Compressed Training Adaptive Equalization: Algorithms and Analysis

Baki B. Yilmaz, *Student Member, IEEE,* and Alper T. Erdogan, *Senior Member, IEEE,*

***Abstract*—We propose "*compressed training adaptive equalization*" as a novel framework to reduce the quantity of training symbols in a communication packet. It is a semi-blind approach for communication systems employing time-domain/frequency-domain equalizers, and founded upon the idea of exploiting the magnitude boundedness of digital communication symbols. The corresponding algorithms are derived by combining the least-squares-cost-function measuring the training symbol reconstruction performance and the infinity-norm of the equalizer outputs as the cost for enforcing the special constellation boundedness property along the whole packet. In addition to providing a framework for developing effective adaptive equalization algorithms based on convex optimization, the proposed method establishes a direct link with compressed sensing by utilizing the duality of the ℓ_1 and ℓ_∞ norms. This link enables the adaptation of recently emerged ℓ_1-norm-minimization-based algorithms and their analysis to the channel equalization problem. In particular, we show for noiseless/low noise scenarios, the required training length is on the order of *the logarithm of the channel spread*. Furthermore, we provide approximate performance analysis by invoking the recent MSE results from the sparsity-based data processing literature. Provided examples illustrate the significant training reductions by the proposed approach and demonstrate its potential for high bandwidth systems with fast mobility.**

***Index Terms*—Equalizers; Adaptive Equalizers; Signal Reconstruction; Adaptive Signal Processing**

I. Introduction

The Inter Symbol Interference (ISI) caused by the multi-path propagation environment imposes a significant barrier to reliable information transmission [1]. As a counter measure, communication receivers generally employ equalizer filters as a low complexity solution to revert the distortion caused by the communication medium. These filters need to be adaptive to keep track of the changing channel conditions, especially in a mobile environment.

Adaptation is usually enabled through the regular transmission of training or pilot symbols along with the information symbols, which are used by the receiver to adjust its equalizer coefficients. Short channel coherence times of high mobility environments, stringent latency constraints and the goal of high bandwidth utilization favor adaptive algorithms that can work with short training times. It is the goal of this article to introduce a novel adaptive equalization framework with a significantly low training requirement that can be employed in communication systems using either time-domain or frequency-domain equalization schemes.

Substantial effort has been spent on developing blind algorithms to completely eliminate the need for training symbols. The most well-known approach is the Constant Modulus Algorithm (CMA) [2]–[4]. The non-convex cost function of CMA and the ill or misconvergence issues that it caused led to the development of the convex framework proposed in [5] and later extended in [6]–[9]. This framework is founded upon the minimization of the ℓ_∞-norm of the equalizer output under a fixed tap constraint on the equalizer coefficients. The algorithms developed through this framework are free of convergence issues and have superior performance, especially for short data lengths.

Blind algorithms are appealing in terms of full bandwidth utilization. However, their data requirements for convergence are relatively large, especially under the short channel coherence time constraints set by mobile environments. Furthermore, the unresolved ambiguity in the delay and sign of the equalizer outputs is a concern. Semi-blind approaches address these concerns by allowing some limited training data to supplement the blind adaptation process. Several semi-blind approaches have been proposed: in the area of semi-blind channel identification, Carvalho et al. in [10] utilized the maximum likelihood approach to estimate the channel. In the area of semi-blind channel equalization, [11] introduced a gradient-descent method by combining knowledge about the constant power property of signals with the available training information. Likewise, the convex combination of CMA and the least-squares costs was proposed in [12]. Cetin et al. in [13] offered a semi-blind channel estimation approach that exploits the side information about the sparsity of the channel impulse response in addition to the constant modulus property and the training information.

In this article, we propose a new semi-blind adaptive algorithm framework for both Single-Carrier Time-Domain Equalization (SC-TDE) and Single-Carrier Frequency-Domain Equalization (SC-FDE) based communication systems. This framework is built upon the convex optimization settings combining the least-squares cost function to exploit training symbols and the infinity norm-based cost function to exploit the magnitude boundedness property of the digital communication sources. The resulting convex optimization problems are the duals of the ℓ_1-norm optimization settings encountered in data processing applications exploiting sparsity, in particular, compressed sensing. This link is particularly useful for applying the algorithm development and analysis methods of sparsity-driven research to the adaptive equal-

This work is supported in part by the TUBITAK 112E057 project.

Baki B. Yilmaz is a PhD candidate in the Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 30332 USA e-mail: b.berkayyilmaz@gatech.edu.

Alper T. Erdogan is with the Department of Electrical and Electronics Engineering, Koc University, Istanbul, Turkey e-mail: alper@stanfordalumni.org.

ization problem. In fact, using this connection, we propose various convex optimization settings for semi-blind adaptive equalization, enabling significant reduction in the amount of training relative to the existing methods. Furthermore, we also invoke results from compressed sensing to analyze the performance of these algorithms. We provide both theoretical and empirical results for the minimum training length requirement and the dependence of the algorithm performance on the amount of training. To emphasize the goal of training size reduction and the connection with compressed sensing, we refer to the resulting framework as *compressed training adaptive equalization* [14], [15].

Regarding the training length requirement for adapting the receivers, the most explicit result is obtained for the supervised channel estimation in [16]. The authors show that if the channel is estimated based only on the training information, the optimal (capacity maximizing) training length is equal to the channel length L_C, which is also the minimum training length required to reliably estimate the channel under this constraint. There are no concrete analytical results, to the best knowledge of the authors, on how much further this training length can be reduced by exploiting data symbols outside the training region via semi-blind approaches. In this article, we show that it is possible to reduce the training length to the order of the logarithm of the channel length, i.e., $\log(L_C)$, for the semi-blind adaptive equalization scenario. These drastic reduction enables high bandwidth communication systems targeting scenarios with high mobility and stringent latency constraints.

As a word of caution, the adaptive equalization approach proposed in this article exploits the special boundedness structure of the PAM/QAM constellations to "compress training data length" while it does not assume that the channel is sparse. Therefore, the contributions of this article should not be confused with approaches exploiting the potential sparse-tap structure of communication channels (or equalizers), such as the sparse channel estimation algorithms in [17], [18]. Our approach is more general in the sense that the channel/equalizer impulse responses are allowed to be dense. The proposed approach aims to sparsify the impulse response for the channel and equalizer cascade system without a priori assumption about the sparseness of the original channel and/or the corresponding equalizer.

Notation	Meaning
n	Discrete-time sequence sample index
k	Receiver branch index
$a_n * b_n$	The convolution of discrete time sequences a_n and b_n
$\mathcal{T}(\mathbf{a}, \mathbf{b}^T)$	Toeplitz matrix whose first column is $\mathbf{a}$ and whose first row is $\mathbf{b}^T$
$(\bullet)^H$	Hermitian transpose
$\langle \mathbf{a}, \mathbf{b} \rangle$	$\mathbf{b}^H\mathbf{a}$, inner product of $\mathbf{a}$ and $\mathbf{b}$
$\mathbf{0}_{M\times N}$	$M \times N$ matrix with all-zero entries

TABLE I
NOTATION TABLE.

This article is organized as follows: we start by introducing the compressed training framework through a time-domain equalization scenario. Therefore, in Section II, we first describe the adaptive setting for the Single-Carrier Time-Domain Equalization (SC-TDE) problem. In Section III, we introduce the proposed compressed training adaptive equalization framework corresponding to the SC-TDE setting and discuss its link to the ℓ_1-norm-based sparse optimization settings. In Section IV, we present the performance analysis corresponding to the proposed compressed training algorithms. Section V extends the proposed approach to the Single-Carrier Frequency-Domain Equalization (SC-FDE) problem. In Section VI, we discuss the implementation of proposed algorithms. In Section VII, the numeric experimental results for the proposed compressed training adaptive equalization algorithms are provided. Finally, Section VIII concludes the article.

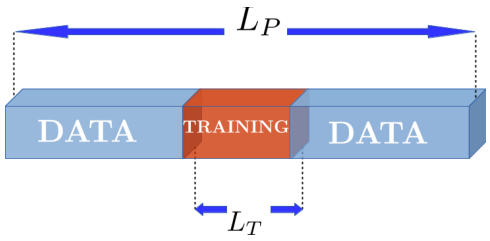


Fig. 1. A Generic Transmit Packet for a Communication System.

II. Adaptive Equalization Setting

In this article, we consider both time-domain and frequency-domain equalization problems for single carrier communications systems. We start with the SC-TDE case in this section and later extend the proposed approach to SC-FDE in Section V. In the SC-TDE case, we assume a standard communication setting where a transmitter sends packets of L_P constellation symbols, $\{s_0, s_1, \ldots, s_{L_P-1}\}$, to a receiver, as illustrated in Fig. 1. Each packet contains L_T training symbols represented as $\{p, p+1, \ldots, p+L_T-1\}$. The main goal of the approach proposed in this article is to reduce (or compress) the training length to increase the spectral efficiency.

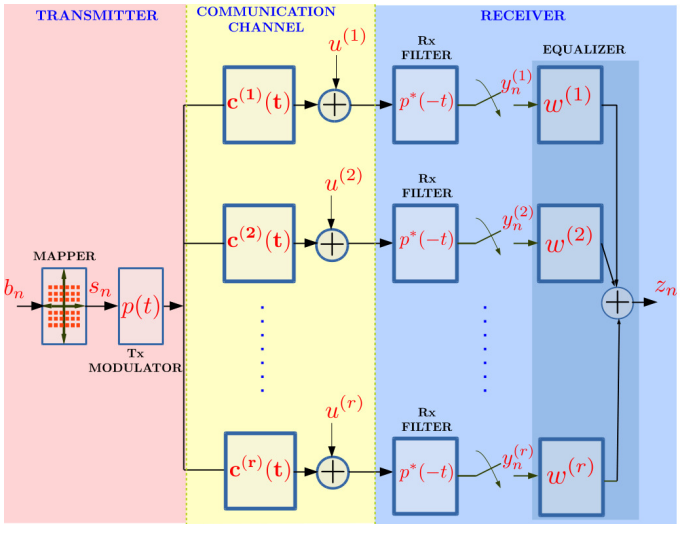


Fig. 2. Equalization Setting for a SC-TDE Receiver with Multiple Diversity Branches.

A baseband equivalent model for time-domain equalization is illustrated in Fig. 2:

- At the transmitter, the binary sequence $\{b_n\}$ is mapped to transmit symbols $\{s_n\}$ by the constellation mapper. We

assume that either a real valued PAM or complex valued QAM constellation is used at the transmitter.

- The transmit sequence $\{s_n\}$ is used to modulate the transmit pulse $p(t)$ to obtain the continuous time signal $x(t) = \sum_n s_n p(t - nT_s)$, where T_s is the symbol period.
- The effect of the communication medium is represented as a linear channel with finite dispersion, which is assumed to be time-invariant over the transmission block and whose output is corrupted by noise. This implies that the corresponding discrete time-equivalent channel has a Finite Impulse Response (FIR).
- We assume a receiver with r diversity branches, where each branch corresponds to a separate antenna and/or an oversampling phase. The channel corresponding to the k^{th} branch is modeled by impulse response $c^{(k)}(t)$ and additive white Gaussian noise $u^{(k)}(t)$.
- The front-end receiver consists of a fixed filter with impulse response $p^*(-t)$ that matches the transmit pulse shape filter $p(t)$, which is typically chosen to minimize self ISI.
- Assuming symbol-spaced sampling at each branch, the discrete time equivalent channel between the transmit symbol sequence $\{s_n\}$ and the sampled receiver sequence $\{y_n^{(k)}\}$ for the k^{th} receiver branch is modeled by

$$y_n^{(k)} = \sum_{l=0}^{L_C-1} h_l^{(k)} s_{n-l} + v_n^{(k)}, \quad k = 1, \dots r. \tag{1}$$

