Detection of Large-Scale Wireless Systems via Sparse Error Recovery

Jun Won Choi, *Member, IEEE*, and Byonghyo Shim, *Senior Member, IEEE*

***Abstract*—In this paper, we propose a new detection algorithm for large-scale wireless systems, referred to as post sparse error detection (PSED) algorithm, that employs a sparse error recovery algorithm to refine the estimate of a symbol vector obtained by the conventional linear detector. The PSED algorithm operates in two steps: First, sparse transformation converting the original nonsparse system into the sparse system whose input is an error vector caused by the symbol slicing; and second, the estimation of the error vector using the sparse recovery algorithm. From the asymptotic mean square error analysis and empirical simulations performed on large-scale wireless systems, we show that the PSED algorithm brings significant performance gain over classical linear detectors while imposing relatively small computational overhead.**

***Index Terms*—Sparse signal recovery, compressive sensing, large-scale systems, orthogonal matching pursuit, sparse transformation, linear minimum mean square error, error correction.**

I. Introduction

As A paradigm guaranteeing the perfect reconstruction of a sparse signal from a small set of linear measurements, compressive sensing (CS) has generated a great deal of interest in recent years. Basic premise of the CS is that the sparse signals can be reconstructed from the compressed measurements as long as the signal to be recovered is sparse (i.e., number of nonzero elements in the vector is small) and the measurement process approximately preserves the energy of the original sparse vector [2], [3]. The CS paradigm works well in many signal processing applications where the signal vector to be reconstructed is sparse in itself or sparse in a transformed domain. In particular, CS technique has been applied to wireless communication applications in the context of sparse channel estimation [4]–[8] and wireless multiuser detection [9]–[11], where the multi-dimensional quantities being estimated exhibit sparsity structure. However, not much work is available for the information vector detection (possible exception can be [12]) mainly because the information vectors being transmitted in a typical communication system are by no means sparse so that the sparse recovery algorithm would not perform better than the conventional receiver algorithm, not to mention having no performance guarantee.

It is to these types of wireless detection problem that this paper is addressed. This problem, which is seemingly unconnected to the CS principle, is prevalent and embraces many of current and future detection problems in wireless communication scenarios such as multiuser detection for Internet of Things (IoT). As an example, consider a scenario where n_t devices simultaneously transmit information to the access point (AP) (see Fig. 1). One of L quasi-orthogonal signature sequences $\mathbf{q}_1, \ldots, \mathbf{q}_L \in \mathbb{R}^{n_r}$ is assigned to each device for the purpose of limiting the inter-user interference. Assuming that the channel is under flat fading, the received vector at the AP is

$$\begin{aligned}\mathbf{y} &= h_1 \mathbf{q}_{i_1} s_1 + \cdots + h_{n_t} \mathbf{q}_{i_{n_t}} s_{n_t} + \mathbf{n} \\ &= \mathbf{H}\mathbf{s} + \mathbf{n} \end{aligned} \tag{1}$$

where $\mathbf{H} = [h_1 \mathbf{q}_{i_1} \ \cdots \ h_{n_t} \mathbf{q}_{i_{n_t}}]$ is the composite channel matrix, $\mathbf{s} = [s_1 \ \cdots s_{n_t}]$ is the vector of transmit symbols for n_t active devices, i_n is the sequence index for the nth device, h_n is the complex channel gain from the nth device to the AP, and $\mathbf{n}$ is the $n_r \times 1$ measurement noise vector. In this setting, the multiuser detection problem boils down to an estimation of the symbol vector $\mathbf{s}$ from $\mathbf{y}$ with the knowledge of $\mathbf{H}$.[1] In the IoT scenarios, the number of devices being served n_t is very large so that it is of importance to develop low complexity multi-user detection algorithm.

Roughly speaking, traditional way of detecting the input signals falls into two categories: *linear detection* and *nonlinear detection* techniques. Linear detection techniques, such as zero forcing (ZF) or linear minimum mean square error (LMMSE) estimation, are simple to implement and easy to use but the performance is in general worse than the nonlinear detectors [9]. Nonlinear detection schemes usually perform better than the linear detection but it requires significant computational overhead. Over the years, various nonlinear detection algorithms for large-scale systems have been proposed. These include fixed-complexity sphere decoder [13], approximate message passing (AMP)-based detector [14]–[16], and sum-product-based detector [17].

Manuscript received August 8, 2016; revised July 7, 2017; accepted August 22, 2017. Date of publication September 4, 2017; date of current version September 26, 2017. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ami Wiesel. This work was supported in part by the research grant from Qualcomm Incorporated and in part by the Institute for Information & Communications Technology Promotion Grant funded by the Korea government (MSIP) (No. 2015-0-00294) and the Ministry of Education (NRF-2017R1D1A1A09000602). This paper was presented in part at the IEEE Global Telecommunications Conference, Austin, TX, USA, December 2014. *(Corresponding author: Byonghyo Shim.)*

J. W. Choi is with the Department of Electrical Engineering, Hanyang University, Seoul 133-791, South Korea (e-mail: junwchoi@hanyang.ac.kr).

B. Shim is with the Institute of New Media and Communications and the School of Electrical and Computer Engineering, Seoul National University, Seoul 151-742, South Korea (e-mail: bshim@snu.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2017.2749214

[1] The channel gains can be estimated using the preamble or pilot signals.

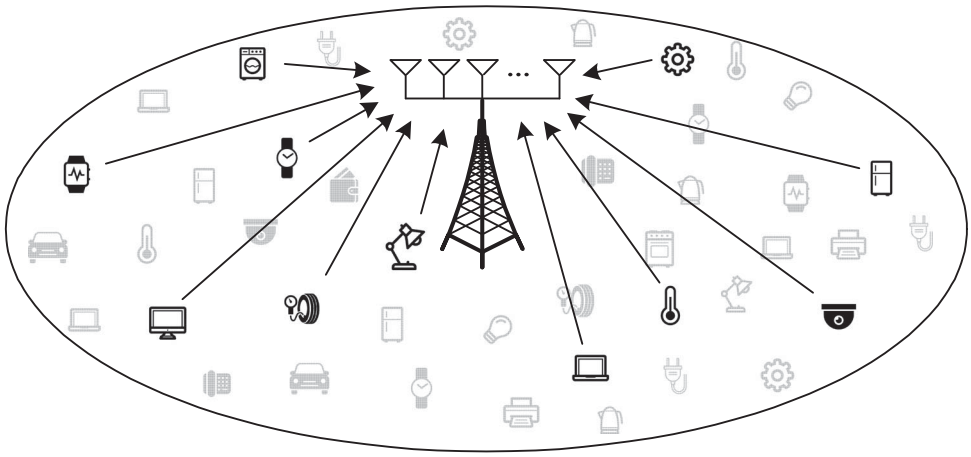


Fig. 1. Multiple access system for IoT applications.

Our approach provides a comparably simple yet effective solution to the large-scale wireless detection problems by deliberately combining the linear detection and a (nonlinear) sparse signal recovery algorithm. The proposed algorithm, henceforth dubbed as *post sparse error detection* (PSED), is based on the simple observation that the conventional linear detection algorithm performs reasonably well and thus the number of errors in the detector output is quite small. In a nutshell, the PSED algorithm operates in two steps.[2] In the first step, the conventional linear detection is performed to generate a *rough* estimate of the transmit symbol vector. Since the performance of conventional detector is acceptable in the operating regime, the error vector obtained by the slicing of detected symbol vector can be readily modeled as a sparse signal. Now, by a simple linear transform, a new system model whose input is the sparse error vector can be obtained. In the second step, the sparse recovery algorithm to estimate this sparse error vector is employed. By cancelling the recovered error vector from the conventional detector output which corresponds to the sum of symbol and error vectors, we obtain more reliable estimate of the original symbol vector.

In our analysis based on the random matrix theory, we show that the asymptotic performance of the proposed PSED algorithm, measured in terms of mean square error (MSE), decays exponentially with signal-to-noise ratio (SNR), which is in contrast to the linear or sublinear decaying behavior of the conventional linear detectors. In fact, we show from the empirical simulations that the PSED scheme outperforms the conventional linear detectors by a large margin.

The rest of this paper is organized as follows. In Section II, we describe the system model and the proposed PSED algorithm. In Section III, the asymptotic performance analysis of the PSED scheme is provided using random matrix theory. In Section V, we present the simulation results and conclude the paper in Section VI.

[2]In [18], two-step processing approach to correct error of real-valued signals when the codeword is corrupted by gross errors on a fraction of entries and a small noise on all the entries has been proposed.

II. Post Sparse Error Detection

A. System Model and Conventional Detectors

The relationship between the transmit symbol in the devices and the received signal vector in the AP in IoT environments can be expressed as

$$\mathbf{y} = \sqrt{P}\mathbf{H}\mathbf{s} + \mathbf{v} \tag{2}$$

where $\mathbf{y} \in \mathbb{C}^{n_r}$ is the received signal vector, $\mathbf{s} \in \mathbb{C}^{n_t}$ is the vector of transmit symbols whose entries are chosen from a set Ω of finite symbol alphabet, $\mathbf{H} \in \mathbb{C}^{n_r \times n_t}$ is the channel matrix, $\mathbf{v} \sim \mathcal{CN}(0, \sigma_v^2 \mathbf{I}_{n_r})$ is the noise vector, and P is the transmitted power. As mentioned, there are two options, linear detection and nonlinear detection schemes, in recovering the transmit information from the received signals. In the linear detection scheme, an estimate $\tilde{\mathbf{s}}$ of the transmit symbol vector is obtained by applying the weight matrix $\mathbf{W} \in \mathbb{C}^{n_r \times n_t}$ to the received vector $\mathbf{y}$

$$\tilde{\mathbf{s}} = \mathbf{W}^H \mathbf{y}. \tag{3}$$

Well-known linear detection schemes include:

- Matched filter (MF): $\mathbf{W} = \frac{1}{\sqrt{P}}\mathbf{H}$
- ZF receiver: $\mathbf{W} = \frac{1}{\sqrt{P}}\mathbf{H}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}$
- LMMSE receiver: $\mathbf{W} = \mathbf{H}\left(\mathbf{H}^H\mathbf{H} + \frac{\sigma_v^2}{P}\mathbf{I}\right)^{-1}$.

The linear detection is simple to implement and computationally efficient, but the performance is typically not better than the nonlinear detection scheme. The nonlinear detectors, such as ML and maximum a posteriori (MAP) detectors, exploit the additional side information that an element of the transmit vector is chosen from the set of finite alphabets. For the lattice that the symbol vector spans, a tree search is performed to find out the solution minimizing the cost function. In the ML detector, for example, a symbol vector $\mathbf{s}$ minimizing the ML cost function $J(\mathbf{s}) = \min_{\mathbf{s}\in\Omega^{n_t}} \|\mathbf{y} - \sqrt{P}\mathbf{H}\mathbf{s}\|^2$ is chosen among all possible candidates. When compared to the linear detection schemes, the nonlinear detector performs better but requires higher computational cost. Sphere decoding (SD) algorithm, for example, performs an efficient ML detection using the closest lattice point

search (CLPS) in a hypersphere with a small radius [20]. In spite of the substantial reduction in complexity over the brute force enumeration, computational burden of the SD algorithm is still a major problem, since the expected complexity is exponential with the problem size [21]. Due to these reasons, in many future wireless scenarios where the dimension of the system matrix is much larger than that of today's wireless systems, both linear and nonlinear principles have their own pros and cons and may not offer an elegant tradeoff between performance and complexity.

