Orthogonal Symmetric Toeplitz Matrices for Compressed Sensing: Statistical Isometry Property

Kezhi Li, *Member, IEEE,* Lu Gan, *Member, IEEE,* and Cong Ling, *Member, IEEE*

***Abstract*—Recently, the statistical restricted isometry property (RIP) has been formulated to analyze the performance of deterministic sampling matrices for compressed sensing. In this paper, we propose the usage of orthogonal symmetric Toeplitz matrices (OSTM) for compressed sensing and study their statistical RIP by taking advantage of Stein's method. In particular, we derive the statistical RIP performance bound in terms of the largest value of the sampling matrix and the sparsity level of the input signal. Based on such connections, we show that OSTM can satisfy the statistical RIP for an overwhelming majority of signals with given sparsity level, if a Golay sequence used to generate the OSTM. Such sensing matrices are deterministic, Toeplitz, and efficient to implement. Simulation results show that OSTM can offer reconstruction performance similar to that of random matrices.**

***Index Terms*—Compressed sensing, restricted isometry property, sensing matrices, sequences, signal sampling, Toeplitz matrices.**

I. INTRODUCTION

COMPRESSED sensing (CS) [1]–[4] is a novel theory which has drawn much attention since its advent several years ago. The CS theory is based on the assumption that a signal is compressible or sparse. Consider a discrete-time, length-N signal x that can be represented (or approximated) by only K ($K \ll N$) coefficients. CS is accomplished by computing a measurement vector y through the following linear transformation [1], [2]:

$$y = \Phi x, \tag{1}$$

where y represents an $M \times 1$ sampled vector and Φ is an $M \times N$ measurement matrix. It was proved in [1], [2] that under certain conditions, x can be well recovered from only $M = \mathcal{O}(K \log(N/K))$ measurements through non-linear optimization.

Despite the fast development, there still exists a wide gap between the theory and practice of CS. The first family of sensing matrices for l_1-based reconstruction consists of random Gaussian/Bernoulli matrices. However, huge memory buffering for storage and high computational complexity due to unstructured nature restrict its applications. As a result, some papers proposed structurally random matrices, such as Hadamard or Toeplitz matrices [5], [6]. The restrict isometry property (RIP) guarantees the recovery of the sparsest solution from the compressive measurements utilizing certain efficient and robust algorithms such as l_1 minimization, matching pursuit and its variants [4]. In practice, deterministic sensing matrices are highly desirable. Deterministic matrices with the RIP have been proposed in [7]–[9]. However, their guaranteed RIP performance is not comparable to that of random matrices.

Recently, a statistical version of the RIP has been developed by several authors [10]–[12], where a deterministic sensing matrix Φ is required to preserve the l_2 norm within a small fraction, with respect to for *any* K-sparse input vector x whose support is taken uniformly at random. In particular, Tropp derived explicit bounds for the extreme singular values of random collections of columns from a general dictionary [10]. Though no specific sensing matrices were proposed, it implied dictionaries with small coherence μ can be used as sensing matrices with good statistical RIP. Meanwhile, Calderbank *et al.* derived the performance bound of statistical RIP for a large class of deterministic matrices [12]. An essential property of their sensing matrices is that the columns from a group under pointwise multiplication, such as those constructed from the extended BCH codes and discrete chirps. Their sensing matrices permit low-complexity reconstruction, and the required conditions on sensing matrices are easily checkable.

In this paper, we propose a new class of deterministic sensing matrices based on orthogonal symmetric Toeplitz matrices (OSTM) [13], and investigate their statistical RIP by exploiting Stein's method. Stein's method of exchangeable pairs is a powerful tool for concentration inequalities, especially for those involving random permutations [14]. We hope this will inspire more research on the application of this powerful machinery to CS. Furthermore, we show that the statistical RIP of an OSTM will be near-optimal if it is generated from of a Golay sequence [15]. The proposed deterministic sensing matrices offer a few advantages: they are easy to generate as only N numbers need to be stored; both sampling and reconstruction are more efficient to implement since the Toeplitz structure permits a fast Fourier transform (FFT)-based implementation [16]; and they are well suited to some applications that are inherently Toeplitz, such as system identification, channel estimation [17] and Terahertz imaging [18], [19]. Empirical results show that these deterministic matrices can provide reconstruction performance similar to that of the random matrices.

This paper was presented in part at the IEEE Information Theory Workshop (ITW), Taormina, Italy, Oct. 2009, and has been partially submitted to the International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2011.

K. Li and C. Ling are with the Department of Electrical and Electronic Engineering, Imperial College London, London, SW7 2AZ, UK (e-mail: k.li08@imperial.ac.uk, cling@ieee.org).

L. Gan is with the School of Engineering and Design, Brunel University, London UB8 3PH, UK (e-mail: lu.gan@brunel.ac.uk).

A. Relation to Prior Work

In the conference version of this paper [13], we derived a weaker bound for the statistical RIP of OSTM using Chebyshev's inequality. However, the bound was unable to capture the impact of the sign sequence. In this paper, we make a significant improvement by deriving an exponential bound using Stein's method, which enables us to tell the superiority of Golay sequences as the underlying sign sequences.

The proposed OSTM might be viewed as a derandomized version of Romberg's random convolution [20]. More precisely, the sensing matrices in [20] are generated from randomly sampled rows of the matrix $\Phi_N = N^{-1/2}F_N^*\Sigma F_N$ where F_N is the $N \times N$ discrete Fourier matrix, and Σ is a diagonal matrix whose entries, roughly speaking, are random phases uniformly distributed on $[0, 2\pi]$. The analysis was based on the mutual coherence μ. In our deterministic sensing matrices, we will fix the diagonal of Σ (obtained from a Golay sequence) to optimize the isometry property of sensing matrices.

We notice that while random permutation is a technique of proof in statistical RIP, it is an inherent feature of the structurally random sensing matrices proposed in [5], [21], where the signals are scrambled by uniform permutation before sampling.

Random Toeplitz matrices with entries drawn independently from the same distribution have been shown to good for CS [6], [16]. For instance, the entries of the Toeplitz matrix in [6] are drawn from the Bernoulli distribution. It is shown that there exist constants $c_1, c_2 > 0$ depending only on δ_{3K} such that for any $M \geq c_1 K^3 \ln(N/K)$, the sensing matrix Φ satisfies RIP of order $3K$ for every $\delta_{3K} \in (0, 1/3)$ with probability at least

$$1 - \exp\left(-c_2 \frac{M}{K^2}\right). \tag{2}$$

It is seen that the failure probability is relatively high because of the square of K in denominator.

B. Organization

The rest of this paper is organized as follows. In Section II, we briefly review the concepts of RIP and statistical RIP. The construction of OSTM is given in Section III. Then in Section IV, we derive an exponential bound for the statistical RIP for the proposed deterministic sensing matrices, by using Stein's method. Section V is devoted to the usage of Golay's sequences to achieve near-optimal performance for OSTM. Simulation results are given in Section VI, followed by conclusions in Section VII.

II. RIP AND STATISTICAL RIP

It was established in [22] that for a matrix Φ to be a CS sensing matrix, it is sufficient that it satisfies the RIP, which makes sensing matrix act as a near isometry on all K-sparse vectors.

Definition 1 (RIP [22]): Let Ω denote the set of all length-N vectors x with K non-zero coefficients. An $M \times N$ measurement matrix Φ has the restricted isometry property (RIP) with parameters (K, δ) for $\delta \in (0, 1)$ if it satisfies [22]

$$(1-\delta)\|x\|^2 \leq \|\Phi x\|^2 \leq (1+\delta)\|x\|^2, \text{ for all } x \in \Omega.$$

Note that the RIP implies that for *all* $N \times K$ sub-matrices of Φ, the eigenvalues of their Gram matrices lie in the interval of $[1-\delta, 1+\delta]$. This is a very restrictive condition and the currently known measurement matrices satisfying the RIP with (near) optimal number of measurements fall into two categories [23]: (*i*) Random matrices with i.i.d. sub-Gaussian variables, e.g., normalized i.i.d. Gaussian or Bernoulli matrices; (*ii*) Random partial bounded orthogonal matrices in which the sensing operators are obtained by choosing M rows uniformly at random from a normalized $N \times N$ Fourier or Walsh-Hadamard transform matrices.

In some applications, deterministic sensing matrices are highly desirable. However, the construction of deterministic RIP matrices is a challenging task. As an alternative, statistical versions of the RIP were proposed for deterministic sensing matrices [11], [12]. In this paper, we follow the statistical RIP formulation given by Calderbank *et al.* in [12].

Definition 2 (Statistical RIP): Let Φ be a normalized $M \times N$ deterministic matrix and denote ϕ_i $(1 \leq i \leq N)$ as its i-th column. The input signal x is a K-sparse random vector with non-zero coefficients $x_1,\ x_2, \cdots, x_K$ whose positions are chosen uniformly at random. Under such a model, the measurement vector can be expressed as

$$y(\pi) = \Phi x = \sum_{i=1}^{N} \phi_{\pi(i)} x_i, \tag{3}$$

where π is drawn from the uniform distribution over the set of all permutations of $\{1, \cdots, N\}$. Φ is said to have the statistical RIP provided that the following inequality holds for any K-sparse signals with high probability:

$$\left|\|y(\pi)\|^2 - \|x\|^2\right| \leq \delta\|x\|^2 \tag{4}$$

with respect to uniform permutation π.

