Training for Antenna Selection in Time-Varying Channels: Optimal Selection, Energy Allocation, and Energy Efficiency Evaluation

Vinod Kristem, Neelesh B. Mehta, *Senior Member, IEEE*, and Andreas F. Molisch, *Fellow Member, IEEE*

***Abstract*—Single receive antenna selection (AS) is a popular method for obtaining diversity benefits without the additional costs of multiple radio receiver chains. Since only one antenna receives at any time, the transmitter sends a pilot multiple times to enable the receiver to estimate the channel gains of its N antennas to the transmitter and select an antenna. In time-varying channels, the channel estimates of different antennas are outdated to different extents. We analyze the symbol error probability (SEP) in time-varying channels of the N-pilot and $(N+1)$-pilot AS training schemes. In the former, the transmitter sends one pilot for each receive antenna. In the latter, the transmitter sends one additional pilot that helps sample the channel fading process of the selected antenna twice. We present several new results about the SEP, optimal energy allocation across pilots and data, and optimal selection rule in time-varying channels for the two schemes. We show that due to the unique nature of AS, the $(N+1)$-pilot scheme, despite its longer training duration, is much more energy-efficient than the conventional N-pilot scheme. An extension to a practical scenario where all data symbols of a packet are received by the same antenna is also investigated.**

***Index Terms*—Training, antenna selection, diversity methods, fading channels, estimation, error analysis, phase shift keying, time-varying channels, energy management.**

I. Introduction

SINGLE receive antenna selection (AS) provides a low hardware complexity solution for exploiting the spatial diversity benefits of multiple receive antennas [2]. It dynamically selects the antenna with the 'best' instantaneous channel gain to the transmitter, and only processes the signal received from it. AS enables a receiver with N antennas to employ only one radio frequency (RF) chain instead of N RF chains. Consequently, AS has been adopted in contemporary wireless systems such as the IEEE 802.11n wireless local area network (WLAN) standard [3], [4]. Despite its lower hardware complexity, AS with perfect channel state information (CSI) achieves full diversity order [5], [6].

In practice, in order to select the best antenna, a pilot-based training scheme is used to estimate the channel gains from the transmitter to all the receive antennas. The low hardware complexity, which is a key motivator for AS, constrains how training gets done for AS since only one antenna can receive at any time instant. Conventionally, in systems such as IEEE 802.11n [7], this is overcome by making the transmitter send a pilot symbol N times, which enables the receiver to *sequentially* estimate the channel gains of its N antennas to the transmitter [8]. The receive antenna is then selected based on these estimates. We shall, therefore, refer to this as the N-pilot scheme. For example, a multiple access control (MAC)-based training protocol is used in IEEE 802.11n for transmit and/or receive antenna selection. In it, a sequence of consecutive training packets, in which pilots are embedded in the physical layer header, are transmitted to obtain the CSI of all antennas [7]. The duration of the AS training phase depends on the data carried by the packets, and can be up to several milliseconds. Antenna switching times are typically much less and, thus, do not contribute much to the training delay.

The longer training overhead required has several ramifications for AS in time-varying channels, given the sensitivity of the system to Doppler. The receiver must not only grapple with the fact that the channel estimates of the antennas, which are used for AS and coherent demodulation, are noisy, but also that they are outdated by different amounts. As a result, selecting the antenna with the largest estimated channel gain is no longer optimal. It was shown in [8] that the SEP-optimal selection rule linearly weights the antennas by a weight that is smaller for antennas that are estimated earlier.

However, several open questions remain about AS in time-varying channels, especially with regard to its energy efficiency. Training for AS not only takes up more time but is also energy-inefficient since a pilot must be transmitted multiple times. It is, therefore, of interest to determine how to optimally allocate energy among pilots and data so as to minimize the SEP. In [9], the optimal allocation for the N-pilot scheme was determined, but only for the time-invariant scenario. While optimal power allocation for pilots and data has been investigated in [10]–[14], AS was not considered.

The training scheme also affects the energy-efficiency of AS, as was shown in [9], which proposed an alternate $(N+1)$-pilot training scheme for AS. In it, the transmitter sends an extra pilot symbol after the first N pilots. The inherent robustness of AS to selection errors enables the transmitter to significantly reduce the energy it allocates to the first N pilots,

Manuscript received July 30, 2012; revised December 24, 2012; February 14 and March 20, 2013. The associate editor coordinating the review of this paper and approving it for publication was Y. Chen.

A part of this paper has appeared in the Asilomar Conf. on Signals, Systems, and Computers, Nov. 2011 [1].

V. Kristem and A. F. Molisch are with the Dept. of Electrical Eng. Univ. of Southern California, CA, USA (e-mail: kristem@usc.edu, molisch@usc.edu).

N. B. Mehta is with the Dept. of Electrical Communication Eng. Indian Institute of Science (IISc), Bangalore, India (e-mail: nbmehta@ece.iisc.ernet.in).

Digital Object Identifier 10.1109/TCOMM.2013.032713.120551

which are used for selecting the antenna. Instead, it boosts the energy of the extra pilot, which yields an accurate channel estimate for the selected antenna, which helps accurately demodulate the data symbols. Optimization of the allocation of energy across the pilots and data symbols led to energy savings as large as 2.5 dB over the N-pilot scheme. Saving energy by introducing an extra pilot – without changing the total energy – is unique to AS. In conventional systems, only the total energy allocated to the pilots matters, but not the number of pilots [15].

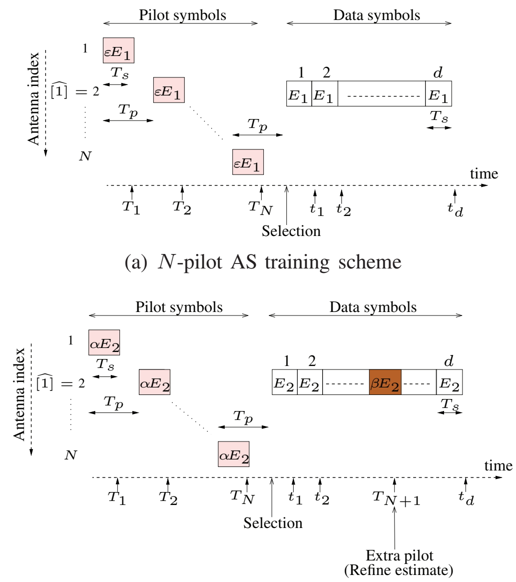
However, [9] focused only on time-invariant channels. In the time-varying scenario, it is not obvious whether the $(N+1)$-pilot scheme remains more energy efficient. One can even argue that sending an extra pilot is counter-productive because extending the training duration makes the channel estimates more outdated. Furthermore, for this scheme, it is not known how to optimally allocate the energy across pilots and data in the time-varying scenario.

In this paper, we analyze, optimize, and compare in *time-varying* channels the performance of the N-pilot and the $(N+1)$-pilot training schemes, which are two known training schemes for AS.[1] We develop a practically relevant perspective about the energy-efficiency of AS training schemes. We show a seemingly counter-intuitive result that the energy-efficiency gain from the $(N+1)$-pilot scheme increases as the Doppler spread increases – despite its larger training overhead. We also derive the following results about AS in time-varying channels:

- We derive a novel, closed-form optimal allocation of pilot and data energies for the N-pilot scheme.
- We derive the optimal antenna selection rule for the $(N+1)$-pilot scheme. As per the rule, an affine function of the estimated channel gain is evaluated for each antenna, and the one with the largest value is selected. To the best of our knowledge, this is the first instance of an affine rule being shown to be optimal in the context of AS.
- We derive novel closed-form expressions for the fading-averaged SEP of the $(N+1)$-pilot scheme for MPSK, and optimize the energy allocation across pilots and data symbols. The model we consider combines the models of [8] and [9]. However, the analysis and the outcomes are quite different.
- We also investigate the problem under the additional practical constraint that all the data symbols in the packet are received by the same antenna. This constraint is motivated by the fact that the time required to switch reception from one antenna to another is not insignificant. We propose a novel AS rule for packet reception that is based on the aforementioned affine selection rule.

While the model we analyze does not consider all the practical aspects of AS such as its use in conjunction with multicarrier modulation, scheduling, energy consumption associated with switching, and RF insertion losses [18], a key thing to note is that our optimal AS rule, SEP analysis, and energy allocation do take into account the imperfection of the channel estimates due to noise and training delays and their impact on both selection and data demodulation.

[1]In the AS model in [16], it was assumed that all the N antennas can receive the pilots simultaneously, but not during the data transmission phase. This requires the use of N low noise amplifiers or analog power detectors at the receiver, and is not considered in this paper. Prediction using discrete prolate spheroidal sequences was facilitated by the AS training scheme considered in [17]. However, we do not consider this scheme since it requires at least $2N$ pilots.



(a) N-pilot AS training scheme

(b) $(N+1)$-pilot AS training scheme

Fig. 1. Training for receive antenna selection: illustration of N-pilot and $(N+1)$-pilot schemes.

The paper is organized as follows. The system model is developed in Sec. II. The AS training schemes are analyzed and optimized in Sec. III. Our results and conclusions follow in Sec. IV and Sec. V, respectively. Mathematical derivations and proofs are relegated to the Appendix.

II. System Model

We use the following notation henceforth. $\Pr(A)$, $\mathbf{E}[A]$, and $\mathbf{var}[A]$ denote the probability, expectation, and variance of A, respectively. Similarly, $\Pr(A|B)$, $\mathbf{E}[A|B]$, and $\mathbf{var}[A|B]$ denote the conditional probability, expectation, and variance of A given B, respectively. The complex conjugate of x is denoted by x^*. The Hermitian transpose and transpose are denoted by $(\cdot)^H$ and $(\cdot)^T$, respectively. The cardinality of a set A is denoted by $|A|$. The covariance matrix of two random vectors $\mathbf{X}$ and $\mathbf{Y}$ is denoted by $\Sigma_{\mathbf{XY}} \triangleq \mathbf{E}\left[(\mathbf{X}-\mathbf{E}[\mathbf{X}])(\mathbf{Y}-\mathbf{E}[\mathbf{Y}])^H\right]$. The indicator function is denoted by $I_{\{x\}}$ – it equals 1 if x is true, and is 0 otherwise. $X \sim \mathcal{CN}(\sigma^2)$ means that X is a zero-mean circularly symmetric complex Gaussian random variable (RV) with variance σ^2.