  Here, $\{h_n^{(k)}, n \in \{0, \dots, L_C - 1\}\}$ is the discrete-time equivalent channel FIR filter coefficients, which absorbs the communication channel as well as transmit-receiver filters. We should stress that we do not assume that $\{h_n^{(k)}\}$ is sparse. In fact, all the coefficients are allowed to be non-zero. Here, $v_n^{(k)}$ represents zero-mean complex Gaussian noise samples with variance σ_v^2. If we assume that the pulse shape $p(t)$ is chosen to have zero auto-correlation at lag, then $\{v_n^{(k)}\}$ is an independent and identically distributed (i.i.d.) sequence.
- The receiver employs a bank of equalizer filters $\{w_n^{(k)}, n \in \{0, \dots, L_E - 1\}\}$, $k = 1, \dots, r$, to compensate the effect of the communication channel. We define the equalizer vector $\mathbf{w} = \begin{bmatrix} \mathbf{w}^{(1)^T} & \mathbf{w}^{(2)^T} & \dots & \mathbf{w}^{(r)^T} \end{bmatrix}^T$,where $\mathbf{w}^{(k)} = \begin{bmatrix} w_0^{(k)} & w_1^{(k)} & \dots & w_{L_E-1}^{(k)} \end{bmatrix}^T$, $k = 1, \dots r$.
- The equalizer output is represented by $\{z_n\}$, where $z_n = \sum_{k=1}^r w_n^{(k)} * y_n^{(k)}$. In terms of the equalizer vector, it can be written as $z_n = \mathbf{y}_n^T \mathbf{w}$, where $\mathbf{y}_n = \begin{bmatrix} \mathbf{y}_n^{(1)^T} & \dots & \mathbf{y}_n^{(r)^T} \end{bmatrix}^T$, and $\mathbf{y}_n^{(k)} = \begin{bmatrix} y_n^{(k)} & y_{n-1}^{(k)} & \dots & y_{n-L_E+1}^{(k)} \end{bmatrix}^T$ for $k = 1, \dots, r$.
- The overall channel impulse response from the transmit symbol sequence $\{s_n\}$ to the equalizer output sequence $\{z_n\}$ is defined as $\{g_n, n \in \{0, \dots, L_G - 1\}\}$, such that

$$z_n = g_n * s_n \quad \text{where} \quad g_n = \sum_{k=1}^{r} w_n^{(k)} * h_n^{(k)}, \tag{2}$$

  and $L_G = L_E + L_C - 1$. We define the overall channel impulse response vector as $\mathbf{g} = \begin{bmatrix} g_0 & g_1 & \dots & g_{L_G-1} \end{bmatrix}^T$. In terms of the equalizer vector $\mathbf{w}$, we can write

$$\mathbf{g} = \underbrace{\begin{bmatrix} \mathbf{H}^{(1)} & \mathbf{H}^{(2)} & \dots & \mathbf{H}^{(r)} \end{bmatrix}}_{\mathbf{H}} \mathbf{w}, \tag{3}$$

  where $\mathbf{H}^{(k)}$'s are Toeplitz matrices given by

$$\mathbf{H}^{(k)} = \mathcal{T}\left(\begin{bmatrix} \mathbf{h}^{(k)} \\ \mathbf{0}_{L_E-1 \times 1} \end{bmatrix}, \begin{bmatrix} h_0^{(k)} & \mathbf{0}_{1 \times L_E - 1} \end{bmatrix} \right) \tag{4}$$

  and $\mathbf{h}^{(k)} = \begin{bmatrix} h_0^{(k)} & h_1^{(k)} & \dots & h_{L_C-1}^{(k)} \end{bmatrix}^T$ for $k = 1, \dots, r$. The main goal of the equalizer is to reduce ISI and, therefore, to sparsify $\mathbf{g}$. Note that the corresponding channel or equalizer impulse responses are not necessarily sparse. The condition of having zero ISI corresponds to $g_n = \delta_{n-d}$ for some equalization delay d. The same condition can be rewritten as $\mathbf{g} = \mathbf{e}_{d+1}$, where $\mathbf{e}_{d+1}$ is a member of the standard basis, whose $(d+1)^{th}$ entry is equal to 1 and whose remaining entries are equal to 0.
- Given an equalization delay of d, we define the equalizer output vector for the whole packet as $\mathbf{z}_P = \begin{bmatrix} z_d & z_{d+1} & \dots & z_{d+L_P-1} \end{bmatrix}^T$. In terms of the equalizer vector, we can write $\mathbf{z}_P = \underbrace{\begin{bmatrix} \mathbf{Y}_P^{(1)} & \mathbf{Y}_P^{(2)} & \dots & \mathbf{Y}_P^{(r)} \end{bmatrix}}_{\mathbf{Y}_P} \mathbf{w}$,where $\mathbf{Y}_P^{(k)}$'s are Toeplitz matrices given by

$$\mathbf{Y}_P^{(k)} = \mathcal{T}\left(\begin{bmatrix} y_d^{(k)} & \dots & y_{d+L_P-1}^{(k)} \end{bmatrix}^T, {\mathbf{y}_d^{(k)}}^T \right) \tag{5}$$

  for $k = 1, \dots, r$.
- The equalizer output vector corresponding to the training region can be written as $\mathbf{z}_T = \begin{bmatrix} z_{p+d} & z_{p+d+1} & \dots & z_{p+d+L_T-1} \end{bmatrix}^T$. In terms of the equalizer vector, we can write $\mathbf{z}_T = \underbrace{\begin{bmatrix} \mathbf{Y}_T^{(1)} & \mathbf{Y}_T^{(2)} & \dots & \mathbf{Y}_T^{(r)} \end{bmatrix}}_{\mathbf{Y}_T} \mathbf{w}$,where $\mathbf{Y}_T^{(k)}$'s are Toeplitz matrices given by

$$\mathbf{Y}_T^{(k)} = \mathcal{T}\left(\begin{bmatrix} y_{p+d}^{(k)} & \dots & y_{p+d+L_T-1}^{(k)} \end{bmatrix}^T, {\mathbf{y}_{p+d}^{(k)}}^T \right) \tag{6}$$

  for $k = 1, \dots, r$.
- It is well known [19] that if

  A1. the channels corresponding to all receiver branches do not share a common zero,

  A2. the fractionally spaced equalizer satisfies the sufficient length requirement $L_E \geq L_C - 1$,

  then perfect equalization is possible under the no-noise condition. (**A1** and **A2**) will be our standing assumptions for the rest of the article.

III. Compressed Training Adaptive Equalization

As noted in the introduction, the proposed compressed training adaptive equalization aims to reduce the training data

length by exploiting the special bounded magnitude constellation structure of communication sources. We first provide the underlying principles of the proposed approach in Section III-A. Then, we introduce the corresponding algorithmic framework for the noiseless case in Section III-B and for the noisy case in Section III-C.

A. Motivation for the Proposed Approach

The compressed training approach exploits the special boundedness structure of digital communication constellations. It is founded upon the observation that the ℓ_∞-norm of the equalizer output sequence $\{z_n\}$ reflects the ℓ_1-norm of the overall channel g_n. In other words, the peak equalizer output is a measure of the sparseness of g_n (and therefore the ISI level): *the lower the peak equalizer output is, the more sparse the equalized channel.*

To clarify this claim, we start by writing the equation relating the equalizer output to the transmit symbols and the equalized channel impulse response: $z_n = \sum_{k=0}^{L_G-1} g_k s_{n-k}$. Based on this expression, we can write

$$|z_n| = \left| \sum_{k=0}^{L_G-1} g_k s_{n-k} \right| \leq \sum_{k=0}^{L_G-1} |g_k s_{n-k}| \quad (7)$$

$$\leq \|\mathbf{g}\|_1 \|\mathbf{s}_n\|_\infty, \quad (8)$$

where $\mathbf{s}_n = \begin{bmatrix} s_n & s_{n-1} & \dots & s_{n-L_G+1} \end{bmatrix}^T$. Note that (7) is due to the triangle inequality and (8) is due to Holder's inequality. The inequality in (8) also implies $\|\mathbf{z}_P\|_\infty \leq \|\mathbf{g}\|_1 \|\mathbf{s}_n\|_\infty$. If we assume M-PAM constellation for the transmit symbols, i.e., $s_n \in \{-M+1, -M+3, \dots, M-3, M-1\}$, then $\|\mathbf{s}_n\|_\infty \leq (M-1)$. Therefore, we obtain the bound

$$\|\mathbf{z}_P\|_\infty \leq (M-1)\|\mathbf{g}\|_1. \quad (9)$$

The following discussion shows that under a suitable assumption, the inequality in (9) can be written as an equality, given that the transmit sequence is sufficiently rich in the sense that there exists an index l for which either $s_{l-k} = sign(g_k)(M-1)\ \forall k \in \{0, \dots, L_G-1, \}$ such that

$$z_l = \sum_{k=0}^{L_G-1} g_k s_{l-k} = \sum_{k=0}^{L_G-1} g_k sign(g_k)(M-1)$$

$$= (M-1) \sum_{k=0}^{L_G-1} |g_k| = (M-1)\|\mathbf{g}\|_1, \quad (10)$$

or $s_{l-k} = -sign(g_k)(M-1)\ \forall k \in \{0, \dots, L_G-1\}$, such that $z_l = -(M-1)\|\mathbf{g}\|_1$ holds. Then, due to (9), we conclude that

$$\|\mathbf{z}_P\|_\infty = |z_l| = (M-1)\|\mathbf{g}\|_1, \quad (11)$$

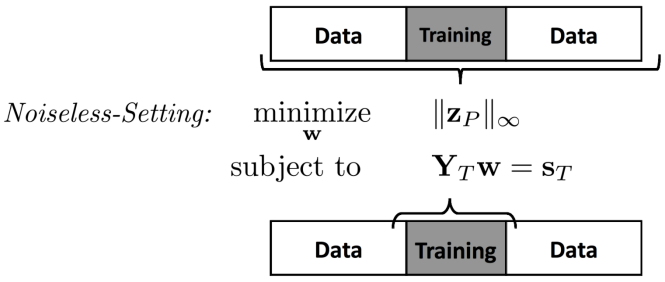
i.e., the ℓ_∞-norm of the equalizer output vector reflects the ℓ_1-norm of the equalized channel impulse response. As a result, the minimization of the peak absolute value of the equalizer outputs amounts to the sparsification of the combined channel-equalizer impulse response or equivalently to the reduction of the ISI.

In the past, based on this observation, a convex optimization-based blind algorithm framework was developed, which was centered around the minimization of the ℓ_∞-norm of the equalizer output under a linear constraint on the equalizer coefficients [5]–[8]. The linear constraints fixed some of the equalizer coefficients or their linear combinations as constants to avoid the trivial, all-zeros solution for the equalizer vector.

In this article, we propose a semi-blind framework based on the ℓ_∞-norm cost function, where the fixed tap constraints of the blind framework cited above are replaced with constraints taking advantage of the limited training information. The algorithms of this framework are derived from special convex optimization settings involving ℓ_∞-norm-based cost functions utilizing whole packet data in addition to the training symbols. These convex settings incorporate training symbols through a linear constraint in the noiseless case and through a least-squares cost function in the noisy scenario, as described in detail in the following subsections.

B. Noiseless Case

1) Proposed Approach: In the noiseless case, perfect reconstruction of the original source signal is possible using an FIR equalizer. This appears mostly as a case of theoretical interest; however, it can also be motivated as the asymptotic case for high Signal-to-Noise Ratio (SNR) or ISI-limited scenarios. As the adaptive equalization approach for the noiseless scenario, we propose the following optimization setting:



where $\mathbf{s}_T = \begin{bmatrix} s_p & s_{p+1} & \dots & s_{n+L_T-1} \end{bmatrix}^T$ is the vector containing training symbols.

For the setting proposed above:

- The constraint $\mathbf{Y}_T\mathbf{w} = \mathbf{s}_T$ corresponds to the utilization of the training information: an ideal equalizer should be able to reconstruct the training symbols in $\mathbf{s}_T$. In the noiseless case, the proposed setting is meaningful only when it corresponds to the under-determined case, i.e., when we try to obtain rL_E equalizer coefficients using only $L_T < rL_E$ training symbols. This compresses the training data to a value less than the number of equalizer coefficients (unknowns). We address the issue of choosing the minimum training length in Section IV.
- The cost function $\|\mathbf{z}_P\|_\infty$ reflects the exploitation of the special constellation structure, and it aims to utilize whole communication packet, including the portion outside of the training region. The under-determined set of equations in the constraint determines the feasible set, including perfect equalizers and undesired solutions. As discussed in the previous section, the minimization of the infinity

norm-based cost function is equivalent to the sparsification of the impulse response of the channel-equalizer cascade. Therefore, the use of this cost function aims to select perfect equalizers from the feasible set, leading to minimized ISI.

- From an algorithmic perspective, we can obtain the solution of the convex problem in *Noiseless-Setting* through various alternatives. For example, it can be stated in the form of a linear program

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \mathbf{e}_1^T\mathbf{x} \\ \text{subject to} \quad & \begin{bmatrix} -\mathbf{1} & \mathbf{Y}_P \\ \mathbf{1} & -\mathbf{Y}_P \\ \mathbf{0} & \mathbf{Y}_T \\ \mathbf{0} & -\mathbf{Y}_T \end{bmatrix} \mathbf{x} \leq \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{s}_T \\ -\mathbf{s}_T \end{bmatrix} \end{aligned}$$

where $\mathbf{e}_1$ is a standard basis vector and $\mathbf{x} = \begin{bmatrix} t & \mathbf{w}^T \end{bmatrix}^T$ with t as a slack variable.

