Pilot-to-data power ratio for maximizing the capacity of MIMO-OFDM

Taeyoon Kim, Student Member, IEEE, and Jeffrey G. Andrews, Member, IEEE,
Submitted to IEEE Trans. on Communications

Abstract

Orthogonal frequency division multiplexing (OFDM) can be used with a multiple-input multiple-output (MIMO) system to improve communication quality and capacity. Pilot-symbol-aided or decision-directed channel estimation must be used to track the variations of the channel. However, while pilot symbols facilitate channel estimation, they reduce the transmit energy for data symbols per OFDM symbol under a fixed total transmit power constraint. We analyze the effects of pilot-symbol-aided channel estimation on the lower bound of the system capacity in MIMO-OFDM systems with three different types of pilot patterns: independent, scattered, and orthogonal. We derive the optimal pilot-to-data power ratio (PDR) for each of these patterns and quantify the resulting capacity. The analysis shows that independent and orthogonal pilot patterns have the same capacity, outperforming the scattered pilot pattern. This paper implies that implementing the optimal PDR in an actual MIMO-OFDM system should prove relatively straightforward, since there is a fairly broad range of PDR values over which near optimal capacity is achieved. In this range – which calls for more power to be allocated to pilot symbols than data symbols – the capacity is increased by 10-20% by simply allocating an appropriate portion of power to the pilot and data subcarriers.

Index Terms

Pilot-to-data power ratio (PDR), Ergodic capacity, multiple-input multiple-output (MIMO), Orthogonal frequency division multiplexing (OFDM), pilot-symbol-aided channel estimation.

The authors are with the Wireless Networking and Communications Group in the Department of Electrical and Computer Engineering at The University of Texas at Austin, Austin, TX 78712 USA. E-mail: {tykim, jandrews}@ece.utexas.edu. This research was supported in part by National Instruments and an equipment donation by Intel.
Orthogonal frequency division multiplexing (OFDM) is one of the most promising physical layer technologies for high data rate wireless communications due to its robustness to frequency selective fading, high spectral efficiency, and low computational complexity. OFDM can be used in conjunction with a multiple-input multiple-output (MIMO) transceiver to increase the diversity gain and/or the system capacity [1], [2], [3], [4]. Because the OFDM system effectively provides numerous parallel narrowband channels, MIMO-OFDM is considered a key technology in emerging high-data rate systems such as 4G, IEEE 802.16, and IEEE 802.11n [5], [6], [7].

In MIMO-OFDM systems, channel state information (CSI) is essential at the receiver in order to coherently detect the received signal and to perform diversity combining or spatial interference suppression. In order to attain instantaneous CSI at the receiver, pilot-symbol-aided or decision-directed channel estimation must be used to track the variations of the frequency selective fading channel [8], [9]. Pilot symbols facilitate channel estimation, but in addition to consuming bandwidth, they reduce the transmitted energy for data symbols per OFDM symbol under a fixed total transmit power condition. This suggests a tradeoff between the power allowed to data symbols and the accuracy of the channel estimation in MIMO-OFDM systems when the total transmit power is fixed. The pilot symbols can be transmitted in various ways in MIMO-OFDM systems including independent (time-multiplexed), scattered (frequency-multiplexed), or orthogonal (code-multiplexed) pilot patterns [10]. In this paper, the pilot-to-data power ratio (PDR) for MIMO-OFDM systems with these three different pilot patterns is optimized from a Shannon capacity point of view. The results herein allow system designers to have a well-justified basis for allocating more power to the pilot symbols, despite the fact that they don’t carry actual information themselves.

To date, there have been numerous previous studies on the problem of optimizing pilot signals for wireless communication systems, but none that optimize the PDR for MIMO-OFDM capacity with comparing various pilot schemes. Optimizing the PDR for DS/CDMA systems has been considered in [11] for the single user case. In multiuser cases, the optimal PDR for the DS/CDMA
uplink with multiuser detection has been examined in [12]. Both works optimized the PDR of DS/CDMA systems with respect to the bit error rate (BER) performance. For OFDM systems with a single-input single-output (SISO), optimizing training tones for minimizing the mean-square error (MSE) in channel estimates has been proposed in [13], and optimal placement and power of pilot signals for maximizing capacity has been analyzed in [14], [15], and the BER-minimizing pilots for OFDM systems was derived in [16].

For MIMO systems, there have been comparatively fewer studies, but some recent research has begun to develop guidelines for appropriate pilot signal design. For single-carrier MIMO systems, the effects of pilot-assisted channel estimation were analyzed in [17], and an optimal training signal was developed in [18]. Both papers optimize the training signal to maximize the capacity. In [19], the pilot signal is optimized for block transmissions with MIMO systems to maximize the capacity. The problem of optimizing training tones for MIMO-OFDM systems has been addressed in [20]. The metric for optimization in [20] was the MSE of the channel estimation. In this paper, we optimize the PDR of MIMO-OFDM systems by directly maximizing the capacity and show the analysis of the capacity lower bound with the effects of the channel estimation. The analysis shows that the correlation between different channel links in the estimated channel can be removed by using three different pilot patterns which are termed as independent, scattered, and orthogonal, and an optimal PDR is derived for each of those three cases. Our analysis can be viewed as a generalization of prior results on SISO-OFDM [14], as our results reduce to that special case when only one antenna is present at both the transmitter and receiver.

The rest of this paper is organized as follows. Section II describes the system model of the analyzed MIMO-OFDM system, and introduces the required notation. The principal results of the paper are in Section III and IV, which find expressions for the MMSE of the channel estimation error, and then uses these results to find an information-theoretic capacity lower bound for each of the three different pilot patterns. In doing so, we find an optimal PDR for each case. In Section V, we demonstrate and interpret the analysis results by simulation. Finally, concluding remarks are given in Section VI.
II. System Model

The system under consideration is depicted in Fig. 1, which shows a spatial multiplexing MIMO-OFDM system with $M_T$ transmit antennas and $M_R$ receive antennas. The number of subcarriers is $K$, and the number of nonzero taps of the impulse response is $L$ for each channel. Suppose the received signal in the frequency domain at time index $n$ is denoted by the $M_R K \times 1$ vector $Y(n) = [y_T^T(n) \cdots y_{M_r}^T(n)]^T$ where $y_j(n) = [y_j(n, 1) \cdots y_j(n, K)]^T$ and $y_j(n, k)$ is the $n^{th}$ received signal on the $k^{th}$ subcarrier of the $j^{th}$ receive antenna. The received signal at the $j^{th}$ receive antenna is given by

$$y_j(n) = \sum_{i=1}^{M_t} x_i W_k h_{i,j} + n_j = X W h_j + n_j,$$

where $X = [x_1 \cdots x_{M_t}]$ is the transmit symbol matrix where $x_i = mdiag(x_i(k))_{k=1}^{K}$, the function $mdiag(\cdot)$ is a modified diagonal function defined as follows,

