

# Sparsifying Dictionary Analysis for FIR MIMO Channel-Shortening 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**—In this paper, we propose a general framework that transforms the problems of designing sparse finite-impulse response channel-shortening equalizers and target impulse response filters for multiple antenna systems into the problems of sparsest-approximation of a vector in different dictionaries. Additionally, we compare several choices of the sparsifying dictionaries in terms of the worst-case coherence metric, which determines their sparsifying effectiveness. Furthermore, a reduced complexity design approach is proposed, which is realized by exploiting the asymptotic equivalence of Toeplitz and circulant matrices. Finally, the significance of our proposed approach is demonstrated through numerical experiments.

## I. INTRODUCTION

In many practical communications settings, the duration of the channel impulse response (CIR) is far too long. This significantly affects the performance and complexity of the communications transceivers. For example, the complexity of maximum-likelihood sequence estimation (MLSE) increases exponentially with the number of users (or streams) and with the memory of the multiple-input multiple-output (MIMO) channel, which makes its use in preventing inter-symbol interference (ISI) prohibitively expensive [1]. Moreover, in block-based multicarrier communications, interblock and intrablock interferences are eliminated by inserting a cyclic prefix in every block whose length is equal to the MIMO channel memory, which in turn can result in a significant reduction of the achievable throughput [2].

Channel shortening, a generalization of equalization, is an elegant solution for dealing with this problem. Channel shortening equalizers (CSEs) are designed to approximate the original channel with a shorter channel, i.e., the combined impulse response of the MIMO channel and the CSE is approximately equivalent to a short MIMO target impulse response (TIR). CSEs have been well investigated in several studies; e.g., [3]–[9]. In [3], conditions for optimum TIR are derived under the unit-tap and the unit-energy constraints. This framework is generalized to MIMO systems in [4]. In [5], the CSE coefficients are inferred blindly from the received data without channel knowledge. In [6], a CSE design approach is proposed to reduce the complexity of trellis detection, which converts the MIMO tree structure into a much smaller trellis, assuming that the channel input is zero-mean Gaussian

process. However, none of these designs impose a sparsity constraint on the CSE design.

Recently, sparse equalizers have received increased attention to reduce implementation cost at acceptable performance loss. In [7], an exhaustive search method is proposed to design a sparse filter. However, its computational cost increases exponentially with the filter order. In [8], using greedy algorithms, a framework for designing sparse CSE and TIR is proposed. This framework achieved better performance by designing the TIR taps to be non-contiguous compared to the designs in [3], where the TIR taps are assumed to be contiguous. However, this approach is limited to single-input single-output (SISO) systems and involves inversion of large matrices and Cholesky factorization, whose computational cost could be large for channels with large delay spreads.

To the best of the authors' knowledge, this paper is the first one to propose sparse MIMO FIR channel-shortening equalizers for MIMO ISI channels. The main contributions of this papers in this context are as follows. First, we extend our framework proposed in [9] to accommodate the more general case of MIMO systems<sup>1</sup>. The extended framework transforms the original problem of designing sparse MIMO filters into one of sparse approximation of a vector using different dictionaries. Then, this framework is used to find the sparsifying dictionary that leads to the sparsest FIR filter, hence reducing its implementation complexity, subject to performance constraints. Second, we numerically study the coherence of the sparsifying dictionaries that we propose as part of our analysis and identify one dictionary that has small coherence. Then, we use simulations to validate that the dictionary with the smallest coherence results in the sparsest FIR design. Finally, numerical results demonstrate the significance of our approach compared to conventional sparse TIR designs, e.g., in [10], in terms of both performance and computational complexity.

**Notations:** We use the following standard notation in this paper:  $I_N$  denotes the identity matrix of size  $N$ . Upper- and lower-case bold letters denote matrices and vectors, respectively. Underlined upper-case bold letters, e.g.,  $\underline{\mathbf{X}}$ , denote frequency-domain vectors. The notations  $(\cdot)^{-1}$ ,  $(\cdot)^*$ ,  $(\cdot)^T$  and  $(\cdot)^H$  denote the matrix inverse, the ma-

<sup>1</sup>We remark that the proposed framework in [4] follows as a special case of the framework proposed here by setting the performance loss to zero (no sparsity constraint), while [9] follows by setting the number of inputs and outputs to be equal to one.

trix (or element) complex conjugate, the matrix transpose and the complex-conjugate transpose operations, respectively.  $E[\cdot]$  denotes the expected value operator.  $\|\cdot\|_\ell$  and  $\|\cdot\|_F$  denote the  $\ell$ -norm and Frobenius norm, respectively.  $\otimes$  denotes the Kronecker product of matrices. 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 schematic of the relationship between CIR, CSE and TIR is shown in Figure 1, and the key matrices used in this paper are summarized in Table I. We consider a linear time-invariant MIMO ISI channel with  $n_i$  inputs and  $n_o$  outputs. The received samples from all  $n_o$  channel outputs at sample time  $k$  are grouped into a  $n_o \times 1$  column vector  $\mathbf{y}_k$  as follows:

$$\mathbf{y}_k = \sum_{l=0}^v \mathbf{H}_l \mathbf{x}_{k-l} + \mathbf{n}_k, \quad (1)$$

where  $\mathbf{H}_l$  is the  $l^{\text{th}}$  channel matrix coefficient of dimension  $(n_o \times n_i)$ , and  $\mathbf{x}_{k-l}$  is the size  $n_i \times 1$  input vector at time  $k-l$ . The parameter  $v$  is the maximum order of all of the  $n_o n_i$  CIRs. Over a block of  $N_f$  output samples, 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, respectively. Note that  $\mathbf{y}_{k:k-N_f+1}$  is a vector of length  $N_f$ , i.e.,  $\mathbf{y} = [y_k \ y_{k-1} \ \dots \ y_{k-N_f+1}]^T$ . Additionally,  $\mathbf{H}$  is a block Toeplitz matrix whose first block row is formed by  $\{\mathbf{H}_l\}_{l=0}^v$  followed by zero matrices. 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 a block of length  $N_f$ . Using (2), the  $n_i(N_f+v) \times n_i(N_f+v)$  input auto-correlation and the  $n_o N_f \times n_o N_f$  noise auto-correlation matrices are, respectively, defined by  $\mathbf{R}_{xx} \triangleq E[\mathbf{x}_{k:k-N_f-v+1} \mathbf{x}_{k:k-N_f-v+1}^H]$  and  $\mathbf{R}_{nn} \triangleq E[\mathbf{n}_{k:k-N_f+1} \mathbf{n}_{k:k-N_f+1}^H]$ . 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.,  $\mathbf{R}_{xx} = \mathbf{I}_{n_i(N_f+v)}$  and  $\mathbf{R}_{nn} = \frac{1}{SNR} \mathbf{I}_{n_o N_f}$ . Moreover, the output-input cross-correlation and the output auto-correlation matrices are, respectively, defined as