The statistical RIP defined above is a weaker condition than the RIP. The statistical RIP has been analyzed in [12] for a large class of deterministic matrices, as summarized by the following theorem:

Theorem 1 ([12]): Let Φ be a deterministic $M \times N$ sensing matrix satisfying the following properties:

- The columns of Φ form a group under point-wise multiplication;
- The rows of Φ are orthogonal and all row sums are equal to zero, i.e., $\sum_{i=1}^{N} \phi_i = 0$;
- For all columns apart from the first one, the square of the column's l_2 norm is less than $N^{2-\eta}$, where $0 < \eta \leq 1$.

Let x be a K-sparse signal where the positions of the K non-zero entries are equiprobable. Then, for $\frac{K-1}{N-1} < \delta < 1$ and $\eta > 1/2$, the following inequality holds:

$$\begin{aligned} &\mathbf{P}\left(\left|\|\Phi x\|^2 - \|x\|^2\right| < \delta\|x\|^2\right) \\ &\quad \geq 1 - 2\exp\left[-\frac{[\delta - (K-1)/(N-1)]^2 M^\eta}{8K}\right]. \end{aligned} \tag{5}$$

The proof of the exponential bound was based on McDiarmid's inequality [12]. Note that (5) implies that

$$\mathbf{P}\left(\left|\|\Phi x\|^2 - \|x\|^2\right| > \delta\|x\|^2\right) < \mathcal{O}\left(\exp\left[-\frac{\delta^2 M^\eta}{K}\right]\right). \quad (6)$$

III. OSTM

From above one can see that sensing matrices play an important part in compressed sensing because their properties directly affect the performance and complexity. In this section we first give the definition of circulant sensing matrices, and then that of OSTM.

Definition 3 (Circulant sensing matrices): An $M \times N$ circulant sensing matrix Φ is obtained by selecting M rows from an $N \times N$ circulant matrix Φ_N and multiplying a normalized parameter $\sqrt{N/M}$. The $N \times N$ matrix Φ_N has the following properties:

- It is circulant, which means that each row is obtained from the preceding row by rotation to the right by one element.
- All the row vectors and column vectors are real and orthogonal, i.e.

$$\sum_{j=1}^{N} \phi_{js}\phi_{jt} = 0 \quad \text{if } s \neq t, \quad (7)$$

  and have normalized amplitude

$$\sum_{s=1}^{N} \phi_{js}^2 = 1. \quad (8)$$

- The sum of elements in any row or column is ±1.

There are several ways to generate circulant matrices. In this paper, we are interested in OSTM introduced by Bottcher [24]. Denote by OST_N the set of real-valued $N \times N$ OSTM. Bottcher proved that the set OST_N is finite and its cardinality is given by

$$|OST_N| = \begin{cases} 3 \cdot 2^{N/2} - 2 & \text{if } N \text{ is even} \\ 2\sqrt{2} \cdot 2^{N/2} - 2 & \text{if } N \text{ is odd.} \end{cases}$$

The set OST_N consists of circulant and skew-circulant matrices. In this paper, we focus on the case of even N where all matrices in set OST_N are circulant:

Definition 4 (Circulant OSTM [24]): Let N be an even number and $N \geq 4$, and denote by F_N the $N \times N$ Fourier matrix. The set OST_N contains $2^{N/2+1}$ circulants whose first column is given by

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix} = \frac{1}{N} F_N^* \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \vdots \\ \sigma_N \end{pmatrix}$$

where $(\sigma_1, \sigma_2, \cdots, \sigma_N) = (\gamma, \varepsilon_1, \ldots, \varepsilon_{N/2-1}, \beta, \varepsilon_{N/2-1}, \ldots, \varepsilon_1)$ and the parameters $(\gamma, \varepsilon_1, \ldots, \varepsilon_{N/2-1}, \beta) \in \{-1, 1\}^{N/2+1}$.

A circulant OSTM which has the following structure:

$$\mathbf{\Phi}_N = \begin{pmatrix} a & b & c & \ldots & f & g & f & \ldots & c & b \\ b & a & b & c & \ldots & f & g & f & \ldots & c \\ c & b & a & b & c & \ldots & f & g & f & \ldots \\ \vdots & & & & & & & & \ldots & \\ f & & & & & & & & \ldots & \\ g & f & & & & & & & \ldots & \\ f & g & f & & & & & & \ldots & \\ \vdots & & & & & & & & \ldots & \\ b & c & & & & & & & \ldots & \end{pmatrix}.$$

In this paper, $\sigma \triangleq (\sigma_1, \sigma_2, \cdots, \sigma_N)$ is referred to as the *sign sequence*. Thus, the first row (i.e., the transpose of the first column) of OSTM is given by the inverse FFT (IFFT) of the sign sequence. One can generate other rows by successive rotation. Let the diagonal matrix $\Sigma = \mathrm{diag}(\sigma)$. It is not difficult to see that OSTM Φ_N takes the following form:

$$\Phi_N = \frac{1}{N} F_N^* \Sigma F_N. \quad (9)$$

The reason is that the elements of Φ_N defined in (9) have the expression of $\phi_{p,q} = \frac{1}{N}\sum_{k=0}^{N-1} \sigma_{k+1} \exp\{\frac{2\pi i}{N} k(q-p)\}$, while the first row of OSTM obtained from IFFT of the sign sequence is given by $a_n = \frac{1}{N}\sum_{k=0}^{N-1} \sigma_{k+1} \exp\{\frac{2\pi i}{N} k(n-1)\}$ for $n = 1, \cdots, N$. Obviously, $\phi_{p,q} = a_{q-p+1 \mod N}$ which corresponds to the circulant property.

In this paper, an $M \times N$ sensing matrix based on OSTM is generated in the following way:

1) Apply IFFT to the sign sequence to obtain the first row of OSTM.
2) Follow the circulant property to construct the $N \times N$ matrix Φ_N.
3) Choose M rows and normalize it by multiplying $\sqrt{N/M}$ to form the $M \times N$ sensing matrix Φ.

After the second step, it can be proved that the $N \times N$ matrix Φ_N is orthogonal and Toeplitz. Obviously, among $2^{N/2+1}$ such matrices, some are better than others for the purpose of CS. For example, the sign sequences $(1, 1, \cdots, 1)$ and $(-1, -1, \cdots, -1)$ are very bad since they generate a diagonal matrice Φ_N. In the next two Sections, we will analyze the impact of the sign sequence on the statistical RIP, and then identify good sign sequences.

IV. ANALYSIS OF STATISTICAL RIP

In this Section, we apply Stein's method to analyze the statistical RIP and demonstrate the uniqueness of recovery. Appendix A gives a brief summary of concentration inequalities based on Stein's method.

The main theorem is presented below.

Theorem 2: Let x be a length-N, K-sparse signal with non-zero coefficients x_1, x_2, x_3, $\cdots$ x_K. Assume that x has zero-mean (i.e., $\sum_{i=1}^{K} x_i = 0$) and the positions of the K non-zero entries are equiprobable. Let Φ be an $M \times N$ deterministic sensing matrix obtained by selecting M rows arbitrarily from an $N \times N$ circulant matrix satisfying the three conditions in Definition 3, normalized by factor $\sqrt{N/M}$. Then

$$\mathbb{E}\left(\|\Phi x\|^2\right) = \|x\|^2, \quad (10)$$

and

$$\begin{aligned}&\mathbf{P}\left(\left|\|\Phi x\|^2-\|x\|^2\right|<\delta\|x\|^2\right)\\ \geq&1-2\exp\left(-\frac{M\delta^2}{8C_1\cdot K}\right)\end{aligned}\tag{11}$$

when $K \leq C_0(\delta)\frac{M}{\log N}$ (this condition is required such that the failure probability will tend to 0 as $N \to \infty$), $C_0 \sim \delta^2$ is a constant depending on δ only, and $C_1 = \sum_{\zeta=1}^{M} \varphi_\zeta^2$, determined by the M largest values in Φ.

The proof, based on Stein's method, is given in Appendix B.

Remarks:

1) The theorem holds true for circulant matrices statisfying Definition 3 in general and for OSTM in particular.
2) The restriction that x has zero mean is needed only to simplify the derivation so that $\mathbb{E}(\|\Phi x\|^2)$ is fixed. In practice, for x with non-zero mean, we can add an all-ones row vector $\begin{bmatrix}1 & 1 & \cdots & 1\end{bmatrix}$ to measure the DC component of the signal.
3) The bound holds for arbitrary row selection, in contrast to random row selection in other sensing matrices. In particular, we may choose consecutive rows so that the matrix is Toeplitz.
4) It will be shown in the next Section that OSTM can achieve $C_1 \sim \mathcal{O}(1)$ when the sign sequence is derived from a Golay sequence.
5) When $C_1 \sim \mathcal{O}(1)$, OSTM can achieve a bound on the similar order of magnitude in (5) with the tightest condition $\eta = 1$ [12].

When $C_1 \sim \mathcal{O}(1)$, it can be derived from (11) that

$$\mathbf{P}\left(\left|\|\Phi x\|^2-\|x\|^2\right|>\delta\|x\|^2\right)<2\exp\left(-\mathcal{O}\left(\frac{M\delta^2}{K}\right)\right).\tag{12}$$

Clearly (12) decays exponentially with $\frac{M}{K}$, while the bound in (6) decays exponentially with $\frac{M^\eta}{K}$ for $1/2 < \eta \leq 1$.