Consider a system with one transmit antenna, N receive antennas, and one receive RF chain. Let $h_k(t)$ denote the frequency-flat channel between the transmitter and the k^{th} receive antenna at time t. It is modeled as a zero-mean circularly symmetric complex Gaussian RV with unit variance. The channel gains of different receive antennas are assumed to be independent and identically distributed (i.i.d.). The temporal channel correlation $\rho(t_1,t_2) = \mathbf{E}[h_k(t_1)h_k^*(t_2)]$ is assumed to be known both at the transmitter and the receiver.

It is assumed to be real, as is the case for several models such as the classical Jakes' model and Aulin's model, which avoids the singularities associated with the Jakes' model [19, Chp. 5]. Noise is modeled as a zero-mean circularly symmetric complex Gaussian RV with variance N_0, and is independent across the antennas and time. We now describe the two AS training schemes.

A. N-pilot AS Training Scheme

To enable the receiver to estimate the channel gains of all N antennas, the transmitter first transmits N pilot symbols, each of energy εE_1 and duration T_s, as illustrated in Fig. 1(a). Here, E_1 is the energy allocated to a data symbol and $\varepsilon \geq 0$ is the energy scaling factor for the pilot symbols. Two consecutive pilot symbols are separated in time by a duration T_p.

The signal $r_k(T_k)$ received by the k^{th} receive antenna at time T_k is given by

$$r_k(T_k) = \sqrt{\varepsilon E_1} p h_k(T_k) + n_k(T_k), \tag{1}$$

where p is a complex pilot symbol with $|p| = 1$ and $n_k(T_k)$ is Gaussian noise. The minimum mean square error (MMSE) channel estimate $\widehat{h}_k(t_i)$ of $h_k(t_i)$ is given by

$$\widehat{h}_k(t_i) = \frac{\sqrt{\varepsilon E_1} p^* \rho\left(T_k, t_i\right)}{\varepsilon E_1 + N_0} r_k(T_k). \tag{2}$$

Note that $\widehat{h}_k(t_i) \sim \mathcal{CN}(\sigma_{k,i}^2)$, where

$$\sigma_{k,i}^2 = \frac{\varepsilon E_1 \rho\left(T_k, t_i\right)^2}{\varepsilon E_1 + N_0}. \tag{3}$$

The receiver selects one antenna, which we denote by $\widehat{[1]}$. Here, the notation $\widehat{\cdot}$ emphasizes the fact that, in our model, selection is also based on imperfect estimates.

The pilots are followed by d data symbols, each transmitted with energy E_1. They are all received by the same selected antenna. The delay T_p between the N^{th} pilot and the data symbols takes into account the aforementioned antenna switching and selection delays. The received signal $y_{\widehat{[1]}}(t_i)$ at the selected antenna for a data symbol s, transmitted at time t_i, equals

$$y_{\widehat{[1]}}(t_i) = h_{\widehat{[1]}}(t_i) s + n_{\widehat{[1]}}(t_i). \tag{4}$$

The data symbol s is drawn with equal probability from the MPSK constellation of size M and has energy $|s|^2 = E_1$. The total energy constraint then takes the form

$$(N\varepsilon + d) E_1 = E_T. \tag{5}$$

Let $\gamma \triangleq \frac{E_T}{N_0}$.

B. $(N+1)$-Pilot AS Training Scheme

The $(N+1)$-pilot scheme is illustrated in Fig. 1(b). As before, the transmitter now first sequentially transmits N pilot symbols so that all the N channels can be estimated. However, these N pilot symbols, which we shall refer to as *selection pilots*, are now transmitted with energy αE_2. Here, E_2 is the energy per data symbol.

The pilot signal received by the k^{th} receive antenna at time T_k is given by

$$r_k(T_k) = \sqrt{\alpha E_2} p h_k(T_k) + n_k(T_k), \tag{6}$$

where $n_k(T_k)$ is Gaussian noise. As in (2), the MMSE channel estimate $\widehat{h}_k(t_i)$ of $h_k(t_i)$ is given by

$$\widehat{h}_k(t_i) = \frac{\sqrt{\alpha E_2} p^* \rho\left(T_k, t_i\right)}{\alpha E_2 + N_0} r_k(T_k). \tag{7}$$

As before, $\widehat{h}_k(t_i) \sim \mathcal{CN}(\sigma_{k,i}^2)$, where

$$\sigma_{k,i}^2 \triangleq \mathbf{var}\left[\widehat{h}_k(t_i)\right] = \frac{\alpha E_2 \rho\left(T_k, t_i\right)^2}{\alpha E_2 + N_0}. \tag{8}$$

Based on the channel estimates $\widehat{h}_1(t_i), \ldots, \widehat{h}_N(t_i)$, the receiver selects the antenna that will be used to receive the extra pilot symbol and the first data symbol. The selection rule is derived in Sec. III-B2. As before, let $\widehat{[1]}$ denote the index of the selected antenna.

Extra Pilot: The key difference in the $(N+1)$-pilot scheme is that an extra pilot symbol is transmitted at time T_{N+1} with energy βE_2, and is received only by the *selected antenna*. The extra pilot helps in refining the channel estimate of the selected antenna. Note that, in general, $\beta \neq \alpha$. The received signal for the extra pilot is

$$r_{\widehat{[1]}}\left(T_{N+1}\right) = \sqrt{\beta E_2} p h_{\widehat{[1]}}\left(T_{N+1}\right) + n_{\widehat{[1]}}(T_{N+1}), \tag{9}$$

where $n_{\widehat{[1]}}(T_{N+1})$ is Gaussian noise. In general, this pilot can be transmitted anywhere in the d data symbol packet.[2]

Data Reception: The selection pilots are followed by d data symbols, each of energy E_2. Notice that there is no delay between the extra pilot and the data symbols. This is because AS and switching happen after the transmission of the N^{th} pilot, and the $(N+1)^{\text{th}}$-pilot and all the data symbols are received by the same selected antenna. The signal received by the selected antenna for the i^{th} data symbol is

$$y_{\widehat{[1]}}(t_i) = h_{\widehat{[1]}}(t_i) s + n_{\widehat{[1]}}(t_i), \tag{10}$$

where $|s|^2 = E_2$. The total energy constraint is now

$$E_T = (N\alpha + \beta + d) E_2. \tag{11}$$

As before, let $\gamma \triangleq \frac{E_T}{N_0}$.

III. SEP Analysis and Optimal Power Allocation for Training Schemes

We now analyze the fading-averaged SEP for MPSK for receive AS with imperfect CSI for the two AS training schemes, and optimize their parameters to minimize the SEP.

We first focus on the simpler and tractable problem of selecting an antenna and optimizing the energy allocation to minimize the SEP of the i^{th} data symbol, where $1 \leq i \leq d$. We do this without considering the impact of the selection on the SEPs of the other data symbols. The analytical insights obtained from this investigation shall help tackle in Sec. III-C the problem of selection and optimal energy allocation when all the d data symbols of the packet are received by the same selected antenna.

[2] In general, more than one pilot symbol can be embedded in the data packet, but the analysis becomes intractable.

A. N-pilot Training Scheme

Given the analyses in [8], [15], which develop the optimal selection rule and the SEP, we only focus on the derivation of the optimal value of ε that minimizes the SEP subject to a total energy constraint. The maximum likelihood (ML) decision variable $\mathcal{D}$ for decoding the i^{th} data symbol is $\mathcal{D} = \hat{h}^*_{\widehat{[1]}}(t_i) y_{\widehat{[1]}}(t_i)$.

The SEP-optimal selection rule for the i^{th} symbol in time-varying channels is given by [8]

$$\widehat{[1]} = \arg\max_{1\le k\le N} w_{k,i} \left|\hat{h}_k(t_i)\right|^2, \tag{12}$$

where $w^*_{k,i}(\gamma) = \frac{1}{1-\sigma^2_{k,i}+(N\varepsilon+d)\gamma^{-1}}$. Note that the optimal weights depend on t_i and are, thus, different for different data symbols in the packet.

Upon substituting the optimal weights in [8, (28)], the SEP of the i^{th} data symbol with optimal selection is given by[3]

$$P_i^{\varepsilon}(\gamma) = \frac{1}{\pi}\sum_{k=1}^{N}\sum_{l=0}^{N-1}\sum_{m=1}^{\left|\mathcal{S}_l^k\right|}\int_0^{\frac{M-1}{M}\pi}(-1)^l\left[1+\frac{\sigma^2_{k,i}\frac{\sin^2\left(\frac{\pi}{M}\right)}{\sin^2\theta}}{1-\sigma^2_{k,i}+(N\varepsilon+d)\gamma^{-1}} + \sum_{n\in\mathcal{S}_l^k(m)}\frac{1-\sigma^2_{n,i}+(N\varepsilon+d)\gamma^{-1}}{1-\sigma^2_{k,i}+(N\varepsilon+d)\gamma^{-1}}\frac{\sigma^2_{k,i}}{\sigma^2_{n,i}}\right]^{-1} d\theta, \tag{13}$$

where $\mathcal{S}_l^k$ denotes the set of all l-element subsets formed from the set $\{1,2,\ldots,N\}\setminus\{k\}$ and $\mathcal{S}_l^k(m)$ is the m^{th} l-element subset. We now state the first main result in the paper.

Theorem *1:* The optimal value of ε that minimizes the SEP of the N-pilot training scheme is given by

$$\varepsilon^*(\gamma) = \sqrt{\frac{d(\gamma+d)}{N(\gamma+N)}}. \tag{14}$$

Proof: The proof is given in Appendix A. ■

Note that the optimal value does not depend on the constellation size M, the index of the data symbol i, and the channel correlation coefficients.

