Transactions Papers

A Novel, Balanced, and Energy-Efficient Training Method for Receive Antenna Selection

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

***Abstract*—In receive antenna selection (AS), only signals from a subset of the antennas are processed at any time by the limited number of radio frequency (RF) chains available at the receiver. Hence, the transmitter needs to send pilots multiple times to enable the receiver to estimate the channel state of all the antennas and select the best subset. Conventionally, the sensitivity of coherent reception to channel estimation errors has been tackled by boosting the energy allocated to all pilots to ensure accurate channel estimates for all antennas. Energy for pilots received by unselected antennas is mostly wasted, especially since the selection process is robust to estimation errors. In this paper, we propose a novel training method uniquely tailored for AS that transmits one extra pilot symbol that generates accurate channel estimates for the antenna subset that actually receives data. Consequently, the transmitter can selectively boost the energy allocated to the extra pilot. We derive closed-form expressions for the proposed scheme's symbol error probability for MPSK and MQAM, and optimize the energy allocated to pilot and data symbols. Through an insightful asymptotic analysis, we show that the optimal solution achieves full diversity and is better than the conventional method.**

***Index Terms*—Training, antenna selection, diversity methods, fading channels, estimation, error analysis, quadrature amplitude modulation, quadrature phase shift keying, energy efficient.**

I. Introduction

RECEIVE antenna selection (AS) provides a low hardware complexity solution for exploiting the spatial diversity benefits of receiving with multiple antennas [1]–[3]. In AS, the receiver does not receive and process signals from all its antennas. Instead, it dynamically selects a subset of the antennas with the 'best' instantaneous channel conditions to the transmitter, and processes signals through them. This enables the receiver to employ fewer of the expensive radio frequency (RF) chains. Consequently, AS has been adopted in next generation wireless systems such as IEEE 802.11n [4]. Despite its lower hardware complexity, AS achieves the full diversity order with perfect channel state information (CSI) [5], [6].

In practice, the CSI needs to be acquired using a pilot-based training scheme [7]. The basic operation of AS imposes the constraint that only L antennas can be estimated at any time with L RF chains. Therefore, in receive AS, the transmitter needs to transmit pilots multiple times so that the receiver can estimate the channels of all the available antennas and choose the antennas with the best channels. As an example, consider an AS system with $N = 6$ antennas and $L = 2$ RF chains. To estimate the channels of all 6 antennas, at least 3 pilot transmissions are needed. The first pilot helps estimate channel gains of antennas #1 and #2, the second pilot helps estimate channel gains of antennas #3 and #4, and the third pilot helps estimate channel gains of antennas #5 and #6. In general, with N antennas and L RF chains, at least $\lceil N/L \rceil$ pilot transmissions are required, where $\lceil . \rceil$ denotes the ceil function.

In AS, estimation errors may cause a suboptimal antenna subset to get selected and will also impair coherent demodulation. While selection is quite robust to such errors, as has been observed empirically in [8], coherent demodulation is not. This forces the transmitter to increase/boost, if possible, the energy allocated to the pilot symbols. Since the transmitter does not know a priori which antennas will be selected, it needs to uniformly boost the energy of all the pilots used to estimate the channel gains of all the antennas. This process is energy-inefficient because it obtains highly accurate channel estimates for unselected antennas, as well. Under a total energy constraint, it also draws energy away from data symbols, and, thus, increases their symbol error probability (SEP).

In this paper, we propose a novel training method for AS that significantly improves the energy-efficiency of AS. *In the proposed method, the transmitter sends an extra pilot symbol after the first $\lceil N/L \rceil$ pilots, which we call selection pilots. While the selection pilots help in selecting the best antenna subset, the extra pilot helps in refining the channel*

Manuscript received May 20, 2009; revised April 15, 2010; accepted June 18, 2010. The associate editor coordinating the review of this paper and approving it for publication was H.-C. Yang.

V. Kristem is with Beceem Communications, Bangalore, India. He was with the Dept. of Electrical Communication Engineering (ECE), Indian Institute of Science (IISc), Bangalore, India, during the course of this work (e-mail: vinod.kristem@gmail.com).

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

A. F. Molisch is with the Dept. of Electrical Eng., University of Southern California, Los Angeles, CA, USA (e-mail: andreas.molisch@ieee.org).

A part of this paper has appeared in ICC 2010.

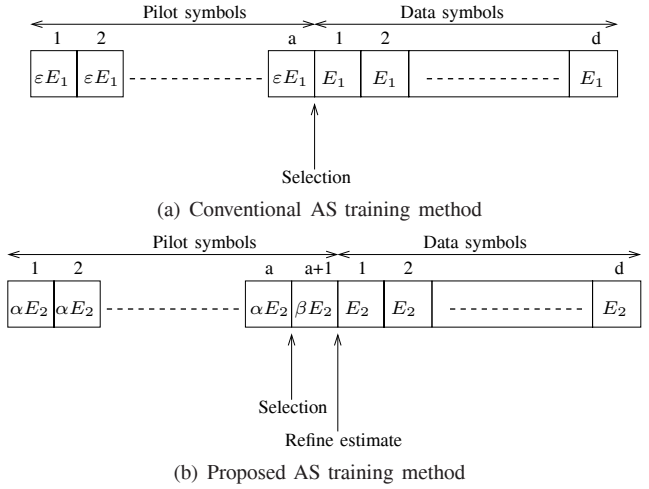
Digital Object Identifier 10.1109/TWC.2010.072610.090744

estimates of the selected antenna subset that will be used for data reception. A key point to note is that the impact of the extra pilot on the SEP of receive AS is vastly different compared to a non-AS system, which has as many RF chains as antennas. In the latter case, sending two pilots – one with energy E_s and the other with energy E_e – leads to the same estimation error as sending a single pilot with an energy of $E_s + E_e$. With AS there exists a unique trade-off between E_s and E_e since the first pilot can be used for selection and the second one for refining the estimates for coherent demodulation. This is because increasing E_s improves the probability that the optimal antenna subset is selected, while increasing E_e specifically reduces the error in the channel estimates of the selected antenna subset used for coherent demodulation. Intuitively, the robustness of AS to selection errors suggests that the transmitter can significantly reduce the total energy it allocates to the many selection pilots. Instead, it need only boost the energy of the extra pilot in order to get accurate estimates for demodulation. The energy thus saved can be transferred to the data symbols to reduce their SEP.

This trade-off has not been explored in the AS literature, to the best of our knowledge. When more pilot symbols than required (*i.e.*, $\lceil N/L \rceil$) are transmitted, it has been conventionally assumed that this affects the performance of AS only through the product of the number of pilot symbols per antenna and the energy per pilot symbol. This is the case, for example, in [9], which analyzed the impact of imperfect estimates on both subset selection and coherent demodulation. While [10]–[15] also considered the impact of channel estimation errors on AS, the way in which estimates are obtained was not modified. Similar results for Selection Diversity receivers in the presence of estimation errors were also derived in [16], [17]. While optimal power allocation for pilots and data has been considered for multiple antenna systems in [18]–[21] and for a time division code division multiple access (TD-CDMA) system in [22], it was not done for AS; consequently, the above trade-off did not occur.

The paper makes the following additional contributions in understanding and exploiting the above trade-off. Under a total energy constraint, it derives closed-form expressions for the fading-averaged SEP of MPSK and MQAM constellations of the proposed method as a function of the fractions of energy allocated to the various pilot and data symbols. *The analysis accounts for imperfect estimation in both the selection and data transmission phases*, and enables the fraction of energy allocated to pilots and data to be optimized. The analytical expressions clearly bring out the dependence of the SEP on the number of available and selected antennas, the number of data symbols, and the total energy available to the transmitter. To provide a fair comparison, we also optimize the energies allocated to pilots and data in the conventional AS training method of [9], and develop a new closed-form solution for its optimal energy allocation. The analysis, which is verified using extensive Monte Carlo simulations, shows that an energy gain as large as 3 dB over the conventional AS training method can be achieved.

In the asymptotic scenario where the total available energy is large, we show that the proposed AS training scheme achieves the full diversity order. The asymptotic analysis shows that the optimal energy allocation for the proposed method is unique and strictly better than that of the conventional optimized method. An approximate, but closed-form, expression for the optimal allocation is also derived.



(a) Conventional AS training method

(b) Proposed AS training method

Fig. 1. Training for receive antenna selection: conventional and proposed AS training methods.

The outline of this paper is as follows. The AS training and data transmission models are described in Sec. II and analyzed in Sec. III, and are followed by extensive simulation results in Sec. IV. We conclude in Sec. V. Several mathematical details are relegated to the Appendix.

II. System Model

Consider a system with one transmit antenna, N receive antennas, and L receive RF chains. Let h_k denote the frequency-flat, block-fading channel between the transmitter and the k^{th} receive antenna. It is modeled as a circularly symmetric complex Gaussian random variable (RV) with unit variance. Therefore, the channel gain $|h_k|$ is a Rayleigh RV. The channel gains for different receive antennas are assumed to be independent and identically distributed (i. i. d.), which is the case when the receive antennas are spaced sufficiently apart in a rich scattering environment [23].

We first describe the transmitter and receiver behavior for the conventional AS training method, and then describe how the proposed method differs from it.

A. Conventional AS Training Method

To enable the receiver to estimate the channel gains of all the N links, the transmitter sequentially transmits $a = \lceil \frac{N}{L} \rceil$ pilot symbols [9], [15] each with energy εE_1, sequentially, as shown in Figure 1(a), where E_1 is the energy allocated to a data symbol and $\varepsilon \geq 0$ is the energy scaling factor for pilots. Each pilot symbol is received by at most L receive antennas.

The signal received by the k^{th} receive antenna is given by

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

where p is the (complex) pilot with $|p| = 1$, n_k is a circular symmetric complex Gaussian RV with variance N_0 that is independent across different antennas, k, and time. The

minimum mean square estimate (MMSE) channel estimate for the k^{th} antenna is [24]

$$\hat{h}_k = \frac{\sqrt{\varepsilon E_1}p^*}{\varepsilon E_1 + N_0} r_k. \tag{2}$$

Since the channel estimates of different antenna elements are independent, the receiver selects the L antennas with the highest estimated channel gains. Let $\hat{\Omega}_L$ denote the selected subset of antennas. Note that $\hat{\Omega}_L$ depends on $\left\{\hat{h}_k\right\}_{k=1}^{N}$.

The pilot symbols are followed by d data symbols, each transmitted with energy E_1. All the d data symbols are received by the same antenna subset $\hat{\Omega}_L$. The received signal at the k^{th} antenna for the l^{th} data symbol, $s^{(l)}$, equals

$$y_k^{(l)} = h_k s^{(l)} + n_k^{(l)}, \quad k \in \hat{\Omega}_L. \tag{3}$$

The data symbols are equi-probable and derived from either the MPSK or MQAM constellations. For MPSK, $s^{(l)} \in \left\{\sqrt{E_1}\exp(\frac{j2\pi m}{M}), m = 0, \ldots, M-1\right\}$. For MQAM, $s^{(l)} = \sqrt{\frac{3E_1}{2(M-1)}}(a_I + ja_Q)$, where $a_I, a_Q \in \{2i - 1 - \sqrt{M}, i = 1, \ldots, \sqrt{M}\}$. To simplify the notation, we shall henceforth drop the symbol index l, unless required otherwise.

