This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2009 proceedings

Optimal Weighted Antenna Selection For Imperfect Channel Knowledge From Training

Vinod Kristem
Student Member, IEEE

Neelesh B. Mehta
Senior Member, IEEE

Andreas F. Molisch
Fellow, IEEE

Abstract*—Receive antenna selection (AS) reduces the hardware complexity of multi-antenna receivers by dynamically connecting an *instantaneously best* antenna element to the available radio frequency (RF) chain. Due to the hardware constraints, the channels at various antenna elements have to be sounded *sequentially* to obtain estimates that are required for selecting the "best" antenna and for coherently demodulating data. Consequently, the channel state information at different antennas is outdated by different amounts. We show that, for this reason, simply selecting the antenna with the highest estimated channel gain is not optimum. Rather, the channel estimates of different antennas should be weighted differently, depending on the training scheme. We derive closed-form expressions for the symbol error probability (SEP) of AS for MPSK and MQAM in time-varying Rayleigh fading channels for arbitrary selection weights, and validate them with simulations. We then derive an explicit formula for the optimal selection weights that minimize the SEP. We find that when selection weights are not used, the SEP need not improve as the number of antenna elements increases, which is in contrast to the ideal channel estimation case. However, the optimal selection weights remedy this situation and significantly improve performance.

I. Introduction

Antenna selection (AS) is a popular technique to reduce the hardware costs at the transmitter or receiver of a wireless link [1]–[6]. It uses fewer radio frequency (RF) chains than the actual number of antenna elements, and only processes signals from a dynamically selected subset of antennas. This is advantageous since antenna elements are typically cheap and easy to implement, while the RF chains are expensive. Consequently, many next generation wireless communications standards such as IEEE 802.11n, 3GPP long term evolution (LTE), and the IEEE 802.16m Advanced WiMax have standardized or are standardizing AS at the transmitter, receiver, or both. In this paper, we concentrate on receive AS, and, in particular, the case that a single antenna element is selected for down conversion. While a receiver can have more RF chains, the model we consider is practically relevant and analytically insightful as it achieves the same full diversity order as one with more RF chains under perfect channel state information (CSI).

While the topic of receive AS has been explored extensively in the literature, most papers assume perfect CSI at the receiver. Imperfect CSI can lead to inaccurate selection and imperfect data decoding, both of which increase the symbol error probability (SEP) or the packet error rate (PER). In practice, the CSI needs to be acquired using a pilot-based training scheme.

The low hardware complexity, which is a key motivator for AS, imposes unique constraints on how training gets done for AS: given the limited number of RF chains, only one antenna can be activated at any instant. Consequently, the transmitter needs to send pilot(s) multiple times to enable the receiver to *sequentially* receive pilots with different antennas and estimate their corresponding links to the transmitter. The receive antenna is then selected based on these estimates. Depending on the system design, the pilots can be several milliseconds apart [7]. Thus, the CSI at the receiver is imperfect not just because of noise in the channel estimates but also because of training delays. Even more importantly, *the CSI at different antenna elements is outdated by different amounts*.

In this paper, we analyze and optimize the performance of AS over time-varying Rayleigh fading channels given a practical training model. We show that imperfect CSI has a significant impact on the AS performance, and argue that the selection criterion should account for the training delays (amount of outdatedness) encountered in any practical AS system. The selection criterion we propose uses weighted versions of the channel estimates to select the best antenna.

We derive general closed-form expressions for the symbol error probability (SEP) of MPSK and MQAM as a function of the selection weights and the antenna sounding pattern. While receive AS with imperfect channel estimates has been explored earlier, most papers consider only estimation errors due to noise [8]–[10]. As we shall see, training delays coupled with noisy estimates lead to an error floor in the SEP, which does not occur when just noisy (or perfect) estimates are considered. While [11], [12] consider outdated channel estimates, they do not model the unequal outdatedness of the CSI of the antennas. While [20] considers unequal outdatedness, it selects the antenna with the highest estimated channel gain.

The paper is organized as follows. The system model is developed in Sec. II, followed by SEP analysis and optimization in Sec. III. The results and conclusions follow in Sec. IV and Sec. V, respectively.

II. Model

Consider a system with one transmit antenna, N receive antennas, and one RF chain at the receiver. Let $h_k(t)$ denote the frequency-flat channel between the transmitter and the

V. Kristem and N. B. Mehta are with the Electrical Communication Engineering Dept. in the Indian Institute of Science (IISc), Bangalore, India. A. F. Molisch is with the University of Southern California, Los Angeles, CA, USA.
Emails: vinod.kristem@gmail.com, nbmehta@ece.iisc.ernet.in, andreas.molisch@ieee.org

k^{th} receive antenna at time t. It is modeled as a circularly symmetric complex Gaussian random variable (RV) with unit variance. Furthermore, 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 [13].

A. Channel Estimation

We consider a transmission format in which multiple pilot symbols precede multiple symbols of the data packet, as shown in Fig. 1. The transmitter transmits a pilot symbol p_p (of duration T_s) to each receive antenna sequentially so that all N channels can be estimated and the optimum antenna selected to receive the data symbols block. The k^{th} receive antenna is estimated at time T_k. Two consecutive pilot symbols are separated in time by a duration T_p. Note that the order in which antennas are trained does not matter since the channel gains of different antennas are i. i. d.

