

# A General Framework for the Design and Analysis of Sparse FIR Linear Equalizers

Abubakr O. Al-Abbasi\*, Ridha Hamila\*, Waheed U. Bajwa†, and Naofal Al-Dhahir‡

\* Dept. of Electrical Engineering, Qatar University, Qatar

† Dept. of Electrical and Computer Engineering, Rutgers University, USA

‡ Dept. of Electrical Engineering, University of Texas at Dallas, USA

**Abstract**—Complexity of linear finite-impulse-response (FIR) equalizers is proportional to the square of the number of nonzero taps in the filter. This makes equalization of channels with long impulse responses using either zero-forcing or minimum mean square error (MMSE) filters computationally expensive. Sparse equalization is a widely-used technique to solve this problem. In this paper, a general framework is provided that transforms the problem of sparse linear equalizers (LEs) design into the problem of sparsest-approximation of a vector in different dictionaries. In addition, some possible choices of sparsifying dictionaries in this framework are discussed. Furthermore, the worst-case coherence of some of these dictionaries, which determines their sparsifying strength, are analytically and/or numerically evaluated. Finally, the usefulness of the proposed framework for the design of sparse FIR LEs is validated through numerical experiments.

## I. INTRODUCTION

In numerous signal processing applications such as equalization and interference cancellation, long FIR filters have to be implemented at high sampling rates. This results in high complexity, which grows proportional to the square of the number of nonzero taps. One approach to reduce this complexity is to implement only the most significant FIR filter taps, i.e., sparse filters. However, reliably determining the locations of these dominant taps is often very challenging.

Several design approaches have been investigated in the literature to reduce the complexity of long FIR filters. In [1], the number of nonzero coefficients is reduced by selecting only the significant taps of the equalizer. Nonetheless, knowledge of the whole equalizer tap vector is required which increases the computational complexity. In [2], an  $\ell_1$ -norm minimization problem is formulated to design a sparse filter. However, since the resulting filter taps are not exactly sparse, a strict thresholding step is required to force some of the nonzero taps to 0. An algorithm, called sparse chip equalizer, for finding the locations of sparse equalizer taps is given in [3] but this approach assumes that the channel itself is sparse. In [4], a general optimization problem for designing a sparse filter is formulated that involves a quadratic constraint on filter performance. Nonetheless, the number of iterations of the proposed backward selection algorithm becomes large as the desired sparsity of the filter increases. In addition, the approach in [4] also involves inversion of a large matrix in the case of long Channel Impulse Responses (CIRs). In [5],

a framework for designing sparse FIR equalizers is proposed. Using greedy algorithms, the proposed framework achieved better performance than just choosing the largest taps of the MMSE equalizer, as in [1]. However, this approach involves Cholesky factorization, whose computational cost could be large in the case of channels with large delay spreads. In addition, no theoretical guarantees are provided.

In this paper, we develop a general framework for the design of sparse FIR equalizers that transforms the original problem into one of sparse approximation of a vector using different dictionaries. The developed framework can then be used to find the sparsifying dictionary that leads to the sparsest FIR filter subject to an approximation constraint. We also investigate the coherence of the sparsifying dictionaries that we propose as part of our analysis and identify one that has the smallest coherence. Then, we use simulations to validate that the dictionary with the smallest coherence gives the sparsest FIR linear equalizer. Moreover, the numerical results demonstrate the significance of our approach compared to conventional sparse FIR equalizers (e.g., [1]) in terms of both performance and computational complexity.

**Notations:** We use the following standard notation in this paper:  $\mathbf{I}_N$  denotes the identity matrix of size  $N$ . Upper and lower case bold letters denote matrices and vectors, respectively. The notations  $(\cdot)^{-1}$ ,  $(\cdot)^*$  and  $(\cdot)^H$  denote the matrix inverse, the matrix (or element) complex conjugate and the complex-conjugate transpose operations, respectively.  $E[\cdot]$  denotes the expected value operator. The components of a vector starting from  $k_1$  and ending at  $k_2$  are given as subscripts to the vector separated by a colon, i.e.,  $\mathbf{x}_{k_1:k_2}$ .

## II. SYSTEM MODEL

A linear, time invariant, dispersive and noisy communication channel is considered. The standard complex-valued equivalent baseband signal model is assumed. At time  $k$ , the received sample  $y_k$  can be expressed as

$$y_k = \sum_{l=0}^v h_l x_{k-l} + n_k, \quad (1)$$

where  $h_l$  is the CIR whose memory is  $v$ ,  $n_k$  is the additive noise symbol and  $x_{k-l}$  is the transmitted symbol at time  $(k-l)$ . At any time  $k$ , an FIR filter of length  $N_f$  is applied to the received samples in order to recover the transmitted symbols with some possible time delay. For simplicity, we assume a symbol-spaced equalizer but our proposed design framework can be easily extended to the general fractionally-spaced case.