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

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

## III. SPARSE FIR MIMO CHANNEL SHORTENING

In FIR MIMO channel shortening [4], the goal is to design a MIMO equalizer with  $N_f$  matrix taps, which is denoted by the  $n_o N_f \times n_i$  matrix  $\mathbf{W} \triangleq [\mathbf{W}_0 \ \mathbf{W}_1 \ \dots \ \mathbf{W}_{N_f-1}]^T$ , to equalize  $\mathbf{H}$  to a MIMO

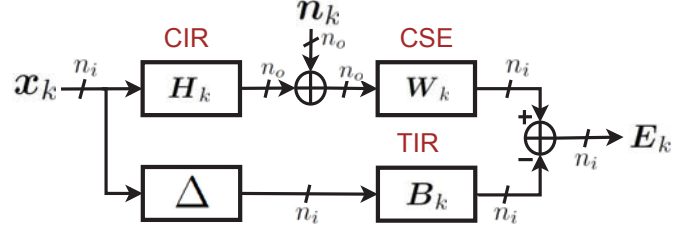


Figure 1. A schematic illustrating the relation between CIR, CSE, and TIR.

TIR matrix  $\tilde{\mathbf{B}} \triangleq [\tilde{\mathbf{B}}_0 \ \tilde{\mathbf{B}}_1 \ \dots \ \tilde{\mathbf{B}}_{N_b}]^T$  with  $(N_b + 1)$  matrix taps  $\tilde{\mathbf{B}}_i$ , each of size  $n_i \times n_i$ . By defining the matrix  $\mathbf{B}^H = [\mathbf{0}_{n_i \times n_i \Delta} \ \tilde{\mathbf{B}}^H \ \mathbf{0}_{n_i \times n_i s}]$ , where  $0 \leq \Delta \leq N_f + v - N_b - 1$  and  $s \triangleq N_f + v - N_b - \Delta - 1$ , the mean square error (MSE) of the error signal can be written as follows (see Figure 1) [4]

$$\begin{aligned} \xi(\mathbf{B}, \mathbf{W}) &\triangleq \text{Trace} \left\{ E[\mathbf{E}_k \mathbf{E}_k^H] \right\} \\ &= \underbrace{\text{Trace} \left\{ \mathbf{B}^H \mathbf{R}_\Delta \mathbf{B} \right\}}_{\triangleq \xi_m(\mathbf{B})} + \underbrace{\text{Trace} \left\{ \mathbf{G}^H \mathbf{R}_{yy} \mathbf{G} \right\}}_{\triangleq \xi_{ex}(\mathbf{B}, \mathbf{W})}, \end{aligned} \quad (5)$$

where  $\mathbf{R}_\Delta \triangleq \mathbf{R}_{xx} - \mathbf{R}_{xy} \mathbf{R}_{yy}^{-1} \mathbf{R}_{yx}$  and  $\mathbf{G} \triangleq \mathbf{W} - \mathbf{R}_{yy}^{-1} \mathbf{R}_{yx} \mathbf{B}$ . The second term of the MSE is equal to zero under the optimum CSE matrix filter coefficients, i.e.,  $\mathbf{W} = \mathbf{R}_{yy}^{-1} \mathbf{R}_{yx} \mathbf{B}$ , and the resulting MSE can then be expressed as follows (defining  $\mathbf{R}_\Delta \triangleq \mathbf{A}_\Delta^H \mathbf{A}_\Delta$ ):

$$\begin{aligned} \xi_m(\mathbf{B}) &= \text{Trace} \left\{ \mathbf{B}^H \mathbf{A}_\Delta^H \mathbf{A}_\Delta \mathbf{B} \right\} = \|\mathbf{A}_\Delta \mathbf{B}\|_F^2 \\ &= \left\| \mathbf{A}_\Delta \begin{bmatrix} \mathbf{b}^{(1)} & \mathbf{b}^{(2)} & \dots & \mathbf{b}^{(n_i)} \end{bmatrix} \right\|_F^2 \\ &= \left\| (\mathbf{I}_{n_i} \otimes \mathbf{A}_\Delta) \mathbf{vec}(\mathbf{B}) \right\|_F^2 \\ &= \left\| \mathbf{A}_\Delta \mathbf{b}^{(1)} \right\|_2^2 + \dots + \left\| \mathbf{A}_\Delta \mathbf{b}^{(n_i)} \right\|_2^2, \end{aligned} \quad (6)$$

where  $\mathbf{b}^{(i)}$  is the  $i^{\text{th}}$  column of  $\mathbf{B}$  and  $\mathbf{vec}$  is an operator that maps a  $n \times n$  matrix to a vector by stacking the columns of the matrix. To compute the TIR matrix filter taps  $\mathbf{B}$  that minimize  $\xi_m(\mathbf{B})$ , we minimize  $\xi_m(\mathbf{B})$  under the identity tap constraint (ITC) where we restrict the  $\ell^{\text{th}}$  matrix coefficient of  $\tilde{\mathbf{B}}$  to be equal to the identity matrix, i.e.,  $\tilde{\mathbf{B}}_\ell = \mathbf{I}_{n_i}$ . Towards this goal, we rewrite  $\xi_m(\mathbf{B})$  as follows:

$$\xi_m(\mathbf{B}, \ell) = \sum_{i=1}^{n_i} \left\| \mathbf{A}_\Delta^{(\cdot, n_i \Delta + n_i \ell + i)} \mathbf{b}^{(i \setminus m+i)} + \mathbf{a}_{m+i} \right\|_2^2, \quad (7)$$

where  $\mathbf{A}_\Delta^{(\cdot, n_i \Delta + n_i \ell + i)}$  is formed by all columns of  $\mathbf{A}_\Delta$  except the  $(m+i)^{\text{th}}$  column, i.e.,  $\mathbf{a}_{m+i}$ , and  $\mathbf{b}^{(i \setminus m+i)}$  is formed by all elements of  $\mathbf{b}^{(i)}$  except the  $(m+i)^{\text{th}}$  entry, which is set equal to one. Then, we formulate the following problem for the design of sparse TIR matrix filter taps  $\mathbf{B}$ :

Table I  
CHANNEL SHORTENING NOTATION AND KEY MATRICES USED IN THIS PAPER

Notation	Meaning
$\mathbf{H}$	Channel matrix of size $n_o N_f \times n_i (N_f + v)$
$\mathbf{R}_{xx}$	Input auto-correlation matrix of size $n_i (N_f + v) \times n_i (N_f + v)$
$\mathbf{R}_{xy}$	Input-Output cross-correlation matrix of size $n_i (N_f + v) \times n_o (N_f)$
$\mathbf{R}_{yy}$	Output auto-correlation matrix of size $n_o N_f \times n_o N_f$
$\mathbf{R}_{nn}$	Noise auto-correlation matrix of size $n_o N_f \times n_o N_f$
$\mathbf{R}_\Delta$	$\triangleq \mathbf{R}_{xx} - \mathbf{R}_{xy} \mathbf{R}_{yy}^{-1} \mathbf{R}_{yx}$ matrix of size $n_i (N_f + v) \times n_i (N_f + v)$
$\mathbf{W}$	CSE matrix of size $n_o N_f \times n_i$
$\mathbf{B}$	TIR matrix of size $n_i (N_f + v) \times n_i$

$$\hat{\mathbf{b}}^{(i \setminus m+i)} \triangleq \operatorname{argmin} \left\| \mathbf{b}^{(i \setminus m+i)} \right\|_0 \quad \text{subject to} \quad (8)$$

$$\left\| \mathbf{A}_\Delta^{(:, \setminus m+i)} \mathbf{b}^{(i \setminus m+i)} + \mathbf{a}_{m+i} \right\|_2^2 \leq \gamma_{eq,i},$$

where  $\gamma_{eq,i}$  is used to control the performance-complexity tradeoff. Once  $\hat{\mathbf{b}}^{(i \setminus m+i)}$ ,  $\forall i \in n_i$ , is calculated, we insert the identity matrix  $\mathbf{B}_\ell$  in the  $\ell^{th}$  location to form the sparse TIR matrix coefficients,  $\mathbf{B}_s$ . Then, the optimum CSE matrix taps, in the minimum MSE (MMSE) sense, are determined from (5) to be

$$\mathbf{W}_{opt} = \mathbf{R}_{yy}^{-1} \overbrace{\mathbf{R}_{yx} \mathbf{B}_s}^{:=\beta}. \quad (9)$$

Since  $\mathbf{W}_{opt}$  is not sparse in general, we further propose a sparse implementation for the CSE matrix taps as follows. After computing  $\mathbf{B}_s$ , the MSE will be a function only of  $\mathbf{W}$  and can be expressed as (defining  $\mathbf{R}_{yy} \triangleq \mathbf{A}_y^H \mathbf{A}_y$ )

$$\begin{aligned} \xi(\mathbf{B}_s, \mathbf{W}) &= \xi_m(\mathbf{B}_s) + \operatorname{Trace} \left\{ \mathbf{G}^H \mathbf{A}_y^H \mathbf{A}_y \mathbf{G} \right\} \\ &= \xi_m(\mathbf{B}_s) + \underbrace{\left\| \mathbf{A}_y \mathbf{W} - \mathbf{A}_y^{-H} \beta \right\|_F^2}_{\triangleq \xi_{ex}(\mathbf{W})}. \end{aligned} \quad (10)$$

By minimizing  $\xi_{ex}(\mathbf{W})$ , we further minimize  $\xi(\mathbf{B}_s, \mathbf{W})$ . This is achieved by a reformulation for  $\xi_{ex}(\mathbf{W})$  to get a vector form of  $\mathbf{W}$ , as in the case of (8), as follows:

$$\xi_{ex}(\mathbf{w}_f) = \left\| \left( \mathbf{I}_{n_i} \otimes \mathbf{A}_y^H \right) \mathbf{vec}(\mathbf{W}) - \mathbf{vec}(\mathbf{A}_y^{-H} \beta) \right\|_2^2. \quad (11)$$

Afterward, we solve the following problem to compute the CSE matrix filter taps

$$\mathbf{w}_f \triangleq \operatorname{argmin} \|\mathbf{w}_f\|_0 \quad \text{subject to} \quad \xi_{ex}(\mathbf{w}_f) \leq \bar{\gamma}_{eq}, \quad (12)$$

where  $\mathbf{w}_f = \mathbf{vec}(\mathbf{W})$  and  $\bar{\gamma}_{eq} > 0$  is used to control the performance-complexity tradeoff.