$$mdiag(A_i)^{M-1}_{i=0} = \begin{bmatrix} A_0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & A_{M-1} \end{bmatrix},$$

where $A_i$ can be a series of matrices or scalars, $h_j = [h_{1,j} \cdots h_{M_t,j}]^T$ is the channel impulse response matrix at $j^{th}$ receive antenna where $h_{i,j} = [h_{i,j}(1) \cdots h_{i,j}(L)]^T$ and $h_{i,j}(l)$ denotes the channel response of the $l^{th}$ path from the $i^{th}$ transmit antenna to the $j^{th}$ receive antenna, and $n_j = [n_j(1), \ldots, n_j(K)]^T$ is additive white Gaussian noise (AWGN) which is zero mean Gaussian with covariance $\sigma_n^2 I$. $W = I_{M_t} \otimes W_k$ where $\otimes$ is the Kronecker product, $I_{M_t}$ is an identity matrix of size $(M_t \times M_t)$, $W_k$ is the DFT matrix of size $(K \times L)$, and $[W_k]_{k,l} = \exp(-j2\pi(k-1)(l-1)/K)$. It is assumed that the taps of $h_{i,j}$ are independent and identically distributed (i.i.d.) complex Gaussian random variables with zero mean and variance equal to $\frac{1}{L}$.

At each transmit antenna, data and pilot symbols are modulated on a set of subcarriers by the OFDM modulator. $D$ and $P$ are the subcarrier index sets for the data and pilot signals, respectively. Let $x_{d,i}$ and $x_{p,i}$ be the matrix of data symbols and pilot symbols transmitted from
the \(i^{th}\) transmit antenna, in other words, \(x_{d,i} = \text{mdiag} (x_i(k))_{k \in \mathcal{D}}\) and \(x_{p,i} = \text{mdiag} (x_i(k))_{k \in \mathcal{P}}\). The data and pilot signals have different transmit power, i.e. \(x_{d,i} x_{d,i}^H = \sigma_d^2 I\) and \(x_{p,i} x_{p,i}^H = \sigma_p^2 I\). The PDR is the power ratio between pilot and data symbol tone defined as

\[
\eta = \frac{\sigma_p^2}{\sigma_d^2}.
\]

Finally, let the DFT matrix for data subchannels and pilot subchannels be \(W_d \in \mathbb{C}^{D \times L}\) and \(W_p \in \mathbb{C}^{P \times L}\) where \(D\) and \(P\) are the number of data symbols and pilot symbols in one OFDM symbol.

### III. CHANNEL ESTIMATION WITH DIFFERENT PILOT PATTERNS

Now that the system model has been established, in this section, we analyze the channel estimation error according to three different pilot patterns to see how the channel estimation error affects the capacity of a MIMO-OFDM system.

#### A. MMSE Channel Estimation

We assume that the channel is estimated by the MMSE estimator using pilot symbols as in [18]. The received signal for channel estimation is composed of the signals on subcarriers reserved for pilot symbols, and can be expressed as follows,

\[
Y_p = \begin{bmatrix} y_{p,1}^T & y_{p,2}^T & \cdots & y_{p,M_r}^T \end{bmatrix}^T = Zh + N_p,
\]

where \(y_{p,j} = \text{vec} (y_j(q))_{q \in \mathcal{P}}\), \(Z = I_{M_r} \otimes X_p F_p\), \(F_p = I_{M_r} \otimes W_p\), \(X_p = [x_{p,1} \cdots x_{p,M_r}]\), \(h = [h_1^T \cdots h_{M_r}^T]^T\), \(N_p = [n_{p,1}^T \cdots n_{p,M_r}^T]^T\), and \(n_{p,j} = \text{vec} (n_j(k))_{k \in \mathcal{P}}\).

With \(E[hh^H] = \frac{1}{L} I\), the MMSE channel estimates can be expressed as,

\[
\hat{h}_p = Z^H (ZZ^H + L\sigma_n^2 I)^{-1} Y_p
\]

\[
= (I_{M_r} \otimes F_p^H X_p^H) \left[ (I_{M_r} \otimes X_p F_p F_p^H X_p^H) + L\sigma_n^2 I \right]^{-1} Y_p
\]

by using \((A \otimes B)(C \otimes D) = AB \otimes CD\). The channel estimation error is

\[
\tilde{h}_p = h - \hat{h}_p
\]
From (6), the estimated channel can be expressed as,

$$
\hat{h}_p = \left( I_{M_r} \otimes F_p^H X_p^H \right) \cdot \left( I_{M_r} \otimes X_p F_p F_p^H X_p^H + L\sigma_n^2 I \right)^{-1} \cdot \left( I_{M_r} \otimes X_p F_p h + N_p \right).
$$

(8)

Since

$$
(mdiag(A_n)_{n=1}^N)^{-1} = mdiag(A_n^{-1})_{n=1}^N,
$$

(9)

$$
mdiag(A_n)_{n=1}^N \cdot mdiag(B_n)_{n=1}^N = mdiag(A_n B_n)_{n=1}^N,
$$

(10)

and $$(I \otimes A) \cdot (I \otimes B) = I \otimes (A \cdot B)$$, $\hat{h}_p^{(s)}$ can be written as $mdiag(\hat{h}_p^{(s)})_{j=1}^{M_r}$ where $s = 1, 2$. Thus, we can write $\hat{h}_p$ as follows,

$$
\hat{h}_p = mdiag(\hat{h}_p^{(1)})_{j=1}^{M_r} \cdot mdiag(\hat{h}_p^{(2)})_{j=1}^{M_r} \cdot \hat{h}_p^{(3)},
$$

(11)

where $\hat{h}_p^{(3)} = \left[ \begin{array}{c} \hat{h}_p^{(3)}(1)^T \\ \vdots \\ \hat{h}_p^{(3)}(M_r)^T \end{array} \right]$, $\hat{h}_p^{(3)} = X_p F_p h_j + n_{p,j}$, and $n_{p,j}$ is the AWGN noise in frequency domain at the $j^{th}$ receive antenna. Then, the estimated channel at the $j^{th}$ receive antenna is

$$
\hat{h}_{p,j} = (F_p^H X_p^H) \cdot (\hat{h}_p^{(2)})^{-1} \cdot (X_p F_p^H X_p^H + L\sigma_n^2 I),
$$

(12)

where $\hat{h}_p^{(2)}$ is a matrix of size $(P \times P)$ and can be written as

$$
\hat{h}_p^{(2)} = \sum_{s=1}^{M_t} x_{p,s} W_p W_p^H x_{p,s}^H + L\sigma_n^2 I.
$$

(13)

Thus, the estimated channel between $i^{th}$ transmit antenna and $j^{th}$ receive antenna using the MMSE channel estimator can be expressed as

$$
\hat{h}_{i,j} = U_i(X_p F_p h_j + n_{p,j}) = \sum_{m=1}^{M_t} U_i x_{p,m} W_p h_{m,j} + U_i n_{p,j},
$$