B. Sparse Transform via Conventional Detection

When we use conventional detectors to the system model in (2), it is clear that the detector output cannot be always identical to the original information vector $\mathbf{s}$ due to the errors. In fact, in the practical SNR regime, the detector output might contain an error, causing mismatches for a few entries of $\mathbf{s}$.[3] In order to exploit the sparsity of the detection errors, in our approach we convert the non-sparse system into the sparse one. Conventional detection together with the symbol slicing serves our purpose since the estimated symbol vector is *roughly* accurate, and hence, the resulting error vector (defined as the difference between the original symbol vector and the sliced estimate) can be modeled as a sparse signal. Denoting the estimate of a symbol vector as $\tilde{\mathbf{s}}$ and its sliced version as $\hat{\mathbf{s}}$, we have

$$\hat{\mathbf{s}} = Q(\tilde{\mathbf{s}}) = \mathbf{s} - \mathbf{e} \tag{4}$$

where $Q(\cdot)$ is the slicing function and $\mathbf{e}$ is the error vector. As mentioned, in an operational regime of communication systems, the number of nonzero entries (i.e., *real errors*) in $\mathbf{e}$ would be small so that the error vector is readily modeled as a sparse vector. Suppose the dimension of the symbol vector $\mathbf{s}$ is 16 and the symbol error rate is 10%, then the probability that 5 or less elements being in error is 99.7%. As long as the error vector is sparse, by transmitting this error vector, one can construct a sparse system expressed in terms of the error vector $\mathbf{e}$. This task, henceforth referred to as the *sparse transform*, can be realized by the re-generation of the received signal from the detected symbol $\hat{\mathbf{s}}$ followed by the subtraction as

$$\mathbf{y}' = \mathbf{y} - \sqrt{P}\mathbf{H}\hat{\mathbf{s}}, \tag{5}$$

where $\mathbf{y}'$ is the newly obtained received vector (see Fig. 2). Then, from (2), (4), and (5), the new measurement vector $\mathbf{y}'$ is

$$\begin{aligned}\mathbf{y}' &= \sqrt{P}\mathbf{H}(\mathbf{s} - \hat{\mathbf{s}}) + \mathbf{v} \\ &= \sqrt{P}\mathbf{H}\mathbf{e} + \mathbf{v}.\end{aligned} \tag{6}$$

Interestingly, by adding trivial operations (matrix-vector multiplication and subtraction), one can convert the original non-sparse system into the sparse system whose input is an error vector associated with the conventional detector. Note that the symbol slicing is essential in sparsifying the error vector and we have two options:

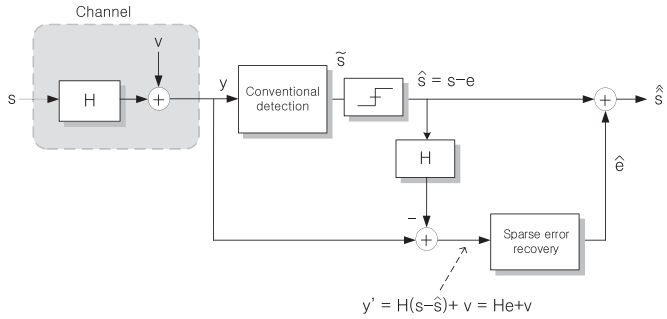


Fig. 2. Overall structure of the proposed PSED detection algorithm.

- Hard-slicing: hard-slicing literally performs the hard decision of symbol estimate. The slicer function maps the input to the closest value in the symbol set Γ (i.e., $Q(z) = \arg\min_{\gamma\in\Gamma} \|z - \gamma\|_2$). By exploiting the discrete property of the transmit vectors in (6), we can enforce the sparsity of the input (error vector) for the modified system. Main benefit of the hard slicing is that it entirely removes the residual interference and noise when the estimate lies in the decision region of the original symbol.
- Soft-slicing: when the *a prior* information on the source exists, soft slicing can be considered. One possible way is to use an MMSE-based soft slicing ($\hat{s}_i = E[s_i|\tilde{s}_i]$), where s_i denotes the i-th element of $\mathbf{s}$ [22]. For example, when the decoder feeds back the log-likelihood ratio (LLR) information L_b on the information bit, one can transform this into the symbol prior information $P_r(s_i)$. Using the nonuniform symbol probability $P(s_i)$, we can estimate $\hat{s}_i$ as

$$\begin{aligned}\hat{s}_i &= E[s_i|\tilde{s}_i] \\ &= \sum_{s_i\in\Omega} s_i P_r(s_i|\tilde{s}_i) = \frac{\sum_{s_i\in\Omega} s_i P_r(\tilde{s}_i|s_i)P_r(s_i)}{\sum_{s_i\in\Omega} P_r(\tilde{s}_i|s_i)P_r(s_i)}.\end{aligned} \tag{7}$$

  When the linear detectors in (3) are used to obtain $\tilde{\mathbf{s}}$, $P_r(\tilde{s}_i|s_i)$ is given by

$$P_r(\tilde{s}_i|s_i) = \frac{1}{2\pi\sigma_s^2}\exp\left(-\frac{1}{\sigma_s^2}\left(\tilde{s}_i - \sqrt{P}\mathbf{w}_i^H\mathbf{h}_i s_i\right)^2\right), \tag{8}$$

  where $\sigma_s^2 = \mathbf{w}_i^H\left(P\mathbf{H}\mathbf{H}^H + \sigma_v^2\mathbf{I}\right)\mathbf{w}_i - P(\mathbf{w}_i^H\mathbf{h}_i)^2$, and $\mathbf{w}_i$ and $\mathbf{h}_i$ are the i-th column of $\mathbf{W}$ and $\mathbf{H}$, respectively. In the hard slicing, all entries of $\mathbf{e}$ are zero except for those associated with the detection errors. Whereas, entries unassociated with the detection errors might have small nonzero magnitude in the soft-slicing, yielding so called approximately sparse vector $\mathbf{e}$.[4]

In our analysis and derivation that follow, we will mainly focus on the hard-slicing based PSED technique for analytical simplicity.

[3] In the operating regime of communication systems, symbol error rate (SER) is typically less than 10%.

[4] By approximately sparse signal, we mean a vector containing most of energy in only a few elements.

TABLE I
OPERATIONS OF THE PSED DETECTOR

Input:	$\mathbf{y}, \mathbf{H}$ Output: $\hat{\mathbf{s}}_{\text{final}}$
Step 1:	Perform conventional detection to obtain $\tilde{\mathbf{s}}$.
Step 2:	Perform sparse transform, i.e., $\mathbf{y}' = \mathbf{y} - \sqrt{P}\mathbf{H}\hat{\mathbf{s}}$ where $\hat{\mathbf{s}} = Q(\tilde{\mathbf{s}})$.
Step 3:	Apply the sparse recovery algorithm to $\mathbf{y}'$ to estimate $\hat{\mathbf{e}}$.
Step 4:	Correct the detection errors in $\hat{\mathbf{s}}$, i.e., $\hat{\hat{\mathbf{s}}} = \hat{\mathbf{s}} + \hat{\mathbf{e}}$.
Step 5:	Generate the final symbol estimate, i.e., $\hat{\mathbf{s}}_{\text{final}} = Q(\hat{\hat{\mathbf{s}}})$.

C. Recovery of Sparse Error Vector

Once the non-sparse system is converted into the sparse one, we can use the sparse recovery algorithm to estimate the error vector. To be specific, using newly obtained measurement vector $\mathbf{y}'$ and the channel matrix $\mathbf{H}$, sparse recovery algorithm estimates the error vector $\mathbf{e}$ (see Fig. 2). There are many algorithms designed to recover the sparse vector in the presence of noise. Well-known examples include basis pursuit de-noising (BPDN) [23] and orthogonal matching pursuit (OMP) [24]. We will say more about this in Section II-D.

Once the output $\hat{\mathbf{e}}$ of the sparse recovery algorithm is obtained, we add it to the sliced detector output $\hat{\mathbf{s}}$, generating the refined symbol vector $\hat{\hat{\mathbf{s}}}$

$$\hat{\hat{\mathbf{s}}} = \hat{\mathbf{s}} + \hat{\mathbf{e}} = (\mathbf{s} - \mathbf{e}) + \hat{\mathbf{e}} = \mathbf{s} + (\hat{\mathbf{e}} - \mathbf{e}). \tag{9}$$

If the error estimate is accurate, i.e., $\hat{\mathbf{e}} \approx \mathbf{e}$, then the magnitude of the error difference $\epsilon = \hat{\mathbf{e}} - \mathbf{e}$ would be small so that the re-estimated symbols $\hat{\hat{\mathbf{s}}}$ becomes more accurate than the initial estimate $\hat{\mathbf{s}}$. As long as the sparsity of the error vector is ensured (i.e., number of nonzero elements in error vector $\mathbf{e}$ is small), an output of the sparse recovery algorithm $\hat{\mathbf{e}}$ would be faithful (i.e., $\|\hat{\mathbf{e}} - \mathbf{e}\|_2^2 < \|\mathbf{e}\|_2^2$) so that the refined symbol vector will be more reliable than the original estimate. We provide more deliberate analysis in Section III.

It is worth mentioning that the sparse error recovery process is conceptually analogous to the decoding process of the linear block code [26]. First, the sliced symbol vector $\hat{\mathbf{s}} = \mathbf{s} - \mathbf{e}$ can be viewed as a received vector (often denoted by $\mathbf{r}$ in the coding theory) which is typically expressed as the sum of the transmit codeword and the error vector. Also, the new observation vector $\mathbf{y}' = \sqrt{P}\mathbf{H}\mathbf{e} + \mathbf{v}$ is similar to the syndrome, the product of the error vector and the transpose of parity check matrix. Note that the syndrome is a sole function of the error vector and does not depend on the transmit codeword. Similarly, the new observation vector $\mathbf{y}'$ is a function of $\mathbf{e}$ and independent of the transmit vector $\mathbf{s}$. Furthermore, in the linear block code, the decoded error pattern is correct only when the cardinality of the error vector is within the error correction capability t (i.e., $\|\mathbf{e}\|_0 < t$) and similar behavior occurs to the problem at hand since an output of the sparse recovery algorithm will be reliable only when the error vector $\mathbf{e}$ is sparse (i.e., $\|\mathbf{e}\|_0 \ll n_t$). Finally, the error correction is performed in the decoding process by adding the reconstructed error vector $\hat{\mathbf{e}}$ to the received vector $\mathbf{r}$ and the same is true for the proposed algorithm (see (9)). In Table I and II, we summarize the proposed PSED algorithm and similarity between the PSED algorithm and the decoding process of the linear block code.