#### IV. PROPOSED SPARSE APPROXIMATION FRAMEWORK

In this section, we provide a general framework for designing both sparse MIMO CSE filter with  $N_f$  matrix taps and MIMO TIR filter with  $(N_b + 1)$  nonzero matrix taps that can be considered as the problem of sparse approximation using

different dictionaries. Mathematically, this framework poses the FIR filter design problem as follows

$$\hat{\mathbf{z}}_s \triangleq \operatorname{argmin}_z \|\mathbf{z}\|_0 \quad \text{subject to} \quad \|\mathbf{K}(\Phi \mathbf{z} - \mathbf{d})\|_2^2 \leq \epsilon, \quad (13)$$

where  $\Phi$  is the dictionary that will be used to sparsely approximate  $\mathbf{d}$ , while  $\mathbf{K}$  is a known matrix and  $\mathbf{d}$  is a known data vector, both of which change depending upon the sparsifying dictionary  $\Phi$ . Notice that  $\hat{\mathbf{z}}_s$  corresponds to one of the elements in  $\{\hat{\mathbf{b}}^{(i)}, \mathbf{w}_f\}$  and  $\epsilon$  is the corresponding element in  $\{\gamma_{eq,i}, \bar{\gamma}_{eq}\}$ . Hence, one can use any factorization for  $\mathbf{R}_{yy}$  and  $\mathbf{R}_\Delta$ , e.g., in (5), to formulate a sparse approximation problem. Using the Cholesky or eigen decomposition for  $\mathbf{R}_{yy}$  or  $\mathbf{R}_\Delta$  we will have different choices for  $\mathbf{K}$ ,  $\Phi$  and  $\mathbf{d}$ . For instance, by defining the Cholesky factorization [11] of  $\mathbf{R}_\Delta$  in (6) as  $\mathbf{R}_\Delta \triangleq \mathbf{L}_\Delta \mathbf{L}_\Delta^H$ , or in the equivalent form  $\mathbf{R}_\Delta \triangleq \mathbf{P}_\Delta \Sigma_\Delta \mathbf{P}_\Delta^H = \Omega_\Delta \Omega_\Delta^H$  (where  $\mathbf{L}_\Delta$  is a lower-triangular matrix,  $\mathbf{P}_\Delta$  is a lower-unit-triangular (unitriangular) matrix and  $\Sigma_\Delta$  is a diagonal matrix), the problem in (13) can, respectively, take one of the forms below:

$$\begin{aligned} \min_{\mathbf{b} \in \mathbb{C}^{n_i(N_f+v)}} \|\mathbf{b}\|_0 \quad \text{s.t.} \quad & \left\| \left( \mathbf{L}_\Delta^{(:, \setminus m+i)} \right)^H \mathbf{b}^{(i \setminus m+i)} + \mathbf{l}_{m+i} \right\|_2^2 \leq \gamma_{eq,i}, \text{ and} \\ \min_{\mathbf{b} \in \mathbb{C}^{n_i(N_f+v)}} \|\mathbf{b}\|_0 \quad \text{s.t.} \quad & \left\| \left( \Omega_\Delta^{(:, \setminus m+i)} \right)^H \mathbf{b}^{(i \setminus m+i)} + \mathbf{p}_{m+i} \right\|_2^2 \leq \gamma_{eq,i}. \end{aligned} \quad (14)$$

Recall that  $\Omega_\Delta^{(:, \setminus m+i)}$  is formed by all columns of  $\Omega_\Delta$  except the  $(m+i)^{th}$  column,  $\mathbf{p}_{m+i}$  is the  $(m+i)^{th}$  column of  $\Omega_\Delta$ , and  $\mathbf{b}^{(i \setminus m+i)}$  is formed by all entries of  $\mathbf{b}^{(i)}$  except the  $(m+i)^{th}$  entry which we have constrained to be equal to 1. Similarly, by writing the Cholesky factorization of  $\mathbf{R}_{yy}$  as  $\mathbf{R}_{yy} \triangleq \mathbf{L}_y \mathbf{L}_y^H$  or the eigen decomposition of  $\mathbf{R}_{yy}$  as  $\mathbf{R}_{yy} \triangleq \mathbf{U}_y \mathbf{D}_y \mathbf{U}_y^H$ , we can formulate the problem in (13) as follows

$$\begin{aligned} \min_{\mathbf{w}_f \in \mathbb{C}^{n_o N_f}} \|\mathbf{w}_f\|_0 \quad \text{s.t.} \quad & \left\| \left( \mathbf{I}_{n_i} \otimes \mathbf{L}_y^H \right) \mathbf{w}_f - \mathbf{vec}(\mathbf{L}_y^{-H} \beta) \right\|_2^2 \leq \bar{\gamma}_{eq}, \quad (15) \\ \min_{\mathbf{w}_f \in \mathbb{C}^{n_o N_f}} \|\mathbf{w}_f\|_0 \quad \text{s.t.} \quad & \left\| \left( \mathbf{I}_{n_i} \otimes \mathbf{D}_y^{\frac{1}{2}} \mathbf{U}_y^H \right) \mathbf{w}_f - \mathbf{vec}(\mathbf{D}_y^{-\frac{1}{2}} \mathbf{U}_y^H \beta) \right\|_2^2 \leq \bar{\gamma}_{eq}, \quad (16) \\ \min_{\mathbf{w}_f \in \mathbb{C}^{n_o N_f}} \|\mathbf{w}_f\|_0 \quad \text{s.t.} \quad & \left\| \left( \mathbf{I}_{n_i} \otimes \mathbf{L}_y^{-1} \right) \left( \left( \mathbf{I}_{n_i} \otimes \mathbf{R}_{yy} \right) \mathbf{w}_f - \mathbf{vec}(\beta) \right) \right\|_2^2 \leq \bar{\gamma}_{eq}. \quad (17) \end{aligned}$$

Note that the sparsifying dictionaries in (15), (16) and (17) are  $(\mathbf{I}_{n_i} \otimes \mathbf{L}_y^H)$ ,  $(\mathbf{I}_{n_i} \otimes \mathbf{D}_y^{\frac{1}{2}} \mathbf{U}_y^H)$  and  $(\mathbf{I}_{n_i} \otimes \mathbf{R}_{yy})$ , respectively. Furthermore, the matrix  $\mathbf{K}$  is an identity matrix in all cases except in (17), where it is equal to  $(\mathbf{I}_{n_i} \otimes \mathbf{L}_y^{-1})$ .