2) Connection to Sparse Optimization Literature: In this section, we show that the approach proposed with the *Noiseless-Setting* is, in effect, equal to the well-known sparse reconstruction problem in the compressed sensing literature:

If we make the following assumption,

A3. L_P is sufficiently long such that $\{s_n\}$ contains all vectors of length L_G of the form (III-A) corresponding to all possible sign patterns,

then (11) can be utilized to show that the cost function in *Noiseless-Setting* is equivalent to

$$\|\mathbf{z}_P\|_\infty = (M-1)\|\mathbf{g}\|_1. \tag{12}$$

Here, $(M-1)$ is the constellation-based constant and can be omitted. Therefore, assuming that (**A3**) holds, we can replace $\|\mathbf{z}_P\|_\infty$ with $\|\mathbf{g}\|_1$ as the cost function.

Regarding the constraint part of *Noiseless-Setting*, we observe that $\mathbf{Y}_T^{(k)} = \mathbf{S}\mathbf{H}^{(k)}$, where $\mathbf{S} = \mathcal{T}(\begin{bmatrix} s_{p+d} & \dots & s_{p+d+L_T-1} \end{bmatrix}^T, {\mathbf{s}_{p+d}}^T)$and $\mathbf{H}^{(k)}$ is as defined in (4). Furthermore,

$$\mathbf{Y}_T\mathbf{w} = \begin{bmatrix} \mathbf{Y}_T^{(1)} & \mathbf{Y}_T^{(2)} & \dots & \mathbf{Y}_T^{(r)} \end{bmatrix} \mathbf{w} = \mathbf{S}\mathbf{H}\mathbf{w} = \mathbf{S}\mathbf{g}. \tag{13}$$

As a result, in terms of the impulse response of the overall mapping, we can rewrite the *Noiseless-Setting* as

$$\begin{aligned} \textit{Noiseless-Equivalent-Setting:} \quad & \underset{\mathbf{g}}{\text{minimize}} \quad \|\mathbf{g}\|_1 \\ & \text{subject to} \quad \mathbf{S}\mathbf{g} = \mathbf{s}_T \end{aligned}$$

We can make the following comments about *Noiseless-Equivalent-Setting* obtained above:

- *Noiseless-Equivalent-Setting* appears as a hypothetical setting because $\mathbf{g}$ is not an accessible optimization parameter requiring the knowledge of $\mathbf{H}$, which is not available, and $\mathbf{S}$ is also unknown. However, under the assumption (**A3**), the optimal solution $\mathbf{w}_{opt}$ of the *Noiseless-Setting*, which is the actual adaptive algorithm, corresponds to the optimal solution $\mathbf{g}_{opt}$ of the hypothetical *Noiseless-Equivalent-Setting*, i.e., $\mathbf{g}_{opt} = \mathbf{H}\mathbf{w}_{opt}$.
- *Noiseless-Equivalent-Setting* is in the form of the *Sparse Reconstruction Problem* (or basis pursuit) in compressed sensing [20], where $\mathbf{S}$ is equivalent to the measurement matrix and $\mathbf{s}_T$ is equivalent to the measurements, and the goal is to reconstruct the original sparse signal $\mathbf{g}$ from its measurements. This link enables analysis of the proposed approach:
  - *Noiseless-Setting* is equivalent to finding a sparse overall impulse response $\{g_n\}$ that reproduces the original training symbols.
  - In the sparse reconstruction problem, the goal is to reduce the number of rows of the measurement matrix, which is equivalent to compressing the number of measurements. In the proposed adaptive equalization setting, the number of rows of $\mathbf{Y}_T$ ($\mathbf{S}$) is equal to the training length. Through this link, we can conclude that the goal of the proposed approach is to establish perfect equalization while compressing the required training information. Therefore, we refer to the proposed framework as "Compressed Training".
  - We note that $\mathbf{s}_T$ is the $(d+1)^{th}$ column of $\mathbf{S}$. Therefore, the standard basis vector $\mathbf{e}_{d+1}$ (and therefore $g_n = \delta_{n-d}$) is a member of the feasible set for *Noiseless-Equivalent-Setting*. In fact, we want it to be the unique global minimum for our problem.
  - In compressive sensing, the main issue is determining the minimum number of required measurements and, in our case, we are also interested in establishing a calculation for the minimum training length required for our setting. We consider this problem in Section IV.
- The assumption (**A3**) implies that the packet length is exponentially related to the overall channel length. More explicitly, as the worst-case scenario, $L_P = \mathcal{C}^{L_G}$, where $\mathcal{C}$ is the number of corner points of the given constellation. However, as illustrated by the numerical examples, the practical requirement is far less than this pessimistic upper bound.

C. Noisy Case

1) Proposed Approach: The existence of noise is more realistic, in which case the exact reconstruction of the training sequence at the equalizer output is a misleading goal. Instead, we aim to keep the corresponding equalizer outputs near the vicinity of the training symbols. This is achieved by replacing the equality constraint in *Noiseless-Setting* with an inequality constraint bounding the Euclidean distance between the equalizer outputs and the training sequence. Therefore, the resulting optimization setting can be written as

$$\begin{aligned} \textit{Noisy-Setting-I:} \quad & \underset{\mathbf{w}}{\text{minimize}} \quad \|\mathbf{z}_P\|_\infty \\ & \text{subject to} \quad \|\mathbf{Y}_T\mathbf{w} - \mathbf{s}_T\|_2 \leq \epsilon. \end{aligned}$$

Note that we can also exchange the cost and the constraint functions above to obtain

$$\begin{aligned} \textit{Noisy-Setting-II:} \quad & \underset{\mathbf{w}}{\text{minimize}} \quad \|\mathbf{Y}_T\mathbf{w} - \mathbf{s}_T\|_2 \\ & \text{subject to} \quad \|\mathbf{z}_P\|_\infty \leq \gamma. \end{aligned}$$

as an alternative setting, where γ is the algorithmic parameter dependent on the prior knowledge about the maximum magnitude of input symbols and the noise level.

We can make the following observations regarding the proposed settings for the noisy case:

- In this case, the matrix $\mathbf{Y}_T$ can be either under-determined or (over)determined. In the under-determined case, it can be used to obtain a reasonable equalizer, which would not be possible by solely using the training data. In the overdetermined case, the proposed approach is expected to require fewer training symbols than the conventional least-squares approach to achieve the same target performance (e.g., probability of error) level.
- Despite the less direct connection to the original ℓ_1-norm-based CLASSO problem, (See Appendix B), relative to the noiseless case, we can still obtain sensible bounds for the performance of the corresponding algorithms (See Section IV).
- Similar to the Lagrangian forms used in compressed sensing [21], we can define an alternative optimization setting for the noisy case by combining the constraint and the cost functions of *Noisy-Setting-II* into a single Lagrangian function

Noisy-Setting-III: $\underset{\mathbf{w}}{\text{minimize}} \quad \|\mathbf{Y}_T\mathbf{w}-\mathbf{s}_T\|_2+\lambda\|\mathbf{z}_P\|_\infty.$

Here, the Lagrangian parameter λ is a reflector of the confidence in the use of the ℓ_∞-based blind cost function relative to the training-based least-squares cost. The selection of the parameter depends on the number of available training symbols.

2) Connection to Sparse Optimization Literature: In the presence of noise, the equalizer output z_n is written as $z_n = g_n * s_n + \sum_{k=1}^{r} w_n^{(k)} * v_n^{(k)}$. Therefore, we have

$$\begin{aligned}\mathbf{Y}_T\mathbf{w} &= (\mathbf{SH}+\mathbf{V}_T)\,\mathbf{w} = \mathbf{SHw}+\mathbf{V}_T\mathbf{w}\\ &= \mathbf{Sg}+\mathbf{V}_T\mathbf{w}, \end{aligned} \tag{14}$$

where $\mathbf{V}_T = \begin{bmatrix} \mathbf{V}_T^{(1)} & \mathbf{V}_T^{(2)} & \dots & \mathbf{V}_T^{(r)} \end{bmatrix}$ is the noise matrix for the training region. Here, $\mathbf{V}_T^{(k)} = \mathcal{T}(\begin{bmatrix} v_{p+d}^{(k)} & \dots & v_{p+d+L_T-1}^{(k)} \end{bmatrix}^T, \mathbf{v}_{p+d}^{(k)\,T})$ is the noise matrix for branch k, defined for $k=1,\dots,r$ where $\mathbf{v}_n^{(k)} = \begin{bmatrix} v_n^{(k)} & v_{n-1}^{(k)} & \dots & v_{n-L_E+1}^{(k)} \end{bmatrix}^T$. Therefore, the constraint part of *Noisy-Setting-I* can be written as

$$\|\mathbf{Sg}+\mathbf{V}_T\mathbf{w}-\mathbf{s}_T\|_2 \le \epsilon. \tag{15}$$

Due to the existence of the noise-dependent term, which is convolved with $\mathbf{w}$, unlike the noiseless case, it is not possible to rewrite *Noisy-Setting-II* solely in terms of $\mathbf{g}$ by dropping the dependence on $\mathbf{w}$. However, if we neglect the effect of noise on the constraint part (which is a rather loose assumption) and define $\hat{\mathbf{s}}_T = \mathbf{s}_T - \mathbf{V}_T\mathbf{w}$, we obtain the setting

Approximately-Equivalent-Noisy-Setting II:

$$\begin{aligned}&\underset{\mathbf{g}}{\text{minimize}} && \|\mathbf{Sg}-\hat{\mathbf{s}}_T\|_2\\ &\text{subject to} && \|\mathbf{g}\|_1 \le 1.\end{aligned}$$

If we also neglect the dependence of $\hat{\mathbf{s}}_T$ on $\mathbf{w}$, the form of this setting resembles the original ℓ_1-norm-based CLASSO problem, which is given in Appendix B.

D. Extension to Complex Square QAM Constellations

All the optimization settings introduced in the previous sections are for real PAM constellations. In the complex square M-QAM constellation case, the transmit symbols take their values from the set $Q = \{a+jb : a,b \in \{-\sqrt{M}+1, -\sqrt{M}+3, \dots \sqrt{M}-3, \sqrt{M}-1\}\}$. The optimization settings for PAM can be adapted to the complex QAM scenario by replacing $\mathbf{z}_P$ with the vector $\tilde{\mathbf{z}}_P \triangleq \begin{bmatrix} \mathbb{R}e\{\mathbf{z}_P^T\} & \mathbb{I}m\{\mathbf{z}_P^T\} \end{bmatrix}^T$. For example, the optimization in *Noiseless-Setting* can be converted for QAM simply as

$$\begin{aligned}\textit{Noiseless-Setting-QAM:} \quad &\underset{\mathbf{w}}{\text{minimize}} && \|\tilde{\mathbf{z}}_P\|_\infty\\ &\text{subject to} && \mathbf{Y}_T\mathbf{w}=\mathbf{s}_T,\end{aligned}$$

and similarly *Noisy-Setting-III* can be adapted to the QAM case as

Noisy-Setting-III-QAM:

$$\underset{\mathbf{w}}{\text{minimize}} \ \|\mathbf{Y}_T\mathbf{w}-\mathbf{s}_T\|_2+\lambda\|\tilde{\mathbf{z}}_P\|_\infty.$$

The reasoning behind this slight modification for the QAM case can be explained using the equation

$$\begin{bmatrix} \mathbb{R}e\{z_n\} \\ \mathbb{I}m\{z_n\} \end{bmatrix} = \begin{bmatrix} \mathbb{R}e\{\mathbf{g}\}^T & -\mathbb{I}m\{\mathbf{g}\}^T \\ \mathbb{I}m\{\mathbf{g}\}^T & \mathbb{R}e\{\mathbf{g}\}^T \end{bmatrix} \begin{bmatrix} \mathbb{R}e\{\mathbf{s}_n\} \\ \mathbb{I}m\{\mathbf{s}_n\} \end{bmatrix}.$$

Under the assumption (**A3**), we can write

$$\begin{aligned}\|\mathbb{R}e\{\mathbf{z}_P\}\|_\infty &= (\sqrt{M}-1)\left\|\underbrace{\begin{bmatrix} \mathbb{R}e\{\mathbf{g}\}^T & \mathbb{I}m\{\mathbf{g}\}^T \end{bmatrix}}_{\tilde{\mathbf{g}}}\right\|_1\\ &= (\sqrt{M}-1)\,\|\tilde{\mathbf{g}}\|_1\,.\end{aligned}$$