This paper was made possible by grant number NPRP 06-070-2-024 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

For these  $N_f$ -long received samples of interest, the input-output relation in (1) can be written compactly as

$$\mathbf{y}_{k:k-N_f+1} = \mathbf{H} \mathbf{x}_{k:k-N_f-v+1} + \mathbf{n}_{k:k-N_f+1}, \quad (2)$$

where  $\mathbf{y}_{k:k-N_f+1}$ ,  $\mathbf{x}_{k:k-N_f-v+1}$  and  $\mathbf{n}_{k:k-N_f+1}$  are column vectors grouping the received, transmitted and noise samples. Additionally,  $\mathbf{H}$  is an  $N_f \times (N_f + \nu)$  Toeplitz matrix whose first row is formed by  $\{h_l\}_{l=0}^{L-v}$  followed by zero entries. It is useful, as will be shown in the sequel, to define the output auto-correlation and the input-output cross-correlation matrices based on the block of length  $N_f$ . Using (2), the input correlation and the noise correlation matrices are, respectively, defined by  $\mathbf{R}_{xx} \triangleq E \left[ \mathbf{x}_{k:k-N_f-v+1} \mathbf{x}_{k:k-N_f-v+1}^H \right]$  and  $\mathbf{R}_{nn} \triangleq E \left[ \mathbf{n}_{k:k-N_f+1} \mathbf{n}_{k:k-N_f+1}^H \right]$ . Both the input and noise processes are assumed to be white; hence, their auto-correlation matrices are assumed to be (multiples of) the identity matrix, i.e.,  $\mathbf{R}_{xx} = \mathbf{I}_{N_f+v}$  and  $\mathbf{R}_{nn} = \frac{1}{SNR} \mathbf{I}_{N_f}$ . Moreover, the output-input cross-correlation and the output auto-correlation matrices are, respectively, defined as

$$\mathbf{R}_{yx} \triangleq E \left[ \mathbf{y}_{k:k-N_f+1} \mathbf{x}_{k:k-N_f-v+1}^H \right] = \mathbf{H} \mathbf{R}_{xx}, \quad \text{and} \quad (3)$$

$$\mathbf{R}_{yy} \triangleq E \left[ \mathbf{y}_{k:k-N_f+1} \mathbf{y}_{k:k-N_f+1}^H \right] = \mathbf{H} \mathbf{R}_{xx} \mathbf{H}^H + \mathbf{R}_{nn}. \quad (4)$$

### III. SPARSE FIR LINEAR EQUALIZERS DESIGN

#### A. Initial formulation

The received samples are passed through an FIR filter with length  $N_f$ . Hence, the error symbol at time  $k$  is given by

$$e_k = x_{k-\Delta} - \hat{x}_k = x_{k-\Delta} - \mathbf{w}^H \mathbf{y}_{k:k-N_f+1}, \quad (5)$$

where  $\Delta$  is the decision delay, typically  $0 \leq \Delta \leq N_f + v - 1$ , and  $\mathbf{w}$  denotes the equalizer taps vector whose dimension is  $N_f \times 1$ . Using the orthogonality principle of linear least-squares estimation, the MSE, denoted as  $\xi(\mathbf{w})$ , equals [5]

$$\xi(\mathbf{w}) \triangleq E \left[ |e_k|^2 \right] = \varepsilon_x - \mathbf{w}^H \mathbf{R}_{yx} - \mathbf{R}_{yx}^H \mathbf{w} + \mathbf{w}^H \mathbf{R}_{yy} \mathbf{w},$$

where  $\varepsilon_x \triangleq E \left[ x_{k-\Delta}^2 \right]$ . By writing  $x_{k-\Delta} = \mathbf{1}_\Delta^H \mathbf{x}_{k:k-N_f-v+1}$  and  $\mathbf{r}_\Delta = \mathbf{R}_{yx} \mathbf{1}_\Delta$ , where  $\mathbf{1}_\Delta$  denotes  $(N_f + v)$ -dimensional vector that is zero everywhere except in the  $(\Delta + 1)$ -th element where it is one, it follows that

$$\xi(\mathbf{w}) = \underbrace{\varepsilon_x - \mathbf{r}_\Delta^H \mathbf{R}_{yy}^{-1} \mathbf{r}_\Delta}_{\xi_m} + \underbrace{(\mathbf{w} - \mathbf{R}_{yy}^{-1} \mathbf{r}_\Delta)^H \mathbf{R}_{yy} (\mathbf{w} - \mathbf{R}_{yy}^{-1} \mathbf{r}_\Delta)}_{\xi_e(\mathbf{w})}. \quad (6)$$

