Abstract—Finite-length delay-optimized multi-input multi-output (MIMO) equalizers that optimally shorten the impulse response memory of frequency-selective MIMO channels are derived. The MIMO equalizers are designed to minimize the average energy of the error sequence between the equalized MIMO channel impulse response and an MIMO target impulse response (TIR) with shorter memory. Two criteria for optimizing the MIMO TIR are analyzed and compared. The presented analytical framework encompasses a multitude of previously-studied finite-length equalization techniques.

Index Terms—Equalizers, FIR digital filters, mean square error methods, MIMO systems.

I. INTRODUCTION

The combination of maximum-likelihood sequence estimation (MLSE) with receiver diversity is an effective technique for achieving high performance over noisy frequency-selective fading channels impaired by cochannel interference [8], [14], [16], [15], [7]. With the addition of transmitter diversity, the resulting multi-input multi-output (MIMO) frequency-selective channel was shown [9], [10] to have a significantly higher capacity than its single-input multi-output (SIMO) or single-input single-output (SISO) counterparts. The use of maximum-likelihood multiuser detection techniques on these frequency-selective MIMO channels significantly outperforms single-user detection techniques that treat signals from other users as colored noise. However, MLSE complexity increases exponentially with the number of inputs (or transmit antennas) and with the memory of the MIMO channel1 making its implementation over severe intersymbol interference (ISI) channels very costly [6].

The discrete matrix multitone (DMMT) was shown in [9] to be a practical transceiver structure that asymptotically achieves the MIMO channel capacity when combined with powerful codes. It uses the discrete Fourier transform in its computationally-efficient fast Fourier transform (FFT) implementation to partition the frequency responses of the underlying frequency-selective channels of the MIMO systems into a large number of parallel, independent, and (approximately) memoryless frequency subchannels. To eliminate interblock and intrablock interference, a cyclic prefix whose length is equal to the MIMO channel memory is inserted in every block. On severe ISI MIMO channels, the cyclic prefix overhead reduces the achievable DMMT throughput significantly unless a large FFT size is used which in turn increases the computational complexity, processing delay, and memory requirements. An elegant solution is to implement a time-domain equalizer to shorten the channel memory and hence reduce the cyclic prefix overhead. This channel-shortening equalizer (CSE) has been well studied for SISO systems (e.g., see [2] and the references therein). In this letter, we present design procedures for the finite-length minimum-mean-square-error (MMSE) CSE for MIMO channels. As we shall show, this CSE performs spatial and temporal noise whitening, multichannel matched filtering, and MIMO channel memory shortening simultaneously subject to the finite-number-of-taps constraint.

The rest of this paper is organized as follows. In Section II, we present our input–output model and the mathematical formulation of the MIMO CSE problem. Two MIMO CSE design procedures are derived in Section III and their computational complexities are evaluated. Furthermore, several equalization structures are shown to follow as special cases of the MIMO CSE. Numerical results are presented in Section IV and the paper is concluded in Section V.

Notation: The notation to be adopted throughout this paper conforms to the following convention.

- Scalars are denoted in lower case: \( a \).
- Unless otherwise stated, vectors are column vectors and are denoted lower case bold: \( \mathbf{x} \).
- \( \mathbf{e} \text{ } \) denotes the \( i \text{th} \) unit vector (has a one in the \( i \text{th} \) position and zeros everywhere else).
- In situations where the components of the vectors are to be emphasized, the first and last components, separated by a colon, are given as a subscript to the vector: \( \mathbf{x}_{k:N-1} \).
- Matrices are upper case bold: \( \mathbf{A} \).
- \( \mathbf{I}_N \) denotes the identity matrix of size \( N \).
- \( \mathbf{0}_{N \times M} \) denotes an all-zeros matrix with \( N \) rows and \( M \) columns.
- \( |\mathbf{A}| \) denotes the determinant of matrix \( \mathbf{A} \).
- \( \text{Trace}(\mathbf{A}) \) denotes the trace of a matrix \( \mathbf{A} \).
- \( \text{E}[\cdot] \) denotes the expected value operator.
- A diagonal matrix with elements \( \{d_0, d_1, \ldots, d_{N-1}\} \) on the main diagonal will be denoted by \( \text{diag}(d_0, d_1, \ldots, d_{N-1}) \).

