Exploiting Spatio-Temporal Correlations in MIMO Wireless Channel Prediction

Ian C. Wong and Brian L. Evans

Wireless Networking and Communications Group
The University of Texas at Austin, Austin, Texas 78712
Email: {iwong, bevans}@ece.utexas.edu

***Abstract*— We investigate prediction algorithms that exploit both temporal and spatial correlations in MIMO correlated narrowband fading wireless channels. We first derive the optimal two dimensional minimum mean square error (2D-MMSE) prediction filter that maximally exploits the available temporal and spatial correlations. We then propose a lower complexity 2-step prediction algorithm, which first exploits temporal correlations using a classical single-input single-output (SISO) MMSE time-domain prediction filter for each entry in the MIMO channel, followed by an MMSE spatial smoothing step to exploit the spatial correlations. Compared to the SISO and specular prediction approaches, this approach either achieves lower MSE with a slight increase in complexity, or comparable MSE with lower complexity, in a wide range of wireless channel conditions. The same advantage holds for the average mutual information.**

I. Introduction

In MIMO wireless communications, exploiting channel state information (CSI) at the transmitter allows for increased link robustness and/or increased capacity [1]. This CSI is typically fed back from the receiver using reciprocity in time-division duplex (TDD) mode, or explicit feedback (full or partial) in frequency-division duplex (FDD) mode. In mobile environments where the channel changes rapidly, the CSI that arrives at the transmitter would be outdated due to feedback delay. This problem can be overcome by using *channel prediction*.

Channel prediction algorithms have been studied extensively in the SISO case (see e.g. [2] [3] [4]), and have been shown to effectively overcome the outdated CSI problem for single antenna systems. Recently, the channel prediction concept has been extended to the MIMO case by several researchers. In [5] and [6], linear predictors that exploit temporal correlations similar to the ones used in the SISO case [2] [3] were applied directly to each of the channel coefficients in the MIMO channel. This approach, which we term as the *SISO approach*, is appealing when the channel responses at each antenna element in the transmit and receive arrays are uncorrelated, e.g. when the classical IID (spatially white) channel model is used. Unfortunately, real world channels typically experience spatial correlation, which was not exploited in these simple SISO prediction based approaches. In [7], linear prediction is performed on each of the dominant right singular vectors of the stacked channel matrix. Their method showed lower mean square error versus the SISO approach, but is highly complex due to the required singular value decomposition (SVD) of a large matrix. We call this the *specular approach* as termed by the authors.

In this paper, we initially derive the optimal 2-dimensional minimum mean square error (2D-MMSE) narrowband MIMO channel prediction filter that maximally exploits the available spatial and temporal correlations. Since designing the 2D-MMSE filter is highly computationally complex, we propose a reduced complexity 2-step prediction approach that is able to exploit the temporal and spatial correlations without significant computational burden. This approach first uses a SISO MMSE time-domain prediction filter for each entry in the MIMO channel, followed by an MMSE spatial smoothing filter to exploit the spatial correlations to improve the prediction performance. This proposed algorithm performs similarly to the specular approach of [7] with much lower computational complexity, and outperforms the SISO approach with only a slight increase in complexity. Thus, our approach presents an attractive prediction performance versus computational complexity tradeoff in correlated MIMO channels.

II. Narrowband MIMO channel model

Consider an N_t transmit antenna and N_r receive antenna MIMO wireless system, with narrowband, time-varying, $N_r \times N_t$ channel matrix $\mathbf{H}(n)$ modeled as [1]

$$\mathbf{H}(n) = \mathbf{R}_{rx}^{1/2}\mathbf{H}_w(n)\mathbf{R}_{tx}^{1/2} \quad (1)$$

where $\mathbf{H}_w(n)$ is the spatially white $N_r \times N_t$ IID MIMO channel matrix, $\mathbf{R}_{tx}$ and $\mathbf{R}_{rx}$ are the Hermitian-symmetric positive definite $N_r \times N_r$ and $N_t \times N_t$ transmit and receive covariance matrices, and where $(\cdot)^{1/2}$ represents the Hermitian square-root of a matrix [1].