From (12), we can easily arrive at the following corollary.

Corollary 1: Suppose that x and Φ follow the same definitions in Theorem 2 with $M \geq \frac{K \log N}{C_0 \delta}$. Then, we have

$$\mathbf{P}\left(\left|\|\Phi x\|^2-\|x\|^2\right|\leq\delta\|x\|^2\right)>1-\frac{1}{N},$$

if the sparsity level K satisfies

$$K \leq C_0(\delta)\frac{M}{\log N},\tag{13}$$

where $C_0(\delta)$ has the same definition as that in (11).

Recall that for i.i.d Gaussian and Bernoulli matrices, the RIP holds with high probability when [4]

$$K \leq C_2(\delta)\frac{M}{\log\left(N/M\right)},\tag{14}$$

where $C_2(\delta)$ is a constant depending only on δ. One can observe that (13) takes a similar form to (14). Note that there are no existing solutions of deterministic $M \times N$ sensing matrices that could achieve the RIP bound of (14). In contrast, Corollary 1 suggests that when K is on the similar order, the statistical RIP can be satisfied with high probability.

Meanwhile, it is known that if Φ satisfies $2K$-RIP, unique and stable reconstruction of K-sparse signals will be guaranteed [2]. However, in general statistical RIP in itself does not guarantee unique reconstruction [12]. The probability that $\{\beta \in \mathrm{R}^N; \Phi\alpha = \Phi\beta\} = \{\alpha\}$ should be estimated, and this probability needs to be very close to 1 when α obeys the uniform distribution among all K-sparse vectors in R^N of the same norm. This is indeed the case as shown in the following theorem.

Theorem 3: Suppose that x, Φ and δ follow the same definitions as in Theorem 2 with $M \geq \frac{K \log N}{C_0(\delta)}$. Then x is the only K-sparse vector that satisfies the equation $y = \Phi x$ with probability at least $1-\epsilon$ with respect to the random choice of x, where

$$\epsilon = N\exp\left(-\frac{3(1-\delta-\frac{K}{M})^2M^2}{8(K^2+K+\frac{K^3}{N})}\right).\tag{15}$$

In other words, the probability of unique recovering a K-sparse signal x whose support is randomly selected exceeds $1-\epsilon$. The detailed proof can be found Appendix C.

Now our task is to minimize the M largest values in Φ. In general, the bound of C_1 might be on the same order of N/M. Fortunately, C_1 will be reduced dramatically to $\mathcal{O}(1)$ if we carefully select the sign sequence for OSTM. This will be shown in the next Section.

V. OPTIMIZATION OF OSTM USING GOLAY SEQUENCES

Since the first row of Φ is the IFFT of the sign sequence $\sigma = (\gamma, \varepsilon_1, \ldots, \varepsilon_{N/2-1}, \beta, \varepsilon_{N/2-1}, \ldots, \varepsilon_1)$, minimizing the M largest values amounts to making the frequency spectrum of the sign sequence as flat as possible. One might wonder the usage of the classical *m* sequence whose spectrum is completely flat due to its ideal autocorrelation property. However, the sign sequence has the mirror symmetry such that the *m* sequence is not well suited. Specifically, if we choose an *m* sequence for $(\varepsilon_1, \ldots, \varepsilon_{N/2-1})$, the spectrum of σ will be far from flat.

In this Section, we use the Golay sequence to form the first $N/2$ entries $(\gamma, \varepsilon_1, \ldots, \varepsilon_{N/2-1})$ of the sign sequence. Then let $\beta = -\gamma$ and complete the second half of the sequence which is a mirror image of $(\varepsilon_1, \ldots, \varepsilon_{N/2-1})$.

Golay sequences were introduced by Golay [15] and have since found numerous applications such as peak-to-average power control for orthogonal frequency-division multiplexing (OFDM) [25], [26]. The binary Golay sequences are known to exist for all lengths $2^{\alpha_1}10^{\alpha_2}26^{\alpha_3}$, $\alpha_1, \alpha_2, \alpha_3$ non-negative integers [15].

Definition 5: Let the aperiodic autocorrelation function of a sequence x be defined by

$$R_x(l) = \sum_{j=0}^{N-l-1} x_j x_{j+l}, \quad l = 0, \cdots, N-1.\tag{16}$$

Let $a = (a_0, a_1, \cdots, a_{N-1})$ and $b = (b_0, b_1, \cdots, b_{N-1})$ be a pair of binary sequences with values 1 or -1 only. Then a and b are a Golay complementary pair if

$$R_a(l) + R_b(l) = 0,\tag{17}$$

for $l = 1, \cdots, N-1$. A sequence in any complementary pairs is called a Golay sequence.

It is helpful to see (17) in polynomial form. A sequence a can be associated with the polynomial $a(z) = a_{N-1}z^{N-1} + \cdots + a_1 z + a_0$ in indeterminate z with coefficients ± 1. The Golay sequence pair (a, b) satisfy [15]

$$a(z)a(z^{-1}) + b(z)b(z^{-1}) = 2N. \tag{18}$$

Equations (17) and (18) are equivalent expressions because $a(z)a(z^{-1}) = R_a(0) + \sum_{k=1}^{N-1} R_a(k)(z^k + z^{-k})$. Further, restricting z to lie on the unit circle in the complex plane, we have

$$|a(z)|^2 + |b(z)|^2 = 2N, \quad |z| = 1. \tag{19}$$

This means that the absolute value of each polynomial on the unit circle is bounded by $\sqrt{2N}$ [25]. The reason why we choose a Golay sequence is exactly this property, which means the envelope of its Fourier transform is limited to a small range relative to the mean. Formally, in our context let $S_s(\omega) = \sum_{i=1}^{N/2} s_i e^{j\omega(i-1)}$ for a Golay sequence $s = (s_1, \ldots, s_{N/2})$. Then one has the property

$$\max_{0 \le \omega < 2\pi} |S_s(\omega)|^2 \le N. \tag{20}$$

Since the bound on the Fourier transform of a Golay sequence holds for all envelop, the bound will not change when we add several binary values to form a new sequence.

Theorem 4: Let $\sigma = (\gamma, \varepsilon_1, \ldots, \varepsilon_{N/2-1}, \beta, \varepsilon_{N/2-1}, \ldots, \varepsilon_1)$ be the sign sequence of OSTM Φ_N. Denote by φ_ζ the ζth largest value in a row of Φ_N. If we set $(\gamma, \varepsilon_1, \ldots, \varepsilon_{N/2-1})$ as a Golay sequence and let $\beta = -\gamma$, then we have the bound

$$C_1 = \sum_{\zeta=1}^{M} \varphi_\zeta^2 \le 4. \tag{21}$$

Proof: First, note that the maximum value φ_1 satisfies

$$\varphi_1^2 = \frac{N}{M} \max_k \{a_k^2\} = \frac{1}{MN} \max_k \left| \sum_{i=1}^{N} \sigma_i e^{j\omega_k (i-1)} \right|^2 \tag{22}$$

where $\omega_k = 2\pi(k-1)/N$. The factor $\frac{N}{M}$ comes from normalization. Due to the mirror-symmetrical structure of the sign sequence, it is not difficult to tell that the second half of the sum is a complex conjugate of the first half. The terms pertaining to σ_i appear in pairs with conjugate values. Then because

$$\left| \sum_{i=1}^{N} \sigma_i e^{j\omega_k (i-1)} \right|^2 \le 2 \left| \sum_{i=1}^{N/2} \sigma_i e^{j\omega_k (i-1)} \right|^2 + 2 \left| \sum_{i=N/2+1}^{N} \sigma_i e^{j\omega_k (i-1)} \right|^2 \tag{23}$$

using the conjugate property, we obtain

$$\varphi_1^2 \le \frac{4}{MN} \max_k \left| \sum_{i=1}^{N/2} s_i e^{j\omega_k (i-1)} \right|^2 = \frac{4}{MN} \max_k |S_s(\omega_k)|^2 \le \frac{4}{M} \tag{24}$$

where the last inequality is due to (20). Now it can be concluded that

$$C_1 = \sum_{\zeta=1}^{M} \varphi_\zeta^2 \le \varphi_1^2 \cdot M \le 4, \tag{25}$$

which gives the near-optimal C_1.

■

Remarks:

1) If $\sigma = (1, 1, \cdots, 1)$, then $a_1 = 1$ and C_1 can be as large as N. Namely, the bound will be very bad.
2) On the other hand, by Parseval's theorem we have the lower bound $C_1 \ge \frac{N}{M} \cdot M \cdot \frac{1}{N} \sum_{k=1}^{N} a_k^2 = \frac{1}{N} \sum_{k=1}^{N} \sigma_k^2 = 1$ for any OSTM. Therefore, setting s as a Golay sequence is optimal within a factor of 4.

VI. SIMULATION RESULTS

Extensive simulations have been carried out to compare the reconstruction performance of different random and deterministic sensing matrices. For illustration purposes, we first present here some results for different 64×512 and 128×1024 sampling matrices with various reconstruction algorithms. Then, recovery performance is compared to that of random convolution [20].