Since  $\xi_m$  does not depend on  $\mathbf{w}$ , the MSE  $\xi(\mathbf{w})$  is minimized by minimizing the term  $\xi_e(\mathbf{w})$ . Hence, the optimum selection for  $\mathbf{w}$ , in the MMSE sense, is the well-known Wiener solution  $\mathbf{w}_{opt} = \mathbf{R}_{yy}^{-1} \mathbf{r}_\Delta$ . However, in general, this optimum choice is undesirable since  $\mathbf{w}_{opt}$  is not sparse and its implementation complexity increases proportional to  $(N_f)^2$  which can be computationally expensive [6]. However, any choice for  $\mathbf{w}$  other than  $\mathbf{w}_{opt}$  increases  $\xi_e(\mathbf{w})$ , which leads to performance loss. This suggests that we can use the excess error  $\xi_e(\mathbf{w})$  as a design constraint to achieve a desirable performance-complexity tradeoff. Specifically, we formulate the following problem for the design of sparse FIR equalizers:

$$\hat{\mathbf{w}}_s \triangleq \underset{\mathbf{w} \in \mathbb{C}^{N_f}}{\text{argmin}} \|\mathbf{w}\|_0 \quad \text{subject to} \quad \xi_e(\mathbf{w}) \leq \delta_{eq}, \quad (7)$$

where  $\|\mathbf{w}\|_0$  is the number of nonzero elements in its argument,  $\|\cdot\|_2$  denotes the  $\ell_2$ -norm and  $\delta_{eq}$  can be chosen as a

Table I  
EXAMPLES OF DIFFERENT SPARSIFYING DICTIONARIES.

Cholesky Factorization			Eigen Decomposition		
$\mathbf{R}_{yy} = \mathbf{L}\mathbf{L}^H$ or $\mathbf{R}_{yy} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^H$			$\mathbf{R}_{yy} = \mathbf{U}\mathbf{D}\mathbf{U}^H$		
$\mathbf{A}$	$\mathbf{\Phi}$	$\mathbf{b}$	$\mathbf{A}$	$\mathbf{\Phi}$	$\mathbf{b}$
$\mathbf{I}$	$\mathbf{L}^H$	$\mathbf{L}^{-1} \mathbf{r}_\Delta$	$\mathbf{I}$	$\mathbf{D}^{\frac{1}{2}} \mathbf{U}^H$	$\mathbf{D}^{-\frac{1}{2}} \mathbf{U}^H \mathbf{r}_\Delta$
$\mathbf{L}^{-1}$	$\mathbf{R}_{yy}$	$\mathbf{r}_\Delta$	$\mathbf{D}^{-\frac{1}{2}} \mathbf{U}^H$	$\mathbf{R}_{yy}$	$\mathbf{r}_\Delta$
$\mathbf{I}$	$\mathbf{\Lambda}^{\frac{1}{2}} \mathbf{P}^H$	$\mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{P}^{-1} \mathbf{r}_\Delta$	$\mathbf{D}^{\frac{1}{2}}$	$\mathbf{U}^H$	$\mathbf{D}^{-1} \mathbf{U}^H \mathbf{r}_\Delta$

function of the noise variance. While one can attempt to use convex-optimization-based approaches (after replacing  $\|\cdot\|_0$  with its convex approximation  $\|\cdot\|_1$  in (7) to reduce the search space and to make it more tractable [7]) in order to estimate the sparse approximation vector  $\hat{\mathbf{w}}_s$ , there exists a number of greedy algorithms with low complexity that can be used in an efficient manner. Starting with this initial formulation, we now discuss a general framework for sparse FIR LEs design such that the performance loss does not exceed a predefined limit.

#### B. Proposed sparse approximation framework

Unlike earlier works, including the one by one of the co-authors [5], we provide a general framework for designing sparse FIR linear equalizers that can be considered as the problem of sparse approximation using different dictionaries. Mathematically, this framework poses the problem of sparse FIR equalizers design as follows:

$$\hat{\mathbf{w}}_s \triangleq \underset{\mathbf{w} \in \mathbb{C}^{N_f}}{\text{argmin}} \|\mathbf{w}\|_0 \quad \text{subject to} \quad \|\mathbf{A}(\mathbf{\Phi}\mathbf{w} - \mathbf{b})\|_2^2 \leq \delta_{eq}, \quad (8)$$