We assume that each of the elements of $\mathbf{H}_w(n)$ are uncorrelated wide-sense stationary (WSS) complex Gaussian stochastic processes with zero mean and identical[1] time correlation function $r_t(\Delta) \triangleq \mathbb{E}\{[\mathbf{H}_w(n+\Delta)]_{i,j}[\mathbf{H}_w(n)]^*_{i,j}\} \quad \forall i,j$ where $[\cdot]_{i,j}$ indicates the (i,j)th component of the matrix argument. Notice that the spatio-temporal correlation function is separable, and can be written as

$$\mathbb{E}\{[\mathbf{H}(n+\Delta)]_{i,j}[\mathbf{H}(n)]^*_{k,l}\} = r_t(\Delta)[\mathbf{R}_{tx}^T]_{j,l}[\mathbf{R}_{rx}]_{i,k}, \quad (2)$$

[1]We chose to make this assumption to simplify the presentation and notation, but is not crucial for the development of the algorithm. Our algorithm can be easily extended to the more general case where each transmit-receive antenna pair have different temporal correlations.

We further assume that we have available causal estimates of the MIMO channel $\hat{\mathbf{H}}(n) = \mathbf{H}(n) + \mathbf{E}(n)$ where $\mathbf{E}(n)$ is the estimation error matrix, whose elements are IID zero-mean, circular symmetric, complex Gaussian with variance σ^2. We shall occasionally use the vectorized version of the MIMO channel matrix $\hat{\mathbf{h}}(n) = \text{vec}(\hat{\mathbf{H}}(n))$, where $\text{vec}(\cdot)$ is the operator that stacks the columns of its argument one of top of the other into a vertical vector, making $\hat{\mathbf{h}}(n)$ the $N_r N_t \times 1$ vector version of channel matrix $\hat{\mathbf{H}}(n)$. We let $N = N_r N_t$ in the subsequent discussion for notational brevity.

III. 2D-MMSE Prediction

We denote the $N \times L$ matrix of L current and previous estimates of the MIMO channel spaced Δ_t apart as

$$\hat{\boldsymbol{\mathcal{H}}} = [\hat{\mathbf{h}}(n), \hat{\mathbf{h}}(n-\Delta_t), \ldots, \hat{\mathbf{h}}(n-(L-1)\Delta_t)] \tag{3}$$

and its vectorized version as $\hat{\mathfrak{h}} = \text{vec}(\hat{\boldsymbol{\mathcal{H}}})$ with length NL. We predict the MIMO channel $\hat{\mathbf{h}}(n+\Delta_t)$ using the NL 2D-MMSE filter $\mathbf{W}_{2D}$, given as

$$\hat{\mathbf{h}}(n+\Delta_t) = \mathbf{W}_{2D}\hat{\mathfrak{h}} \tag{4}$$

where $\mathbf{W}_{2D} = \arg\min_{\mathbf{W}} \mathbb{E}\{\| \mathbf{h}(n+\Delta_t) - \mathbf{W}\hat{\mathfrak{h}} \|^2\}$.

Using the orthogonality principle (see e.g. [8]), the 2D-MMSE filter should satisfy $\mathbb{E}\{(\mathbf{h}(n+\Delta_t) - \mathbf{W}_{2D}\hat{\mathfrak{h}})\hat{\mathfrak{h}}^H\} = \mathbf{0}$, where after some simplification becomes

$$\mathbf{W}_{2D} = \left(\mathbf{r}_t^T \otimes \mathbf{R}_s\right)\left(\mathbf{R}_t^T \otimes \mathbf{R}_s + \sigma^2\mathbf{I}\right)^{-1} \tag{5}$$

where $\mathbf{R}_t$ is the Hermitian-symmetric and Toeplitz $L \times L$ temporal autocorrelation matrix with entries $[\mathbf{R}_t]_{i,j} = r_t((i-j)\Delta_t)$, $\mathbf{r}_t^T = [r_t(\Delta_t), \ldots, r_t(L\Delta_t)]$ is the vector of Δ_t-ahead cross-correlations, and $\mathbf{R}_s = \mathbf{R}_{tx}^T \otimes \mathbf{R}_{rx}$ is the spatial autocorrelation matrix. The MMSE is given by

$$\epsilon_{2D} = \text{Tr}\left\{\mathbf{R}_s - \left(\mathbf{r}_t^T \otimes \mathbf{R}_s\right)\left(\mathbf{R}_t^T \otimes \mathbf{R}_s + \sigma^2\mathbf{I}\right)^{-1}\left(\mathbf{r}_t \otimes \mathbf{R}_s\right)^H\right\} \tag{6}$$

We can also use this prediction filter to predict other future time instances as well. If the prediction length exceeds Δ_t, we can iterate the prediction process by treating the latest predicted channel as the current channel, and dropping the channel furthest in the past. This can be repeated as necessary, although error propagation is a problem, especially for very long prediction lengths. The channel for time instances that are not a multiple of Δ_t can then be determined using interpolation. The prediction filter could also be designed for other time instances explicitly by simply changing the time lag in cross-correlation vector $\mathbf{r}_t$, but it would be much more complex because (5) needs to be computed separately for each time lag.