Pilot based channel estimation is imperfect due to:

1. Noise-induced channel estimation errors: The pilot signal received by the k^{th} antenna is

$$r_k(T_k) = p_p h_k(T_k) + n_k(T_k),$$

where the noise $n_k(t)$ is an ergodic stationary circularly symmetric complex Gaussian process with zero mean and power N_0 that is independent of $h_k(t)$. Therefore, the channel estimate for the k^{th} receive antenna is

$$\hat{h}_k(T_k) = \frac{p_p^* r_k(T_k)}{|p_p|^2} = h_k(T_k) + e_k, \tag{1}$$

where the noise-induced channel estimation error $e_k = \frac{n_k(T_k)}{|p_p|^2}$ has a variance $\sigma_e^2 = \frac{N_0}{E_p}$. Here, E_p is the pilot symbol energy.

2. Outdated channel estimates: Due to the time-varying nature of the wireless links, the N channels will have changed by the time data transmission starts. Specifically, the channel for receive antenna i at time $t+\delta$ can be written in terms of the channel at time t as [14]

$$h_i(t+\delta) = \rho_i(\delta) h_i(t) + \sqrt{1 - |\rho_i(\delta)|^2}\, n_i'(t+\delta), \tag{2}$$

where $\rho_i(\delta)$ is the channel correlation coefficient. The variation $n_i'(t+\delta)$, $i = 1, 2, \ldots, N$, is a circularly symmetric complex Gaussian RV with unit variance that is independent of $h_i(t)$. The channel correlation coefficient depends on the time difference δ as well as the Doppler spectrum (which, in turn, depends on the velocity, angular spectrum, and antenna pattern of the mobile station [13]). Our derivations in Sec. III are valid for arbitrary Doppler spectra; for the simulations in Sec. IV, we use the classical Jakes spectrum [15] in which

$$\rho_i(\delta) = J_0(2\pi f_d \delta), \tag{3}$$

where $J_0(.)$ is the zeroth order Bessel function of the first kind [17] and f_d is the maximum Doppler frequency.[1]

[1]The regressive model uses the simplifying assumption that the channel realizations at different times can be computed based only on the correlation with the channel at time $t = 0$, and not as a realization of a stochastic process with a continuous correlation function. The approximation is good so long as $2\pi f_d \delta$ is smaller than 2.4, the first root of $J_0(.)$.

B. Weighted Antenna Selection

The standard selection criterion is to pick the antenna with the highest (estimated) channel gain. However, this is not optimal when the CSI of different antennas is outdated by different amounts – it is possible that the antenna with the highest channel gain could have severely outdated CSI and should not be selected. We thus propose selecting the antenna based on *weighted* channel gain estimates as follows:

$$[\hat{1}] = \underset{1 \le k \le N}{\operatorname{argmax}}\ w_k \left|\hat{h}_k\right|^2. \tag{4}$$

Antenna $[\hat{1}]$ is used for receiving entire data packet.

C. Data Reception

The pilots are followed by D data symbols, each of duration T_s and average energy E_s. When the i^{th} data symbol, s_i, is transmitted, the signal received by antenna $[\hat{1}]$ at time t_i, after matched filtering, is given by

$$y_{[\hat{1}]}(t_i) = h_{[\hat{1}]}(t_i) s_i + n_{[\hat{1}]}(t_i). \tag{5}$$

The data symbols are equi-probable and are derived from either the MPSK or MQAM constellations.

III. SEP Analysis and Optimization

A. SEP for Given Antenna Selection Weights

We now analyze the SEP for an MPSK or MQAM symbol transmitted at time t_i for receive AS with imperfect and outdated CSI. Henceforth, we simplify our notation as follows: we denote $\hat{h}_k(T_k)$ by $\hat{h}_k$, $n_k'(t)$ by n_k', $\rho_i(t_j - T_i)$ by $\rho_i^{(j)}$, and $n_k(t)$ by n_k. $\mathbf{E}[A]$ and $\mathbf{var}[A]$ shall denote the expectation and variance of event A, respectively. And, $\mathbf{E}[A|B]$ and $\mathbf{var}[A|B]$ shall denote the conditional expectation and variance of A given B, respectively.

The imperfect channel estimates that are used for selection are also used for data decoding. Therefore, the decision variable, $\mathcal{D}$, for the signal received by antenna $[\hat{1}]$ is:

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

Using (1) and (2), we can write the channel at time t_i in terms of its estimate. Hence,

$$\mathcal{D} = \hat{h}_{[\hat{1}]}^* \left(\rho_{[\hat{1}]}^{(i)} \left(\hat{h}_{[\hat{1}]} - e_{[\hat{1}]} \right) s_i + \sqrt{1 - \left|\rho_{[\hat{1}]}^{(i)}\right|^2}\, n_{[\hat{1}]}' s_i + n_{[\hat{1}]} \right). \tag{6}$$

Thus, the decision variables for symbols transmitted at different times will be different, and so will their error probabilities.

We first state the following Lemma about the statistics of $\mathcal{D}$, which shall be useful in the rest of the paper.