(14)

where $U_i = W_p^H x_{p,i}^H \left( \sum_{s=1}^{M_t} x_{p,s} W_p W_p^H x_{p,s}^H + L\sigma_n^2 I \right)^{-1}$. As shown in (14), the estimated channel between $i^{th}$ transmit antenna and $j^{th}$ receive antenna has components related not only to $h_{i,j}$ but also $h_{m,j}$ $(m \neq i)$. Moreover, it has the term $\sum_{s=1}^{M_t} x_{p,s} W_p W_p^H x_{p,s}^H$ which sums the power of pilot symbols from all transmit antennas, and decreases the performance of channel estimation. This is because the pilot symbols on the same subcarriers are transmitted from all transmit antennas at the same time.
B. Pilot patterns

Since it was seen in the previous subsection that interference from other antennas and subcarriers caused a large amount of interference, we consider three pilot transmit schemes that eliminate this interference in order to allow better channel estimation.

1) Independent pilot pattern

In this scheme, only one antenna can transmit pilot symbols over one OFDM symbol time while all other antennas send zeros on the subcarriers reserved for pilot symbols. This scheme assumes that the channel remains constant over $M_t$ consecutive OFDM symbols. Clearly, by allowing only one antenna to transmit pilot signals, we can guarantee the orthogonality between channels. For example, in the case of $M_t = 2$, the received signals at two consecutive times $n$ are

$$Y(n) = Z_n h + N(n),$$

where $n = 1, 2$, $Z_n = I_2 \otimes X_{id}(n)F_p$, $X_{id}(1) = [x_{p,1} \ 0]$, and $X_{id}(2) = [0 \ x_{p,2}]$. Then, from above and (5), the estimated channel from $i^{th}$ transmit antenna to all receive antennas can be expressed as

$$\hat{h}_{i,M_r} = Z_i^H(Z_iZ_i^H + L\sigma_n^2 I)^{-1}Y(i).$$

Thus, from (14), the estimated channel between $i^{th}$ transmit antenna and $j^{th}$ receive antenna using the independent pilot pattern is

$$\hat{h}_{i,j} = W_p^H x_{p,i}^H (x_{p,i} W_p W_p^H x_{p,i}^H + L\sigma_n^2 I)^{-1}(x_{p,i} W_p h_{i,j} + n_{p,i,j}),$$

where $n_{p,i,j}$ is the noise vector at the $j^{th}$ receive antenna at $i^{th}$ OFDM symbol time. Since $\hat{h}_{i,j} = h_{i,j} + \tilde{h}_{i,j}$ and $E[h_{i,j} h_{k,l}^H] = 0$ unless $i = k$ and $j = l$, it can be easily shown from (17) that $E[\hat{h}_{i,j} \hat{h}_{k,l}^H] = E[\hat{h}_{i,j} \hat{h}_{k,l}^H] = 0$ unless $i = k$ and $j = l$.

From (7), the channel estimation error covariance can be expressed as

$$E[\hat{h}_p \hat{h}_p^H] = E[(h_p - \hat{h}_p)(h_p - \hat{h}_p)^H],$$

$$= \frac{1}{L} \left( I - \sum_{i=1}^{M_t} Z_i^H(Z_iZ_i^H + L\sigma_n^2 I)^{-1}Z_i \right),$$
where
\[
\sum_{i=1}^{M_t} Z_i^H (Z_iZ_i^H + L\sigma_n^2 I)^{-1} Z_i = \sum_{i=1}^{M_t} I_{M_t} \otimes (F_p^H X_{id}(i)(\sigma_p^2 W_p W_p^H + L\sigma_n^2 I)^{-1} X_{id}(i)F_p)
\]
\[
= I_{M_t, M_t} \otimes (\sigma_p^2 W_p^H (\sigma_p^2 W_p W_p^H + L\sigma_n^2 I)^{-1} W_p).
\]

(20)

Thus, by using matrix inversion lemma,
\[
E \left [ \hat{h}_p \hat{h}_p^H \right ] = \frac{1}{L} \left ( I - I_{M_t, M_t} \otimes (\sigma_p^2 W_p^H (\sigma_p^2 W_p W_p^H + L\sigma_n^2 I)^{-1} W_p) \right )
\]
\[
= \frac{1}{L} I_{M_t, M_t} \otimes \left ( I + \frac{\sigma_p^2}{L\sigma_n^2} W_p^H W_p \right )^{-1}.
\]

(21)

One thing to note is that in this independent pilot pattern scheme, we transmit pilot symbols from only one antenna at a time. The saved power from other transmit antennas can be used for data symbols.

2) Scattered pilot pattern

In this scheme, pilot symbols are transmitted on different subcarriers for different antennas, and zeros are transmitted on the subcarriers which are used for the pilot symbols of other transmit antennas. The pilot symbols can be allocated as follows,

\[
S_i = \left [ \begin{array}{cccccc}
\cdots & s_d & p & s_d & s_d & \cdots & s_d & p & s_d & s_d & \cdots \\
\end{array} \right ]
\]

where \( S_i \) is the OFDM signal at \( i \)th transmit antenna in frequency domain, \( p \) is the vector of length \( M_t \) with a \( s_p \) in \( i \)th column and with 0’s elsewhere, \( s_p \) is pilot symbols, and \( s_d \) is data symbols. Thus, each transmit antenna has different subcarrier index set \( P_i \) for pilot symbols. The estimated channel using the scattered pilot pattern is the same as (5), but it has different \( Z \) as follows,

\[
Z_{sc} = I_{M_t} \otimes X_{p(sc)} F_{p(sc)},
\]

where \( F_{p(sc)} = I_{M_t} \otimes W_{p(sc)} \), and \( X_{p(sc)} = \left [ x_{p(sc),1} \cdots x_{p(sc),M_t} \right ] \) where \( x_{p(sc),i} \) is a sparse diagonal matrix with a pilot symbol \( x_i \) occurring periodically at every \( M_t \) position, offset by \( i \) positions and 0’s elsewhere, for example if \( M_t = 4, P = 3, \) and \( i = 2 \), then \( x_{p(sc),2} = \text{diag}(0 \ x_i \ 0 \ 0 \ 0 \ x_i \ 0 \ 0 \ 0 \ x_i \ 0 \ 0 \).
This scheme uses more subcarriers for pilot symbols, but it requires the channel to be constant over just one OFDM symbol, rather than $M_t$ OFDM symbols. Therefore, it is better for high Doppler channel case, i.e. fast fading channel. Alternatively, in slowly fading channels, it might be possible to send several OFDM symbols without pilot symbols, but that case is not specifically considered in this paper. Per OFDM symbol, this scheme requires $M_tP$ subcarriers for pilot symbols at each transmit antenna. Thus, the number of subcarriers for data symbols $D_{sc}$ is $D - (M_t - 1)P$.

