Fast and Efficient Compressive Sensing using Structurally Random Matrices

Thong T. Do, *Student Member, IEEE,* Lu Gan, *Member, IEEE,* Nam Nguyen and Trac D. Tran, *Member, IEEE*

This work has been supported in part by the National Science Foundation under Grant CCF-0728893.

Thong T. Do, Nam Nguyen and Trac D. Tran are with the Johns Hopkins University, Baltimore, MD, 21218 USA.

Lu Gan is with the Brunel University, London, UK.

Abstract

This paper introduces a fast and efficient framework for practical compressive sensing. Our framework is mainly based on a novel design called *Structurally Random Matrix* (SRM). It is highly promising for large-scale, real-time compressive sensing applications because it can be realized as a product of simple and fast operators and thus, there is no need for storing the sensing matrix explicitly. The introduced framework is flexible and provides relevant features such as universality, block-based processing and hardware friendliness to analog and optical domain implementation. Despite all of these practical advantages, the framework can be shown to approach optimal performance, i.e. the number of measurements for exact signal reconstruction is at the minimum bound. Simulation results with several interesting SRM under various practical settings are also presented to verify the validity of the theory as well as to illustrate the promising potentials of the proposed framework.

Index Terms

compressed sensing, compressive sensing, random projection, sparse reconstruction, fast and efficient algorithm

I. Introduction

Compressed sensing (CS) [1], [2] has attracted a lot of interests over the past few years as a revolutionary signal sampling paradigm. Suppose that $\boldsymbol{x}$ is a length-N signal. It is said to be K-sparse (or compressible) if $\boldsymbol{x}$ can be well approximated using only $K \ll N$ coefficients under some linear transform:

$$\boldsymbol{x} = \boldsymbol{\Psi\alpha},$$

where $\boldsymbol{\Psi}$ is the sparsifying basis and $\boldsymbol{\alpha}$ is the transform coefficient vector that has K (significant) nonzero entries.

According to the CS theory, such a signal can be acquired through the following random linear projection:

$$\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{e},$$

where $\boldsymbol{y}$ is the sampled vector with $M \ll N$ data points, $\boldsymbol{\Phi}$ represents a $M \times N$ random matrix and $\boldsymbol{e}$ is the acquisition noise. The CS framework is attractive as it implies that $\boldsymbol{x}$ can be faithfully recovered from only $M = \mathcal{O}(K \log N)$ measurements, suggesting the potential of significant cost reduction in digital data acquisition.

While the sampling process is simply a random linear projection, the reconstruction to find the sparsest signal from the received measurements is highly non-linear process. More precisely, the reconstruction algorithm is to solve the l_1-minimization of a transform coefficient vector:

$$\min \|\boldsymbol{\alpha}\|_1 \quad \text{s.t.} \quad \boldsymbol{y} = \boldsymbol{\Phi\Psi\alpha}.$$

Linear programming [1], [2] and other convex optimization algorithms [3], [4], [5] have been proposed to solve the l_1 minimization. Furthermore, there also exists a family of greedy pursuit algorithms [6], [7], [8], [9], [10] offering another promising option for sparse reconstruction. These algorithms all need to compute $\boldsymbol{\Phi\Psi}$ and $(\boldsymbol{\Phi\Psi})^T$ multiple times. Thus, computational complexity of the system depends on the structure of sensing matrix $\boldsymbol{\Phi}$ and its transpose $\boldsymbol{\Phi}^T$.

Preferably, the sensing matrix $\mathbf{\Phi}$ should be highly incoherent with sparsifying basis $\mathbf{\Psi}$, i.e. rows of $\mathbf{\Phi}$ do not have any sparse representation in the basis $\mathbf{\Psi}$. Incoherence between two matrices is mathematically quantified by the mutual coherence coefficient [11].

Definition I.1. The mutual coherence of an orthonormal matrix $N \times N$ $\mathbf{\Phi}$ and another orthonormal matrix $N \times N$ $\mathbf{\Psi}$ is defined as:

$$\mu(\mathbf{\Phi}, \mathbf{\Psi}) = \max_{1 \leq i,j \leq N} |\langle \Phi_i, \Psi_j \rangle|$$

where $\mathbf{\Phi}_i$ are rows of $\mathbf{\Phi}$ and $\mathbf{\Psi}_j$ are columns of $\mathbf{\Psi}$, respectively.

If $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are two orthonormal matrices, $\|\mathbf{\Phi}\mathbf{\Psi}_j\|_2 = \|\mathbf{\Psi}_j\|_2 = 1$. Thus, it is easy to see that for two orthonormal matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}$, $1/\sqrt{N} \leq \mu \leq 1$. Incoherence implies that the mutual coherence or the maximum magnitude of entries of the product matrix $\mathbf{\Phi}\mathbf{\Psi}$ is relatively small. Two matrices are completely incoherent if their mutual coherence coefficient approaches the lower bound value of $1/\sqrt{N}$.

A popular family of sensing matrices is a random projection or a random matrix of i.i.d random variables from a sub-Gaussian distribution such as Gaussian or Bernoulli [12], [13]. This family of sensing matrix is well-known as it is universally incoherent with all other sparsifying basis. For example, if $\mathbf{\Phi}$ is a random matrix of Gaussian i.i.d entries and $\mathbf{\Psi}$ is an arbitrary orthonormal sparsifying basis, the sensing matrix in the transform domain $\mathbf{\Phi}\mathbf{\Psi}$ is also Gaussian i.i.d matrix. The universal property of a sensing matrix is important because it enables us to sense a signal directly in its original domain without significant loss of sensing efficiency and without any other prior knowledge. In addition, it can be shown that random projection approaches the optimal sensing performance of $M = \mathcal{O}(K \log N)$.

However, it is quite costly to realize random matrices in practical sensing applications as they require very high computational complexity and huge memory buffering due to their completely unstructured nature [14]. For example, to process a 512×512 image with $64K$ measurements (i.e., 25% of the original sampling rate), a Bernoulli random matrix requires nearly gigabytes storage and giga-flop operations, which makes both the sampling and recovery processes very expensive and in many cases, unrealistic.

Another class of sensing matrices is a uniformly random subset of rows of an orthonormal matrix in which the partial Fourier matrix (or the partial FFT) is a special case [13], [14]. While the partial FFT is well known for having fast and efficient implementation, it only works well in the transform domain or in the case that the sparsifying basis is the identity matrix. More specifically, it is shown in [[14], *Theorem* 1.1] that the minimal number of measurements required for exact recovery depends on the incoherence of $\mathbf{\Phi}$ and $\mathbf{\Psi}$:

$$M = \mathcal{O}(\mu_n^2 K \log N) \tag{1}$$

where μ_n is the normalized mutual coherence: $\mu_n = \sqrt{N}\mu$ and $1 \leq \mu_n \leq \sqrt{N}$. With many well-known sparsifying basis such as wavelets, this mutual coherence coefficient might be large and thus, resulting in performance loss. Another approach is to design a sensing matrix to be incoherent with a given sparsifying basis. For example, *Noiselets* is designed to be incoherent with the Haar wavelet basis in [15], i.e. $\mu_n = 1$ when $\mathbf{\Phi}$ is Noiselets transform and $\mathbf{\Psi}$

is the Haar wavelet basis. Noiselets also has low-complexity implementation $\mathcal{O}(N \log N)$ although it is unknown if noiselets is also incoherent with other bases.

II. Compressive Sensing with Structurally Random Matrices

A. Overview

One of remaining challenges for CS in practice is to design a CS framework that has the following features:

- *Optimal or near optimal sensing performance*: the number of measurements for exact recovery is almost minimal, i.e. on the order of $\mathcal{O}(K \log N)$;
- *Universality*: sensing performance is equally good with all sparsifying bases;
- *Low complexity, fast implementation that can support block-based processing*: this is necessary for large-scale, realtime sensing applications;
- *Easy and cheap to implement in hardware and optics domain*: Preferably, entries of the sensing matrix should only take values in the set $\{0, 1, -1\}$.

In this paper, we propose a framework that aims to satisfy the above wish-list. Lying at the heart of our framework is the concept of *Structurally Random Matrix*(SRM) that is defined as a product of three matrices:

$$\mathbf{\Phi} = \sqrt{\frac{N}{M}} \boldsymbol{DFR} \tag{2}$$

where:

- $\boldsymbol{R} \in N \times N$ is either a uniform random permutation matrix or a diagonal random matrix whose diagonal entries R_{ii} are i.i.d Bernoulli random variables with identical distribution $P(R_{ii} = \pm 1) = 1/2$. A uniformly random permutation matrix scrambles signal's sample locations globally while a diagonal matrix of Bernoulli random variables flips signal's sample signs locally. Hence, we often refer the former as the *global randomizer* and the latter as the *local randomizer*.
- $\boldsymbol{F} \in N \times N$ is an orthonormal matrix that,in practice, is selected to be fast computable such as popular fast transforms: FFT, DCT, WHT or their block diagonal versions. The purpose of the matrix $\boldsymbol{F}$ is to spread *information* (or energy) of the signal's samples over all measurements
- $\boldsymbol{D} \in M \times N$ is a subsampling matrix/operator. The operator $\boldsymbol{D}$ selects a random subset of rows of the matrix $\boldsymbol{FR}$. If the probability of selecting a row P(a row is selected) is M/N, the number of rows selected would be M in average. In matrix representation, $\boldsymbol{D}$ is simply a random subset of M rows of the identity matrix of size $N \times N$. The scale coefficient $\sqrt{\frac{N}{M}}$ is to normalize the transform so that energy of the measurement vector is almost similar to that of the input signal vector.

The proposed sensing algorithm can be described step by step as follows:

- Step 1 (Signal pre-randomization): Randomize a target signal by either flipping its sample signs or uniformly permuting its sample locations. This step corresponds to multiplying the signal with the matrix $\boldsymbol{D}$
- Step 2 (Signal transform): Apply a fast transform $\boldsymbol{F}$ to the randomized signal

- Step 3 (Signal subsampling): randomly pick up M measurements out of N transform coefficients. This step corresponds to multiplying the transform coefficients with the matrix $\boldsymbol{D}$

Conventional CS reconstruction algorithm is employed to recover the transform coefficient vector $\boldsymbol{\alpha}$ by solving the l_1 minimization:

$$\widehat{\boldsymbol{\alpha}} = \text{argmin}\|\boldsymbol{\alpha}\|_1 \quad s.t. \quad \boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\alpha}. \tag{3}$$

Finally, the signal is recovered as $\widehat{\boldsymbol{x}} = \boldsymbol{\Psi}\widehat{\boldsymbol{\alpha}}$. The framework can achieve perfect reconstruction if $\widehat{\boldsymbol{x}} = \boldsymbol{x}$.

From the best of our knowledge, the proposed sensing algorithm is distinct from currently existing methods such as random projection [16], random filters [17], structured Toeplitz [18], random convolution [19] via the step of pre-randomization. The main idea of this step is to deliberately scramble the structure of the signal, converting the signal to be sampled into a white noise-like one. Detail analysis in the following section will show that pre-randomization is necessary for obtaining universally incoherent sensing. The intuition behind this pre-randomization strategy is that scrambling a signal into a white noise-like form enables the sensing process to be independent of the signal's sparsifying basis.

The remaining of the paper is organized as follows. We first discuss about incoherence between SRMs and sparsifying transforms in Section III. More specifically, Section III-A will give us a rough intuition of why SRM could work as well as a random Gaussian matrix. Detail quantitative analysis of the incoherence for SRM with local randomizer and global randomizer is presented in Section III-B and Section III-C, respectively. Based on these incoherence results, theoretical performance of the proposed framework is analyzed in Section IV and then followed by experiment validation in Section V. Finally, Section VI concludes the paper with detail discussion of practical advantages of the proposed framework and relationship between the proposed framework and other related works.