Similarly, under the same assumption,

$$\begin{aligned}\|\mathbb{I}m\{\mathbf{z}_P\}\|_\infty &= (\sqrt{M}-1)\left\|\begin{bmatrix} \mathbb{I}m\{\mathbf{g}\}^T & \mathbb{R}e\{\mathbf{g}\}^T \end{bmatrix}\right\|_1\\ &= (\sqrt{M}-1)\,\|\tilde{\mathbf{g}}\|_1\,.\end{aligned}$$

Consequently, we have

$$\begin{aligned}\|\tilde{\mathbf{z}}_P\|_\infty &= \left\|\begin{bmatrix} \|\mathbb{R}e\{\mathbf{z}_P\}\|_\infty & \|\mathbb{I}m\{\mathbf{z}_P\}\|_\infty \end{bmatrix}\right\|_\infty\\ &= (\sqrt{M}-1)\,\|\tilde{\mathbf{g}}\|_1\,.\end{aligned} \tag{16}$$

As a result, the minimization of $\|\tilde{\mathbf{z}}_P\|_\infty$ amounts to the sparsification of $\tilde{\mathbf{g}}$ and, therefore, the corresponding $\mathbf{g}$, as in real PAM constellations. Furthermore, the equivalent optimization settings for QAM can be written in terms of $\tilde{\mathbf{g}}$. As an example, the equivalent of *Noiseless-Setting-QAM* in this case can be written as

Noiseless-Equivalent-Setting-QAM:

$$\underset{\tilde{\mathbf{g}}}{\text{minimize}} \ \|\tilde{\mathbf{g}}\|_1 \text{ subject to } \begin{bmatrix} \mathbb{R}e\{\mathbf{S}\} & -\mathbb{I}m\{\mathbf{S}\} \\ \mathbb{I}m\{\mathbf{S}\} & \mathbb{R}e\{\mathbf{S}\} \end{bmatrix} \tilde{\mathbf{g}} = \tilde{\mathbf{s}}_T.$$

Algorithm 1 Compressed Training Based Adaptive (CoTA) Algorithm Update Rule for SC-TDE (iteration i)

1: (Equalizer Outputs) $\mathbf{z}[i] \leftarrow \mathbf{Y}_P\mathbf{w}[i]$,
2: (Real Equalizer Output Peaks) $\mathbf{p}_{re}[i] \leftarrow \sum_{n=1}^{L_P} 1_{|(\tilde{\mathbf{z}}_P[i])_n|>\alpha\|\tilde{\mathbf{z}}_P[i]\|_\infty}\text{sign}((\tilde{\mathbf{z}}_P[i])_n)\mathbf{e}_n$
3: (Imag Equalizer Output Peaks) $\mathbf{p}_{im}[i] \leftarrow \sum_{n=L_P+1}^{2L_P} 1_{|(\tilde{\mathbf{z}}_P[i])_n|>\alpha\|\tilde{\mathbf{z}}_P[i]\|_\infty}\text{sign}((\tilde{\mathbf{z}}_P[i])_n)\mathbf{e}_{n-L_P}$
4: (2-norm Gradient) $\mathbf{u}_2[i] \leftarrow \frac{\mathbf{Y}_T^H(\mathbf{Y}_T\mathbf{w}[i]-\mathbf{s}_T)}{\left\|\mathbf{Y}_T^H(\mathbf{Y}_T\mathbf{w}[i]-\mathbf{s}_T)\right\|_2}$
5: (∞-norm Weighted Subgradient) $\mathbf{u}_\infty[i] \leftarrow \frac{\mathbf{\Pi}^{-1}\mathbf{Y}_P^H(\mathbf{p}_{re}[i]+j\mathbf{p}_{im}[i])}{\|\mathbf{\Pi}^{-1}\mathbf{Y}_P^H(\mathbf{p}_{re}[i]+j\mathbf{p}_{im}[i])\|_2}$
6: (Update Vector) $\mathbf{u}[i] \leftarrow \mathbf{w}[i] - \mu[i](\mathbf{u}_2[i] + \lambda\mathbf{u}_\infty[i])$
7: (Nesterov Step) $\mathbf{w}[i+1] = \mathbf{u}[i] + \frac{i-1}{i+2}(\mathbf{u}[i] - \mathbf{u}[i-1])$
8: (Normalization) $\mathbf{w}[i+1] \leftarrow \frac{\mathbf{w}[i+1]^H\mathbf{Y}_T^H\mathbf{s}_T}{\mathbf{w}[i+1]^H\mathbf{Y}_T^H\mathbf{Y}_T\mathbf{w}[i+1]}\mathbf{w}[i+1]$

E. Iterative Algorithm

In this section, we provide an iterative algorithm, referred to as **Co**mpressed **T**raining **A**daptive equalization(CoTA), corresponding to the complex *Noisy-Setting-III-QAM*. The update equations for this algorithm are in the table Algorithm 1:

- Step 1 is the equalizer output computation.
- Steps 2-3 are used to determine the potential real and imaginary peaks of the equalizer outputs as required for the subgradient calculation in Step 5 in reference to [8]. Here $\|\tilde{\mathbf{z}}_P[i]\|_\infty$ represents the absolute sample peak of the cascaded real and imaginary components of the equalizer outputs. Due to noise, the true peak(s) may not overlap with the sample peak. Therefore, the operation performed at these steps marks all indices of $\tilde{\mathbf{z}}_P$ for which the absolute values are within the α (an empirical algorithm parameter) factor of the sample peak as potential peak locations. In fact, $1_{|(\tilde{\mathbf{z}}_P[i])_n|>\alpha\|\tilde{\mathbf{z}}_P[i]\|_\infty}$ is the indicator function of this condition. Therefore, the n^{th} component of $\mathbf{p}_{re}[i]$ ($\mathbf{p}_{im}[i]$) is set to the sign of the real (imaginary) part of the n^{th} component of $\mathbf{z}_P[i]$ if its absolute value is greater than $\alpha\|\tilde{\mathbf{z}}_P[i]\|_\infty$; otherwise, it is set to zero.
- Step 4 is the gradient for the ℓ_2-norm component, and Step 5 is the weighted subgradient (as defined in [8]) for the ℓ_∞-norm component of *Noisy-Setting-III-QAM*. Here $\mathbf{\Pi} = \mathbf{Y}_P^H\mathbf{Y}_P L_P^{-1}$ is the sample covariance matrix for the equalizer inputs.
- Steps 6-7 correspond to the "Nesterov Method" [22]-based accelerated equalizer vector update, where $\mathbf{u}$ is an intermediate algorithm variable with $\mathbf{u}[0] = \mathbf{0}$ initialization.
- Step 8 is the normalization of the equalizer vector to reduce bias.

The CoTA algorithm has a flexible structure for implementing the decision directed mode. After the initial eye opening is achieved, the algorithm can continuously expand the training region by appending the reliable decisions.

IV. Analysis of the Proposed Approach

In this section, we provide analysis results for the compressed training adaptive equalization settings proposed in the previous section, mainly by invoking links to the relevant problems and methods in the compressed sensing literature. The main goal is to investigate the effect of training length on performance.

A. Noiseless Case

In the noiseless case, the proposed approach reduces the required training length to less than the number of unknown parameters, rL_E. However, the basic question is how much we can reduce the training length, while maintaining the perfect equalization solution $\mathbf{g} = \mathbf{e}_{d+1}$ as the unique solution to the *Noiseless-Equivalent-Setting*. In that respect, the corollary given in [14], which is derived based on mutual coherence [23], relates the probability of obtaining a perfect equalization solution as a function of L_T and L_G when the entries of $\mathbf{S}$ are i.i.d. Bernoulli. However, the probability upper bound obtained in this work is conservative, as the mutual-coherence-based approach assumes an arbitrary right-hand side of the constraint in *Noiseless-Equivalent-Setting*. We can obtain a tighter bound by exploiting the fact that the right side of the *Noiseless-Equivalent-Setting* is fixed to $\mathbf{s}_T$, which is equal to $\mathbf{S}_{:,d+1}$, as outlined by the following corollary. In contrast to the corollary given in [14], which is specific to the BPSK constellation, the following result is valid for more general PAM and QAM constellations:

Corollary 4.1: Consider a scenario where the communication packet in Fig. 1 contains uniformly distributed M-QAM (or M-PAM) information symbols and the training symbols constructed from the corners of the same constellation. Then, the corresponding *Noiseless-Equivalent-Setting* has $\mathbf{e}_{d+1}$ as the unique solution with probability greater than

$$1 - (L_G - 1)\mathcal{C}^{-L_T+1} \tag{17}$$

where $\mathcal{C}$ is 4 for QAM and 2 for PAM constellations.

Proof: See Appendix A. ■

Corollary 4.1 is simply the reflection of the phase transition results for the sparse reconstruction problem in compressed sensing [20] as it arises as the equivalent problem (*Noiseless-Equivalent-Setting*) to our original adaptive equalization problem in *Noiseless-Setting*. The intuition behind this result, which is also a sketch of the proof in Appendix A, can be stated as follows: (i) $\mathbf{s}_T$ in *Noiseless-Equivalent-Setting* is the $d+1^{\text{th}}$ column of $\mathbf{S}$; therefore, $\mathbf{g} = \mathbf{e}_{d+1}$ is a solution of the constraint part. (ii) $\mathbf{g} = \mathbf{e}_{d+1}$ is the unique optimal solution of *Noiseless-Equivalent-Setting* if no other column of $\mathbf{S}$ aligns with its $d+1^{\text{th}}$ column. (iii) The probability that another column of $\mathbf{S}$ aligns with its $d+1^{\text{th}}$ column decreases exponentially with increasing L_T, as in (17).

As a result, the phase transition property proposed by Corollary 4.1 implies that the amount of training required in the noiseless (or high SNR) case is approximately $log_{\mathcal{C}}(L_G) + \varrho$, where ϱ represents a safety margin for the completion of phase

transition. Assuming that the equalizer length L_E is roughly equal to the channel spread L_C, we can prescribe

$$\boxed{L_T \approx \log_{\mathcal{C}}(2L_E) + \varrho} \tag{18}$$

in terms of the number of equalizer coefficients or, equivalently,

$$\boxed{L_T \approx \log_{\mathcal{C}}(2L_C) + \varrho} \tag{19}$$

in terms of the channel spread. This corresponds to a low training size relative to the number of equalizer coefficients, which is equal to rL_E, especially if the channel and equalizer lengths are relatively long. The results of this subsection on the phase transition property and related minimum training length are illustrated in Section VII (in Fig. 4.)

B. Noisy Case

The performance analysis provided in the previous section prescribes a training length selection rule that is suitable for noiseless or high-SNR regimes. When the SNR is relatively low, the impact of noise needs to be taken into account by both the algorithm choice, as proposed in Section III-C, and by its corresponding analysis. For this purpose, we consider the following performance metrics.

- Mean Square ISI level given by

$$MSISI = E(\|\mathbf{g} - \mathbf{e}_{d+1}\|_2^2), \tag{20}$$

- Mean Square Error given by

$$\begin{aligned} MSE &= E(|z_n - s_{n-d}|^2) \\ &= E(\|\mathbf{g} - \mathbf{e}_{d+1}\|_2^2) + E(\|\mathbf{V}_T\mathbf{w}\|_2^2)L_T^{-1} \end{aligned} \tag{21}$$

  where the second term on the right hand side of (21) represents the filtered noise power at the output of the equalizer.

To obtain expressions for these metrics, we utilize the connection with the original CLASSO problem and exploit the existing MSE analysis results for CLASSO, which are summarized in Appendix B. The treatment in Appendix C uses this approach to obtain

- $MSISI(\|\mathbf{g}^* - \mathbf{e}_{d+1}\|_2^2) \approx \frac{\log(L_G+1)}{L_T - \log(L_G+1)} \frac{E(\|\mathbf{V}_T\mathbf{w}^*\|_2^2)}{L_T}$,
- $MSE \approx \left(\frac{\log(L_G+1)}{L_T - \log(L_G+1)} + 1\right) \frac{E(\|\mathbf{V}_T\mathbf{w}^*\|_2^2)}{L_T}$.

In Appendix C, we also provide examples to illustrate the accuracy of these expressions in predicting the empirical performance. As a major conclusion, MSISI and MSE expressions in the noisy case also confirm that the training length should be greater than $\log(L_G)$.