where  $\mathbf{\Phi}$  is the dictionary that will be used to sparsely approximate  $\mathbf{b}$ , while  $\mathbf{A}$  is a known matrix and  $\mathbf{b}$  is a known data vector, both of which change depending upon the sparsifying dictionary  $\mathbf{\Phi}$ . Note that by completing the square in (7), the problem reduces to the one shown in (8). Hence, one can use any decomposition for  $\mathbf{R}_{yy}$  to come up with a sparse approximation problem. By writing the Cholesky or eigenvalue decompositions for  $\mathbf{R}_{yy}$ , we can have different choices for  $\mathbf{A}$ ,  $\mathbf{\Phi}$  and  $\mathbf{b}$ . Some of these possible choices are shown in Table I. Note that the framework parameters (i.e.,  $\mathbf{A}$ ,  $\mathbf{\Phi}$  and  $\mathbf{b}$ ) in the left list of Table I result by defining the Cholesky factorization [8] either in the form  $\mathbf{R}_{yy} \triangleq \mathbf{L}\mathbf{L}^H$  or  $\mathbf{R}_{yy} \triangleq \mathbf{P}\mathbf{\Lambda}\mathbf{P}^H$  (where  $\mathbf{L}$  is a lower-triangular matrix,  $\mathbf{P}$  is a lower-unit-triangular (unitriangular) matrix and  $\mathbf{\Lambda}$  is a diagonal matrix). On the other hand, the columns on the right result by letting  $\mathbf{R}_{yy} \triangleq \mathbf{U}\mathbf{D}\mathbf{U}^H$ , where  $\mathbf{U}$  is a unitary matrix whose columns are the eigenvectors of the matrix  $\mathbf{R}_{yy}$  and  $\mathbf{D}$  is a diagonal matrix with the corresponding eigenvalues on the diagonal. For instance, by assuming  $\mathbf{L}^H$ ,  $\mathbf{D}^{\frac{1}{2}} \mathbf{U}^H$  and  $\mathbf{R}_{yy}$  as sparsifying dictionaries, the problem in (8) can, respectively, take one of the forms shown below

$$\min_{\mathbf{w} \in \mathbb{C}^{N_f}} \|\mathbf{w}\|_0 \quad \text{s.t.} \quad \|(\mathbf{L}^H \mathbf{w} - \mathbf{L}^{-1} \mathbf{r}_\Delta)\|_2^2 \leq \delta_{eq}, \quad (9)$$

$$\min_{\mathbf{w} \in \mathbb{C}^{N_f}} \|\mathbf{w}\|_0 \quad \text{s.t.} \quad \left\| \left( \mathbf{D}^{\frac{1}{2}} \mathbf{U}^H \mathbf{w} - \mathbf{D}^{-\frac{1}{2}} \mathbf{U}^H \mathbf{r}_\Delta \right) \right\|_2^2 \leq \delta_{eq}, \quad \text{and} \quad (10)$$

$$\min_{\mathbf{w} \in \mathbb{C}^{N_f}} \|\mathbf{w}\|_0 \quad \text{s.t.} \quad \|\mathbf{L}^{-1} (\mathbf{R}_{yy} \mathbf{w} - \mathbf{r}_\Delta)\|_2^2 \leq \delta_{eq}. \quad (11)$$

Note that we can reduce the decomposition complexity by approximating, for reasonably large  $N_f$ , the Toeplitz  $\mathbf{R}_{yy}$  by a circulant matrix whose eigenvectors are the Discrete Fourier Transform (DFT) vectors and eigenvalues are the output discrete spectrum of its first column [9]. For a Toeplitz matrix, the most efficient algorithms for Cholesky factorization

are Levinson or Schur algorithms [10], which involve  $\mathcal{O}(N_f^2)$  computations. In contrast, the eigen-decomposition of a circulant matrix can be done efficiently using the fast Fourier transform (FFT) and its inverse with only  $\mathcal{O}(N_f \log(N_f))$  operations.

The preceding discussion shows that the problem of designing sparse FIR equalizers can be cast into one of sparse approximation of a vector by a fixed dictionary. The general form of this problem is given by (8). To solve this problem, we use the well-known Orthogonal Matching Pursuit (OMP) greedy algorithm [11] that estimates  $\hat{\mathbf{w}}_s$  by iteratively selecting a set  $S$  of the sparsifying dictionary columns (i.e., atoms  $\phi'_{iS}$ ) of  $\Phi$  that are most correlated with the data vector  $\mathbf{b}$  and then solving a restricted least-squares problem using the selected atoms. The OMP stopping criterion ( $\rho$ ) is changed here from an upper-bound on the residual error to an upper-bound on the Projected Residual Error (PRE), i.e., “ $\mathbf{A} \times$  Residual Error”. The computations involved in the OMP algorithm are well documented in the sparse approximation literature (e.g., [11]) and are omitted here due to page limitations.