The maximum likelihood (ML) decision variable for decoding a data symbol is $\mathcal{D} = \sum_{k\in\hat{\Omega}_L} \hat{h}_k^* y_k$. The constraint on total energy, E_T, takes the form[1]

$$(a\varepsilon + d)E_1 = E_T. \tag{4}$$

Let $\gamma \triangleq \frac{E_T}{N_0}$. Therefore, $\frac{E_1}{N_0} = \frac{\gamma}{a\varepsilon+d}$. Notice that the choice of ε affects the energy allocated, E_1, to each data symbol.

B. Proposed Method

We now describe the proposed AS training method, which is shown in Figure 1(b). As before, the transmitter now first sequentially transmits $a = \lceil \frac{N}{L} \rceil$ pilot symbols so that all the N channels can be estimated, with L channels getting estimated every time a pilot is transmitted. But, each pilot symbol is now transmitted with energy αE_2, where E_2 is the energy per data symbol and α is now the energy scaling factor for these pilots.

The pilot symbol received by the k^{th} receive antenna is $r_k = \sqrt{\alpha E_2}ph_k + n_k$, where n_k, as before, is a circular symmetric complex Gaussian RV with variance N_0. As in (2), the MMSE channel estimate for the k^{th} antenna is given by

$$\hat{h}_k = \frac{\sqrt{\alpha E_2}p^*}{\alpha E_2 + N_0} r_k = a_1 h_k + e_k, \tag{5}$$

where p is the (complex) pilot with $|p| = 1$, $a_1 \triangleq \frac{\alpha E_2}{\alpha E_2 + N_0}$, and the zero-mean Gaussian noise term $e_k \triangleq \frac{n_k p^* \sqrt{\alpha E_2}}{\alpha E_2 + N_0}$ has a variance $\sigma_e^2 = \frac{\alpha E_2 N_0}{(\alpha E_2 + N_0)^2}$.

Since the channel estimates of different antennas are uncorrelated, the L antennas with the highest estimated channel gains are selected. As before, $\hat{\Omega}_L$ denotes the selected antenna subset, and depends on $\left\{\hat{h}_k\right\}_{k=1}^{N}$.

[1]The case where the total pilot and data energy is less than E_T is suboptimal, and is, therefore, not considered here.

Extra Pilot and Refined Estimates: The key difference in the proposed method is that an extra pilot symbol is transmitted with energy βE_2, and is received by the *selected* L *antennas*. The extra pilot helps in refining the channel estimates of the selected L antennas as explained below. Note that, in general, $\beta \neq \alpha$. The received signal, r'_k, for the extra pilot is

$$r'_k = \sqrt{\beta E_2}ph_k + n'_k, \quad k \in \hat{\Omega}_L, \tag{6}$$

where n'_k is a circular symmetric complex Gaussian RV with variance N_0, and is independent of n_k.

The channel estimate of a selected antenna $k \in \hat{\Omega}_L$ can be refined using the two observations r'_k and r_k. The refined MMSE estimate, $\hat{\hat{h}}_k$, that uses both r_k and r'_k equals [24]

$$\hat{\hat{h}}_k = \frac{\sqrt{\alpha E_2}p^* r_k + \sqrt{\beta E_2}p^* r'_k}{(\alpha+\beta)E_2 + N_0} = a_2 h_k + e'_k, \tag{7}$$

where $a_2 \triangleq \frac{(\alpha+\beta)E_2}{(\alpha+\beta)E_2+N_0}$ and the zero-mean Gaussian noise term $e'_k \triangleq \frac{\sqrt{\alpha E_2}p^* n_k + \sqrt{\beta E_2}p^* n'_k}{(\alpha+\beta)E_2+N_0}$ has a variance of $\sigma_{e'}^2 = \frac{(\alpha+\beta)E_2 N_0}{((\alpha+\beta)E_2+N_0)^2}$. Note that e'_k and e_k are correlated.

Data Reception: The pilot symbols are followed by d data symbols, each transmitted with energy E_2. They are all received by the antenna subset $\hat{\Omega}_L$.[2] The received signal for a data symbol s is

$$y_k = h_k s + n''_k, \quad k \in \hat{\Omega}_L. \tag{8}$$

The maximum likelihood decision variable, $\mathcal{D}$, for data decoding is $\mathcal{D} = \sum_{k\in\hat{\Omega}_L} \hat{\hat{h}}_k^* y_k$. The total energy constraint is

$$E_T = (a\alpha + \beta + d)\, E_2. \tag{9}$$

Let $\gamma \triangleq \frac{E_T}{N_0}$. Hence, $\frac{E_2}{N_0} = \frac{\gamma}{a\alpha+\beta+d}$. Now, α and β together affect the energy E_2 allocated to a data symbol. When $\beta = 0$, the SEP of the proposed method is the same as that of the conventional AS training method with $\varepsilon = \alpha$.

While a corresponding scheme can be developed for multiple transmit antennas, its analysis is beyond the scope of this paper. Even the one transmit antenna case considered in this paper will turn out to be analytically interesting and insightful.

III. SEP Analysis and Optimization

We now analyze the fading-averaged SEP for MPSK or MQAM for receive AS with imperfect CSI for the conventional and proposed AS training methods, and optimize their parameters to minimize the SEP. The case where channel coding is used across the d symbols is beyond the scope of this paper due to its analytical intractability. This approach has also been followed in [6], [9], [12], [17], which focus on the SEP to gain insights. While [25] did analyze AS with channel coding, channel estimation errors or training details for AS were not addressed. Henceforth, $\mathbf{E}[A]$ and $\mathbf{var}[A]$ shall denote the expectation and variance, respectively, of RV A. $\mathbf{E}[A|B]$ and $\mathbf{var}[A|B]$ will denote the conditional expectation and variance of A given B, respectively; x^* denotes the complex conjugate of scalar x; Y^H denotes the Hermitian transpose of vector Y.

[2]After receiving the extra pilot, the receiver can, in fact, reselect the antennas based on $\hat{h}_k$ or $\hat{\hat{h}}_k$, depending on whether k is in $\hat{\Omega}_L$ or not. We do not allow this to keep the analysis tractable. Simulations show that the performance gains from such a reselection are negligible.

A. Conventional Method Optimization

We discuss the conventional method only briefly given the analysis in [9]. The main contribution of this section lies in determining the optimal value of ε that minimizes the SEP subject to total energy constraint, and in deriving the expression for the optimal SEP; these were are not considered in [9]. This provides a fair benchmark for our new method, which is analyzed next.

In terms of the notation used in our paper, the SEP expressions for MPSK and MQAM, denoted by $P^{\varepsilon}_{\text{MPSK}}(\gamma)$ and $P^{\varepsilon}_{\text{MQAM}}(\gamma)$, respectively, are [9]:

$$P^{\varepsilon}_{\text{MPSK}}(\gamma) = \frac{1}{\pi}\int_0^{\frac{M-1}{M}\pi}\left(\frac{\sin^2\theta}{\sin^2\theta + c_{\text{PSK}}}\right)^L \times \prod_{n=L+1}^{N}\left(\frac{\sin^2\theta}{\sin^2\theta + \frac{c_{\text{PSK}}L}{n}}\right)d\theta, \quad (10)$$

$$P^{\varepsilon}_{\text{MQAM}}(\gamma) = \frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)\int_0^{\frac{\pi}{2}}\xi(\theta)\left(\frac{\sin^2\theta}{\sin^2\theta + c_{\text{QAM}}}\right)^L \times \prod_{n=L+1}^{N}\left(\frac{\sin^2\theta}{\sin^2\theta + \frac{c_{\text{QAM}}L}{n}}\right)d\theta, \quad (11)$$

where $\xi(\theta) = \frac{1}{\sqrt{M}}$, for $0 \le \theta < \frac{\pi}{4}$, and $\xi(\theta) = 1$, for $\frac{\pi}{4} \le \theta \le \frac{\pi}{2}$; $c_{\text{PSK}} \triangleq \frac{\varepsilon\gamma^2\sin^2\left(\frac{\pi}{M}\right)}{(a\varepsilon+d)((\varepsilon+1)\gamma+a\varepsilon+d)}$, and $c_{\text{QAM}} \triangleq \frac{1.5\varepsilon\gamma^2/(M-1)}{(a\varepsilon+d)((\varepsilon+1)\gamma+a\varepsilon+d)}$.

Lemma *1:* The optimal value of ε, denoted by $\varepsilon^*(\gamma)$, that minimizes the SEP of both MPSK and MQAM is

$$\varepsilon^*(\gamma) = \sqrt{\frac{d(\gamma+d)}{a(\gamma+a)}}. \quad (12)$$

Proof: The proof is given in Appendix A. ∎

When $\gamma \to \infty$, we have $\lim_{\gamma\to\infty}\varepsilon^*(\gamma) \triangleq \varepsilon^*_\infty = \sqrt{\frac{d}{a}}$. Note that this asymptotic result (but not the general expression for any γ in Lemma 1) can also be construed as a special case of the result in [22, (15)], which considered energy allocation between data symbols and mid-amble pilots in a TD-CDMA system to minimize mean square estimation error. Consequently, the asymptotic expression for the optimal SEP of MPSK, $P^{\varepsilon,\infty}_{\text{MPSK}}(\gamma)$, simplifies to

$$P^{\varepsilon,\infty}_{\text{MPSK}}(\gamma) = \gamma^{-N}\frac{(L+1)(L+2)\cdots(N)}{\pi 4^N L^{N-L}}\left(\sqrt{d}+\sqrt{a}\right)^{2N} \times \sin^{-2N}\left(\frac{\pi}{M}\right)\psi\left(\frac{M-1}{M}\pi, N\right), \quad (13)$$

where $\psi(T,N) \triangleq \binom{2N}{N}T + \sum_{j=0}^{N-1}\frac{(-1)^{j+N}}{N-j}\binom{2N}{j}\sin(2(N-j)T)$. The derivation is relegated to Appendix B. The asymptotic expression for the optimal SEP of MQAM can be written similarly.

B. Analysis and Optimization of SEP of Proposed Method

We now analyze the SEP of the proposed method given α and β, and then minimize it. The following result about the statistics of $\mathcal{D}$ shall be useful in deriving the SEP.

Lemma *2:* Conditioned on $\left\{\hat{h}_l, \hat{\hat{h}}_l\right\}_{l\in\hat{\Omega}_L}$ and s, the decision variable, $\mathcal{D}$, is a complex Gaussian RV with conditional mean, $\mu_{\mathcal{D}}$, and variance, $\sigma^2_{\mathcal{D}}$, given by

$$\mu_{\mathcal{D}} \triangleq \mathbf{E}\left[\mathcal{D}\,\middle|\,\left\{\hat{h}_l, \hat{\hat{h}}_l\right\}_{l\in\hat{\Omega}_L}, s\right] = s\sum_{k\in\hat{\Omega}_L}\left|\hat{\hat{h}}_k\right|^2, \quad (14)$$

$$\sigma^2_{\mathcal{D}} \triangleq \mathbf{var}\left[\mathcal{D}\middle|\left\{\hat{h}_l, \hat{\hat{h}}_l\right\}_{l\in\hat{\Omega}_L}, s\right] = ((1-a_2)E_2+N_0)\sum_{k\in\hat{\Omega}_L}\left|\hat{\hat{h}}_k\right|^2. \quad (15)$$

Proof: The proof is given in Appendix C. ∎

Since $\mu_{\mathcal{D}}$ and $\sigma^2_{\mathcal{D}}$ depend only on $\hat{\hat{h}}_k$, we see that all the information in $\hat{h}_k$ is captured by $\hat{\hat{h}}_k$. We now derive the SEP for MPSK and MQAM for the proposed method as a function of α and β.