B. Notations

We reserve a bold letter for a vector, a capital and bold letter for a matrix, a capital and bold letter with one sub-index for a row or a column of a matrix and a capital letter with two sub-indices for an entry of a matrix. We often employ $\boldsymbol{x} \in \mathbb{R}^N$ for the input signal, $\boldsymbol{y} \in \mathbb{R}^N$ for the measurement vector, $\boldsymbol{\Phi} \in \mathbb{R}^{M\times N}$ for the sensing matrix, $\boldsymbol{\Psi} \in \mathbb{R}^{N\times N}$ for the sparsifying matrix and $\boldsymbol{\alpha} \in \mathbb{R}^N$ for the transform coefficient vector ($\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{\alpha}$). We use the notation supp($\boldsymbol{z}$) to indicate the index set (or coordinate set) of nonzero entries of the vector $\boldsymbol{z}$. Occasionally, we also use $\mathcal{T}$ to alternatively refer to this index set of nonzero entries (i.e., $\mathcal{T}$=supp($\boldsymbol{z}$)). In this case, $\boldsymbol{z}_{\mathcal{T}}$ denotes the portion of vector $\boldsymbol{z}$ indexed by the set $\mathcal{T}$ and $\boldsymbol{A}_{\mathcal{T}}$ denotes the submatrix of $\boldsymbol{A}$ whose columns are indexed by the set $\mathcal{T}$.

Let S_{ij}, F_{ij} be the entry at the i^{th} row and j^{th} column of $\boldsymbol{A}\boldsymbol{\Psi}$ and $\boldsymbol{F}$, R_{kk} be the k^{th} entry on the diagonal of the diagonal matrix R, $\boldsymbol{A}_i$ and $\boldsymbol{\Psi}_j$ be the i^{th} row of $\boldsymbol{A}$ and j^{th} column of $\boldsymbol{\Psi}$, respectively.

In addition, we also employ the following notations:

- x_n is on the order of $o(z_n)$, denoted as $x_n = o(z_n)$, if

$$\lim_{n\to\infty} \frac{x_n}{z_n} = 0.$$

- x_n is on the order of $\mathcal{O}(z_n)$, denoted as $x_n = \mathcal{O}(z_n)$, if

$$\lim_{n\to\infty} \frac{x_n}{z_n} = c.$$

  where c is some positive constant.

- A random variable X_n is called asymptotically normally distributed $\mathcal{N}(0, \sigma^2)$, if

$$\lim_{n\to\infty} P(\frac{X_n}{\sigma} \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-y^2}{2}} dy.$$

III. Incoherence Analysis

A. Asymptotical Distribution Analysis

If $\mathbf{\Phi}$ is an i.i.d Gaussian matrix $\mathcal{N}(0, \frac{1}{N})$ and $\mathbf{\Psi}$ is an arbitrarily orthonormal matrix, $\mathbf{\Phi\Psi}$ is also i.i.d Gaussian matrix $\mathcal{N}(0, \frac{1}{N})$, implying that with overwhelming probability, a Gaussian matrix is highly incoherent with all orthonormal $\mathbf{\Psi}$. In other words, i.i.d. Gaussian matrices are universally incoherent with fixed transforms (with overwhelming probability). In this section, we will argue that under some mild conditions, with $\mathbf{\Phi} = \boldsymbol{DFR}$, where $\boldsymbol{D}, \boldsymbol{F}, \boldsymbol{R}$ are defined as in the previous section, entries of $\mathbf{\Phi\Psi}$ are asymptotically normally distributed $\mathcal{N}(0, \sigma^2)$, where $\sigma^2 \leq \mathcal{O}(\frac{1}{N})$. This claim is illustrated in Fig. 1, which depicts the quantile-quantile (QQ) plots of entries of $\mathbf{\Phi\Psi}$, where $N = 256$, $\boldsymbol{F}$ is the 256×256 DCT matrix and $\mathbf{\Psi}$ is the Daubechies-8 orthogonal wavelet basis. Fig. 1(a) and Fig. 1(b) correspond to the case $\boldsymbol{R}$ is the local and global randomizer, respectively. In both cases, the QQ-plots appear straight, as the Gaussian model demands.

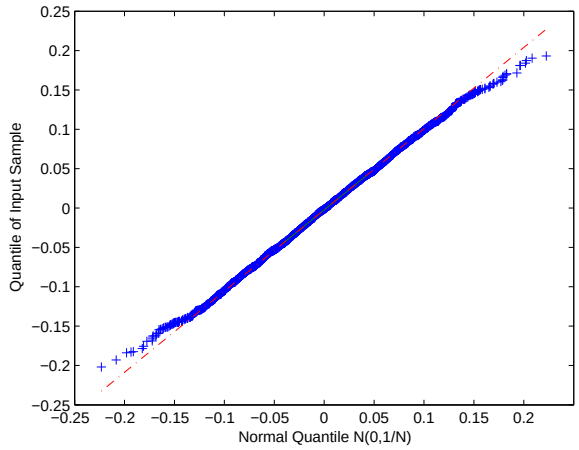
Note that $\mathbf{\Phi}$ is a submatrix of $\boldsymbol{A} = \boldsymbol{FR}$. Thus, asymptotical distribution of the entries of $\boldsymbol{A\Psi}$ is similar to that of entries of $\mathbf{\Phi\Psi}$.

Theorem III.1. *Let $\boldsymbol{A} = \boldsymbol{FR}$, where $\boldsymbol{R}$ is an $N \times N$ random diagonal matrix of i.i.d Bernoulli random variables along its diagonal $P(R_{ii} = \pm 1) = 1/2$. Let $\boldsymbol{F}$ be an $N \times N$ unit-norm row matrix with absolute magnitude of all entries on the order of $\mathcal{O}(\frac{1}{\sqrt{N}})$. Let $\mathbf{\Psi}$ be an $N \times N$ unit-norm column matrix with the maximal absolute magnitude of entries on the order of $o(1)$. Then, entries of $\boldsymbol{A\Psi}$ are asymptotically normally distributed $\mathcal{N}(0, \sigma^2)$ with $\sigma^2 \leq \mathcal{O}(\frac{1}{N})$.*

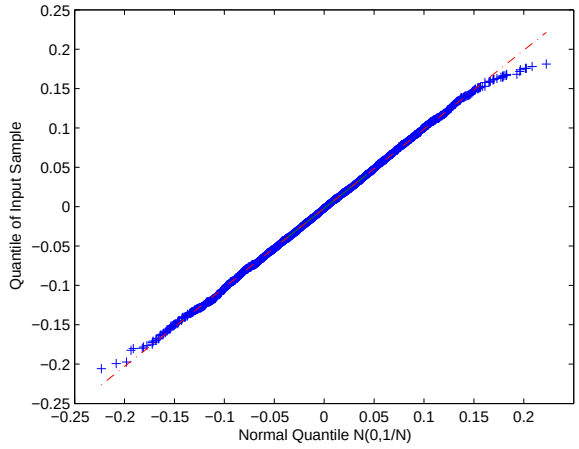
Proof. With notations being defined in Section II-B, we have:

$$S_{ij} = \langle \boldsymbol{A}_i, \mathbf{\Psi}_j \rangle = \sum_{k=1}^{N} F_{ik} \Psi_{kj} R_{kk} \tag{4}$$

Denote $Z_k = F_{ik}\Psi_{kj}R_{kk}$. Because R_{kk} are i.i.d Bernoulli random variables, Z_k are i.i.d zero-mean random variables with $E(Z_k) = 0$. The assumption that $|F_{ik}|$ are on the order of $\mathcal{O}(\frac{1}{\sqrt{N}})$ implies that there exist two positive constants c_1 and c_2 such that:



(a)



(b)

Fig. 1. QQ plots comparing distribution of entries of $\mathbf{\Phi\Psi}$ and Gaussian distribution. (a) $\boldsymbol{R}$ is the local randomizer. (b) $\boldsymbol{R}$ is the global randomizer. The plots all appear nearly linear, indicating that entries of $\mathbf{\Phi\Psi}$ are nearly Normal distributed

$$\frac{c_1}{N}\Psi_{kj}^2 \leq \text{Var}(Z_k) = F_{ik}^2\Psi_{kj}^2 \leq \frac{c_2}{N}\Psi_{kj}^2. \tag{5}$$

The variance of S_{ij}, σ^2, can be bounded as the follows:

$$\frac{c_1}{N} = \frac{c_1}{N}\sum_{k=1}^{N}\Psi_{kj}^2 \leq \sigma^2 = \sum_{k=1}^{N}\text{Var}(Z_k) \leq \frac{c_2}{N}\sum_{k=1}^{N}\Psi_{kj}^2 = \frac{c_2}{N}. \tag{6}$$

Because S_{ij} is a sum of i.i.d zero-mean random variables $\{Z_k\}_{k=1}^N$, according to the Central Limit Theorem (CLT)(see Appendix I), $S_{ij} \rightarrow \mathcal{N}(0, \mathcal{O}(\frac{1}{N}))$. To apply CLT, we need to verify its convergence condition: for a given $\epsilon > 0$ and there exists N that is sufficiently large such that the $\text{Var}(Z_k)$ satisfy:

$$\text{Var}(Z_k) < \epsilon\sigma^2, k = 1, 2, ..., N. \tag{7}$$

To show that this convergence condition is met, we use the counterproof method. Assume there exists ϵ_0 such that $\forall N$, there exists at least $k_0 \in \{1, 2, \ldots, N\}$:

$$\text{Var}(Z_{k_0}) > \epsilon_0 \sigma^2. \tag{8}$$

From (5), (6) and (8), we achieve:

$$\epsilon \frac{c_1}{N} \leq \text{Var}(Z_{k_0}) \leq \frac{c_2}{N} \Psi_{kj}^2. \tag{9}$$

This inequality can be true if all Ψ_{kj} are in the order of $o(1)$. The underlying intuition of the convergence condition is to guarantee that there is no random variable with dominant variance in the sum S_{ij}. In this case, it simply requires that there is no dominant entry on each column of $\boldsymbol{\Psi}$. □

Given a few more restrictions, we can establish a similar result when $\boldsymbol{R}$ is a uniformly random permutation matrix,

Theorem III.2. *Let $\boldsymbol{A} = \boldsymbol{F}\boldsymbol{R}$, where $\boldsymbol{R}$ is an $N \times N$ uniformly random permutation matrix. Let $\boldsymbol{F}$ be an $N \times N$ unit-norm row matrix with the maximal absolute magnitude of entries on the order of $\mathcal{O}(\frac{1}{\sqrt{N}})$. Let $\boldsymbol{\Psi}$ be an $N \times N$ unit-norm column matrix with the maximal absolute magnitude of entries on the order of $o(1)$ and the average sum of entries on each column is on the order of $o(\frac{1}{\sqrt{N}})$. Assume that sum of entries on each row of $\boldsymbol{F}$ is zero. Also, assume that entries on each row of $\boldsymbol{F}$ and on each column of $\boldsymbol{\Psi}$ are not all equal. Then, entries of $\boldsymbol{A}\boldsymbol{\Psi}$ are asymptotically normally distributed $\mathcal{N}(0, \sigma^2)$, where $\sigma^2 \leq \mathcal{O}(\frac{1}{N})$.*

Proof. Let $[\omega_1, \omega_2, ..., \omega_N]$ be a uniform random permutation of $[1, 2, ..., N]$. Note that $\{\omega_k\}_{k=1}^N$ can be viewed as a sequence of random variables with identical distribution. In particular, for a fixed k:

$$P(\omega_k = i) = \frac{1}{N}, i = 1, 2, ..., N.$$

Denote $Z_k = F_{i\omega_k} \Psi_{kj}$ (we omit the dependence of Z_k on i and j to simplify the notation), we have:

$$S_{ij} = \langle \boldsymbol{A}_i, \boldsymbol{\Psi}_j \rangle = \sum_{k=1}^{N} F_{i\omega_k} \Psi_{kj} = \sum_{k=1}^{N} Z_k.$$

Using the assumption that the vector $\boldsymbol{F}_i$ has zero average sum and unit norm, we derive:

$$E(Z_k) = \Psi_{kj} E(F_{i\omega_k}) = \frac{\Psi_{kj}}{N} \sum_{j=1}^{N} F_{ij} = 0.$$

and also,

$$E(Z_k^2) = \Psi_{kj}^2 E(F_{i\omega_k}^2) = \frac{\Psi_{kj}^2}{N} \sum_{j=1}^{N} F_{ij}^2 = \frac{\Psi_{kj}^2}{N}.$$