Since the 2D-MMSE prediction filter exploits all the possible spatio-temporal correlations, we expect it to outperform the previous SISO approaches [5] [6] when spatial correlation is present. However, the computation of $\mathbf{W}_{2D}$ requires the inversion of a block Toeplitz matrix, which entails $O(N^3L^2)$ operations, and is a huge computational burden for cost-effective implementation. The prediction filtering operation itself requires $O(N^2L)$, which is also quite complex.

IV. 2-step Prediction

We reduce the computational burden by exploiting temporal and spatial correlations using two separate filters. We first perform time-domain prediction ignoring the spatial correlation, followed by spatial smoothing to exploit the antenna correlations.

A. Time-domain prediction step

We use a single L-tap time-domain prediction filter $\mathbf{w}_t$ to predict the MIMO channel matrix as

$$\tilde{\mathbf{h}}_t(n+\Delta_t) = \hat{\boldsymbol{\mathcal{H}}}\mathbf{w}_t \tag{7}$$

Note that this is equivalent to the SISO approach of [5], and $\mathbf{w}_t$ is simply the SISO linear prediction filter given as

$$\mathbf{w}_t = (\mathbf{R}_t + \sigma^2\mathbf{I})^{-1}\mathbf{r}_t \tag{8}$$

with MSE

$$\epsilon_t = N(r_t(0) - \mathbf{r}_t^H(\mathbf{R}_t + \sigma^2\mathbf{I})^{-1}\mathbf{r}_t) \tag{9}$$

B. Spatial-smoothing step

We now apply a $N \times N$ MMSE spatial smoothing filter $\mathbf{W}_s$ to exploit the spatial correlation, i.e.

$$\tilde{\mathbf{h}}_s(n+\Delta_t) = \mathbf{W}_s\tilde{\mathbf{h}}_t(n+\Delta_t) \tag{10}$$

Using the orthogonality principle, $\mathbf{W}_s$ should satisfy $\mathbb{E}\left\{\left(\mathbf{h}(n+\Delta_t) - \mathbf{W}_s\tilde{\mathbf{h}}_t(n+\Delta_t)\right)\tilde{\mathbf{h}}_t^H(n+\Delta_t)\right\} = \mathbf{0}$, whose solution is

$$\mathbf{W}_s = \mathbf{R}_{\mathbf{h}\tilde{\mathbf{h}}_t}\mathbf{R}_{\tilde{\mathbf{h}}_t}^{-1} \tag{11}$$

where

$$\begin{aligned}\mathbf{R}_{\mathbf{h}\tilde{\mathbf{h}}_t} &= \mathbb{E}\{\mathbf{h}(n+\Delta_t)\tilde{\mathbf{h}}_t^H(n+\Delta_t)\} \\ &= \left(\mathbf{r}_t^H\mathbf{w}_t\right)\mathbf{R}_s\end{aligned} \tag{12}$$

is the cross-correlation matrix of the true future channel and the time-domain predicted channel, and

$$\begin{aligned}\mathbf{R}_{\tilde{\mathbf{h}}_t} &= \mathbb{E}\{\tilde{\mathbf{h}}_t(n+\Delta_t)\tilde{\mathbf{h}}_t^H(n+\Delta_t)\} \\ &= (\mathbf{w}_t^H\mathbf{R}_t\mathbf{w}_t)\mathbf{R}_s + (\mathbf{w}_t^H\mathbf{w}_t)\sigma^2\mathbf{I}\end{aligned} \tag{13}$$

is the autocorrelation of the time-domain predicted channel (see Appendix I for the derivation). The MSE in this case is

$$\epsilon_{2S} = \text{Tr}\left\{\mathbf{R}_s - \mathbf{R}_{\mathbf{h}\tilde{\mathbf{h}}_t}\mathbf{R}_{\tilde{\mathbf{h}}_t}^{-1}\mathbf{R}_{\mathbf{h}\tilde{\mathbf{h}}_t}^H\right\} \tag{14}$$

In terms of complexity, the time-domain prediction step requires $O(L^2)$ (using e.g. Levinson-Durbin [8]) for the design of the filter, and $O(NL)$ for prediction. The spatial smoothing step requires $O(L^2 + N^3)$ for the design of the filter, and $O(N^2)$ for prediction.