---

1Defined to be the maximum of the memories of impulse responses between all input and output pairs.
II. ANALYSIS

We start in this section by describing the input–output model assumed throughout this paper.

A. Input–Output Model

We consider the general case of a linear, dispersive, and noisy digital communication system with \( n_i \) inputs and \( n_o \) outputs. We use the standard complex-valued baseband equivalent signal model. Assuming a temporal oversampling factor of \( f \), the samples at the \( j \)-th channel output (\( 1 \leq j \leq n_o \)) have the standard form

\[
\mathbf{y}_k^{(j)} = \sum_{i} \sum_{m=0}^{\nu-1} \mathbf{h}_m^{(i,j)} x_{k-m}^{(i)} + \mathbf{n}_k^{(j)} \tag{1}
\]

where \( \mathbf{y}_k^{(j)} \) is the \( j \)-th channel output vector, \( \mathbf{h}_m^{(i,j)} \) is the channel impulse response (CIR) between the \( i \)-th input and the \( j \)-th output, whose memory is denoted by \( p^{(i,j)} \), and \( \mathbf{n}_k^{(j)} \) is the noise vector at the \( j \)-th output. All of these three quantities are \( l \times 1 \) column vectors corresponding to the \( l \) time samples per symbol in the assumed temporally-oversampled channel model. By grouping the received samples from all \( n_o \) channel outputs at symbol time \( k \) into a \( n_o \times 1 \) column vector \( \mathbf{y}_k \), we can relate \( \mathbf{y}_k \) to the corresponding \( n_i \times 1 \) column vector of input samples as follows:

\[
\mathbf{y}_k = \sum_{m=0}^{\nu-1} \mathbf{H}_m \mathbf{x}_{k-m} + \mathbf{n}_k \tag{2}
\]

where \( \mathbf{H}_m \) is the MIMO channel matrix coefficient of size \( n_n \times n_i \), and \( \mathbf{x}_{k-m} \) is a size \( n_i \times 1 \) input vector at time \( k - m \). The parameter \( \nu \) is the maximum length of all of the \( n_o n_i \) CIRs, i.e., \( \nu = \max_i \nu^{(i,j)} \).

Over a block of \( N_f \) symbol periods, (2) can be expressed in matrix notation as follows:

\[
\begin{bmatrix}
\mathbf{y}_{k+N_f-1} \\
\mathbf{y}_{k+N_f-2} \\
\vdots \\
\mathbf{y}_k
\end{bmatrix}
= 
\begin{bmatrix}
\mathbf{H}_0 & \mathbf{H}_1 & \cdots & \mathbf{H}_\nu & 0 & \cdots & 0 \\
0 & \mathbf{H}_0 & \mathbf{H}_1 & \cdots & \mathbf{H}_\nu & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & 0 & \mathbf{H}_0 & \mathbf{H}_1 & \cdots & \mathbf{H}_\nu
\end{bmatrix}
\begin{bmatrix}
\mathbf{x}_{k+N_f-1} \\
\mathbf{x}_{k+N_f-2} \\
\vdots \\
\mathbf{x}_{k-\nu}
\end{bmatrix}
+ 
\begin{bmatrix}
\mathbf{n}_{k+N_f-1} \\
\mathbf{n}_{k+N_f-2} \\
\vdots \\
\mathbf{n}_k
\end{bmatrix}
\tag{3}
\]

or more compactly

\[
\mathbf{y}_{k+N_f-1: k} = \mathbf{H} \mathbf{x}_{k+N_f-1: k-\nu} + \mathbf{n}_{k+N_f-1: k}. \tag{4}
\]