From (18) and (22), the channel estimation error covariance for scattered pilot pattern is given by

\[
E \left[ \tilde{h}_p \tilde{h}_p^H \right] = \frac{1}{L} \left( I + \frac{1}{L\sigma^2_n} \left( I_{M_t} \otimes F_{p(sc)}^H X_{p(sc)}^H X_{p(sc)} F_{p(sc)} \right) \right)^{-1}
\]

\[
= \frac{1}{L} \left( I + \frac{1}{L\sigma^2_n} \left( I_{M_t} \otimes F_{p(sc)}^H \text{diag} \left( \sigma^2_p \hat{I}_i \right) \right) \right)^{-1}
\]

\[
= \frac{1}{L} \left( I_{M_t} \otimes \left( I + \frac{\sigma^2_p}{L\sigma^2_n} \text{diag} \left( W_{p(sc)}^H \hat{I}_i W_{p(sc)} \right) \right)^{M_t} \right)^{-1},
\]

where $\hat{I}$ is a sparse diagonal matrix with a 1 occurring periodically at every $M_t$th position, offset by $i$ positions and 0's elsewhere.

3) Orthogonal pilot pattern

This orthogonal pilot pattern is similar to the orthogonal space-time block codes developed by Tarokh et al. [21]. We can construct the orthogonal pilot pattern for $M_t = 2^n$ cases using the Walsh-Hadamard matrix, e.g. for $M_t = 2$,

\[
\begin{bmatrix}
  x_{p,1} & x_{p,1} \\
  x_{p,2} & -x_{p,2}
\end{bmatrix},
\]

where each row corresponds to a unique transmit antenna, and the columns are for unique time instants. It also assumes that the channel is constant over $M_t$ OFDM symbols, which is typically reasonable in a high data rate system. The received channel can be resolved at the receiver by linear processing of the received signals. For example, in the case of $M_t = 2$ and $M_r = 2$, the received signals at two consecutive times $n$ are

\[
Y(n) = Z(n)h + N(n),
\]
where \( Z(n) = I_2 \otimes X_p(n)F_p \), \( X_p(1) = [x_{p,1} \ x_{p,2}] \), and \( X_p(2) = [x_{p,1} - x_{p,2}] \). Then,
\[
\bar{Y}_i = \frac{1}{2}(Y(1) + (-1)^{i-1}Y(2)) = Z_ih + \frac{1}{2}(N(1) + (-1)^{i-1}N(2)),
\] (26)
where \( i = 1, 2 \), and \( Z_i \) is the same as in (15). From above and (5), the estimated channel can be expressed as,
\[
\hat{h} = \hat{h}_{1,M_t} + \hat{h}_{2,M_t},
\] (27)
\[
\hat{h}_{i,M_t} = Z_i^H(Z_iZ_i^H + \frac{L\sigma_n^2}{2}\mathbf{I})^{-1}\bar{Y}_i,
\] (28)
where \( \hat{h}_{i,M_t} \) is the estimated channel from \( \bar{Y}_i \). Thus, the estimated channel between \( i^{th} \) transmit antenna and \( j^{th} \) receive antenna using the orthogonal pilot pattern is
\[
\hat{h}_{i,j} = W_p^Hx_{p,i}^H(x_{p,i}W_pW_p^Hx_{p,i}^H + \frac{L\sigma_n^2}{M_t}\mathbf{I})^{-1}(x_{p,i}W_p\hat{h}_{i,j} + \bar{n}_{p,j}),
\] (29)
where \( \bar{n}_{p,j} \) is the average noise vector at the \( j^{th} \) receive antenna during \( M_t \) symbols times. As shown in (29), by using the orthogonal pilot scheme, we get noise reduction effect. The noise power in the channel estimation error is reduced by \( 1/M_t \) times.

The channel estimation error covariance for the orthogonal pilot pattern is similar to that for independent pilot pattern, resulting in
\[
E[\hat{h}_p\hat{h}_p^H] = \frac{1}{L}I_{M_s,M_t} \otimes \left( I + \frac{M_t\sigma_p^2}{L\sigma_n^2}W_pW_p^H\right)^{-1}.
\] (30)

IV. CAPACITY LOWER BOUND VS. PDR

In this section, we find a lower bound on capacity and show the relationship between the capacity lower bound and the PDR.

A. Capacity Lower Bound

The capacity is the maximum of the mutual information between the known signals and the unknown signal over the distribution of the transmit data signal \( X_d \) as follows,
\[
C = \sup_{p(X_d)} I(Y_p^q, X_p; Y_d; X_d)
\]
\[
= \sup_{p(X_d)} I(Y_d; X_d|Y_p^q, X_p) + I(Y_p^q, X_p; X_d)_{=0}
\]
\[
= \sup_{p(X_d)} I(Y_d; X_d|Y_p, X_p)
\] (31)
where \( p(X_d) \) is the probability distribution of data symbol \( X_d \), and \( Y_d \) is a received signal vector on subcarriers reserved for data symbols.

For the convenience, we redefine the received signal at the \( j^{th} \) receive antenna in the frequency domain as

\[
Y_j = XW_h + n_j = H_j \hat{X} + n_j
\]

where \( H_j = [H_{1,j} \ H_{2,j} \ \cdots \ H_{M_t,j}] \), \( H_{i,j} = \text{diag}(H_{i,j}(1), \cdots, H_{i,j}(K)) = \text{diag}(W_k h_{i,j}) \),

\[\hat{X} = [\hat{x}_1^T \ \hat{x}_2^T \ \cdots \ \hat{x}_{M_t}^T]^T, \quad \text{and} \quad \hat{x}_i = [x_i(1) \ x_i(2) \ \cdots \ x_i(K)]^T.\]

To find the capacity lower bound, we can rewrite the received signal for data symbols in the frequency domain as

\[
Y = \begin{bmatrix}
\hat{H}_1 \hat{X}_d \\
\vdots \\
\hat{H}_{M_t} \hat{X}_d
\end{bmatrix} + \begin{bmatrix}
\hat{H}_1 \hat{X}_d \\
\vdots \\
\hat{H}_{M_t} \hat{X}_d
\end{bmatrix} + \begin{bmatrix}
n_1 \\
\vdots \\
n_{M_r}
\end{bmatrix} = Y_S + Y_E = Y + N
\]

where \( H = \hat{H} + \tilde{H} \), \( \hat{H} \) is the channel estimation error in frequency domain. When \( X_d \sim \mathcal{CN}(0, \sigma_d^2 I) \), the capacity lower bound can be obtained as follows [18],

\[
C_{\text{lowerbound}} = \inf_{p(Y_E)} \sup_{p(X_d)} I\left(Y_d X_d | \hat{H}\right) \\
= \frac{1}{K} \mathbb{E} \left[ \log_2 \det \left( I + \sigma_d^2 R_{Y_E}^{-1} \hat{H} \hat{H}^H \right) \right]
\]

where \( p(Y_E) \) and \( p(X_d) \) are the probability distribution of \( Y_E \) and \( X_d \), respectively, \( R_{Y_E} \) is the autocorrelation matrix of \( Y_E \), and the worst case is the independent Gaussian distribution. The autocorrelation \( R_{Y_E} \) is

\[
R_{Y_E} = \mathbb{E} \left[ \tilde{Y} \tilde{Y}^H \right] + \mathbb{E} \left[ NN^H \right].
\]