Unlike conventional compressive sensing [12], where the measurement matrix is a fat matrix, the sparsifying dictionary in our framework is a square one with full rank. However, OMP and similar methods can still be used if  $\mathbf{R}_{yy}$  can be decomposed into  $\mathbf{Q}\mathbf{Q}^H$  and the data vector  $\mathbf{b}$  is compressible [13], [4]. Among the proposed dictionaries shown in Table I, only  $\mathbf{U}^H$  is not a valid choice of  $\Phi$  since the data vector  $\mathbf{b}$  associated with it can not be compressed into a lower dimensional space without significant information loss and, in addition, its PRE is large. Notice that it is better to keep the PRE as small as possible to limit the amount of noise in the data.

Our next challenge is to determine the best sparsifying dictionary for use in our framework. We know from the sparse approximation literature that the sparsity of the OMP solution tends to be inversely proportional to the worst-case coherence  $\mu(\Phi)$ ,  $\mu(\Phi) \triangleq \max_{i \neq j} \frac{|\langle \phi_i, \phi_j \rangle|}{\|\phi_i\|_2 \|\phi_j\|_2}$  [14], [15]. Notice that  $\mu(\Phi) \in [0, 1]$ . Next, we investigate the coherence of the dictionaries proposed in Table I.

### C. Worst-Case Coherence Analysis

We carry out a coherence metric analysis to gain some insight into the performance of the proposed sparsifying dictionaries and the behavior of the resulting sparse equalizers. First and foremost, we are concerned with analyzing  $\mu(\Phi)$  to ensure that it does not approach 1 for the proposed sparsifying dictionaries. In addition, we are interested in identifying which  $\Phi$  has the smallest coherence and, hence, gives the sparsest FIR equalizer. We proceed as follows. We estimate an upper bound on the worst-case coherence of  $\mathbf{R}_{yy}$  and evaluate its closeness to 1. Then, through simulation we show that the coherence of other dictionaries, which can be considered as the square roots of  $\mathbf{R}_{yy}$  in the spectral-norm sense, i.e.,  $\|\mathbf{R}_{yy}\|_2 = \|\mathbf{L}\mathbf{L}^H\|_2 \leq \|\mathbf{L}\|_2^2$ ,  $\|\mathbf{R}_{yy}\|_2 \leq \|\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{P}^H\|_2^2$  and  $\|\mathbf{R}_{yy}\|_2 \leq \|\mathbf{D}^{\frac{1}{2}}\mathbf{U}^H\|_2^2$ , will be less than that of  $\mu(\mathbf{R}_{yy})$ . Interestingly,  $\mathbf{R}_{yy}$  has a well-structured (Hermitian Toeplitz) closed form in terms of the CIR coefficients, filter time span

$N_f$  and SNR, i.e.,  $\mathbf{R}_{yy} = \mathbf{H}\mathbf{H}^H + \frac{1}{\text{SNR}}\mathbf{I}$ . It can be expressed in a matrix form as

$$\mathbf{R}_{yy} = \text{Toeplitz}(\overbrace{[\ r_0 \ r_1 \ \dots \ r_v \ 0 \ \dots \ 0 \ ]}^{\phi_1^H}), \quad (12)$$

where  $r_0 = \sum_{i=0}^v |h_i|^2 + (\text{SNR})^{-1}$ ,  $r_j = \sum_{i=j}^v h_i h_{i-j}^*$ ,  $\forall j \neq 0$ .

Assuming high SNR, we can compute  $\mu(\mathbf{R}_{yy})$  in terms of the channel taps only. By noting that the columns of  $\mathbf{R}_{yy}$  are fully defined by the first column, we can get the maximum possible absolute inner product  $\mu(\mathbf{R}_{yy})$  by simultaneously maximizing the entries of  $\phi_1$ , which results in maximizing all columns entries accordingly. While we can pose the problem of computing  $\mu(\mathbf{R}_{yy})$  in terms of maximizing the sum of the inner product  $\langle \phi_i, \phi_j \rangle$ ,  $\forall i \neq j$ , it turns out that it is equivalent to maximizing  $r_1$  due to the special structure of  $\mathbf{R}_{yy}$ . Hence, an upper bound on  $\mu(\mathbf{R}_{yy})$  in the high SNR setting can be derived by solving the following optimization problem

$$\max \sum_{i=1}^v |h_i h_{i-1}^*| \quad \text{s.t.} \quad \sum_{i=0}^v |h_i|^2 = 1. \quad (13)$$