Theorem *1:* With noisy channel estimates, the fading-averaged SEP of MPSK and MQAM is:

$$P^{\alpha,\beta}_{\text{MPSK}}(\gamma) = \frac{1}{\pi}\int_0^{\frac{M-1}{M}\pi}\left(\frac{\sin^2\theta}{a_2 b_{\text{PSK}} + \sin^2\theta}\right)^L \times \prod_{n=L+1}^{N}\left(1+\frac{a_1 b_{\text{PSK}}L/n}{(a_2-a_1)b_{\text{PSK}} + \sin^2\theta}\right)^{-1}d\theta, \quad (16)$$

$$P^{\alpha,\beta}_{\text{MQAM}}(\gamma) = \frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)\int_0^{\frac{\pi}{2}}\xi(\theta)\left(\frac{\sin^2\theta}{a_2 b_{\text{QAM}} + \sin^2\theta}\right)^L \times \prod_{n=L+1}^{N}\left(1+\frac{a_1 b_{\text{QAM}}L/n}{(a_2-a_1)b_{\text{QAM}} + \sin^2\theta}\right)^{-1}d\theta, \quad (17)$$

where $\xi(\theta) = \frac{1}{\sqrt{M}}$, for $0 \le \theta < \frac{\pi}{4}$, and $\xi(\theta) = 1$, for $\frac{\pi}{4} \le \theta \le \frac{\pi}{2}$, $a_1 = \frac{\alpha\gamma}{(a+\gamma)\alpha+\beta+d}$, $a_2 = \frac{(\alpha+\beta)\gamma}{(a+\gamma)\alpha+(\gamma+1)\beta+d}$, $b_{\text{PSK}} = \frac{\gamma\sin^2\left(\frac{\pi}{M}\right)}{(1-a_2)\gamma+a\alpha+\beta+d}$, and $b_{\text{QAM}} = \frac{1.5\gamma/(M-1)}{(1-a_2)\gamma+a\alpha+\beta+d}$.

Proof: The proof is given in Appendix D. ∎

Closed-form expressions can be derived from (16) and (17) using [26, (5A.42), (5A.56)]. However, they are quite involved, and are not shown here. The optimal values of α and β are then found numerically using gradient search. Considerable insight about them can be gained by analyzing the asymptotic energy regime, as we show below.

C. Asymptotic Characterization of SEP ($\gamma \to \infty$)

In the asymptotic regime, the SEP of MPSK and MQAM shall be denoted by $P^{\alpha,\beta,\infty}_{\text{MPSK}}(\gamma)$ and $P^{\alpha,\beta,\infty}_{\text{MQAM}}(\gamma)$, respectively. In the results below, we do not show higher order terms involving γ since their contribution becomes negligible as $\gamma \to \infty$.

Lemma *3:* The asymptotic SEPs of MPSK and MQAM are given by

$$P^{\alpha,\beta,\infty}_{\text{MPSK}}(\gamma) = \gamma^{-N}\frac{(L+1)(L+2)\cdots(N)}{\pi L^{N-L}(4(\alpha+\beta))^N} \times (a\alpha+\beta+d)^N(\alpha+\beta+1)^N\sin^{-2N}\left(\frac{\pi}{M}\right) \times \sum_{k=0}^{N-L}\binom{N-L}{k}\left(\frac{4\beta\sin^2\left(\frac{\pi}{M}\right)}{\alpha(\alpha+\beta+1)}\right)^k\psi\left(\frac{M-1}{M}\pi, N-k\right), \quad (18)$$

$$P_{\text{MQAM}}^{\alpha,\beta,\infty}(\gamma) = \gamma^{-N}\frac{4\,(L+1)\,(L+2)\cdots(N)}{\pi L^{N-L}\,(6(\alpha+\beta))^{N}} \times (a\alpha+\beta+d)^{N}\,(\alpha+\beta+1)^{N}\,(M-1)^{N}\left(1-\frac{1}{\sqrt{M}}\right) \times \sum_{k=0}^{N-L}\left[\binom{N-L}{k}\left(\frac{6\beta/(M-1)}{\alpha(\alpha+\beta+1)}\right)^{k} \times\left(\psi\left(\frac{\pi}{2},N-k\right)-\left(1-\frac{1}{\sqrt{M}}\right)\psi\left(\frac{\pi}{4},N-k\right)\right)\right]. \quad (19)$$

Proof: The proof is given in Appendix E. ∎

The above expressions show that the diversity order is N for any $\alpha > 0$ and $\beta \geq 0$. This is to be expected since the conventional training method, which is a special case of the proposed scheme with regard to SEP, also achieves a diversity order of N.

Even the form in (18) is intractable for determining the optimal values of α and β. We, therefore, minimize an upper bound on the SEP that is obtained by replacing $\sin^2\theta$ with 1 in (16) [27]. Let α_∞^* and β_∞^* denote the optimal values of α and β, respectively, that minimize the SEP upper bound.

From the proof of Lemma 3 in Appendix E (see (41)), it can be shown that $P_{\text{MPSK}}^{\alpha,\beta,\infty}(\gamma) \leq \gamma^{-N} U_{\text{MPSK}}^{\alpha,\beta}(\gamma)$, where

$$U_{\text{MPSK}}^{\alpha,\beta} \triangleq \frac{(L+1)\,(L+2)\cdots(N)}{L^{N-L}}\left(\frac{M-1}{M}\right)\sin^{-2N}\left(\frac{\pi}{M}\right) \times\left(\frac{(a\alpha+\beta+d)(\alpha+\beta+1)}{\alpha+\beta}\right)^{N}\left(1+\frac{\beta\sin^2\left(\frac{\pi}{M}\right)}{\alpha(\alpha+\beta+1)}\right)^{N-L}. \quad (20)$$

Similarly, for MQAM, $P_{\text{MQAM}}^{\alpha,\beta,\infty}(\gamma) \leq \gamma^{-N} U_{\text{MQAM}}^{\alpha,\beta}$, where

$$U_{\text{MQAM}}^{\alpha,\beta} \triangleq \frac{(L+1)\,(L+2)\cdots(N)}{L^{N-L}}\left(\frac{M-1}{M}\right)\left(\frac{1.5}{M-1}\right)^{-N} \times\left(\frac{(a\alpha+\beta+d)(\alpha+\beta+1)}{\alpha+\beta}\right)^{N}\left(1+\frac{\beta\left(\frac{1.5}{M-1}\right)}{\alpha(\alpha+\beta+1)}\right)^{N-L}. \quad (21)$$

As we see below, the upper bounds provide considerable insight about the optimal parameters. Specifically, for MPSK, α_∞^* and β_∞^* satisfy the following properties:

Theorem *2:* 1) α_∞^* and β_∞^* are related by

$$\beta_\infty^* = -\alpha_\infty^* + \sqrt{d-(\alpha_\infty^*)^2\,(a-1)}. \quad (22)$$

2) α_∞^* is a zero of the function $g_1(\alpha)$, where

$$g_1(\alpha) \triangleq \sqrt{d-\alpha^2(a-1)}\left(N\alpha^2\csc^2\left(\frac{\pi}{M}\right)+L\alpha-d\left(\frac{N-L}{a-1}\right)\right) + \alpha^2\left(N\csc^2\left(\frac{\pi}{M}\right)-L\right)-d\left(\frac{N-L}{a-1}\right). \quad (23)$$

3) And, α_∞^* is unique and lies in the following range:

$$0 < \alpha_\infty^* \leq \sqrt{\frac{d(N-L)}{N(a-1)}}\sin\left(\frac{\pi}{M}\right) \leq \varepsilon_\infty^* = \sqrt{\frac{d}{a}}, \quad (24)$$

with the equalities holding only if both the following conditions hold: $M = 2$ (BPSK) and L divides N.

Proof: The proof is given in Appendix F. ∎

For large constellation sizes ($M \to \infty$), the optimal values α_∞^* and β_∞^* can, in fact, be determined in closed-form as shown below.

Corollary *1:*

$$\lim_{M\to\infty}\frac{\alpha_\infty^*}{\sin\left(\frac{\pi}{M}\right)} = \sqrt{\frac{d(N-L)}{N(a-1)}}. \quad (25)$$

Proof: The proof is given in Appendix G. ∎

Corollary 1 motivates the following approximation for any M, which is obtained by substituting (25) in (22):

$$\alpha_\infty^* \approx \sqrt{\frac{d(N-L)}{N(a-1)}}\sin\left(\frac{\pi}{M}\right), \quad (26)$$

$$\beta_\infty^* \approx -\sqrt{\frac{d(N-L)}{N(a-1)}}\sin\left(\frac{\pi}{M}\right)+\sqrt{d}\sqrt{1-\left(1-\frac{L}{N}\right)\sin^2\left(\frac{\pi}{M}\right)}. \quad (27)$$

From the Corollary, it directly follows that the error in the approximation disappears as $M \to \infty$. Interestingly, we shall see that the error in (26) is small even when M is as small as 4.

Corresponding Results for MQAM: Results analogous to Theorem 2 and Corollary 1 are obtained for MQAM by simply replacing $\sin\left(\frac{\pi}{M}\right)$ with $\sqrt{\frac{1.5}{M-1}}$. This follows because the expressions in (20) and (21) are very similar except that $\sin\left(\frac{\pi}{M}\right)$ is replaced by $\sqrt{\frac{1.5}{M-1}}$.

The asymptotic expressions for SEP provide the following insights about the proposed scheme.

1) Variation with System Parameters: While ε_∞^* in the conventional method depends only on d and the number of pilot symbols, a; in the proposed scheme, α_∞^* depends on *all* system parameters N, L, d, M, and a. Furthermore, as d increases, the relative energy allocated to training increases in the conventional and proposed methods. This trend is consistent with the observations in [18], which considered training for multiple antenna systems to maximize average throughput.

2) BPSK ($M = 2$): Theorem 2 shows that this is the only case when the proposed method provides no benefits over the conventional method if L divides N. As expected, for this case, (24) implies that $\alpha_\infty^* = \varepsilon_\infty^*$ and $\beta_\infty^* = 0$. However, the conclusion is different when L does not divide N even for BPSK. In this case, $\alpha_\infty^* < \varepsilon_\infty^*$, and the proposed scheme necessarily improves performance.

3) Asymptotic Energy Gain: The conventional and proposed methods both achieve the full diversity order of N. Hence, we can compare them in terms of the *asymptotic energy gain, Δ, which measures the savings in total energy achieved by the proposed method over the conventional optimized method for the same SEP.* It is easy to see that, for MPSK, $\Delta = \frac{10}{N}\log_{10}\left(P_{\text{MPSK}}^{\varepsilon,\infty}/P_{\text{MPSK}}^{\alpha,\beta,\infty}\right)$ dB.