From Appendix I, The autocorrelation \( R_{Y_E} \) can be written as

\[
R_{Y_E} = \sigma_d^2 \cdot \text{mdig} \left( \sum_{i=1}^{M_t} \text{diag} \left( W_i \bar{I}_{i,j} \mathbb{E} \left[ \tilde{h}_j \tilde{h}_j^H \right] \bar{I}_{i,j}^H W_i^H \right) \right)_{j=1}^{M_r} + \sigma_n^2 I
\]

Now, we consider the channel estimation error covariances of the three candidate pilot patterns to simplify the autocorrelation \( R_{Y_E} \). We use some restrictions for the placement of pilot symbols
in frequency domain to make it easier to find the optimal PDR in next section. Let us assume that the pilot symbols are placed periodically in frequency band, $P$ is a factor of $K$, and $P \geq L$. Then, the matrix $\mathbf{W}_p^H \mathbf{W}_p$ becomes the identity matrix. The channel estimation error covariance for the three pilot patterns are in (21), (23), and (30).

1) Independent pilot pattern

We can simplify (36) as follows,

$$
\mathbf{I}_{i,j} \mathbb{E} \left[ \mathbf{h}_p^H \mathbf{h}_p^H \right] \mathbf{I}_{i,j}^H = \frac{1}{L} \mathbf{I}_{i,j} \left( \mathbf{I}_{M_r M_t} \otimes \left( \mathbf{I} + \frac{\sigma_p^2 L}{\sigma_n^2} \mathbf{W}_p^H \mathbf{W}_p \right)^{-1} \right) \mathbf{I}_{i,j}^H
$$

$$
= \frac{1}{L} \left( \mathbf{I} + \frac{\sigma_p^2 L}{\sigma_n^2} \mathbf{W}_p^H \mathbf{W}_p \right)^{-1}.
$$

(37)

Thus, from the assumption of pilot placement, the autocorrelation of $\mathbf{Y}_E$ can be expressed as

$$
\mathbf{R}_{\mathbf{Y}_E} = \frac{\sigma_d^2}{L} \text{mdag} \left( \sum_{i=1}^{M_t} \text{diag} \left( \mathbf{W}_d \left( \mathbf{I} + \frac{\gamma_p P}{L} \mathbf{I} \right)^{-1} \mathbf{W}_d^H \right) \right)_{j=1}^{M_r} + \sigma_n^2 \mathbf{I}
$$

$$
= \frac{\sigma_d^2}{L} \cdot \text{mdag} \left( \sum_{i=1}^{M_t} \frac{L}{L + \gamma_p} \text{diag} \left( \mathbf{W}_d \mathbf{W}_d^H \right) \right)_{j=1}^{M_r} + \sigma_n^2 \mathbf{I}
$$

$$
= \left( \frac{L \sigma_d^2 M_t}{\gamma_p P + L} + \sigma_n^2 \right) \mathbf{I}_{DM_r},
$$

(38)

where $\gamma_p = \frac{\sigma_p^2}{\sigma_n^2}$. Thus, from (34), the capacity lower bound for independent pilot pattern is

$$
C_{lb(id)} = \frac{1}{K} \mathbb{E} \left[ \log_2 \det \left( \mathbf{I} + \frac{\sigma_d^2 \mathbf{H} \mathbf{H}^H}{\sigma_d^2 M_t \gamma_p + \sigma_n^2} \right) \right].
$$

(39)

2) Scattered pilot pattern

We can simplify (36) as follows,

$$
\mathbf{I}_{i,j} \mathbb{E} \left[ \mathbf{h}_p^H \mathbf{h}_p^H \right] \mathbf{I}_{i,j}^H = \frac{1}{L} \mathbf{I}_{i,j} \left( \mathbf{I}_{M_r M_t} \otimes \left( \mathbf{I} + \frac{\sigma_p^2}{L \sigma_n^2} \text{mdag} \left( \mathbf{W}_p(\text{sc}) \mathbf{I}_i \mathbf{W}_p(\text{sc}) \right)_{i=1}^{M_t} \right)^{-1} \right) \mathbf{I}_{i,j}^H
$$

$$
= \frac{1}{L} \left( \mathbf{I} + \frac{\sigma_p^2}{L \sigma_n^2} \mathbf{W}_p(\text{sc}) \mathbf{I}_i \mathbf{W}_p(\text{sc}) \right)^{-1}
$$

$$
= \frac{1}{L + \gamma_p} \mathbf{I}_L.
$$

(40)

The autocorrelation of $\mathbf{Y}_E$ is

$$
\mathbf{R}_{\mathbf{Y}_E} = \left( \frac{L \sigma_d^2 M_t}{\gamma_p P + L} + \sigma_n^2 \right) \mathbf{I}_{DM_r}.
$$

(41)
The capacity lower bound for scattered pilot pattern is

\[ C_{lb(sc)} = \frac{1}{K} E \left[ \log_2 \det \left( I + \frac{\sigma_n^2 \hat{H}_{sc} \hat{H}_{sc}^H}{(L \sigma_t^2 M_t P \gamma_p + L + \sigma_n^2) I_{D_{sc} M_r}} \right) \right], \quad (42) \]

where \( \hat{H}_{sc} \) is the estimated channel matrix of size \( (D_{sc} M_r \times D_{sc} M_t) \).

3) Orthogonal pilot pattern

Due to noise averaging, the autocorrelation of \( Y_E \) is given by

\[ R_{Y_E} = \left( \frac{L \sigma_t^2 M_t}{M_t P \gamma_p + L + \sigma_n^2} \right) I_{D M_r}. \quad (43) \]

Then, the capacity lower bound for orthogonal pilot pattern is

\[ C_{lb(id)} = \frac{1}{K} E \left[ \log_2 \det \left( I + \frac{\sigma_n^2 \hat{H} \hat{H}^H}{(L \sigma_t^2 M_t P \gamma_p + L + \sigma_n^2) I_{D M_r}} \right) \right]. \quad (44) \]

B. optimal PDR

The estimated channel \( \hat{H} \) is zero mean, thus the variance can be obtained as \( \sigma_n^2 = E[\text{tr}(\hat{H} \hat{H}^H)] \).