V. Adaptive Single-Carrier Frequency-Domain Equalization

In this section, we extend the compressed training approach to Single-Carrier Frequency-Domain Equalization (SC-FDE), which takes advantage of the special convolutive structure of the frequency-selective channels [24], [25]. In this scheme, the linear convolution channel is, in effect, converted to a circulant convolution channel through the inclusion of a cyclic-prefix symbol. As illustrated by Fig. 3, a predetermined Unique-Word (UW) sequence $\{u_n, n \in \{1, \ldots, L_T\}\}$, where L_T is the UW length, is appended to the beginning and to the end of the data sequence $\{d_n, n \in \{1, \ldots L_D\}\}$, where L_D is the length of the data symbols, to form the transmit block of the SC-FDE system. The unique word serves both as the

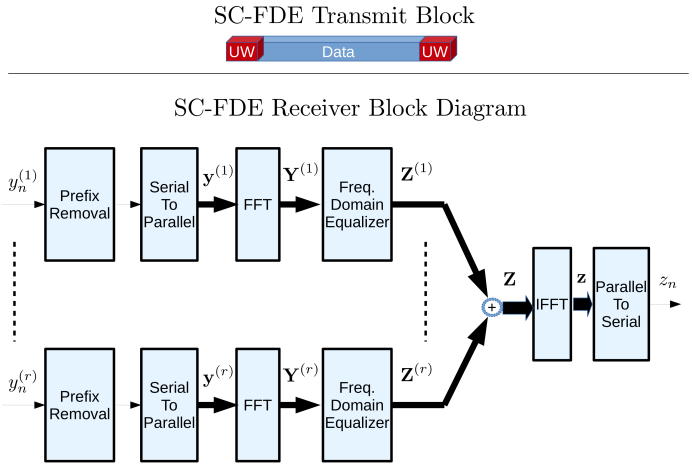


Fig. 3. Single-Carrier Frequency-Domain Equalization System.

training sequence and as the cyclic-prefix block, and its size is selected to be greater than the presumed channel spread. The corresponding SC-FDE receiver branch structure with r diversity branches is also shown in Fig. 3. According to this figure, at each receiver branch, after prefix removal, $L_P = L_T + L_d$ consecutive time-domain receiver samples are vectorized $\mathbf{y}^{(k)} \in C^{L_P}, k = 1, \ldots, r$, and converted to frequency-domain vector $\mathbf{Y}^{(k)}$ through L_P-point FFT operation, represented as $\mathbf{Y}^{(k)} = \mathbf{F}^H\mathbf{y}^{(k)}$. Here, $\mathbf{F}$ is the normalized DFT basis matrix with $F_{lm} = \frac{e^{\frac{j2\pi(l-1)(m-1)}{L_P}}}{\sqrt{L_P}}$, which is unitary. The equalizer operation corresponds to the combination of elementwise-multiplied frequency-domain vectors

$$\mathbf{Z} = \sum_{k=1}^{r} \mathbf{W}^{(k)} \odot \mathbf{Y}^{(k)} \tag{22}$$

where $\odot$ is the elementwise multiplication operator and $\mathbf{W}^{(k)}, k = 1, \ldots, r$ are equalizer coefficient vectors. The equalizer output converted to the time domain is given by $\mathbf{z} = \mathbf{F}\mathbf{Z}$.

In the adaptive setting, we consider the scenario where the equalizer is constrained to be trained using a single block. This is a desirable property for the adaptive algorithm given the channel coherence time constraints mandated by wireless mobile environments. The compressed training approach is a good fit for this task, where the goal is to increase the room for the data symbols by restricting the amount of training symbols in the same block (the presented approach can be easily extended to multi-block-based training). For the adaptive compressed training-based SC-FDE, we devise the following optimization setting:

$$\textit{SC-FDE Setting:} \quad \begin{array}{ll} \underset{\mathbf{W}}{\text{minimize}} & \|\mathbf{E}_d\mathbf{z} - \mathbf{u}\|_2 + \lambda\|\tilde{\mathbf{z}}\|_\infty \\ \text{subject to} & \mathbf{W} \in \mathcal{F}_{L_E} \end{array}$$

where

- $\mathbf{u} = \begin{bmatrix} u_1 & u_2 & \dots, u_{L_T} \end{bmatrix}$ is the vector containing the UW (training) symbols.
- d is the target equalization delay and $\mathbf{E}_d$ is the matrix that extracts the part of the $\mathbf{z}$ matrix corresponding the UW symbols based on the selected choice of delay. $\mathbf{E}_d$ is a submatrix obtained by deleting L_D rows of the $L_P \times L_P$ identity matrix. It can be written as $\mathbf{E}_d = \begin{bmatrix} \mathbf{0} & \mathbf{I}_{L_T} \end{bmatrix} \mathbf{A}_d$, where multiplication of $\mathbf{z}$ with $\mathbf{A}_d$ is equivalent to applying the circular d-advance operation on $\mathbf{z}$ to position the elements of $\mathbf{z}$ corresponding to UW to the last L_T rows, i.e., the compensation of the equalization delay. Multiplication by $\begin{bmatrix} \mathbf{0} & \mathbf{I}_{L_T} \end{bmatrix}$ extracts the UW region from the circularly shifted $\mathbf{z}$.
- $\mathcal{F}_{L_E}$ is the set of frequency domain equalizers with impulse response length less than or equal to L_E, which can be written as $\mathcal{F}_{L_E} = \{\mathbf{W} \in C^{L_P} : \begin{bmatrix} \mathbf{0}_{L_P - L_E \times L_E} & \mathbf{I}_{L_P - L_E} \end{bmatrix} \mathbf{F}\mathbf{W} = \mathbf{0}\}$

The corresponding adaptive algorithm update rule is given in the table Algorithm 2. Here, $\mu[k]$ is the step-size parameter (empirically chosen), $1_{(\bullet)}$ is the indicator function, $\mathbf{U}[k]$ is algorithm's intermediate variable (with $\mathbf{U}[0] = \mathbf{0}$) for accelerated convergence, and α is the peak selection threshold parameter (empirically chosen). The projection (to set $\mathcal{F}_{L_E}$) operator in Step 8 is $\mathcal{P}_{\mathcal{F}_{L_E}}(\mathbf{a}) \triangleq \text{fft}\left(\begin{bmatrix} \mathbf{0}_{N-L_E \times L_E} & \mathbf{I}_{N-L_E} \end{bmatrix} \text{ifft}(\mathbf{a})\right)$, which first transforms its argument to the time domain, zeros out all components except the first L_E and then transforms back to the frequency domain.

Algorithm 2 Compressed Training-Based Adaptive SC-FDE Algorithm Update

1: (FD Eq. Output) $\mathbf{Z}[k] \leftarrow \sum_{l=1}^{r} \mathbf{W}^{(l)}[k] \odot \mathbf{Y}^{(l)}$
2: (TD Eq. Output) $\mathbf{z}[k] \leftarrow \text{ifft}_{L_P}(\mathbf{Z}[k])$
3: (TD Training Error) $\mathbf{e}[k] \leftarrow \mathbf{E}_d \mathbf{z}[k] - \mathbf{u}$
4: (FD Training Error) $\mathbf{E}[k] \leftarrow \text{fft}_{L_P}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{e}[k] \end{bmatrix}\right)$
5: (TD Re. Peaks) $\mathbf{p}_{re}[k] \leftarrow \sum_{l=1}^{L_P} 1_{|\tilde{z}_l[k]| \geq \alpha \|\tilde{\mathbf{z}}[k]\|_\infty} \text{sign}(\tilde{z}_l[k]) \mathbf{e}_l$
6: (TD Im. Peaks) $\mathbf{p}_{im}[k] \leftarrow \sum_{l=L_P+1}^{2L_P} 1_{|\tilde{z}_l[k]| \geq \alpha \|\tilde{\mathbf{z}}[k]\|_\infty} \text{sign}(\tilde{z}_l[k]) \mathbf{e}_{l-L_P}$
7: (FD Transformed Peaks) $\mathbf{P}[k] \leftarrow \text{fft}_{L_P}\left(\underbrace{\mathbf{p}_{re}[k] + i\mathbf{p}_{im}[k]}_{\mathbf{p}[k]}\right)$
8: (Projected Subgradient)

$$\mathbf{G}^{(l)}[k] \leftarrow \mathcal{P}_{\mathcal{F}_{L_E}}\left(\left(\frac{\mathbf{E}[k]}{\|\mathbf{e}[k]\|_2} + \frac{\lambda}{\|\tilde{\mathbf{p}}[k]\|_1}\mathbf{P}[k]\right) \odot \bar{\mathbf{Y}}^{(l)}\right)$$

9: (Subgrad. Update) $\mathbf{U}^{(l)}[k] \leftarrow \mathbf{W}^{(l)}[k] - \mu[k]\mathbf{G}^{(l)}[k]$
10: (Nesterov Step) $\mathbf{W}^{(l)}[k+1] = \mathbf{U}^{(l)}[k] + \frac{k-1}{k+2}(\mathbf{U}^{(l)}[k] - \mathbf{U}^{(l)}[k-1])$

VI. On the Implementation of The Compressed Training Adaptive Algorithms

The compressed training adaptive equalization approach introduced in this article has two major attributes: (i) it reduces the required training length significantly (on the order of the logarithm of the channel spread) and (ii) the corresponding algorithms are based on convex optimization setups. The second feature is critical to obtaining algorithms with low computational complexity and robust convergence behavior. If we look at the computational complexity of the CoTA algorithm for SC-TDE systems (i.e., Algorithm 1), the iteration complexity is mainly dominated by Step 1, which is the computation of the equalizer outputs for the whole packet and requires $L_E \times L_P$ operations. For the CoTA algorithm for SC-FDE systems (i.e., Algorithm 2), the iteration complexity is essentially determined by the fft/ifft operations, which require on the order of $L_P \log(L_P)$ operations. We should note that the proposed convex optimization settings for CoTA parallels the form of the large-scale sparsity-driven optimization settings developed for machine learning and big data analysis. Therefore, there is room to develop even faster and lower-complexity CoTA algorithms by taking advantage of the recent results generated by the surge in algorithm development research in these areas [26].

VII. Numerical Experiments

In this section, we provide numerical examples to demonstrate the performance of the proposed approach relative to the state of the art algorithms.

A. Example 1: Theoretical Verification for the Noiseless Case

In this section, we illustrate the dependence of the performance of the proposed scheme on the training length for the noiseless case. We validate the theoretically established results given in Section IV-A experimentally.

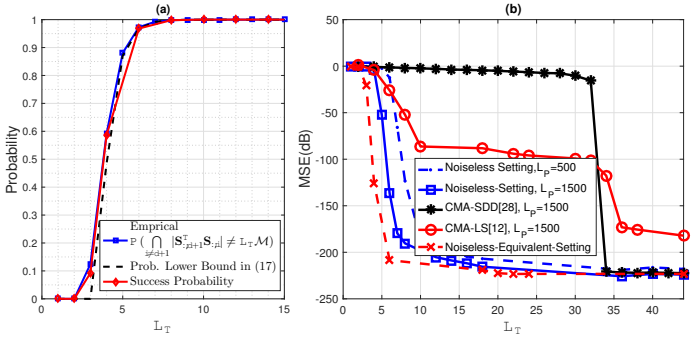


Fig. 4. Performance Dependence on Training Length (a) *Noiseless-Equivalent-Setting*: Probability of perfect equalization versus training length (L_T), (b) *Noiseless-Setting*: MSE vs. L_T performance: comparison with other algorithms and the impact of packet length L_P.

To test the method obtained for the minimum training length, we consider a scenario with a channel with two diversity branches and the channel $L_C = 15$. The channel impulse response is formed by zero mean i.i.d. Gaussian entries with $\mathbb{E}\left(\left|h_i^{(k)}\right|^2\right) = 1$. The 4-QAM constellation is used for the transmit symbols, and the equalizer length is selected as $L_E = 20$. Therefore, the combined channel length is $L_G = L_C + L_E - 1 = 34$.

Fig. 4.(a) shows the probability phase transition event (for perfect equalization) as a function of training length L_T. This figure includes:

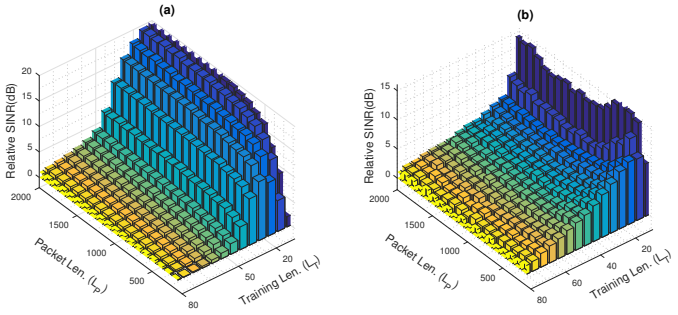


Fig. 5. Average SINR Improvement of the SC-TDE CoTA Relative to **a)** CMA-SDD [28], **b)** CMA-LS [12] for i.i.d. Gaussian channel.

i. The empirical (frequency-based) estimate for the $P(\bigcap_{i\neq d+1}\mathbf{S}_{:,d+1}^T\mathbf{S}_{:,i}\neq\mathcal{M}L_T)$, which is the probability that sufficient conditions for perfect equalization hold (see Appendix A).
ii. The success probability, defined (similar to [27]) as $P(\|\mathbf{g}^*-\mathbf{e}_{d+1}\|_2<10^{-5})$, where $\mathbf{g}^*$ is the optimal solution of *Noiseless-Equivalent-Setting*,
iii. The bound in (17).