D. Sparse Recovery Algorithm

In this subsection, we briefly describe the sparse recovery algorithm used for the proposed PSED scheme. Note that an approach often called Lasso or Basis pursuit de-noising (BPDN) formulates the problem to recover the sparse signal in the noisy scenario as [34]

$$\min_{\mathbf{e}} \ \frac{1}{2}\|\mathbf{y}' - \sqrt{P}\mathbf{H}\mathbf{e}\|_2^2 + \lambda\|\mathbf{e}\|_1$$

where λ is the penalty term to control the amount of weight given to the sparsity of the desired signal $\mathbf{e}$. This problem is in essence a convex optimization problem and there are many algorithms to solve this type of problem [35]. As an alternative to the convex optimization problem, greedy algorithms have received great deal of interest in recent years [4]. In a nutshell, greedy algorithms attempt to find the support of $\mathbf{e}$ (i.e., the set of columns in $\mathbf{H}$ constructing $\mathbf{y}'$) in an iterative fashion, generating a sequence of the estimate for $\mathbf{e}$. In the OMP algorithm, for example, a column of $\mathbf{H}$ that is maximally correlated with the modified measurement vector (residual $\mathbf{r}$) is chosen as an element of the support set $\hat{\mathcal{E}}$ [24]. Then the estimate of the desired signal $\hat{\mathbf{e}}_{\hat{\mathcal{E}}}$ is constructed by projecting $\mathbf{y}'$ onto the subspace spanned by the columns supported by $\hat{\mathcal{E}}$. That is,

$$\hat{\mathbf{e}}_{\hat{\mathcal{E}}} = \frac{1}{\sqrt{P}}\left(\mathbf{H}_{\hat{\mathcal{E}}}^H \mathbf{H}_{\hat{\mathcal{E}}}\right)^{-1}\mathbf{H}_{\hat{\mathcal{E}}}^H \mathbf{y}' \tag{10}$$

Finally, we update the residual $\mathbf{r}$ so that it contains the measurement excluding those included by the estimated support set ($\mathbf{r} = \mathbf{y}' - \sqrt{P}\mathbf{H}_{\hat{\mathcal{E}}}\hat{\mathbf{e}}_{\hat{\mathcal{E}}}$). In case the OMP algorithm identifies the support $\mathcal{E}$ accurately ($\hat{\mathcal{E}} = \mathcal{E}$), one can remove all non-support elements in $\mathbf{e}$ and corresponding columns in $\mathbf{H}$ so that one can obtain the overdetermined system model

$$\mathbf{y}' = \sqrt{P}\mathbf{H}_{\mathcal{E}}\mathbf{e}_{\mathcal{E}} + \mathbf{v}. \tag{11}$$

The final estimate is equivalent to the best possible estimate called the Oracle LS estimator,[5]

$$\hat{\mathbf{e}}_{\mathcal{E}} = \frac{1}{\sqrt{P}}\left(\mathbf{H}_{\mathcal{E}}^H \mathbf{H}_{\mathcal{E}}\right)^{-1}\mathbf{H}_{\mathcal{E}}^H \mathbf{y}'.$$

While the OMP algorithm is simple to implement and also computationally efficient, due to the selection of the single candidate in each iteration, the performance depends heavily on the selection of index. In fact, the output of OMP would be wrong if an incorrect index (index not contained in the support) is chosen in the middle of the search. In order to alleviate this drawback, various alternatives investigating *multiple indices* have been suggested. Recent developments of this approach include compressive sampling matching pursuit (CoSaMP) [36], subspace pursuit (SP) [37], and generalized OMP [38].

In this work, we employ the multipath matching pursuit (MMP) algorithm, a recently proposed near-ML greedy tree search algorithm, in recovering the sparse error vector [27].

[5]In case the *a prior* information on signal and noise is available, one can alternatively use the linear minimum mean square error (LMMSE) estimate. For example, if $\sigma_e^2 = E[|\mathbf{e}_{\mathcal{E}}|^2]$ and $\sigma_v^2 = E[|\mathbf{v}|^2]$, we have $\hat{\mathbf{e}}_{\mathcal{E}} = \frac{1}{\sqrt{P}}(\mathbf{H}_{\mathcal{E}}^H \mathbf{H}_{\mathcal{E}} + \frac{\sigma_v^2}{P\sigma_e^2}\mathbf{I})^{-1}\mathbf{H}_{\mathcal{E}}^H \mathbf{y}'$

TABLE II
SIMILARITIES BETWEEN THE DECODING PROCESS OF THE LINEAR BLOCK CODE AND THE PSED ALGORITHM. IN THE DECODING PROCESS, $\mathbf{r}$, $\mathbf{c}$, AND $\mathbf{s}$ DENOTE RECEIVED, CODEWORD, AND SYNDROME VECTORS, RESPECTIVELY, AND $\mathbf{H}$ IS THE PARITY CHECK MATRIX

Decoding process of the linear block code	Proposed PSED algorithm
Received vector: $\mathbf{r} = \mathbf{c} + \mathbf{e}$	Detected symbol vector: $\hat{\mathbf{s}} = \mathbf{s} - \mathbf{e}$
Syndrome vector: $\mathbf{s} = \mathbf{e}\mathbf{H}^T$	New observation vector: $\mathbf{y}' = \sqrt{P}\mathbf{H}\mathbf{e} + \mathbf{v}$
Recovered codeword: $\hat{\mathbf{c}} = \mathbf{r} + \hat{\mathbf{e}}$	Re-detected symbol vector: $\hat{\hat{\mathbf{s}}} = \hat{\mathbf{s}} + \hat{\mathbf{e}}$

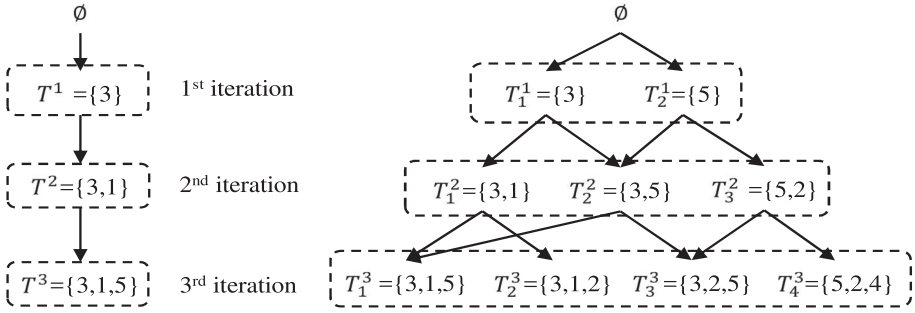


Fig. 3. Comparison between the OMP and the MMP algorithm ($L = 2$ and $K = 3$). While OMP maintains a single candidate T^k in each iteration, MMP investigates multiple promising candidates T_j^k $(1 \le j \le L)$ (the subscript j counts the candidate in the i-th iteration).

While aforementioned greedy recovery algorithms identify the elements of support sequentially and choose a single support estimate, MMP performs the list tree search to find the multiple promising supports (we henceforth refer to it as *support candidates*). Among multiple candidates, the best one minimizing the residual power is chosen as an output in the last minute. As shown in Fig. 3, each support candidate brings forth L child candidates in the MMP algorithm. Since multiple promising candidates are investigated, one can expect that MMP outperforms the sequential greedy algorithm returning a single support candidate. In fact, it has been shown that MMP performs close to the best possible estimator using genie support information (called Oracle estimator) for high SNR regime [27, Theorem 4.6]. The operation of MMP is summarized in Table III.

TABLE III
THE MMP ALGORITHM

Input: measurement $\mathbf{y}$, sensing matrix $\mathbf{\Phi}$, sparsity K, number of path L
Output: estimated signal $\hat{\mathbf{e}}$
Initialization: $k := 0$ (iteration index), $\mathbf{r}^0 := \mathbf{y}$ (initial residual), $S^0 := \{\emptyset\}$
while $k < K$ **do**
  $k := k + 1$, $u := 0$, $S^k := \emptyset$
  for $i = 1$ **to** $|S^{k-1}|$ **do**
    $\tilde{\pi} := \arg\max_{|\pi|=L} \|(\mathbf{H}^H \mathbf{r}_i^{k-1})_\pi\|_2^2$ (*choose L best indices*)
    for $j = 1$ **to** L **do**
      $s_{tmp} := s_i^{k-1} \cup \{\tilde{\pi}_j\}$ (*construct a temporary path*)
      if $s_{tmp} \notin S^k$ **then** (*check if the path already exists*)
        $u := u + 1$ (*candidate index update*)
        $s_u^k := s_{tmp}$ (*path update*)
        $S^k := S^k \cup \{s_u^k\}$ (*update the set of path*)
        $\hat{\mathbf{e}}_u^k := \mathbf{H}_{s_u^k}^\dagger \mathbf{y}$ (*perform estimation*)
        $\mathbf{r}_u^k := \mathbf{y} - \mathbf{H}_{s_u^k} \hat{\mathbf{e}}_u^k$ (*residual update*)
      end if
    end for
  end for
end while
$u^* := \arg\min_u \|\mathbf{r}_u^K\|_2^2$ (*find index of the best candidate*)
$s^* := s_{u^*}^K$
return $\hat{\mathbf{e}} = \mathbf{H}_{s^*}^\dagger \mathbf{y}$

III. PERFORMANCE ANALYSIS

In this section, we analyze the performance of the proposed PSED algorithm. As a performance metric, we consider the normalized MSE between the original symbol vector $\mathbf{s}$ and the output of PSED algorithm $\hat{\mathbf{s}}$, which is defined as

$$\text{MSE} = \frac{1}{n_t} E[\|\mathbf{s} - \hat{\mathbf{s}}\|^2] \tag{12}$$

$$= \frac{1}{n_t} E[\|\mathbf{s} - (\mathbf{s} + \hat{\mathbf{e}} - \mathbf{e})\|^2] \tag{13}$$

$$= \frac{1}{n_t} E[\|\mathbf{e} - \hat{\mathbf{e}}\|^2] \tag{14}$$

One can observe that the MSE associated with the symbol vector $\mathbf{s}$ is equivalent to the MSE associated with the error vector $\mathbf{e}$. Since we use the random matrix theory for analytical tractability, our analysis is accurate when the dimension of the channel matrix is large. In our analysis, we assume that the sparse recovery algorithm identifies the support $\mathbf{e}$ (index set of nonzero elements) of the error vector accurately. This is true when the channel matrix satisfies a property called the restricted isometry property (RIP)[6] and the signal power is sufficiently higher than the poise power (see, e.g., [11, Theorem 4.6]).

When the support of $\mathbf{e}$ is identified, all non-support elements in $\mathbf{e}$ and columns of $\mathbf{H}$ associated with these can be removed

[6]A matrix $\mathbf{H} \in \mathcal{R}^{n_r \times n_t}$ is said to satisfy the RIP condition if $(1 - \delta_K)\|\mathbf{x}\|_2^2 \le \|\mathbf{H}\mathbf{x}\|_2^2 \le (1 + \delta_K)\|\mathbf{x}\|_2^2$.

from the system model. The resulting overdetermined system model is[7]

$$\mathbf{y}' = \sqrt{P}\mathbf{H}\mathbf{e} + \mathbf{v} \tag{15}$$

$$= \sqrt{P}\mathbf{H}_{\mathcal{E}}\mathbf{e}_{\mathcal{E}} + \mathbf{v}. \tag{16}$$

Note that most of greedy sparse recovery algorithms use the linear squares (LS) solution in generating the estimate of error vector $\hat{\mathbf{e}}_{\mathcal{E}}$. In this case, the estimate $\hat{\mathbf{e}}_{\mathcal{E}}$ is given by [28]

$$\hat{\mathbf{e}}_{\mathcal{E}} = \frac{1}{\sqrt{P}}\left(\mathbf{H}_{\mathcal{E}}^H\mathbf{H}_{\mathcal{E}}\right)^{-1}\mathbf{H}_{\mathcal{E}}^H\mathbf{y}'. \tag{17}$$

