} \left[\log p(y|x, \hat{h}) \right] = \mathbb{E}_x \left[\log \frac{1}{\pi(E_0 \sigma_e^2 |x|^2 + \sigma_z^2) e} \right]. \quad (39)$$

It can be easily shown that (39) is convex in $|x|^2$. Based on Jensen's inequality, we have

$$\mathbb{E}_x \left[\log \frac{1}{\pi(E_0 \sigma_e^2 |x|^2 + \sigma_z^2) e} \right] \geq \log \frac{1}{\pi(E_0 \sigma_e^2 + \sigma_z^2) e}, \quad (40)$$

where $\mathbb{E}(|x|^2) = 1$ is used in the above inequality.

Define

$$\mathbf{I}_{\text{low}}(y; x | \hat{h}) = \log \frac{1}{\pi(E_0 \sigma_e^2 + \sigma_z^2) e} + \mathbb{E} \left[\log \frac{1}{p(y|\hat{h}_d)} \right]. \quad (41)$$

Thus $\mathbf{I}_{\text{low}}(y; x | \hat{h}) \leq \mathbf{I}(y; x | \hat{h})$.

The second term in (41) is the conditional differential entropy of y given \hat{h} . Conditioned on \hat{h} , the conditional mean and variance of y are given by

$$u_{y|\hat{h}} = 0, \quad (42a)$$

$$\sigma_{y|\hat{h}}^2 = E_0(|\hat{h}|^2 + \sigma_e^2) + \sigma_z^2. \quad (42b)$$

Given variance $\sigma_{y|\hat{h}}^2$, it is well known that the entropy of $y|\hat{h}$ is maximized if $y|\hat{h} \sim \mathcal{N}(0, \sigma_{y|\hat{h}}^2)$. In this case,

$$\max \mathbb{E} \left[\log \frac{1}{p(y|\hat{h}_d)} \right] = \log(E_0 |\hat{h}|^2 + E_0 \sigma_e^2 + \sigma_z^2) \quad (43)$$

Thus, (41) can be maximized by

$$C_{\text{low}}(\hat{h}) = \max \mathbf{I}_{\text{low}}(y; x | \hat{h}) = \log \left(1 + |\hat{h}|^2 \frac{1}{\sigma_e^2 + \frac{1}{\gamma_0}} \right). \quad (44)$$

Since $\mathbf{I}_{\text{low}}(y; x | \hat{h}) \leq \mathbf{I}(y; x | \hat{h})$, we have $C_{\text{low}}(\hat{h}) \leq C(\hat{h})$.

H. Proof of Proposition 5

Based on (19) and (20), a lower bound on the effective spectral efficiency can be obtained as

$$\eta_{\text{low}} = (1 - \delta) \int_0^{+\infty} \log \left(1 + x \frac{1}{\sigma_e^2 + \frac{1}{\gamma_0}} \right) p_{|\hat{h}|^2}(x) dx, \quad (45)$$

where $p_{|\hat{h}|^2}(x)$ is the pdf of the exponentially distributed random variable $x = |\hat{h}|^2$ with mean $1 - \sigma_e^2$. With the change of integration variable $v = \frac{x}{1 - \sigma_e^2}$, (45) can be alternatively represented as

$$\eta = (1 - \delta) \int_0^{\infty} \exp(-v) \log \left(1 + v \frac{1 - \sigma_e^2}{\sigma_e^2 + \frac{1}{\gamma_0}} \right) dv \quad (46)$$

Solving the above integral based on the definition of the incomplete Gamma function yields (21).