Based on this figure,

- the empirical estimate (i) and the success probability (ii) are very close, and the lower bound in (17) provides a relatively tight (especially for higher values of transition region) lower bound for them.
- The phase transition starts close to $L_T=3$, which confirms the predicted value $log_4(L_G)=2.55$.

In Fig. 4.(b), the empirical mean square error value $\|\mathbf{g}^*-e_{d+1}\|_2^2$ for *Noiseless-Equivalent-Setting* as a function of L_T is shown, and it can be observed that the MSE value drops significantly in the interval 2 to 6. The figure also shows the MSE result for *Noiseless-Setting*, involving optimization with respect to $\mathbf{w}$. Therefore, this figure reflects the reliability of the assumption $\|\tilde{\mathbf{z}}_P\|_\infty=(\sqrt{M}-1)\|\tilde{\mathbf{g}}\|_1$ for different data packet lengths. From the pessimistic view point, we require on the order of 4^{34} samples to guarantee this condition. However, we can obtain satisfactory performance even for packet lengths as small as 500, where the threshold for the training length with reasonable performance is slightly increased by 2 to 3 symbols. Moreover, Fig. **??**.(b) contains the experimental results for the algorithms proposed by [28] and [12], from which we can observe that the proposed framework consistently outperforms, especially when the training symbols are much fewer than the combined channel length of $L_G=34$.

As a result, the sample scenario in this section confirms the logarithmic reduction in the required training length through the use of the proposed compressed training adaptive equalization framework.

B. Example 2: Random Noisy Channel

The setup for the experiment is as follows: we consider a communication channel with two branches; the equalizer length L_E and FIR channel length L_C for each branch are chosen as 20 and 15, respectively, the SNR is set to 28dB, the coefficients of the communication channel are i.i.d. Gaussian with $\mathbb{E}\left(\left|h_l^{(k)}\right|\right)=1$, the training length L_T alternates between 10 and 80, the packet length L_P varies from 200 to 2000 and the 16-QAM constellation is used for the communication symbols.

In Fig. 5, we provide the Signal-to-residual-ISI-and-Noise-Ratio (SINR) gain of the proposed approach relative to the CMA-SDD [28] and CMA-LS [12] algorithms for different L_T, L_P pairs. This figure demonstrates the large gain of the proposed approach, especially for short training lengths, where the proposed algorithm can have an SINR performance advantage of up to 20dB.

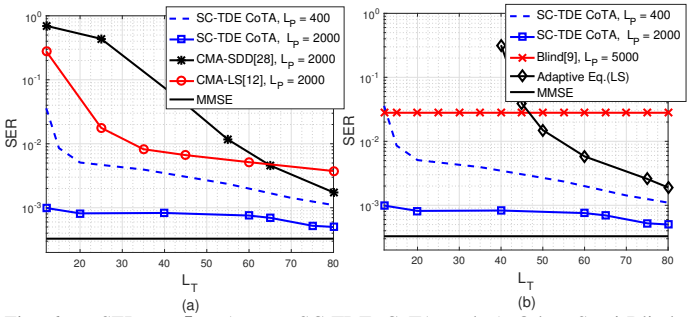


Fig. 6. SER vs L_T Among SC-TDE CoTA and a) Other Semi-Blind Algorithms b) Adaptive and Blind Algorithm for i.i.d. Gaussian Channel.

The same algorithms are compared for the Symbol Error Rate (SER) performance in Fig. 6. We also provide the performance of the best linear Minimum Mean Square Error (MMSE) equalizer (using perfect knowledge of the channel) as a benchmark. These figures also reflect the significant advantages of the proposed algorithm.

C. Example 3: Sample Microwave Noisy Channel Experiment

As the next example, we consider a sample microwave channel obtained from the Signal Processing Information Base [29], whose impulse response (magnitude) is shown in Fig. 7.(a). For the experiment, we consider the full communication scenario given in Fig. 2 with the transmit and receive filters fixed as square root raised cosine filters (with bandwidth extension parameter=0.4). We consider a fractionally spaced equalization with oversampling factor 2. The impulse response in Fig. 7.(a) corresponds to $\frac{T_s}{2}$ spaced samples. The equalizer length L_E is selected as 100 per diversity branch. We consider two scenarios: $L_T=30$ and $L_T=250$. We use the 16-QAM constellation, data packet length L_P of 2500, equalizer delay of 60 and α of 0.8.

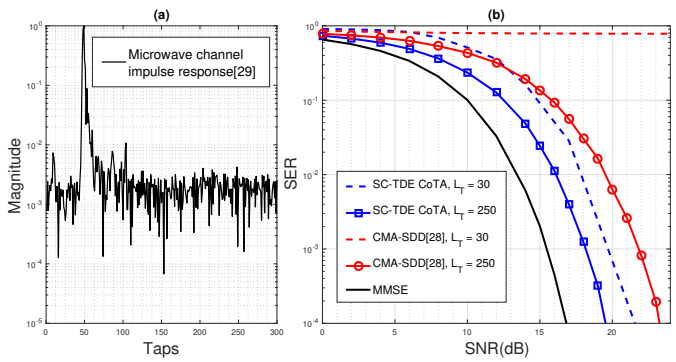


Fig. 7. (a) Sample Microwave Channel Impulse Response (Magnitude) [29] (b) The SER vs SNR Curves for the SC-TDE CoTA, CMA-SDD [28] and MMSE Equalizers for $L_T=30$ and $L_T=250$.

Fig. 7.(b) shows the SER vs SNR performance of the proposed approach relative to the CMA-SDD algorithm [28] and the MMSE equalizer as the benchmark. The proposed approach can achieve reasonable performance levels, even for training lengths as short as 30, and clearly outperforms the other algorithm for both short and long training lengths.

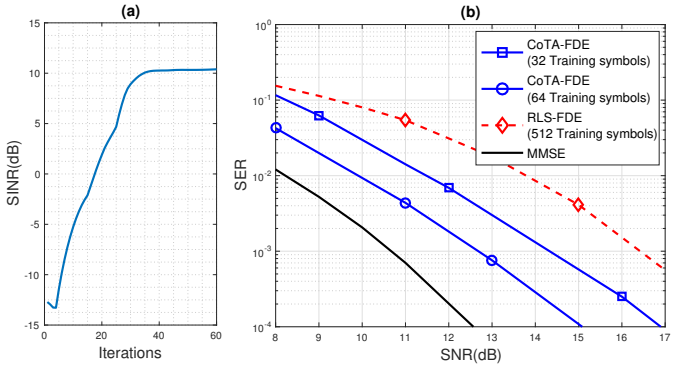


Fig. 8. (a) Sample SINR Convergence Curve for the SC-FDE CoTA Algorithm, (b) SNR vs. SER Performance of SC-FDE COTA and Adaptive RLS-FDE [30] Algorithms

D. Example 4: Single-Carrier Frequency-Domain Equalization

As the final example, we consider the SC-FDE case and evaluate the approach proposed in Section V. We consider the scenario with packet length $N = 512$ and prefix length equal to 32. The transmit constellation is 4-QAM, and the channel is assumed to be of length 32 with random independent and identical complex (circular) Gaussian taps. The proposed approach is applied to a single packet of data, where after the 20^{th} iteration, the algorithm switches to decision directed mode. In Fig. 8.(a), the equalizer output SINR as a function of iterations is shown for 11dB SNR level. It can be observed that the proposed algorithm converges in around 50 iterations. The resulting SNR versus SER curve is shown in Fig. 8.(b). In the same figure, the performance of the nonadaptive MMSE FDE equalizer with perfect channel information is also shown as a benchmark. The performance of the adaptive RLS-FDE [30] is also shown in the same figure. This algorithm uses entire packet of 512 symbols to train FDE and then uses this equalizer for the next data block. This figure reflects that even though the proposed algorithm uses only 32(or 64) training symbols, it performs better than the RLS-SCFDE algorithm using 512 training symbols.

VIII. Conclusion

In this article, we introduced *compressed training adaptive equalization* as a novel semi-blind approach utilizing the magnitude boundedness property of digital communication sources. The proposed framework

- offers a clear prescription for the training length on the order of the logarithm of the channel spread ($\log(L_C)$), which also corresponds to an important reduction compared to the state-of-the-art algorithms,
- enables the formulation of the semi-blind adaptive equalization algorithms through convex settings, facilitating low-complexity implementations with global convergence,
- establishes a link between compressed sensing and adaptive equalization (for both non-sparse and sparse channels) problems, which can be utilized both to derive algorithms and to provide their performance analysis.

The numerical examples provided in the article both confirm the theoretical findings and demonstrate the remarkable performance of the proposed framework in terms of a reduction in the number of training samples required for a given target SNR level, for both noiseless and noisy scenarios. The significant reduction achieved in training length through the proposed framework has three major implications: i) high bandwidth utilization by allocating more room to information symbols, ii) enabling support for highly mobile environments by keeping training length less than the corresponding channel coherence times, iii) supporting applications with low latency requirements via single (sub)packet based adaptation.

Appendix A
Proof of Corollary 4.1

In this appendix, we provide the proof of Corollary 4.1. We start by listing the symbols used in defining the set of corner points of the PAM and QAM constellations:

Symbol	for M-QAM	for M-PAM
$\mathbb{A}$	$\{\frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}\}$	$\{0, \pi\}$
$\mathcal{M}$	$2\left(\sqrt{M}-1\right)^2$	$(M-1)^2$
$\tilde{\mathbb{A}}$	$\{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$	$\{0, \pi\}$

To prove Corollary 4.1, we use the following lemma, which provides sufficient conditions for the unique solution for the *Noiseless-Equivalent-Setting*:

Lemma A.1: Assume that the training symbols, i.e., the elements of $\mathbf{s}_T$, are chosen from the set $\mathbb{S} = \left\{\sqrt{\mathcal{M}}e^{j\theta} | \theta \in \mathbb{A}\right\}$ (which is the corner points of the corresponding constellation used for the data symbols). If no column of $\mathbf{S}$ other than $\mathbf{S}_{:,d+1}$ is equal to $\mathbf{s}_T e^{j\theta}$, where $\theta \in \tilde{\mathbb{A}}$, then $\mathbf{e}_{d+1}$ is the unique solution of the *Noiseless-Equivalent-Setting*.

Proof: The assumption that no column of $\mathbf{S}$ other than $\mathbf{S}_{:,d+1}$ is equal to $\mathbf{s}_T e^{j\theta}$ implies

i. for any $l \neq d+1$, $|\langle \mathbf{s}_T, \mathbf{S}_{:,l} \rangle| < \sqrt{\|\mathbf{s}_T\|_2 \|\mathbf{S}_{:,l}\|_2} \leq \mathcal{M}L_T$, where the first inequality is due to the strictness of the Cauchy-Schwartz under this assumption, and the second inequality is due to $\|\mathbf{S}_{:,l}\|_2 \leq \|\mathbf{s}_T\|_2 = \sqrt{\mathcal{M}L_T}$.
ii. Let $\boldsymbol{\alpha}$ be a solution of the *Noiseless-Equivalent-Setting* and $\boldsymbol{\alpha} \neq \mathbf{e}_{d+1}$; then, $\boldsymbol{\alpha}$ should have more than one non-zero elements.