Notice that several other sparsifying dictionaries can be used to sparsely design the CSE and TIR FIR matrix filter taps. Due to lack of space, we have presented above some of these possible choices and other choices can be derived by applying suitable transformations to (6) and (10).

So far, we have shown that the problem of designing sparse CSE and TIR FIR matrix filter taps can be cast into one of sparse approximation of a vector by a fixed dictionary. The general form of this problem is given by (13). To solve this problem, we use the well-known Orthogonal Matching Pursuit (OMP) greedy algorithm [12] that estimates  $\hat{z}_s$  by iteratively selecting a set  $S$  of the sparsifying dictionary columns (i.e., atoms  $\phi'_{j,s}$ ) of  $\Phi$  that are most correlated with the data vector  $\mathbf{d}$  and then solving a restricted least-squares problem using the selected atoms. The OMP stopping criterion can be either a predefined sparsity level (number of nonzero entries) of  $z_s$  or an upper-bound on the norm of the Projected Residual Error (PRE), i.e., " $\mathbf{K} \times (\Phi \mathbf{z} - \mathbf{d}) \triangleq \mathbf{K} \times \text{Residual Error}$ ". The computations involved in the OMP algorithm are well documented in the sparse approximation literature (e.g., [12]) and are omitted here for brevity.

It is also worth pointing out that we can use the asymptotic equivalence between Toeplitz and circulant matrices to carry out the computations needed for  $\mathbf{R}_{yy}$  and  $\mathbf{R}_\Delta$  factorizations efficiently using the fast Fourier transform (FFT) and inverse FFT [13]. This approximation turns out to be quite accurate as shown in [9]. For a Toeplitz matrix, the most efficient algorithms for Cholesky factorization are Levinson or Schur algorithms [14], which involve  $\mathcal{O}(N_f^2)$  computations. In contrast, the eigen-decomposition of a circulant matrix can be done efficiently using the FFT and its inverse with only  $\mathcal{O}(N_f \log(N_f))$  operations. Hence, based on our results in [9], it can be shown that (defining  $L = n_o N_f$ )

$$\begin{aligned} \mathbf{R}_{yy} &\cong \frac{1}{L^2} \mathbf{F}_L^H \Psi_Y \Psi_Y^H \mathbf{F}_L + n_o \sigma_n^2 \mathbf{I}_L \\ &= \frac{1}{L^2} \mathbf{F}_L^H \left( \Psi_Y \Psi_Y^H + n_o L \sigma_n^2 \mathbf{I}_L \right) \mathbf{F}_L \\ &= \Sigma \Sigma^H, \end{aligned} \quad (18)$$

where  $\mathbf{F}_L$  is an  $L \times L$  DFT matrix,  $\sigma_n^2 = 1/\text{SNR}$ ,  $\Psi_Y = \begin{bmatrix} \mathbf{A}_{\tilde{Y}^1} & \dots & \mathbf{A}_{\tilde{Y}^{n_o}} \end{bmatrix}^H$ ,  $\Psi_Y^H \Psi_Y = \sum_{i=1}^{n_o} \|\tilde{\mathbf{Y}}^i\|^2 = N_f \sum_{i=1}^{n_o} \|\mathbf{H}^i\|^2$ , where  $\|\cdot\|^2$  is defined as the element-wise norm square and  $\mathbf{H}$  is the  $L$ -point DFT of the CIRs,  $\mathbf{H} = \begin{bmatrix} \mathbf{H}^{1T} & \dots & \mathbf{H}^{n_o T} \end{bmatrix}^T$ . To illustrate, for  $n_o = 1$ ,  $\mathbf{R}_{yy}$  in (18) reduces to  $\mathbf{R}_{yy} = \mathbf{F}_{N_f}^H (\mathbf{A}_{\varrho_1}) \mathbf{F}_{N_f}$ , where  $\varrho_1 = N_f \|\mathbf{H}\|^2 + \sigma_n^2 N_f \mathbf{1}_{N_f}$ , and  $\mathbf{H}$  is the  $N_f$ -point DFT of the CIR  $\mathbf{h}$ . Similarly, after some algebraic manipulations,  $\mathbf{R}_\Delta$  can be expressed as

$$\mathbf{R}_\Delta \cong \frac{1}{L} \mathbf{F}_N^H \left( \mathbf{I}_N - \Gamma \mathbf{A}_{\bar{\varrho} \circ \varrho} \Gamma^H \right) \mathbf{F}_N = \Theta \Theta^H \quad (19)$$

where  $\Gamma = \begin{bmatrix} \mathbf{I}_M & \dots & \mathbf{I}_M \end{bmatrix}^H$ ,  $\varrho = \bar{\varrho} + n_o L \sigma_n^2 \mathbf{1}_L$ ,  $\bar{\varrho} = N_f \sum_{i=1}^{n_o} \|\mathbf{H}^i\|^2$ ,  $\circ$  denotes element-wise division and  $N = n_i (N_f + v)$ . Notice that for the proposed sparse CSE and TIR filters, the main computational tasks are the

factorizations of the matrices  $\mathbf{R}_\Delta$ ,  $\mathbf{R}_{yy}$  and the computation of OMP. Reference [12] pointed out that the computation cost, e.g., complex multiplications and additions, of OMP is  $\mathcal{O}(NMS)$ , where  $NM$  is the size of  $\mathbf{W}$  or  $\mathbf{B}$  and  $S$  is the number of nonzero entries of  $z_s$ . Moreover, additional  $\mathcal{O}(S^3)$  computations are required to obtain the restricted least square estimate of  $z_s$ . Hence, the total cost to obtain an estimate of  $z_s$  using our proposed design method is the sum of the factorization cost of the involved matrices in the FIR filter design, OMP cost and the restricted least square cost.

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  $\mu(\Phi)$ ,  $\mu(\Phi) \triangleq \max_{i \neq j} \frac{|\langle \phi_i, \phi_j \rangle|}{\|\phi_i\|_2 \|\phi_j\|_2}$  [15], [16]. Notice that  $\mu(\Phi) \in [0, 1]$ . Next, we investigate the coherence of the dictionaries involved in our setup.