Taking the ratio of (13) and (18), we get

$$\Delta = 10\log_{10}\left(\frac{\left(\sqrt{d}+\sqrt{a}\right)^2(\alpha_\infty^*+\beta_\infty^*)}{(\alpha_\infty^*+\beta_\infty^*+1)\,(a\alpha_\infty^*+\beta_\infty^*+d)}\right)$$

$$-\frac{10}{N}\log_{10}\left[\sum_{k=0}^{N-L}\binom{N-L}{k}\left(\frac{4\beta_\infty^*\sin^2\left(\frac{\pi}{M}\right)}{\alpha_\infty^*\left(\alpha_\infty^*+\beta_\infty^*+1\right)}\right)^k \times\left(\psi\left(\frac{M-1}{M}\pi,N-k\right)\Big/\psi\left(\frac{M-1}{M}\pi,N\right)\right)\right], \quad (28)$$

where α_∞^* and β_∞^* are either found numerically or approximated using (26).

4) Energy Allocation: $E_1 \le E_2$ and $\alpha_\infty^*+\beta_\infty^* \ge \varepsilon_\infty^*$. From (4) and (9), for optimal values of the parameters, it can be seen that $\frac{E_1}{E_2}-1=\frac{(a-1)\alpha_\infty^*+\sqrt{d-(\alpha_\infty^*)^2(a-1)}-a\varepsilon_\infty^*}{a\varepsilon_\infty^*+d}$. This is negative because $d-(\alpha_\infty^*)^2(a-1)-(a\varepsilon_\infty^*-(a-1)\alpha_\infty^*)^2 = -a(a-1)(\varepsilon_\infty^*-\alpha_\infty^*)^2 \le 0$ (since $a \ge 1$). Thus, $E_1 \le E_2$. From (26), $\frac{\alpha_\infty^*+\beta_\infty^*}{\varepsilon_\infty^*} = \sqrt{a}\sqrt{1-\left(1-\frac{L}{N}\right)\sin^2\left(\frac{\pi}{M}\right)}$. Since $a \ge \frac{N}{L}$, it follows that $(\alpha_\infty^*+\beta_\infty^*)/\varepsilon_\infty^* \ge 1$. Furthermore, the ratio increases as M increases.

Note that $\varepsilon_\infty^* E_1$ and $(\alpha_\infty^*+\beta_\infty^*)E_2$ respectively denote the quality of channel estimates used for data decoding in the conventional method and the proposed method. *Thus, the proposed method not only allocates more energy to each data symbol (since $E_2 \ge E_1$), but it also ensures that the quality of estimates is better than the conventional method (since $(\alpha_\infty^*+\beta_\infty^*)E_2 \ge \varepsilon_\infty^* E_1$).*

5) Behavior of E_2/E_1, i.e., Ratio of Data Energies Allocated by The Two Methods: Understanding the behavior of E_2/E_1 will help explain the behavior of SEP and Δ in the next section.

(i) *E_2/E_1 increases with M:* To show this, it is sufficient to show that $\frac{\partial}{\partial\alpha_\infty^*}\frac{E_2}{E_1} \le 0$ since α_∞^* scales as $\sin\left(\frac{\pi}{M}\right)$. From (4) and (9), it can be seen that $\frac{E_2}{E_1}=\frac{a\varepsilon_\infty^*+d}{(a-1)\alpha_\infty^*+\sqrt{d-(\alpha_\infty^*)^2(a-1)}+d}$. Since ε_∞^* does not depend on M, its sufficient to show that the denominator is non-decreasing in α_∞^*. This follows because $\frac{\partial}{\partial\alpha_\infty^*}\left((a-1)\alpha_\infty^*+\sqrt{d-(\alpha_\infty^*)^2(a-1)}\right) = (a-1)\left(1-\frac{\alpha_\infty^*}{\sqrt{d-(\alpha_\infty^*)^2(a-1)}}\right) \ge 0$ since $\alpha_\infty^* \le \sqrt{\frac{d}{a}}$.

(ii) *E_2/E_1 increases with L and decreases with N (when a is fixed):* From (4), (9), and Theorem 2, we see that $\frac{E_2}{E_1}=\frac{a\varepsilon_\infty^*+d}{(a-1)\alpha_\infty^*+\sqrt{d-(\alpha_\infty^*)^2(a-1)}+d}$. Substituting the optimal values of the parameters and simplifying further, we get

$$\frac{E_2}{E_1}=\frac{\sqrt{a}+\sqrt{d}}{\sqrt{(a-1)\tau}+\sqrt{1-\tau}+\sqrt{d}}, \quad (29)$$

where $\tau \triangleq \left(1-\frac{L}{N}\right)\sin^2\left(\frac{\pi}{M}\right)$. To prove the desired result it is sufficient to show that the denominator is non-decreasing in τ. This follows because

$$\frac{\partial}{\partial\tau}\left(\sqrt{(a-1)\tau}+\sqrt{1-\tau}\right)=0.5\left(\sqrt{\frac{a-1}{\tau}}-\sqrt{\frac{1}{1-\tau}}\right).$$

This is non-negative because the inequality $a \ge \frac{N}{L}$ implies that $\left(1-\frac{1}{a}\right) \ge \tau$, which after algebraic manipulations implies that $\sqrt{\frac{a-1}{\tau}} \ge \sqrt{\frac{1}{1-\tau}}$.

(iii) *E_2/E_1 decreases with d:* On substituting the optimal values of parameters, the expression for the ratio simplifies to $\frac{E_2}{E_1}=\frac{\sqrt{a}+\sqrt{d}}{\sqrt{(a-1)\left(1-\frac{L}{N}\right)}\sin\left(\frac{\pi}{M}\right)+\sqrt{1-\left(1-\frac{L}{N}\right)\sin^2\left(\frac{\pi}{M}\right)}+\sqrt{d}}$. Since $E_2 \ge E_1$, as shown in Sec. III-C4, the expression takes the form $\frac{E_2}{E_1}=1+\frac{f_1}{f_2}$, with f_1 is positive and independent of d, and f_2 is positive and monotonically increases with d. Hence, the ratio decreases with d.

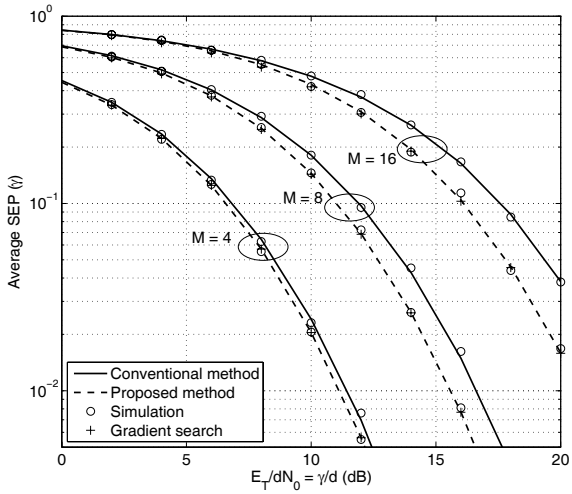


Fig. 2. Effect of constellation size on average SEP of MPSK ($N = 6$, $L = 1$, and $d = 10$).

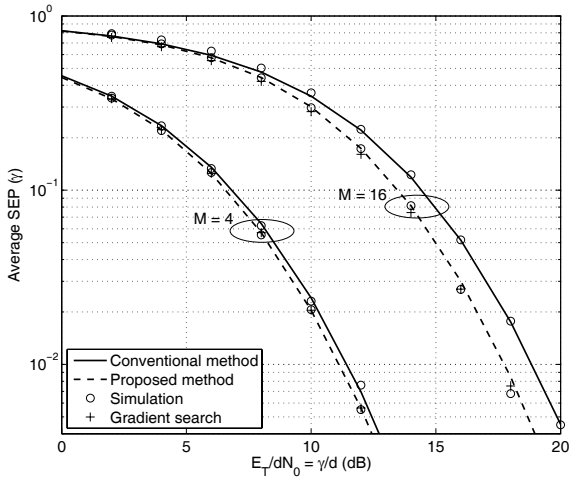


Fig. 3. Effect of constellation size on average SEP of MQAM ($N = 6$, $L = 1$, and $d = 10$).

IV. Simulation Results

We now plot the analytical results derived in Sec. III and validate them with Monte Carlo simulations that use 10^5 samples for each SNR. As specified in the system model, the channel remains constant for $d+a+1$ symbols. We also compare the conventional and proposed AS training methods as a function of all the system parameters N, L, d, and M.

Figures 2 and 3 plot the SEP as a function of normalized total energy, $\frac{E_T}{dN_0}$ $(=\frac{\gamma}{d})$, for MPSK and MQAM, respectively, for a given L and N. While the SEP of both methods expectedly increases with M, the performance gain of the proposed method increases with M. This is because ε_∞^* of the conventional method is insensitive to M despite the fact that

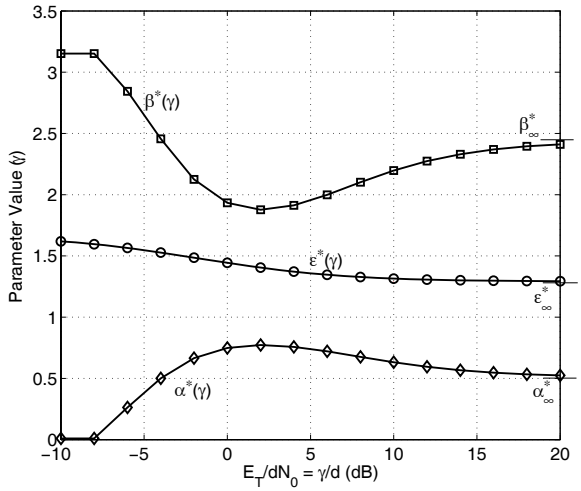


Fig. 4. Optimal values of different parameters as a function of $\frac{E_T}{dN_0}$ (8PSK, $N = 6$, $L = 1$, and $d = 10$).

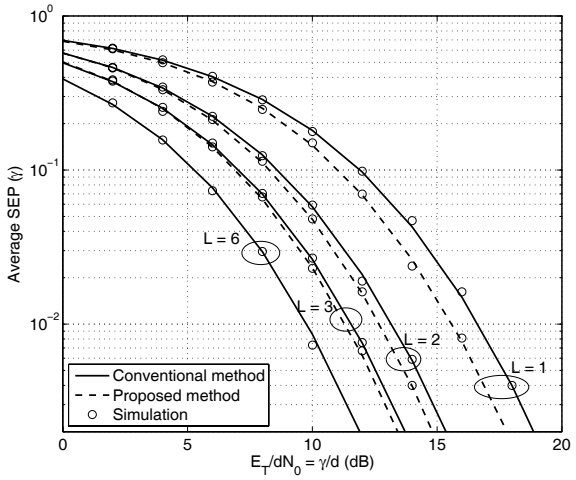


Fig. 5. Effect of number of receive RF chains on average SEP (8PSK, $N = 6$, and $d = 10$).

the SEP of larger constellations is more sensitive to channel estimation errors. The optimal values of α and β that are used in the figures are found by using a gradient search over the SEP formulae of Theorem 1.[3] The figures also plot the SEP obtained by using the approximate values of α^*_∞ and β^*_∞ given in (26). Notice that the SEP using the approximation is accurate even at low values of γ. Furthermore, there is an excellent match between the analytical and simulation results. The small mismatch between analytical and simulation results for MQAM is explained in Appendix D.