Implication (ii) can be shown as follows: let $\boldsymbol{\alpha} \neq \mathbf{e}_{d+1}$ be a solution of the *Noiseless-Equivalent-Setting*; then, we can write

$$\mathbf{s}_T = \sum_{l=1}^{L_G} \alpha_l \mathbf{S}_{:,l}. \tag{23}$$

Taking the inner product of both sides with $\mathbf{s}_T$, we obtain $\sum_{l=1}^{L_G} \alpha_l \langle \mathbf{s}_T, \mathbf{S}_{:,l} \rangle = \|\mathbf{s}_T\|_2^2 = \mathcal{M}L_T$. From this expression, by moving the $(d+1)^{th}$ term of the summation to the right

and noting that $\langle \mathbf{s}_T, \mathbf{S}_{:,d+1} \rangle = \mathcal{M}L_T$, we can write,

$$\sum_{l=1,l\neq d+1}^{L_G} \alpha_l \langle \mathbf{s}_T, \mathbf{S}_{:,l} \rangle = (1-\alpha_{d+1})\mathcal{M}L_T. \tag{24}$$

For the summation term, we can write the inequality,

$$\sum_{l=1,l\neq d+1}^{L_G} \alpha_l \langle \mathbf{s}_T, \mathbf{S}_{:,l} \rangle \leq (\|\boldsymbol{\alpha}\|_1 - |\alpha_{d+1}|) \max_{l\neq d+1} |\langle \mathbf{s}_T, \mathbf{S}_{:,l} \rangle|. \tag{25}$$

Hence, combining the equality in (24) with the inequality in (25), we obtain

$$\|\boldsymbol{\alpha}\|_1 - |\alpha_{d+1}| \geq \frac{(1-\alpha_{d+1})\mathcal{M}L_T}{\max_{l\neq d+1} |\langle \mathbf{s}_T, \mathbf{S}_{:,l} \rangle|} \tag{26}$$

Under the assumption of the Lemma A.1, using property (i) above,

$$\begin{aligned} &\|\boldsymbol{\alpha}\|_1 - |\alpha_{d+1}| > 1 - \alpha_{d+1} \\ \implies &\|\boldsymbol{\alpha}\|_1 > 1 - \alpha_{d+1} + |\alpha_{d+1}| \implies \|\boldsymbol{\alpha}\|_1 > 1, \end{aligned} \tag{27}$$

which implies that $\mathbf{e}_{d+1}$ is the unique optimal solution for the *Noiseless-Equivalent-Setting*. ∎

We can now provide the proof of Corollary 4.1 by showing that the sufficient condition in Lemma A.1 holds with probability more than $1-\frac{L_G-1}{\mathcal{C}^{L_T-1}}$, where $\mathcal{C} = |\mathbb{A}|$, i.e., the cardinality of set $\mathbb{A}$, which is the number of corner points of the chosen constellation (note that $\mathcal{C} = 2$ for PAM and $\mathcal{C} = 4$ for QAM constellations).

Proof: (Corollary 4.1). We define $\boldsymbol{\psi} = \mathbf{S}^H \mathbf{s}_T$ as the vector containing the inner products of the columns of $\mathbf{S}$ with $\mathbf{s}_T$. Using Cauchy-Schwarz inequality, we can write $0 \leq |\psi_l| \leq \mathcal{M}L_T$ for each element of $\boldsymbol{\psi}$. We note that the condition in Lemma A.1 for the uniqueness of the solution for *Noiseless-Equivalent-Setting* is equivalent to

$$|\langle \mathbf{s}_T, \mathbf{S}_{:,l} \rangle| \neq \mathcal{M}L_T, l \in \{1, 2, \cdots, d, d+2, \cdots, L_G\}, \tag{28}$$

i.e., the absolute values of all elements of $\boldsymbol{\psi}$ other than the $(d+1)^{th}$ must be strictly less than $\mathcal{M}L_T$. This will be the condition used to derive the lower bound in (28) for perfect equalization.

We start with the observation that using BPSK in the real case and 4-QAM in the complex case for information symbols constitutes the worst cases because for higher constellations, if a column of $\mathbf{S}$ contains a non-corner point, the corresponding element of $\boldsymbol{\psi}$ clearly satisfies the condition in (28). Therefore, the lower bounds obtained below for the BPSK and 4-QAM constellations are also valid (although pessimistic) for higher constellations.

When we assume BPSK or 4-QAM constellations for non-training symbols, we can utilize the union bound to obtain the desired lower bound. For this purpose, we concentrate on the probability that ψ_{d+2} has an absolute value $\mathcal{M}L_T$. Here ψ_{d+2} is the inner product of the two neighboring columns in the sub-matrix

$$\mathbf{S}_{:,d+1:d+2} = \begin{bmatrix} s_p & s_{p+1} & \cdots & s_{p+L_T-1} \\ s_{p-1} & s_p & \cdots & s_{p+L_T-2} \end{bmatrix}^T, \tag{29}$$

which is constructed using $L_T + 1$ different $i.i.d.$ variables, namely, the set $\{s_{p-1}, s_p, \ldots, s_{p+L_T-1}\}$. Therefore, there are $\mathcal{C}^{L_T+1}$ possibilities for $\mathbf{S}_{:,d+1:d+2}$. Due to Lemma A.1, $|\psi_{d+2}|$ is equal to $\mathcal{M}L_T$ if and only if $\mathbf{S}_{:,d+1} = e^{j\Theta}\mathbf{S}_{:,d+2}$ for some $\Theta \in \tilde{\mathbb{A}}$. Due to the Toeplitz structure in (29), this condition is equivalent to $s_l = s_{l-1}e^{j\Theta}$ for $l = p, \ldots, p + L_T - 1$, for some $\Theta \in \tilde{\mathbb{A}}$, or equivalently, $\mathbf{S}_{:,d+2} = s_{p-1}\begin{bmatrix} 1 & e^{j\Theta} & \ldots & e^{j(L_T-1)\Theta} \end{bmatrix}^T$. Based on that, there are $\mathcal{C}$ choices for s_{p-1} and $\mathcal{C}$ choices for Θ. As a result, there are $\mathcal{C}^2$ choices for $\mathbf{S}_{:,d+1:d+2}$ (among all $\mathcal{C}^{L_T+1}$) choices for which $|\psi_{d+2}| = \mathcal{M}L_T$. As a result, we can write

$$P(|\psi_{d+2}| = \mathcal{M}L_T) = \frac{\mathcal{C}^2}{\mathcal{C}^{L_T+1}} = \frac{1}{\mathcal{C}^{L_T-1}} \tag{30}$$

For ψ_d, which is the inner product of $\mathbf{S}_{:,d+1}$ with its left neighbor, a similar line of argument leads to $P(|\psi_d| = \mathcal{M}L_T) = \mathcal{C}^{-L_T+1}$.

Following the same approach, if we consider $(d-1)^{th}$ or $(d+3)^{th}$ elements of $\boldsymbol{\psi}$: the expression for $\psi_{d+1\mp2}$ contains $L_T + 2$ i.i.d. consecutive elements of $\{s_n\}$, which has $\mathcal{C}^{L_T+2}$ vectors in its sample space. We can show that there are only $\mathcal{C}^3$ vectors out of $\mathcal{C}^{L_T+2}$ choices for which $|\psi_{d+1\mp2}| = \mathcal{M}L_T$. Therefore, we can write $P(|\psi_{d+1\mp2}| = \mathcal{M}L_T) = \frac{\mathcal{C}^3}{\mathcal{C}^{L_T+2}} = \frac{1}{\mathcal{C}^{L_T-1}}$, which is the same as the probability for the $(d+1\mp1)^{th}$ elements of $\boldsymbol{\psi}$. In fact, this result generalizes to any element where we can write

$$P(|\psi_l| = \mathcal{M}L_T) = \mathcal{C}^{-L_T+1}, \forall l \neq (d+1). \tag{31}$$

For the probability that the condition for Lemma A.1 holds, we can write

$$\begin{aligned} &P\Big(\bigcap_{l\neq d+1} \{|\langle \mathbf{s}_T, \mathbf{S}_{:,l} \rangle| \neq \mathcal{M}L_T\}\Big) \\ &= 1 - P\Big(\bigcup_{l\neq d+1} \{|\langle \mathbf{s}_T, \mathbf{S}_{:,l} \rangle| = \mathcal{M}L_T\}\Big) && (32) \\ &\geq 1 - \sum_{l\neq d+1} P(\{|\langle \mathbf{S}_{:,d+1}, \mathbf{S}_{:,l} \rangle| = \mathcal{M}L_T\}) && (33) \\ &= 1 - \sum_{l\neq d+1} P(|\psi_l| = \mathcal{M}L_T) && (34) \end{aligned}$$

where the inequality in (33) is obtained from the union bound. For the case $L_T > \log_2(L_G - 1) + 1$, using the expression obtained in (31), we can write for $l \neq d+1$

$$\begin{aligned} P\Big(\bigcap_{l\neq d+1} \{|\langle \mathbf{s}_T, \mathbf{S}_{:,l} \rangle| \neq \mathcal{M}L_T\}\Big) &\geq 1 - \sum_{l\neq d+1} P(|\psi_l| = \mathcal{M}L_T) && (35) \\ &= 1 - (L_G - 1)\mathcal{C}^{-L_T+1} && (36) \end{aligned}$$

∎

APPENDIX B
SUMMARY OF THE PERFORMANCE ANALYSIS RESULTS FOR ℓ_1-CLASSO

In referencing the results from the compressed sensing literature, we consider a linear model with

$$\mathbf{q} = \boldsymbol{\Phi}\mathbf{x}_o + \mathbf{u} \tag{37}$$

where $\mathbf{q}, \mathbf{u} \in \Re^m$, $\mathbf{x}_o \in \Re^n$ and $\boldsymbol{\Phi} \in \Re^{m\times n}$, and the goal is to obtain an estimate of $\mathbf{x}_o$ from the knowledge of $\boldsymbol{\Phi}$ and $\mathbf{q}$. We also assume that the system matrix $\boldsymbol{\Phi}$ is i.i.d. Gaussian with

zero mean and unity variance entries. Similarly, the noise term $\mathbf{u}$ is also assumed to be Gaussian (independent of $\mathbf{\Phi}$) with i.i.d. entries having zero mean and variance σ_u^2. The performance characterizations reflect the average over the ensembles of these random quantities:

- *Least-Squares Problem:* If we do not assume additional structure on the desired vector $\mathbf{x}_o$, it is estimated by

$$\mathbf{x}_{ls} = \arg\min_{\mathbf{x}} \|\mathbf{\Phi}\mathbf{x} - \mathbf{q}\|_2 \tag{38}$$

  for $m \geq n$. The average performance of such estimate can be explicitly obtained as [31]

$$E(\|\mathbf{x}_{ls} - \mathbf{x}_o\|_2^2) = \frac{n}{m-n-1}\sigma_u^2 \quad \text{for } m \geq n+2. \tag{39}$$

- ℓ_1*-CLASSO Problem:* Let us first introduce the ℓ_1-CLASSO:

  Setting ℓ_1*-CLASSO:*

$$\underset{\mathbf{x}}{\text{minimize}} \quad \|\mathbf{\Phi}\mathbf{x} - \mathbf{q}\|_2 \quad \text{subject to} \quad \|\mathbf{x}\|_1 \leq \psi.$$

  where $\mathbf{x} \in \Re^n$ is the sparse target vector to be reconstructed, $\mathbf{q} \in \Re^m$ is the vector containing noisy measurements, where $m \ll n$, $\mathbf{\Phi} \in \Re^{m \times n}$ is the measurement matrix, and ψ reflects the prior knowledge about the ℓ_1-norm of the optimal solution. If we assume that the desired vector is sparse, with only κ non-zero coefficients, where $\kappa \ll n$, and use the ℓ_1-CLASSO optimization setting in III-C to obtain the estimate $\mathbf{x}_{CLASSO}$, then the corresponding square error expression can be written as [21]

$$E(\|\mathbf{x}_{CLASSO} - \mathbf{x}_o\|_2^2) \approx \frac{r_\kappa}{m - r_\kappa}\sigma_u^2 \tag{40}$$

  where $r_\kappa = 2\kappa\log(\frac{n}{\kappa} + 1)$ and $m > r_\kappa$. We note that "the statistical dimension" r is smaller than the ambient dimension n; therefore, the exploitation of the sparsity of $\mathbf{x}_o$ through the use of the ℓ_1-CLASSO approach (over ordinary least-squares) could significantly improve the performance, as evident from the comparison of (39) and (40).

APPENDIX C
A CLOSE LINK BETWEEN COTA AND COMPRESSIVE SENSING

In this section, we endeavor to clarify the connection between the proposed framework and compressive sensing and to verify our proposal, transferring already established results from compressive sensing literature to the communication equalization problem. In that respect, we provide an estimation for the reconstruction error based on $MSIS$ and MSE by modifying the results given in [31]. As the established results in [31] count on low-noise scenarios, our example follows this fashion and provides a communication channel experiment with high SNR. We are aware that our example is highly unlikely for a real-time communication scenario; however, our intention is to remove all doubt about our claim that compressive sensing results are transferable through our framework.