In addition, note that although $\{\omega_k\}_{k=1}^N$ have the identical distribution, they are correlated random variables because of the uniformly random permutation *without* replacement. Thus, with a pair of k and l such that $1 \leq k \neq$

$l \leq N$, we have:

$$\begin{aligned} E(Z_k Z_l) &= \Psi_{kj}\Psi_{lj} E(F_{i\omega_k} F_{i\omega_l}) \\ &= \frac{\Psi_{kj}\Psi_{lj}}{N(N-1)} \sum_{1\leq p\neq q\leq N} F_{ip}F_{iq} \\ &= \frac{\Psi_{kj}\Psi_{lj}}{N(N-1)} ((\sum_{p=1}^{N} F_{ip})^2 - \sum_{p=1}^{N} F_{ip}^2) \\ &= -\frac{\Psi_{kj}\Psi_{lj}}{N(N-1)}. \end{aligned}$$

The last equation holds because the vector $\boldsymbol{F}_i$ has zero average sum and unit-norm. Then, we derive the expectation and variance of S_{ij} as follows:

$$E(S_{ij}) = 0;$$

$$\begin{aligned} \text{Var}(S_{ij}) &= \sum_{k=1}^{N} E(Z_k^2) + \sum_{1\leq k\neq q\leq N} E(Z_k Z_l) \\ &= \frac{1}{N}\sum_{k=1}^{N} \Psi_{kj}^2 - \frac{1}{N(N-1)} \sum_{1\leq k\neq l\leq N} \Psi_{kj}\Psi_{lj} \\ &= \frac{1}{N} - \frac{1}{N(N-1)} ((\sum_{k=1}^{N} \Psi_{kj})^2 - \sum_{k=1}^{N} \Psi_{kj}^2) \\ &= \frac{1}{N} - \frac{1}{N(N-1)} ((\sum_{k=1}^{N} \Psi_{kj})^2 - 1) \\ &\leq \frac{1}{N} + \frac{1}{N(N-1)} = \mathcal{O}(\frac{1}{N}). \end{aligned}$$

The forth equations holds because the column $\boldsymbol{\Psi}_j$ has unit-norm. The theorem is then a simple corollary of the Combinatorial Central Limit Theorem [20] (see Appendix 1), provided that its convergence condition can be verified that is:

$$\lim_{N\to\infty} N \frac{\max_{1\leq k\leq N}(F_{ik} - \overline{F_i})^2}{\sum_{k=1}^{N}(F_{ik} - \overline{F_i})^2} \frac{\max_{1\leq k\leq N}(\Psi_{kj} - \overline{\Psi_j})^2}{\sum_{k=1}^{N}(\Psi_{kj} - \overline{\Psi_j})^2} = 0, \tag{10}$$

where

$$\overline{F_i} = \frac{1}{N}\sum_{k=1}^{N} F_{ik}; \quad \overline{\Psi_j} = \frac{1}{N}\sum_{k=1}^{N} \Psi_{kj}.$$

Because $\overline{F_i} = 0$, $\|F_i\|_2^2 = 1$ and $\max_{1\leq k\leq N} F_{ik}^2 = \mathcal{O}(\frac{1}{N})$, the equation (10) holds if the following equation holds:

$$\lim_{N\to\infty} \frac{\max_{1\leq k\leq N}(\Psi_{jk} - \overline{\Psi_j})^2}{\sum_{k=1}^{N}(\Psi_{jk} - \overline{\Psi_j})^2} = 0. \tag{11}$$

Because $\{|\overline{\Psi_j}|\}_{j=1}^{N}$ are on the order of $o(\frac{1}{\sqrt{N}})$:

$$\sum_{k=1}^{N}(\Psi_{kj} - \overline{\Psi_j})^2 = \|b\|_2^2 - N\overline{\Psi_j}^2 = 1 - N\overline{\Psi_j}^2 = \mathcal{O}(1). \tag{12}$$

Also, due to $|\overline{\Psi_j}| \leq \max_{1\leq k\leq N} |\Psi_{jk}|$ and $|\Psi_{jk}|$ are on the order of $o(1)$:

$$\max_{1\leq k\leq N}(\Psi_{jk} - \overline{\Psi_j})^2 \leq 4 \max_{1\leq k\leq N} \Psi_{jk}^2 = o(1). \tag{13}$$

Combination of (12) and (13) implies (11) and thus the convergence condition of the Combinatorial Central Limit Theorem is verified. □

The condition that each row of $\boldsymbol{F}$ has zero average sum is to guarantee that entries of $\boldsymbol{F\Psi}$ have zero mean while the condition that entries on each row of $\boldsymbol{F}$ and on each column of $\boldsymbol{\Psi}$ are not all equal is to prevent the degenerate case that entries of $\boldsymbol{F\Psi}$ might become a deterministic quantity. For example, when entries of a row $\boldsymbol{F}_i$ are all equal $\frac{1}{\sqrt{N}}$, $S_{ij} = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \Psi_{kj}$, which is a deterministic quantity, not a random variable. Note that these conditions are not necessary when $\boldsymbol{R}$ is a diagonal matrix of Bernoulli random entries.

If $\boldsymbol{F}$ is a DCT matrix, a (normalized) WHT matrix or a (normalized) DFT matrix, all the rows (except for the first one) have zero average sum due to the symmetry in these matrices. The first row, whose entries are all equal $\frac{1}{\sqrt{N}}$, can be considered as the averaging row, or a lowpass filtering operation. When the input signal is zero-mean, this row might be chosen or not without affecting quality of the reconstructed signal. Otherwise, it should be included in the chose row set to encode the signal's mean. Lastly, the condition that absolute average sum of every column of the sparsifying basis $\boldsymbol{\Psi}$ are on the order of $o(\frac{1}{\sqrt{N}})$ is also close to the reality because the majority of columns of the sparsifying basis $\boldsymbol{\Psi}$ can be roughly viewed as bandpass and highpass filters whose average sum of the coefficients are always zero. For example, if $\boldsymbol{\Psi}$ is a wavelet basis (with at least one vanishing moment), then all columns of $\boldsymbol{\Psi}$ (except one at DC) has column sum of zero.

The aforementioned theorems show that under certain conditions, the majority of entries of $\boldsymbol{A\Psi}$ (also $\boldsymbol{\Phi\Psi}$) behave like Gaussian random variables $\mathcal{N}(0, \sigma^2)$, where $\sigma^2 \leq \mathcal{O}(\frac{1}{N})$. Roughly speaking, this behavior constitutes to a good sensing performance for the proposed framework. However, these asymptotic results are not sufficient for an explicit measurement of sensing performance because in general, entries of $\boldsymbol{A\Psi}$ are not stochastically independent, violating a condition of a Gaussian i.i.d matrix. In fact, the sensing performance might be quantitatively analyzed by employing a powerful analysis framework of a random subset of rows of an orthonormal matrix [14]. Note that $\boldsymbol{A}$ is also an orthonormal matrix when $\boldsymbol{R}$ is either a random permutation matrix or a diagonal matrix of Bernoulli random entries.

Based on the Gaussian tail probability and a union bound for a supreme (i.e., maximum absolute value) of a random sequence, the maximum absolute magnitude of $\boldsymbol{A\Psi}$ can be asymptotically bounded as follows:

$$P(\max_{1\leq i,j\leq N} |S_{ij}| \geq t) \preceq 2N^2 \exp(-\frac{t^2}{2\sigma^2})$$

where $\sigma^2 \leq \frac{c}{N}$ and c is some positive constant and $\preceq$ stands for "asymptotically smaller or equal", i.e., when N goes to infinity, $\preceq$ becomes $\leq$.

If we choose $t = \sqrt{\frac{2c \log 2(N/\delta)^2}{N}}$, the above inequality is equivalent to:

$$P(\max_{1\leq i,j\leq N} |S_{ij}| \leq \sqrt{\frac{c \log 2(N/\delta)^2}{N}}) \succeq 1 - \delta$$

which implies that with probability at least $1 - \delta$, the mutual coherence of $\boldsymbol{A}$ and $\boldsymbol{\Psi}$ is upper bounded by $\mathcal{O}(\sqrt{\frac{\log(N/\delta)}{N}})$, which is close to the optimal value, except the $\log N$ factor.

In the following section, we will employ a more powerful tool from the theory of concentration inequalities to analyze the coherence between $\boldsymbol{A}$ and $\boldsymbol{\Psi}$ when N is finite.

B. Incoherence Analysis With The Local Randomizer

The first theorem is about the mutual coherence of $\boldsymbol{A} = \boldsymbol{FR}$ and $\boldsymbol{\Psi}$ when $\boldsymbol{R}$ is a diagonal matrix of i.i.d Bernoulli random variables.

Theorem III.3. *Let $\boldsymbol{A} = \boldsymbol{FR}$, where $\boldsymbol{R}$ is an $N \times N$ random diagonal matrix of i.i.d Bernoulli random variables along its diagonal $P(R_{ii} = \pm 1) = 1/2$. Let $\boldsymbol{F}$ be an $N \times N$ unit-norm row matrix with the maximal absolute magnitude of entries on the order of $\mathcal{O}(\frac{1}{\sqrt{B}})$, where $1 \leq B \leq N$. Let $\boldsymbol{\Psi}$ be an $N \times N$ unit-norm column matrix.*

- *With probability at least $1 - \delta$, the mutual coherence of $\boldsymbol{A}$ and $\boldsymbol{\Psi}$ is upper bounded by $\mathcal{O}(\sqrt{\frac{\log(N/\delta)}{B}})$.*
- *In addition, if the maximal absolute magnitude of entries of $\boldsymbol{\Psi}$ is on the order of $\mathcal{O}(\frac{1}{\sqrt{N}})$, the mutual coherence is upper bounded by $\mathcal{O}(\sqrt{\frac{\log(N/\delta)}{N}})$, which is independent of B.*

Proof. A common proof strategy for this theorem as well as for other theorems in this paper is to establish a large deviation inequality that implies the quantity of our interest is concentrated around its expected value with high probability. Proof steps include:

- Showing that the quantity of our interest is a sum of independent random variables;
- Bounding the expectation and variance of the quantity;
- Applying a relevant concentration inequality of a sum of random variables;
- Applying a union bound for the supreme of a random sequence.

In this case, the quantity of interest is:

$$S_{ij} = \langle \boldsymbol{A}_i, \boldsymbol{\Psi}_j \rangle = \sum_{k \in \text{supp}(\boldsymbol{F}_i)} F_{ik} \Psi_{kj} R_{kk}$$

Denote $Z_k = F_{ik}\Psi_{kj}R_{kk}$, for $k \in \text{supp}(\boldsymbol{F}_i)$ (in the support set of the row $\boldsymbol{F}_i$). Because R_{kk} are i.i.d Bernoulli random variables, Z_k are also i.i.d random variables with $E(Z_k) = 0$. Z_{kk} are also bounded because $Z_k = \pm F_{ik}\Psi_{kj}$

S_{ij} is a sum of independent, bounded random variables. Applying the Hoeffding's inequality (see Appendix 2) yields:

$$\Pr(|S_{ij}| \geq t) \leq 2\exp(-\frac{t^2}{\sum_{k \in \text{supp}(\boldsymbol{f}_i)} F_{ik}^2 \Psi_{jk}^2}).$$

The next step is to evaluate $\sigma^2 = \sum_{k \in supp(\boldsymbol{f}_i)} F_{ik}^2 \Psi_{jk}^2$. Here, σ^2 can be roughly viewed as an approximation of the variance of S_{ij}.

$$\sigma^2 \leq \max_{1 \leq i,j \leq N} |F_{ij}|^2 \sum_{k \in \text{supp}(\boldsymbol{F}_i)} \Psi_{kj}^2 \leq \max_{1 \leq i,j \leq N} |F_{ij}|^2 = \frac{c}{B} \tag{14}$$