Figure 4 shows the optimal energy allocation (ε^*, α^*, and β^*) as a function of the normalized total energy. $\alpha^*(\gamma)$ and $\beta^*(\gamma)$ are obtained by performing a gradient search that minimizes $P_{\text{MPSK}}^{\alpha-\beta}(\gamma)$. Also shown are the asymptotically optimal values α^*_∞ and β^*_∞ of (26). These are very close to the exact optimal values for $\frac{\gamma}{d} \geq 15$ dB. Henceforth, we, therefore, only use α^*_∞ and β^*_∞ of (26), unless mentioned otherwise. In the conventional method, ε^* monotonically decreases with γ and saturates at $\sqrt{\frac{d}{a}} = 1.291$. However, the proposed method behaves differently. At low γ, $\alpha^*(\gamma)$ is close to zero and $\beta^*(\gamma)$ is large since selection does not matter, only coherent reception does. Once γ crosses a threshold, the system allocates more energy to the selection pilots. This triggers a corresponding decrease in $\beta^*(\gamma)$ since the selection pilot and the last pilot are both used for coherent reception.

Figure 5 shows the effect of the antenna subset size, L, on the SEP. As L increases, the number of pilots that need to be transmitted decreases. Hence, relatively less energy is spent on training. Consequently, for both methods, the SEP improves for all γ.[4]

[3]Since the optimal solution is unique in the asymptotic regime ($\gamma \to \infty$), we know that the gradient search will converge to the global minimum at least for large γ. In our numerical computations, we have observed the same for smaller γ, as well.

[4]The behavior with N turns out to be different. At low γ, a smaller N is better because the energy-starved system cannot afford to spend energy on training. At high γ, where diversity order matters, a larger N does better. Due to the better utilization of energy by the proposed method, this crossover turns out to occur at a smaller value of γ.

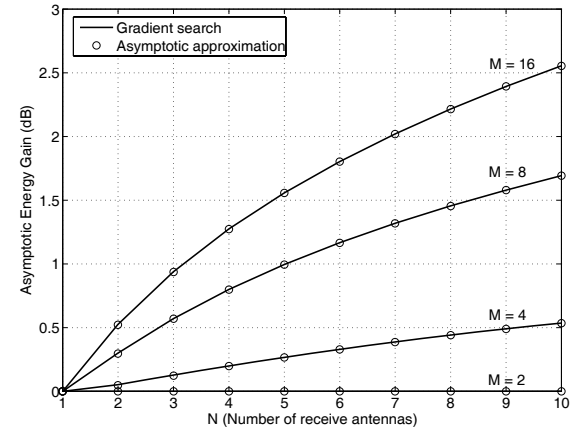


Fig. 6. Effect of number of receive antennas on asymptotic energy gain (MPSK, $d = 10$, and $L = 1$).

Figure 6 plots the asymptotic energy gain, Δ, as a function of N and M. We can see that the energy gain increases with M. This is because: (i) The relative quality of estimates improves with M. (This can be seen from the behavior of the ratio $\frac{\alpha^*_\infty + \beta^*_\infty}{\varepsilon^*_\infty}$, which was discussed in Sec. III-C4.) (ii) The proposed method allocates relatively more energy to the data symbols, as shown in Sec. III-C5. A similar argument, not repeated here due to space constraints, also explains the variation with N.

Figure 7 plots the energy gain as a function of L for fixed N and different values of M. We choose N as 16 in order to illustrate how the energy gain increases as L increases from $\frac{N}{2}$ to $N-1$. This figure can be understood as follows: (i) For $\frac{N}{2} \leq L < N$, a is fixed at 2. Hence, $\frac{\alpha^*_\infty + \beta^*_\infty}{\varepsilon^*_\infty}$ increases with L. (ii) E_2/E_1 increases with L, as shown in Sec. III-C5. Hence, the energy gain increases with L in this region. The same is true even when L increases from 4 to 5 ($a = 4$) and 6 to 7 ($a = 3$). In other regions, as L increases, a decreases. Hence, the energy gain decreases.

Figure 8 studies the joint impact of d and M on the

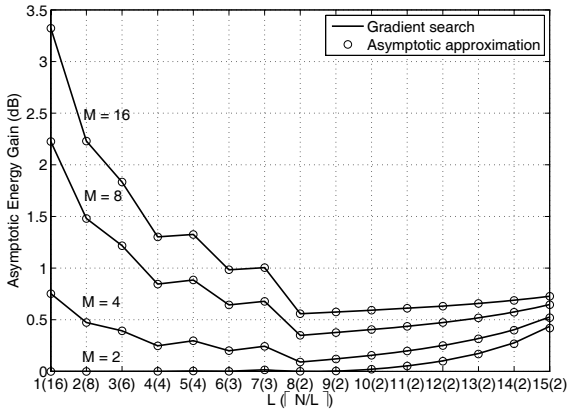


Fig. 7. Effect of number of receive RF chains on asymptotic energy gain (MPSK, $d = 10$, and $N = 16$). In the x-axis, the numbers in brackets are the number of pilot symbols, $a = \lceil N/L \rceil$, transmitted for the given L.

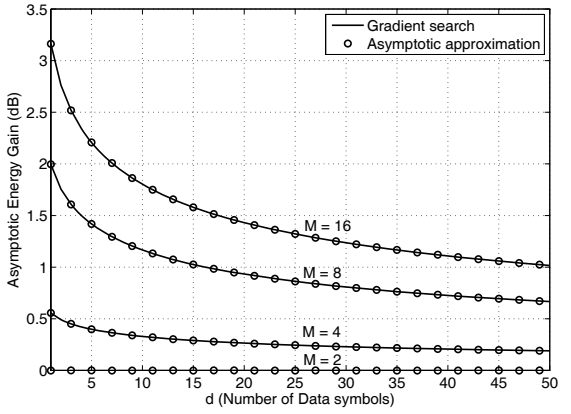


Fig. 8. Effect of number of data symbols on asymptotic energy gain (MPSK, $N = 6$, and $L = 1$).

asymptotic energy gain. The energy gain decreases as d increases because: (i) The relative quality of channel estimates decreases since the ratio $\frac{\alpha_\infty^* + \beta_\infty^*}{\varepsilon_\infty^*}$ does not depend on d. (ii) The relative energy allocated to data symbols decreases since the ratio E_2/E_1 decreases with d (see Sec. III-C5). While the energy gain decreases, it still exceeds 1 dB when $d \leq 15$ for $M \geq 8$. The above three figures again show that the closed-form approximations for α_∞^* and β_∞^* are just as accurate as the optimal values found numerically using gradient search.

V. Conclusions

We proposed a new training method tailored for antenna subset selection that exploited the observation that subset selection is robust to channel estimation errors, while coherent demodulation is not. The method assigns less energy to 'selection pilots', which are used to select the antenna subset. It instead allocates more energy to an extra last pilot to refine the channel estimates for the antenna subset actually used for data reception. We derived closed-form equations for the SEP of both MPSK and MQAM, and showed analytically that it is lower than the SEP of the conventional method that uses the same pilots for selection and channel estimation.

The analysis gave considerable insight about the behavior of the proposed method vis-à-vis the conventional method. While the conventional method's optimal energy allocation depended only on the number of data symbols, d, and the number of pilot symbols, a, the proposed method's optimal parameters depended on N, L, d, and the constellation size, M. Approximate, but closed-form, expressions are derived for the optimal parameters that are exact as $M \to \infty$. Even for $M = 4$, $d = 10$, $N = 6$, and $\gamma = 6$ dB, the approximation error in SEP is just 4%. The energy gain increased as the constellation size or the number of pilots transmitted increased. It decreased when more data symbols were to be received by the same antenna subset.

The considerable energy savings achieved motivate future work that incorporates a time-varying channel model and an outage capacity analysis that quantifies the throughput penalty of including an extra pilot symbol. Finally, the use of data-decision-aided refinement of channel estimates is worth exploring.

Appendix

A. Proof of Lemma 1

From (10), it is clear that the SEP of MPSK monotonically decreases with c_{PSK}. Hence, the ε that maximizes c_{PSK} will minimize the SEP of MPSK. Similarly, from (11), the ε that maximizes c_{QAM} will minimize the SEP of MQAM. Since, c_{PSK} and c_{QAM} differ only by a constant scaling factor, the same solution is optimal for both MPSK and MQAM.

For MPSK, $\frac{\partial}{\partial \varepsilon} c_{\text{PSK}}$ must be 0 at the extremum point, *i.e.*, $\frac{\partial}{\partial \varepsilon}\left(\frac{\varepsilon}{(a\varepsilon+d)((\varepsilon+1)\gamma+a\varepsilon+d)}\right)\Big|_{\varepsilon=\varepsilon^*} = 0$. This results in the equation: $(a\varepsilon^*+d)\left(\varepsilon^*(\gamma+a)+\gamma+d\right) = \varepsilon^*\left(2a\varepsilon^*(\gamma+a)+d(\gamma+a)+a(d+\gamma)\right)$. Solving it shows that $\varepsilon^* = \sqrt{\frac{d(\gamma+d)}{a(\gamma+a)}}$ is the unique positive-valued optimum.

B. Derivation of (13)

From the expression for c_{PSK} given in Sec. III-A, as $\gamma \to \infty$, $\lim_{\gamma\to\infty} \frac{c_{\text{PSK}}}{\gamma} = \frac{\varepsilon \sin^2\left(\frac{\pi}{M}\right)}{(a\varepsilon+d)(\varepsilon+1)}$. Hence, the asymptotic SEP expression simplifies to

$$P_{\text{MPSK}}^{\varepsilon,\infty}(\gamma) = \gamma^{-N} \frac{(L+1)(L+2)\cdots(N)}{\pi L^{N-L}\varepsilon^N} (a\varepsilon+d)^N \times (\varepsilon+1)^N \sin^{-2N}\left(\frac{\pi}{M}\right) \int_0^{\frac{M-1}{M}\pi} \sin^{2N}\theta d\theta. \quad (30)$$

Using the following result from [28, (2.513)] in (30) yields the desired expression:

$$\int_0^T \sin^{2k} t \, dt = \frac{T}{2^{2k}}\binom{2k}{k} + \frac{(-1)^k}{2^{2k-1}} \sum_{j=0}^{k-1} (-1)^j \binom{2k}{j} \frac{\sin[(2k-2j)T]}{2k-2j}. \quad (31)$$

C. Proof of Lemma 2

From (5), we see that h_k is independent of e_k but not $\hat{h}_k$. We, therefore, need to take recourse to the following two standard results on moments of conditional Gaussians: If X, a complex Gaussian RV, and Y, a complex Gaussian random vector, are jointly Gaussian, then

$$\begin{aligned}\mathbf{E}[X|Y] &= \mathbf{E}[X] + \Sigma_{XY}\Sigma_Y^{-1}(Y - \mathbf{E}[Y]),\\ \mathbf{var}[X|Y] &= \mathbf{var}[X] - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{XY}^H,\end{aligned} \tag{32}$$

where Σ_{XY} is the cross-correlation of X and Y and Σ_Y is the covariance of Y.