## V. COHERENCE OF SPARSIFYING DICTIONARY

In [17], we showed that design of sparse FIR filters depends largely on the worst-case coherence,  $\mu(\Phi)$ , of the sparsifying dictionaries. Here, we are also concerned with analyzing  $\mu(\Phi)$  of the sparsifying dictionaries to ensure it does not approach 1. Furthermore, we are interested in identifying which  $\Phi$  has the smallest coherence and, hence, has the potential to give the sparsest FIR design. In our setup, we have many sparsifying dictionaries ( $\mathbf{R}_{yy}$ ,  $\mathbf{R}_\Delta$  and their factors), but, we can classify them into two groups. The first group includes dictionaries resulting from factorization of the posterior error covariance matrix  $\mathbf{R}_\Delta$ , while the second group includes both the output auto-correlation matrix  $\mathbf{R}_{yy}$  itself or any of its factors. Notice that the matrix  $\mathbf{R}_\Delta$  is an asymptotically Toeplitz matrix, while  $\mathbf{R}_{yy}$  is a Hermitian positive-definite square block Toeplitz matrix.

To characterize the upper-bounds on  $\mu(\Phi)$  for each group of dictionaries, we first obtain upper bounds on the worst-case coherence of both  $\mathbf{R}_\Delta$  and  $\mathbf{R}_{yy}$  separately and evaluate their closeness to 1. Then, we demonstrate through simulations that the coherence of their factors will be less than 1 and smaller than that of  $\mu(\mathbf{R}_\Delta)$  and  $\mu(\mathbf{R}_{yy})$ , respectively. The matrix  $\mathbf{R}_\Delta$  can be expressed compactly in terms of the SNR and CIR coefficients as  $\mathbf{R}_\Delta = \left[ \mathbf{R}_{xx}^{-1} + \mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H} \right]^{-1} = \left[ \mathbf{I} + \text{SNR} \left( \mathbf{H}^H \mathbf{H} \right) \right]^{-1}$ . This shows that, at low SNR, the noise dominates, i.e.,  $\mathbf{R}_\Delta \approx \mathbf{I}$ , and thus,  $\mu(\mathbf{R}_\Delta) \rightarrow 0$ . As the SNR increases, the noise effect decreases and the CIR effect increases, which makes  $\mu(\mathbf{R}_\Delta)$  converge to a constant. Typically, this constant, as shown through simulations, does not approach 1.

On the other hand,  $\mathbf{R}_{yy}$  has a well-structured (Hermitian Toeplitz) closed-form in terms of the CIR coefficients, filter time span  $N_f$  and SNR, i.e.,  $\mathbf{R}_{yy} = \mathbf{H} \mathbf{H}^H + \frac{1}{\text{SNR}} \mathbf{I}$ . Also, it is a square matrix with full rank, due to the presence of noise. In [17], we derived an upper-bound on  $\mu(\mathbf{R}_{yy})$  for any given channel length  $v$ . By numerical evaluation, we find that the worst-case coherence of  $\mathbf{R}_{yy}$  (for any  $v$ ) is sufficiently less than 1. This observation points to the likely success of



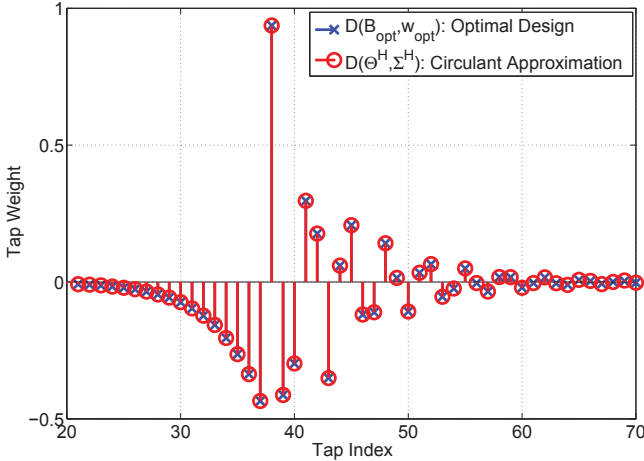


Figure 2. Single impulse response realizations of both optimal (obtained by setting  $\epsilon = 0$ ) and equivalent circulant-approximation based designs for UPDP channel with  $v = 6$ ,  $N_f = 70$  and input SNR = 25 dB.

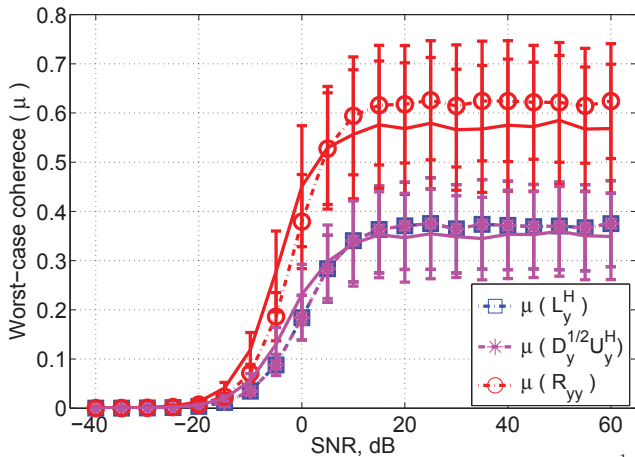


Figure 3. Worst-case coherence for the sparsifying dictionaries  $L_y^H$ ,  $D_y^{1/2} U_y^H$  and  $R_{yy}$  versus input SNR for UPDP with  $n_i = 2$ ,  $n_o = 2$ ,  $v = 8$  and  $N_f = 80$ . Solid lines represent the coherence of the corresponding circulant approximation for  $D_y^{1/2} U_y^H$  (i.e.,  $\Sigma^H$ ) and  $R_{yy}$  (i.e.,  $R_{yy} \cong \Sigma \Sigma^H$ ).

OMP in providing the sparsest solution  $\hat{z}_s$  corresponding to the dictionary that has the smallest  $\mu(\Phi)$ . Next, we confirm these assertions through numerical experiments reported in the next section.