If the maximal absolute magnitude of entries of $\boldsymbol{\Psi}$ is on the order of magnitude of $\mathcal{O}(\frac{1}{\sqrt{N}})$:

$$\max_{1 \leq i,j \leq N} |\Psi_{ij}| = \frac{c}{\sqrt{N}},$$

where c is some positive constant, then

$$\sigma^2 \leq \max_{1\leq i,j\leq N} |\Psi_{ij}|^2 \sum_{1\leq k\leq N} F_{ik}^2 \leq \max_{1\leq i,j\leq N} |\Psi_{ij}|^2 = \frac{c}{N}. \tag{15}$$

Finally, we derive an upper bound of the mutual coherence $\mu = \max_{1\leq i,j\leq N} |S_{ij}|$ by taking a union bound for the supreme of a random sequence:

$$P(\max_{1\leq i,j\leq N} |S_i j| \geq t) \leq 2N^2 \exp(\frac{-t^2}{\sigma^2}).$$

Choose $t = \sqrt{\sigma^2 \log(2N^2/\delta)}$, after simplifying the inequality, we get:

$$P(\max_{1\leq i,j\leq N} |S_{ij}| \leq \sqrt{\sigma^2 \log(2N^2/\delta)}) \geq 1-\delta.$$

Thus, with an arbitrarily $\mathbf{\Psi}$, (14) holds and we achieve the first claim of the Theorem:

$$P(\max_{1\leq i,j\leq N} |S_{ij}| \leq \sqrt{\frac{c\log(2N^2/\delta)}{B}}) \geq 1-\delta.$$

In the case that (15) holds, we achieve the second claim of the Theorem:

$$P(\max_{1\leq i,j\leq N} |S_{ij}| \leq \sqrt{\frac{c\log(2N^2/\delta)}{N}}) \geq 1-\delta.$$

□

Remark III.1. When $\boldsymbol{A}$ is a popular transform such as the DCT or the normalized WHT, the maximal absolute magnitude of entries is on the order of $\mathcal{O}(\frac{1}{\sqrt{N}})$. As a result, the mutual coherence of the $\boldsymbol{A}$ and an *arbitrary* $\mathbf{\Psi}$ is upper bounded by $\mathcal{O}(\sqrt{\frac{\log(N/\delta)}{N}})$, which is also consistent with our asymptotic analysis above. In other words, when at least $\mathbf{\Phi}$ or $\mathbf{\Psi}$ is a *dense and uniform* matrix, i.e. the maximal absolute magnitude of their entries is on the order of $\mathcal{O}(\frac{1}{\sqrt{N}})$, their mutual coherence is nearly minimal, except the $\log N$ factor. Otherwise, mutual coherence between any arbitrary $\mathbf{\Psi}$ and a sparse matrix $\boldsymbol{A}$ (e.g. block diagonal matrix of block size B) might be $\sqrt{\frac{N}{B}}$ times larger.

Cumulative coherence is more subtle way to quantify incoherence between two matrices [21].

Definition III.1. The cumulative coherence of an $N \times N$ $\boldsymbol{A}$ and an $N \times K$ $\boldsymbol{B}$ is defined as:

$$\mu_c(\boldsymbol{A}, \boldsymbol{B}) = \max_{1\leq i\leq N} \sqrt{\sum_{1\leq j\leq K} \langle \boldsymbol{A}_i, \boldsymbol{B}_j \rangle^2}$$

where $\boldsymbol{A}_i$ and $\boldsymbol{B}_j$ are rows of $\boldsymbol{A}$ and columns of $\boldsymbol{B}$, respectively.

The cumulative coherence $\mu_c(\boldsymbol{A}, \boldsymbol{B})$ measures the *average* incoherence between two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ while mutual coherence $\mu(\boldsymbol{A}, \boldsymbol{B})$ measures the entry-wise incoherence. As a result, the cumulative coherence seems to be a better indicator of average sensing performance. In many cases, we are only interested in cumulative coherence between $\boldsymbol{A}$ and $\mathbf{\Psi}_T$, where T is the support of the transform coefficient vector. As will be shown in the following section, the cumulative coherence provides a more powerful tool to obtain a tighter bound of the number of measurements required for exact recovery.

From the definition of cumulative coherence, it is easy to verify that $\mu_c \leq \sqrt{K}\mu$. If we directly apply the result of the Theorem III.3, we obtain a trivial bound of the cumulative coherence: $\mu_c = \mathcal{O}(\sqrt{\frac{K \log N}{B}})$ for any arbitrary basis $\boldsymbol{\Psi}$ and $\mu_c = \mathcal{O}(\sqrt{\frac{K \log N}{N}})$ for any dense and uniform $\boldsymbol{\Psi}$. In fact, we can get rid of the factor $\log N$ by directly measuring the cumulative coherence from its definition.

Theorem III.4. *Let $\boldsymbol{A} = \boldsymbol{FR}$, where $\boldsymbol{R}$ is an $N \times N$ random diagonal matrix of i.i.d Bernoulli random variables along its diagonal $P(R_{ii} = \pm 1) = 1/2$. Let $\boldsymbol{F}$ be an $N \times N$ unit-norm row matrix with the maximal absolute magnitude of entries is on the order of $\mathcal{O}(\frac{1}{\sqrt{B}})$, i.e. $\max_{1 \leq i,j \leq N} |F_{ij}| = \frac{c}{\sqrt{B}}$, where $1 \leq B \leq N$ and c is some positive constant. Let $\boldsymbol{\Psi}$ be an $N \times N$ unit-norm column matrix. With probability at least $1 - \delta$, the cumulative coherence of $\boldsymbol{A}$ and $\boldsymbol{\Psi}_{\mathcal{T}}$, where $|\mathcal{T}| = K$, is upper bounded by $\mathcal{O}(\sqrt{\frac{K}{B}})$ if $K > 16c^2 \log(2N/\delta)$.*

Proof. Denote $\boldsymbol{U} = \boldsymbol{\Psi}_{\mathcal{T}}^*$ and $\boldsymbol{U}_k$ are columns of $\boldsymbol{U}$. Let $\boldsymbol{A}_i$ and $\boldsymbol{\Psi}_j$ ($j \in \mathcal{T}$) be rows of $\boldsymbol{A}$ and columns of $\boldsymbol{\Psi}_{\mathcal{T}}$, respectively.

$$S_i = \sqrt{\sum_{j \in \mathcal{T}} \langle \boldsymbol{A}_i, \boldsymbol{\Psi}_j \rangle^2} = \|\boldsymbol{A}_i \boldsymbol{\Psi}_{\mathcal{T}}\|_2 = \| \sum_{k \in \text{supp}(\boldsymbol{F}_i)} R_{kk} F_{ik} \boldsymbol{U}_k \|_2.$$

Denote $\boldsymbol{V}_k = F_{ik}\boldsymbol{U}_k$ and $\boldsymbol{V}$ is the matrix of columns $\boldsymbol{V}_k$, $k \in \text{supp}(\boldsymbol{F}_i)$. First, we derive upper bound for the Frobenius of $\boldsymbol{V}$:

$$\|\boldsymbol{V}\|_F^2 \leq \max_{1 \leq i,j \leq N} F_{ij}^2 \|U\|_F^2 = \frac{c^2 K}{B}.$$

The last equation holds because $\|\boldsymbol{U}\|_F^2 = K$. Also, the bound for the spectral norm is:

$$\begin{aligned}
\|V\|_2^2 &= \sup_{\|\boldsymbol{\beta}\|_2 = 1} \sum_{k \in \text{supp}(\boldsymbol{F}_i)} |\langle \boldsymbol{\beta}, \boldsymbol{V}_k \rangle|^2 \\
&= \sup_{\|\boldsymbol{\beta}\|_2 = 1} \sum_{k \in \text{supp}(\boldsymbol{F}_i)} F_{ik}^2 (\sum_{j=1}^{K} \boldsymbol{\beta}_j U_{kj})^2 \\
&\leq \max_{1 \leq i,j \leq N} F_{ij}^2 \sup_{\|\boldsymbol{\beta}\|_2 = 1} \sum_{1 \leq k \leq N} |\langle \boldsymbol{\beta}, \boldsymbol{U}_k \rangle|^2 \\
&\leq \frac{c^2}{B} \|\boldsymbol{U}\|_2^2 = \frac{c^2}{B}.
\end{aligned}$$

The last equation holds because $\|\boldsymbol{U}\|_2^2 = 1$. Now, we have:

$$S_i = \| \sum_{k \in \text{supp}(\boldsymbol{F}_i)} R_{kk} F_{ik} \boldsymbol{U}_k \|_2 = \| \sum_{k \in \text{supp}(\boldsymbol{F}_i)} R_{kk} \boldsymbol{V}_k \|_2.$$

Let us denote $\boldsymbol{Z} = \sum_{k \in \text{supp}(\boldsymbol{F}_i)} R_{kk} \boldsymbol{V}_k$.

$\boldsymbol{Z}$ is a Rademacher sum of vectors and $S_i = \|\boldsymbol{Z}\|_2$ is a random variable. To show that S_i is concentrated around its expectation, we first derive bound of $E(\|\boldsymbol{Z}\|_2)$. It is easy to verify that for a random variable X, $E(X) \leq \sqrt{E(X^2)}$. Thus, we will derive the upper bound for the simpler quantity $E(\|\boldsymbol{Z}\|_2^2)$

$$\begin{aligned}
E(\|\boldsymbol{Z}\|_2^2) = E(\boldsymbol{Z}^* \boldsymbol{Z}) &= \sum_{k,l \in \text{supp}(\boldsymbol{F}_i)} E(R_{kk} R_{ll}) \langle \boldsymbol{V}_k, \boldsymbol{V}_l \rangle \\
&= \sum_{k \in supp(\boldsymbol{F}_i)} \langle \boldsymbol{V}_k, \boldsymbol{V}_k \rangle = \|\boldsymbol{V}\|_F^2 = \frac{c^2 K}{B}.
\end{aligned}$$

The third equality holds because R_{kk} are i.i.d Bernoulli random variables and thus, $E(R_{kk}R_{ll}) = 0\ \forall k \neq l$. As a result,

$$E(S_i) = E(\|\boldsymbol{Z}\|_2) \leq c\sqrt{\frac{K}{B}}.$$

Applying Ledoux's concentration inequality of the norm of a Rademacher sum of vectors [22] (see Appendix 2). Noting that $\|\boldsymbol{V}\|_2^2$ can be viewed as the variance of S_i, yields:

$$\Pr(S_i \geq c\sqrt{\frac{K}{B}} + t) \leq 2\exp(-t^2\frac{B}{16c^2})$$

Finally, apply a union bound for the supreme of a random process,we obtain:

$$\Pr(\max_{1\leq i\leq N} S_i \geq c\sqrt{\frac{K}{B}} + t) \leq 2N\exp(-t^2\frac{B}{16c^2}).$$

Choose $t = \sqrt{\frac{K}{B}}$. If $K > 16c^2\log(2N/\delta)$, we get:

$$\Pr(\max_{1\leq i\leq N} S_i \geq \mathcal{O}(\sqrt{\frac{K}{B}})) \leq \delta.$$

□

Remark III.2. When $\boldsymbol{F}$ is some popular transform such as the DCT or the normalized WHT, the maximum absolute magnitude of entries is on the order of $\mathcal{O}(\frac{1}{\sqrt{N}})$. As a result, the cumulative coherence of $\boldsymbol{A}$ and any arbitrary $\boldsymbol{\Psi}_{\mathcal{T}}$,where $|\mathcal{T}| = K$, is upper bounded by $\mathcal{O}(\sqrt{\frac{K}{N}})$ if $K > 16c^2\log(\frac{2N}{\delta})$, where c is some positive constant.

Remark III.3. The above theorem represents the worst-case analysis because $\boldsymbol{\Psi}$ can be an arbitrary matrix (the worst case corresponds to the case when $\boldsymbol{\Psi}$ is the identity matrix). When $\boldsymbol{\Psi}$ is known to be dense and uniform, the upper bound of cumulative coherence, according to the Theorem III.3 and the fact that $\mu_c \leq \mu\sqrt{K}$, is $\mathcal{O}(\sqrt{\frac{K\log N}{N}})$, which is, in general, better than $\mathcal{O}(\sqrt{\frac{K}{B}})$.