Lemma *1:* Conditioned on $\{\hat{h}_l\}_{l=1}^N$ and s_i, $\mathcal{D}$ is a complex Gaussian RV with conditional mean and variance given by

$$\mathbf{E}\left[\mathcal{D} \,\middle|\, \{\hat{h}_l\}_{l=1}^N, s_i\right] = \left|\hat{h}_{[\hat{1}]}\right|^2 \rho_{[\hat{1}]}^{(i)} s_i q^2, \tag{7}$$

$$\mathbf{var}\left[\mathcal{D} \,\middle|\, \{\hat{h}_l\}_{l=1}^N, s_i\right] = \left(1 - \left|\rho_{[\hat{1}]}^{(i)}\right|^2\right) |s_i|^2 \left|\hat{h}_{[\hat{1}]}\right|^2 + \left|\hat{h}_{[\hat{1}]}\right|^2 \left|\rho_{[\hat{1}]}^{(i)}\right|^2 |s_i|^2 \sigma_e^2 q^2 + \left|\hat{h}_{[\hat{1}]}\right|^2 N_0, \tag{8}$$

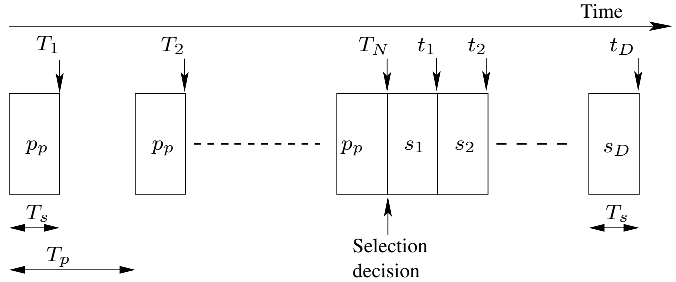

Fig. 1. Training for antenna selection

where $q^2 \triangleq 1/(1+\sigma_e^2)$.

Proof: The proof uses standard results on Gaussian RVs, and is omitted due to space constraints. ■

We are now ready to derive the SEPs for MPSK and MQAM in the following two theorems. Let $\gamma \triangleq \frac{E_s}{N_0}$ (average SNR per branch) and $\varepsilon \triangleq \frac{E_p}{E_s}$.

Theorem *1:* With training delays and noisy channel estimates, the SEP for the i^{th} MPSK symbol received at time t_i is[2]

$$P_i^{\text{MPSK}}(\gamma) = \frac{1}{\pi}\sum_{k=1}^{N}\sum_{r=0}^{N-1}\sum_{\substack{l_0,\ldots,l_r=1\\ l_0=1,l_1\neq\cdots\neq l_r\neq k}}^{N} \frac{(-1)^r}{r!\left(1+\sum_{j=1}^{r}\frac{w_k}{w_{l_j}}\right)}\left[\frac{M-1}{M}\pi - \frac{\tan^{-1}\left(\sqrt{\alpha_k^{(i)}(\gamma,w_{l_1},\ldots,w_{l_r})}\tan\left(\frac{M-1}{M}\pi\right)\right)}{\sqrt{\alpha_k^{(i)}(\gamma,w_{l_1},\ldots,w_{l_r})}}\right], \quad (9)$$

where

$$\alpha_k^{(i)}(\gamma,w_{l_1},\ldots,w_{l_r}) \triangleq 1+\frac{\left(1+\sum_{j=1}^{r}\frac{w_k}{w_{l_j}}\right)}{\varepsilon\left|\rho_k^{(i)}\right|^2\sin^2(\frac{\pi}{M})}\left(\varepsilon\left(1-\left|\rho_k^{(i)}\right|^2\right)+\frac{1+\varepsilon}{\gamma}+\frac{1}{\gamma^2}\right). \quad (10)$$

Proof: The proof is relegated to Appendix A. ■

Theorem *2:* With training delays and noisy channel estimates, the SEP for the i^{th} MQAM symbol received at time t_i is given by equation (12) (next page), where

$$\beta_k^{(i)}(\gamma,w_{l_1},\ldots,w_{l_r}) \triangleq 1+\frac{\left(1+\sum_{j=1}^{r}\frac{w_k}{w_{l_j}}\right)}{\varepsilon\left|\rho_k^{(i)}\right|^2\left(\frac{3}{2(M-1)}\right)}\left(\varepsilon\left(1-\left|\rho_k^{(i)}\right|^2\right)+\frac{1+\varepsilon}{\gamma}+\frac{1}{\gamma^2}\right). \quad (11)$$

Proof: The proof is given in Appendix B. ■

[2]For notational compactness, $\sum_{\substack{l_0,\ldots,l_r=1\\ l_0=1,l_1\neq\cdots\neq l_r\neq k}}^{N}$ will henceforth denote $\sum_{l_0=1}^{1}\sum_{\substack{l_1=1\\(l_1\neq k)}}^{N}\sum_{\substack{l_2=1\\(l_2\neq k,l_2\neq l_1)}}^{N}\cdots\sum_{\substack{l_r=1\\(l_r\neq k,l_r\neq l_1,\cdots,l_r\neq l_{r-1})}}^{N}$.

B. Optimal Selection Weights

The optimal weights $\left\{w_k^{\text{opt}}\right\}_{k=1}^{N}$ that minimize the SEP for MPSK and MQAM at an SNR γ are then given as follows.