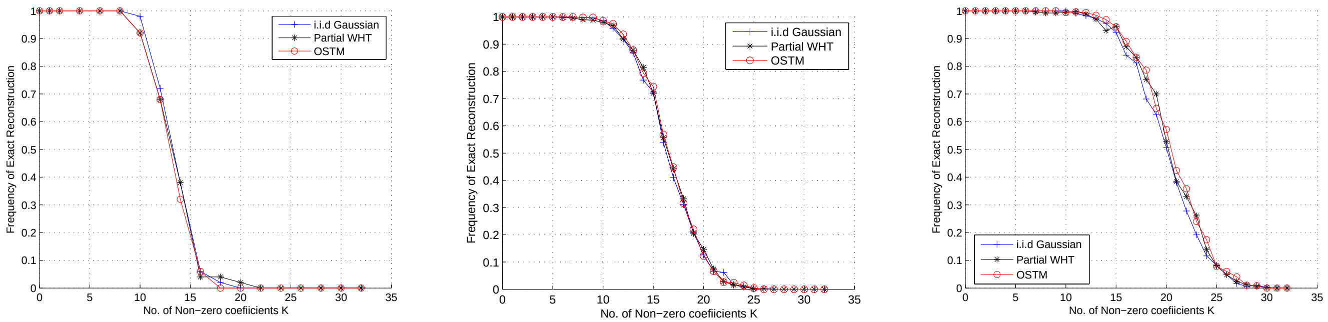
In the first simulation with sparse Gaussian signals, the details are shown as follows:

1) 64×512 *sensing matrices*: These sampling matrices include the random i.i.d. Gaussian matrix, the random partial Walsh-Hadamard transform (WHT) matrix and the OSTM;
2) 128×1024 *sensing matrices*: Sampling matrices under comparison are the random i.i.d. Gaussian matrix, the random partial Discrete Cosine Transform (DCT) matrix and the OSTM.

The reconstruction algorithms are based on the l_1-regularized least-squares algorithm [27], subspace pursuit (SP) [28], and the sparsity adaptive matching pursuit (SAMP) [29], respectively. The non-zero coefficients x_i $(i = 1, \cdots K)$ of the input signal obey the Gaussian distribution.

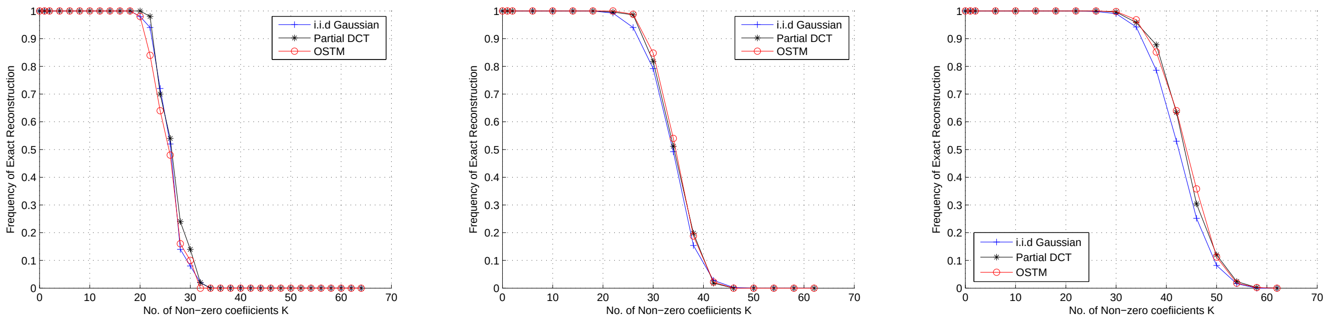
Fig. 1 and Fig. 2 depict the empirical frequencies of exact reconstruction for 64×512 sensing matrices and 128×1024 sensing matrices, respectively. In these simulations, 1000 trials were run for each sparsity level K and the positions of non-zero elements are selected uniformly at random. Besides, we assume that the exact reconstruction is achieved if the signal to noise ratio (SNR) is greater than 50 dB. From these figures, one can observe that the performance of deterministic matrices are quite similar to those of the random matrices.

Simulations have also been run on images. We used the MRI Shepp-Logan phantom image (128×128) as a test image and the results are compared with random convolution. The fast reconstruction algorithm for Toeplitz matrices with same parameters in [16] was applied. 500 trials were run for each sampling level and the results with 10% sampling rate are shown in Fig. 3. Results with different sampling rates can be seen in Table I. The reconstruction SNR shows that the performance of OSTM is a little better than that of random



(a) (b) (c)

Fig. 1. Simulation results for different 64×512 sensing matrices. (a) l_1 regularized least squares algorithm. (b) SP algorithm. (c) SAMP algorithm.



(a) (b) (c)

Fig. 2. Simulation results for different 128×1024 sensing matrices. (a) l_1-regularized least squares algorithm, (b) SP algorithm, (c) SAMP algorithm.

TABLE I
SNR (IN DB) COMPARISON OF RANDOM CONVOLUTION AND OSTM

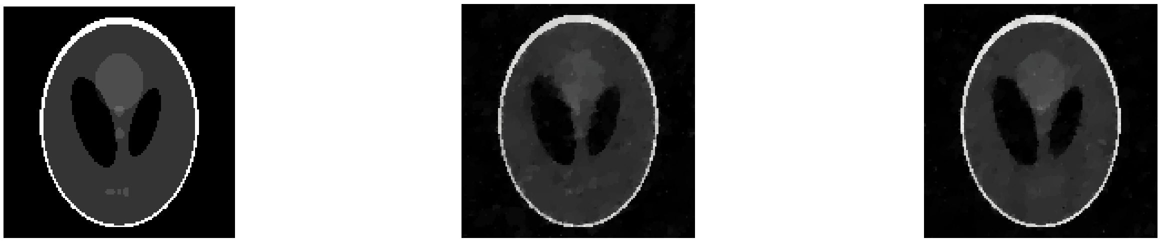
Sampling rate M/N	14%	13%	12%	11%	10%	9%
Random convolution	83.20	54.21	27.55	21.38	15.94	12.83
OSTM	83.20	57.87	28.37	21.49	16.49	13.29

convolution. It is because the sign sequence is carefully chosen instead of a random one.

OSTM is well suited to applications with large values of N, since both sampling and reconstruction will benefit from efficient implementation by means of FFT. The complexity of basis pursuit (BP) is N^3, while the complexity of the SP algorithm is upper bounded by $\mathcal{O}(MNK)$. The computational complexity of the fast reconstruction algorithm in [16] is complicated because it is not only a l_1 minimization problem, but also includes equality fidelity, l_1 and l_2 square penalized fidelity, as well as one of or both l_1 and total variation regularizations. Its theoretical computation speed is faster than that of SP because of using FFT. Fig. 4 demonstrates the recovery peak SNR (PSNR) and consuming time of processing image Lena (256×256) under various sampling rates for random convolution and OSTM using the fast recovery algorithm in [16], structurally random matrices (SRM-SAMP) using SAMP, and Gaussian matrices using block-SP. Because K is unknown in natural image processing, SAMP is applied here with Daubechies' 9/7 wavelet as the sparsifying matrix. SAMP is known for offering a comparable theoretical guarantee as the best optimization-based approach, and both the OMP and the SP can be viewed as SAMP's special cases. Structurally random matrices are used as the sampling operator in SAMP due to their fast and efficient implementations [5]. Results of SP with block size 64×64 and Gaussian random matrices as sensing matrices are shown as a comparison of the processing time. A natural image is often too large to implement SP in Matlab directly so we use block-SP in order to avoid running out of memory. In contrast, the fast reconstruction algorithm in [16] can deal with much longer signals with less simulation time, because the image is viewed as a two-dimensional matrix. The results show that OSTM with the fast recovery algorithm in [16] achieves similar PSNR to that of random convolution, and better than that of SRM-SAMP and block-SP. The fast recovery algorithms reveal an obvious superiority in computational complexity.

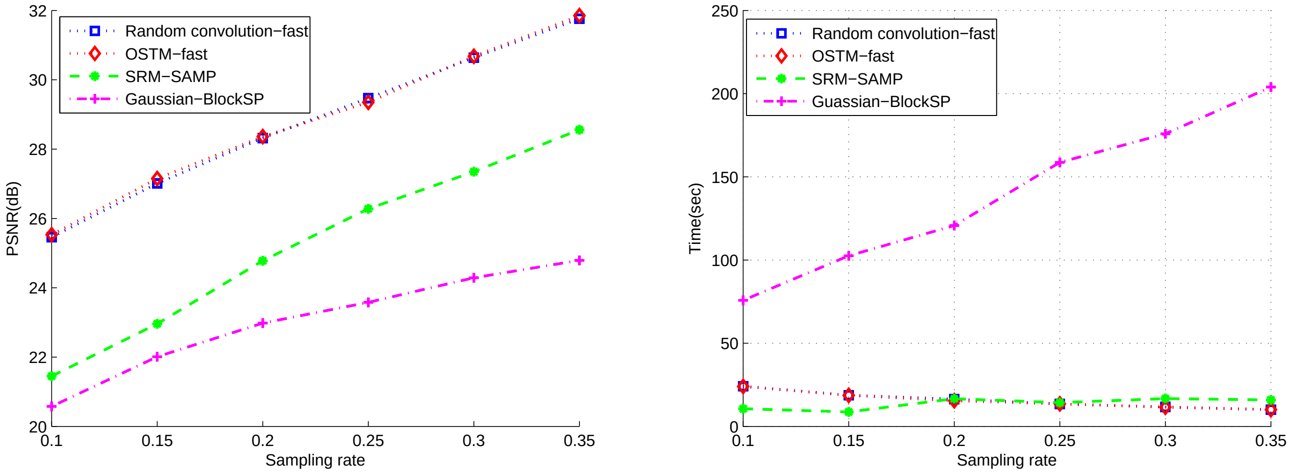
VII. CONCLUSIONS AND FUTURE WORK

We have investigated the statistical RIP of OSTM in compressed sensing. Specifically, we have shown that OSTM satisfies the statistical RIP and unique recovery with high probability except an exponentially small fraction. If the underlying sign sequence is obtained from a Golay sequence, the OSTM will achieve near-optimal statistical RIP. Experimental results show that these deterministic sensing matrices compare



(a) (b) (c)

Fig. 3. Simulation results for different 128×1024 sensing matrices. (a) Original phantom image. (b) Recovered image from random convolution with 10% sampling rate, SNR=15.64 dB. (c) Recovered image from OSTM with 10% sampling rate, SNR=16.38 dB.