V. Performance Analysis and Simulations

We analyze the performance of our proposed prediction algorithms, and compare them with applying SISO linear prediction on each entry in the MIMO channel matrix [5] [6], which we call the *SISO* approach; and with applying SISO linear prediction on the right singular vectors of the stacked channel matrix $\hat{\boldsymbol{\mathcal{H}}}$ (3), which we call the *specular* approach.

TABLE I

COMPLEXITY OF L-ORDER, N_t-TRANSMIT AND N_r-RECEIVE ANTENNA MIMO CHANNEL PREDICTION ALGORITHMS WITH $N = N_r N_t$.

Algorithm	Design	Prediction
2D-MMSE	$O(N^3L^2)$	$O(N^2L)$
Specular [7]	$O(NL^2 + N^2L)$	$O(NL + N^2)$
2-step	$O(L^2 + N^3)$	$O(NL + N^2)$
SISO [5][6]	$O(L^2)$	$O(NL)$

TABLE II

SIMULATION PARAMETERS

Item	Value	Item	Value
Bandwidth	10 kHz	Frame length	1 ms
Carrier frequency	2.0 GHz	Time-domain spacing	2 ms
Symbol period	50 μs	Prediction length	2 ms
Velocity	20 kph	Filter order	32
Spatial corr. coef.	0.9	No. of training symbols	500

A. Complexity analysis

The complexity of the design stage and prediction stage of different prediction algorithms are shown in Table I in decreasing order. Since $L \gg N$ in most practical scenarios, the 2D-MMSE and specular approaches are much more complex than the 2-step and SISO approaches, and 2-step is only slightly more complex than the SISO approach.

B. Mean square error analysis

We have the following proposition (whose proof is in Appendix II) in terms of the theoretical MSE performance

Proposition 1: Given the channel model in Sec. II where $\mathbf{R}_s$ and $\mathbf{R}_t$ are both non-singular, the mean square error of the 2D-MMSE (6), 2-step (14), and SISO (9) prediction approaches have the following relation

$$\epsilon_{2D} \leq \epsilon_{2S} \leq \epsilon_t \tag{15}$$

with equality when $\mathbf{R}_s = \mathbf{I}_N$ (spatially white channel) or when $\sigma^2 = 0$ (noiseless channel).

Both 2D-MMSE and 2-step approaches are guaranteed to outperform the SISO approach, except in the (unrealistic) cases when we have a spatially white or noiseless channel, where the performance is the same. These results are not surprising since 2D-MMSE and 2-step exploit the spatial correlation, which is ignored by the SISO approach. However, the identical performance when $\sigma^2 = 0$ is not immediately apparent. The theoretical MSE relationship of these linear approaches with the non-linear specular approach is not as straightforward. Thus, we use simulations for its performance assessment.

C. Monte-Carlo simulations

We consider a time-division-duplex narrowband 2×2 MIMO system with parameters given in Table II unless otherwise specified. We assume that at the start of each uplink frame, a preamble is present where we estimate the MIMO channel, and through reciprocity then determine an estimate of the downlink channel. We then use $L = 32$ of these previous estimates, spaced Δ_t apart, to predict the channel one round-trip-time ahead, roughly twice that of the frame length.

We generate the channel $\mathbf{H}(n)$ using (1), where the elements of $\mathbf{H}_w(n)$ are generated IID using the modified Jake's simulator [9] with 64 rays. The spatial correlation matrices are generated using the exponential model, i.e., the i, jth entry of the matrix is given by $r_{i,j} = \rho^{|i-j|}$, where $0 \leq \rho \leq 1$ signifies the degree of spatial correlation of the matrix. For the following simulation curves, we generated 10000 noisy channel realizations, with $M + 1$ time-steps Δ_t apart per realization, where the first M time steps are used to estimate the time and space autocorrelation functions, given as

$$\hat{r}_t(\Delta) = \frac{1}{N_t N_r} \sum_{i,j} \sum_{n=0}^{M-1-\Delta} [\hat{\mathbf{H}}(n + \Delta)]_{i,j} [\hat{\mathbf{H}}(n)]^*_{i,j} \tag{16}$$

$$\hat{\mathbf{R}}_s = \frac{1}{M} \sum_{n=0}^{M-1} \hat{\mathbf{h}}(n) \hat{\mathbf{h}}(n)^H, \tag{17}$$