B. $(N+1)$-pilot Scheme: SEP Analysis and Optimal AS Rule

In the $(N+1)$-pilot scheme, the channel estimate of the selected antenna $\widehat{[1]}$ can be refined using the observation $r_{\widehat{[1]}}(T_{N+1})$, which is obtained from the extra pilot, in addition to the observation $r_{\widehat{[1]}}\left(T_{\widehat{[1]}}\right)$, which is obtained from a selection pilot sent earlier.

Lemma *1:* Let $k = \widehat{[1]}$. The refined MMSE channel estimate $\hat{\hat{h}}_k(t_i)$ of the k^{th} antenna's channel gain at time t_i, obtained from the observations $r_k(T_k)$ and $r_k(T_{N+1})$, is

$$\begin{aligned}\hat{\hat{h}}_k(t_i) = {}& p^*\left[\alpha\beta\left(1-\rho(T_k,T_{N+1})^2\right)+(\alpha+\beta)\frac{N_0}{E_2}+\frac{N_0^2}{E_2^2}\right]^{-1}\\ &\times\left[\frac{r_k(T_k)\sqrt{\alpha}}{\sqrt{E_2}}\left(\beta\left(\rho(T_k,t_i)-\rho(T_{N+1},t_i)\rho(T_k,T_{N+1})\right)\right.\right.\\ &\left.+\frac{N_0}{E_2}\rho(T_k,t_i)\right)+\frac{r_k(T_{N+1})\sqrt{\beta}}{\sqrt{E_2}}\left(\frac{N_0}{E_2}\rho(T_{N+1},t_i)\right.\\ &\left.\left.+\alpha\left(\rho(T_{N+1},t_i)-\rho(T_k,t_i)\rho(T_k,T_{N+1})\right)\right)\right].\end{aligned} \tag{15}$$

Furthermore, $\hat{\hat{h}}_k(t_i) \sim \mathcal{CN}(\sigma'^2_{k,i})$, where

$$\begin{aligned}\sigma'^2_{k,i} = {}& \left(\alpha\beta\left(1-\rho(T_k,T_{N+1})^2\right)+(\alpha+\beta)\frac{N_0}{E_2}+\frac{N_0^2}{E_2^2}\right)^{-1}\\ &\times\left[\left(\alpha\rho(T_k,t_i)^2+\beta\rho(T_{N+1},t_i)^2\right)\frac{N_0}{E_2}+\alpha\beta\left(\rho(T_k,t_i)^2\right.\right.\\ &\left.\left.+\rho(T_{N+1},t_i)^2-2\rho(T_{N+1},t_i)\rho(T_k,T_{N+1})\rho(T_k,t_i)\right)\right].\end{aligned} \tag{16}$$

Proof: The proof is given in Appendix B. ■

1) Decision Variable and Statistics: The maximum likelihood (ML) decision variable $\mathcal{D}$ for detecting the i^{th} data symbol received at time t_i is

$$\mathcal{D} = \hat{\hat{h}}^*_{\widehat{[1]}}(t_i) y_{\widehat{[1]}}(t_i). \tag{17}$$

As shown in Appendix C, when conditioned on $\widehat{[1]}$, $\hat{\hat{h}}_{\widehat{[1]}}(t_i)$, and s, the decision variable $\mathcal{D}$ is a complex Gaussian RV with mean and variance given by

$$\mu_{\mathcal{D}} \triangleq \mathbf{E}\left[\mathcal{D}\,\middle|\,\widehat{[1]}=k, \hat{\hat{h}}_k(t_i), s\right] = \left|\hat{\hat{h}}_k(t_i)\right|^2 s, \tag{18}$$

$$\sigma^2_{\mathcal{D}} \triangleq \mathbf{var}\left[\mathcal{D}\,\middle|\,\widehat{[1]}=k, \hat{\hat{h}}_k(t_i), s\right] = \left|\hat{\hat{h}}_k(t_i)\right|^2\left[\left(1-\sigma'^2_{k,i}\right)E_2+N_0\right]. \tag{19}$$

Therefore, the SEP, conditioned on $\widehat{[1]}$ and $\hat{\hat{h}}_{\widehat{[1]}}(t_i)$, can be shown to be [20, (40)]

$$\begin{aligned}&P_i^{\alpha,\beta}\left(\gamma\,\middle|\,\widehat{[1]}=k, \hat{\hat{h}}_k(t_i)\right)\\ &= \frac{1}{\pi}\int_0^{\frac{M-1}{M}\pi}\exp\left(\frac{-\left|\mu_{\mathcal{D}}\right|^2\sin^2\left(\frac{\pi}{M}\right)}{\sigma^2_{\mathcal{D}}\sin^2\theta}\right)d\theta,\\ &= \frac{1}{\pi}\int_0^{\frac{M-1}{M}\pi}\exp\left(\frac{-\left|\hat{\hat{h}}_k(t_i)\right|^2\sin^2\left(\frac{\pi}{M}\right)/\sin^2\theta}{1-\sigma'^2_{k,i}+(N\alpha+\beta+d)\gamma^{-1}}\right)d\theta.\end{aligned} \tag{20}$$

2) Optimal AS Rule: We now derive an expression for the SEP of the i^{th} data symbol if antenna k is selected.

Lemma *2:* The conditional SEP of the i^{th} data symbol given that antenna k is selected and its channel estimate from

[3]The model in [8] is slightly different from the N-pilot scheme used in the paper. Hence, the following minor changes have to be made to derive (13) from [8, (28)]. The SNR in [8] is defined as $\gamma = \frac{E_1}{N_0}$. Since $E_1 = \frac{E_T}{N\varepsilon+d}$, γ in [8, (28)] has to be replaced by $\frac{\gamma}{N\varepsilon+d}$ in our paper. Also, the AS rule in [8] weights the channel estimates $\hat{h}_k(T_k)$ instead of $\hat{h}_k(t_i)$, as done in the paper. Since $\hat{h}_k(t_i)$ is proportional to $\rho(T_k,t_i)\hat{h}_k(T_k)$, the optimal weights to be used in our paper are $w_{k,i} = \frac{\rho(T_k,t_i)^2}{1-\sigma^2_{k,i}+(N\varepsilon+d)\gamma^{-1}} \equiv \frac{\sigma^2_{k,i}}{1-\sigma^2_{k,i}+(N\varepsilon+d)\gamma^{-1}}$.

the selection pilot is $\widehat{h}_k(t_i)$ is given by

$$P_i^{\alpha,\beta}\left(\gamma \,\middle|\, \widehat{[1]}=k, \widehat{h}_k(t_i)\right) = \frac{1}{\pi}\int_0^{\frac{M-1}{M}\pi}\left(1+\frac{c_k}{\sin^2\theta}\right)^{-1} \times \exp\left(\frac{-\left|\widehat{h}_k(t_i)\right|^2 \sin^2\left(\frac{\pi}{M}\right)/\left(c_k+\sin^2\theta\right)}{1-\sigma'^2_{k,i}+(N\alpha+\beta+d)\gamma^{-1}}\right) d\theta, \quad (21)$$

where $c_k \triangleq \left(\sigma'^2_{k,i}-\sigma^2_{k,i}\right)\frac{\sin^2\left(\frac{\pi}{M}\right)}{1-\sigma'^2_{k,i}+(N\alpha+\beta+d)\gamma^{-1}}$.

Proof: The proof is relegated to Appendix D. ■

The above expression is in the form of a single integral, minimizing which is analytically intractable. However, its Chernoff upper bound is much more insightful for deriving the optimal AS rule, as we show below. Replacing θ with $\pi/2$ yields the following upper bound:

$$P_i^{\alpha,\beta}\left(\gamma \,\middle|\, \widehat{[1]}=k, \widehat{h}_k(t_i)\right) \leq \frac{M-1}{M}\frac{1}{1+c_k} \times \exp\left(\frac{-\left|\widehat{h}_k(t_i)\right|^2 \sin^2\left(\frac{\pi}{M}\right)/\left(1+c_k\right)}{1-\sigma'^2_{k,i}+(N\alpha+\beta+d)\gamma^{-1}}\right). \quad (22)$$

From this, the optimal selection rule easily follows, and is captured in the result below.[4]

Theorem *2:* The optimal selection rule that minimizes the SEP of the i^{th} data symbol, which is transmitted at time t_i, is

$$\widehat{[1]} = \arg\max_{1\leq k\leq N}\left(a_k + b_k\left|\widehat{h}_k(t_i)\right|^2\right), \quad (23)$$

where $a_k = \ln(1+c_k)$, $b_k = \frac{\sin^2\left(\frac{\pi}{M}\right)}{(1+c_k)\left(1-\sigma'^2_{k,i}+(N\alpha+\beta+d)\gamma^{-1}\right)}$, and c_k has been defined in Lemma 2.

Proof: The proof is relegated to Appendix E. ■

Comments: Note that the optimal AS rule is an *affine* function of the channel estimate. An antenna k is now associated with a weight b_k and an offset a_k. This is different from the optimal *linear weighted* selection rule in (12) for the N-pilot scheme and the no-weighting rule. Note also that the selected antenna depends on the index of the data symbol i. For the time-invariant channel, the affine rule can be shown to reduce to the conventional no-weighting rule. For $\beta = 0$ and $\alpha = \varepsilon$, it can be seen that the two training schemes deliver the same SEP, and the affine selection rule is equivalent to the linear weighted selection rule in (12).

Observe that the optimal selection rule requires the knowledge of the temporal channel correlation coefficient $\rho(T_k, t_i)$ at the receiver. It can be estimated from the channel gain because it depends on array geometry, antenna radiation pattern, and the scattering environment, and, thus, changes on a much slower time scale than the channel gains [21], [22].

3) SEP with Optimal Selection: We now derive the SEP of the i^{th} data symbol for the $(N+1)$-pilot scheme with optimal selection.

[4]Note that the rule minimizes an upper bound on the SEP and not the SEP itself. However, for the sake of brevity, we henceforth do not distinguish between the two.