(a) (b)

Fig. 4. Reconstruction of the Lena image (256×256) under various sampling rates for random convolution and OSTM using fast recovery algorithm in [16], SAMP with structurally random matrices, and Gaussian matrices with block-SP. (a) Reconstruction PSNR in dB. (b) Consuming time in seconds. Note that the reconstructing times of random convolution and OSTM decrease with the sampling rate, because the fast reconstruction algorithm terminates earlier when more measurements are obtained.

favorably with existing random matrices. Our proof was based on Stein's method.

There are many intriguing questions that future work should consider. First, the noise resilience of these deterministic operators needs to be analyzed and evaluated. Secondly, the applications of OSTM-based sensing matrices can be exploited. Thirdly, new reconstruction algorithms for OSTM-based sensing need to be developed. Furthermore, it is interesting to further consider the statistical model of the input signal in the statistical RIP formulation.

APPENDIX

A. Concentration inequality based on Stein's method

Stein's method is a powerful tool to obtain bounds on the distance between two probability distributions with respect to a probability metric. A fundamental idea in Stein's method is the *exchangeable pair*. Two random variables Z and Z' are said to form an exchangeable pair if the joint probability density function of (Z, Z') is equal to that of (Z', Z). In [14], Stein's method was used to derive measure concentration, as stated in the following theorem:

Theorem 5 ([14]): Suppose that (Z, Z') is an exchangeable pair of random variables. Let $F(Z, Z')$ be an antisymmetric function, i.e., $F(Z, Z') = -F(Z', Z)$ with $E(F(Z, Z')|Z) = f(Z)$. Define $\Delta(Z)$ as

$$\Delta(Z) = \frac{1}{2}\mathbb{E}\left(|(f(Z) - f(Z'))F(Z, Z')|\,\big|Z\right).$$

Then, $\mathbb{E}(f(Z)) = 0$ and if there exist non-negative constants a_0 and a_1 such that $\Delta(Z) \leq a_0 + a_1 f(Z)$ almost surely, then for any $t \geq 0$, we have

$$\mathbf{P}(f(Z) \geq t) \leq \exp\left(-\frac{t^2}{2a_0 + 2a_1 t}\right) \tag{26}$$

and

$$\mathbf{P}(f(Z) \leq -t) \leq \exp\left(-\frac{t^2}{2a_0}\right). \tag{27}$$

As the statistical RIP is based on the permutation operator π, in this paper we will derive a new bound of (4) by applying Theorem 5 to the concentration inequality for random permutation (i.e., when $Z = \pi$).

B. Proof of Theorem 2

For the random permutation operator π in (3), we can define its exchangeable pair as $\pi' = \pi \circ (I, J)$ [14], where I and J are chosen uniformly and independently at random from $\{1, \cdots, N\}$. (I, J) denotes the transposition of I and J, i.e., $\pi'(I) = \pi(J)$, $\pi'(J) = \pi(I)$ and $\pi'(i) = \pi(i)$ for $i \neq I, J$.

Define the antisymmetric function $\mathbf{F}(\pi, \pi')$ as follows:

$$\begin{aligned}
\mathbf{F}(\pi, \pi') &= \frac{N}{4}\left(\|y(\pi)\|^2 - \|y(\pi')\|^2\right) \\
&= \frac{N}{4}\left(\|y(\pi)\|^2 - \|y(\pi) - r\|^2\right) \\
&= \frac{N}{4}\left(\|y(\pi)\|^2 - \|y(\pi)\|^2 + 2y(\pi)^H r(\pi) - \|r(\pi)\|^2\right) \\
&= \frac{N}{4}\left(2y^H(\pi)r(\pi) - \|r(\pi)\|^2\right)
\end{aligned} \tag{28}$$

where $y(\pi)$ is given by (3) and $r(\pi)$ is the difference between $y(\pi)$ and $y(\pi')$ that can be expressed as

$$\begin{aligned}
r(\pi) &= y(\pi) - y'(\pi) \\
&= (\phi_{\pi(I)} - \phi_{\pi(J)})(x_I - x_J) \\
&= \phi_{\pi(I)}x_I + \phi_{\pi(J)}x_J - \phi_{\pi(I)}x_J - \phi_{\pi(J)}x_I.
\end{aligned} \tag{29}$$

Next, according to Theorem 5, we need to find

$$\begin{aligned}
f(\pi) &= \mathbb{E}\left(F(\pi, \pi')|\pi\right) \\
&= \frac{N}{4}\mathbb{E}_{I,J}\left(2y^H(\pi)r(\pi) - \|r(\pi)\|^2\right).
\end{aligned}$$

From the expression of $r(\pi)$ in (29), by summing over all choices of $1 \leq I, J \leq N$, we have

$$\mathbb{E}_{I,J}\left(2y^H(\pi)r(\pi)\right) = \frac{4\|y(\pi)\|^2}{N}, \tag{30}$$

$$\mathbb{E}_{I,J}\left(\|r(\pi)\|^2\right) = \frac{4\|x\|^2}{N}, \tag{31}$$

where we have used the assumptions that $\sum_{i=1}^N \phi_i = \mathbf{0}$ and $\sum_{i=1}^N x_i = 0$. Substituting the above results into $f(\pi)$ yields

$$f(\pi) = \|y(\pi)\|^2 - \|x\|^2. \tag{32}$$

By Theorem 5, we know that $\mathbb{E}(f(\pi)) = 0$, which implies that (10) holds.

To get the concentration inequality in (11), we thus need to bound

$$\begin{aligned}
\Delta(\pi) &= \frac{1}{2}\mathbb{E}\left(\left|\left(f(\pi) - f(\pi')\right)F(\pi, \pi')\right| \big|\pi\right) \\
&= \frac{N}{8}\mathbb{E}\left(\left(2y^H(\pi)r(\pi) - \|r(\pi)\|^2\right)^2 \big|\pi\right).
\end{aligned} \tag{33}$$

Here,

$$\begin{aligned}
2y^H(\pi)r(\pi) &= 2(\sum_{k=1}^N \phi^T_{\pi(k)}x_k)(\phi_{\pi(I)} - \phi_{\pi(J)})(x_I - x_J) \\
&= 2[\sum_{k\neq I,J}^N \phi^T_{\pi(k)}x_k + \phi^T_{\pi(I)}x_I + \phi^T_{\pi(J)}x_J] \\
&\quad (\phi_{\pi(I)} - \phi_{\pi(J)})(x_I - x_J).
\end{aligned} \tag{34}$$

Note that

$$\begin{aligned}
&(\phi^T_{\pi(I)}x_I + \phi^T_{\pi(J)}x_J)(\phi_{\pi(I)} - \phi_{\pi(J)})(x_I - x_J) \\
&\quad = x_I(\phi^T_{\pi(I)}\phi_{\pi(I)} - \phi^T_{\pi(I)}\phi_{\pi(J)})(x_I - x_J) \\
&\qquad + x_J(\phi^T_{\pi(J)}\phi_{\pi(I)} - \phi^T_{\pi(J)}\phi_{\pi(J)})(x_I - x_J) \\
&\quad = (1 - \phi^T_{\pi(I)}\phi_{\pi(J)})(x_I - x_J)^2 = \frac{\|r(\pi)\|^2}{2}.
\end{aligned} \tag{35}$$

As a result,

$$\begin{aligned}
&\mathbb{E}\left(2y^H(\pi)r(\pi) - \|r(\pi)\|^2\right)^2 \\
=&\mathbb{E}\left(2(\sum_{k\neq I,J}^N \phi^T_{\pi(k)}x_k)(\phi_{\pi(I)} - \phi_{\pi(J)})(x_I - x_J)\right)^2.
\end{aligned} \tag{36}$$

To bound (36), we need three lemmas in the following.