From (16) and (17), the MSE is expressed as

$$\mathrm{MSE}_{psed} = \frac{1}{n_t}E[\|\mathbf{e}-\hat{\mathbf{e}}\|^2] \tag{18}$$

$$= \frac{1}{n_t}E[\|\mathbf{e}_{\mathcal{E}}-\hat{\mathbf{e}}_{\mathcal{E}}\|^2] \tag{19}$$

$$= E_{\mathbf{H}}\left[\mathrm{tr}E_{\mathcal{E},\mathbf{v}}\left[\frac{1}{n_tP}\left(\mathbf{H}_{\mathcal{E}}^H\mathbf{H}_{\mathcal{E}}\right)^{-1}\mathbf{H}_{\mathcal{E}}^H\mathbf{v}\mathbf{v}^H \cdot \mathbf{H}_{\mathcal{E}}\left(\mathbf{H}_{\mathcal{E}}^H\mathbf{H}_{\mathcal{E}}\right)^{-1}\Big|\mathbf{H}\right]\right], \tag{20}$$

where $E_{\mathbf{x}}[\cdot]$ is the expectation with respect to the random variable $\mathbf{x}$. One thing to notice is that $\mathbf{H}_{\mathcal{E}}$ and $\mathbf{v}$ are correlated with each other since the error pattern associated with $\mathbf{H}_{\mathcal{E}}$ depends on the realization of the noise vector $\mathbf{v}$. Since this makes the evaluation of (20) very difficult, we take an alternative approach and investigate a lower bound of the MSE. First, let $\tilde{\mathcal{E}}$ be the set of indices whose elements are randomly chosen from the set of all column indices $\Omega = \{1, 2, \ldots, n\}$ and the cardinality of $\tilde{\mathcal{E}}$ is the same as that of $\mathcal{E}$ (i.e., $|\tilde{\mathcal{E}}| = |\mathcal{E}|$). Then, we conjecture that

$$\mathrm{MSE}_{psed} \geq E_{\mathbf{H}}\left[\mathrm{tr}E_{\mathcal{E},\mathbf{v}}\left[\frac{1}{n_tP}\left(\mathbf{H}_{\tilde{\mathcal{E}}}^H\mathbf{H}_{\tilde{\mathcal{E}}}\right)^{-1}\mathbf{H}_{\tilde{\mathcal{E}}}^H\mathbf{v}\mathbf{v}^H \cdot \mathbf{H}_{\tilde{\mathcal{E}}}\left(\mathbf{H}_{\tilde{\mathcal{E}}}^H\mathbf{H}_{\tilde{\mathcal{E}}}\right)^{-1}\Big|\mathbf{H}\right]\right] \tag{21}$$

$$= E_{\mathbf{H}}\left[\frac{1}{\mathrm{SNR}}\mathrm{tr}E_{\mathcal{E},\mathbf{v}}\left[\frac{1}{n_t}\left(\mathbf{H}_{\tilde{\mathcal{E}}}^H\mathbf{H}_{\tilde{\mathcal{E}}}\right)^{-1}\Big|\mathbf{H}\right]\right]. \tag{22}$$

where $\mathrm{SNR} = \frac{P}{\sigma_v^2}$. Note that this conjecture is justified by the fact that columns associated with $\mathcal{E}$ are caused by the detection errors while those associated with $\tilde{\mathcal{E}}$ are randomly chosen so that the former would yield higher MSE than the latter. To judge the effectiveness of this lower bound, we perform empirical simulations of two quantities (i.e., (20) and the right-hand side of (21)) for i.i.d. complex Gaussian random matrix. We observe from Fig. 4 that the conjecture holds true empirically and also the lower bound in (21) is tight across the board.

We next investigate an asymptotic behavior of the term $E[\frac{1}{\mathrm{SNR}}\mathrm{tr}E[\frac{1}{n_t}(\mathbf{H}_{\tilde{\mathcal{E}}}^H\mathbf{H}_{\tilde{\mathcal{E}}})^{-1}\Big|\mathbf{H}]]$. We assume that the elements of the channel matrix $\mathbf{H}$ are i.i.d. complex Gaussian ($h_{ij} \sim \mathcal{CN}(0, 1/n_r)$) and the dimension of the channel matrix $\mathbf{H}$ is sufficiently large (i.e., $n_t, n_r \to \infty$) with a constant ratio $\frac{n_t}{n_r} \to \beta$. Let $\mathbf{H}_{\tilde{\mathcal{E}}}$ be the submatrix generated by randomly choosing $|\mathcal{E}|$ columns of $\mathbf{H}$. Then we have

$$\frac{1}{|\mathcal{E}|}\mathrm{tr}\left(\mathbf{H}_{\tilde{\mathcal{E}}}^H\mathbf{H}_{\tilde{\mathcal{E}}}\right)^{-1} \to \frac{1}{1-\beta'}$$

where $\beta' = \frac{|\mathcal{E}|}{n_r}$ [29, Section 2.3.3]. Notice that when the dimension of the matrix is large, the quantity converges to the deterministic limit depending only on the ratio β'. Therefore, it suffices to evaluate a single matrix realization in (22) and hence the outer expectation with respect to the channel realization is unnecessary.

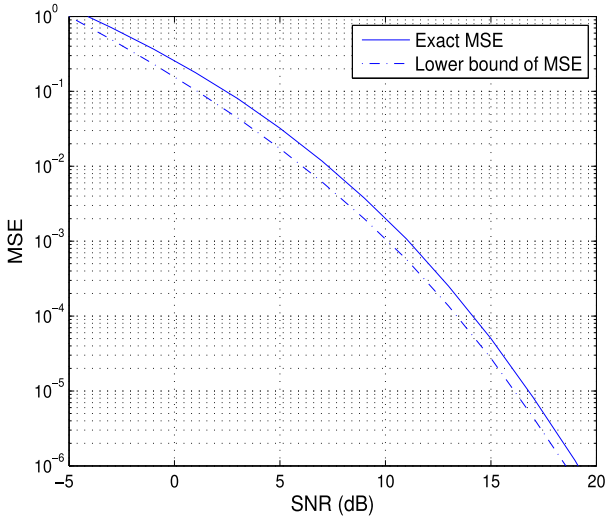


Fig. 4. Comparison of the exact MSE in (20) and lower bound in the right-hand side of (21) for i.i.d. complex Gaussian random matrix. We set $n_r = n_t = 128$ and use BPSK modulation.

We next investigate the distribution of $|\mathcal{E}|$ (cardinality of $\mathcal{E}$) which corresponds to the number of errors in the output streams of the linear detector. Here we consider the LMMSE detector as a conventional linear detector. Note that the event of detection errors is related to the post-detection signal-to-interference-and-noise-ratio (SINR) at the output of the LMMSE detector. While the SINR for each output stream relies on a channel realization, when the system size is sufficiently large (i.e., $n_t, n_r \to \infty$ and $\frac{n_t}{n_r} \to \beta$), the SINR for all output streams approaches to the same quantity [29]

$$\mathrm{SINR}_i^{(\mathrm{LMMSE})} \longrightarrow \mathrm{SINR}_\infty^{(\mathrm{LMMSE})} = \mathrm{SNR} - \frac{\mathcal{F}(\mathrm{SNR}, \beta)}{4}$$

where $\mathcal{F}(x,z) = \left(\sqrt{x(1+\sqrt{z})^2+1} - \sqrt{x(1-\sqrt{z})^2+1}\right)^2$ and $\mathrm{SINR}_i^{(\mathrm{LMMSE})}$ is the SINR for the ith stream.

In addition, one can show that the output streams of the LMMSE receiver are asymptotically uncorrelated with each other (see Appendix A). Thus, the detection problem for each output stream can be considered separately and the number of errors $|\mathcal{E}|$ is approximated as a Binomial distribution with the success probability P_e, where P_e is the probability of error event

[7] $\mathbf{H}_D$ is a submatrix of $\mathbf{H}$ that only contains columns indexed by D. For example, if $D = \{1, 4, 5\}$, then $\mathbf{H}_D = [\mathbf{h}_1\ \mathbf{h}_4\ \mathbf{h}_5]$ where $\mathbf{h}_j$ is the j-th column of $\mathbf{H}$.

for each output stream. That is,

$$Pr(|\mathcal{E}|=t) \approx \binom{n_t}{t} P_e^t (1-P_e)^{n_t-t}$$

When n_t is large, the Binomial distribution with parameter n_t and P_e approaches to the Normal distribution $\mathcal{N}(n_t P_e, n_t P_e(1-P_e))$ by DeMoivre-Laplace theorem [30] and hence one can show that (see Appendix B)

$$P_r\left(\left|\frac{|\mathcal{E}|}{n_r} - P_e\beta\right| > \epsilon\right) \to 0 \tag{23}$$

$$P_r\left(\left|\frac{|\mathcal{E}|}{n_t} - P_e\right| > \epsilon\right) \to 0. \tag{24}$$

Our discussions so far can be summarized as follows.

1) $\frac{1}{|\mathcal{E}|}\mathrm{tr}\left(\mathbf{H}_{\tilde{\mathcal{E}}}^H \mathbf{H}_{\tilde{\mathcal{E}}}\right)^{-1}$ converges to $\frac{1}{1-\beta'} = \frac{1}{1-\frac{|\mathcal{E}|}{n_r}}$ irrespective of the channel realization.
2) $\frac{|\mathcal{E}|}{n_r}$ and $\frac{|\mathcal{E}|}{n_t}$ converge in probability to $P_e\beta$ and P_e, respectively.

Using these, one can show that

$$\begin{aligned}
&\frac{1}{\mathrm{SNR}}\mathrm{tr}E\left[\frac{1}{n_t}\left(\mathbf{H}_{\tilde{\mathcal{E}}}^H \mathbf{H}_{\tilde{\mathcal{E}}}\right)^{-1}\middle|\mathbf{H}\right] \\
&\quad= \frac{1}{\mathrm{SNR}}E\left[\frac{|\mathcal{E}|}{n_t}\frac{1}{|\mathcal{E}|}\mathrm{tr}\left(\mathbf{H}_{\tilde{\mathcal{E}}}^H \mathbf{H}_{\tilde{\mathcal{E}}}\right)^{-1}\middle|\mathbf{H}\right] && (25)\\
&\quad\to \frac{1}{\mathrm{SNR}}E\left[\frac{|\mathcal{E}|}{n_t}\frac{1}{1-\frac{|\mathcal{E}|}{n_r}}\right] && (26)\\
&\quad= \frac{1}{\mathrm{SNR}}\frac{n_r}{n_t}E\left[\frac{\frac{|\mathcal{E}|}{n_r}}{1-\frac{|\mathcal{E}|}{n_r}}\right] && (27)\\
&\quad\geq \frac{1}{\mathrm{SNR}}\frac{n_r}{n_t}\frac{E\frac{|\mathcal{E}|}{n_r}}{1-E\frac{|\mathcal{E}|}{n_r}} && (28)\\
&\quad= \frac{1}{\mathrm{SNR}}\frac{1}{\beta}\frac{P_e\beta}{1-P_e\beta} && (29)\\
&\quad= \frac{1}{\mathrm{SNR}}\frac{P_e}{1-P_e\beta}. && (30)
\end{aligned}$$

where (28) is from Jensen's inequality. Noting that $\frac{|\mathcal{E}|}{n_r} \ll 1$, we see that the obtained lower bound is tight.[8]

In the high SNR regime, $P_e \ll 1$ and hence

$$\mathrm{MSE}_{psed} > \frac{1}{\mathrm{SNR}}\mathrm{tr}E\left[\frac{1}{n_t}\left(\mathbf{H}_{\tilde{\mathcal{E}}}^H \mathbf{H}_{\tilde{\mathcal{E}}}\right)^{-1}\middle|\mathbf{H}\right] \gtrapprox \frac{P_e}{\mathrm{SNR}}. \tag{31}$$

Since P_e is a function of SINR, the obtained lower bound of MSE_{psed} is a function of SNR and SINR. Indeed, when $n_t, n_r \to \infty$ and $\frac{n_t}{n_r} \to \beta$, the residual interference plus noise for the LMMSE detector approaches to Normal distribution [31], [32] so that P_e can be readily expressed as a function of the SINR. For example, for the binary phase shift keying (BPSK), the error probability P_e in symbol detection approaches [33]

$$P_e = \mathrm{Erf}(\mathrm{SINR}_\infty^{(\mathrm{LMMSE})}) = Q\left(\sqrt{2\mathrm{SINR}_\infty^{(\mathrm{LMMSE})}}\right) \tag{32}$$

where $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}}\exp(-\frac{t}{2})dt$. Using $Q(x) > \frac{x}{1+x^2}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, we have

$$\begin{aligned}
\mathrm{MSE}_{psed} &> \frac{Q\left(\sqrt{2\mathrm{SINR}_\infty^{(\mathrm{LMMSE})}}\right)}{\mathrm{SNR}} \\
&> \frac{1}{\mathrm{SNR}}\frac{\sqrt{2\mathrm{SINR}_\infty^{(\mathrm{LMMSE})}}}{1+2\mathrm{SINR}_\infty^{(\mathrm{LMMSE})}} \\
&\quad\cdot\frac{1}{\sqrt{2\pi}}\exp\left(-\mathrm{SINR}_\infty^{(\mathrm{LMMSE})}\right) && (33)
\end{aligned}$$