and the next time step being the channel Δ_t ahead that we wish to predict. Figs. 1-3 compares the normalized MSE, given as $\text{NMSE}(\Delta_t) = \frac{\mathbb{E}\{\|\mathbf{H}(n+\Delta_t) - \hat{\mathbf{H}}(n+\Delta_t)\|_F^2\}}{\mathbb{E}\{\|\mathbf{H}(n+\Delta_t)\|_F^2\}}$, for the 2D-MMSE approach assuming perfect knowledge of the autocorrelation matrices $\mathbf{R}_t$ and $\mathbf{R}_s$ (this essentially serves as the benchmark for the rest of the algorithms); the 2-step and SISO approaches using estimated autocorrelation matrices; and the specular approach using the autocorrelation method [8] to determine the autoregressive parameters. We also assume knowledge of the noise variance σ^2 in all of the above algorithms. Unless otherwise indicated, we assume a velocity of 20 kph, spatial correlation coefficient of $\rho = 0.9$, and $M = 500$ training symbols for the autocorrelation estimates.

In Figs. 1-3, the 2D-MMSE approach has the lowest MSE among all the algorithms for all parameter configurations as expected. We also observe that the 2-step prediction outperforms the SISO approach uniformly for all the configurations. The specular approach, on the other hand, performs differently in relation to the linear methods for different configurations. In Fig. 1, we see that the MSE decreases as we increase SNR, but increases for higher velocities (higher Doppler frequency). This is consistent with intuition since noisy and fast varying channels are harder to predict. We also observe that the specular approach performs better than the 2-step approach in cases where the SNR is high and the velocity is high. In Fig. 2, we observe that increasing the correlation ρ increases the prediction MSE. This can be explained by the fact that lower correlations increases the ”diversity” of our channel observations, and thus exposes the structure of the channel variations more than highly correlated observations (this is consistent with the observations in [10]). We also observe that for low correlations, 2D-MMSE, 2-step, and SISO prediction perform almost similarly, which is consistent with the fact that for totally uncorrelated channels $\rho = 0$, all three prediction

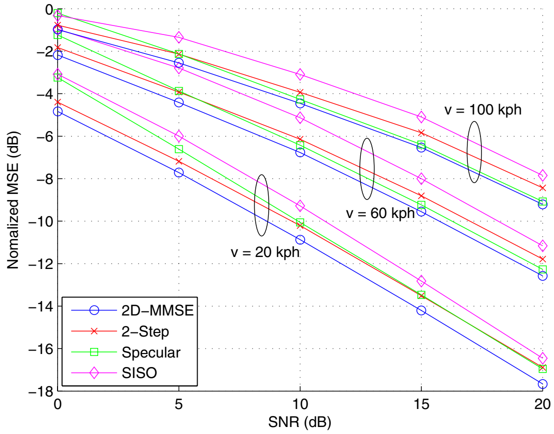

Fig. 1. Normalized prediction MSE performance for 2×2-MIMO. Simulation parameters are given in Table II.

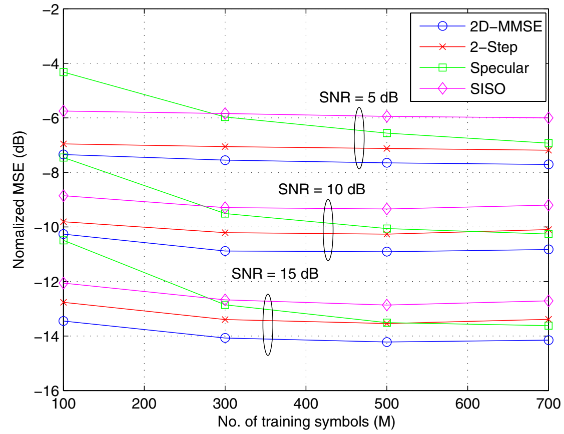

Fig. 3. Normalized prediction MSE performance for 2×2-MIMO. Simulation parameters are given in Table II.