Theorem *3:* The fading-averaged SEP of the i^{th} data symbol transmitted at time t_i with the optimal affine AS rule is

$$P_i^{\alpha,\beta}(\gamma) = \frac{1}{\pi}\sum_{k=1}^{N}\sum_{l=0}^{N-1}\frac{(-1)^l}{\sigma^2_{k,i}}\sum_{m=1}^{\left|\mathcal{S}_l^k\right|}\exp\left(-\sum_{z\in\mathcal{S}_l^k(m)}\frac{a_k-a_z}{b_z\sigma^2_{z,i}}\right) \times \int_0^{\frac{M-1}{M}\pi}\left(1+\frac{c_k}{\sin^2\theta}\right)^{-1}\frac{\exp\left(-g_k f\left(\mathcal{S}_l^k(m)\right)\right)}{f\left(\mathcal{S}_l^k(m)\right)}d\theta, \quad (24)$$

where $g_k = \max_{l=1,\ldots,N, l\neq k}\frac{a_l-a_k}{b_k}$ and $f\left(\mathcal{S}_l^k(m)\right) = \frac{1}{\sigma^2_{k,i}} + \sum_{z\in\mathcal{S}_l^k(m)}\frac{b_k}{b_z\sigma^2_{z,i}} + \frac{\sin^2\left(\frac{\pi}{M}\right)}{(\sin^2\theta+c_k)\left(1-\sigma'^2_{k,i}+(N\alpha+\beta+d)\gamma^{-1}\right)}$.

Proof: The proof is given in Appendix F. ■

In the above expression, we see that the SEP depends on α and β through several variables such as a_k, b_k, c_k, $\sigma^2_{k,i}$, and $\sigma'^2_{k,i}$. Altogether, the SEP expression is considerably more involved than for the N-pilot scheme. Consequently, it is not possible to derive closed-form expressions for the optimal values of α and β, unlike the N-pilot scheme. The optimal values are, therefore, found numerically by doing a gradient search using the closed-form SEP expression. Note that this is computationally much simpler than determining the optimal values using, for example, brute-force multiple Monte Carlo simulations of the SEP over a grid of values of α and β. Also, the optimal values of α and β depend on the index of the data symbol i. This issue will be discussed in detail in Sec. IV.

C. Selection to Minimize Error Rates for Packet Reception

In practice, the switching times required to switch reception between antennas can be significant relative to the symbol duration. Furthermore, in the $(N+1)$-pilot scheme, once the $(N+1)^{\text{th}}$ pilot is received by the selected antenna, it cannot be changed during data transmission since its channel estimate is good only for the selected antenna. These constraints motivate the practical restriction that *all the data symbols of a packet must be received by the same antenna*. The above constraint introduces a new twist to the selection problem since different symbols of a packet experience different training delays. The same antenna may not be SEP-optimal for receiving data symbols transmitted at different times.

Since the same antenna is used to receive all the d symbols, the performance metric now becomes the fading-averaged packet error rate (PER). As was the case for the analytically simpler N-pilot scheme [8], deriving a closed-form PER expression and finding the optimal selection rule for packet reception is not possible. The PER cannot be obtained from the fading-averaged SEPs of the different symbols because the channels seen by different symbols are correlated. We, therefore, use the results about optimizing the SEP to propose the following heuristic solution to this problem.

Consider an ideal receiver in which the antennas can be switched on a symbol-by-symbol basis. In this case, the PER is minimized by using for each symbol index its SEP-optimal antenna. This involves evaluating for each symbol index its optimal selection weight pair a_k and b_k and then determining the best antenna for each symbol as per Theorem 2. Thereafter, *the antenna that is SEP-optimal for the most number of data*

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

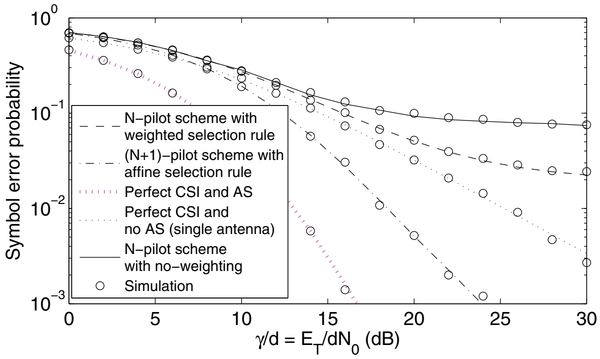


Fig. 2. Benchmarking of the SEPs of the AS training schemes (8PSK, $N = 4$, $f_dT_p = 0.03$, $T_p = 10T_s$, $d = 10$, and first data symbol ($i = 1$)).

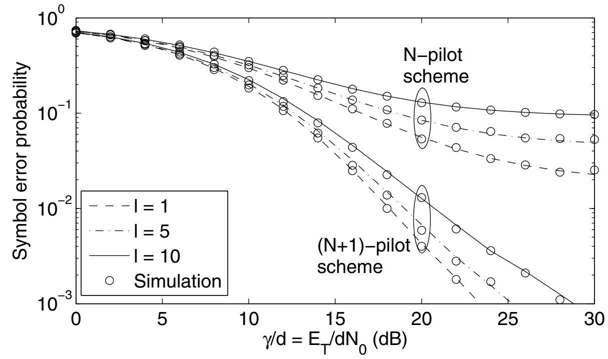


Fig. 3. Comparison of the SEPs of the first, fifth, and tenth data symbols of the N-pilot and $(N + 1)$-pilot AS training schemes (8PSK, $N = 6$, $T_p = 10T_s$, $f_dT_p = 0.03$, and $d = 10$).

symbols in the packet is selected for receiving the entire packet. This rule applies to both the AS training schemes.

The proposed AS rule for packet reception is computationally more intensive than just selecting the antenna with the highest estimated channel gain. This is a small price to pay for the significantly better performance, as we shall see in Sec. IV. A computationally simpler approach is to compute the SEP-optimal antenna only once every l data symbols.

IV. Simulation Results

We now present graphically the results derived in Sec. III, and study the effect of parameters such as f_dT_p, M, and N on the SEP and the PER. We also validate these results with the Monte Carlo simulations that use 10^5 samples for each value of $\gamma = E_T/N_0$. The modified Jakes' simulator of [23] is used to generate multiple time-varying Rayleigh fading channels with zero cross-correlation for the different receive antennas. The correlation $\rho(t_1, t_2)$ between channel gains at time instants t_1 and t_2 is given by $\rho(t_1, t_2) = \mathbf{E}\left[h_k(t_1)h_k(t_2)^*\right] = J_0\left(2\pi f_d(t_1 - t_2)\right)$, where $J_0(.)$ is the zeroth order Bessel function of the first kind and f_d is the Doppler spread. We compare the performance of the AS training schemes for f_dT_p ranging from 0.01 (low mobility) to 0.06 (high mobility).[5] We first study the SEP and then the PER. We first consider the scenario where the $(N+1)^{\text{th}}$ pilot is sent before the data symbols. Later, in Figure. 10, we also optimize the location of this pilot.

Figure 2 plots the SEP of the first data symbol ($i = 1$) of the N-pilot and $(N + 1)$-pilot schemes as a function of $\frac{\gamma}{d}$ for optimal values of their respective parameters. Also plotted are the SEPs of the following benchmark schemes: (i) N-pilot scheme without weighting, (ii) Ideal AS with perfect noise-free and delay-free CSI, and (iii) Ideal single antenna system with perfect CSI. In the ideal schemes above, no energy is assumed to be spent on the pilots. Analytical and simulation results are shown for all the schemes.

[5]For a carrier frequency of $f_c = 2.4$ GHz, which corresponds to the band in which many WLANs operate, the Doppler spread at a pedestrian speed of 3 kmph is $f_d = f_c\frac{v}{c} = 6.67$ Hz. For $T_p = 2$ ms, this corresponds to $f_dT_p = 0.013$. Note that T_p is of the order of milliseconds in the AS training protocols used in the IEEE 802.11n WLAN standard [7] and the Long Term Evolution (LTE) cellular standard [18]. In cellular systems, which can handle higher Doppler spreads, the value of f_dT_p can be larger.

We see that the $(N + 1)$-pilot scheme with its optimal affine selection rule outperforms the N-pilot scheme with its optimal weighted selection rule, which, in turn, outperforms the N-pilot scheme with the no-weighting selection rule. Furthermore, the performance gains increase as the SNR increases. This is because at high SNRs, the outdated estimates drive the SEP. The $(N + 1)$-pilot scheme, in which the last pilot is used for refining the channel estimate of the selected antenna, gives relatively less outdated estimates than the N-pilot scheme. At an SEP of 0.01, the $(N + 1)$-pilot system is 8 dB better than the idea single antenna system. Thus, antenna selection is beneficial even when the RF insertion loss, which arises due to the introduction of a selection switch, is of the order of a few dBs. For RF micro-electro-mechanical systems switches, the insertion loss can be less than a dB.

The SEPs of the first, fifth, and tenth data symbols are plotted in Figure 3 as a function of the symbol-normalized total energy $\frac{\gamma}{d}$. This is done for both the AS training schemes. The antenna is selected to minimize the SEP of the data symbol under consideration. We see that the performance gains of the $(N + 1)$-pilot scheme over the N-pilot scheme increase for later data symbols. This is intuitive because these symbols suffer more from the outdated estimates in the N-pilot scheme than in the $(N + 1)$-pilot scheme, which obtains a refined estimate of the selected antenna just before data transmission starts. Furthermore, the $(N + 1)$-pilot scheme leads to an order of magnitude lower SEP compared to the N-pilot scheme. Since the trends are similar for the different data symbols, we focus on the first out of d data symbols below.