Theorem *3:* For $1 \le k \le N$, the optimal selection weights that minimize the SEP of an MQAM or MPSK symbol transmitted at time t_i are given by

$$w_{k,i}^{\text{opt}}(\gamma) = \frac{\left|\rho_k^{(i)}\right|^2}{\left(\varepsilon\left(1-\left|\rho_k^{(i)}\right|^2\right)+\frac{1+\varepsilon}{\gamma}+\frac{1}{\gamma^2}\right)}. \quad (13)$$

Proof: The proof is given in the Appendix C. ■

The optimal weights depend on the channel correlation coefficients, which further depend on the geometry of the antennas and the scattering environment. Since these change on a much slower time scale than small-scale fading, the correlation coefficients can be accurately estimated and used [16]. Notice that the optimal weights are inversely proportional to the training delays. When the training delays are the same, *i.e.*, $\rho_k^{(i)} = \rho$, the optimal weights do not depend on k and i, which is equivalent to selection without weighting. For large training delays, *i.e.*, $\rho_k^{(i)} \ll 1$, the optimal weights, after removing common factors, simplify to $w_{k,i}^{\text{opt}}(\gamma) = \left|\rho_k^{(i)}\right|^2$. Note that the result above implies that the optimal selection weights – and, hence, the selected antenna – can be different for data symbols transmitted at different times. When decoding a packet consisting of multiple symbols, an additional practical constraint that the receive antenna is the same for all symbols may need to be imposed. However, this is beyond the scope of this paper.

C. Asymptotic Behavior of SEP with Optimal Weights

We now consider the asymptotic behavior of our SEP expressions in (9) and (12) as the SNR, γ, increases. Let $P_{i,\text{asm}}^{\text{MPSK}} \triangleq \lim_{\gamma\to\infty} P_i^{\text{MPSK}}(\gamma)$, $P_{i,\text{asm}}^{\text{MQAM}} \triangleq \lim_{\gamma\to\infty} P_i^{\text{MQAM}}(\gamma)$, $\lim_{\gamma\to\infty}\alpha_k^{(i)}(\gamma,w_{l_1},\ldots,w_{l_r}) \triangleq \alpha_{k,\text{asm}}^{(i)}(w_{l_1},\ldots,w_{l_r})$, and $\lim_{\gamma\to\infty}\beta_k^{(i)}(\gamma,w_{l_1},\ldots,w_{l_r}) \triangleq \beta_{k,\text{asm}}^{(i)}(w_{l_1},\ldots,w_{l_r})$.

When training delays are absent, $\rho_k^{(i)} = 1$, for all k and i. From (10) and (11), we can see that $\alpha_{k,\text{asm}}^{(i)}(.) = 1$ and $\beta_{k,\text{asm}}^{(i)}(.) = 1$, for all k and i. Using these asymptotic values in (9) and (12), we get $P_{i,\text{asm}}^{\text{MPSK}} \equiv 0$ and $P_{i,\text{asm}}^{\text{MQAM}} \equiv 0$.

For non-zero training delays, we have $\rho_k^{(i)} < 1$. From (13), we can see that the optimal weights (after removing common factors) equal $\lim_{\gamma\to\infty} w_{k,i}^{\text{opt}}(\gamma) = \left|\rho_k^{(i)}\right|^2\Big/\left(1-\left|\rho_k^{(i)}\right|^2\right)$, $\alpha_{k,\text{asm}}^{(i)}(.) = 1-\csc^2\left(\frac{\pi}{M}\right)\left(r+1-\frac{1}{\left|\rho_k^{(i)}\right|^2}-\sum_{j=1}^{r}\frac{1}{\left|\rho_{l_j}^{(i)}\right|^2}\right)$, $\beta_{k,\text{asm}}^{(i)}(.) = 1-\frac{2(M-1)}{3}\left(r+1-\frac{1}{\left|\rho_k^{(i)}\right|^2}-\sum_{j=1}^{r}\frac{1}{\left|\rho_{l_j}^{(i)}\right|^2}\right)$.

Upon substituting the asymptotic values in the SEP formulae for MPSK and MQAM, it can be shown that $P_{i,\text{asm}}^{\text{MPSK}}$ and $P_{i,\text{asm}}^{\text{MQAM}}$ are not identically 0. Hence, an irreducible error floor exists at high SNR, and depends on the correlations $\{\rho_k^{(i)}\}_{k=1}^{N}$.

$$P_i^{\mathrm{MQAM}}(\gamma) = 2\left(1-\frac{1}{\sqrt{M}}\right)\sum_{k=1}^{N}\sum_{r=0}^{N-1}\sum_{\substack{l_0,\ldots,l_r=1\\ l_0=1, l_1\neq\cdots\neq l_r\neq k}}^{N}\left(1-\frac{1}{\sqrt{\beta_k^{(i)}(\gamma, w_{l_1},\ldots,w_{l_r})}}\right)\frac{(-1)^r}{r!\left(1+\sum_{j=1}^{r}\frac{w_k}{w_{l_j}}\right)}$$
$$-\left(1-\frac{1}{\sqrt{M}}\right)^2\sum_{k=1}^{N}\sum_{r=0}^{N-1}\sum_{\substack{l_0,\ldots,l_r=1\\ l_0=1, l_1\neq\cdots\neq l_r\neq k}}^{N}\frac{(-1)^r}{r!\left(1+\sum_{j=1}^{r}\frac{w_k}{w_{l_j}}\right)}\left(1-\frac{4\tan^{-1}\left(\sqrt{\beta_k^{(i)}(\gamma, w_{l_1},\ldots,w_{l_r})}\right)}{\pi\sqrt{\beta_k^{(i)}(\gamma, w_{l_1},\ldots,w_{l_r})}}\right). \quad (12)$$