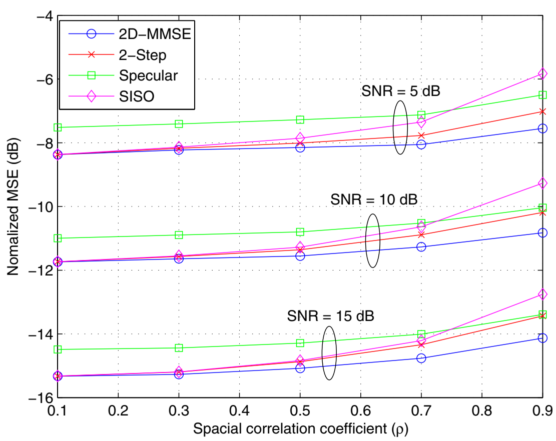

Fig. 2. Normalized prediction MSE performance for 2×2-MIMO. Simulation parameters are given in Table II.

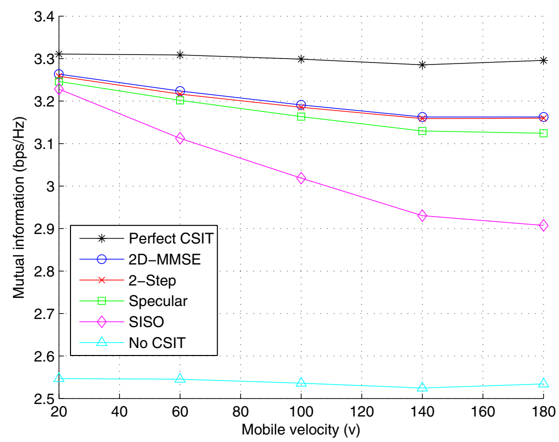

Fig. 4. Mutual information of a 2×2-MIMO system with SNR = 5 dB. Simulation parameters are given in Table II.

methods actually end up with the same prediction filter [cf. Prop. 1]. As for the specular approach, we see the advantage of its use only in high correlation and high SNR regimes. In Fig. 3, we see the effect of training length M on the MSE, and observe that training lengths beyond $M = 300$ no longer improves the performance significantly. We also observe that the specular approach is more sensitive to the training length than the linear approaches. We observed similar trends for different $N_r \times N_t$ MIMO configurations, but the results are omitted here due to space constraints.

Fig. 4 shows the ergodic mutual information of a $2\times$ 2-MIMO system assuming perfect channel state information at the receiver and different assumptions on the channel state information at the transmitter (CSIT), computed as $\frac{1}{C}\sum_{i=1}^{C}\log_2\left(\det\left(\mathbf{I}_{N_r}+\frac{1}{N_t\sigma^2}\mathbf{H}_i\mathbf{R}_i\mathbf{H}_i^H\right)\right)$ where $C =$ 10000 is the number of generated IID channel realizations, $\mathbf{H}_i$ corresponds to the "future" channel for the ith realization, and $\mathbf{R}_i$ is the transmit symbol covariance matrix corresponding to the following scenarios:

- Perfect CSIT: $\mathbf{R}_i = \mathbf{V}_i\mathbf{\Gamma}_i\mathbf{V}_i^H$ where $\mathbf{V}_i$ is the matrix whose columns are the eigenvectors of $\mathbf{H}_i^H\mathbf{H}_i$, and $\mathbf{\Gamma}_i$ is the diagonal gain matrix computed via waterfilling [1].
- Predicted CSIT: Same as the perfect CSIT case but using the corresponding predicted channel matrix $\hat{\mathbf{H}}_i$.
- No CSIT: $\mathbf{R}_i = \mathbf{I}_{N_t}$

Observe that increasing the velocity causes a degradation in mutual information for all the prediction algorithms. However, the 2D-MMSE and 2-step prediction algorithms outperform the previous SISO and specular approaches. Note also that having perfect/predicted CSIT outperforms no CSIT.

VI. Conclusion

We derived two prediction algorithms that exploit both temporal and spatial correlations for the MIMO narrowband correlated fading channel. The derived 2D-MMSE prediction algorithm exploits all the available temporal and spatial correlations using a 2-dimensional filter, but is computationally prohibitive for practical implementation. Nevertheless, it provides a good benchmark for what is achievable using linear MMSE prediction for this class of MIMO channels. The proposed 2-step prediction algorithm first exploits temporal correlations using a SISO MMSE time-domain prediction filter for each entry in the MIMO channel matrix, then exploits spatial correlations using an MMSE spatial smoothing filter. We have shown that the 2-step prediction approach presents an attractive trade-off between complexity and performance, since it outperforms the SISO prediction approach with a slight increase in complexity; and it performs similarly to the specular approach with a much lower computational complexity.