Figure 4 studies the effect of Doppler spread on the SEP of the N-pilot and $(N + 1)$-pilot schemes for optimal values of their respective parameters. We see that, for $f_dT_p = 0.01$ and SEP of 0.001, the $(N+1)$-pilot scheme requires 3 dB less total energy than the N-pilot scheme. This corresponds to energy savings of 50% over the N-pilot scheme, which is significant. Furthermore, these energy savings increase as the Doppler spread increases or the constellation size increases (not shown in the figures due to space constraints). This is again because the $(N + 1)$-pilot scheme produces less outdated estimates for coherent demodulation than the N-pilot scheme, where any of the N estimates – and not just the one from the $(N + 1)^{\text{th}}$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

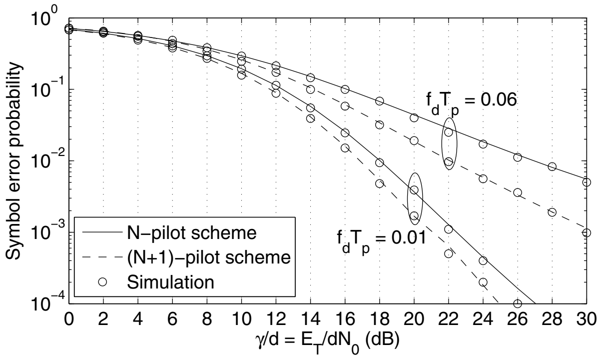


Fig. 4. Effect of Doppler spread on the SEPs of the N-pilot and $(N+1)$-pilot AS training schemes (8PSK, $N = 4$, $T_p = 10T_s$, $d = 10$, and first data symbol ($i = 1$)).

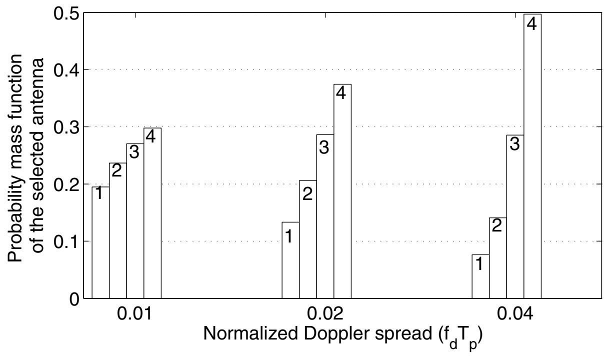


Fig. 5. $(N+1)$-pilot scheme: Probability of selecting different antennas by the optimal affine selection rule for the first data symbol (8PSK, $N = 4$, $T_p = 10T_s$, and $\gamma/d = 20$ dB, $i = 1$, and $d = 50$).

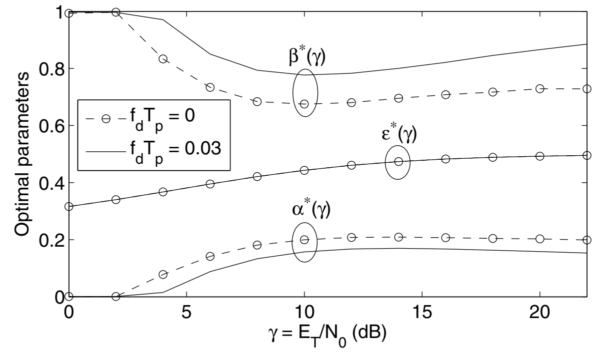


Fig. 6. Optimal values of the parameters of the N-pilot and $(N+1)$-pilot AS training schemes (8PSK, $N = 4$, $T_p = 10T_s$, and $d = 1$).

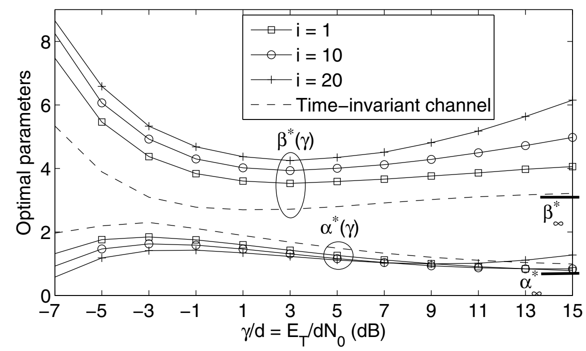


Fig. 7. Optimal energy allocation for the $(N+1)$-pilot scheme for different data symbols of a packet (8PSK, $N = 4$, $T_p = 10T_s$, $f_dT_p = 0.03$, and $d = 20$).

pilot – could be used for coherent reception. Notice again the good match between the analytical and simulation curves. We, therefore, show only the analytical curves henceforth.

Figure 5 plots the probability mass function (pmf) of the index of selected antenna for the $(N+1)$-pilot scheme. Three different values of the Doppler spread are considered. From the figure, it is clear that in low mobility regimes, the pmf is almost uniform. In high mobility regimes, the antenna that is trained last has higher probability of getting selected, since its estimates are the least outdated.

Figure 6 plots the optimal energy allocation – $\varepsilon^*(\gamma)$ for the N-pilot scheme and $\alpha^*(\gamma)$ and $\beta^*(\gamma)$ for the $(N+1)$-pilot scheme – as a function of $\gamma/d = E_T/(dN_0)$ for two values of f_dT_p. We see that $\varepsilon^*(\gamma)$ does not depend on the Doppler spread (cf. (14)). It is a monotonically non-decreasing function of γ. However, $\alpha^*(\gamma)$ and $\beta^*(\gamma)$ do depend on the Doppler spread. As f_dT_p increases, the system allocates less energy to the selection pilots and more energy to the $(N+1)^{\text{th}}$ pilot. This is because the $(N+1)^{\text{th}}$ pilot provides relatively less outdated estimates than the N selection pilots that precede it.

Figure 7 compares the SEP-optimal parameters $\alpha^*(\gamma)$ and $\beta^*(\gamma)$ for different data symbols of a packet. Also plotted are the optimal parameters for the time-invariant channel ($f_d = 0$) in the asymptotic regime of large γ. In this regime, closed-form expressions for the optimal parameters of the $(N+1)$-pilot scheme are known [9]. We shall denote them by α^*_∞ and β^*_∞. They are given as $\alpha^*_\infty = \sqrt{\frac{d}{N}}\sin\left(\frac{\pi}{M}\right)$ and $\beta^*_\infty = -\sqrt{\frac{d}{N}}\sin\left(\frac{\pi}{M}\right) + \sqrt{d}\sqrt{1-\left(1-\frac{1}{N}\right)\sin^2\left(\frac{\pi}{M}\right)}$. Notice that the optimal values of the parameters do not vary much for different data symbols, and are close to α^*_∞ and β^*_∞, which does not depend on i. We exploit this behavior below, which avoids having to compute the optimal α and β numerically from (24).

All Symbols Received by Same Antenna: We now investigate the scenario where all the data symbols of a packet are received by the same selected antenna. Figure 8 compares the uncoded PER of the AS training schemes for 8PSK and $d = 50$ and two different Doppler spreads. The antenna is selected using the criterion proposed in Sec. III-C. Also plotted is the PER of the $(N+1)$-pilot scheme when α^*_∞ and β^*_∞ are used to allocate energy across pilots and data. We see that with these parameters, the $(N+1)$-pilot scheme is markedly more energy-efficient than the N-pilot scheme, and is just 2-3 dB away from the PER with optimal parameters obtained by a brute-force simulations-based search. As before, the energy savings relative to the N-pilot scheme increase as the Doppler spread increases.

Figure 9 makes the same comparison when a practical coded system, in which the rate 1/3 mother convolutional code specified in the 3GPP wideband code division multiple access (WCDMA) standard [4] is used to generate the d data symbols. Its generator polynomial is $[133\ 171\ 165]$ in octal notation. Hard decision Viterbi decoding is used at the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

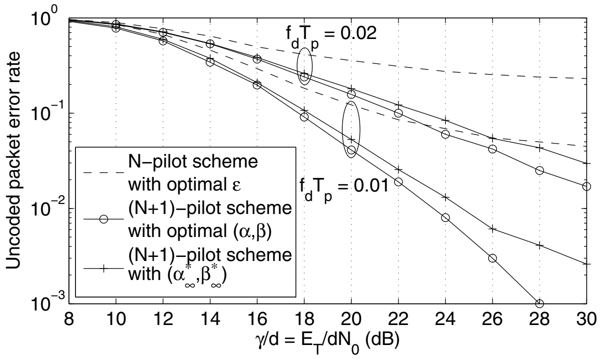


Fig. 8. Uncoded PER comparison of N-pilot and $(N+1)$-pilot AS training schemes for different Doppler spreads (8PSK, $N=4$, $T_p=10T_s$, and $d=50$).

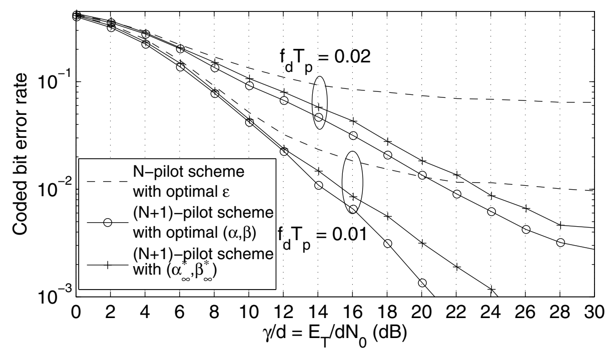


Fig. 9. Performance of N-pilot and $(N+1)$-pilot AS training schemes for coded packet reception (8PSK, $N=4$, $T_p=10T_s$, and $d=100$).