We can normalize the estimated channel as \( \hat{H} = \frac{1}{\sigma_n} \hat{H} \). From (38), (41), and (43), the variance can be expressed as

\[ \sigma_n^2 = \sigma_n^2 - \sigma_n^2 = \begin{cases} M_r M_t D - \frac{M_t M_r L D}{P \gamma_p + L} & \text{for independent pilot} \\ M_r M_t D_{sc} - \frac{M_t M_r L D_{sc}}{P \gamma_p + L} & \text{for scattered pilot} \\ M_r M_t D - \frac{M_t M_r L D}{M_t P \gamma_p + L} & \text{for orthogonal pilot} \end{cases} \quad (45) \]

The capacity lower bound can be written as

\[ C_{\text{lowerbound}} = \frac{1}{K} E \left[ \log_2 \det \left( I + \rho_{\text{eff}} \hat{H} \hat{H}^H \right) \right], \quad (46) \]

where

\[ \rho_{\text{eff}} = \begin{cases} \frac{D_{sc} P M_r M_t \sigma_n^2}{\sigma_n^2 (P \sigma_p^2 + L (\sigma_n^2 M_t + \sigma_n^2))} & \text{for independent pilot} \\ \frac{D_{sc} P M_r M_t \sigma_n^2}{\sigma_n^2 (P \sigma_p^2 + L (\sigma_n^2 M_t + \sigma_n^2))} & \text{for scattered pilot} \\ \frac{D_{sc} P M_r M_t \sigma_n^2}{\sigma_n^2 (P \sigma_p^2 + L (\sigma_n^2 M_t + \sigma_n^2))} & \text{for orthogonal pilot} \end{cases} \quad (47) \]
As shown in (46), the capacity lower bound can be maximized only by maximizing the effective SNR $\rho_{\text{eff}}$ using the optimal PDR. From [14], under the assumptions that $D = mP(m \geq 1)$, all of the following placements are optimal.

$$\mathbf{P} = \{i, i + m + 1, i + 2(m + 1), \ldots, i + (P - 1)(m + 1)\}$$  \hspace{1cm} (48)

where $i$ can be any integer between 1 and $m + 1$. With this placement, we can find the optimal PDR. Let $\alpha$ denote the fraction of total transmit power that is allocated to data symbols. Then the power of data and pilot symbols for three pilot patterns can be expressed as follows,

1) Independent pilot pattern

In the independent pilot pattern case, the total transmit power is allocated to data symbols over $M_t$ transmit antennas and pilot symbols for one transmit antenna at each OFDM symbol time. However, since we use the estimated channel over $M_t$ OFDM symbol times, the power allocation is for $M_tD$ data symbols and $M_tP$ pilot symbols with total $P_{\text{tot}} + P(M_t - 1)\sigma^2_{\text{p(id)}}$ transmit power and can be expressed as

$$DM_t\sigma^2_{d(id)} = \alpha'(P_{\text{tot}} + P(M_t - 1)\sigma^2_{\text{p(id)}})$$  \hspace{1cm} (49)

$$PM_t\sigma^2_{\text{p(id)}} = (1 - \alpha')(P_{\text{tot}} + P(M_t - 1)\sigma^2_{\text{p(id)}}),$$  \hspace{1cm} (50)

where $\alpha'$ is the fraction of power $P_{\text{tot}} + P(M_t - 1)\sigma^2_{\text{p(id)}}$ for data symbols, and $P_{\text{tot}}$ is the total transmit power of the system. If we consider the fraction of total transmit power for data symbols, it can be given by

$$\sigma^2_{d(id)} = \frac{\alpha P_{\text{tot}}}{DM_t}, \quad \sigma^2_{\text{p(id)}} = \frac{(1 - \alpha)P_{\text{tot}}}{P}.$$  \hspace{1cm} (51)

From (50) and (51), we can find

$$\alpha = \frac{\alpha'}{M_t - \alpha'(M_t - 1)}$$  \hspace{1cm} (52)

Therefore, if we find the optimal $\alpha$ to maximize the capacity, then we also have the maximum capacity with $\alpha'$ corresponding to $\alpha$.

2) Scattered pilot pattern

$$\sigma^2_{d(sc)} = \frac{\alpha P_{\text{tot}}}{D_{sc}M_t}, \quad \sigma^2_{\text{p(sc)}} = \frac{(1 - \alpha)P_{\text{tot}}}{PM_t}.$$  \hspace{1cm} (53)
3) Orthogonal pilot pattern

\[ \sigma^2_{d(\text{or})} = \frac{\alpha P_{\text{tot}}}{DM_t}, \quad \sigma^2_{p(\text{or})} = \frac{(1 - \alpha)P_{\text{tot}}}{PM_t}. \] (54)

Then, the PDR for all three pilot patterns are

\[ \eta = \begin{cases} 
(1 - \alpha)D & \text{for independent / orthogonal pilot} \\
(1 - \alpha)D_{sc} & \text{for scattered pilot} 
\end{cases} \] (55)

Using \( \alpha \), the effective SNR can be written as

\[ \rho_{\text{eff}} = \begin{cases} 
\frac{\alpha(1-\alpha)D P_{\text{tot}}^2 M_r}{\sigma_n^2 (\sigma_n^2 + D(\sigma_n^2 + L\sigma_n^2))} & \text{for independent / orthogonal pilot} \\
\frac{\alpha(1-\alpha)D_{sc} P_{\text{tot}}^2 M_r}{\sigma_n^2 (\sigma_n^2 + D_{sc}(\sigma_n^2 + L\sigma_n^2))} & \text{for scattered pilot} 
\end{cases} \] (56)

If we consider a single-input single-output system by setting \( M_t \) and \( M_r \) equal to 1, (56) is similar to the effective SNR in [14]. From (56), we can find that the capacity with independent pilot pattern is the same as the one with orthogonal pilot pattern. In independent and orthogonal pattern cases, we get gain of \( M_t \) in pilot SNR when estimating the channel by transmitting \( M_t \) times more power for pilot symbols and averaging the noise over \( M_t \) OFDM symbols, respectively. Now we can find the \( \alpha \) which maximizes the effective SNR over \( 0 < \alpha < 1 \) by differentiating \( \rho_{\text{eff}} \) and setting it equal to zero. The following Lemma shows the optimal PDR and the capacity lower bound under the assumption that \( P \) is a factor of \( K \) and \( P \geq L \).