APPENDIX I
STATISTICS OF TIME-DOMAIN PREDICTED MIMO CHANNELS

From (7), we can write the time-domain predicted MIMO channel as a linear combination of the columns of $\hat{\boldsymbol{\mathcal{H}}}$

$$\tilde{\mathbf{h}}_t(n+\Delta_n) = \sum_{i=0}^{L-1} w_i \hat{\mathbf{h}}(n - i\Delta_n) \tag{18}$$

where w_i is the ith element of $\mathbf{w}_t$ given in (8). Thus, using (18) in (12) and (13) respectively, we have

$$\begin{aligned}
\mathbf{R}_{\mathbf{h}\tilde{\mathbf{h}}_t} &= \mathbb{E}\left\{\mathbf{h}(n+\Delta_n)\left(\sum_{i=0}^{L-1} w_i^* \hat{\mathbf{h}}^H(n-i\Delta_n)\right)\right\} \\
&= \sum_{i=0}^{L-1} w_i^* \mathbb{E}\left\{\mathbf{h}(n+\Delta_n)\hat{\mathbf{h}}^H(n-i\Delta_n)\right\} \\
&= \sum_{i=0}^{L-1} w_i^* r_n((i+1)\Delta_n)\mathbf{R}_s \\
&= (\mathbf{w}_t^H \mathbf{r}_t)\mathbf{R}_s
\end{aligned}$$

$$\begin{aligned}
\mathbf{R}_{\tilde{\mathbf{h}}_t} &= \mathbb{E}\left\{\left(\sum_{i=0}^{L-1} w_i \hat{\mathbf{h}}(n-i\Delta_n)\right)\left(\sum_{i=0}^{L-1} w_i^* \hat{\mathbf{h}}^H(n-i\Delta_n)\right)\right\} \\
&= \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} w_i w_j^* \mathbb{E}\left\{\hat{\mathbf{h}}(n-i\Delta_n)\hat{\mathbf{h}}^H(n-j\Delta_n)\right\} \\
&= \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} w_i w_j^* r_n((j-i)\Delta_n)\mathbf{R}_s + \sum_{i=0}^{L-1}|w_i|^2\sigma^2\mathbf{I} \\
&= (\mathbf{w}_t^H \mathbf{R}_t \mathbf{w}_t)\mathbf{R}_s + (\mathbf{w}_t^H \mathbf{w}_t)\sigma^2\mathbf{I}
\end{aligned}$$

APPENDIX II
PROOF OF PROPOSITION 1

Proof: We can write both the 2-step and SISO prediction filters in the same form as (4), where $\mathbf{W}_{2D}$ is replaced by

$$\mathbf{W}_{2S} = (\mathbf{w}_t^T \otimes \mathbf{W}_s) \tag{19}$$

$$\mathbf{W}_t = (\mathbf{w}_t^T \otimes \mathbf{I}_N) \tag{20}$$

for 2-step and SISO respectively. Since $\mathbf{W}_{2D}$ is chosen to minimize the MSE, then $\epsilon_{2D} \leq \epsilon_{2S}$ and $\epsilon_{2D} \leq \epsilon_t$. Also, since $\mathbf{W}_s$ is chosen to minimize the MSE for the time-domain predicted channel [cf. Sec. IV-B], and SISO prediction (7) is simply (10) with $\mathbf{W}_s = \mathbf{I}_N$, then $\epsilon_{2S} \leq \epsilon_t$. To show equality of the MSE in the special cases of $\mathbf{R}_s = \mathbf{I}_N$ and $\sigma^2 = 0$, we show the stronger condition that the prediction filters themselves are identical in these cases.

$$\begin{aligned}
\mathbf{W}_{2D}|_{\mathbf{R}_s=\mathbf{I}_N} &= \left(\mathbf{r}_t^T \otimes \mathbf{I}_N\right)\left(\mathbf{R}_t^T \otimes \mathbf{I}_N + \sigma^2\mathbf{I}_{LN}\right)^{-1} \\
&= \left(\mathbf{r}_t^T \otimes \mathbf{I}_N\right)\left((\mathbf{R}_t^T + \sigma^2\mathbf{I}_L)\otimes\mathbf{I}_N\right)^{-1} \\
&= \left(\mathbf{r}_t^T \otimes \mathbf{I}_N\right)\left((\mathbf{R}_t^T + \sigma^2\mathbf{I}_L)^{-1}\otimes\mathbf{I}_N\right) \\
&= \left(\mathbf{r}_t^T\left((\mathbf{R}_t + \sigma^2\mathbf{I}_L)^{-1}\right)^T\right)\otimes(\mathbf{I}_N) \\
&= \mathbf{W}_t
\end{aligned}$$