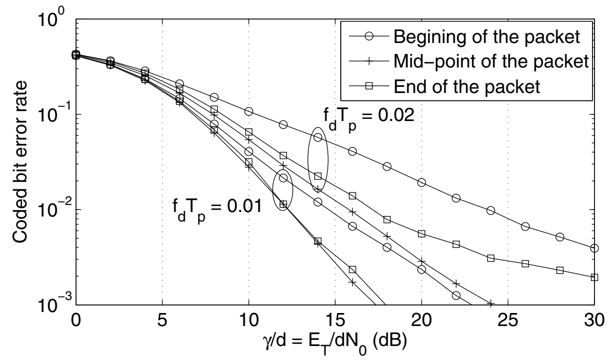


Fig. 10. Effect of the extra pilot location on the coded BER of the $(N+1)$-pilot scheme (8PSK, $N=4$, $T_p=10T_s$, $d=100$, and rate 1/3 mother convolutional code).

receiver. Each coded packet is 300 bits long and is 8PSK modulated. As before, the criterion in Sec. III-C is used to select the antenna. The bit error rate (BER) is plotted in order to enable a direct comparison with the several results presented thus far, which plotted the SEP. As in the uncoded packet case, the $(N+1)$-pilot scheme with α_∞^* and β_∞^* is much more energy-efficient than the N-pilot scheme. It reduces the error floor by an order of magnitude, and is just 2-3 dB away from the BER with optimal parameters obtained by a brute-force simulations-based search. At $f_dT_p=0.01$ and BER of 0.001, the performance loss for $l=5$ and $l=10$ is less than 0.5 dB, when the SEP-optimal antenna is computed only every l^{th} data symbol.

Effect of Location of the $(N+1)^{th}$ pilot: Figure 10 plots the BER of a coded packet for different locations of the extra pilot in the packet. Effectively, the extra pilot enables the time-varying channel gain of the selected antenna to be sampled twice. Since all the symbols of the packet contribute to the BER, the BER is the lowest when the extra pilot is in the middle of the packet. Also, the BER is lower when the extra pilot is at the end of the packet rather than at the start of the packet. This is because the former helps interpolate the fading process unlike the later, which involves an extrapolation. The observations are a little different for the SEP of individual uncoded data symbols. Here, the closer the extra pilot is to the data symbol under consideration, the better is the channel estimate used for decoding, and the smaller is the SEP. The corresponding figure is not shown due to space constraints.

V. Conclusions

We analyzed and compared the conventional N-pilot scheme and the $(N+1)$-pilot scheme for the practically-motivated scenario of time-varying channels and imperfect and differently outdated channel estimates. Unlike the N-pilot scheme, for which the optimal selection rule is a linear weighting rule, the optimal AS rule for the $(N+1)$-pilot scheme is affine, with the scaling weight and offset for an antenna being a function of the outdatedness of its channel gain estimate. We also considered coded and uncoded packet reception, in which the receiver is practically constrained to receive the entire packet using the same antenna, and proposed a novel AS rule for it. We saw that the $(N+1)$-pilot scheme, despite its longer training duration, is much more energy efficient than the N-pilot scheme. These savings are unique to AS, and arise because of the ability of the $(N+1)$-pilot scheme to generate a fresh and accurate estimate of the channel of the selected antenna, without wasting energy on generating equally accurate channel estimates of the unselected antennas.

Appendix

A. Proof of Theorem 1

Substituting $\sigma_{k,i}^2=\frac{\varepsilon\rho(T_k,t_i)^2}{\varepsilon+(N\varepsilon+d)\gamma^{-1}}$ in (13) and simplifying further yields the following SEP expression:

$$P_i^{\varepsilon}(\gamma)=\frac{1}{\pi}\sum_{k=1}^{N}\sum_{l=0}^{N-1}\sum_{m=1}^{\left|\mathcal{S}_l^k\right|}\int_0^{\frac{M-1}{M}\pi}(-1)^l\left(1+\frac{\sin^2\left(\frac{\pi}{M}\right)\csc^2\theta}{-1+\eta\rho\left(T_k,t_i\right)^{-2}}\right.$$
$$\left.+\sum_{n\in\mathcal{S}_k^l(m)}\frac{-1+\eta\rho\left(T_n,t_i\right)^{-2}}{-1+\eta\rho\left(T_k,t_i\right)^{-2}}\right)^{-1}d\theta, \quad (25)$$

where csc is cosecant and

$$\eta=\frac{1}{\varepsilon}\left(1+\left(N\varepsilon+d\right)\gamma^{-1}\right)\left(\varepsilon+\left(N\varepsilon+d\right)\gamma^{-1}\right). \quad (26)$$

Thus, the SEP depends on ε only through η.

In order to find the optimal ε, we differentiate the SEP expression above with respect to ε and equate it to 0. Since $\frac{dP_i^{\varepsilon}(\gamma)}{d\varepsilon}=\frac{dP_i^{\varepsilon}(\gamma)}{d\eta}\frac{d\eta}{d\varepsilon}$, we have $\frac{dP_i^{\varepsilon}(\gamma)}{d\eta}=0$ or $\frac{d\eta}{d\varepsilon}=0$. It can be easily verified from (26) that $\frac{d\eta}{d\varepsilon}=0$ occurs only at

$\varepsilon^*(\gamma) = \sqrt{\frac{d(\gamma+d)}{N(\gamma+N)}}$. Also, notice that $\left.\frac{d^2\eta}{d\varepsilon^2}\right|_{\varepsilon=\varepsilon^*(\gamma)} > 0$ and $\left.\frac{d^2 P_i^{\varepsilon}(\gamma)}{d\varepsilon^2}\right|_{\varepsilon=\varepsilon^*(\gamma)} = \frac{dP_i^{\varepsilon}(\gamma)}{d\eta}\left.\frac{d^2\eta}{d\varepsilon^2}\right|_{\varepsilon=\varepsilon^*(\gamma)}$.

We now argue that $\frac{dP_i^{\varepsilon}(\gamma)}{d\eta}$ is strictly positive. In (25), we see that η always appears in the form of a product with the reciprocal of the channel correlation coefficient. Hence, increasing η is equivalent to decreasing the channel correlation coefficient, which increases the SEP. Thus, $\frac{dP_i^{\varepsilon}(\gamma)}{d\eta} > 0$. Hence, $\frac{dP_i^{\varepsilon}(\gamma)}{d\varepsilon} = 0$ has a unique minimum at $\varepsilon^*(\gamma)$.

B. Proof of Lemma 1

The MMSE estimate of $X \triangleq h_k(t_i)$, obtained from the observations $\mathbf{Y} \triangleq [r_k(T_k)\ r_k(T_{N+1})]^T$, is given by [24]

$$\widehat{\widehat{h}}_k(t_i) = \mathbf{E}\left[X \,\middle|\, \mathbf{Y}\right] = \Sigma_{X\mathbf{Y}}\Sigma_{\mathbf{YY}}^{-1}(\mathbf{Y} - \mathbf{E}\,[\mathbf{Y}]). \qquad (27)$$

Since $\widehat{\widehat{h}}_k(t_i)$ is linear combination of elements of circularly symmetric complex Gaussian random vector $\mathbf{Y}$, it is also a circularly symmetric complex Gaussian RV. Its variance is

$$\sigma_{k,i}'^2 = \mathbf{var}\left[X \,\middle|\, \mathbf{Y}\right] = \mathbf{var}\,[X] - \Sigma_{X\mathbf{Y}}\Sigma_{\mathbf{YY}}^{-1}\Sigma_{X\mathbf{Y}}^H. \qquad (28)$$

From (6), (9), and the definition of ρ, we get $\mathbf{var}\,[X] = 1$, $\mathbf{E}\,[\mathbf{Y}] = [0\ 0]^T$,

$$\Sigma_{X\mathbf{Y}} = \left[\sqrt{\alpha E_2}p^*\rho\,(T_k, t_i)\ \ \sqrt{\beta E_2}p^*\rho\,(T_{N+1}, t_i)\right], \qquad (29)$$

$$\Sigma_{\mathbf{YY}} = \begin{bmatrix} \alpha E_2 + N_0 & \sqrt{\alpha\beta}E_2\rho\,(T_k, T_{N+1}) \\ \sqrt{\alpha\beta}E_2\rho\,(T_k, T_{N+1}) & \beta E_2 + N_0 \end{bmatrix}. \qquad (30)$$

Substituting the above results in (27) and (28), and simplifying further yields (15) and (16).

C. Derivation of (19)

From (17), the following relations follow easily:

$$\mathbf{E}\left[\mathcal{D} \,\middle|\, \widehat{[1]} = k, \widehat{\widehat{h}}_k(t_i), s\right] = \widehat{\widehat{h}}_k^*(t_i)s\mathbf{E}\left[h_k(t_i) \,\middle|\, \widehat{\widehat{h}}_k(t_i)\right], \qquad (31)$$

$$\mathbf{var}\left[\mathcal{D} \,\middle|\, \widehat{[1]} = k, \widehat{\widehat{h}}_k(t_i), s\right] = \left|\widehat{\widehat{h}}_k(t_i)\right|^2\left(\mathbf{var}\left[h_k(t_i) \,\middle|\, \widehat{\widehat{h}}_k(t_i)\right] E_2 + N_0\right). \qquad (32)$$

From (15), it is clear that $h_k(t_i)$ conditioned on $\widehat{\widehat{h}}_k(t_i)$ is a complex Gaussian RV. Hence, using (27) and (28), with $X \triangleq h_k(t_i)$ and $Y \triangleq \widehat{\widehat{h}}_k(t_i)$, the conditional mean and variance of $h_k(t_i)$ are given by

$$\mathbf{E}\left[h_k(t_i) \,\middle|\, \widehat{\widehat{h}}_k(t_i)\right] = \frac{\mathbf{E}\left[h_k(t_i)\widehat{\widehat{h}}_k^*(t_i)\right]}{\mathbf{var}\left[\widehat{\widehat{h}}_k(t_i)\right]}\widehat{\widehat{h}}_k(t_i) = \widehat{\widehat{h}}_k(t_i),$$

$$\mathbf{var}\left[h_k(t_i) \,\middle|\, \widehat{\widehat{h}}_k(t_i)\right] = \mathbf{var}\,[h_k(t_i)] - \frac{\left|\mathbf{E}\left[h_k(t_i)\widehat{\widehat{h}}_k^*(t_i)\right]\right|^2}{\mathbf{var}\left[\widehat{\widehat{h}}_k(t_i)\right]},$$
$$= 1 - \sigma_{k,i}'^2. \qquad (33)$$

In the above derivation, we used $\mathbf{E}\left[h_k(t_i)\widehat{\widehat{h}}_k^*(t_i)\right] = \mathbf{var}\left[\widehat{\widehat{h}}_k(t_i)\right] = \sigma_{k,i}'^2$, which follows from (15).