We define the \( n_i (N_f + \nu) \times n_i (N_f + \nu) \) input autocorrelation matrix \( \mathbf{R}_{xx} \), and the \( n_o (N_f) \times n_o (N_f) \) noise autocorrelation matrix \( \mathbf{R}_{nn} \); both assumed nonsingular. Then, the input–output cross correlation and the output autocorrelation matrices are given by

\[
\begin{align*}
\mathbf{R}_{xy} &\triangleq \mathbb{E} [\mathbf{x}_{k+N_f-1: k-\nu}^H \mathbf{y}_{k+N_f-1: k}] \\
\mathbf{R}_{yy} &\triangleq \mathbb{E} [\mathbf{y}_{k+N_f-1: k}^H \mathbf{y}_{k+N_f-1: k}] = \mathbf{H}^H \mathbf{R}_{xx} \mathbf{H} + \mathbf{R}_{nn}
\end{align*}
\]

B. Problem Formulation

Given the MIMO channel matrix \( \mathbf{H} \) which has \( (\nu + 1) \) matrix taps, our objective is to design an MIMO filter with \( N_f \) matrix taps, which we denote by the \( (n_o N_f \times n_i) \) matrix \( \mathbf{W} \), to equalize \( \mathbf{H} \) to an MIMO target impulse response (TIR) matrix \( \mathbf{B} \), with \( (n_o + 1) \) matrix taps \( \mathbf{B} \) (each of size \( n_i \times n_i \)) where \( N_f \ll \nu \) (see Fig. 1). The MIMO

---

**TABLE I**

**SUMMARY OF KEY MATRICES USED IN THE PAPER AND THEIR SIZES**

<table>
<thead>
<tr>
<th>Matrix</th>
<th>Name</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathbf{H} )</td>
<td>Channel Matrix</td>
<td>( l n_o N_f \times n_i (N_f + \nu) )</td>
</tr>
<tr>
<td>( \mathbf{R}_{xx} )</td>
<td>Input Auto-Correlation Matrix</td>
<td>( n_i (N_f + \nu) \times n_i (N_f + \nu) )</td>
</tr>
<tr>
<td>( \mathbf{R}_{nn} )</td>
<td>Noise Auto-Correlation Matrix</td>
<td>( l n_o N_f \times l n_o N_f )</td>
</tr>
<tr>
<td>( \mathbf{R}_{xy} )</td>
<td>Input–Output Cross-Correlation Matrix</td>
<td>( n_i (N_f + \nu) \times l n_o N_f )</td>
</tr>
<tr>
<td>( \mathbf{R}_{yy} )</td>
<td>Output Auto-Correlation Matrix</td>
<td>( l n_o N_f \times l n_o N_f )</td>
</tr>
<tr>
<td>( \mathbf{W} )</td>
<td>Channel Shortening Equalizer Matrix</td>
<td>( n_i (N_f + \nu) \times n_i )</td>
</tr>
<tr>
<td>( \mathbf{B} )</td>
<td>Augmented Target Impulse Response Matrix</td>
<td>( n_i (N_f + \nu) \times n_i (N_f + \nu) )</td>
</tr>
<tr>
<td>( \mathbf{R} )</td>
<td>Defined in Equation (10)</td>
<td>( n_i (N_f + 1) \times n_i (N_f + 1) )</td>
</tr>
</tbody>
</table>
Fig. 1. Block diagram of the MIMO channel shortening equalizer (dimensions of various signals are indicated between the brackets).