C. Incoherence Analysis With The Global Randomizer

The asymptotical analysis above reveals a significant technical difference for two cases: when $\boldsymbol{R}$ is the local randomizer and when $\boldsymbol{R}$ is the global randomizer. With the local randomizer, entries of $\boldsymbol{A\Psi}$ are sums of *independent* random variables while with global randomizer they are sums of *dependent* random variables. Stochastic dependence among random variables makes it much harder to set up similar arguments of their sum's concentration. In this case, we will show that the incoherence of $\boldsymbol{A}$ and $\boldsymbol{\Psi}$ might depend on an extra quantity, the *heterogeneity coefficient* of the matrix $\boldsymbol{\Psi}$.

Definition III.2. Assume $\boldsymbol{\Psi}$ is an $N \times N$ matrix. Let $\mathcal{T}_k$ be the support of the column $\boldsymbol{\Psi}_k$. Define:

$$\rho_k = \frac{\max_{1\leq i\leq N}|\Psi_{ki}|}{\sqrt{\frac{1}{|\mathcal{T}_k|}\sum_{i\in\ \mathcal{T}_k}\Psi_{ki}^2}}. \tag{16}$$

The column-wise heterogeneity coefficient of the matrix $\mathbf{\Psi}$ is defined as:

$$\rho_{\mathbf{\Psi}} = \max_{1\leq k\leq N} \rho_k. \tag{17}$$

Obviously, $1 \leq \rho_k \leq \sqrt{|\mathcal{T}_k|}$. ρ_k illustrates the difference between the largest entry's magnitude and the average energy of *nonzero* entries. Roughly speaking, it indicates heterogeneity of nonzero entries of the vector $\mathbf{\Psi}_k$. If nonzero entries of a column $\mathbf{\Psi}_k$ are homogeneous, i.e. they are on the same order of magnitude, ρ_k is on the order of a constant. If all nonzero entries of a matrix are homogeneous, the heterogeneity coefficient is also on the order of a constant, $C_{\mathbf{\Psi}} = \mathcal{O}(1)$ and $\mathbf{\Psi}$ is referred as a uniform matrix. Note that a uniform matrix is not necessarily dense, for example, a block-diagonal matrix of DCT or WHT blocks

The following theorem indicates that when the global randomizer is employed, the mutual coherence between $\boldsymbol{A}$ and $\mathbf{\Psi}$ is upper-bounded by $\mathcal{O}(\rho_{\mathbf{\Psi}}\sqrt{\frac{\log(N/\delta)}{B}})$, where B is the block size of $\mathbf{\Phi}$ and $\mathbf{\Psi}$ is an arbitrarily matrix with the heterogeneity coefficient $\rho_{\mathbf{\Psi}}$.

Theorem III.5. *Let $\boldsymbol{A} = \boldsymbol{F}\boldsymbol{R}$, where $\boldsymbol{R}$ is an $N \times N$ uniformly random permutation matrix. Let $\boldsymbol{F}$ be an $N \times N$ unit-norm row matrix with the maximal absolute magnitude of entries is on the order of $\mathcal{O}(\frac{1}{\sqrt{B}})$. Assume that all rows of $\boldsymbol{F}$ have zero average sum. Let $\mathbf{\Psi}$ be an $N \times N$ unit-norm column matrix with $\mathcal{T}_k$ and $\rho_{\mathbf{\Psi}}$ defined as in (16) and (17). Assume that $\rho_k \geq 4\log(2N^2/\delta)$ $\forall k \in \{1, 2, \ldots, N\}$.*

- *With probability at least $1-\delta$, the mutual coherence of $\boldsymbol{A}$ and $\mathbf{\Psi}$ is upper-bounded by $\mathcal{O}(\rho_{\Psi}\sqrt{\frac{\log(N/\delta)}{B}})$.*
- *In addition, if $\mathbf{\Psi}$ is dense and uniform, i.e. the maximum absolute magnitude of its entries is on the order of $\mathcal{O}(\frac{1}{\sqrt{N}})$ and $B \geq 4\log(2N^2/\delta)$, the mutual coherence is upper-bounded by $\mathcal{O}(\sqrt{\frac{\log(N/\delta)}{N}})$, which is independent of B.*

Proof. Let $[\omega_1, \omega_2, \ldots, \omega_N]$ be a uniformly random permutation of $[1, 2, \ldots, N]$. With the same notation of $\boldsymbol{A}_i, \mathbf{\Psi}_j, \boldsymbol{F}_i, S_{ij}, F_{ik}, \Psi_j$ as denoted previously:

$$S_{ij} = \langle \boldsymbol{A}_i, \mathbf{\Psi}_j \rangle = \sum_{k=1}^{N} F_{i\omega_k}\Psi_{jk}.$$

As in the proof of the Theorem III.2, $\{\omega_k\}_{k=1}^N$ can be viewed as a sequence of dependent random variables with identical distribution, i.e. for a fixed $k \in \{1, 2, \ldots, N\}$:

$$P(\omega_k = i) = \frac{1}{N}, \quad i \in \{1, 2, \ldots, N\}.$$

The condition of $\boldsymbol{F}$ is equivalent to $\max_{1\leq i,j\leq N}|F_{ij}| = \frac{c}{\sqrt{B}}$, where c is some positive constant. Define $\{w_{k\omega_k}\}_{k=1}^N$ as the follows:

$$w_{k\omega_k} = \begin{cases} \frac{\sqrt{B|\mathcal{T}_k|}}{2c\rho_{\mathbf{\Psi}}}F_{i\omega_k}\Psi_{jk} + \frac{1}{2} & \text{if } \Psi_{jk} \neq 0 \\ \frac{\sqrt{B|\mathcal{T}_k|}}{2c\rho_{\mathbf{\Psi}}}F_{i\omega_k}\Psi_{jk} & \text{if } \Psi_{jk} = 0. \end{cases}$$

It is easy to verify that $0 \leq w_{k\omega_k} \leq 1$. Define W_k as the sum of dependent random variables $w_{k\omega_k}$

$$W_k = \sum_{k=1}^{N} w_{k\omega_k} = \frac{\sqrt{B|T_k|}}{2c\rho_\Psi} \sum_{k=1}^{N} F_{i\omega_k}\Psi_{jk} + \frac{|T_k|}{2}$$
$$= \frac{\sqrt{B|T_k|}}{2c\rho_\Psi} S_{ij} + \frac{|T_k|}{2}.$$

Note that $\{F_{i\omega_k}\}_{k=1}^{N}$ are zero-mean random variables because $\boldsymbol{F}_i$ has zero average sum. Thus, $E(S_{ij}) = 0$ and $E(W_k) = \frac{|T_k|}{2}$. Then, applying the Sourav's theorem of concentration inequality for a sum of *dependent* random variables [23] (see Appendix 2) results in:

$$P\{\frac{\sqrt{B|\mathcal{T}_k|}}{2c\rho_{\boldsymbol{\Psi}}}|S_{ij}| \geq \epsilon\} \leq 2\exp(-\frac{\epsilon^2}{2|\mathcal{T}_k| + 2\epsilon}).$$

Denote $t = \frac{2c\rho_{\boldsymbol{\Psi}}}{\sqrt{B|\mathcal{T}_k|}}\epsilon$. The above inequality is equivalent to:

$$P\{|S_{ij}| \geq t\} \leq 2\exp(-\frac{B|\mathcal{T}_k|}{4c^2\rho_\Psi^2}\frac{t^2}{2|\mathcal{T}_k| + \frac{t}{c\rho_{\boldsymbol{\Psi}}}\sqrt{B|\mathcal{T}_k|}}).$$

By choosing $t = 4c\rho_{\boldsymbol{\Psi}}\sqrt{\frac{1}{B}\log(\frac{2N^2}{\delta})}$, we achieve:

$$P\{|S_{ij}| \geq t\} \leq 2\exp(\frac{-4|\mathcal{T}_k|\log(\frac{2N^2}{\delta})}{2|\mathcal{T}_k| + 4\sqrt{|\mathcal{T}_k|\log(\frac{2N^2}{\delta})}}).$$

If $|\mathcal{T}_k| \geq 4\log(\frac{2N^2}{\delta})$, the denominator inside the exponent is smaller than $4|\mathcal{T}_k|$. Thus,

$$P\{|S_{ij}| \geq 2c\rho_{\boldsymbol{\Psi}}\sqrt{\frac{1}{B}\log(\frac{2N^2}{\delta})}\} \leq 2\exp(-\log(\frac{2N^2}{\delta})) = \frac{\delta}{N^2}.$$

Finally, after taking the union bound for the supreme of a random sequence and simplifying the inequality, we obtain the first claim of the Theorem:

$$P\{\max_{1\leq i,j\leq N}|S_{ij}| \leq \mathcal{O}(\rho_{\boldsymbol{\Psi}}\sqrt{\frac{\log(N/\delta)}{B}})\} \geq 1-\delta.$$

If $\boldsymbol{\Psi}$ is known to be dense and uniform, i.e. $\max_{1\leq i,j\leq N}|\Psi_{ij}| = \frac{c_1}{\sqrt{N}}$, where c_1 is some positive constant. We then define $\{w_{k\omega_k}\}_{k=1}^{N}$ as the following:

$$w_{k\omega_k} = \begin{cases} \frac{\sqrt{BN}}{2cc_1}F_{ik}\Psi_{j\omega_k} + \frac{1}{2} & \text{if } F_{ik} \neq 0 \\ \frac{\sqrt{BN}}{2cc_1}F_{ik}\Psi_{j\omega_k} & \text{if } F_{ik} = 0. \end{cases}$$

Note that $0 \leq w_{k\omega_k} \leq 1$ and $E(w_{k\omega_k}) = \frac{B}{2}$. Repeat the same arguments above, we have:

$$P\{|S_{ij}| \geq t\} \leq 2\exp(-\frac{NB}{4c^2c_1^2}\frac{t^2}{2B + \frac{t}{cc_1}\sqrt{NB}}).$$

Similarly, choose $t = 4cc_1\sqrt{\frac{1}{N}\log(\frac{2N^2}{\delta})}$, we can derive:

$$P\{|S_{ij}| \geq t\} \leq 2\exp(\frac{-4B\log(\frac{2N^2}{\delta})}{2B + 4\sqrt{B\log(\frac{2N^2}{\delta})}}).$$

If $B \geq 4\log(\frac{2N^2}{\delta})$, the denominator inside the exponent is smaller than $4B$. Thus,

$$P\{|S_{ij}| \geq 2cc_1\sqrt{\frac{1}{N}\log(\frac{2N^2}{\delta})}\} \leq \frac{\delta}{N^2}.$$

After taking the union bound of the supreme of a random sequence, we achieve the second claim of the Theorem. □

Remark III.4. The first part of theorem implies that when $\boldsymbol{F}$ is a dense and uniform matrix (e.g. DCT or normalized WHT) and $\boldsymbol{\Psi}$ is a uniform matrix (not necessarily dense), the mutual coherence closely approaches the minimum $\mathcal{O}(\sqrt{\frac{\log(N/\delta)}{N}})$. Although in this theorem, the mutual coherence depends on the heterogeneity coefficient, one will see in the experimental Section **??** that this dependence is almost negligible in practice.

As a consequence of this theorem, when at least $\boldsymbol{A}$ or $\boldsymbol{\Psi}$ is dense and uniform, the mutual coherence of $\boldsymbol{A}$ and $\boldsymbol{\Psi}$ is roughly on the order of $\mathcal{O}(\sqrt{\frac{\log N}{N}})$, which is quite close to the lower bound $\frac{1}{\sqrt{N}}$, except for the $\log N$ factor. Otherwise, the coherence linearly depends on the block size B of $\boldsymbol{F}$ and is on the order of $\mathcal{O}(\sqrt{\frac{\log N}{B}})$. As a matter of fact, this bound is almost optimal because when $\boldsymbol{\Psi}$ is the identity matrix, the mutual coherence is actually equal the maximum absolute magnitude of entries of $\boldsymbol{A}$, which is on the order of $\mathcal{O}(\frac{1}{\sqrt{B}})$.