The solution of (13) gives the worst CIR vector  $\mathbf{h}$  which is then used to estimate an upper-bound on  $\mu(\mathbf{R}_{yy})$  for any given channel length  $v$ . This solution has a symmetric structure that can be obtained by solving a simpler equivalent problem formulated as below

$$\max \left| \mathbf{h}^H \mathbf{R} \mathbf{h} \right| \quad \text{s.t.} \quad \mathbf{h}^H \mathbf{h} = 1, \quad (14)$$

where  $\mathbf{h} = [h_0 \ h_1 \ \dots \ h_v]^H$  is the length- $(v+1)$  CIR vector and  $\mathbf{R}$  is a matrix that has ones along the super and sub-diagonals. The solution of (14) is the eigenvector corresponding to the maximum (or minimum, since  $\mu(\mathbf{R}_{yy})$  is defined in terms of absolute value) eigenvalue of  $\mathbf{R}$ . Interestingly, the eigenvalues  $\lambda_s$  and eigenvectors  $\mathbf{h}_j^{(s)}$  of the matrix  $\mathbf{R}$  have the following simple closed forms [16]

$$\lambda_s = 2 \cos\left(\frac{\pi s}{v+2}\right), \quad \mathbf{h}_j^{(s)} = \sqrt{\frac{2}{v+2}} \sin\left(\frac{j\pi s}{v+2}\right), \quad (15)$$

where  $s, j = 1, \dots, v+1$ . Finally, by numerically evaluating  $\mathbf{h}_j^{(s)}$  for the maximum  $\lambda_s$  we find that the worst-case coherence of  $\mathbf{R}_{yy}$  (for any  $v$ ) is sufficiently less than 1, which points to the likely success of OMP in providing the sparsest solution  $\hat{\mathbf{w}}_s$  which is corresponding to the dictionary that has the smallest  $\mu(\mathbf{R}_{yy})$ . Next, we will report the results of our numerical experiments to evaluate the performance of our proposed framework under different sparsifying dictionaries.

## IV. SIMULATION RESULTS

Throughout the simulations, the used CIRs are unit-energy symbol-spaced FIR filters with  $v$  nonzero taps generated as zero-mean uncorrelated complex Gaussian random variables. We assume  $v = 5$  and  $N_f = 35$  [17]. To quantify the performance of the sparsifying dictionaries involved in our analysis in terms of coherence, we plot the worst-case coherence versus the input SNR in Figure 1. Note that a smaller value of  $\mu(\Phi)$  indicates that a reliable sparse approximation is more likely. Clearly,  $\mathbf{R}_{yy}$  has higher  $\mu(\Phi)$  which reflects higher similarities between its columns compared to  $\mathbf{D}^{\frac{1}{2}}\mathbf{U}^H$  and  $\mathbf{L}^H$  (both have the same  $\mu(\Phi)$ ). The coherence increases with SNR up to a certain limit and then saturates. This can be interpreted by

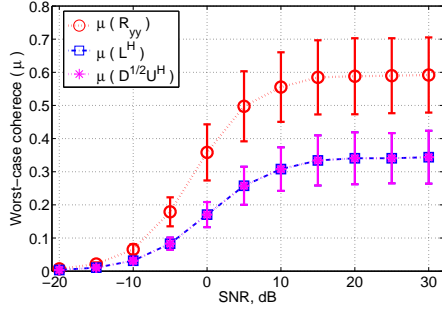


Figure 1. Worst-case coherence for the sparsifying dictionaries versus input SNR. Each point represents the mean of 5000 channel realizations.

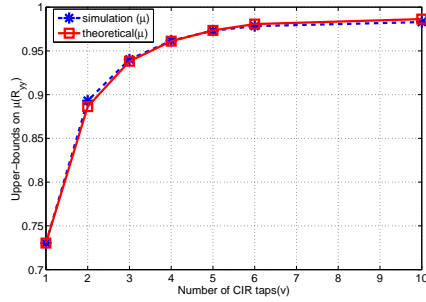


Figure 2. Upper-bounds on  $\mathbf{R}_{yy}$  worst-case coherence versus channel length under unit-energy channel constraint.

the fact that, at high SNR, the noise effects are negligible and, therefore, the sparsifying dictionaries (e.g.,  $\mathbf{R}_{yy} \approx \mathbf{H}\mathbf{H}^H$ ) do not depend on the SNR. Hence, the coherence converges to a constant. In contrast, at low SNR, the noise effects dominate the channel effects. Hence, the channel can be approximated as a memoryless (i.e., 1 tap) channel. Then, the dictionaries (e.g.,  $\mathbf{R}_{yy} \approx \frac{1}{\text{SNR}}\mathbf{I}$ ) can be approximated as a multiple of the identity matrix, i.e.,  $\mu(\Phi) \rightarrow 0$ . Figure 2 shows theoretical bounds, estimated through (15), and empirical upper bounds on the worst-case coherence  $\mu(\mathbf{R}_{yy})$ . This figure shows that the maximum coherence is sufficiently less than 1 and the mismatch between the theoretical and simulation results is negligible (only 0.67%).