$$\begin{aligned}
\mathbf{W}_{2D}|_{\sigma^2=0} &= \left(\mathbf{r}_t^T \otimes \mathbf{R}_s\right)\left(\mathbf{R}_t^T \otimes \mathbf{R}_s\right)^{-1} \\
&= \left(\mathbf{r}_t^T \otimes \mathbf{R}_s\right)\left((\mathbf{R}_t^T)^{-1} \otimes \mathbf{R}_s^{-1}\right) \\
&= \left(\mathbf{r}_t^T(\mathbf{R}_t^{-1})^T\right)\otimes(\mathbf{I}_N) \\
&= \mathbf{W}_t|_{\sigma^2=0}
\end{aligned}$$

For the 2-step method, the following equations show that the spatial smoothing filter reduces to the identity matrix when $\mathbf{R}_s = \mathbf{I}_N$ or when $\sigma^2 = 0$

$$\begin{aligned}
\mathbf{W}_s|_{\mathbf{R}_s=\mathbf{I}_N} &= \left(\mathbf{r}_t^H\mathbf{w}_t\right)\left((\mathbf{w}_t^H\mathbf{R}_t\mathbf{w}_t)\mathbf{I}_N + (\mathbf{w}_t^H\mathbf{w}_t)\sigma^2\mathbf{I}_N\right)^{-1} \\
&= \frac{\left(\mathbf{r}_t^H\mathbf{w}_t\right)}{\left(\mathbf{w}_t^H\left(\mathbf{R}_t + \sigma^2\mathbf{I}_L\right)\mathbf{w}_t\right)}\mathbf{I}_N \\
&= \frac{\left(\mathbf{r}_t^H\mathbf{w}_t\right)}{\left(\mathbf{r}_t^H\mathbf{w}_t\right)}\mathbf{I}_N = \mathbf{I}_N
\end{aligned}$$

$$\begin{aligned}
\mathbf{W}_s|_{\sigma^2=0} &= \left(\mathbf{r}_t^H\mathbf{w}_t\right)\left((\mathbf{w}_t^H\mathbf{R}_t\mathbf{w}_t)\mathbf{I}_N\right)^{-1} \\
&= \frac{\left(\mathbf{r}_t^H\mathbf{w}_t\right)}{\left(\mathbf{r}_t^H\mathbf{w}_t\right)}\mathbf{I}_N = \mathbf{I}_N
\end{aligned}$$

We thus have $\mathbf{W}_{2S} = \mathbf{W}_t$ in (19) in either case. ■

REFERENCES

[1] A. Paulraj, R. Nabar, and D. Gore, *Introduction to space-time wireless communications*. Cambridge University Press, 2003.

[2] A. Duel-Hallen, S. Hu, and H. Hallen, "Long-Range Prediction of Fading Signals," *IEEE Signal Processing Mag.*, vol. 17, no. 3, pp. 62–75, May 2000.

[3] T. Ekman, M. Sternad, and A. Ahlen, "Unbiased power prediction of Rayleigh fading channels," in *Proc. IEEE Vehicular Technology Conference*, Sept. 2002, pp. 280 – 284.

[4] I. C. Wong and B. L. Evans, "Joint Channel Estimation and Prediction for OFDM Systems," in *Proc. IEEE Global Telecommunications Conference*, St. Louis, MO, Dec. 2005.

[5] S. Zhou and G. Giannakis, "How accurate channel prediction needs to be for transmit-beamforming with adaptive modulation over Rayleigh MIMO channels?" *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1285–1294, July 2004.

[6] Z. Luo, H. Gao, Y. Liu, and J. Gao, "Robust pilot-symbol-aided MIMO channel estimation and prediction," in *Proc. IEEE Global Telecommunications Conference*, vol. 6, Dec. 2004, pp. 3646–3650 Vol.6.

[7] M. Guillaud and D. Slock, "A specular approach to MIMO frequency-selective channel tracking and prediction," in *Proc. IEEE Signal Processing Advances in Wireless Communications*, July 2004, pp. 59–63.

[8] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. John Wiley and Sons, 1996.

[9] Y. R. Zheng and C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 920–928, June 2003.

[10] T. Svantesson and A. Swindlehurst, "A performance bound for prediction of a multipath MIMO channel," *IEEE Trans. Signal Processing*, submitted.