IV. Compressive Sampling Performance Analysis

Section III demonstrates that under some mild conditions, the matrix $\boldsymbol{A}$ and $\boldsymbol{\Psi}$ are highly incoherent, implying that the matrix $\boldsymbol{A\Psi}$ is almost dense. When $\boldsymbol{A\Psi}$ is dense, energy of nonzero transform coefficients $\boldsymbol{\alpha}_T$ is distributed over all measurements. Commonly speaking, this is good for signal recovery from a small subset of measurements because if energy of some transform coefficients were concentrated in few measurements that happens to be bypassed in the sampling process, there is no hope for exact signal recovery even when employing the most sophisticated reconstruction method. This section shows that a random subset of rows of the matrix $\boldsymbol{A} = \boldsymbol{FR}$ yields almost optimal measurement matrix $\boldsymbol{\Phi}$ for compressive sensing.

A. Main Assumptions for Theoretical Analysis

We first discuss main assumptions for theoretical results in the next section to hold. A signal $\boldsymbol{x}$ is assumed to be sparse in some sparsifying basis $\boldsymbol{\Psi}$: $\boldsymbol{x} = \boldsymbol{\Psi\alpha}$, where the vector of transform coefficients $\boldsymbol{\alpha}$ has no more than K nonzero entries. The sign sequence of nonzero transform coefficients $\boldsymbol{\alpha}_T$ which is denoted as $\boldsymbol{z}$, is assumed to be a random vector of i.i.d Bernoulli random variables (i.e. $P(z_i = \pm 1) = \frac{1}{2}$). Let $\boldsymbol{y} = \boldsymbol{\Phi x}$ be the measurement vector, where $\boldsymbol{\Phi} = \sqrt{\frac{N}{M}}\boldsymbol{DFR}$ is a structurally random matrix. For a general analysis, F is assumed to be a block diagonal (and uniform) matrix with block size B ($1 \leq B \leq N$). If $\boldsymbol{R}$ is the global randomizer, we also need the additional assumption that $\boldsymbol{\Psi}$ is uniform so that the theorem III.5 holds with the heterogeneity coefficient of $\boldsymbol{\Psi}$, $\rho_{\boldsymbol{\Psi}}$, is on the order of a constant. Note that the sensing operation can be equivalently accomplished via applying the Algorithm 1.

B. Theoretical Results

Theorem IV.1. *With probability at least $1-\delta$, the proposed sensing framework can recover K-sparse signals exactly if the number of measurements $M \geq \mathcal{O}(\frac{N}{B}K\log^2(\frac{N}{\delta}))$. If $\boldsymbol{F}$ is a dense and uniform rather than block-diagonal(e.g. DCT or normalized WHT matrix), the number of measurement needed is on the order of $\mathcal{O}(K\log^2(\frac{N}{\delta}))$.*

Proof. This is a simple corollary of the theorem of Candès et. al. [[14] *Theorem* 1.1] (1) because (i) $\boldsymbol{A} = \boldsymbol{FR}$ is an orthonormal matrix, and (ii) our incoherence results between $\boldsymbol{A}$ and $\boldsymbol{\Psi}$ in the Theorem III.3 and Theorem III.5. □

Remark IV.1. If $\boldsymbol{\Psi}$ is dense and uniform, the number of measurements for exact recovery is always $\mathcal{O}(K\log^2(\frac{N}{\delta}))$ regardless of the block size B. This implies that we can use the identity matrix for the transform $\boldsymbol{F}$ (B = 1). For example, when input signal is known in advance to be spectrally sparse, compressively sampling it in the time domain is as efficient as in any other transform domain.

Compared with the framework that uses random projection, there is an upscale factor of $\log N$ for the number of measurements for exact recovery. In fact, by employing the above result of cumulative coherence, we can eliminate this upscale factor and thus, successfully showing optimal performance guarantee.

Theorem IV.2. *Assume that the sparsity $K > 16c^2\log(\frac{2N}{\delta})$. With probability at least $1-\delta$, the proposed framework employing the local randomizer can reconstruct K-sparse signals exactly if the number of measurements $M \geq \mathcal{O}(\frac{N}{B}K\log(\frac{N}{\delta}))$.If $\boldsymbol{F}$ is a dense and uniform matrix (e.g. DCT or normalized WHT), the minimal number of required measurements is $M = \mathcal{O}(K\log(\frac{N}{\delta}))$.*

Proof. The proof is based on the result of cumulative coherence in the Theorem III.4 and a modification of the proof framework of the compressed sensing [14].

Denote $\boldsymbol{U} = \sqrt{\frac{N}{M}}\boldsymbol{FR\Psi}$, $\boldsymbol{U}_{\mathcal{T}} = \sqrt{\frac{N}{M}}\boldsymbol{FR\Psi}_{\mathcal{T}}$, $\boldsymbol{U}_{\Omega} = \sqrt{\frac{N}{M}}\boldsymbol{DFR\Psi}$ and $\boldsymbol{U}_{\Omega\mathcal{T}} = \sqrt{\frac{N}{M}}\boldsymbol{DFR\Psi}_{\mathcal{T}}$, where the support $\Omega = \{k|\boldsymbol{D}_{kk} = 1, k = 1, 2, .., N\}$. Let $\boldsymbol{v}_k$, $k \in \{1, 2, ..., N\}$, be columns of $\boldsymbol{U}^*_{\mathcal{T}}$. Denote $\mu_c = \max_{1\leq k\leq N}\|\boldsymbol{v}_k\|_2$, where $\mu_c = \mu_c(\boldsymbol{A}, \boldsymbol{\Psi}_{\mathcal{T}})$ is the cumulative coherence of $\boldsymbol{A} = \sqrt{\frac{N}{M}}\boldsymbol{FR}$ and $\boldsymbol{\Psi}_{\mathcal{T}}$. According to the above incoherence analysis, $\mu_c \leq \mathcal{O}(\sqrt{\frac{KN}{BM}})$. Also, denote μ as the mutual coherence of $\boldsymbol{A}$ and $\boldsymbol{\Psi}_{\mathcal{T}}$, $\mu \leq \mathcal{O}(\sqrt{\frac{N\log N}{BM}})$.

As indicated in [12], [14], to show l_1 minimization exact recovery, it is sufficient to verify the *Exact Recovery Principle*.

Exact Recovery Principle. *With high probability, $|\pi_k| < 1$ for all $k \in \mathcal{T}^c$, where $\mathcal{T}^c$ is the complementary set of the set $\mathcal{T}$ and $\pi = \boldsymbol{U}^*_{\Omega}\boldsymbol{U}_{\Omega\mathcal{T}}(\boldsymbol{U}^*_{\Omega\mathcal{T}}\boldsymbol{U}_{\Omega\mathcal{T}})^{-1}\boldsymbol{z}$, where $\boldsymbol{z}$ is the sign vector of nonzero transform coefficients $\boldsymbol{\alpha}_{\mathcal{T}}$.*

Also note that $\pi_k = \langle\nu_k(\boldsymbol{U}^*_{\Omega\mathcal{T}}\boldsymbol{U}_{\Omega\mathcal{T}})^{-1}, z\rangle$, where $\boldsymbol{\nu}_k$ is the k^{th} row of $\boldsymbol{U}^*_{\Omega}\boldsymbol{U}_{\Omega\mathcal{T}}$, for some $k \in \mathcal{T}^c$. The proof contains three major steps:

- *Claim 1* (Bound the norm of ν_k): With high probability, $\|\boldsymbol{\nu}_k\|$ on the order of $\mathcal{O}(\mu_c)$
- *Claim 2* (Bound the spectral norm of $\boldsymbol{U}^*_{\Omega T}\boldsymbol{U}_{\Omega\mathcal{T}}$): With probability $1-\delta$, $\|\boldsymbol{U}^*_{\Omega\mathcal{T}}\boldsymbol{U}_{\Omega\mathcal{T}}\| \geq \frac{1}{2}$.

- *Claim 3* (Bound the norm of $w_k = \boldsymbol{\nu}_k(\boldsymbol{U}^*_{\Omega\mathcal{T}}\boldsymbol{U}_{\Omega T})^{-1}$): With high probability, $\|w_k\|$ is on the order of $\mathcal{O}(\mu_c)$. Finally, exploiting the assumption that $\boldsymbol{z}$ is a random vector of i.i.d Bernoulli random variables to show that with probability $1 - \mathcal{O}(\delta)$, $|\pi_k| = |\langle w_k, z\rangle| < 1$

□

We first present proof for the Claim 1.

Proof. Let $\boldsymbol{U}_k$ be columns of $\boldsymbol{U}$. For $k \in \mathcal{T}^c$:

$$\boldsymbol{\nu}_k = \frac{1}{M}\sum_{i=1}^{N} D_{ii}U_{ik}\boldsymbol{v}_i = \sum_{i=1}^{N}(D_{ii} - \frac{M}{N})U_{ik}\boldsymbol{v}_i$$

where the second equality holds because $\sum_{i=1}^{N} U_{ik}\boldsymbol{v}_i = \boldsymbol{U}^*_T\boldsymbol{U}_k = 0$ that results from the orthogonality of columns of $\boldsymbol{U}$. Let $Z_i = (D_{ii} - \frac{M}{N})$. Because D_{ii} are i.i.d binary random variables with $P(D_{ii} = 1) = \frac{M}{N}$, Z_i are zero mean i.i.d random variables and $E(Z_i^2) = \frac{M}{N}(1 - \frac{M}{N})$. Let $\boldsymbol{W}$ be the matrix of columns $\boldsymbol{W}_i = U_{ik}\boldsymbol{v}_i$, $i \in \{1, 2, \ldots, N\}$. Then, $\boldsymbol{\nu}_k$ can be viewed as a random weighted sum of column vectors $\boldsymbol{w}_i$:

$$\boldsymbol{\nu}_k = \frac{1}{M}\sum_{i=1}^{N} Z_i\boldsymbol{W}_i$$

and $\|\boldsymbol{\nu}_k\|$ is a random variable. We have:

$$E(\|\boldsymbol{\nu}_k\|^2) = \sum_{1\le i,j\le N} E(Z_iZ_j)\langle \boldsymbol{W}_i, \boldsymbol{W}_j\rangle = \sum_{1\le i\le N} E(Z_i^2)\|\boldsymbol{W}_i\|^2,$$

where the last equality holds due to $E(Z_iZ_j) = 0$ if $i \neq j$. Thus,

$$\begin{aligned} E(\|\boldsymbol{\nu}_k\|^2) &= \frac{M}{N}(1 - \frac{M}{N})\sum_{1\le i\le N} V_{ki}^2\|U_i\|^2 \\ &\le \frac{M}{N}(1 - \frac{M}{N})\mu_c^2\sum_{1\le i\le N} U_{ik}^2 \le \mu_c^2. \end{aligned}$$

where the last inequality holds due to $\|\boldsymbol{U}_k\|^2 = \frac{N}{M}$. This implies that $E(\|\boldsymbol{\nu}_k\|) \le \mu_c$. To show that $\|\boldsymbol{\nu}_k\|$ is concentrated around its mean, we use the Talagrand's theorem of concentration inequality [24]. First, we have:

$$\begin{aligned} \|\boldsymbol{W}\|_2^2 &= \sup_{\|\beta\|=1}\sum_{i=1}^{N}|\langle\beta, \boldsymbol{W}_i\rangle|^2 = \sup_{\|\beta\|=1}\sum_{i=1}^{N} U_{ik}^2|\langle\beta, v_i\rangle|^2 \\ &\le \mu^2\sup_{\|\beta\|=1}\sum_{i=1}^{N}|\langle\beta, v_i\rangle|^2 = \mu^2\|\boldsymbol{U}_T\|_2^2 = \frac{N}{M}\mu^2. \end{aligned}$$

where the last equation holds because $\|\boldsymbol{U}_T\|_2^2 = \frac{N}{M}$. Thus, we derive the upper bound of the variance σ^2:

$$\sigma^2 = E(Z_k^2)\|W\|_2^2 \le \frac{M}{N}(1 - \frac{M}{N})\frac{N}{M}\mu^2 \le \mu^2.$$