Now substituting these results on conditional mean and variance in (31) and (32), yields (18) and (19).

D. Proof of Lemma 2

Unconditioning the SEP expression in (20) over $\widehat{\widehat{h}}_k(t_i)$ given $\widehat{h}_k(t_i)$, we get

$$P_i^{\alpha,\beta}\left(\gamma \,\middle|\, \widehat{[1]} = k, \widehat{h}_k(t_i)\right) = \frac{1}{\pi} \times \int_0^{\frac{M-1}{M}\pi} \mathcal{M}_{\left|\widehat{\widehat{h}}_k(t_i)\right|^2 \,\middle|\, \widehat{h}_k(t_i)}\left(\frac{-\sin^2\left(\frac{\pi}{M}\right)\csc^2\theta}{1-\sigma_{k,i}'^2 + (N\alpha+\beta+d)\,\gamma^{-1}}\right)d\theta, \qquad (34)$$

where $\mathcal{M}_{\left|\widehat{\widehat{h}}_k(t_i)\right|^2 \,\middle|\, \widehat{h}_k(t_i)}(.)$ is the moment generating function (MGF) of $\left|\widehat{\widehat{h}}_k(t_i)\right|^2$ conditioned on $\widehat{h}_k(t_i)$. Recall that k is assumed to be the index of the selected antenna.

It is easy to see that $\widehat{\widehat{h}}_k(t_i)$ conditioned on $\widehat{h}_k(t_i)$ is a complex Gaussian RV. Again using (27) and (28), the conditional moments of $\widehat{\widehat{h}}_k(t_i)$ are given by

$$\mathbf{E}\left[\widehat{\widehat{h}}_k(t_i) \,\middle|\, \widehat{h}_k(t_i)\right] = \frac{\mathbf{E}\left[\widehat{\widehat{h}}_k(t_i)\widehat{h}_k^*(t_i)\right]}{\sigma_{k,i}^2}\widehat{h}_k(t_i) = \widehat{h}_k(t_i),$$

$$\mathbf{var}\left[\widehat{\widehat{h}}_k(t_i) \,\middle|\, \widehat{h}_k(t_i)\right] = \sigma_{k,i}'^2 - \frac{\left|\mathbf{E}\left[\widehat{\widehat{h}}_k(t_i)\widehat{h}_k^*(t_i)\right]\right|^2}{\sigma_{k,i}^2},$$
$$= \sigma_{k,i}'^2 - \sigma_{k,i}^2. \qquad (35)$$

The above results follow because $\mathbf{E}\left[\widehat{\widehat{h}}_k(t_i)\widehat{h}_k^*(t_i)\right] = \sigma_{k,i}^2$, which, in turn, follows from (7) and (15). Hence, $\left|\widehat{\widehat{h}}_k(t_i)\right|^2$ conditioned on $\widehat{h}_k(t_i)$ is a non-central Chi-square distributed RV, whose MGF is given by [25]

$$\mathcal{M}_{\left|\widehat{\widehat{h}}_k(t_i)\right|^2 \,\middle|\, \widehat{h}_k(t_i)}(x) = \frac{1}{1 - x\left(\sigma_{k,i}'^2 - \sigma_{k,i}^2\right)}\exp\left(\frac{\left|\widehat{h}_k(t_i)\right|^2 x}{1 - x\left(\sigma_{k,i}'^2 - \sigma_{k,i}^2\right)}\right).$$

Substituting this in (34) and simplifying further yields the desired result in (21).

E. Proof of Theorem 2

The SEP-optimal antenna is clearly the one that minimizes the upper bound for the conditional SEP

$P_i^{\alpha,\beta}\left(\gamma \,\middle|\, \widehat{[1]} = k, \widehat{h}_k(t_i)\right)$ in (22). Hence,

$$\widehat{[1]} = \arg\min_{k=1,\ldots,N} \left\{ \frac{1}{1+c_k} \exp\left(\frac{-\left|\widehat{h}_k(t_i)\right|^2 \sin^2\left(\frac{\pi}{M}\right)/(1+c_k)}{1-\sigma_{k,i}'^2 + (N\alpha+\beta+d)\gamma^{-1}} \right) \right\},$$
$$= \arg\min_{k=1,\ldots,N} \left\{ \ln(1+c_k)^{-1} - \frac{\left|\widehat{h}_k(t_i)\right|^2 \sin^2\left(\frac{\pi}{M}\right)/(1+c_k)}{1-\sigma_{k,i}'^2 + (N\alpha+\beta+d)\gamma^{-1}} \right\},$$

where the last step follows because logarithm is a monotonically increasing function and, thus, does not affect the ordering. Simplifying further yields (23).

F. *Proof of Theorem 3*

Using the law of total probability, the SEP of the $(N+1)$-pilot scheme is given by

$$P_i^{\alpha,\beta}(\gamma) = \sum_{k=1}^{N} \mathbf{E}\left[P_i^{\alpha,\beta}\left(\gamma \,\middle|\, \widehat{[1]} = k, \widehat{h}_k(t_i)\right) \Pr\left(\widehat{[1]} = k \middle| \widehat{h}_k(t_i)\right) \right], \tag{36}$$

where we have used the observation that if antenna k is selected, then the SEP depends only on its estimate $\widehat{h}_k(t_i)$ and not the estimates of the other antennas. For each term in the summation, the expectation is taken over the channel estimate of the selected antenna.

The expression for the first term $P_i^{\alpha,\beta}\left(\gamma \,\middle|\, \widehat{[1]} = k, \widehat{h}_k(t_i)\right)$ inside the expectation in (36) is already derived in (21). The second term inside the expectation is evaluated as follows. From the optimal selection rule in (23), it follows that

$$\Pr\left(\widehat{[1]} = k \middle| \widehat{h}_1(t_i), \ldots, \widehat{h}_N(t_i)\right) = \prod_{l=1, l\neq k}^{N} I_{\left\{a_k + b_k \left|\widehat{h}_k(t_i)\right|^2 > a_l + b_l \left|\widehat{h}_l(t_i)\right|^2\right\}}. \tag{37}$$

Hence,

$$\Pr\left(\widehat{[1]} = k \middle| \widehat{h}_k(t_i)\right) = \prod_{l=1, l\neq k}^{N} \Pr\left(a_k + b_k \left|\widehat{h}_k(t_i)\right|^2 > a_l + b_l \left|\widehat{h}_l(t_i)\right|^2 \,\middle|\, \left|\widehat{h}_k(t_i)\right|^2 \right).$$

Since $\left|\widehat{h}_l(t_i)\right|^2$ is an exponential RV with mean $\sigma_{l,i}^2$, the l^{th} term in the product above is given by

$$\Pr\left(a_l + b_l \left|\widehat{h}_l(t_i)\right|^2 < a_k + b_k \left|\widehat{h}_k(t_i)\right|^2 \,\middle|\, \left|\widehat{h}_k(t_i)\right|^2 \right) = \left[1 - \exp\left(-\frac{a_k - a_l}{b_l \sigma_{l,i}^2} - \frac{b_k \left|\widehat{h}_k(t_i)\right|^2}{b_l \sigma_{l,i}^2} \right) \right] I_{\left\{\left|\widehat{h}_k(t_i)\right|^2 \geq \frac{a_l - a_k}{b_k}\right\}}. \tag{38}$$

Therefore,

$$\Pr\left(\widehat{[1]} = k \,\middle|\, \widehat{h}_k(t_i)\right) = \left(\prod_{l=1, l\neq k}^{N} I_{\left\{\left|\widehat{h}_k(t_i)\right|^2 \geq \frac{a_l - a_k}{b_k}\right\}} \right) \times \left(\prod_{l=1, l\neq k}^{N} \left[1 - \exp\left(-\frac{a_k - a_l}{b_l \sigma_{l,i}^2} - \frac{b_k \left|\widehat{h}_k(t_i)\right|^2}{b_l \sigma_{l,i}^2} \right) \right] \right).$$

Now, using the identity $\prod_{l=1, l\neq k}^{N} (1 - \exp(-x_l)) = \sum_{l=0}^{N-1} (-1)^l \sum_{m=1}^{\left|\mathcal{S}_l^k\right|} \exp\left(-\sum_{z\in\mathcal{S}_l^k(m)} x_z \right)$, the above equation simplifies to

$$\Pr\left(\widehat{[1]} = k \,\middle|\, \widehat{h}_k(t_i)\right) = I_{\left\{\left|\widehat{h}_k(t_i)\right|^2 \geq g_k\right\}} \sum_{l=0}^{N-1} (-1)^l \times \sum_{m=1}^{\left|\mathcal{S}_l^k\right|} \exp\left(-\sum_{z\in\mathcal{S}_l^k(m)} \left(\frac{a_k - a_z}{b_z \sigma_{z,i}^2} + \frac{b_k}{b_z \sigma_{z,i}^2} \left|\widehat{h}_k(t_i)\right|^2 \right) \right), \tag{39}$$

where $g_k \triangleq \max_{l=1,\ldots,N, l\neq k} \frac{a_l - a_k}{b_k}$.