In high SNR regime, we have [19]

$$\mathrm{SINR}_\infty^{(\mathrm{LMMSE})} \approx \begin{cases} \sqrt{\mathrm{SNR}} & \beta = 1 \\ (1-\beta)\mathrm{SNR} + \frac{\beta}{1-\beta} & \beta < 1 \end{cases} \tag{34}$$

and hence

$$\mathrm{MSE}_{psed} \gtrapprox \begin{cases} \frac{1}{2\sqrt{\pi}}\frac{1}{\mathrm{SNR}^{5/4}}\exp\left(-\sqrt{\mathrm{SNR}}\right) & \text{if } \beta = 1, \\ \frac{1}{2\sqrt{\pi(1-\beta)}}\frac{1}{\mathrm{SNR}^{3/2}}\exp\left(-(1-\beta)\mathrm{SNR}\right) & \text{if } \beta < 1 \end{cases}. \tag{35}$$

On the other hand, MSE of the conventional linear MMSE detector is given by [29]

$$\begin{aligned}
\mathrm{MSE}_{conv} &= \frac{1}{n_t}E_{\mathbf{H}}\left[\mathrm{tr}(\mathbf{I}+\mathrm{SNR}\mathbf{H}^H\mathbf{H})^{-1}\right] \\
&\longrightarrow 1 - \frac{\mathcal{F}(\mathrm{SNR},\beta)}{4\beta\mathrm{SNR}}. && (36)
\end{aligned}$$

Further, in high SNR regime, MSE_{conv} can be expressed as[19, Chap. 6.3]

$$\begin{aligned}
\mathrm{MSE}_{conv} &\to 1 - \frac{\mathcal{F}(\mathrm{SNR},\beta)}{4\beta\mathrm{SNR}} \\
&\approx \begin{cases} \frac{1}{\sqrt{\mathrm{SNR}}} & \text{if } \beta = 1, \\ \frac{1}{(1-\beta)\mathrm{SNR}} & \text{if } \beta < 1. \end{cases} && (37)
\end{aligned}$$

It is interesting to compare (35) and (37), asymptotic MSEs of the proposed and the conventional schemes when the dimension of the matrix goes to infinity. While the MSE of the conventional method decays linearly or sublinearly with SNR, the MSE of the PSED decays exponentially with SNR. In Fig. 5, we plot the MSE in (35) and (37) and the empirical MSEs as a function of the SNR when the complex Gaussian system matrix is employed. In Fig. 5(a)–(c), we use the BPSK modulation, and in Fig. 5(d), we use quadrature amplitude modulation (QAM)

[8] $f(x) = \frac{x}{1-x}$ is a convex function for $0 < x < 1$ and hence $E\left[\frac{x}{1-x}\right] \geq \frac{E[x]}{1-E[x]}$. In our case, $x = \frac{|\mathcal{E}|}{n_r}$ and hence $x \ll 1$ so that $f(x) \approx x$ and the lower bound is tight.

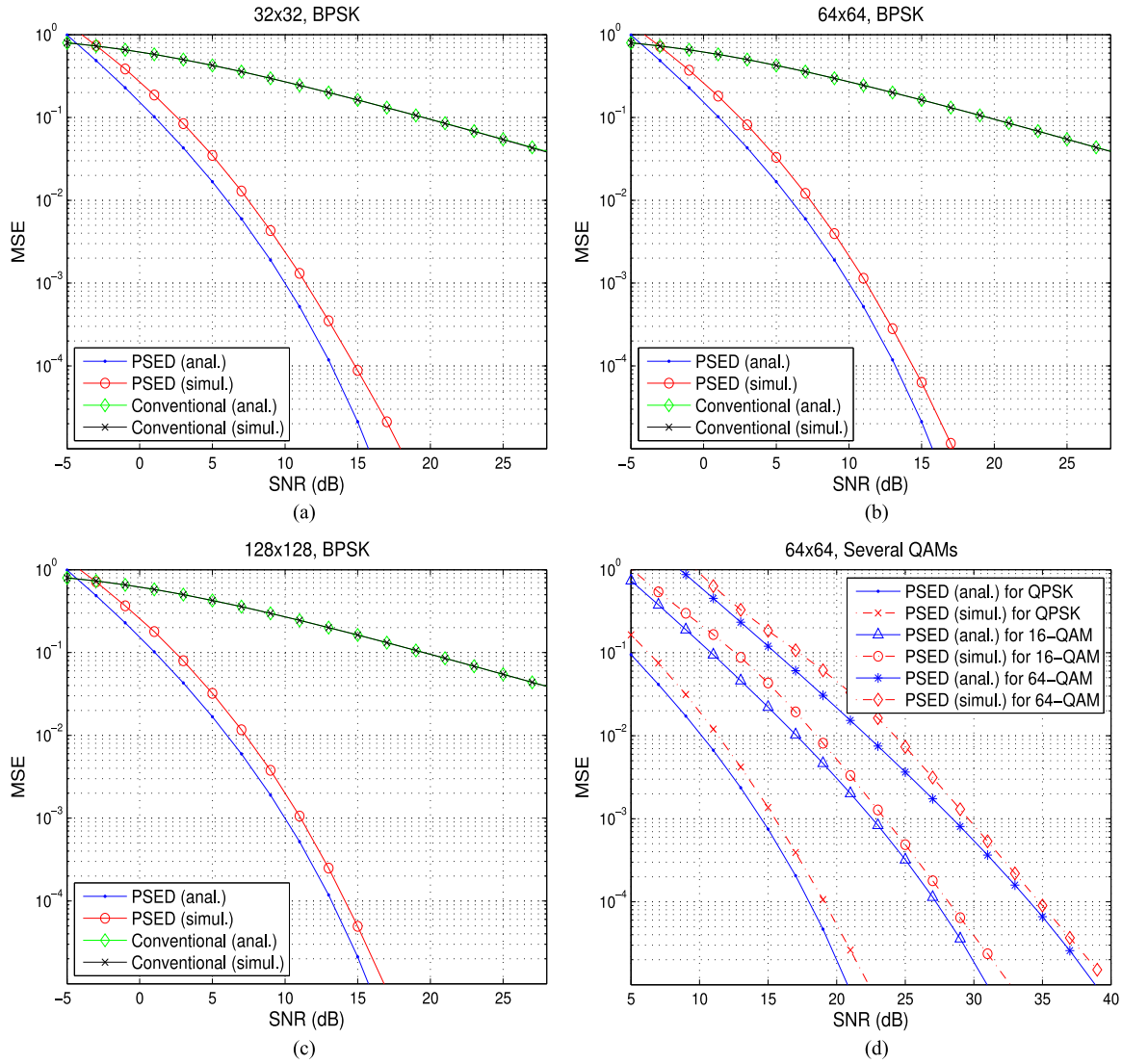


Fig. 5. The MSE performance of the PSED technique and the conventional detector: (a) $n_r = n_t = 32$, (b) $n_r = n_t = 64$, and (c) $n_r = n_t = 128$ when BPSK modulation is used. (d) The MSE performance for $n_r = n_t = 64$ system is shown for several QAM modulations.

modulations with the modulation order $4, 16$ and 64. For all cases considered, we observe that the obtained bound is close to the simulation results. We also observe that the performance difference between the PSED algorithm and the conventional detector is substantial and further the difference increases with SNR. This is mainly because the accuracy of the estimated error vector $\hat{\mathbf{e}}$ for the sparse recovery algorithm improves with SNR so that the magnitude of the error difference $\epsilon = \hat{\mathbf{e}} - \mathbf{e}$ is reduced sharply.

IV. Computational Cost of Sparse Recovery

In this subsection, we analyze the computational complexity of the PSED algorithm. We consider two versions of the PSED algorithm; PSED with LMMSE detection (PSED-LMMSE) and PSED with MF detection (PSED-MF). As mentioned, additional operations caused by the proposed method are 1) sparse transform and 2) sparse error recovery. While the computational overhead of the sparse transform is fixed (matrix multiplication and subtraction), that for the sparse error vector recovery depends on the tree branching parameter L of MMP and the sparsity K. Since the number of iterations of MMP is set to the sparsity K, K matrix inverse operations (from 1×1 to $K \times K$ matrix inverse) are required. Noting that the sparsity K is much smaller than the dimension of symbol vector n_t and the MMP algorithm performs well with small value of L (in our simulations, we set $L = 2$), additional burden of the PSED is very marginal.

Table IV summarizes the number of complex multiplications required for PSED and the conventional detectors. When the LMMSE detector is used, inversion of the covariance matrix is needed in the weight generation process. In Table IV, we denote the number of complex multiplications to perform the inversion of an $n \times n$ matrix by $I_v(n)$. Since the required complexity to invert a matrix is cubic in the dimension n of the matrix (i.e., $O(n^3)$), additional matrix inversion overhead (from 1×1 to $K \times K$ dimensional matrix) of PSED-LMMSE is small

TABLE IV
THE TOTAL NUMBER OF COMPLEX MULTIPLICATIONS OF SEVERAL DETECTORS

Operation	MF detector	PSED-MF	LMMSE detector	PSED-LMMSE
Filter weight generation	0	0	$2n_r n_t^2 + I_v(n_t)$	$2n_r n_t^2 + I_v(n_t)$
Filtering	$n_r n_t$	$n_r n_t$	$n_r n_t$	$n_r n_t$
Sparse transform	0	$n_r n_t$	0	$n_r n_t$
Sparse recovery (matching)	0	$\sum_{k=1}^{K} n_r(n_t-k+1)$	0	$\sum_{k=1}^{K} n_r(n_t-k+1)$
Sparse recovery (orthogonal projection)	0	$\sum_{k=1}^{K}(2n_r k^2 + I_v(k) + kn_r)L$	0	$\sum_{k=1}^{K}(2n_r k^2 + I_v(k) + kn_r)L$
Sparse recovery (residual generation)	0	$\sum_{k=1}^{K} kn_r L$	0	$\sum_{k=1}^{K} kn_r L$
Total	$n_r n_t$	$2n_r n_t + \sum_{k=1}^{K} n_r(n_t - k + 1) + (2n_r k^2 + I_v(k) + 2kn_r)L$	$2n_r n_t^2 + I_v(n_t) + n_r n_t$	$2n_r n_t^2 + I_v(n_t) + 2n_r n_t + \sum_{k=1}^{K} n_r(n_t - k + 1) + (2n_r k^2 + I_v(k) + 2kn_r)L$

TABLE V
COMPARISON OF COMPUTATIONAL COMPLEXITY IN TERMS OF NUMBER OF COMPLEX MULTIPLICATIONS OF DETECTION SCHEMES (NUMBERS IS IN THOUSANDS)

	32×32 ($\times 10^3$)	64×64 ($\times 10^3$)	128×128 ($\times 10^3$)	256×256 ($\times 10^3$)
MF	1	4	16	66
PSED-MF	12	135	1759	23439
LMMSE	78	618	4918	39245
PSED-LMMSE	88	744	6644	62553
K-best	88 (m=15)	778 (m=40)	6666 (m=95)	62305 (m=260)
	128×16 ($\times 10^3$)	128×32 ($\times 10^3$)	64×16 ($\times 10^3$)	64×32 ($\times 10^3$)
MF	2	4	1	2
PSED-MF	13	47	6	23
LMMSE	69	277	35	144
PSED-LMMSE	78	316	39	163
K-best	77 (m=15)	317 (m=15)	40 (m=7)	163 (m=6)

when compared to the $n_t \times n_t$ dimensional covariance matrix inversion of LMMSE.[9] For example, additional complexity of PSED-LMMSE for $n_t = n_r = 32$ is only 13% and that for $n_t = n_r = 64$ is 20%. When the MF is used (PSED-MF), there is no inversion and thus the additional complexity associated with PSED might be relatively higher than that of PSED-LMMSE. In spite of the increased computational overhead, due to the low-complexity operations, overall complexity of the PSED-MF is much smaller than the computational burden of the LMMSE detector.