In addition, it is obvious that $|Z_k| \le 1$ and thus

$$B = \max_{1\le i\le N}\|\boldsymbol{W}_i\|_2 \le \mu\mu_c.$$

The Talagrand's theorem [24] (see Appendix 2) shows that:

$$P(\|\boldsymbol{\nu}_k\| - E(\|\boldsymbol{\nu}_k\|) \ge t) \le 3\exp(\frac{-t}{cB}\log(1 + \frac{Bt}{\sigma^2 + BE(\|\boldsymbol{\nu}_k\|)})),$$

where c is some positive constant. Replacing $E(\|\boldsymbol{\nu}_k\|)$, σ^2 and B by their upper bounds in the right-hand side, we obtain:

$$P(\|\boldsymbol{\nu}_k\| - E(\|\boldsymbol{\nu}_k\|) \geq t) \leq 3\exp(\frac{-t}{c\mu\mu_c}\log(1+\frac{\mu\mu_c t}{\mu^2+\mu\mu_c^2})).$$

The next step is to simplify the right-hand side of the above inequality by replacing the denominator inside the log by two times the dominant term and note that $log(1+x) \geq \frac{x}{2}$ when $x \leq 1$. In particular, there are two cases:

- Case 1: $\mu\mu_c^2 \geq \mu^2$ or equivalently, $\mu_c^2 \geq \mu$, denote $\overline{\sigma}^2 = \mu\mu_c^2$ and $t = a\overline{\sigma}$. If $\mu\mu_c t \leq 2\mu\mu_c^2$ or equivalently, $a \leq 2(1/\mu)^{\frac{1}{2}}$,
$$P(\|\boldsymbol{\nu}_k\| - E(\|\boldsymbol{\nu}_k\|) \geq t) \leq 3\exp(-\gamma a^2).$$
- Case 2: $\mu^2 \geq \mu\mu_c^2$, denote $\overline{\sigma}^2 = \mu^2$ and $t = a\overline{\sigma}$. If $\mu\mu_c t \leq 2\mu^2$ or equivalently, $a \leq 2/\mu_c$
$$P(\|\boldsymbol{\nu}_k\| - E(\|\boldsymbol{\nu}_k\|) \geq t) \leq 3\exp(-\gamma a^2).$$
where γ is some positive constant.

In conclusion, we just derive that

$$P(\|\boldsymbol{\nu}_k\| \geq \mu_c + a\overline{\sigma}) \leq 3\exp(-\gamma a^2), \tag{18}$$

where γ is a positive constant and a is an arbitrary number that satisfies the above conditions. □

The Theorem 1.2 in [14] shows that the *Claim 2* holds when $M \geq \mu_c^2 \max(c_1 \log K, c_2 \log(3/\delta))$, where c_1 and c_2 are some known positive constants.

Finally, we present proof for the Claim 3.

Proof. First, we show that:

$$P(\sup_{k\in\mathcal{T}^c}\|\boldsymbol{W}_k\| \geq 2\mu_c + 2a\overline{\sigma}) \leq 3N\exp(-\gamma a^2) + P(\|U_{\Omega\mathcal{T}}^* U_{\Omega\mathcal{T}}\| \leq \frac{1}{2}). \tag{19}$$

where $\boldsymbol{W}_k = \boldsymbol{\nu}_k(\boldsymbol{U}_{\Omega\mathcal{T}}^*\boldsymbol{U}_{\Omega\mathcal{T}})^{-1}$.

Let $\mathcal{A}$ be the event that $\{\|U_{\Omega\mathcal{T}}^* U_{\Omega\mathcal{T}}\| \geq \frac{1}{2}\}$ or equivalently, $\{\|(U_{\Omega\mathcal{T}}^* U_{\Omega\mathcal{T}})^{-}1\| \leq 2\}$ and $\mathcal{B}$ be the event that $\{\sup_{k\in\mathcal{T}^c}\|\nu_k\| \leq \mu_c + a\overline{\sigma}\}$. Note that

$$\sup_{k\in\mathcal{T}^c}\|\boldsymbol{W}_k\| \leq \|(U_{\Omega\mathcal{T}}^* U_{\Omega\mathcal{T}})^{-1}\| \sup_{k\in\mathcal{T}^c}\|\boldsymbol{\nu}_k\|.$$

Thus,

$$P(\sup_{k\in\mathcal{T}^c}\|\boldsymbol{W}_k\| \geq 2\mu_c + 2a\overline{\sigma}) \leq P(\overline{\mathcal{A}\cap\mathcal{B}}) \leq P(\overline{\mathcal{A}}) + P(\overline{\mathcal{B}}).$$

Note that $P(\overline{\mathcal{B}}) \leq 3N\exp(-\gamma a^2)$ implies (19) holds.

The last step is to show that $\sup_{k\in\mathcal{T}^c}|\langle\boldsymbol{W}_k, \boldsymbol{z}\rangle| \leq 1$ with high probability. Note that because $\boldsymbol{z}$ is assumed to be a vector of i.i.d Bernoulli random variables, $|\langle\boldsymbol{W}_k, \boldsymbol{z}\rangle|$ is concentrated around its zero mean. In particular, according to the Hoeffding's inequality:

$$P(|\langle\boldsymbol{W}_k, \boldsymbol{z}\rangle| \geq 1) \leq 2\exp(-\frac{1}{2\|\boldsymbol{W}_k\|^2}).$$

$$\Rightarrow P(|\langle \boldsymbol{W}_k, \boldsymbol{z}\rangle| \geq 1 | \sup_{k\in\mathcal{T}^c} \|\boldsymbol{W}_k\| \leq \lambda) \leq 2N \exp(-\frac{1}{2\lambda^2}).$$

Note that with two arbitrary probabilistic events $\mathcal{A}$ and $\mathcal{B}$:

$$P(\mathcal{A}) = P(\mathcal{A}|\mathcal{B})P(\mathcal{B}) + P(\mathcal{A}|\overline{\mathcal{B}})P(\overline{\mathcal{B}}) \leq P(\mathcal{A}|\mathcal{B}) + P(\overline{\mathcal{B}}).$$

Now, let $\mathcal{A}$ be the event $\{\sup_{k\in\mathcal{T}^c} |\langle \boldsymbol{W}_k, \boldsymbol{z}\rangle| \geq 1\}$ and $\mathcal{B}$ be the event $\{\sup_{k\in\mathcal{T}^c} \|\boldsymbol{W}_k\| \leq \lambda\}$, we can show that

$$P(\sup_{k\in\mathcal{T}^c} |\langle \boldsymbol{W}_k, \boldsymbol{z}\rangle| \geq 1) \leq 2N \exp(-\frac{1}{2\lambda^2}) + P(\sup_{k\in\mathcal{T}^c} \|\boldsymbol{W}_k\| \geq \lambda). \tag{20}$$

Choose $\lambda = 2\mu_c + 2a\overline{\sigma}$, according to (19) and (20), the probability of our interest $P(\sup_{k\in\mathcal{T}^c} |\langle \boldsymbol{W}_k, \boldsymbol{z}\rangle| \geq 1)$ is upper bounded by:

$$3N \exp(-\gamma a^2) + 2N \exp(-\frac{1}{2\lambda^2}) + \delta.$$

To show that $\{\sup_{k\in\mathcal{T}^c} |\langle \boldsymbol{W}_k, \boldsymbol{z}\rangle| \leq 1\}$ with probability $1 - \mathcal{O}(\delta)$, it is sufficient to show that the above upper bound is not greater than 3δ. In particular, choose $a^2 = \gamma^{-1}\log(3N/\delta)$ that makes the first term to be equal δ.

To make the second term less than δ, it is required that

$$\frac{1}{2\lambda^2} \geq \log(\frac{2N}{\delta}). \tag{21}$$

- Case 1: $\mu_c^2 \geq \mu$. The condition that (18) holds is $a \leq 2(1/\mu)^{\frac{1}{2}}$ that is equivalent to:

  $$1 \geq \frac{1}{4}\gamma^{-2}\mu^2 \log^2(3N/\delta).$$

  It is easy to see $\mu_c \geq a\overline{\sigma}$, where $\overline{\sigma} = (\mu\mu_c^2)^{1/2}$. In this case, $\lambda \leq 4\mu_c$. Thus, (21) holds if

  $$1 \geq 32\mu_c^2 \log(\frac{2N}{\delta}). \tag{22}$$

- Case 2: $\mu \geq \mu_c^2$. The condition that (18) holds is $a \leq 2/\mu_c$ or equivalently,

  $$1 \geq \frac{1}{4}\gamma^{-2}\mu_c^2 \log(3N/\delta).$$

  If $\mu_c \geq a\overline{\sigma}$, where $\overline{\sigma} = \mu$, $\lambda \leq 4\mu_c$ and the condition is again (22). Otherwise, $\lambda \leq 4a\overline{\sigma}$. In this case, (21) holds if

  $$1 \geq 32\gamma^{-1}\mu^2 \log(\frac{2N}{\delta}).$$

In conclusion, the Exact Recovery Principle is verified if $1 \geq \max(c_1\mu^2 \log^2(3N/\delta), c_2\mu_c^2 \log(3N/\delta))$, where c_1 and c_2 are known positive constants.

Finally, note that $\mu^2 \leq \mathcal{O}(\frac{N\log N}{BM})$ and $\mu_c^2 \leq \mathcal{O}(\frac{NK}{BM})$ and the assumption that $K \geq 16c^2\log(\frac{2N}{\delta})$, the sufficient condition for exact recovery is $M \geq \mathcal{O}(\frac{N}{B}K\log(\frac{N}{\delta}))$. When $\boldsymbol{F}$ is dense and uniform, the condition becomes $M \geq \mathcal{O}(K\log(\frac{N}{\delta}))$. □

TABLE I

SRMs EMPLOYED IN THE EXPERIMENT WITH SPARSE SIGNALS

Notation	**R**	**F**
WHT64-L	Local randomizer	64×64 block diagonal WHT
WHT64-G	Global randomizer	64×64 block diagonal WHT
WHT256-L	Local randomizer	256×256 block diagonal WHT
WHT256-G	Global randomizer	256×256 block diagonal WHT

V. NUMERICAL EXPERIMENTS

A. Simulation with Sparse Signals

In this section, we evaluate the sensing performance of several structurally random matrices and compare it with that of completely random projection. We also explore the connection among sensing performance (probability of exact recovery), streaming capacity (block size of $\boldsymbol{F}$) and structure of the sparsifying basis $\boldsymbol{\Psi}$ (e.g. sparsity and heterogeneity).

In the first simulation, the input signal $\boldsymbol{x}$ of length $N = 256$ is sparse in the DCT domain, i.e. $\boldsymbol{x} = \boldsymbol{\Psi\alpha}$, where the sparsifying basis $\boldsymbol{\Psi}$ is the 256×256 IDCT matrix. Its transform coefficient vector $\boldsymbol{\alpha}$ has K nonzero entries whose magnitudes are Gaussian distributed and locations are at uniformly random, where $K \in \{10, 20, 30, 40, 50, 60\}$. With the signal $\boldsymbol{x}$, we generate a measurement vector of length $M = 128$: $\boldsymbol{y} = \boldsymbol{\Phi x}$, where $\boldsymbol{\Phi}$ is some structurally random matrix or a completely Gaussian random matrix. SRMs under consideration are summarized in Table I.

The Orthogonal Matching Pursuit algorithm [6], is used to recover the signal from its measurements $\boldsymbol{y}$. For each value of sparsity $K \in \{10, 20, 30, 40, 50, 60\}$, we repeat the experiment 500 times and count the probability of exact recovery. The performance curve is plotted in Fig. 2(a). Numerical values on the x-axis denote signal sparsity K while those on the y-axis denote the probability of exact recovery. We then repeat similar experiments when an input signal is sparse in some sparse and non-uniform basis $\boldsymbol{\Psi}$. Fig. 2(b) and Fig. 2(c) illustrate the performance curves when $\boldsymbol{\Psi}$ is the Daubechies-8 wavelet basis and the identity matrix, respectively.