We further compare the sparse FIR equalizer designs based on the dictionaries  $\mathbf{D}^{\frac{1}{2}}\mathbf{U}^H$ ,  $\mathbf{L}^H$  and  $\mathbf{R}_{yy}$ , denoted as  $\mathbf{w}_s(\mathbf{D}^{\frac{1}{2}}\mathbf{U}^H)$ ,  $\mathbf{w}_s(\mathbf{L}^H)$  and  $\mathbf{w}_s(\mathbf{R}_{yy})$ , respectively, to study the effect of  $\mu$  on their performance. The case of  $\Phi = \Lambda^{\frac{1}{2}}\mathbf{P}^H$  is not presented here since its performance is almost equivalent to  $\mathbf{L}^H$ . The OMP algorithm is used to compute the sparse approximations. The OMP stopping criterion is set to be a function of the PRE such that: Performance Loss ( $\eta$ ) =  $10 \log_{10} \left( \frac{\text{SNR}(\mathbf{w}_s)}{\text{SNR}(\mathbf{w}_{opt})} \right) \leq 10 \log_{10} \left( 1 + \frac{\delta_{eq}}{\xi_m} \right) \triangleq \eta_{max}$ . Here,  $\delta_{eq}$  is computed based on an acceptable  $\eta_{max}$  and, then, the coefficients of  $\hat{\mathbf{w}}_s$  are computed through (8). The percentage of the active taps is calculated as the ratio between the number of nonzero taps to the total number of filter taps, i.e.,  $N_f$ . For the MMSE equalizer, where none of the coefficients is zero, the number of active filter taps is equal to the filter span. The decision delay is set to be  $\Delta \approx \frac{N_f + v}{2}$  [17].

Figure 3 plots the percentage of the active taps versus the performance loss  $\eta_{max}$  for the proposed sparse FIR-LEs. We observe that a lower active taps percentage is obtained when the coherence of the sparsifying dictionary is small. For instance, allowing for 0.25 dB SNR loss results in a significant reduction in the number of active LE taps. Approximately two-thirds (two-fifths) of the taps are eliminated when using

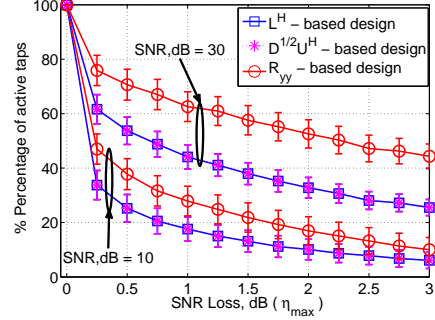


Figure 3. Percentage of active taps versus the performance loss ( $\eta_{max}$ ) for the sparse LEs (5000 channel realizations).

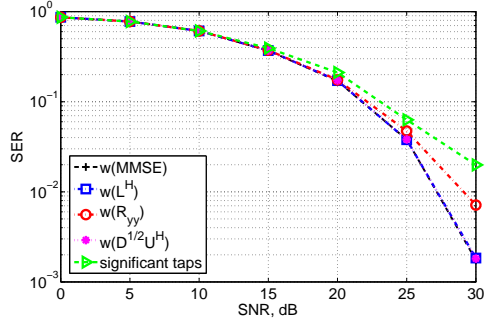


Figure 4. SER comparison between the MMSE non-sparse LE, the proposed sparse LEs  $\mathbf{w}_s(\mathbf{D}^{\frac{1}{2}}\mathbf{U}^H)$ ,  $\mathbf{w}_s(\mathbf{L}^H)$ ,  $\mathbf{w}_s(\mathbf{R}_{yy})$  and the “significant-taps” based LE with sparsity level = 0.25 and 16-QAM modulation.