IV. Simulations

We now present graphically the results derived in Sec. III and study the effect of N, f_dT_p, and $\{w_k\}_{k=1}^N$ on the SEP. We also compare these with Monte Carlo simulations (with 10^4 samples generated for each SNR ($\gamma \triangleq E_s/N_0$)), and use the simulator of [19] to generate the time-varying Rayleigh channels. From (3), the correlation values for $k = 1, 2, \ldots, N$, $i = 1, 2, \ldots, D$, equal $\rho_k^{(i)} = J_0\left(2\pi f_d((N-k)T_p + iT_s)\right)$. The figures are plotted for $T_p = 10T_s$ and $E_p = E_s$. Unless mentioned otherwise, SEP of the first data symbol is plotted.

Figures 2 and 3 plot the SEP as a function of the SNR for MPSK and MQAM, respectively, with $N = 4$ antennas. One can see that the SEP always decreases to 0 as the SNR increases when $f_dT_p = 0$, even with noisy estimates. On the other hand, an error floor exists when $f_dT_p > 0$, which increases as f_dT_p increases. Also shown is the effect of different selection weights on the SEP. For $f_dT_p \approx 0$, $w_k^{opt} \approx 1$, and, hence, all the six curves coincide. For large f_dT_p, $w_{k,i}^{\mathrm{opt}}(\gamma) = |\rho_k|^2$ performs almost as well as optimal weighting, which follows from Sec.III-B. Notice the excellent match between analytical and simulation results.[3] Given the unequal outdatedness of the CSI, it is instructive to get an idea of how often each antenna is selected after optimal weighting. At an SNR of 12 dB and $f_dT_p = 0.06$, antenna 1 ($\rho_1 = 0.6866$) and antenna 4 ($\rho_4 = 0.9996$) get selected 1.4% and 58.8% of the time, respectively.

Figure 4 compares the SEP of MPSK for $N = 2$, 4, and 8 receive antennas as a function of SNR at $f_dT_p = 0.01$ for the no-weighting and optimal weighting selection schemes. In the no-weighting scheme, we can see that increasing the number of receive antennas can lead to worse performance. This can be understood as follows. As the training delays increase, selection becomes more inaccurate (and unequally so) due to more outdated estimates. However, the optimal weights, which account for this, remedy this problem. Furthermore, they also reduce the error floors by one to two orders of magnitude.

An alternative view is presented in Figure 5, which compares the MPSK SEP of $N = 2$, 4, and 8 antennas at an SNR of 10 dB as a function of the Doppler spread. As in Figure 4, T_p is fixed. For no-weighting, $N = 8$ outperforms others when $0 \le f_dT_p \le 0.017$. However, for $0.017 \le f_dT_p \le 0.044$, and $f_dT_p \ge 0.044$, $N = 4$ and $N = 2$, respectively, are the best choices. In contrast, for optimal weighting, $N = 8$ is always best choice.[4] However, for higher Doppler spreads, the performance difference between smaller and larger number of antennas decreases.

We also compared the SEP of 8PSK data symbols transmitted at different times. At $f_dT_p = 0.06$ and 12 dB SNR, the SEP of the 1st, 5th, and 10th data symbols calculated from (9) are $0.1741, 0.2155$, and 0.2910 respectively. The difference is more significant at higher values of SNR. A figure is not shown for want of space.

V. Conclusions

In this paper we analyzed receive AS with channel estimates that are affected both by noise and the time variations of the fading channel. For training (pilot) structures that are typical for AS systems, the channel estimates of different antenna elements are outdated by different amounts. This has important consequences for the selection criteria and the overall system performance. Our most important results and insights are the following: (i) We propose a new selection scheme that weights the channel estimates before antenna selection; (ii) We provide closed-form equations for the SEP in such a system; and (iii) We find that selecting the antenna with the highest estimated channel gain is not optimum, and leads to a high error floor. But, the optimal selection weights do drastically reduce the error floor. Increasing the number of antenna elements can worsen performance when unweighted AS is used, while, optimal selection weighting does remedy this effect.

Our work thus provides useful insights into optimum selection in the presence of channel estimation errors. Future work will include an analysis of block or packet error rates, and the analysis of different distributed pilot patterns.

Appendix

A. Proof of Theorem 1

Using Lemma 1 and [18, eq.(40)], the SEP for MPSK, conditioned on $\{\hat{h}_l\}_{l=1}^N$, is given by

$$P_i\left(\mathrm{Err}\,\middle|\,\{\hat{h}_l\}_{l=1}^N\right) = \frac{1}{\pi}\int_0^{\frac{M-1}{M}\pi}\exp\left(\frac{-\left|\hat{h}_{[\hat{1}]}\right|^2 b_{[\hat{1}]}^{(i)}}{\sin^2\theta}\right)d\theta, \quad (14)$$

[3]The small mismatch between analytical and simulation results for MQAM is explained in Appendix B.