CSE $W$ is optimized to minimize the equalization mean square error (MSE) defined by $\text{MSE} \triangleq \text{trace}(R_{ee})$, where $R_{ee}$ is the autocorrelation matrix of the equalization error vector $E_k$ given by [3]

$$E_k = \tilde{B}^*x_{k+N_f-1; k-k};$$

For notational convenience, we define the augmented MIMO TIR matrix $\tilde{B}$ as follows:

$$\tilde{B}^{+}[B_0^\dagger B_1^\dagger \ldots B_{N_b}^\dagger 0_{n_i \times n_s}^\dagger]$$

where $\Delta$ is the decision delay that lies in the range $0 \leq \Delta \leq (N_f + \nu - N_b - 1)$ and $s^{\dagger}[N_f + \nu - N_b - \Delta - 1].$

Using the Orthogonality Principle, it can be shown [3] that the optimum MIMO CSE and TIR filters are related by

$$W_{q_k}^* \triangleq B_{k}^{-1}R_{xx}^{-1}R_{yy}^{-1}$$

The last line above shows explicitly that the MIMO CSE consists of a noise whitener $R_{xx}^{-1}$, an MIMO matched filter $H^*$, and a bank of FIR channel-shortening filters. It remains to optimize $B$ such that MSE is minimized. This optimization is carried out in the next section under two criteria.

III. MIMO CHANNEL SHORTENING ALGORITHMS

The $n_i \times n_i$ error autocorrelation matrix $R_{ee}$ can be expressed as follows:

$$R_{ee}^{\dagger} = E[E_k E_k^\dagger]$$

where $\tilde{R}$ is a submatrix of $R^\dagger$ determined by $\Delta$. The optimum MIMO TIR is determined by computing $B$ that minimizes the trace (or determinant) of $R_{ee}$ subject to some constraint on its matrix coefficients. We consider two constraints.

A. Identity Tap Constraint (ITC)

Under the ITC, we restrict the $m$th matrix coefficient of $B$ to be equal to the identity matrix. Therefore, we solve the following optimization problem

$$B_{\text{opt}}^{\text{ITC}} = \arg\min_{B} \text{trace}(R_{ee}) \quad \text{subject to} \quad B^*B = I_{n_i}$$

where $B^*B = I_{n_i}$. It can be shown [11] that the optimum MIMO TIR and the corresponding error autocorrelation matrix are given by

$$B_{\text{opt}}^{\text{ITC}} = R^{-1}\hat{\Phi}(\hat{\Phi}^*R^{-1}\hat{\Phi})^{-1}$$

B. Orthonormality Constraint (ONC)

Under the ONC, we constrain $B$ to have orthonormal rows, i.e., $B^*B = I_{n_i}$. Therefore, we solve the following optimization problem:

$$B_{\text{opt}}^{\text{ONC}} = \arg\min_{B} \text{trace}(R_{ee}) \quad \text{subject to} \quad B^*B = I_{n_i}$$

If we define the eigendecomposition

$$R \equiv U\Sigma U^* = \text{Diag}(\sigma_0, \ldots, \sigma_{n_i(N_b+1)-1})U^*$$

where $\sigma_0 \geq \sigma_1 \geq \ldots \geq \sigma_{n_i(N_b+1)-1}$, then the optimum MIMO TIR and the resulting error autocorrelation matrix are given by (see Appendix for a proof)

$$B_{\text{opt}}^{\text{ONC}} = U[e_0, e_1, \ldots, e_{n_i(N_b+1)-1}]$$

$$R_{ee,\text{opt}} = \text{Diag}(\sigma_0, \sigma_1, \ldots, \sigma_{n_i(N_b+1)-1})$$

In words, the optimum MIMO TIR matrix is given by the $n_i$ eigenvectors of $\tilde{R}$ that correspond to its $n_i$ smallest eigenvalues. The delay parameter $\Delta (0 \leq \Delta \leq N_f + \nu - N_b - 1)$ that affects $\tilde{R}$ is optimized to minimize the trace (or determinant) of $R_{ee,\text{opt}}$. Under both criteria, the performance measure adopted in this paper is equalization SNR defined as follows:

$$\text{SNR}_{eq} \triangleq \frac{1}{\eta_i(N_f + \nu)} \text{trace}(R_{xx})$$

We conclude this section by noting that several equalization structures follow as special cases of the finite-length MIMO CSE presented here including:

4 It can be shown [3] that the same $B_{\text{opt}}$ minimizes the trace and the determinant of $R_{ee}$.

5 This constraint implies that the energy of the aggregate CIR into any particular output $j$ formed by concatenating the CIRs from all $n_i$ inputs to output $j$ is equal to unity for all $1 \leq j \leq n_o$. Moreover, these aggregate CIRs are orthogonal for different outputs.
Fig. 2. Variation of the equalization SNR of the MIMO CSE under the ONC and ITC constraints versus delay for $\nu = 6$, $N_b = 2$, and $N_f = 12$.