**Lemma 1:** If we assume that \( P \) is a factor of \( K \) and \( P \geq L \), the optimal fraction of total transmit power that is allocated to data symbols is given by

\[ \alpha_{\text{max}} = \arg \max_{0 < \alpha < 1} \rho_{\text{eff}} = \begin{cases} 
\frac{1}{2} & \text{if } A \text{ is true} \\
\frac{-\gamma + \sqrt{\beta \gamma + \gamma^2}}{\beta} & \text{otherwise} 
\end{cases} \] (57)

where

\[ \begin{align*}
A &: L = D, \beta = P_{\text{tot}}(L - D), \gamma = D(P_{\text{tot}} + L\sigma_n^2) & \text{for independent / orthogonal pilot} \\
A &: LM_t = D_{sc}, \beta = P_{\text{tot}}(LM_t - D_{sc}), \gamma = D_{sc}(P_{\text{tot}} + LM_t\sigma_n^2) & \text{for scattered pilot}
\end{align*} \]

**Proof:** To find the optimal \( \alpha \), we need to solve the optimization problems for the three pilot pattern cases by finding \( \max \rho_{\text{eff}} \) subject to the power constraints:

\[ DM_t\sigma_d^2 + P\sigma_p^2 = P_{\text{tot}} \quad \text{for independent pilot pattern} \]
\[ D_{sc} M_t \sigma_d^2 + P M_t \sigma_p^2 = P_{tot} \quad \text{for scattered pilot pattern} \]
\[ D M_t \sigma_d^2 + P M_t \sigma_p^2 = P_{tot} \quad \text{for orthogonal pilot pattern} \]

(58)

where \( \rho_{eff} \) for three pilot pattern cases are in (47). This optimization problem can be replaced by the following optimization problem with (56);

\[
\max_{0<\alpha<1} \rho_{eff},
\]

(59)

where \( \rho_{eff} \) is the effective SNR in (56). We can find the \( \alpha \) which maximizes the effective SNR over \( 0 < \alpha < 1 \) by differentiating \( \rho_{eff} \) and setting it equal to zero. This results in a simple optimization problem similar to [18].

The capacity lower bound with this optimal \( \alpha_{max} \) is

\[
C_{\text{lowerbound}} = \frac{1}{K} E \left[ \log_2 \det (I + \rho_{max-\text{eff}} \hat{H} \hat{H}^H) \right],
\]

(60)

where the maximum effective SNR is the effective SNR in (56) with the optimal \( \alpha_{max} \) in Lemma 1. From Lemma 1, we can find the optimal PDR of the MIMO-OFDM system.

**Theorem 1:** Under the assumption that \( P \) is a factor of \( K \) and \( P \geq L \), the optimal PDR of MIMO-OFDM system which maximize the capacity is

1) Independent / orthogonal pilot pattern

\[
\eta_{opt} = \begin{cases} 
\frac{D}{P} & \text{if } L = D \\
\frac{\sqrt{LD(L \sigma_n^2 + P_{tot})(P_{tot} + \sigma_n^2 D)}}{P(P_{tot} + L \sigma_n^2)} & \text{otherwise}
\end{cases}
\]

(61)

2) Scattered pilot pattern

\[
\eta_{opt} = \begin{cases} 
\frac{D_{sc}}{P} & \text{if } LM_t = D_{sc} \\
\frac{\sqrt{LD_{sc} M_t (LM_t \sigma_n^2 + P_{tot})(P_{tot} + \sigma_n^2 D_{sc})}}{P(P_{tot} + LM_t \sigma_n^2)} & \text{otherwise}
\end{cases}
\]

(62)

**Proof:** From Lemma 1 and the definition of PDR in (55), the result follows easily.

V. RESULTS

In this section, we demonstrate the capacity lower bound of MIMO-OFDM systems using pilot symbol based MMSE channel estimation with the optimized PDR using Monte Carlo simulation over the channel realizations, for each of the three considered pilot symbol constructions. For the simulation, \( L \) is 3 for fixed \( L \) case, and \( K \) is 64. For independent and orthogonal pilot
patterns, 4 pilot symbols are placed periodically in the subchannels at each transmit antenna. For the scattered pattern, $4 \times M_t$ pilot symbols are transmitted at each transmit antenna.

A. Ergodic capacity versus pilot power

Fig. 2 shows the ergodic capacity of a $2 \times 2$ MIMO-OFDM system with three different pilot patterns according to the percentage of pilot power ($1-\alpha$) for 2 different SNR cases (SNR=5,15dB). The case of the perfect channel knowledge allocates the entire transmit power to data symbols and uses 60 subcarriers for data symbols. The capacity of the independent and orthogonal pilot patterns show the same results. At each SNR, the capacity lower bound has maximum value when the percentage of pilot power is about 0.2 for independent and orthogonal cases and about 0.27 for scattered case. This means that we can maximize the capacity by allocating 20% and 27% of the total transmit power to pilot symbols in these cases. For this configuration of MIMO-OFDM systems, this percentage results in the PDR of $\sigma_p^2/\sigma_d^2 = 3.75$ and 3.5, respectively. The capacity is decreased as the pilot power is decreased when the percentage of pilot power is below those percentages, even though the data power is increased, since low pilot power results in poor channel estimation. Similarly, when the pilot power is further increased, the capacity decreases despite the improved channel estimation, since the data power is decreased under a fixed total transmit power constraint. Fig. 3 shows the ergodic capacity of a $4 \times 4$ MIMO-OFDM system. The capacity gap between the independent/orthogonal pattern and scattered pattern is larger than in a $2 \times 2$ MIMO-OFDM system. The results also show that the $4 \times 4$ system with a scattered pilot pattern has a larger capacity gap between the optimal PDR case and the perfect channel case than the $2 \times 2$ system. This gap can be expressed with a capacity efficiency, and it will be discussed more in Section V-B with Fig. 6. An overall observation is that the capacity is not especially sensitive to the PDR as long as it is in a certain region. For example, a quasi-optimal region is about 0.1 $\sim$ 0.3 in $2 \times 2$ MIMO-OFDM systems with independent or orthogonal pilot patterns. Fig. 4 shows the comparison of the capacity of $2 \times 2$, $2 \times 4$, and $4 \times 4$ MIMO-OFDM systems with three different pilot patterns versus the percentage of pilot power when SNR is 10dB. The optimal PDR of scattered pilot pattern is more sensitive
to the number of transmit antennas than that of independent pilot and orthogonal pilot.

B. Ergodic capacity with optimal PDR

The ergodic capacity with the optimal PDR in Theorem 1 for 2×2 and 4×4 MIMO-OFDM systems with three different pilot patterns is shown in Fig. 5. In Fig. 5, the ergodic capacity of the independent/orthogonal pattern case with the optimal PDR shows higher capacity than the scattered pattern case. It is because we get Mt gain in pilot SNR when estimating the channel in the independent/orthogonal pattern case and use more subcarriers for data symbols than in the scattered pattern case. Figs. 6 and 7 show the capacity efficiency defined as below,

$$\mu = \frac{\text{capacity with PDR } \eta}{\text{capacity with perfect channel knowledge}}.$$  \hfill (63)

This capacity efficiency tells us how much capacity we can get with PDR \( \eta \) when the maximum achievable capacity is the open loop capacity with perfect channel knowledge. We can use this metric to compare the achievable capacity with different PDR. In Figs. 6 and 7, we compare the optimal PDR case with the \( \sigma_d^2 = \sigma_p^2 \) case for 2×2 and 4×4 MIMO-OFDM systems considering three different pilot patterns. In the case of the independent pilot pattern with equal power, \( \sigma_p^2(id) \) is \( M_t \) times \( \sigma_p^2(or) \). It is because only one antenna transmits pilot symbols at a time. We can find that the capacity efficiency with optimal PDR is higher than that of equal power. For example, in 2×2 MIMO-OFDM systems with SNR=10dB, we can get about 19% more capacity efficiency for the scattered pattern and about 10% more capacity efficiency for the independent/orthogonal pattern. It means that higher capacity results by simply using this optimal PDR.