[4]Note that the conclusions would be different if one assumed a fixed overall energy budget, so that using more pilot tones reduces the available energy for the payload data. In that case, a very large number of antennas is not optimum even when weighted selection is used.

where $b_k^{(i)} \triangleq \frac{E_s\left|\rho_k^{(i)}\right|^2 q^4 \sin^2\left(\frac{\pi}{M}\right)}{\left|\rho_k^{(i)}\right|^2 E_s\sigma_e^2 q^2+N_0+\left(1-\left|\rho_k^{(i)}\right|^2\right)E_s}$. This simplifies to $b_k^{(i)} = \frac{\varepsilon\left|\rho_k^{(i)}\right|^2 \sin^2\left(\frac{\pi}{M}\right)}{\left(1+\frac{1}{\gamma\varepsilon}\right)\left(\varepsilon\left(1-\left|\rho_k^{(i)}\right|^2\right)+\frac{1+\varepsilon}{\gamma}+\frac{1}{\gamma^2}\right)}$.

From (1), the probability density function, $f(x)$, and the cumulative distribution function, $F(x)$, of $\left|\hat{h}_k\right|^2$ are given by $f(x) = \frac{1}{1+\sigma_e^2}\exp\left(\frac{-x}{1+\sigma_e^2}\right)$ and $F(x) = 1-\exp\left(\frac{-x}{1+\sigma_e^2}\right)$. Therefore, the SEP averaged over $\{\hat{h}_i\}_{i=1}^N$ equals

$$P_i\left(\text{Err}\right) = \frac{1}{\pi}\sum_{k=1}^{N}\int_0^\infty \int_0^{\frac{M-1}{M}\pi}\exp\left(\frac{-xb_k^{(i)}}{\sin^2\theta}\right) \times f(x)\prod_{\substack{l=1\\l\neq k}}^{N}F\left(\frac{w_k x}{w_l}\right)d\theta\, dx.$$

Expanding $\prod_{\substack{l=1\\l\neq k}}^{N}F\left(\frac{w_k x}{w_l}\right)$ and simplifying,

$$P_i\left(\text{Err}\right) = \frac{1}{\pi\left(1+\sigma_e^2\right)}\sum_{k=1}^{N}\sum_{r=0}^{N-1}\sum_{\substack{l_0,\dots,l_r=1\\l_0=1,l_1\neq\cdots\neq l_r\neq k}}^{N}\frac{(-1)^r}{r!} \times\int_0^{\frac{M-1}{M}\pi}\left(\frac{b_k^{(i)}}{\sin^2\theta}+\frac{1+\sum_{j=1}^{r}\frac{w_k}{w_{l_j}}}{1+\sigma_e^2}\right)^{-1}d\theta. \quad (15)$$

Using the following identity, which follows from [17, 2.562], and simplifying further yields the desired expression.

$$\text{For } a,b>0, \int_0^\zeta\left(\frac{a}{\sin^2\theta}+b\right)^{-1}d\theta = \frac{1}{b}\left[\zeta-\sqrt{\frac{a}{a+b}}\tan^{-1}\left(\sqrt{\frac{a+b}{a}}\tan\zeta\right)\right]. \quad (16)$$

B. Proof of Theorem 2

Using Lemma 1 and [18, eq. (48)], the SEP for MQAM, conditioned on $\{\hat{h}_i\}_{i=1}^N$, is given by[5]

$$P_i\left(\text{Err}|\{\hat{h}_i\}_{i=1}^N\right) = \frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)\int_0^{\frac{\pi}{2}}\exp\left(\frac{-\left|\hat{h}_{[\hat{1}]}\right|^2 c_{[\hat{1}]}^{(i)}}{\sin^2\theta}\right)d\theta - \frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)^2\int_0^{\frac{\pi}{4}}\exp\left(\frac{-\left|\hat{h}_{[\hat{1}]}\right|^2 c_{[\hat{1}]}^{(i)}}{\sin^2\theta}\right)d\theta, \quad (17)$$

where $c_k^{(i)} = \frac{\varepsilon\left|\rho_k^{(i)}\right|^2\left(\frac{3}{2(M-1)}\right)}{\left(1+\frac{1}{\gamma\varepsilon}\right)\left(\varepsilon\left(1-\left|\rho_k^{(i)}\right|^2\right)+\frac{1+\varepsilon}{\gamma}+\frac{1}{\gamma^2}\right)}$. Using steps similar to Theorem 1, we can derive $P_i^{\text{MQAM}}(\gamma)$.

[5]This expression implicitly assumes that the variance of $\mathcal{D}$ is the same for all MQAM symbols. However, with imperfect estimation, this is no longer the case, as can be seen from (8). However, this approximation is quite good and is commonly used [8].

C. Proof of Theorem 3

Let $\mathbf{w} = \{w_k\}_{k=1}^N$. Let $\widetilde{\mathbf{w}}_i = \{\widetilde{w}_{i,k}\}_{k=1}^N$ denote the optimal $\mathbf{w}$ for the i^{th} data symbol as per (13). We outline below the key steps that show that $\frac{\partial}{\partial w_p}P_i^{\text{MPSK}}(\gamma)\Big|_{\mathbf{w}=\widetilde{\mathbf{w}}_i} = 0$.