Lemma 1: Let x be a length-N, K-sparse signal and $\Phi = \left[\phi_1, \phi_2, \cdots, \phi_N\right]$ be an $M \times N$ normalized sensing matrix. Then (36) is bounded by

$$\begin{aligned}
&\mathbb{E}\left(2y^H(\pi)r(\pi) - \|r(\pi)\|^2\right)^2 \\
&\leq \frac{16}{N}\|x\|^2 \cdot \sum_{j=1}^K \sum_{\substack{i=1,\\ i\neq w_j}}^N (\phi^T_{\pi(w_j)}\phi_{\pi(i)})^2 \cdot \mathbb{E}(x_I - x_J)^2,
\end{aligned} \tag{37}$$

where $w_j \in \{1, 2, \cdots, N\}$, $j = 1 \ldots K$ denotes the index of K non-zero values of x.

Proof: According to (36),

$$\begin{aligned}
&\mathbb{E}\left(2y^H(\pi)r(\pi) - \|r(\pi)\|^2\right)^2 \\
=&\mathbb{E}\left[4\left(\sum_{k\neq I,J}^N(\phi^T_{\pi(k)}x_k\phi_{\pi(I)}) - \sum_{k\neq I,J}^N(\phi^T_{\pi(k)}x_k\phi_{\pi(J)})\right)^2 (x_I - x_J)^2\right].
\end{aligned} \tag{38}$$

When permutations π and π' are fixed, ϕ_i's and x are also fixed, except for the pair I and J. The expectation will be taken over all (I, J) pairs. Because the two terms in the brackets share no common values, the expectation can be calculated separately.

The first term can be bounded by

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{k\neq I,J}^N(\phi^T_{\pi(k)}x_k\phi_{\pi(I)}) - \sum_{k\neq I,J}^N(\phi^T_{\pi(k)}x_k\phi_{\pi(J)})\right)^2\right] \\
&\leq \mathbb{E}\left[2\sum_{k\neq I,J}^N(\phi^T_{\pi(k)}x_k\phi_{\pi(I)})^2 + 2\sum_{k\neq I,J}^N(\phi^T_{\pi(k)}x_k\phi_{\pi(J)})^2\right] \\
&\leq \mathbb{E}\left[4\sum_{k\neq I,J}^N(\phi^T_{\pi(k)}x_k\phi_{\pi(I)})^2\right].
\end{aligned} \tag{39}$$

By the Cauchy-Schwarz inequality, we have

$$\begin{aligned}&\sum_{k\neq I,J}^{N}(\phi^T_{\pi(k)}x_k\phi_{\pi(I)})^2 \le \sum_{\substack{j=1,\\ j\neq I,J}}^{N}(\phi^T_{\pi(j)}\phi_{\pi(I)})^2\sum_{\substack{j=1,\\ j\neq I,J}}^{N}x_j^2\\ &\le \sum_{\substack{j=1,\\ j\neq I}}^{N}(\phi^T_{\pi(j)}\phi_{\pi(I)})^2\sum_{j=1}^{N}x_j^2 = \sum_{\substack{j=1,\\ j\neq I}}^{N}(\phi^T_{\pi(j)}\phi_{\pi(I)})^2\|x\|^2.\end{aligned} \tag{40}$$

Substitute this into (39), we obtain

$$\begin{aligned}&\mathbb{E}\left[4\sum_{k\neq I,J}^{N}(\phi^T_{\pi(k)}x_k\phi_{\pi(I)})^2\right]\\ &\le \frac{4}{N}\|x\|^2\sum_{i=1}^{N}\sum_{\substack{j=1,\\ j\neq i}}^{N}(\phi^T_{\pi(j)}\phi_{\pi(i)})^2.\end{aligned} \tag{41}$$

Recall that the signal x is assumed to be sparse, i.e., there are only K non-zero values among $\{x_i, i = 1\cdots N\}$. Denote the indices of these K non-zero values as $w_1, w_2, \ldots, w_K$, $w_j \in (1, N)$. Then, (39) can be rewritten as

$$\begin{aligned}&\mathbb{E}\left[4\sum_{k\neq I,J}^{N}(\phi^T_{\pi(k)}x_k\phi_{\pi(I)})^2\right]\\ &=\mathbb{E}\left[4\sum_{w_j\neq I,J}(\phi^T_{\pi(w_j)}x_{w_j}\phi_{\pi(I)})^2\right]\\ &\le \frac{4}{N}\|x\|^2\sum_{i=1}^{N}\sum_{\substack{j=1,\\ j\neq i}}^{K}(\phi^T_{\pi(w_j)}\phi_{\pi(i)})^2\\ &\le \frac{4}{N}\|x\|^2\sum_{j=1}^{K}\sum_{\substack{i=1,\\ i\neq j}}^{N}(\phi^T_{\pi(w_j)}\phi_{\pi(i)})^2\end{aligned} \tag{42}$$

which leads to Lemma 1 immediately. ■

Lemma 2: Let x be a length-N, K-sparse signal and Φ be an $M \times N$ normalized sensing matrix. Then

$$\sum_{j=1}^{K}\sum_{\substack{i=1,\\ i\neq w_j}}^{N}(\phi^T_{\pi(w_j)}\phi_{\pi(i)})^2 \le \frac{N}{M}\left[\sum_{j\in(1,N)}^{K}\sum_{i\in(1,N)}^{M}v_{ij}^2\right], \tag{43}$$

where v_{ij}'s are the entries of Φ indexed by the selected M rows and K non-zero values of x (cf. Fig. 5).

Proof: Fig. 5 shows the $M\times N$ sensing matrix. s and t are any two of the column indexes corresponding to indexes w_s and w_t of nonzero values in x. p, q are any two row indexes, $p \neq q$. The cross elements of two columns and two rows are v_{ps} and v_{qs}. Here we notice that the left side of (43) consists of the quadratic sum of two vectors, so it can be divided into two parts: *non-cross term* and *cross term*.

Consider terms containing v_{ps}. The *non-cross term* in (43) pertaining v_{ps} is

$$v_{ps}^2\cdot\sum_{\substack{i=1\\ i\neq s}}^{N}v_{pi}^2 = v_{ps}^2\cdot\left(\frac{N}{M}-v_{ps}^2\right). \tag{44}$$

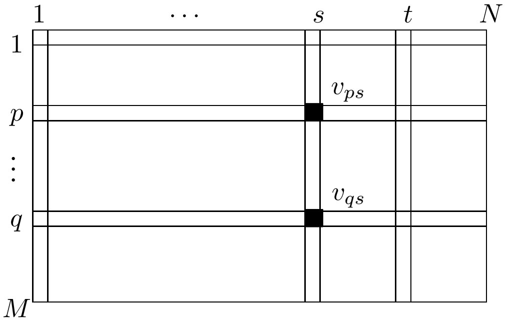


Fig. 5. The schematic diagram of sensing matrix Φ with M rows. p and q are row indexes. s and $t \in \{w_i | i = 1, \ldots, K\}$ are indexes of columns corresponding to non-zero values in x. v_{ps} denotes the element in the sensing matrix within pth row and sth column. v_{qs} shares the similar definition.

Equation (44) holds for any row in the sensing matrix, as the quadratic sum is always N/M. When it comes to the *cross term*, the result in (43) will be a little complicated. Similarly, we consider the terms pertaining v_{ps} and v_{qs} as follows:

$$\begin{aligned}&v_{ps}v_{qs}\cdot\left(\sum_{\substack{i=1\\ i\neq s}}^{N}v_{pi}v_{qi}\right)\\ =&v_{ps}v_{qs}\cdot\left(v_p\cdot v_q^T - v_{ps}v_{qs}\right)\\ =&-\left(v_{ps}v_{qs}\right)^2,\end{aligned} \tag{45}$$

where v_p and v_q are two row vectors. The last equality exploits the orthogonal property of the sensing matrix.

According to (44) and (45), the result which includes all elements in column s will be

$$\frac{N}{M}\left(\sum_{i=1}^{M}v_{is}^2\right)-\sum_{i=1}^{M}v_{is}^4-\sum_{\substack{i=1,j=1\\ i\neq j}}^{M}v_{is}^2v_{js}^2 \le \frac{N}{M}\left(\sum_{i=1}^{M}v_{is}^2\right). \tag{46}$$

Repeating this for all columns, we derive (43) easily. ■

Lemma 3: Let x be a length-N, K-sparse signal and Φ be an $M \times N$ the normalized sensing matrix. v_{ij} is given in the definition in Lemma 2. Then, the bound of $\sum_{j\in(1,N)}^{k}\sum_{i\in(1,N)}^{M}v_{ij}^2$ is

$$\sum_{j\in(1,N)}^{K}\sum_{i\in(1,N)}^{M}v_{ij}^2 \le K\sum_{\zeta=1}^{M}\varphi_\zeta^2 \le K\cdot M\varphi_1^2 \tag{47}$$

where φ_ζ denotes the ζth largest value of a row in Φ.

Remark 1: Since the elements of every row in Φ are the same in circulant matrices, φ_ζ is the ζth largest value for any row.