## VI. SIMULATION RESULTS

In our simulations, the used MIMO CIRs are unit-energy symbol-spaced FIR filters with  $v$  taps generated as zero-mean uncorrelated complex Gaussian random variables with a uniform power-delay-profile<sup>2</sup> (UPDP). For simplicity, we assume  $n_i = n_o = 2$ . However, our framework can work for any given number of inputs and outputs. Each point in the reported figures represents the mean of 5000 channel realizations. We use the notation  $D(\chi_t, \chi_e)$  to refer to a TIR designed using sparsifying dictionary  $\chi_t$  and a CSE designed using sparsifying dictionary  $\chi_e$ .

<sup>2</sup>This type of CIRs can be considered as a worst-case assumption since the inherent sparsity of other channel models, e.g., [18] and [19], leads to even better channel shortening performance.

In Figure 2, we plot the impulse responses of the CSEs versus the tap index to quantify the accuracy of approximating the block Toeplitz matrices by their equivalent block circulant matrices. This figure shows single realizations of the impulse responses obtained from the optimum solution (in the MMSE sense) and the equivalent circulant approximation. We observe that both solutions are typically matched. The effect of filter length  $N_f$  is shown in [9].

To investigate the coherence of the sparsifying dictionaries used in our analysis, we plot the worst-case coherence versus the input SNR in Figure 3 for sparsifying dictionaries  $L_y$ ,  $D_y^{1/2} U_y^H$  and  $R_{yy}$ . Note that a smaller value of  $\mu(\Phi)$  indicates that a sparser approximation is more likely. At high SNR levels, the noise effects are negligible and, hence, the sparsifying dictionaries (e.g.,  $R_{yy} \approx \mathbf{H}\mathbf{H}^H$ ) do not depend on the SNR. As a result, the coherence converges to a constant. On the other hand, at low SNR, the noise effects dominate the channel effects. Hence, the channel can be approximated as a memoryless (i.e., single tap) channel. Then, the dictionaries (e.g.,  $R_{yy} \approx \frac{1}{SNR} \mathbf{I}$ ) can be approximated as a multiple of the identity matrix, i.e.,  $\mu(\Phi) \rightarrow 0$ . A similar trend for  $\mu(\mathbf{R}_\Delta)$  and its factors has been observed in [9].

Next, we compare different sparse FIR MIMO CSE and MIMO TIR designs based on different sparsifying dictionaries to study the effect of  $\mu(\Phi)$  on their performance. The OMP algorithm is used to compute the sparse approximations. The OMP stopping criterion is set to be a predefined sparsity level (number of nonzero entries) or a function of the PRE such that: Performance Loss ( $\eta$ ) =  $10 \log_{10} \left( \frac{SNR(\hat{z}_s)}{SNR(z_{opt})} \right) \leq 10 \log_{10} \left( 1 + \frac{\epsilon}{\xi_m} \right) \triangleq \eta_{max}$ , where  $z_{opt}$  is the optimum solution where no sparsity constraint is imposed. Here,  $\epsilon$  is computed based on an acceptable  $\eta_{max}$  and, then, the coefficients of  $\hat{z}_s$  are computed using (13). The percentage of the active taps is calculated as the ratio between the number of nonzero taps to the total number of filter taps. For optimum equalizers, where none of the coefficients is typically zero, the number of active filter taps is equal to the filter span [20]. The decision delay,  $\Delta$ , and the identity-tap location,  $m$ , are chosen to be around  $(N_f + v)/2$  to maximize the equivalent SNR, i.e.,  $SNR = 1/\xi(\mathbf{B}, \mathbf{W})$  [4].

Figure 4 plots the percentage of the active taps versus the performance loss  $\eta_{max}$  for the proposed sparse FIR MIMO-CSEs and the proposed approach in [10], which we refer to as the ‘‘significant taps’’ approach. In that approach, all of the FIR filter taps are computed and only the  $\nu$ -significant ones are retained. 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 CSE taps. Approximately 80% of the taps are eliminated when using  $D_y^{1/2} U_y^H$  and  $L_y^H$  at SNR equal to 10. The sparse MIMO-CSE designed based on  $R_{yy}$  needs more active taps to maintain the same SNR loss as that of the other sparse MIMO-CSEs due to its higher coherence. This suggests that the smaller the worst-case coherence of the dictionary in our setup, the sparser is the equalizer. Moreover, a lower sparsity level (active taps percentage) is achieved at higher SNR levels,

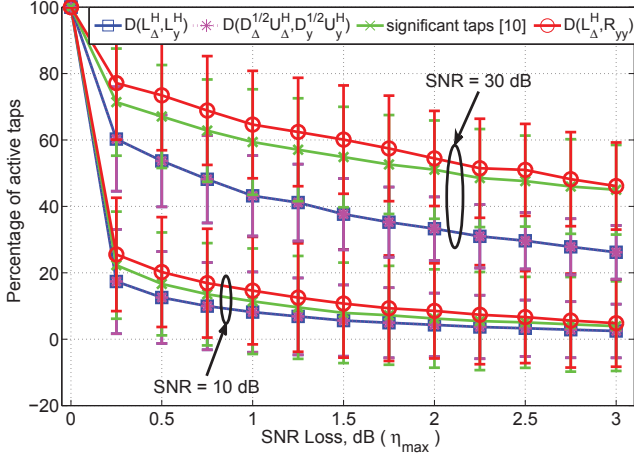


Figure 4. Percentage of active MIMO CSE taps versus the performance loss ( $\eta_{max}$ ) for the sparse MIMO TIR with  $N_b = 5$ , SNR (dB) = 10, 30,  $n_o = 2$ ,  $n_i = 2$ ,  $v = 8$  and  $N_f = 80$ .