V. SIMULATIONS AND DISCUSSIONS

A. Simulation Setup

In this section, we examine the performance of the proposed PSED algorithm (PSED-LMMSE and PSED-MF). For comparison, we test conventional linear receivers (MF and LMMSE detectors), K-best detection algorithm [25], and IO-LAMA [16]. Also, as a performance lower bound, we consider the ML detector implemented via sphere decoding (SD) algorithm. Note that the K-best detector is a sub-optimal MIMO detector and its complexity does not vary with channel realizations and SNR. It has been shown that IO-LAMA asymptotically achieves the performance of individual parallel detection over random i.i.d. Gaussian channels [16] Note also that we only present the result of the ML detector for 32×32 dimensional system since it is very hard to obtain the results when the dimension is higher than this.

In order to check the worst-case performance, we always turn on the sparse recovery algorithm of the PSED even in the low SNR regime where the error signal would not be sparse. As mentioned, we use MMP in recovering the sparse error vector.

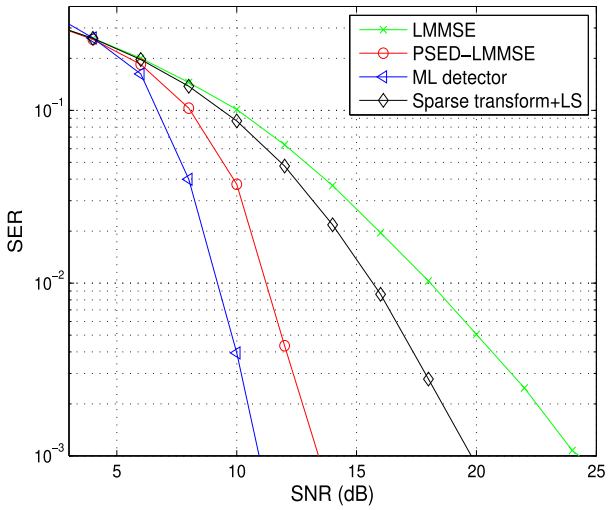


Fig. 6. The SER as a function of SNR for $(n_r, n_t) = (32, 32)$ system.

[9]Note that gaussian elimination method for $n \times n$ inversion requires $(2n^3 + 3n^2 - 5n)/6$ multiplications [40]. See [41] for more efficient implementation of matrix inversion.

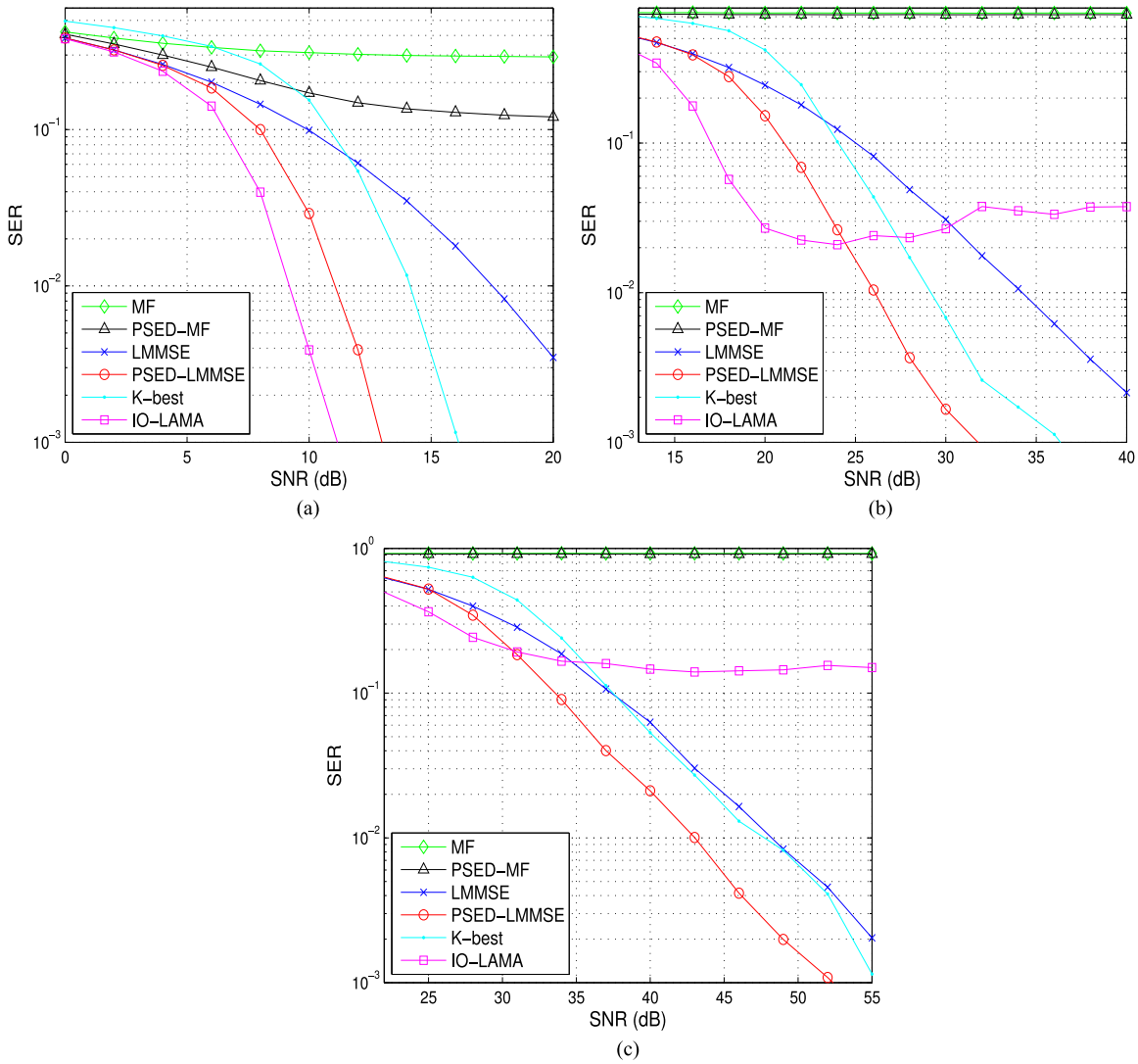


Fig. 7. The SER as a function of SNR for (a) QPSK, (b) 16-QAM, and (c) 64-QAM modulations. The uncoded system with the size $(n_r, n_t) = (64, 64)$ is considered.

The branching parameter L is fixed to 2. Ideally, the sparsity parameter K should be equal to the number of errors of the conventional detector used in the first stage. Since the number of errors is unknown, we compute analytic symbol error rate P_e (see (32)) and set $K = P_e n_t$ in our simulations. In the K-best detector, we set the number of the symbols survived for each layer such that overall complexity is comparable to that of the PSED-LMMSE (see the complexity comparison in Table V). The SNR per each receive antenna is defined as

$$\mathrm{SNR}_r = \frac{n_t}{n_r}\mathrm{SNR}. \tag{38}$$

B. Simulation Results

1) Performance for i.i.d. Gaussian System Matrix: First, we evaluate the performance of the proposed detector for the system matrix $\mathbf{H}$ whose entries are chosen from i.i.d. complex Gaussian (i.e., $h_{i,j} \sim \mathcal{CN}(0, \frac{1}{n_r})$). We assume that the channel information is perfectly known in the receiver.

In order to demonstrate the efficacy of the sparse recovery algorithm, we compare the symbol error rate (SER) performance of several detectors for $(n_r, n_t) = (32, 32)$ uncoded system. In Fig. 6, we observe that the proposed PSED-LMMSE method offers substantial performance gain over the conventional LMMSE detector. "Sparse transform + LS" refers to the method employing sparse transform followed by LS estimation of error vector $\mathbf{e}$. Since this scheme does not rely on the sparse recovery and thus cannot suppress the noise effectively, performance of this approach is much worse than that of the PSED-LMMSE. Note that the performance of PSED-LMMSE is within 2 dB from that of the optimal ML detector.

In Fig. 7, we plot the SER performance of the PSED algorithm and the conventional MIMO detectors for the uncoded system with the size $(n_r, n_t) = (64, 64)$. In Fig. 7(a), (b), and (c), we provide the SER performance when QPSK, 16-QAM, 64-QAM modulations are used. We observe that the proposed PSED-LMMSE outperforms the conventional LMMSE detector by a large margin for all cases under consideration. Note

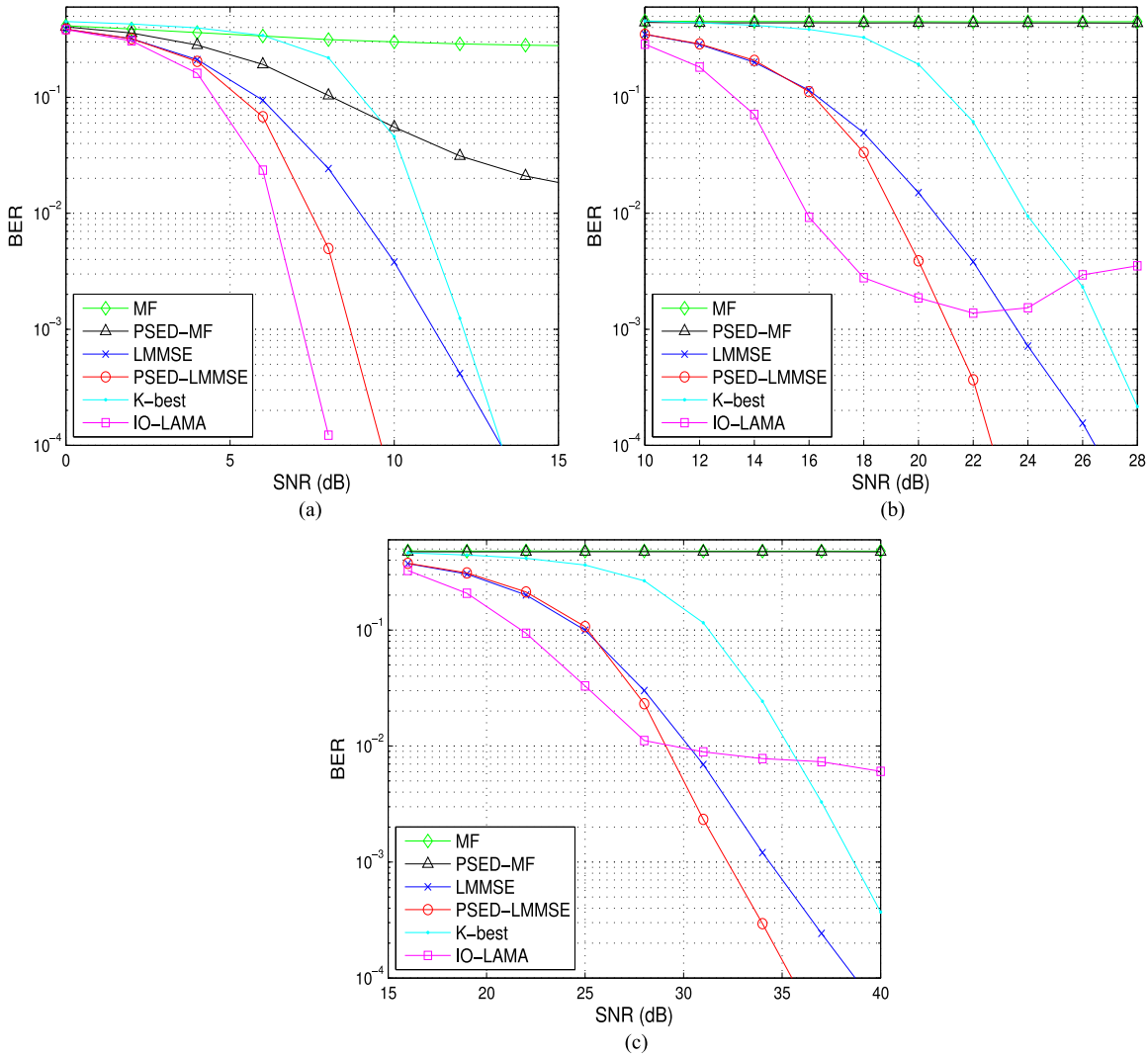


Fig. 8. The SER as a function of SNR for (a) QPSK, (b) 16-QAM, and (c) 64-QAM modulations. The coded system with the size $(n_r, n_t) = (64, 64)$ is considered.