Proof: Consider the column v_j. Obviously,

$$\sum_{i\in(1,N)}^{M}v_{i,j}^2 \le \sum_{\zeta=1}^{M}{\varphi_\zeta}^2. \tag{48}$$

Since the each column is constituted from the same sequence of numbers, (48) applies to all columns. Moreover, since in our case the sensing matrix is real-valued, Lemma 3 follows straightforwardly. ■

Let $C_1 = \sum_{\zeta=1}^{M} {\varphi_\zeta}^2$. Applying the lemmas above, $\Delta(\pi)$ can be bounded as follows

$$\begin{aligned}\Delta(\pi) &= \frac{N}{8}\mathbb{E}_{I,J}\left(\left(2y^H r - \|r\|^2\right)^2 |\pi\right)\\ &\le \frac{N}{8}\cdot\frac{16K\cdot C_1}{M}\|x\|^2\mathbb{E}\left[(x_I - x_J)^2\right]\\ &\le \frac{N}{8}\cdot\frac{16K\cdot C_1}{M}\|x\|^2\cdot\frac{2\|x\|^2}{N}\\ &\le \frac{4C_1\cdot K}{M}\cdot\|x\|^4.\end{aligned}\tag{49}$$

Now we are ready to derive the final result. Using (26) and (27) in Theorem 5, we have

$$\begin{aligned}\mathbf{P}\left(f(\pi)\ge t\right) &\le \exp\left(-\frac{t^2}{2\frac{4C_1\cdot K}{M}\cdot\|x\|^4}\right)\\ \mathbf{P}\left(\|y\|^2 - \|x\|^2 \ge t\right) &\le \exp\left(-\frac{Mt^2}{8C_1\cdot K\|x\|^4}\right).\end{aligned}\tag{50}$$

Let $t = \delta\|x\|^2$. (11) can be easily obtained:

$$\begin{aligned}&\mathbf{P}\left(\left|\|\Phi x\|^2 - \|x\|^2\right| < \delta\|x\|^2\right)\\ \ge& 1 - 2\exp\left(-\frac{M\delta^2}{8C_1\cdot K}\right),\end{aligned}\tag{51}$$

which completes the proof of Theorem 2.

C. Proof of Theorem 3

Proof: Here we prove the unique reconstruction of statistical RIP of OSTM by using Stein's method. The proof shares the same definition of random permutation operator π and the exchangeable pair π, π'. Let $\tau = \{\tau_1, \tau_2, \cdots, \tau_K\}$ be a set of K random columns of Φ, and w be a random selected column of Φ apart from τ. Table II refers to the cases of locations of I, J. Because we need to calculate the expectation of F, the results of different cases will be added together according to the locations of I, J. In Table II, we define the cases depending on when I, J are in τ, w or not, because the locations of I, J will make the result of $\left(\|y(\pi)\|^2 - \|y(\pi')\|\right)$ very different.

Define the antisymmetric function $F(\pi, \pi')$ as follows:

$$F(\pi,\pi') = \frac{N(N+1)}{3N-2}\left(\|y(\pi)\|^2 - \|y(\pi')\|^2\right)\tag{52}$$

where $y = \phi_\tau^T\phi_w$.

According to Theorem 5, we need to find

$$\begin{aligned}f(\pi) &= \mathbb{E}\left(F(\pi,\pi')|\pi\right)\\ &= \mathbb{E}_{I,J}\left(\|y(\pi)\|^2 - \|y(\pi')\|^2\right).\end{aligned}\tag{53}$$

The expectation is taken over all (I, J) pairs (there are $N(N-1)$ of them). Given the positions of τ and w, we analyze the (I, J) pairs case by case.

Case 1: When I, J are not in τ, w, $\|y(\pi)\|^2 - \|y(\pi')\|^2 = 0$. This is because the value of $\|\phi_\tau^T\phi_w\|_2$ will not change if I, J are exchanged.

Case 2: When $I \in \tau$, $J \notin \tau$ and $J \notin w$, which means J is in the $(N-K-1)$ columns (the case of $J \in \tau$, $I \notin \tau$ and $I \notin w$ is exactly similar). Write $\|\phi_w\|^2 = \Upsilon^2$. Then,

$$\begin{aligned}\sum_{i=1}^{N}\|\phi_i^T\phi_w\|^2 &= \sum_{i=1}^{N}\left(\sum_{j=1}^{M} v_{jw}v_{ji}\right)^2\\ &= \sum_{j=1}^{N}\left(v_{jw}^2\cdot\sum_{i=1}^{N}v_{ji}^2\right) + \sum_{p\neq q}\left(v_{pw}v_{qw}\sum_{i=1}^{N}v_{pi}v_{qi}\right)\\ &= \sum_{j=1}^{N}\left(v_{jw}^2\cdot\frac{N}{M}\right) + \sum_{p\neq q}\left(v_{pw}v_{qw}\cdot 0\right)\\ &= \sum_{j=1}^{N}\left(v_{jw}^2\right)\cdot\frac{N}{M}\\ &= \frac{N}{M}\Upsilon^2\end{aligned}\tag{54}$$

due to Condition 2) and Condition 3). Thus,

$$\mathbb{E}_I\left(\|\phi_I^T\phi_w\|^2\right) = \frac{1}{K}\|y\|^2,\tag{55}$$

$$\begin{aligned}\mathbb{E}_J\left(\|\phi_J^T\phi_w\|^2\right) &= \frac{\left(\|\phi_{all}^T\phi_w\|^2 - \|\phi_\tau^T\phi_w\|^2 - |\phi_w^T\phi_w|^2\right)}{N-K-1}\\ &= \frac{(\frac{N}{M}-1)\Upsilon^2 - \|y\|^2}{N-K-1}.\end{aligned}\tag{56}$$

Substituting the above results into $f(\pi)$ yields the sum of the second case in the table

$$\begin{aligned}&\sum\left(\|y(\pi)\|^2 - \|y(\pi')\|^2\right) = 2K(N-K-1)\cdot\\ &\left[\left(\frac{\|y\|^2}{K} + \frac{\|y\|^2}{N-K-1}\right) - \frac{N-M}{M(N-K-1)}\Upsilon^2\right]\\ &= 2(N-1)\|y\|^2 - 2(N-M)\cdot\frac{K}{M}\cdot\Upsilon^2.\end{aligned}\tag{57}$$

Case 3: When $J \in w$, $I \notin \tau$ and $I \notin w$, which means I is in the $(N-K-1)$ columns (the case of $I \in w$, $J \notin \tau$ and $J \notin w$ is exactly similar).

Case 4: When $I \in \tau$ and $J \in w$ (the case of $J \in \tau$, $I \in w$ is exactly similar).

Now consider the combination of the third and fourth cases. For instance, J is in w, and I is any column except for w. The sum of the two cases is

$$\begin{aligned}&\sum\left(\|y(\pi)\|^2 - \|y(\pi')\|^2\right)\\ &= \left[(N-K-1)\|y\|^2 - \sum_{\substack{I\text{ is in}\\(N-K-1)}}\|\phi_\tau^T\phi_I\|^2\right] +\\ &\left[K\|y\|^2 - \sum_{\substack{I\text{ is in}\\(\tau)}}\|\phi_\tau^T\phi_I\|^2\right]\\ &\doteq (N-1)\|y\|^2 - \left(\frac{N}{M}\Upsilon^2\cdot K - \|y\|^2\right)\\ &= N\|y\|^2 - N\cdot\frac{K}{M}\Upsilon^2\end{aligned}\tag{58}$$

TABLE II

$\|y(\pi)\|^2 - \|y(\pi')\|^2$ IN DIFFERENT CASES OF (I, J) PAIRS

Case	Location of I, J	$\|y(\pi)\|^2 - \|y(\pi')\|^2$	Number of Terms
1	I, J are not in τ, w	0	$(N-K-1)(N-K-2)$
2	I is in τ, or J is in τ	$\|\phi_I^T\phi_w\|^2 - \|\phi_J^T\phi_w\|^2$	$2K(N-K-1)$
3	I or J is in w	$\|\phi_\tau^T\phi_I\|^2 - \|\phi_\tau^T\phi_J\|^2$	$2(N-K-1)$
4	I is in τ and J is w, or J is in τ and I is w	$\|\phi_\tau^T\phi_J\|^2 - \|\phi_\tau^T\phi_I\|^2$	$2K$
5	I, J are in τ	0	$K(K-1)$

The last two equations hold because the norm of each column is very close to the norm of column w when $M \to \infty$.

Case 5: I, J are both in τ. Similar to Case1, after I, J are exchanged, the value of $\|\phi_\tau^T\phi_w\|_2$ will not change. So the result will also be 0 in this case.