C. Optimal PDR

Figs. 8 and 9 show the relation between the optimal PDR, the number of channel taps, and the number of transmit antennas for the independent/orthogonal pattern and scattered pattern, respectively. As shown in (61) and Fig. 8, independent and orthogonal pilot patterns are not sensitive to \( M_t \). In this case, the number of transmit antennas only affects to \( \sigma_n^2 \) in (61). It results in the little change of optimal PDR according to \( M_t \). The optimal PDR heavily depends on the channel duration \( L \), since higher \( L \) means that we need more pilot power to estimate the
channel correctly. From Fig. 9, we can find that the optimal PDR of the scattered pilot pattern depends on both $M_t$ and $L$. If $M_t$ is increased, more power is needed for the pilot symbols in order to estimate the channel and the number of subcarriers for data symbols is decreased. The effect of increasing $L$ is similar for the independent/orthogonal pilot pattern case.

VI. CONCLUSION

In this paper, a capacity lower bound for MIMO-OFDM systems with MMSE channel estimation was derived, and the optimal PDR of the system for maximizing the capacity lower bound was formed and analyzed for three likely pilot patterns. There is a tradeoff between the power for data symbols and the accuracy of the channel estimation in MIMO-OFDM systems when the total transmit power is fixed. The optimal PDR for maximizing capacity results showed that independent and orthogonal pilot patterns show the same performance, and their performance is superior to a frequency-scattered pilot pattern. From the analysis, it is shown that the capacity of a $2 \times 2$ MIMO-OFDM system can be increased about 10%~30% compared with the case of equal power allocation by simply using the optimal PDR. The optimal PDR depends not only on the system configuration, but also on the channel delay spread, and it increases with the delay spread and number of antennas. Nevertheless, it should be quite practical to use the results of this paper to design the PDR for typical cases, since the system capacity is not highly sensitive to the PDR over a fairly broad range for a given system configuration.

REFERENCES


APPENDIX I

THE AUTOCORRELATION OF $Y_E$, $R_{Y_E}$

From (35), we need to analyze $\tilde{H}_j \tilde{H}_j^H$ to find the autocorrelation of $Y_E$.

\[ \tilde{H}_j = H_j - \hat{H}_j \]

\[ = \begin{bmatrix} \text{diag} \left( W_d h_{1,j} \right) & \cdots & \text{diag} \left( W_d h_{M_t,j} \right) \end{bmatrix} - \begin{bmatrix} \text{diag} \left( W_d \hat{h}_{1,j} \right) & \cdots & \text{diag} \left( W_d \hat{h}_{M_t,j} \right) \end{bmatrix} \]

\[ = \begin{bmatrix} \text{diag} \left( W_d \tilde{h}_{1,j} \right) & \text{diag} \left( W_d \tilde{h}_{2,j} \right) & \cdots & \text{diag} \left( W_d \tilde{h}_{M_t,j} \right) \end{bmatrix} \]

(A.1)

Thus,

\[ \tilde{H}_j \tilde{H}_j^H = \left( H_j - \hat{H}_j \right) \left( H_j - \hat{H}_j \right)^H \]

\[ = \text{diag} \left( W_d \tilde{h}_{1,j} \tilde{h}_{1,j}^H W_d^H \right) + \cdots + \text{diag} \left( W_d \tilde{h}_{M_t,j} \tilde{h}_{M_t,j}^H W_d^H \right) \quad (A.2) \]

From (A.2),

\[ W_d \left( \tilde{h}_{i,j} \tilde{h}_{i,j}^H \right) W_d^H = W_d I_{i,j} \left( \tilde{h}_{p} \tilde{h}_{p}^H \right) I_{i,j}^H W_d^H \quad (A.3) \]

where $I_{i,j} = \begin{bmatrix} 0_{(i-1)M_t + M_r(j-1)L} & I_L & 0 \end{bmatrix}$, the size of $I_i$ is $(L \times L M_t M_r)$, $0_{(i-1)L + M_r(j-1)L}$ is an $(L \times ((i-1)L + M_r(j-1)L))$ all zero matrix, and $I_L$ is a $(L \times L)$ identity matrix. Thus, the autocorrelation of $Y_E$ is

\[ R_{Y_E} = \sigma_d^2 E \left[ \text{mdiag} \left( \tilde{H}_j \tilde{H}_j^H \right)_{j=1}^{M_r} \right] + \sigma_n^2 I \]

\[ = \sigma_d^2 \cdot \text{mdiag} \left( \sum_{i=1}^{M_t} \text{diag} \left( W_d I_{i,j} E \left[ \tilde{h}_{p} \tilde{h}_{p}^H \right] I_{i,j}^H W_d^H \right) \right)_{j=1}^{M_r} + \sigma_n^2 I \]
Fig. 1. System Model
Fig. 2. Capacity versus pilot power for $2 \times 2$ MIMO-OFDM systems with three different pilot patterns, $L=3$. 

- Perfect channel
- Independent pilot
- Scattered pilot
- Orthogonal pilot

SNR=15dB
SNR=5dB
Fig. 3. Capacity versus pilot power for $4 \times 4$ MIMO-OFDM systems with three different pilot patterns, $L = 3$. 
Fig. 4. Comparison of the capacity of $2\times2$, $2\times4$, and $4\times4$ MIMO-OFDM systems versus pilot power, SNR=10dB, L=3
Fig. 5. Capacity with optimal PDR versus SNR for 2×2 and 4×4 MIMO-OFDM systems, L=3
Fig. 6. Capacity efficiency versus SNR for $2 \times 2$ MIMO-OFDM systems with optimal PDR and equal power ($\sigma_p^2 = \sigma_d^2$)
Fig. 7. Capacity efficiency versus SNR for 4×4 MIMO-OFDM systems with optimal PDR and equal power ($\sigma_p^2 = \sigma_d^2$)
Fig. 8. Optimal PDR of independent pilot or orthogonal pilot MIMO-OFDM systems with various number of $L$ and $M_t$. 
Fig. 9. Optimal PDR of scattered pilot MIMO-OFDM systems with various number of $L$ and $M_t$. 