Substituting (21) and (39) in the SEP expression in (36) and simplifying, we get

$$P_i^{\alpha,\beta}(\gamma) = \frac{1}{\pi} \sum_{k=1}^{N} \sum_{l=0}^{N-1} (-1)^l I_{\left\{\left|\widehat{h}_k(t_i)\right|^2 \geq g_k\right\}} \times \sum_{m=1}^{\left|\mathcal{S}_l^k\right|} \mathbf{E}\left[\exp\left(-\sum_{z\in\mathcal{S}_l^k(m)} \left(\frac{a_k - a_z}{b_z \sigma_{z,i}^2} + \frac{b_k \left|\widehat{h}_k(t_i)\right|^2}{b_z \sigma_{z,i}^2} \right) \right) \times \int_0^{\frac{M-1}{M}\pi} \left(1 + \frac{c_k}{\sin^2\theta} \right)^{-1} \exp\left(\frac{-\left|\widehat{h}_k(t_i)\right|^2 \frac{\sin^2\left(\frac{\pi}{M}\right)}{(c_k + \sin^2\theta)}}{1 - \sigma_{k,i}'^2 + (N\alpha+\beta+d)\gamma^{-1}} \right) d\theta \right].$$

Averaging over $\left|\widehat{h}_k(t_i)\right|^2$, which is an exponential RV, and consolidating the various exponential terms, we get

$$P_i^{\alpha,\beta}(\gamma) = \frac{1}{\pi} \sum_{k=1}^{N} \sum_{l=0}^{N-1} \frac{(-1)^l}{\sigma_{k,i}^2} \sum_{m=1}^{\left|\mathcal{S}_l^k\right|} \exp\left(-\sum_{z\in\mathcal{S}_l^k(m)} \frac{a_k - a_z}{b_z \sigma_{z,i}^2} \right) \times \int_0^{\frac{M-1}{M}\pi} \left(1 + \frac{c_k}{\sin^2\theta} \right)^{-1} \int_{g_k}^{\infty} \exp\left(-f\left(\mathcal{S}_l^k(m)\right) x \right) dx\, d\theta,$$

where f is defined in the theorem statement. Evaluating the integral over x and simplifying yields (24).

References

[1] V. Kristem, N. B. Mehta, and A. F. Molisch, "Energy-efficient training for antenna selection in time-varying channels," in *Proc. 2011 Asilomar Conf. on Signals, Syst., and Comput.*, pp. 13–17.

[2] A. F. Molisch and M. Z. Win, "MIMO systems with antenna selection," *IEEE Microwave Mag.*, vol. 5, pp. 46–56, Mar. 2004.

[3] IEEE, "IEEE standard for information technology – local and metropolitan area networks – specific requirements – part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 5: Enhancements for higher throughput," Tech. Rep. 802.11n-2009, IEEE Computer Society, Oct. 2009.

[4] 3GPP, "Technical specification group radio access network; evolved universal terrestrial radio access (E-UTRA); physical layer procedures (release 8)," TS 36.213 (v8.2.0), 3rd Generation Partnership Project (3GPP), 2008.

[5] A. Ghrayeb and T. M. Duman, "Performance analysis of MIMO systems with antenna selection over quasi-static fading channels," *IEEE Trans. Veh. Technol.*, vol. 52, pp. 281–288, Mar. 2003.

[6] M. Z. Win and J. H. Winters, "Virtual branch analysis of symbol error probability for hybrid selection/maximal-ratio combining in Rayleigh fading," *IEEE Trans. Commun.*, vol. 49, pp. 1926–1934, Nov. 2001.

[7] H. Zhang, A. Molisch, and J. Zhang, "Applying antenna selection in WLANs for achieving broadband multimedia communications," *IEEE Trans. Broadcast.*, vol. 52, pp. 475–482, Dec. 2006.

[8] V. Kristem, N. B. Mehta, and A. F. Molisch, "Optimal receive antenna selection in time-varying fading channels with practical training constraints," *IEEE Trans. Commun.*, vol. 58, pp. 2023–2034, Jul. 2010.

[9] V. Kristem, N. B. Mehta, and A. F. Molisch, "A novel, balanced, and energy-efficient training method for receive antenna selection," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 2742–2753, Sep. 2010.

[10] M. Chiani, A. Conti, and C. Fontana, "Improved performance in TD-CDMA mobile radio system by optimizing energy partition in channel estimation," *IEEE Trans. Commun.*, vol. 51, pp. 352–355, Mar. 2003.

[11] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, pp. 951–963, 2003.

[12] Y. Peng, S. Cui, and R. You, "Optimal pilot-to-data power ratio for diversity combining with imperfect channel estimation," *IEEE Commun. Lett.*, vol. 10, pp. 97–99, Feb. 2006.

[13] A. Maaref and S. Aissa, "Optimized rate-adaptive PSAM for MIMO MRC systems with transmit and receive CSI imperfections," *IEEE Trans. Commun.*, vol. 57, pp. 821–830, Mar. 2009.

[14] K. Yadav, N. Singh, and M. C. Srivatsava, "Analysis of adaptive pilot symbol assisted modulation with optimized pilot spacing for Rayleigh fading channel," in *Proc. 2008 ICSCN*, pp. 405–410.

[15] W. M. Gifford, M. Z. Win, and M. Chiani, "Antenna subset diversity with non-ideal channel estimation," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 1527–1539, May 2008.

[16] A. B. Narasimhamurthy and C. Tepedelenlioglu, "Antenna selection for MIMO OFDM systems with channel estimation error," in *Proc. 2007 Globecom*, pp. 3290–3294.

[17] H. A. Abou-Saleh, A. F. Molisch, T. Zemen, S. D. Blostein, and N. B. Mehta, "Receive antenna selection for time-varying channels using discrete prolate spheroidal sequences," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 2616–2627, Jul. 2012.

[18] N. B. Mehta, A. F. Molisch, and S. Kashyap, "Antenna selection in LTE: from motivation to specification," *IEEE Commun. Mag.*, vol. 50, pp. 144–150, Oct. 2012.

[19] J. D. Parsons, *The Mobile Radio Propagation Channel*. John Wiley & Sons, 2001.

[20] M.-S. Alouini and A. J. Goldsmith, "A unified approach for calculating error rates of linearly modulated signals over generalized fading channels," *IEEE Trans. Commun.*, vol. 47, pp. 1324–1334, Sept. 1999.

[21] C. Tepedelenlioglu, A. Abdi, G. B. Giannakis, and M. Kaveh, "Estimation of Doppler spread and signal strength in mobile communications with applications to handoff and adaptive transmission," *Wireless Commun. and Mobile Computing*, vol. 1, pp. 221–242, 2001.

[22] A. Dogandzic and B. Zhang, "Estimating Jakes' Doppler power spectrum parameters using the Whittle approximation," *IEEE Trans. Signal Process.*, vol. 53, pp. 987–1005, Mar. 2005.

[23] Y. Li and Y. Guan, "Modified Jakes model for simulating multiple uncorrelated fading waveforms," in *Proc. 2000 ICC*, pp. 46–49.

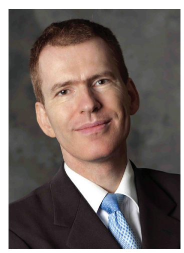
[24] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory, Vol. 1*, 1st edition. Prentice Hall Signal Processing Series, 1993.

[25] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th edition. Dover, 1972.

Vinod Kristem received his Bachelor of Technology degree in Electronics and Communications Eng. from the National Institute of Technology (NIT), Warangal, India in 2007. He received his Master of Engineering degree in Telecommunications from the Dept. of Electrical Communication Eng., Indian Institute of Science, Bangalore, India in 2009. From 2009–2011, he was with Beceem Communications Pvt. Ltd., Bangalore, India (which was recently acquired by Broadcom Corp.), where he worked on channel estimation and physical layer measurements for WiMAX and LTE. He is currently working toward his Ph.D. degree with the Dept. of Electrical Eng., University of Southern California, Los Angeles, USA. His research interests include the design and analysis of receiver algorithms for wireless communication networks, antenna selection in MIMO systems, and cooperative localization in wireless networks.

Neelesh B. Mehta (S'98-M'01-SM'06) received his Bachelor of Technology degree in Electronics and Communications Eng. from the Indian Institute of Technology (IIT), Madras in 1996, and his M.S. and Ph.D. degrees in Electrical Engineering from the California Institute of Technology, Pasadena, CA, USA in 1997 and 2001, respectively. He is now an Associate Professor in the Dept. of Electrical Communication Eng., Indian Institute of Science (IISc), Bangalore, India. Prior to joining IISc, he was a research scientist in USA from 2001 to 2007 in AT&T Research Laboratories, NJ, Broadcom Corp., NJ, and Mitsubishi Electric Research Laboratories (MERL), MA.

His research includes work on link adaptation, multiple access protocols, cellular system design, MIMO, cooperative communications, energy harvesting communications, and cognitive radio. He was also actively involved in the 3GPP RAN1 from 2003 to 2007. He has served as TPC co-chair of tracks in Chinacom 2008, VTC 2009 (Fall), ICC 2013, and Globecom 2013. He has co-authored 40 IEEE transactions papers, 65+ conference papers, and three book chapters, and is a co-inventor in 20 issued US patents. He is an Editor of IEEE WIRELESS COMMUNICATIONS LETTERS and serves as the Director of Conference Publications on the Board of Governors of IEEE ComSoc.



Andreas F. Molisch (S'89, M'95, SM'00, F'05) received the Dipl. Ing., Ph.D., and habilitation degrees from the Technical University of Vienna, Vienna, Austria, in 1990, 1994, and 1999, respectively. He subsequently was with AT&T (Bell) Laboratories Research (USA); Lund University, Lund, Sweden, and Mitsubishi Electric Research Labs (USA). He is now a Professor of electrical engineering with the University of Southern California, Los Angeles.

His current research interests are the measurement and modeling of mobile radio channels, ultra-wideband communications and localization, cooperative communications, multiple-input-multiple-output systems, wireless systems for healthcare, and novel cellular architectures. He has authored, coauthored, or edited four books (among them the textbook *Wireless Communications*, Wiley-IEEE Press), 16 book chapters, some 150 journal papers, and numerous conference contributions, as well as more than 70 patents and 60 standards contributions.

Dr. Molisch has been an Editor of a number of journals and special issues, General Chair, Technical Program Committee Chair, or Symposium Chair of multiple international conferences, as well as Chairman of various international standardization groups. He is a Fellow of the IEEE, Fellow of the AAAS, Fellow of the IET, an IEEE Distinguished Lecturer, and a member of the Austrian Academy of Sciences. He has received numerous awards, most recently the Donald Fink Prize of the IEEE, and the Eric Sumner Award of the IEEE.