that the performance gain of the PSED-LMMSE is pronounced in high SNR regime. This is because 1) stronger sparsity, i.e., lower K leads to better recovery performance and thus 2) the gain obtained from the post sparse error detection/correction process increases sharply in this regime. Specifically, in the low SNR regime, number of errors in a symbol vector is relatively large (that is, sparsity of the input error vector is high) so that the sparse recovery algorithm would not work properly and the gain obtained from the PSED technique would be marginal. On the other hand, in the high SNR regime, input error vector of the the sparse recovery algorithm is sparse and thus the PSED would return highly reliable error vector estimate, resulting in the substantial performance gain. Note also that in contrast to PSED-LMMSE, the performance of PSED-MF is not so appealing when n_t is equal to n_r. This is because the MF detector does not perform well for the SNR range of interest so that the input error vector after the sparse transformation is not sparse, resulting in poor reconstruction performance of the sparse recovery algorithm. We observe that IO-LAMA outperforms the proposed PSED detector when QPSK modulation is used, but it suffers from error floor phenomenon in high order modulation scenarios (i.e., 16QAM and 64QAM). Note that the proposed scheme outperforms the K-best detector for all cases considered. In Table V, we provide the computational complexity (i.e., the number of complex multiplications) of detection schemes under consideration. When compared to the complexity of the conventional LMMSE detector, we see that the computational complexity required to perform the post processing in the PSED detectors is quite marginal. Note also that in spite of significant performance gain, the complexity of PSED-LMMSE is comparable to that of K-best detector.

In Fig. 8, we investigate the bit error rate (BER) performance of PSED-LMMSE when the forward error correction code (FEC) is employed. We use the convolutional channel code with generator polynomial $(171, 131)$ and the Viterbi decoder to perform channel decoding [46]. The size of code block is set

TABLE VI
COMPLEXITY AND PERFORMANCE OF LMMSE AND PSED-LMMSE DETECTORS FOR DIFFERENT SYSTEM SIZES

Modulation	Detector	Metric	32×32	64×64	128×128	256×256
QPSK	LMMSE	Compl. ($\times 10^3$)	78	618	4918	39245
		Required SNR (dB)	18.1	17.5	17.2	17.1
	PSED-LMMSE	Compl. ($\times 10^3$)	88	744	6644	62553
		Required SNR (dB)	11.8	11.3	11.2	11.2
16QAM	LMMSE	Compl. ($\times 10^3$)	78	618	4918	39245
		Required SNR (dB)	35.3	34	32.8	31.0
	PSED-LMMSE	Compl. ($\times 10^3$)	88	744	6644	62553
		Required SNR (dB)	28.2	25.3	23.3	22.2

to 1536 bits and a random interleaver is used between the channel encoder and the symbol modulator. For fair comparison, we used the hard-sliced symbol in the channel decoding. Overall, we observe that the performance gain of the PSED-LMMSE over the conventional LMMSE detection scheme is well maintained for all modulation orders considered. We also see that the K-best detector performs worse than the LMMSE detector for the SNR range of interest. IO-LAMA detector works well in QPSK systems but it exhibits error floor phenomenon for high order modulation scenarios.

Finally, we compare the performance and complexity of the LMMSE detector and PSED-LMMSE detector for various system dimensions. In Table VI, we provide the number of complex multiplications and the required SNR at 10^{-2} SER for $(n_r, n_t) = (32, 32), (64, 64), (128, 128)$, and $(256, 256)$. Note that for all dimensions we considered, the proposed PSED-LMMSE maintains large performance gain over the LMMSE. When the system dimension increases, computational complexity of the PSED-LMMSE also increases due to the large number computations needed for the sparse recovery algorithm.

2) Performance Evaluation for Massive MIMO Scenarios: We finally investigate the performance of the PSED in realistic massive MIMO scenario. We consider the uplink scenario where n_t users with single antenna transmit the information to the basestation with n_r transmit antennas. The channel vector between the kth user to the basestation is given by [47]

$$\mathbf{h}_k = \frac{1}{\sqrt{P}} \sum_{p=1}^{P} g_{kp} \mathbf{a}(\theta_p) \tag{39}$$

where g_{kp} is the channel gain for the kth user, P is the total number of multi-paths, θ_p is the angle of arrival for the pth multi-path signal, and $\mathbf{a}(\theta_p)$ is the steering vector expressed as

$$\mathbf{a}(\theta_p) = \left[1 \; e^{-j2\pi\gamma \sin(\theta_p)} \; \cdots \; e^{-j(n_r-1)2\pi\gamma \sin(\theta_p)}\right]. \tag{40}$$

Note that γ is the ratio of the antenna spacing to the wavelength. The corresponding channel matrix $\mathbf{H}$ is given by

$$\mathbf{H} = \frac{1}{\sqrt{P}} \mathbf{A}\mathbf{G}. \tag{41}$$

We assume that the signal powers received from each user are equal and hence the entries of the matrix G are i.i.d. Gaussian with zero mean and unit variance. In our simulations, we set $\lambda = 0.3$ and $\theta_p = -\pi/2 + (p-1)\pi/P$ where $p = 1, 2, \ldots, P$ where P is the number of the multi-paths (P is set to 32) [47].

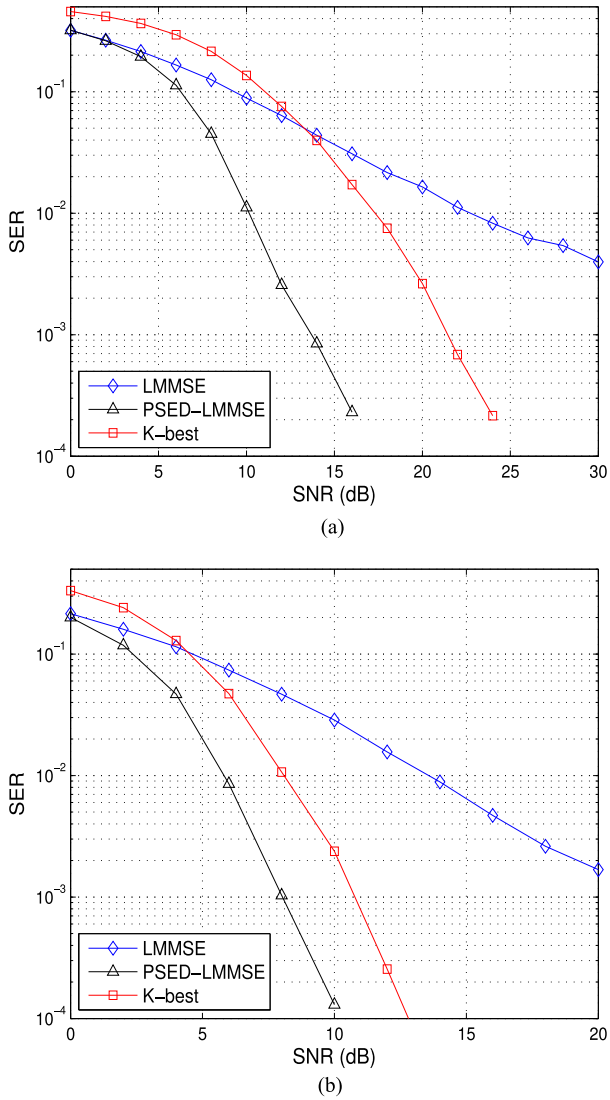


Fig. 9. The SER vs. SNR plot of the massive MIMO detectors for a) $(n_r, n_t) = (64, 32)$ and b) $(n_r, n_t) = (128, 32)$.

Fig. 9 provides the performance of the proposed detector for $n_t = 32$ (i.e., 32 mobile users) and (a) $n_r = 64$ and (b) $n_r = 128$ basestation antennas, respectively. We observe that the proposed PSED-LMMSE detector achieves significant performance gain over the K-best detector as well as the conventional LMMSE detector. We also observe that as the number of basestation

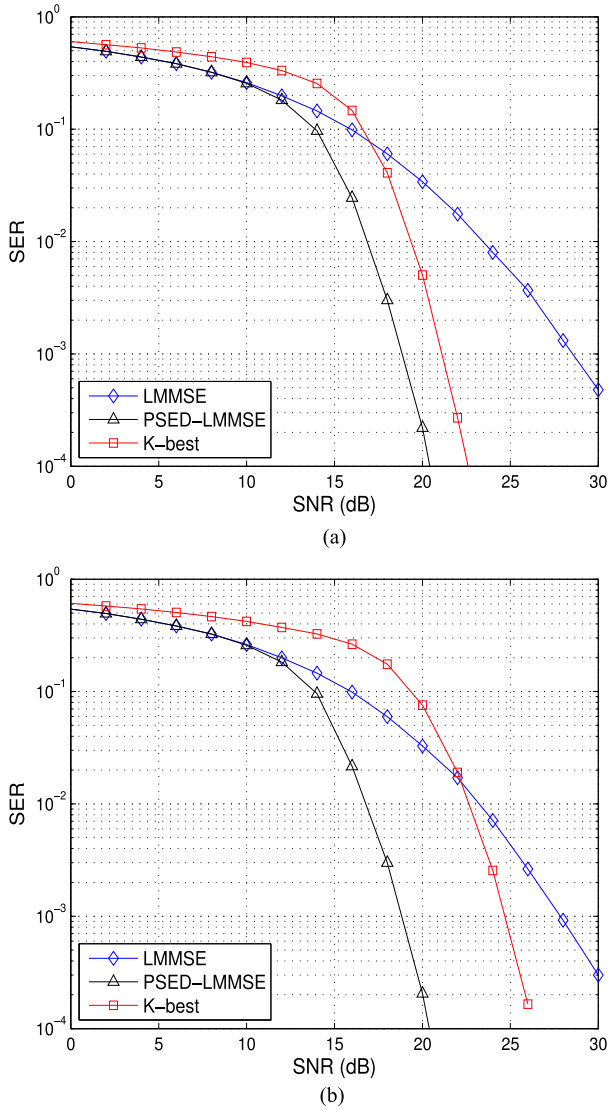


Fig. 10. The SER vs. SNR plot of the IoT detectors for a) $(n_r, n_t) = (128, 128)$ and b) $(n_r, n_t) = (256, 256)$.

antennas increases, the performance gain of the proposed method over the conventional detectors improves significantly.

3) Performance Evaluation for IoT Multiuser Detection Scenarios: We evaluate the performance of the proposed detector for the IoT scenario where n_t devices access the AP and the receiver algorithm in the AP detects the data symbols from all users simultaneously. Each device is encoded with a pseudo noise (PN) based signature sequence with length n_r and we assume that the signal transmission of all users is synchronous.

Fig. 10 shows the detection performance of the proposed detector as a function of SNR. Two configurations $(n_r, n_t) = (128, 128)$ and $(256, 256)$ are considered in the simulations. Note that we do not include the results of the MF and PSED-MF detectors since these schemes perform poor. We show that the proposed PSED-LMMSE achieves significant performance gain over the existing detectors, resulting in 2.3 dB and 5.5 dB gain at 10^{-2} SER.