In our case, $X \triangleq h_k$ and $Y \triangleq \begin{bmatrix}\hat{h}_k\\ \hat{\hat{h}}_k\end{bmatrix}$. We can show from (5) and (7) that $\Sigma_Y = \begin{bmatrix}a_1 & a_1\\ a_1 & a_2\end{bmatrix}$ and $\Sigma_{XY} = \begin{bmatrix}a_1 & a_2\end{bmatrix}$, where, as mentioned, $a_1 = \frac{\alpha E_2}{\alpha E_2 + N_0}$ and $a_2 = \frac{(\alpha+\beta)E_2}{(\alpha+\beta)E_2+N_0}$. Hence,

$$\mathbf{E}\left[h_k \,\middle|\, \hat{h}_k, \hat{\hat{h}}_k\right] = \begin{bmatrix}a_1 & a_2\end{bmatrix}\begin{bmatrix}a_1 & a_1\\ a_1 & a_2\end{bmatrix}^{-1}\begin{bmatrix}\hat{h}_k\\ \hat{\hat{h}}_k\end{bmatrix} = \hat{\hat{h}}_k,$$

and

$$\begin{aligned}\mathbf{var}\left[h_k \,\middle|\, \hat{h}_k, \hat{\hat{h}}_k\right] &= 1 - \begin{bmatrix}a_1 & a_2\end{bmatrix}\begin{bmatrix}a_1 & a_1\\ a_1 & a_2\end{bmatrix}^{-1}\begin{bmatrix}a_1\\ a_2\end{bmatrix}\\ &= 1 - a_2.\end{aligned}$$

Therefore, the decision variable's conditional mean and variance are

$$\mathbf{E}\left[\mathcal{D}\,\middle|\, \left\{\hat{h}_l, \hat{\hat{h}}_l\right\}_{l\in\hat{\Omega}_L}, s\right] = \sum_{k\in\hat{\Omega}_L}\hat{\hat{h}}_k^*\mathbf{E}\left[y_k \,\middle|\, \hat{h}_k, \hat{\hat{h}}_k, s\right],$$

$$\mathbf{var}\left[\mathcal{D}\,\middle|\, \left\{\hat{h}_l, \hat{\hat{h}}_l\right\}_{l\in\hat{\Omega}_L}, s\right] = \sum_{k\in\hat{\Omega}_L}\left|\hat{\hat{h}}_k\right|^2\mathbf{var}\left[y_k \,\middle|\, \hat{h}_k, \hat{\hat{h}}_k, s\right].$$

From (8), we have $\mathbf{E}\left[y_k \,\middle|\, \hat{h}_k, \hat{\hat{h}}_k, s\right] = \hat{\hat{h}}_k s$ and $\mathbf{var}\left[y_k \,\middle|\, \hat{h}_k, \hat{\hat{h}}_k, s\right] = \mathbf{var}\left[h_k \,\middle|\, \hat{h}_k, \hat{\hat{h}}_k\right]|s|^2 + N_0 = (1 - a_2)E_2 + N_0$. Hence, the desired result follows.

D. Proof of Theorem 1

MPSK: The standard SEP expression for MPSK when $\mathcal{D}$ is a Gaussian RV [27, (40)] is

$$P_{\text{MPSK}}\left(\left\{\hat{h}_l, \hat{\hat{h}}_l\right\}_{l\in\hat{\Omega}_L}\right) = \frac{1}{\pi}\int_0^{\frac{M-1}{M}\pi}\exp\left(\frac{-|\mu_{\mathcal{D}}|^2\sin^2\left(\frac{\pi}{M}\right)}{\sigma_{\mathcal{D}}^2\sin^2\theta}\right)d\theta. \tag{33}$$

The notation above clearly shows the dependence of P_{MPSK} on the observables $\hat{h}_l, \hat{\hat{h}}_l$, and $\hat{\Omega}_L$. Note that $\hat{\Omega}_L$, in turn, depends on $\hat{h}_l$. From Lemma 2, the above equation can be simplified to

$$P_{\text{MPSK}}\left(\left\{\hat{h}_l, \hat{\hat{h}}_l\right\}_{l\in\hat{\Omega}_L}\right) = \frac{1}{\pi}\int_0^{\frac{M-1}{M}\pi}\exp\left(-\frac{b_{\text{PSK}}}{\sin^2\theta}\sum_{k\in\hat{\Omega}_L}\left|\hat{\hat{h}}_k\right|^2\right)d\theta, \tag{34}$$

where $b_{\text{PSK}} = \frac{E_2\sin^2\left(\frac{\pi}{M}\right)}{(1-a_2)E_2+N_0} = \frac{\gamma\sin^2\left(\frac{\pi}{M}\right)}{(1-a_2)\gamma+a\alpha+\beta+d}$. Let $Y \triangleq \sum_{k\in\hat{\Omega}_L}\left|\hat{\hat{h}}_k\right|^2$. Averaging over $\left\{\hat{\hat{h}}_l\right\}_{l\in\hat{\Omega}_L}$, the SEP simplifies to

$$P_{\text{MPSK}}\left(\left\{\hat{h}_l\right\}_{l\in\hat{\Omega}_L}\right) = \frac{1}{\pi}\int_0^{\frac{M-1}{M}\pi}\mathcal{M}_{Y\,|\,\{\hat{h}_l\}_{l\in\hat{\Omega}_L}}\left(\frac{-b_{\text{PSK}}}{\sin^2\theta}\right)d\theta, \tag{35}$$

where $\mathcal{M}_{Y\,|\,\{\hat{h}_l\}_{l\in\hat{\Omega}_L}}(.)$ is the moment generating function (MGF) of $\sum_{k\in\hat{\Omega}_L}\left|\hat{\hat{h}}_k\right|^2$ conditioned on $\left\{\hat{h}_l\right\}_{l\in\hat{\Omega}_L}$ [30]. We know that $\hat{\hat{h}}_k$ conditioned on $\hat{h}_k$ is a Gaussian RV. Using the conditional results in (32), we can show that $\mathbf{E}\left[\hat{\hat{h}}_k \,\middle|\, \hat{h}_k\right] = \hat{h}_k$ and $\mathbf{var}\left[\hat{\hat{h}}_k \,\middle|\, \hat{h}_k\right] = a_2 - a_1$. Hence, Y is a non-central Chi-square distributed RV, whose conditional MGF is [31]

$$\mathcal{M}_{Y\,|\,\{\hat{h}_l\}_{l\in\hat{\Omega}_L}}(x) = (1-(a_2-a_1)x)^{-L}\exp\left(\frac{\sum_{k\in\hat{\Omega}_L}\left|\hat{h}_k\right|^2 x}{1-(a_2-a_1)x}\right). \tag{36}$$

Substituting (36) in (35) and averaging over the channel estimates of the selected subset of antennas, $\left\{\hat{h}_l\right\}_{l\in\hat{\Omega}_L}$, yields

$$\begin{aligned}P_{\text{MPSK}} = \frac{1}{\pi}\int_0^{\frac{M-1}{M}\pi}&\left(1 + \frac{(a_2-a_1)b_{\text{PSK}}}{\sin^2\theta}\right)^{-L}\\ &\times\mathcal{M}_{\sum_{k\in\hat{\Omega}_L}|\hat{h}_k|^2}\left(\frac{-b_{\text{PSK}}}{\sin^2\theta + (a_2-a_1)b_{\text{PSK}}}\right)d\theta,\end{aligned} \tag{37}$$

where $\mathcal{M}_{\sum_{k\in\hat{\Omega}_L}|\hat{h}_k|^2}$ is the MGF of $\sum_{k\in\hat{\Omega}_L}\left|\hat{h}_k\right|^2$. Using the virtual branch combining technique of [6], the MGF of $\sum_{k\in\hat{\Omega}_L}\left|\hat{h}_k\right|^2$ becomes [9]

$$\mathcal{M}_{\sum_{k\in\hat{\Omega}_L}|\hat{h}_k|^2}(x) = (1-a_1x)^{-L}\prod_{n=L+1}^{N}\left(1 - \frac{a_1Lx}{n}\right)^{-1}.$$

Substituting this in (37) and simplifying further gives the fading-averaged SEP expression.[5]

MQAM: Using Lemma 2, the standard MQAM SEP expression is [27, (48)][6]

$$\begin{aligned}&P_{\text{MQAM}}\left(\left\{\hat{h}_l, \hat{\hat{h}}_l\right\}_{l\in\hat{\Omega}_L}\right)\\ &= \frac{4}{\pi}\left(1 - \frac{1}{\sqrt{M}}\right)\int_0^{\frac{\pi}{2}}\xi(\theta)\exp\left(-\frac{b_{\text{QAM}}}{\sin^2\theta}\sum_{k\in\hat{\Omega}_L}\left|\hat{\hat{h}}_k\right|^2\right)d\theta,\end{aligned} \tag{38}$$

where $\xi(\theta) = \frac{1}{\sqrt{M}}$, for $0 \le \theta < \frac{\pi}{4}$, and $\xi(\theta) = 1$, for $\frac{\pi}{4} \le \theta \le \frac{\pi}{2}$, and $b_{\text{QAM}} = \frac{1.5E_2}{(M-1)((1-a_2)E_2+N_0)} = \frac{1.5\gamma/(M-1)}{(1-a_2)\gamma+a\alpha+\beta+d}$. Proceeding along lines similar to that of MPSK, gives the required result. This involves a two step process that first averages over $\left\{\hat{\hat{h}}_l\right\}_{l\in\hat{\Omega}_L}$ given $\left\{\hat{h}_l\right\}_{l\in\hat{\Omega}_L}$, and then averages over $\left\{\hat{h}_l\right\}_{l\in\hat{\Omega}_L}$. We skip intermediate steps to avoid repetition and to conserve space.

[5]Note that the initial steps in this proof such as computing the moments of the decision variable and using (33) are similar to [9]. However, it differs thereafter from [9] because the SEP in our case depends on both $\hat{h}_l$ and $\hat{\hat{h}}_l$.

[6]This expression implicitly assumes in its derivation that the variance of $\mathcal{D}$ is the same for all symbols, which is not so for the MQAM constellation in the presence of estimation errors. However, as the simulation results in [9], [15], and this paper show, the expression is accurate.

E. Proof of Lemma 3

From the expressions for a_2, a_1, and b_{PSK} given in Sec. III-B, as $\gamma \to \infty$, we have the following:

$$\lim_{\gamma\to\infty}(a_2 - a_1)b_{\text{PSK}} = \frac{\beta \sin^2\left(\frac{\pi}{M}\right)}{\alpha(\alpha+\beta+1)}, \tag{39}$$

and

$$\lim_{\gamma\to\infty}\frac{a_1 b_{\text{PSK}}}{\gamma} = \lim_{\gamma\to\infty}\frac{a_2 b_{\text{PSK}}}{\gamma} = \frac{(\alpha+\beta)\sin^2\left(\frac{\pi}{M}\right)}{(\alpha+\beta+1)(a\alpha+\beta+d)}. \tag{40}$$

Hence,

$$P_{\text{MPSK}}^{\alpha,\beta,\infty}(\gamma) = \gamma^{-N}\frac{(L+1)(L+2)\cdots(N)}{\pi L^{N-L}} \times \left(\frac{(a\alpha+\beta+d)(\alpha+\beta+1)}{\alpha+\beta}\right)^{N} \sin^{-2N}\left(\frac{\pi}{M}\right) \times \int_0^{\frac{M-1}{M}\pi} \sin^{2L}\theta \left(\frac{\beta\sin^2\left(\frac{\pi}{M}\right)}{\alpha(\alpha+\beta+1)} + \sin^2\theta\right)^{N-L} d\theta. \tag{41}$$

Expanding $\left(\frac{\beta\sin^2\left(\frac{\pi}{M}\right)}{\alpha(\alpha+\beta+1)} + \sin^2\theta\right)^{N-L}$ in a binomial series and using (31) gives the desired result. The asymptotic expression for SEP of MQAM can be derived similarly.