Sum up all above and applying Stein's method, we have

$$\begin{aligned}\mathbb{E}\left(f(\pi)\right) = &\frac{N(N-1)}{3N-2}\cdot\frac{1}{N(N-1)}\cdot \\ &\mathbb{E}\left[(3N-2)\|y(\pi)\|^2 - (3N-2M)\cdot\frac{K}{M}\cdot\Upsilon^2\right] \\ = &\,\mathbb{E}\left(\|y(\pi)\|^2\right) - \frac{3N-2M}{3N-2}\cdot\frac{K}{M}\cdot\Upsilon^2 = 0,\end{aligned} \tag{59}$$

which indicates that $\mathbb{E}\left(\|y(\pi)\|^2\right) = \frac{3N-2M}{3N-2}\cdot\frac{K}{M}\cdot\Upsilon^2$. When $N \gg M$, the term $\frac{3N-2M}{3N-2} \to 1$, meaning that $\mathbb{E}\left(\|y(\pi)\|^2\right) \doteq \frac{K}{M}\cdot\Upsilon^2$. In addition, we have the upper bound on $\Delta(\pi)$ in terms of elements in Table II:

$$\begin{aligned}\Delta(\pi) = &\frac{1}{2}\mathbb{E}\left(|(f(\pi)-f(\pi'))\mathbf{F}(\pi,\pi')|\,|\pi\right). \\ \leq &\frac{N(N-1)}{2(3N-2)}\cdot\frac{1}{N(N-1)}\cdot\{\frac{4}{M^2}\Upsilon^4\cdot 2K(N-K-1) \\ &+\frac{4K^2}{M^2}\Upsilon^4\cdot 2(N-K-1)+\frac{4K^2}{M^2}\Upsilon^4\cdot 2K\} \\ \leq &\frac{4(N-K-1)\Upsilon^4}{M^2(3N-2)}\cdot(K+K^2+\frac{K^3}{N-K-1}).\end{aligned} \tag{60}$$

If $N \gg K$, the result becomes

$$\frac{4\Upsilon^4}{3M^2}\cdot(K^2+K+\frac{K^3}{N}).$$

Based on the above result, Stein's method indicates that

$$\begin{aligned}&\mathbf{P}\left(\|y(\pi)\|^2 - \frac{3N-2M}{3N-2}\cdot\frac{K}{M}\cdot\Upsilon^2 \geq t\right) \\ &\leq \exp\left(-\frac{t^2}{2\cdot\frac{4\Upsilon^4}{3M^2}\cdot(K^2+K+\frac{K^3}{N})}\right).\end{aligned} \tag{61}$$

Substituting $t = (1-\delta)\Upsilon^2 - \frac{3N-2M}{3N-2}\cdot\frac{K}{M}\cdot\Upsilon^2$ into the inequality above, we have

$$\begin{aligned}&\mathbf{P}\left(\|y(\pi)\|^2 \geq (1-\delta)\Upsilon^2\right) \\ &\leq \exp\left(-\frac{(1-\delta-\frac{3N-2M}{3N-2}\cdot\frac{K}{M})^2\Upsilon^4}{2\cdot\frac{4\Upsilon^4}{3M^2}\cdot(K^2+K+\frac{K^3}{N})}\right) \\ &\leq \exp\left(-\frac{3(1-\delta-\frac{K}{M})^2M^2}{8(K^2+K+\frac{K^3}{N})}\right),\end{aligned} \tag{62}$$

where δ is as same as in Theorem. 2.

Note that the results so far hold for a fixed choice of column w. Since the bound does not depend on the identity of w, we can apply the union bound over all N possible columns w. So the probability that there exists a w such that

$$\|y(\pi)\|^2 \leq (1-\delta)\Upsilon^2$$

is at most $1 - N\exp\left(-\frac{3(1-\delta-\frac{K}{M})^2M^2}{8(K^2+K+\frac{K^3}{N})}\right) = 1-\epsilon$. If (13) holds, $\epsilon \sim \mathcal{O}\left(\exp\left(-(\log N)^2+\log N\right)\right)$. Thus, for small δ and sufficiently large N, $\|y(\pi)\|^2$ has a very high probability of being smaller than $(1-\delta)\Upsilon^2$; specifically, the probability is $1-\epsilon$.

The following proof uses an argument in [10]. The procedure of the last step is similar to the proof of UStRIP in [12]. As said before, τ is a set of K indices sampled randomly from $\{1,\ldots,N\}$. Then let S be any other subset of $\{1,\ \ldots,$ N$\}$ of size less than or equal to K, and $\mathbf{P}_\tau$ be the orthogonal projection operator on the range $\mathcal{R}_\tau$. With probability at least $1-\epsilon$,

$$\begin{aligned}\|\mathbf{P}_\tau\phi_S\|^2 &= (\phi_\tau^\dagger\phi_S)^\dagger(\phi_\tau^\dagger\phi_\tau)^{-1}(\phi_\tau^\dagger\phi_S) \\ &\leq \frac{\|\phi_\tau^\dagger\phi_S\|^2}{(\sigma_{\min}(\phi_\tau))^2} \\ &\leq \frac{\|y\|^2}{1-\delta} \leq \Upsilon^2,\end{aligned} \tag{63}$$

where $\sigma_{\min}(\phi_\tau)$ denotes the smallest singular value of ϕ_τ.

The result implies that there exists a vector in range of ϕ_S that is outside the range of ϕ_τ. Precisely, $\dim(\mathrm{range}(\phi_\tau) \cap \mathrm{range}(\phi_S)) < K$. Due to the structure of OSTM, ϕ_τ is nonsingular with probability exceeding $1-\epsilon$, so that

$$\dim(\mathrm{range}(\phi_\tau)) = K$$

with probability exceeding $1-\epsilon$. Consider the set of signals in $\mathrm{range}(\phi_\tau)$ that can be represented using a different set of K columns. This set can be written as a finite union of subspaces with dimension strictly less than K. Therefore, it has zero volume with respect to any nonatomic measure. In other words, it is almost sure that the signal has no other representation as a linear combination of K columns from Φ [10]. Since the τ in ϕ_τ are randomly chosen which satisfy near-isometry with probability exceeding $1-\epsilon$, no two signals with support τ have the same values in the measurement space. Thus, the theorem is proved.

■

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, Jul. 2006.

[2] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, pp. 489–509, Feb. 2006.

[3] E. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5406–5425, 2006.

[4] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure and Applied Math.*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[5] T. T. Do, T. D. Tran, and L. Gan, "Fast compressive sampling with structually random matrices," in *Proc. ICASSP*, Mar 2008.

[6] W. U. Bajwa, J. D. Haupt, G. M. Raz, S. J. Wright, and R. D. Nowak, "Toeplitz-structured compressed sensing matrices," in *Statistical Signal Processing, 2007. SSP '07. IEEE/SP 14th Workshop on*, Aug. 2007, pp. 294 –298.

[7] R. A. Devore, "Deterministic construction of compressed sensing matrices," *Journal of Complexity*, vol. 23, pp. 918–925, 2007.

[8] M. A. Iwen, "Simple deterministically constructible RIP matrices with sublinear Fourier sampling requirements," in *Proc. CISS*, 2008.

[9] V. Saligrama. Deterministic designs with deterministic guarantees: Toeplitz compressed sensing matrices, sequence design and system identification. [Online]. Available: http://arxiv.org/abs/0806.4958v2.pdf

[10] J. A. Tropp, "On the conditioning of random subdictionaries," *Appl. Comput. Harmonic Anal.*, vol. 25, pp. 1–24, 2008.

[11] S. Guverich and R. Hadani, "The statistical restricted isometry property and the Wigner semicircle distribution of incoherent dictionaries," *Annals of Applied Probability*, 2009, submitted.

[12] R. Calderbank, S. Howard, and S. Jafarpour, "Construction of a large class of deterministic sensing matrices that satisfy a statistical isometry property," *IEEE J. Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 358 –374, April 2010.

[13] K. Li, C. Ling, and L. Gan, "Statistical restricted isometry property of orthogonal symmetric Toeplitz matrices," in *IEEE Information Theory Workshop, 2009*, Oct. 2009, pp. 183–187.

[14] S. Chatterjee, "Stein's method for concentration inequalities," *Probability Theory Related Fields*, vol. 138, pp. 305–321, 2007.

[15] M. Golay, "Complementary series," *IEEE Trans. Inf. Theory*, vol. 7, no. 2, pp. 82 –87, April 1961.

[16] W. Yin, S. P. Morgan, J. Yang, and Y. Zhang, "Practical compressive sensing with Toeplitz and circulant matrices," Rice University CAAM Technical Report TR10-01. Submitted to VCIP 2010.

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

[18] W. L. Chan, K. Charan, D. Takhar, K. F. Kelly, R. G. Baraniuk, and D. M. Mittleman, "A single-pixel terahertz imaging system based on compressive sensing," *Appl. Phys. Lett.*, vol. 93, p. 121105, 2008.

[19] Y. C. Shen, L. Gan, and M. Stringer et al., "Terahertz pulsed spectroscopic imaging using optimized binary masks," *Appl. Phys. Lett.*, vol. 95, p. 231112, 2009.

[20] J. Romberg, "Compressive sensing by random convolution," *SIAM J. Imaging Sciences*, vol. 2, pp. 1098–1128, 2009.

[21] L. Gan, T. Do, and T. D. Tran, "Fast compressive imaging using scrambled block hadamard ensemble," in *European Signal Processing Conf. (EUSIPCO), Lausanne, Switzerland*, Aug. 2008.

[22] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4203–4215, 2005.

[23] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Communications on Pure and Applied Mathematics*, vol. 61, pp. 1025–1045, 2008.

[24] A. Bottcher, "Orthogonal symmetric toeplitz matrices," *Complex Analysis and Operator Theory*, vol. 2, no. 2, pp. 285–298, 2008.

[25] B. Popovic, "Synthesis of power efficient multitone signals with flat amplitude spectrum," *IEEE Trans. Commun.*, vol. 39, pp. 1031–1033, 1991.

[26] J. Davis and J. Jedwab, "Peak-to-mean power control in OFDM, Golay complementary sequences, and Reed-Muller codes," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2397–2417, Nov. 1999.

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

[28] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing: Closing the gap between performance and complexity," *IEEE Trans. Inf. Theory*, vol. 55, pp. 2230–2249, May 2009.

[29] T. Do, L. Gan, N. Nguyen, and T. Tran, "Sparsity adaptive matching pursuit algorithm for practical compressed sensing," Oct. 2008, pp. 581 –587.