$\mathbf{w}_s(\mathbf{D}^{\frac{1}{2}}\mathbf{U}^H)$  and  $\mathbf{w}_s(\mathbf{L}^H)$  at SNR equals to 10(30). The sparse LE  $\mathbf{w}_s(\mathbf{R}_{yy})$  needs more active taps to maintain the same SNR loss as that of the other sparse LEs due to its higher coherence. This suggests the smaller the worst-case coherence, the sparser is the equalizer. Moreover, a lower sparsity level (active taps percentage) is achieved at higher SNR levels which is consistent with the previous findings (e.g., in [18]). Furthermore, reducing the number of active taps decreases the filter equalization design complexity and, consequently, the power consumption since a smaller number of complex multiply-and-add operations are required.

In Figure 4, we compare the symbol error rate (SER) performance of our proposed sparse LEs with the proposed approach in [1] which we refer to it as the “significant-taps” approach. In that approach, all of the MMSE LE taps are computed and only the  $K$  significant ones are retained. Assuming a 25% sparsity level, both the  $\mathbf{w}_s(\mathbf{D}^{\frac{1}{2}}\mathbf{U}^H)$  and  $\mathbf{w}_s(\mathbf{L}^H)$  sparse LEs achieve the lowest SER followed by  $\mathbf{w}_s(\mathbf{R}_{yy})$ , while the “significant-taps” performs the worst. In addition to this performance gain, the complexity of the proposed sparse LEs is less than that of the “significant-taps” LE since only an inversion of an  $N_s \times N_s$  matrix is required (not  $N_f \times N_f$  as in the “significant-taps” approach) where  $N_s$  is the number of nonzero taps. Although the  $\mathbf{w}_s(\mathbf{D}^{\frac{1}{2}}\mathbf{U}^H)$  and  $\mathbf{w}_s(\mathbf{L}^H)$  LEs achieve almost the same SER, the former has a lower decomposition complexity since its computation can be done efficiently with only the FFT and its inverse.

## V. CONCLUSIONS

In this paper, we proposed a general framework for sparse FIR equalizer design based on a sparse approximation formulation using different dictionaries. In addition, we investigated the coherence of the proposed dictionaries and showed that the dictionary with the smallest coherence gives the sparsest equalizer design. The significance of our approach was shown analytically and quantified through simulations.

## REFERENCES

- [1] M. Melvasalo, P. Janis, and V. Koivunen, "Sparse equalization in high data rate WCDMA systems," in *IEEE 8th SPAWC*, 2007, pp. 1–5.
- [2] T. Baran, D. Wei, and A. Oppenheim, "Linear programming algorithms for sparse filter design," *IEEE Trans. on Sig. Processing*, vol. 58, no. 3, pp. 1605–1617, 2010.
- [3] G. Kutz and A. Chass, "Sparse chip equalizer for DS-CDMA downlink receivers," *IEEE Commun. Letters*, vol. 9, no. 1, pp. 10–12, 2005.
- [4] D. Wei, C. Sestok, and A. Oppenheim, "Sparse filter design under a quadratic constraint: Low-complexity algorithms," *IEEE Trans. on Sig. Processing*, vol. 61, no. 4, pp. 857–870, 2013.
- [5] A. Gomaa and N. Al-Dhahir, "A new design framework for sparse FIR MIMO equalizers," *IEEE Trans. on Commun.*, vol. 59, no. 8, pp. 2132–2140, 2011.
- [6] J. Proakis and M. Salehi, *Digital Communications, 5th Edition*. New York, NY, USA: McGraw-Hill, 2007.
- [7] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. on Info. Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.
- [8] R. A. Horn and C. R. Johnson, Eds., *Matrix Analysis*. New York, NY, USA: Cambridge University Press, 1986.
- [9] J. Pearl, "On coding and filtering stationary signals by discrete fourier transforms (corresp.)," *IEEE Trans. on Info. Theory*, vol. 19, no. 2, pp. 229–232, 1973.
- [10] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*, 1st ed. New York, NY, USA: John Wiley & Sons, Inc., 1996.
- [11] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. on Info. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [12] D. L. Donoho, "Compressed sensing," *IEEE Trans. on Info. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [13] X. Feng *et al.*, "Sparse equalizer filter design for multi-path channels," Master's thesis, Massachusetts Institute of Technology, 2012.
- [14] W. U. Bajwa and A. Pezeshki, "Finite frames for sparse signal processing," in *Finite Frames*. Springer, 2013.
- [15] J. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. on Info. Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.
- [16] G. H. Golub, "CME 302: Eigenvalues of Tridiagonal Toeplitz Matrices," *Stanford University, USA*.
- [17] J. Cioffi, "EE379A notes Chapter 3," *Stanford University, USA*.
- [18] G. Kutz and D. Raphaeli, "Determination of tap positions for sparse equalizers," *IEEE Trans. on Commun.*, vol. 55, no. 9, pp. 1712–1724, 2007.