F. Proof of Theorem 2

1) Proof of (22): At the optimal values of α and β, we have $\frac{\partial}{\partial\alpha}U_{\text{MPSK}}^{\alpha,\beta}\Big|_{\alpha_\infty^*,\beta_\infty^*} = 0$ and $\frac{\partial}{\partial\beta}U_{\text{MPSK}}^{\alpha,\beta}\Big|_{\alpha_\infty^*,\beta_\infty^*} = 0$. Hence, from (20), α_∞^* and β_∞^* satisfy the following two equations:

$$\frac{(N-L)\beta(a\alpha+\beta+d)(2\alpha+\beta+1)\sin^2\left(\frac{\pi}{M}\right)}{\alpha^2(\alpha+\beta)(\alpha+\beta+1)} = N\left(1+\frac{\beta\sin^2\left(\frac{\pi}{M}\right)}{\alpha(\alpha+\beta+1)}\right)\left(a+\frac{\beta(a-1)-d}{(\alpha+\beta)^2}\right), \tag{42}$$

and

$$\frac{(N-L)(a\alpha+\beta+d)(\alpha+1)\sin^2\left(\frac{\pi}{M}\right)}{\alpha(\alpha+\beta)(\alpha+\beta+1)} = N\left(1+\frac{\beta\sin^2\left(\frac{\pi}{M}\right)}{\alpha(\alpha+\beta+1)}\right)\left(-1+\frac{\alpha(a-1)+d}{(\alpha+\beta)^2}\right). \tag{43}$$

Taking the ratio of (42) and (43), we get

$$\frac{\beta(2\alpha+\beta+1)}{\alpha(\alpha+1)} = \frac{a(\alpha+\beta)^2+\beta(a-1)-d}{-(\alpha+\beta)^2+\alpha(a-1)+d}. \tag{44}$$

When 1 is added to both sides, it becomes obvious that $\alpha+\beta$ and $\alpha+\beta+1$ are common factors of both sides of the equation, which can be canceled since $\alpha+\beta>0$ for the optimal values. We then get $\beta^2+2\alpha\beta+a\alpha^2-d=0$. This implies (22).

2) Derivation of (23): The partial derivative with respect to α of (20) takes the form

$$\frac{\partial}{\partial\alpha}U_{\text{MPSK}}^{\alpha,\beta} = f(\alpha,\beta)\Bigg[\left(N\alpha\left(\alpha(\alpha+\beta+1)+\beta\sin^2\left(\frac{\pi}{M}\right)\right)\right) \times\left(a(\alpha+\beta)^2+\beta(a-1)-d\right)\Bigg) - \Bigg((\alpha+\beta)(a\alpha+\beta+d) \times(N-L)\beta\sin^2\left(\frac{\pi}{M}\right)(2\alpha+\beta+1)\Bigg)\Bigg]. \tag{45}$$

Here,

$$f(\alpha,\beta) \triangleq \frac{(L+1)(L+2)\cdots(N)}{L^{N-L}}\left(\frac{M-1}{M}\right)\sin^{-2N}\left(\frac{\pi}{M}\right) \times \frac{(a\alpha+\beta+d)^{N-1}(\alpha+\beta+1)^{N-2}}{\alpha^2(\alpha+\beta)^{N+1}}\left(1+\frac{\beta\sin^2\left(\frac{\pi}{M}\right)}{\alpha(\alpha+\beta+1)}\right)^{N-L-1},$$

is positive, for $\alpha>0$ and $\beta\geq 0$. Using the relationship in (22) between the optimal α and β, we can simplify (45) to

$$\frac{\partial}{\partial\alpha}U_{\text{MPSK}}^{\alpha,\beta}\Big|_{\alpha_\infty^*,\beta_\infty^*} = \beta_\infty^*(2\alpha_\infty^*+\beta_\infty^*+1)f(\alpha_\infty^*,\beta_\infty^*) \times(a-1)\sin^2\left(\frac{\pi}{M}\right)g_1(\alpha_\infty^*), \tag{46}$$

where $g_1(.)$ is as defined in (23). Hence, the solution of $\frac{\partial}{\partial\alpha}U_{\text{MPSK}}^{\alpha,\beta}\Big|_{\alpha_\infty^*,\beta_\infty^*} = 0$ is either $\beta_\infty^*=0$ or $g_1(\alpha_\infty^*)=0$. We first characterize the positive roots of $g_1(\alpha)=0$.

3) At Least One Positive Root of $g_1(\alpha)$ Lies in $\left(0,\sqrt{\frac{d(N-L)}{N(a-1)}}\sin\left(\frac{\pi}{M}\right)\right]$: $g_1(\alpha)$ is continuous and differentiable for $0\leq\alpha\leq\sqrt{\frac{d}{a}}$. Furthermore, $g_1(0) = -(\sqrt{d}+1)d\left(\frac{N-L}{a-1}\right) < 0$. Let $\alpha_1 \triangleq \sqrt{\frac{d(N-L)}{N(a-1)}}\sin\left(\frac{\pi}{M}\right)$. Then, $g_1(\alpha_1)\geq 0$ because $\left(\sqrt{d-\alpha_1^2(a-1)}+\alpha_1\right)g_1(\alpha_1) = \frac{adL^2\alpha_1}{N}\left(\frac{N}{L}-\frac{\frac{N}{L}-1}{a-1}\sin^2\left(\frac{\pi}{M}\right)\right)\geq 0$. The last expression is non-negative because $a=\lceil\frac{N}{L}\rceil\geq\frac{N}{L}$, which implies that $\frac{\frac{N}{L}}{a}\geq\frac{\frac{N}{L}-1}{a-1}\geq\frac{\frac{N}{L}-1}{a-1}\sin^2\left(\frac{\pi}{M}\right)$. *Note that $g_1(\alpha_1)=0$ only when L divides N and $\sin\left(\frac{\pi}{M}\right)=1$, which occurs only when $M=2$ (BPSK). In this case, α_1 is itself the root of $g_1(\alpha)$ and equals ε_∞^*.* For all other cases, which will be the focus of the steps below, it follows from the Intermediate value theorem [32] that at least one root of $g_1(\alpha)$ lies in the interval $[0,\alpha_1]$. Since $\alpha_1<\sqrt{\frac{d(N-L)}{N(a-1)}}\leq\varepsilon_\infty^*$ (see Lemma 1), this root of $g_1(\alpha)$ is necessarily different from ε_∞^*.

4) Positive Root of $g_1(\alpha)$ is Unique: $g_1(\alpha)$ cannot have an even number of real roots in the interval $\left[0,\sqrt{\frac{d}{a}}\right]$ because $g_1(0)<0$, $g_1(\alpha_1)>0$, and $g_1\left(\sqrt{\frac{d}{a}}\right)>0$. If $g_1(\alpha)$ has three or more real roots in the interval $\left[0,\sqrt{\frac{d}{a}}\right]$, then $\frac{\partial}{\partial\alpha}g_1(\alpha)$ *must have more than one real root in the same interval*. We will now rule out this possibility. From (23),

$$\frac{\partial}{\partial\alpha}g_1(\alpha) = \frac{g_2(\alpha)}{\sqrt{d-\alpha^2(a-1)}} + 2\alpha\left(N\csc^2\left(\frac{\pi}{M}\right)-L\right), \tag{47}$$

where

$$g_2(\alpha) \triangleq \alpha^3\left(-3N(a-1)\csc^2\left(\frac{\pi}{M}\right)\right) + \alpha^2(-2L(a-1)) + \alpha\left(2Nd\csc^2\left(\frac{\pi}{M}\right)+d(N-L)\right)+Ld. \tag{48}$$

Notice that the second term, $2\alpha\left(N\csc^2\left(\frac{\pi}{M}\right)-L\right)$, in the right hand side (RHS) of (47) is positive for $\alpha>0$. We make the following three observations about $g_2(\alpha)$: (i) $g_2(0)=Ld>0$. (ii) $g_2(\alpha)$ has only one positive real root. This follows from Descartes' rule of signs [32] since $g_2(\alpha)$ has one sign change. And, (iii) $g_2(\alpha)$ is concave for $\alpha\geq 0$

since $\frac{\partial^2}{\partial\alpha^2}g_2(\alpha) = -2(a-1)\left(9N\alpha\csc^2\left(\frac{\pi}{M}\right)+2L\right) < 0$, for $\alpha \geq 0$.

Let α_p be the positive root of $g_2(\alpha)$. If $\alpha_p > \sqrt{\frac{d}{a}}$, then, $\frac{\partial}{\partial\alpha}g_1(\alpha) > 0$ for all $\alpha \in \left[0, \sqrt{\frac{d}{a}}\right]$ since each of the terms in the RHS of (47) is positive. This implies that $\frac{\partial}{\partial\alpha}g_1(\alpha)$ has no real root in the interval $\left[0, \sqrt{\frac{d}{a}}\right]$, in which case we are done.

Else, let $\alpha_p \leq \sqrt{\frac{d}{a}}$. Then, the following two observations hold: (i) $\frac{\partial}{\partial\alpha}g_1(\alpha) > 0$ for $\alpha \in [0, \alpha_p]$. This is because the concavity of $g_2(\alpha)$ implies that the first term in the RHS of (47) is positive in this interval. And, (ii) in the interval $\left[\alpha_p, \sqrt{\frac{d}{a}}\right]$, $\frac{\partial}{\partial\alpha}g_1(\alpha)$ is concave, *i.e.*, $\frac{\partial^3}{\partial\alpha^3}g_1(\alpha) < 0$. This follows from the reasoning below.

We have

$$\frac{\partial^3}{\partial\alpha^3}g_1(\alpha) = \frac{\frac{\partial^2}{\partial\alpha^2}g_2(\alpha)}{\sqrt{d-\alpha^2(a-1)}} + \frac{2\alpha(a-1)\frac{\partial}{\partial\alpha}g_2(\alpha)}{(d-\alpha^2(a-1))^{3/2}} + \frac{(a-1)(d+2\alpha^2(a-1))g_2(\alpha)}{(d-\alpha^2(a-1))^{5/2}}. \quad (49)$$

The first term in the RHS of (49) is negative since $g_2(\alpha)$ is concave. The second and third terms are also negative because $g_2(\alpha)$, which is concave for $\alpha > 0$, must be both negative and monotonically decreasing in $\left[\alpha_p, \sqrt{\frac{d}{a}}\right]$. The above two observations together imply that $\frac{\partial}{\partial\alpha}g_1(\alpha)$ has at most one real root in the interval $\left[0, \sqrt{\frac{d}{a}}\right]$.