Appropriately clustering the terms in the integrand of (15), and differentiating with respect to w_p, we get

$$\frac{\partial}{\partial w_p}P_i\left(\text{Err}\right)\Big|_{\mathbf{w}=\widetilde{\mathbf{w}}_i} = -\sum_{r=1}^{N-1}\sum_{\substack{l_1,\dots,l_r=1\\l_1\neq\cdots\neq l_r\neq p}}^{N}\frac{(-1)^r}{r!}\left(\frac{b_p^{(i)}}{\sin^2\theta}+\frac{1+\sum_{j=1}^{r}\frac{\widetilde{w}_{i,p}}{\widetilde{w}_{i,l_j}}}{1+\sigma_e^2}\right)^{-2} \times\left(\frac{\sum_{j=1}^{r}\frac{1}{\widetilde{w}_{i,l_j}}}{1+\sigma_e^2}\right) + \sum_{\substack{k=1\\k\neq p}}^{N}\sum_{r=1}^{N-1}\sum_{q=1}^{r}\sum_{\substack{l_1,\dots,l_{q-1},l_{q+1},\dots,l_r=1\\l_1\neq\cdots\neq l_{q-1}\neq p\neq l_{q+1}\cdots\neq l_r\neq k}}^{N}\frac{(-1)^r}{r!} \times\left(\frac{b_k^{(i)}}{\sin^2\theta}+\frac{1+\frac{\widetilde{w}_{i,k}}{\widetilde{w}_{i,p}}+\sum_{\substack{j=1\\j\neq q}}^{r}\frac{\widetilde{w}_{i,k}}{\widetilde{w}_{i,l_j}}}{1+\sigma_e^2}\right)^{-2}\frac{\widetilde{w}_{i,k}}{\widetilde{w}_{i,p}^2\left(1+\sigma_e^2\right)}. \quad (18)$$

Using $b_k^{(i)}\widetilde{w}_{i,p} = b_p^{(i)}\widetilde{w}_{i,k}$, the second term in (18) becomes

$$\sum_{r=1}^{N-1}\sum_{q=1}^{r}\sum_{\substack{k=1\\k\neq p}}^{N}\sum_{\substack{l_1,\dots,l_{q-1},l_{q+1},\dots,l_r=1\\l_1\neq\cdots\neq l_{q-1}\neq p\neq l_{q+1}\cdots\neq l_r\neq k}}^{N}\frac{(-1)^r}{r!\widetilde{w}_{i,k}(1+\sigma_e^2)} \times\left(\frac{b_p^{(i)}}{\sin^2\theta}+\frac{1+\frac{\widetilde{w}_{i,p}}{\widetilde{w}_{i,k}}+\sum_{\substack{j=1\\j\neq q}}^{r}\frac{\widetilde{w}_{i,p}}{\widetilde{w}_{i,l_j}}}{1+\sigma_e^2}\right)^{-2}.$$

Replacing the variable k with l_q and clubbing summations, the above expression simplifies to

$$\sum_{r=1}^{N-1}\sum_{q=1}^{r}\sum_{\substack{l_1,\dots,l_r=1\\l_1\neq\cdots\neq l_r\neq p}}^{N}\frac{(-1)^r}{r!\widetilde{w}_{i,l_q}(1+\sigma_e^2)}\left(\frac{b_p^{(i)}}{\sin^2\theta}+\frac{1+\sum_{j=1}^{r}\frac{\widetilde{w}_{i,p}}{\widetilde{w}_{i,l_j}}}{1+\sigma_e^2}\right)^{-2}.$$

Upon interchanging the second and third summations, the above expression can be shown to be exactly equal to the first term in (18). Hence, $\frac{\partial}{\partial w_p}P_i^{\text{MPSK}}(\gamma)\Big|_{\mathbf{w}=\widetilde{\mathbf{w}}_i} = 0$. Similarly, one can also show that $\frac{\partial}{\partial w_p}P_i^{\text{MQAM}}(\gamma)\Big|_{\mathbf{w}=\widetilde{\mathbf{w}}_i} = 0$.

We now give an alternate proof for the optimal selection weights for MQAM. The proof uses physical arguments and is therefore quite succinct. A similar approach can also be used for MPSK. Let antenna k be selected and used for data reception. Equation (17) can be rewritten as

$$P_i\left(\text{Err}|\hat{h}_k\right) = \frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)\int_0^{\frac{\pi}{2}}\xi(\theta)\exp\left(\frac{-\left|\hat{h}_k\right|^2 c_k^{(i)}}{\sin^2\theta}\right)d\theta,$$

where $\xi(\theta)=1/\sqrt{M}$ for $0\le\theta<\pi/4$, and $\xi(\theta)=1$, for $\pi/4\le\theta\le\pi/2$. From this special form of the SEP, it follows that the

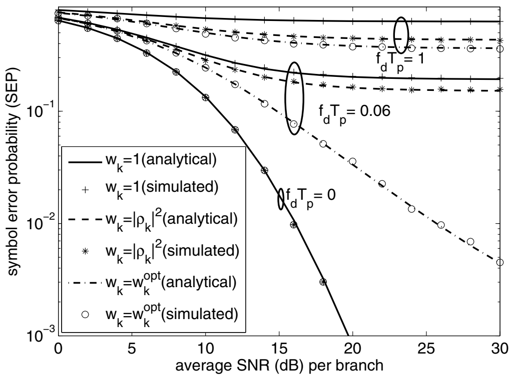

Fig. 2. Effect of normalized Doppler spread and weights (8PSK and $N = 4$).