A. Mean Square ISI Level

Based on the resemblance between the *Approximately-Equivalent-Noisy-Setting* and ℓ_1-CLASSO problem, and using the inequality in (40), we can obtain an approximation for the MSISI as

$$E(\|\mathbf{g}^* - \mathbf{e}_{d+1}\|_2^2) \approx \frac{r_1}{L_T - r_1}\frac{E(\|\mathbf{V}_T\mathbf{w}^*\|_2^2)}{L_T} \tag{41}$$

where the desired overall impulse response $\mathbf{e}_{d+1}$ has only 1 non-zero coefficient, i.e., $\kappa = 1$; therefore, $r_1 = 2\log(L_G+1)$ in (40), where $\mathbf{g}^*$ is the combined channel vector corresponding to the optimal solution $\mathbf{w}^*$ of *Noisy-Setting-II*. We note that the assumptions used to obtain the expression in (40) do not hold for *Noisy-Setting-II*:

- It is assumed in the analysis that the measurement matrix $\mathbf{\Phi}$ of *Setting* ℓ_1*-CLASSO* is i.i.d. Gaussian. However, the matrix $\mathbf{S}$ in *Noisy-Setting-II* is actually a Toeplitz matrix randomly constructed from the used constellation symbols.
- The noise term $\mathbf{V}_T\mathbf{w}$ is actually dependent on the optimization parameter, and its covariance in general is not diagonal.

Despite these assumption violations, the expression given in (41) provides a reasonable estimate of the true performance, especially for high SNR levels, as illustrated by the examples considered in Appendix C-C.

B. Mean Square Error

Using the relation between the Mean Square ISI level and MSE given in (21) and the MSISI approximation in (41), we can provide an approximate expression for MSE as

$$MSE \quad \approx \quad \left(\frac{r_1}{L_T - r_1} + 1\right)\frac{E(\|\mathbf{V}_T\mathbf{w}^*\|_2^2)}{L_T}. \tag{42}$$

C. Illustration of MSISI and MSE Variation as a Function of Training Length

To illustrate the performance of the proposed scheme for the noisy case, we consider a random channel setting with $L_C = 30$ whose entries are zero mean i.i.d. Gaussian entries with $\mathbb{E}\left(\left|h_i^{(k)}\right|^2\right) = 1$. The equalizer length is chosen as $L_E = 35$. Furthermore, a receiver with two diversity branches is considered. The 16-QAM constellation is used, and the packet length is $L_P = 10000$. The channel output is perturbed with Gaussian noise for an SNR level of $50dB$, where $\text{SNR} = 10\log\left(\sigma_s^2 \left\|\left[\mathbf{h}^{(1)^T} \quad \dots \quad \mathbf{h}^{(r)^T}\right]\right\|_2^2 r^{-1}\sigma_u^{-2}\right)$.

The Mean Square ISI versus training length L_T corresponding to this scenario is shown in Fig. 9. In this figure, we plot

- The empirical and theoretical MSISI curves for the least-squares-based adaptive equalization algorithm,
- The empirical MSISI curve for **Algorithm 1**,
- The empirical MSISI curve for the solution of

$$\begin{aligned} &\underset{\mathbf{w}}{\text{minimize}} \quad \|\mathbf{Y}_T\mathbf{w} - \mathbf{s}_T\|_2 \\ &\text{s.t.} \quad \left\|\left[\begin{array}{cc} \mathbb{R}e\{\mathbf{H}\mathbf{w}\}^T & \mathbb{I}m\{\mathbf{H}\mathbf{w}\}^T \end{array}\right]\right\|_1 \leq 1, \end{aligned} \tag{43}$$

which is a hypothetical case eliminating the impact of the noise on the ℓ_∞-norm part. Furthermore, it assumes sufficiently long packet length, so $\|\tilde{\mathbf{z}}_P\|_\infty$ can be replaced with $\|\tilde{\mathbf{g}}\|_1$. This formulation is clearly not adaptive, as it requires knowledge of $\mathbf{H}$. However, it is used as a benchmark for comparison to evaluate the impact of noise and finite packet length effects.

- The empirical MSISI curve for the solution of

$$\begin{aligned} &\underset{\mathbf{w}}{\text{minimize}} \quad \|\mathbf{Y}_T\mathbf{w} - \mathbf{s}_T\|_2 \\ &\text{s.t.} \quad \left\| \begin{bmatrix} \mathbb{R}e\{\mathbf{SHw}\}^T & \mathbb{I}m\{\mathbf{SHw}\}^T \end{bmatrix} \right\|_\infty \leq \gamma_{\mathcal{C}}, \end{aligned} \tag{44}$$

where $\gamma_{\mathcal{C}}$ is $(\sqrt{M}-1)$ for M-QAM ($M-1$ for M-PAM). The performance degradation of this hypothetical setting relative to (43) reflects the impact of the packet length in terms of violating the replacement of $\|\tilde{\mathbf{z}}_P\|_\infty$ with $\|\tilde{\mathbf{g}}\|_1$.

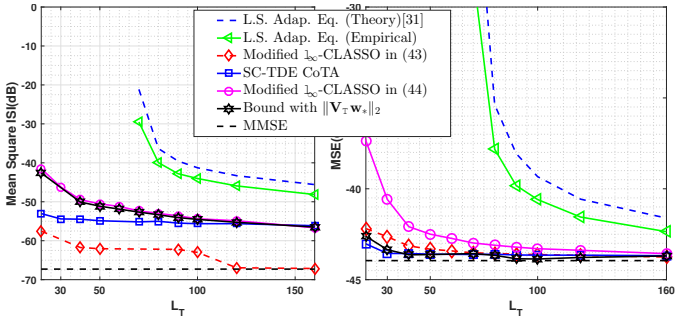


Fig. 9. Mean Square ISI and MSE versus L_T for SNR=50dB.

- The expression in (41) for the theoretical approximation of the expected behavior.

In Fig. 9, the corresponding MSE curves are shown for the same SNR level. Based on these two figures, we can make the following observations:

- For the under-determined case, i.e., $L_T < 70$, the proposed compressed training algorithm successfully equalizes the channel, and reasonable MSE levels can be achieved for a training length as short as 20 symbols.
- For the overdetermined case, the proposed approach provides significant performance improvement, especially close to the border (determined) condition.
- The expressions (41) and (42) provide reasonable estimates of the MSISI and MSE performance of the test settings (43) and (44) with no noise in the constraint part.
- The performance comparison between the CoTA approach and the test settings (43, 44) suggests that the existence of noise in the constraint (ℓ_∞-norm part) does not have a significant impact on the performance with the appropriate choice of parameters, while the choice of data length appears to be sufficient in this specific example (we address the impact of the packet length in Section VII).

REFERENCES

[1] J. G. Proakis, *Digital Communications*. New York, NY: McGraw Hill, 2000.

[2] A. J. van der Veen and A. Paulraj, "An analytical constant modulus allgorithm," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1136–1155, May 1996.

[3] I. Fijalkow, A. Touzni, and J. R. Treichler, "Fractionally spaced equalization using cma: Robustness to channel noise and lack of disparity," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 56–66, 1997.

[4] R. Johnson, P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, and R. A. Casas, "Blind equalization using the constant modulus criterion: A review," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1927–1950, 1998.

[5] S. Vembu, S. Verdu, R. Kennedy, and W. Sethares, "Convex cost functions in blind equalization," *IEEE Transactions on Signal Processing*, vol. 42, pp. 1952–1960, Aug. 1994.

[6] Z. Ding and Z. Luo, "A fast linear programming algorithm for blind equalization," *IEEE Transactions on Communications*, vol. 48, pp. 1432–1436, Sep. 2000.

[7] Z. Q. Luo, M. Meng, K. M. Wong, and J. K. Zhang, "A fractionally spaced blind equalizer based on linear programming," *IEEE Transactions on Signal Processing*, vol. 50, no. 7, pp. 1650–1660, Jul. 2002.

[8] A. T. Erdogan and C. Kizilkale, "Fast and low complexity blind equalization via subgradient projections," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2513–2524, Jul. 2005.

[9] A. T. Erdogan, "A fractionally spaced blind equalization algorithm with global convergence," *Signal Processing*, vol. 88, no. 1, pp. 200–209, 2008.

[10] E. D. Carvalho and D. T. M. Slock, "Maximum-likelihood blind fir multichannel estimation with gaussian prior for the symbols," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP' 97)*, vol. 5. Munich, Germany: IEEE, Apr. 1997, pp. 3593–3596.

[11] V. Zarzoso and P. Comon, "Blind and semi-blind equalization based on the constant power criterion," *IEEE Transactions on Signal Processing*, vol. 53, no. 11, pp. 4363–4375, Nov. 2005.

[12] ——, "Semi-blind constant modulus equalization with optimal step size," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP' 05)*, vol. 3, Philadelphia, PA, Mar. 2005, pp. iii/577–iii/580.

[13] M. Cetin and B. M. Sadler, "Semi-blind sparse channel estimation with constant modulus symbols," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP' 05)*, vol. 3, Philadelphia, PA, Mar. 2005, pp. iii/561–iii/564.

[14] B. B. Yilmaz and A. T. Erdogan, "Compressed training adaptive equalization," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP' 16)*. IEEE, Mar. 2016, pp. 4920–4924.

[15] ——, "Compressed training adaptive MIMO equalization," in *Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC' 16)*. IEEE, Jul. 2016, pp. 1–6.

[16] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "Optimal training for frequency-selective fading channels," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP' 01)*, vol. 4. IEEE, May 2001, pp. 2105–2108.

[17] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Transactions on Communications*, vol. 50, no. 3, pp. 374–377, 2002.

[18] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5862 –5875, Nov. 2010.

[19] Z. Ding, "On convergence analysis of fractionally spaced adaptive blind equalizers," *IEEE Transactions on Signal Processing*, vol. 45, pp. 650–657, Mar. 1997.

[20] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Birkhäuser, 2013, vol. 1, no. 3.

[21] S. Oymak, "Convex relaxation for low-dimensional representation: Phase transitions and limitations," Ph.D. dissertation, California Institute of Technology, Jun. 2015.

[22] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $o(1/k^2)$." *Soviet Mathematics Doklady*, vol. 27, pp. 372–376, 1983.

[23] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010.

[24] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Communications Magazine*, vol. 40, no. 4, pp. 58–66, Apr. 2002.

[25] F. Pancaldi, G. M. Vitetta, R. Kalbasi, N. Al-Dhahir, M. Uysal, and H. Mheidat, "Single-carrier frequency domain equalization," *IEEE Signal Processing Magazine*, vol. 25, no. 5, pp. 37–56, September 2008.

[26] V. Cevher, S. Becker, and M. Schmidt, "Convex optimization for big data: Scalable, randomized, and parallel algorithms for big data analytics," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 32–43, 2014.

[27] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp, "Living on the edge: Phase transitions in convex programs with random data," *Information and Inference*, p. iau005, 2014.

[28] S. Chen, "Semi-blind fast equalization of qam channels using concurrent gradient-newton cma and soft decision-directed scheme," *International Journal of Adaptive Control and Signal Processing*, vol. 24, no. 6, pp. 467–476, 2010.

[29] "Rice university, signal processing information base." [Online]. Available: http://spib.linse.ufsc.br/data/comm/microwave/chan1.mat

[30] M. V. Clark, "Adaptive frequency-domain equalization and diversity combining for broadband wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1385–1395, 1998.

[31] S. Oymak, C. Thrampoulidis, and B. Hassibi, "The squared-error of generalized lasso: A precise analysis," *Proc. Annual Allerton Conference on Communication, Control, and Computing (Allerton'13)*, pp. 1002–1009, Oct. 2013.

Baki B. Yilmaz (S'16) received the B.Sc. and M.Sc. degrees in Electrical and Electronics Engineering from Koc University, Turkey in 2013 and 2015 respectively. He joined Georgia Institute of Technology in Fall 2016 and he is currently pursuing his PhD in School of Electrical and Computer Engineering, focusing on quantifying covert/side-channel information leakage and capacity. He also works on channel equalization and sparse reconstruction. His research interests span areas of electromagnetics, signal processing and information theory.

Alper T. Erdogan (M'00-SM'12) was born in Ankara, Turkey, in 1971. He received the B.S. degree from the Middle East Technical University, Ankara, Turkey, in 1993, and the M.S. and Ph.D. degrees from Stanford University, CA, in 1995 and 1999, respectively. He was a Principal Research Engineer with the Globespan-Virata Corporation (formerly Excess Bandwidth and Virata Corporations) from September 1999 to November 2001. He joined the Electrical and Electronics Engineering Department, Koc University, Istanbul, Turkey, in January 2002, where he is currently a Professor. His research interests include wireless, fiber and wireline communications, adaptive signal processing, optimization, system theory and control, and information theory. Dr. Erdogan is the recipient of several awards including TUBITAK Career Award (2005), Werner Von Siemens Excellence Award (2007), TUBA GEBIP Outstanding Young Scientist Award (2008), TUBITAK Encouragement Award (2010) and Outstanding Teaching Award (2017). He served as an Associate Editor for the IEEE Transactions on Signal Processing, and he was a member of IEEE Signal Processing Theory and Methods Technical Committee.