- the finite-length MIMO MMSE-DFE studied in [3] follows as a special case of ITC by setting $B_0 = I_{n_t}$ and $\Phi = [I_{n_t} \ 0_{n_t \times n_t \times N_b}]^t$;
- the finite-length MIMO MMSE-LE follows as a special case of ITC by setting $B_0 = I_{n_t}$ and $B_i = 0_{n_t \times n_t}$ ($1 \leq i \leq N_b$);
- the finite-length MIMO partial response LE follows as a special case by setting $B$ equal to a desired fixed TIR;
- the finite-length single-user receive-diversity-combining MMSE-DFE studied in [13], [12] follows as a special case of ITC by setting $n_f = 1$ and $B_0 = 0$;
- the finite-length MMSE CSE for SISO systems (see [2] and the references therein) follows as a special case by setting $n_t = n_o = 1$. In this case, the ITC and the ONC criteria become identical to the unit-tap and the unit-energy constraints of [2], respectively.

C. Computational Complexity

In this section, we enumerate the computational tasks required to compute $B_{\text{opt}}^{\text{ONC}}$ and $W_{\text{opt}}$ under both ONC and ITC and their complexities as measured by the number of complex multiplies. We assume uncorrelated input and noise processes.

1) Computation of $R$ requires inversion of the block-Toeplitz matrix $R_{yf}$ with a complexity of $O(\nu(N_b^3 + N_f^3))$ operations using the efficient algorithm in [1]. Computation of the matrix product $R_{xy}R_{yf}^{-1}R_{gf}$ requires $n_t(N_f + \nu)(n_o N_f)^2$ complex multiplies.

2) Under ITC, computing $B_{\text{opt}}^{\text{ITC}}$ using (12) requires $O((n_t(N_b + 1))^3)$ operations to compute $(R)^{-1}$, $O(n_t^3)$ operations to compute $(\Phi^t R^{-1} \Phi)^{-1}$, and $n_t^3(N_b + 1)$ complex multiplies to compute $B_{\text{opt}}^{\text{ITC}}$.

3) Under ONC, the complexity of computing $B_{\text{opt}}^{\text{ONC}}$ using (16) is dominated by the complexity of the eigendecomposition in (15) which requires $O(n_t^3(N_b + 1)^3)$ operations.

4) Finally, computing $W_{\text{opt}}$ using (9) requires $n_t(N_f + \nu)(n_o N_f)^2 + n_t^3(N_b + 1)^2$ complex multiplies.

It is worth mentioning that this complexity estimate assumes a given value of $\Delta$, i.e., the complexity of searching for the optimum $\Delta$ is not included. We found that the complexities in the MIMO CSE and TIR computation under both ONC and ITC were comparable.

IV. NUMERICAL RESULTS

The CIRs used in our numerical results are unit-energy symbol-spaced (i.e., $l = 1$) FIR filters with seven taps (i.e., $\nu = 6$) generated as complex zero-mean uncorrelated Gaussian random variables. The input and noise processes are assumed to be uncorrelated. The performance results are calculated by averaging over 100 channel realizations. Fig. 2 depicts the variation of the equalization SNR, as defined by (18), versus delay under both ONC and ITC constraints. We assume a two-input two-output system with input SNR equal to 20 dB on all four channels. The CIRs are equalized to length-3 TIRs (i.e., $N_b = 2$) using an MIMO CSE with $N_f = 12$ and $l = 1$. Three main observations can be made based on Fig. 2. First, a suboptimum choice of the delay parameter $\Delta$ can result in substantial performance degradation. Second, at any given delay, using the ONC results in better performance than the ITC. Third, as long as $\Delta$ is optimized, the effect of optimizing the index parameter under the ITC on performance is marginal. In Figs. 3 and 4, we plot the variation of the equalization SNR versus $N_f$ and...
Fig. 3. Variation of the equalization SNR of the MIMO CSE under the ONC constraint versus $N_f$ for $\nu = 6$ and $N_b = 2$.