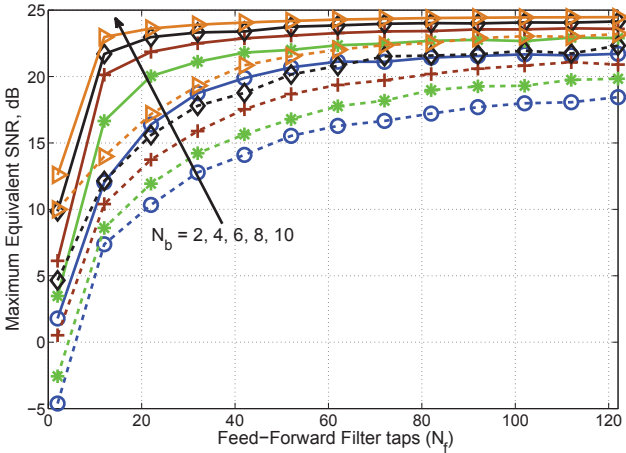


Figure 5. Maximum equivalent SNR versus CSE taps for UPDP CIR with SNR = 25 dB. Solid lines represent the  $D(\Theta^H, \Sigma^H)$  approach while the dashed lines represent the “significant-taps” approach proposed in [10].

which is consistent with the previous findings (e.g., in [21]). Furthermore, reducing the number of active taps decreases the filter equalization design complexity and, consequently, power consumption since a smaller number of complex multiply-and-add operations are required.

In Figure 5, we compare our proposed sparse TIR design with the “significant taps” approach in terms of equivalent SNR, i.e.,  $SNR = 1/\xi(\mathbf{B}, \mathbf{W})$ , where we plot it versus the number of CSE taps  $N_f$  for the UPDP channel. We vary  $N_b$ , the number of TIR taps, from 2 (lower curve) to 10 (upper curve) with step 2. The equivalent SNR increases as  $N_b$  increases for all TIR designs, as expected, and our sparse TIR outperforms, for all scenarios, the “significant taps” approach. Notice that as  $N_b$  increases, the sparse TIR becomes more accurate in approximating the actual CIR.

## VII. CONCLUSIONS

We proposed a general framework for designing sparse FIR MIMO CSE and TIR filters based on a sparse approximation formulation using different dictionaries. In addition, we showed how to reduce the computational complexity of the

design of the sparse equalizers by exploiting the asymptotic equivalence of Toeplitz and circulant matrices. We further numerically evaluated the coherence of the proposed dictionaries involved in our design and showed that the dictionary with the smallest coherence gives the sparsest filter design. The numerical experiments prove superior performance of our approach compared to conventional high-complexity methods.

## REFERENCES

- [1] G. Bauch and A. Naguib, “Map equalization of space-time coded signals over frequency selective channels,” in *Proc. of IEEE WCNC*, 1999, pp. 261–265 vol.1.
- [2] A. M. Tonello, S. D’Alessandro, and L. Lampe, “Cyclic prefix design and allocation in bit-loaded OFDM over power line communication channels,” *IEEE Trans. on Commun.*, vol. 58, no. 11, pp. 3265–3276, 2010.
- [3] N. Al-Dhahir and J. M. Cioffi, “Efficiently computed reduced-parameter input-aided MMSE equalizers for ML detection: A unified approach,” *IEEE Trans. on Info. Theory*, vol. 42, no. 3, pp. 903–915, 1996.
- [4] N. Al-Dhahir, “FIR channel-shortening equalizers for MIMO ISI channels,” *IEEE Trans. on Commun.*, vol. 49, no. 2, pp. 213–218, 2001.
- [5] J. Balakrishnan, R. K. Martin, and C. R. Johnson Jr, “Blind, adaptive channel shortening by sum-squared auto-correlation minimization (SAM),” *IEEE Trans. on Sig. Processing*, vol. 51, no. 12, pp. 3086–3093, 2003.
- [6] F. Rusek and A. Prlja, “Optimal channel shortening for MIMO and ISI channels,” *IEEE Trans. on Wireless Commun.*, vol. 11, no. 2, pp. 810–818, 2012.
- [7] A. Chopra and B. L. Evans, “Design of sparse filters for channel shortening,” *Journal of Signal Processing Systems*, vol. 66, no. 3, pp. 259–272, 2012.
- [8] 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.
- [9] A. O. Al-Abbasi, R. Hamila, W. U. Bajwa, and N. Al-Dhahir, “Design and Analysis Framework for Sparse FIR Channel Shortening,” in *Proc. of IEEE International Conference on Communication*, May, 2016.
- [10] S. Roy, T. M. Duman, and V. K. McDonald, “Error rate improvement in underwater MIMO communications using sparse partial response equalization,” *IEEE Journal of Oceanic Engineering*, vol. 34, no. 2, pp. 181–201, 2009.
- [11] R. A. Horn and C. R. Johnson, Eds., *Matrix Analysis*. New York, NY, USA: Cambridge University Press, 1986.
- [12] 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.
- [13] 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.
- [14] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*, 1st ed. New York, NY, USA: John Wiley & Sons, Inc., 1996.
- [15] W. U. Bajwa and A. Pezeshki, “Finite frames for sparse signal processing,” in *Finite Frames*. Springer, 2013.
- [16] J. Tropp, “Greed is good: algorithmic results for sparse approximation,” *IEEE Trans. on Info. Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.
- [17] A. O. Al-Abbasi, R. Hamila, W. U. Bajwa, and N. Al-Dhahir, “A General Framework for the Design and Analysis of Sparse FIR Linear Equalizers,” in *Proc. of IEEE GlobalSIP Conference*, 2015, pp. 1–5.
- [18] “Guidelines for The Evaluation of Radio Transmission Technologies for IMT-2000 [Available online].” *Recommendation ITU-R M.1225*, 1997.
- [19] IEEE-P802.15, “TG3C Channel Modeling Sub-committee Final Report,” 2009.
- [20] N. Al-Dhahir and J. Cioffi, “MMSE decision-feedback equalizers: finite-length results,” *IEEE Trans. on Info Theory*, vol. 41, no. 4, pp. 961–975, Jul 1995.
- [21] G. Kutz and D. Raphaeli, “Determination of tap positions for sparse equalizers,” *IEEE Trans. on Commun.*, vol. 55, no. 9, pp. 1712–1724, 2007.