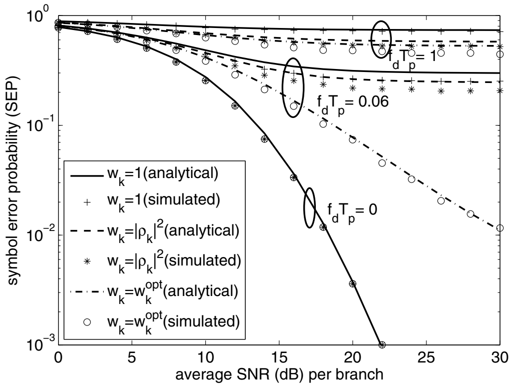

Fig. 3. Effect of normalized Doppler spread and weights (16QAM and N=4).

minimum SEP obtained by selecting the best antenna equals $\frac{4}{\pi}\left(1-\frac{1}{\sqrt{M}}\right)\int_0^{\frac{\pi}{2}}\xi(\theta)\min_{k=1,\ldots,N}\exp\left(\frac{-\left|\hat{h}_k\right|^2 c_k^{(i)}}{\sin^2\theta}\right)d\theta$.

Thus, the optimal antenna to use for data reception is the one that maximizes $\arg\max_k\left(\left|\hat{h}_k\right|^2 c_k^{(i)}\right)$.

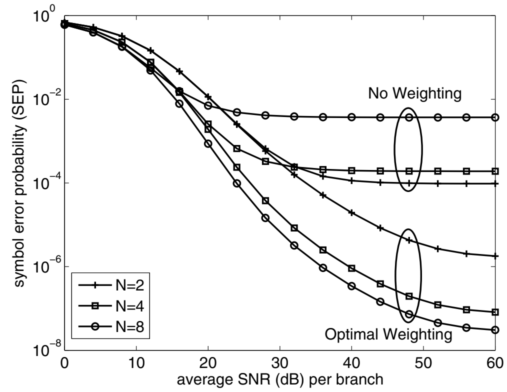

Fig. 4. Effect of number of receive antennas and weights for fixed Doppler spread and T_p (8PSK and $f_dT_p = 0.01$).

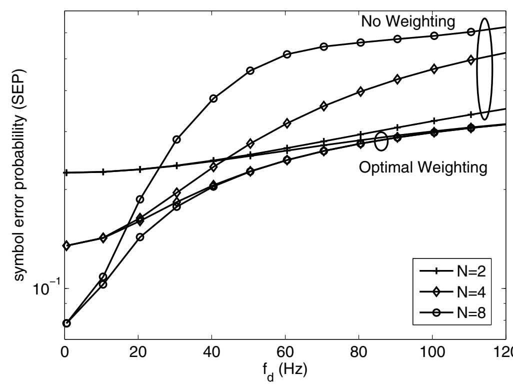

Fig. 5. Effect of number of receive antennas, Doppler spread, and weights for fixed T_p (8PSK, T_p = 1ms, and 10 dB SNR).

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.*, vol. 42, pp. 68–73, Oct. 2004.

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

[4] 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.

[5] A. Gorokhov, D. Gore, and A. Paulraj, "Receive antenna selection for MIMO flat fading channels: Theory and algorithms," *IEEE Trans. Inform. Theory.*, vol. 49, pp. 2687–2696, 2003.

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

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

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

[9] P. Theofilakos and A. G. Kanatas, "Robustness of receive antenna subarray formation to hardware and signal non-idealities," in *Proc. VTC (Spring)*, pp. 324–328, 2007.

[10] 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.

[11] S. Han and C. Yang, "Performance analysis of MRT and transmit antenna selection with feedback delay and channel estimation error," in *Proc. WCNC*, pp. 1135–1139, 2007.

[12] 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. Intl. Conf. Wireless. Commun., Networking and Mobile Computing*, pp. 1–4, 2006.

[13] A. F. Molisch, *Wireless Communications*. Wiley-IEEE Press, 2005.

[14] D. V. Marathe and S. Bhashyam, "Power control for multi-antenna Gaussian channels with delayed feedback," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, pp. 1598–1602, 2005.

[15] W. C. Jakes, *Microwave Mobile Communications*. IEEE Press, 1993.

[16] W. J. Szajnowski, "Estimating the correlation coefficient of highly correlated clutter," Elec. Let., vol. 13, pp. 318–319, May. 1977.

[17] L. S. Gradshteyn and L. M. Ryzhik, *Tables of Integrals, Series and Products*. Academic Press, 2000.

[18] 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, Sept. 1999.

[19] P. Dent, G. E. Bottemley, and T. Croft, "Jakes fading model revisited," *Electron. Lett.*, vol. 29, pp. 1162–1163, June 1933.

[20] V. Kristem and N. B. Mehta, "Receive Antenna Selection with Imperfect Channel Knowledge From Training," *National. Conf. Commun.*, pp. 37–41, 2009.