Fig. 4. Variation of the equalization SNR of the MIMO CSE under the ONC constraint versus $N_b$ for $\nu = 6$ and $N_f = 12$.

$N_b$ under ONC with optimized delay. As expected, increasing $N_f$ and $N_b$ results in improved performance which comes at the expense of the increased complexity in implementing the additional equalizer taps and the increased number of states in the MLSE receiver or increased cyclic prefix overhead in DMMT.

V. CONCLUSIONS

We derived a general framework for analyzing finite-length MIMO MMSE equalizers that shorten the memory of linear frequency-selective MIMO channels in addition to performing noise whitening and multichannel matched filtering. Closed-form
expressions for the optimum FIR delay-optimized MIMO equalizer and MIMO TIR were derived under an orthonormality and an identity-tap constraint on the MIMO TIR. The orthonormality constraint was shown to result in a lower equalization mean square error. The presented framework accommodates fractionally-spaced equalizers, arbitrary decision delay setting, colored noise, transmit/receive diversity, and variable multipath delay spread across the different diversity paths. Applications to turbo equalization are discussed in [5] and [4].

PROOF OF (16) AND (17)

Starting from (14) and (15), we have

$$\text{trace}(R_{ee}) = \sum_{k=0}^{n_t-1} \frac{\rho_k}{\sigma_k^2} \sum_{i=0}^{n_r(N_r+1)-1} \epsilon_i^T B^T U \Sigma U^T \epsilon_i \| b_k \|^2 = \sum_{k=0}^{n_t-1} \frac{\rho_k}{\sigma_k^2} \sum_{i=0}^{n_r(N_r+1)-1} \epsilon_i^T U^T \epsilon_i \| b_k \|^2 = \sum_{k=0}^{n_t-1} \frac{\rho_k}{\sigma_k^2} \sum_{i=0}^{n_r(N_r+1)-1} \epsilon_i^T U^T \epsilon_i b_k b_k^T U \Sigma U^T \epsilon_i \| b_k \|^2$$

where $b_k^T \equiv B e_k$. Therefore

$$\text{trace}(R_{ee}) = \sum_{k=0}^{n_t-1} \frac{\rho_k}{\sigma_k^2} \sum_{i=0}^{n_r(N_r+1)-1} \epsilon_i^T U e_k^T U^T e_i \| b_k \|^2$$

At any given $\Delta$, to minimize this sum, we must choose $b_k$ to minimize each term $\text{MSE}_k$. Since the eigenvalues $\sigma_i$ are arranged in descending order, it follows that $\text{trace}(R_{ee})$ is minimized by setting $b_k = U e_{(n_rN_r)+k}$ which is equivalent to (16) and results in

$$\text{trace}(R_{ee,\text{opt}}) = \sum_{k=0}^{n_t-1} \frac{\rho_k}{\sigma_k^2} \sum_{i=0}^{n_r(N_r+1)-1} \epsilon_i^T U e_k^T U^T e_i \| b_{(n_rN_r)+k} \|^2$$

which is identical to (17).

ACKNOWLEDGMENT

The author would like to thank A. F. Naguib, S. N. Diggavi, R. Jana, V. Tarokh, and A. R. Calderbank of AT&T Shannon Laboratory, N. Seshadri of Broadcom, and G. Bauch of Munich University of Technology for many interesting discussions on the subject of this paper.

REFERENCES