Irrespective of α_p, we have shown that $\frac{\partial}{\partial\alpha}g_1(\alpha)$ cannot have more that one real root in the interval $\left[0, \sqrt{\frac{d}{a}}\right]$. Hence, there is only one unique positive root of $g_1(\alpha)$.

5) $\alpha^*_\infty < \varepsilon^*_\infty$ (i.e., $\beta^*_\infty \neq 0$): The results thus far have established that $\frac{\partial}{\partial\alpha}U^{\alpha,\beta}_{\text{MPSK}} = 0$ occurs at exactly two values of α: one value lies between 0 and $\alpha_1 < \varepsilon^*_\infty$ and the other one equals ε^*_∞. Since $f(0,\beta) > 0$, it follows from (45) that $\frac{\partial}{\partial\alpha}U^{\alpha,\beta}_{\text{MPSK}}\Big|_{\alpha=0} < 0$. Thus, $\frac{\partial}{\partial\alpha}U^{\alpha,\beta}_{\text{MPSK}}$ increases from a negative value at $\alpha = 0$ to zero at exactly one point in the interval $(0, \alpha_1)$ and then decreases back to zero at $\alpha = \varepsilon^*_\infty$. Hence, $\frac{\partial^2}{\partial\alpha^2}U^{\alpha,\beta}_{\text{MPSK}}$ must be negative at $\alpha = \varepsilon^*_\infty$, which implies that it is a maxima that is not of interest to us. Furthermore, this also proves that the optimum solution must lie in $(0, \alpha_1)$ and is unique.

At the same time, we know that the SEP of the proposed method equals that of the conventional AS training method when $\alpha = \varepsilon^*_\infty$ and $\beta = 0$. Since the optimal value of α is different from ε^*_∞, this implies that the optimal SEP of the proposed method is strictly lower than the conventional method. In fact, the proof above shows that any feasible value of $\beta > 0$ (and the corresponding α) will lead to an SEP upper bound that is lower than that for the conventional method for large γ. A similar behavior is observed in the non-asymptotic regime as well.

G. Proof of Corollary 1

We will now show that $g_1\left(\sqrt{\frac{d(N-L)}{N(a-1)}}\sin\left(\frac{\pi}{M}\right)\right) \to 0$ as $M \to \infty$. This will imply that $\sqrt{\frac{d(N-L)}{N(a-1)}}\sin\left(\frac{\pi}{M}\right)$ is a root of $g_1(\alpha)$, and, hence, must be α^*_∞.

From (23), after simplification, we get

$$g_1\left(\sqrt{\frac{d(N-L)}{N(a-1)}}\sin\left(\frac{\pi}{M}\right)\right) = Ld\sqrt{\frac{N-L}{N(a-1)}}\sin^2\left(\frac{\pi}{M}\right) \times\left(-\sqrt{\frac{(N-L)}{N(a-1)}} + \sqrt{\cot^2\left(\frac{\pi}{M}\right)+\frac{L}{N}}\right). \quad (50)$$

As $M \to \infty$, we know that $\sin^2\left(\frac{\pi}{M}\right) \to 0$. Therefore, $g_1\left(\sqrt{\frac{d(N-L)}{N(a-1)}}\sin\left(\frac{\pi}{M}\right)\right) \to 0$.

REFERENCES

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

[2] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Commun. Mag.*, pp. 68-73, Oct. 2004.

[3] N. B. Mehta and A. F. Molisch, "Antenna selection in MIMO systems," in *MIMO System Technol. Wireless Commun.* (G. Tsulos, editor), ch. 6. CRC Press, 2006.

[4] "Draft amendment to wireless LAN media access control (MAC) and physical layer (PHY) specifications: enhancements for higher throughput," Tech. Rep. P802.11n/D0.04, IEEE, Mar. 2006.

[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] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: general model, design criteria, and signal processing," *IEEE Signal Process. Mag.*, pp. 12-25, Nov. 2004.

[8] A. F. Molisch, M. Z. Win, and J. H. Winters, "Performance of reduced-complexity transmit/receive-diversity systems," in *Proc. PIMRC*, pp. 739-742, Oct. 2002.

[9] 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.

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

[11] T. Gucluoglu and E. Panayirci, "Performance of transmit and receive antenna selection in the presence of channel estimation errors," *IEEE Commun. Lett.*, vol. 12, pp. 371-373, May 2008.

[12] P. Polydorou and P. Ho, "Error performance of MPSK with diversity combining in non-uniform Rayleigh fading and non-ideal channel estimation," in *Proc. VTC (Spring)*, pp. 627-631, May 2000.

[13] W. Xie, S. Liu, D. Yoon, and J.-W. Chong, "Impacts of Gaussian error and Doppler spread on the performance of MIMO systems with antenna selection," in *Proc. Int. Conf. Wireless. Commun., Netw., Mobile Comput.*, pp. 1-4, 2006.

[14] K. Zhang and Z. Niu, "Adaptive receive antenna selection for orthogonal space-time block codes with channel estimation errors with antenna selection," in *Proc. Globecom*, pp. 3314-3318, 2005.

[15] V. Kristem, N. B. Mehta, and A. F. Molisch, "Optimal weighted antenna selection for imperfect channel knowledge from training," *IEEE Trans. Commun.*, vol. 58, pp. 2023-2034, July 2010.

[16] L. Xiao and X. Dong, "Error performance of selection combining and switched combining systems in Rayleigh fading channels with imperfect channel estimation," *IEEE Trans. Veh. Technol.*, vol. 54, pp. 2054-2065, Nov. 2005.

[17] R. Annavajjala and L. B. Milstein, "Performance analysis of optimum and suboptimum selection diversity schemes on Rayleigh fading channels with imperfect channel estimates," *IEEE Trans. Veh. Technol.*, vol. 56, pp. 1119-1130, May 2007.

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

[19] 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.
[20] 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.
[21] 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. ICSCN*, pp. 405-410, Jan. 2008.
[22] 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.
[23] A. F. Molisch, *Wireless Communications*. Wiley-IEEE Press, 2005.
[24] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, vol. 1. Prentice Hall Signal Processing Series, 1st edition, 1993.
[25] X. N. Zeng and A. Ghrayeb, "Performance bounds for combined channel coding and space-time block coding with receive antenna selection," *IEEE Trans. Veh. Technol.*, vol. 55, pp. 1441-1446, July 2006.
[26] M. Simon and M.-S. Alouini, *Digital Communication over Fading Channels*. Wiley-Interscience, 2nd edition, 2005.
[27] M.-S. Alouini and A. Goldsmith, "A unified approach for calculating error rates of linearly modulated signals over generalized fading channels," *IEEE Trans. Commun.*, vol. 47, pp. 1324-1334, Sep. 1999.
[28] L. S. Gradshteyn and L. M. Ryzhik, *Tables of Integrals, Series and Products*. Academic Press, 2000.
[29] M. K. Simon and D. Divsalar, "Some new twists to problems involving the Gaussian probability integral," *IEEE Trans. Commun.*, vol. 46, pp. 200-210, Feb. 1998.
[30] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. McGraw Hill, 3rd edition, 1991.
[31] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, 9th edition, 1972.
[32] H. S. Hall and S. R. Knight, *Higher Algebra*. Macmillan, 4th edition, 1891.

Vinod Kristem received his Bachelor of Technology degree in Electronics and Communications Engineering from the National Institute of Technology (NIT), Warangal in 2007. He received his Master of Engineering degree in Telecommunications from Indian Institute of Science, Bangalore, India in 2009. Since then, he has been with Beceem Communications Pvt Ltd, Bangalore, India, working on channel estimation and physical layer measurements for WiMAX and LTE. His research interests include the design and analysis of algorithms for wireless communication networks, MIMO systems, and cooperative communications.

Neelesh B. Mehta (S'98-M'01-SM'06) received his Bachelor of Technology degree in Electronics and Communications Engineering 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 in 1997 and 2001, respectively. He was a visiting graduate student researcher at Stanford University in 1999 and 2000. He is now an Assistant Professor at the Dept. of Electrical Communication Engineering, Indian Institute of Science (IISc), Bangalore, India. Until 2002, he was a research scientist in the Wireless Systems Research group in AT&T Laboratories, Middletown, NJ. In 2002-2003, he was a Staff Scientist at Broadcom Corp., Matawan, NJ, and was involved in GPRS/EDGE cellular handset development. From 2003-2007, he was a Principal Member of Technical Staff at the Mitsubishi Electric Research Laboratories, Cambridge, MA, USA.

His research includes work on link adaptation, multiple access protocols, WCDMA downlinks, system-level performance analysis of cellular systems, MIMO and antenna selection, and cooperative communications. He was also actively involved in radio access network physical layer (RAN1) standardization activities in 3GPP. He has served on several TPCs. He was a tutorial co-chair for SPCOM 2010, and was a TPC co-chair for WISARD 2010 and WISARD 2011, the Transmission technologies track of VTC 2009 (Fall), and the Frontiers of Networking and Communications symposium of Chinacom 2008. He is an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and an executive committee member of the IEEE Bangalore Section and the Bangalore chapter of the IEEE Signal Processing Society.

Andreas F. Molisch (S'89, M'95, SM'00, F'05) received the Dipl. Ing., Dr. techn., and habilitation degrees from the Technical University Vienna (Austria) in 1990, 1994, and 1999, respectively. From 1991 to 2000, he was with the TU Vienna, becoming an associate professor there in 1999. From 2000-2002, he was with the Wireless Systems Research Department at AT&T (Bell) Laboratories Research in Middletown, NJ. From 2002-2008, he was with Mitsubishi Electric Research Labs, Cambridge, MA, USA, most recently as Distinguished Member of Technical Staff and Chief Wireless Standards Architect. Concurrently he was also Professor and Chairholder for radio systems at Lund University, Sweden. Since 2009, he is Professor of Electrical Engineering at the University of Southern California, Los Angeles, CA, USA.

Dr. Molisch has done research in the areas of SAW filters, radiative transfer in atomic vapors, atomic line filters, smart antennas, and wideband systems. His current research interests are measurement and modeling of mobile radio channels, UWB, cooperative communications, and MIMO systems. Dr. Molisch has authored, co-authored or edited four books (among them the textbook *Wireless Communications*, Wiley-IEEE Press), eleven book chapters, some 130 journal papers, and numerous conference contributions, as well as more than 70 patents.

Dr. Molisch is an Area editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS for Antennas and Propagation and co-editor of special issues of several journals. He has been member of numerous TPCs, vice-chair of the TPCs of VTC 2005 spring and VTC 2010 spring, general chair of ICUWB 2006, TPC co-chair of the wireless symposium of Globecomm 2007, TPC chair of Chinacom2007, general chair of Chinacom 2008, TPC co-chair of the Commun. Theory Workshop 2009, tutorial co-chair of WCNC 2009, and workshop co-chair at ICC 2010. He has participated in the European research initiatives "COST 231," "COST 259," and "COST273," where he was chairman of the MIMO channel working group, he was chairman of the IEEE 802.15.4a channel model standardization group. From 2005-2008, he was also chairman of Commission C (signals and systems) of URSI (International Union of Radio Scientists), and since 2009, he is the Chair of the Radio Communications Committee of the IEEE Communications Society. Dr. Molisch is a Fellow of the IEEE, a Fellow of the IET, an IEEE Distinguished Lecturer, and recipient of several awards.