REFERENCES

- [1] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Consum. Electron.*, vol. 44, pp. 1122–1128, Sept. 1998.
- [2] S. Ohno and G. B. Giannakis, "Capacity maximizing pilots and precoders for wireless OFDM over rapidly fading channels," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2138–2145, Sept. 2004.
- [3] M. Karami, A. Olfat, and N. C. Beaulieu, "Pilot symbol parameter optimization based on imperfect channel state prediction for OFDM systems," *IEEE Trans. Commun.*, vol. 61, pp. 2557–2567, June 2013.
- [4] K. M. Z. Islam, T. Y. Al-Naffouri, and N. Al-Dhahir, "On optimum pilot design for comb-type OFDM transmission over doubly-selective channels," *IEEE Trans. Commun.*, vol. 59, pp. 930–935, Apr. 2011.
- [5] R. J. Baxley, J. E. Kleider, and G. T. Zhou, "Pilot design for OFDM with null edge subcarriers," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 396–405, Jan. 2009.
- [6] Z. Wu, J. He, and G. Gu, "Design of optimal pilot-tones for channel estimation in MIMO-OFDM systems," in *Proc. 2005 IEEE Wireless Commun. and Network. Conf.*, vol. 1, pp. 12–17.
- [7] J. Wu and C. Xiao, "Optimal diversity combining based on linear estimation of Rician fading channels," *IEEE Trans. Commun.*, vol. 56, pp. 1612–1615, Oct. 2008.
- [8] H. Sheng and A. M. Haimovich, "Impact of channel estimation on ultra-wideband system design," *IEEE J. Sel. Topics Signal Process.*, vol. 1, pp. 498–507, Oct. 2007.
- [9] A. Vosoughi and Y. Jia, "How does channel estimation error affect average sum-rate in two-way amplify-and-forward relay networks?" *IEEE Trans. Wireless Commun.*, vol. 11, pp. 1676–1687, May 2012.
- [10] H. Gacanin, M. Salmela, and F. Adachi, "Performance analysis of analog network coding with imperfect channel estimation in a frequency-selective fading channel," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 742–750, Feb. 2012.
- [11] T. Whitworth, M. Ghogho, and D. McLernon, "Optimized training and basis expansion model parameters for doubly-selective channel estimation," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 1490–1498, Mar. 2009.
- [12] F. Qu and L. Yang, "On the estimation of doubly-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 1261–1265, Apr. 2010.
- [13] J. Wu and N. Sun, "Optimum sensor density in distortion tolerant wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 2056–2064, June 2012.
- [14] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 6th ed. Academic Press, 2000.
- [15] H. Gazzah, P. A. Regalia, and J. P. Delmas, "Asymptotic eigenvalue distribution of block Toeplitz matrices and application to blind SIMO channel identification," *IEEE Trans. Inf. Theory*, vol. 47, pp. 1243–1251, Mar. 2001.
- [16] R. M. Gray. *Toeplitz and Circulant Matrices: A Review*. NOW Publishers, 2006. Available: <http://ee.stanford.edu/~gray/>.
- [17] H. O. Beca, "An orthogonal set based on Bessel functions of the first kind," *Univ. Beograd. Public. Elektrotehnicki. Fak. Ser. Mat.*, pp. 85–90, Mar. 1980.
- [18] J. Wu, C. Xiao, and N. C. Beaulieu, "Optimal diversity combining based on noisy channel estimation," in *Proc. 2004 IEEE Intern. Conf. Commun.*, vol. 1, pp. 214–218.



Ning Sun received the B.S. and M.S. degrees in Electrical Engineering from Shandong University of Science and Technology, Qingdao, China, in 2005 and 2008, respectively, and the Ph.D. degree in Electrical Engineering from the University of Arkansas, Fayetteville, USA, in 2013. From 2008 to 2009, she was a research assistant at the Underwater Acoustic Communication Institute, Soongsil University, Seoul, South Korea. She is currently a system engineer at LG Electronics, Chicago, Illinois, U.S.A. Her research interests include distortion-tolerant wireless communications, high mobility communications, and wireless networks.



Jingxian Wu (S'02-M'06) received the B.S. (EE) degree from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 1998, the M.S. (EE) degree from Tsinghua University, Beijing, China, in 2001, and the Ph.D. (EE) degree from the University of Missouri at Columbia, MO, USA, in 2005.

He is currently an Associate Professor with the Department of Electrical Engineering, University of Arkansas, Fayetteville. His research interests mainly focus on wireless communications and wireless networks, including ultra-low power communications, energy efficient communications, high mobility communications, and cross-layer optimization, etc. He is an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, an Associate Editor of the *IEEE Access*, and served as an Associate Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY from 2007 to 2011. He served as a Cochair for the 2012 Wireless Communication Symposium of the IEEE International Conference on Communication, and a Cochair for the 2009 Wireless Communication Symposium of the IEEE Global Telecommunications Conference. Since 2006, he has served as a Technical Program Committee Member for a number of international conferences, including the IEEE Global Telecommunications Conference, the IEEE Wireless Communications and Networking Conference, the IEEE Vehicular Technology Conference, and the IEEE International Conference on Communications.