VI. Conclusion

In this paper, we proposed a novel detection algorithm exploiting sparsity of error vector in the conventional linear detection in improving the detection quality of symbol vectors in large-scale wireless systems. Our approach operates in two steps. In the first step, we transform the conventional communication system into the system whose input is the sparse error vector, which is accomplished by the conventional linear detection followed by the symbol quantization. For the transformed measurement vector, we next apply the sparse error recovery algorithm followed by the error cancellation to obtain the refined estimate of the transmit symbol vector. Our approach is simple to implement with comparably small computational cost, yet offers substantial gain over the conventional detection schemes. We observed from the asymptotic performance analysis and empirical simulations that the proposed PSED algorithm achieves significant performance gain over the conventional detection schemes.

Appendix A
Proof that the Output Streams of the Linear MMSE Detector are Asymptotically Uncorrelated

The ith output stream of the linear MMSE detector is given by $\hat{s}_i = \mathbf{h}_i^H \left(\mathbf{H}\mathbf{H}^H + \frac{1}{\mathrm{SNR}}\mathbf{I}\right)^{-1}$. The correlation between the ith and jth output streams is given by

$$E[\hat{s}_i \hat{s}_j^*] = P\mathbf{h}_i^H \left(\mathbf{H}\mathbf{H}^H + \frac{1}{\mathrm{SNR}}\mathbf{I}\right)^{-1} \mathbf{h}_j, \tag{42}$$

where $i \neq j$. If we use a matrix inversion lemma $\mathbf{x}^H(\mathbf{A} + \tau\mathbf{x}\mathbf{x}^H)^{-1} = \frac{\mathbf{x}^H\mathbf{A}^{-1}}{1+\tau\mathbf{x}^H\mathbf{A}^{-1}\mathbf{x}}$, we can show that

$$E[\hat{s}_i \hat{s}_j^*] = \frac{\mathbf{h}_i^H \left(\mathbf{H}_{[i,j]}\mathbf{H}_{[i,j]}^H + \frac{1}{\mathrm{SNR}}\mathbf{I}\right)^{-1} \mathbf{h}_j}{\left(1 + \mathbf{h}_i^H \left(\mathbf{H}_{[i]}\mathbf{H}_{[i]}^H + \frac{1}{\mathrm{SNR}}\mathbf{I}\right)^{-1} \mathbf{h}_i\right)} \cdot \frac{1}{\left(1 + \mathbf{h}_j^H \left(\mathbf{H}_{[j]}\mathbf{H}_{[j]}^H + \frac{1}{\mathrm{SNR}}\mathbf{I}\right)^{-1} \mathbf{h}_j\right)}, \tag{43}$$

where $\mathbf{H}_{[\mathcal{A}]}$ is the submatrix of $\mathbf{H}$ with the columns specified by the index set $\mathcal{A}$ are removed. According to Lemma 4 in [43], when $n_t, n_r \to \infty$, the numerator in (43) converges to zero and the denominator converges to one.

Appendix B
Proof of (23) and (24)

As mentioned, the distribution of $|E|$ is approximated by $\mathcal{N}(n_t P_e, n_t P_e(1-P_e))$. Thus, the quantity $\frac{|E|}{n_t}$ follows $\mathcal{N}(P_e, \frac{P_e(1-P_e)}{n_t})$ and

$$Pr\left(\left|\frac{|E|}{n_t} - P_e\right| > \epsilon\right) = 2Q\left(\frac{\epsilon}{\sqrt{\frac{P_e(1-P_e)}{n_t}}}\right) \tag{44}$$

$$< 2\exp\left(-\frac{\epsilon^2 n_t^2}{2P_e(1-P_e)}\right), \tag{45}$$

for any positive ϵ. When $n_t \to \infty$, (45) goes to zero and thus $\frac{|E|}{n_r} (= \frac{\beta |E|}{n_t})$ also converges to $P_e \beta$.

ACKNOWLEDGMENT

The authors would like to thank the associate editor (Prof. A. Wiesel) and anonymous reviewers for their valuable comments and suggestions that improved the quality of the paper.

REFERENCES

[1] J. W. Choi and B. Shim, "New approach for massive MIMO detection using sparse error recovery," in *Proc. Global Telecommun. Conf.*, 2014, pp. 3754–3759.

[2] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[3] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[4] J. Choi, B. Shim, Y. Ding, B. Rao, and D. Kim, "Compressed sensing for wireless communcations: Useful tips and tricks," in *Proc. IEEE Comm. Survey Tuts.*, vol. 19, no. 3, pp. 1527–1550, 3rd Quart. 2017.

[5] W. Li and J. C. Preisig, "Estimation of rapidly time-varying sparse channels," *IEEE J. Ocean. Eng.*, vol. 32, no. 4, pp. 927–939, Oct. 2007.

[6] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1708–1721, Mar. 2010.

[7] C. Wen, S. Jin, K. Wong, J. Chen, and P. Ting, "Channel estimation for massive MIMO using Gaussian-mixture Bayesian learning," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1356–1368, Mar. 2015.

[8] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.

[9] B. Shim and B. Song, "Multiuser detection via compressive sensing," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 972–974, Jul. 2012.

[10] H. F. Schepker and A. Dekorsy, "Compressive sensing multi-user detection with block-wise orthogonal least squares," in *Proc. IEEE 75th Veh. Technol. Conf.*, 2012, pp. 1–5.

[11] S. Park, H. Seo, H. Ji, and B. Shim, "Joint active user detection and channel estimation for massive machine-type communications," in *Proc. IEEE Signal Process. Adv. Wireless Commun Conf.*, Jul. 2017, pp. 573–576.

[12] B. Shim, S. Kwon, and B. Song, "Sparse detection with integer constraint using multipath matching pursuit," *IEEE Commun. Lett.*, vol. 18, no. 10, pp. 1851–1854, Oct. 2014.

[13] L. G. Barbero and J. S. Thompson, "Fixing the complexity of the sphere decoder for MIMO detection," *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2131–2142, Jun. 2008.

[14] S. Wu, L. Kuang, Z. Ni, J. Lu, D. D. Huang, and Q. Guo, "Low-complexity iterative detection for large-scale multiuser MIMO-OFDM systems using approximate message passing," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 902–915, Oct. 2014.

[15] C. Wen, C. Wang, S. Jin, K. Wong, and P. Ting, "Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs," *IEEE Trans. Signal Process.* vol. 64, no. 10, pp. 2541–2556, May 2016.

[16] C. Jeon, R. Ghods, and C. Studer, "Optimality of large MIMO detection via approximate message passing," in *Proc. Int. Symp. Inf. Theory*, 2015, pp. 1227–1231.

[17] S. Wang, Y. Li, and J. Wang, "Multiuser detection in massive spatial modulation MIMO with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2156–2168, Apr. 2015.

[18] J. Barbier, F. Krzakala, L. Zdeborova, and P. Zhang, "Robust error correction for real-valued signals via message-passing decoding and spatial coupling," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2013, pp. 1–5.

[19] S. Verdu, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[20] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Math. Comput.*, vol. 44, pp. 463–471, Apr. 1985.

[21] J. Jalden and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1474–1484, Apr. 2005.

[22] B. Shim, J. Choi, and I. Kang, "Towards the performance of ML and the complexity of MMSE—A hybrid approach for multiuser detection," *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2508–2519, Jul. 2012.

[23] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[24] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conf.*, Nov. 1993, pp. 40–44.

[25] Z. Guo and P. Nilsson, "Algorithm and implementation of the K-best sphere decoding for MIMO detection," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 491–503, Mar. 2006.

[26] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1995.

[27] S. Kwon, J. Wang, and B. Shim, "Multipath matching pursuit," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2986–3001, May 2014.

[28] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[29] A. M. Tulino and S. Verdu, *Random Matrix Theory and Wireless Communications*. Breda, The Netherlands: Now Publ., 2004.

[30] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*. New York, NY, USA: McGraw-Hills, 2002.

[31] J. Zhang, E. K. P. Chong, and D. N. C. Tse, "Output MAI distribution of linear MMSE multiuser receivers in DS-CDMA systems," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1128–1144, Mar. 2001.

[32] D. Guo, S. Verdu, and L. K. Rasmussen, "Asymptotic normality of linear multiuser receiver outputs," *IEEE Trans. Inf. Theory*, vol. 48, no. 12, pp. 3080–3095, Dec. 2002.

[33] J. G. Proakis, *Digital Communications*, 2nd ed. New York, NY, USA: McGraw-Hill, 1989.

[34] R. Tibshirani, "Regression shrinkage and selection via Lasso," *J. Roy. Statist. Soc B.*, vol. 58, pp. 267–288, 1996.

[35] S. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale ℓ_1-regularized least squares," *IEEE J. Sel. Topics Signal Process.*, vol 1, no. 4, pp. 606–617, Dec. 2007.

[36] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, pp. 301–321, 2009.

[37] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[38] J. Wang, S. Kown, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6202–6216, Dec. 2012.

[39] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Process*. 4th ed. New York, NY, USA: McGraw-Hill, 2002.

[40] R. W. Farebrother, *Linear Least Squares Computations*. New York, NY, USA: Marcel Dekker, 1998.

[41] R. P. Brent and P. Zimmermann, *Modern Computer Arithmetic*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[42] R. Ward, "Compressed sensing with cross validation," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5773–5782, Dec. 2009.

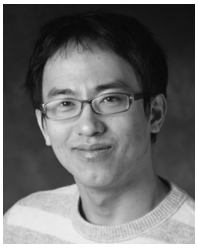
[43] J. Hoydis, S. T. Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[44] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[45] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Construction Approx.*, vol. 28, no. 3, pp. 253–263, 2008.

[46] A. J. Viterbi, "Error bounds for convolutional codes and asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 2, pp. 260–269, Apr. 1967.

[47] H. Q. Ngo, T. L. Marzetta, and E. G. Larsson, "Analysis of the pilot contamination effect in very large multicell multiuser MIMO systems for physical channel models," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, May 2011, pp. 3464–3467.



Jun Won Choi (M'04) received the B.S. and M.S. degrees in electrical and computer engineering from Seoul National University, Seoul, South Korea, and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, respectively. In 2010, he joined Qualcomm, San Diego, CA, USA and participated in wireless communication systems/algorithm design for commercializing LTE and LTE-A modem chipsets. Since 2013, he has been a faculty member in the Department of Electrical Engineering, Hanyang University, Seoul, South Korea, and leading the Signal Processing and Machine Learning Research Group. His research interests include sparsity-based signal processing, wireless communications, machine learning, and intelligent transportation systems.

Byonghyo Shim (S'95–M'97–SM'09) received the B.S. and M.S. degrees in control and instrumentation engineering from Seoul National University, Seoul, South Korea, in 1995 and 1997, respectively. He received the M.S. degree in mathematics and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 2004 and 2005, respectively.

From 1997 and 2000, he was with the Department of Electronics Engineering, Korean Air Force Academy, as an Officer (First Lieutenant) and an Academic Full-time Instructor. From 2005 to 2007, he was with Qualcomm Inc., San Diego, CA, USA, as a Staff Engineer. From 2007 to 2014, he was with the School of Information and Communication, Korea University, Seoul, South Korea, as an Associate Professor. Since September 2014, he has been with the Seoul National University, where he is currently a Professor in the Department of Electrical and Computer Engineering. His research interests include wireless communications, statistical signal processing, estimation and detection, compressed sensing, and information theory. He received the M. E. Van Valkenburg Research Award from the Electrical and Computer Engineering Department of the University of Illinois (2005), Hadong Young Engineer Award from the IEIE (2010), and Irwin Jacobs Award from Qualcomm and KICS (2016). He is an elected member of Signal Processing for Communications and Networking Technical Committee of the IEEE Signal Processing Society. He has been an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE WIRELESS COMMUNICATIONS LETTERS, *Journal of Communications and Networks*, and a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.