These experiments verify that when performance of the SRM is comparable to that of a completely random matrix when the transform matrix $\boldsymbol{F}$ is dense (all of its entries are non-zero) or when the sparsifying matrix $\boldsymbol{\Psi}$ of the input signal is dense (e.g. DCT). This implies that if we know the signal is sparse in a dense domain $\boldsymbol{\Psi}$, we can sense the signal directly in its original domain (i.e., $\boldsymbol{F} = \boldsymbol{I}$) without performance loss. In addition, if we have no prior knowledge of a sparsifying transform, employing a SRM with the dense matrix $\boldsymbol{F}$ guarantees optimal performance.

However, when both sparsifying matrix and SRM are sparse, sensing performance might drop quickly as illustrated in Fig. 2(c), revealing a trade-off between sensing performance and streaming capacity. In this case, Fig. 2(b) shows that the SRM with the global randomizer seems to work much better than the SRM with the local randomizer.

TABLE II

SRMS EMPLOYED IN THE EXPERIMENT WITH COMPRESSIBLE SIGNALS

Notation	**R**	**F**
DCT32-G	Global randomizer	32×32 block diagonal DCT
WHT32-G	Global randomizer	32×32 block diagonal WHT
DCT512-L	Local randomizer	512×512 block diagonal DCT
WHT512-L	Local randomizer	512×512 block diagonal WHT

B. Simulation with Compressible Signals

In this simulation, signals of interest are natural images of size 512×512 such as the 512×512 Lena, Barbara and Boat images. The sparsifying basis $\mathbf{\Psi}$ used for these natural images is the well-known Daubechies $9/7$ wavelet transform. All images are implicitly regarded as 1-D signals of length 512^2. The GPSR software in [3] is used for signal reconstruction.

For such a large scale simulation, it takes a huge amount of system resources to implement the sensing method of a completely random matrix. Thus, for the purpose of benchmark, we adopt a more practical scheme of partial FFT in the wavelet domain (WPFFT). The WPFFT is to sense wavelet coefficients in the wavelet domain using the method of partial FFT. Theoretically, WPFFT has optimal performance as the Fourier matrix is completely incoherent with the identity matrix. The WPFFT is a method of sensing a signal in the transform domain that also requires substantial amount of system resources. SRMs under consideration are summarized in Table II.

For the purpose of comparison, we also implement two popular sensing methods: partial FFT in the time domain (PFFT)[1] and the Scrambled/Permutted FFT (SFFT) in [25], [26] that corresponds to a dense SRM using the global randomizer.

The performance curves of these sensing ensembles are plotted in Fig. 3(a), Fig. 3(b) and Fig. 3(c), which correspond to the input signal Lena, Barbara and Boat images, respectively. Numerical value on the x-axis represents sampling rate, which is the number of measurements over the total number of samples. Value on y-axis is the quality of reconstruction (PSNR in dB). Lastly, Fig. 4 shows the visually reconstructed 512×512 Lena image from 25% of measurements using WPFFT, WHT32-G and WHT512-L ensembles

As clearly seen in Fig. 3, the PFFT is not an efficient sensing matrix for smooth signals like images because Fourier matrix and wavelet basis are highly coherent. On the other hand, the SRM method, which can roughly be viewed as the PFFT preceded by the pre-randomization process, is very efficient. In particular, with a dense SRM like SFFT, the performance difference between the SRM method and the benchmark one, WPFFT, is less than 1 dB. In addition, performance of DCT512-L and WHT512-L that are fully streaming capable SRM, degrades about 1.5 dB, which is a reasonable sacrifice as the buffer size required is less than 0.2 percent of the total length of the original signal. Less degradation is obtainable when the buffer size is increased. Also, in all cases, there is no

observable difference of performance between DCT and normalized WHT transforms. It implies that orthonormal matrices whose entries have the same order of absolute magnitude generate comparable performance. In addition, highly sparse SRM using the global randomizer such as DCT32-G and WHT32-G has experimental performance comparable to that of the dense SRM. Note that these SRM are highly sparse because their density are only 2^{-13}. This observation again verifies that a SRM using the global randomizer might, in general, outperform a SRM using the local randomizer. We leave the theoretical justification of this observation for our future research.

VI. Discussion and Conclusion

A. Complexity Discussion

We compare the computation and memory complexity between the proposed SRM and other random sensing matrices such as Gaussian or Bernoulli i.i.d. matrices. In implementation, the i.i.d Bernoulli matrix is obviously preferred than i.i.d Gaussian one as the former has integer entries $1, -1$ and requires only 1 bit to represent each entry. A $M \times N$ i.i.d. Bernoulli sensing matrix requires MN bits for storing the matrix and MN additions and multiplications for sensing operation. A $M \times N$ structurally random matrix only requires $2N + N \log N$ bits for storage and $N + N \log N$ additions and multiplications for sensing operation. With SRM, its computational complexity and memory space required is independent with the number of measurements M. Note that with SRM, we do not need to store matrices $\boldsymbol{D}$, $\boldsymbol{F}$, $\boldsymbol{R}$ explicitly. We only need to store the diagonals of $\boldsymbol{D}$ and of $\boldsymbol{R}$ and the fast transform $\boldsymbol{F}$, resulting in significant saving of both memory space and computational complexity.

Sparse signal recovery algorithms often require to compute $\boldsymbol{A}$ and $\boldsymbol{A}^T$ in each iteration for reconstructing the original sparse signal $\boldsymbol{x}$ from the compressed measurement vector $\boldsymbol{y}$, where $\boldsymbol{A} = \boldsymbol{\Phi\Psi}$. Speed of these reconstruction algorithm often depends critically on whether matrix-vector multiplications $\boldsymbol{Au}$ and $\boldsymbol{A}^T\boldsymbol{u}$ can be computed quickly [3]. For the sake of simplicity, let's now assume Ψ is identity matrix. $\boldsymbol{Au} = \boldsymbol{\Phi u}$ requires $MN = \mathcal{O}(KN \log N)$ additions and multiplications for a random sensing matrix $\boldsymbol{\Phi}$ and $\mathcal{O}(N \log N)$ additions and multiplications for a SRM. This implies that SRM can speed up the reconstruction algorithm with at least K folds. With compressible signals (e.g., images), the number of measurements acquired tends to be proportional with the signal dimension, for example, $M = N/4$, then computational complexity reduction if using SRM is $\frac{N}{4 \log N}$.

Table III summarizes practical advantages of employing a SRM over a random sensing matrix.

B. Relationship with Other Related Works

When $\boldsymbol{R}$ is a local randomizer, the SRM matrix is a little reminiscent to the so-called Fast Johnson-Lindenstrauss Transform (FJLT) [27]. However, the SRM is much easier and less expensive to implement due to the simplicity of the matrix $\boldsymbol{D}$. In FJLT, this matrix is a completely random matrix with sparse distribution . It is unknown if there exists an efficient implementation of such a sparse random matrix. As a result, SRMs are more appropriate for practical applications because of their simple implementation and its optimal performance guarantee.

In [25], [26], the Scrambled/Permuted FFT is experimentally proposed as a heuristic low-complexity sensing method that is efficient for sensing of a large signal. To the best of our knowledge, however, there has not been

TABLE III

PRACTICAL FEATURE COMPARISON

Features	SRMs	Completely Random Matrices
No. of measurements for exact recovery	$\mathcal{O}(K \log N)$	$\mathcal{O}(K \log N)$
Sensing complexity	$N \log N$	$\mathcal{O}(KN \log N)$
Reconstruction complexity at each iteration	$\mathcal{O}(N \log N)$	$\mathcal{O}(KN \log N)$
Implementation in hardware and optics	Very easy	Difficult
Fast computability	Yes	No
Block-based processing	Yes	No

any theoretical analysis for Scrambled FFT. It turns out the Scrambled FFT is superseded by our unified SRM framework.

Random Convolution convolving the input signal with a random pulse followed by randomly subsampling measurements is proposed in [19] as a promising sensing method for large scale, real signals. Although there are a few other methods that exploit the same idea of convolving a signal with a random pulse, for examples: Random Filter in [17] and Toeplitz structured sensing matrix in [18], only random convolution method can be shown to approach optimal sensing performance. The main difference among the random convolution method and other similar methods is while other methods such as Random Filter and Toeplitz-based CS methods subsample measurements structurally, the random convolution method applies a technique of randomly subsampling measurements that is also employed in the proposed SRM framework. In addition, in random convolution, randomness is introduced in the Fourier domain by randomizing phases of Fourier coefficients. These techniques help to decouple stochastic dependence among measurements and thus, enabling us to establish stronger claims of sensing performance.

Although sharing a few common features, the proposed SRM framework is distinct from all aforementioned methods, including random convolution. One of major differences is that in SRM, signal pre-randomization is performed directly in its original domain (via the global randomizer or the local randomizer), rather than in Fourier domain as in the random convolution method. As a result, the sensing system becomes much simpler and less expensive to implement (without performance loss). In addition, it extends the random convolution method by verifying that not only Fourier transform but also a wide variety of other popular fast transforms, especially ones that are easier to implement such as WHT, can be used to obtain optimal performance. Last but not least, the SRM framework presents a systematic method to design optimal and flexible sensing matrices with practical features.

APPENDIX I

Central Limit Theorem. *Let $Z_1, Z_2, \ldots, Z_N$ be mutually independent random variables. Assume $E(Z_k) = 0$ and denote $\sigma^2 = \sum_{k=1}^{N} Var(Z_k)$. If for a given $\epsilon \geq 0$ and N sufficiently large, the following inequalities hold:*

$$Var(Z_k) < \epsilon\sigma^2 \quad k = 1, 2, \ldots, N$$

then distribution of the normalized sum $S = \sum_{k=1}^{N} Z_k$ *converges to* $\mathcal{N}(0, \sigma^2)$

Combinatorial Central Limit Theorem. *Given two sequences* $\{a_k\}_{k=1}^N$ *and* $\{b_k\}_{k=1}^N$. *Assume the* a_k *are not all equal and* b_k *are also not all equal. Let* $[\omega_1, \omega_2, \ldots, \omega_N]$ *be a uniform random permutation of* $[1, 2, ..., N]$. *Denote* $Z_k = a_{\omega_k}$ *and*

$$S = \sum_{k=1}^{N} Z_k b_k;$$

S *is asymptotically normally distributed* $\mathcal{N}(E(S), Var(S))$ *if*

$$\lim_{N\to\infty} N \frac{\max_{1\le k\le N}(Z_k - \overline{Z})^2}{\sum_{k=1}^{N}(Z_k - \overline{Z})^2} \frac{\max_{1\le k\le N}(b_k - \overline{b})^2}{\sum_{k=1}^{N}(b_k - \overline{b})^2} = 0;$$

where

$$\overline{b} = \frac{1}{N}\sum_{k=1}^{N} b_k \quad \textit{and} \quad \overline{Z} = \frac{1}{N}\sum_{k=1}^{N} Z_k.$$

APPENDIX II

Hoeffding's Concentration Inequality. *Suppose* $X_1, X_2, ..., X_N$ *are independent random variables and* $a_k \le X_K \le b_k$ $(k = 1, 2, ..., N)$. *Define a new random variable* $S = \sum_{k=1}^{N} X_k$. *Then for any* $t > 0$

$$P(|S - E(S)| \ge t) \le 2e^{-\frac{2t^2}{\sum_{k=1}^{N}(b_k - a_k)^2}}.$$

Ledoux's Concentration Inequality. *Let* $\{\eta_i\}_{1\le i\le N}$ *be a sequence of independent random variables such that* $|\eta_i| \le 1$ *almost surely and* $\boldsymbol{v}_1$, $\boldsymbol{v}_2$,..., $\boldsymbol{v}_N$ *be vectors in Banach space. Define a new random variable:* $S = \|\sum_{i=1}^{N} \eta_i v_i\|$. *Then for any* $t > 0$,

$$P(S \ge E(S) + t) \le 2\exp(-\frac{t^2}{16\sigma^2})$$

where σ^2 *denote the variance of* S *and* $\sigma^2 = \sup_{\|\boldsymbol{u}\|\le 1} \sum_{i=1}^{N} |\langle \boldsymbol{u}, \boldsymbol{v}_i\rangle|^2$.

Talagrand's Concentration Inequality. *Let* Z_k *be zero-mean i.i.d random variables and bounded* $|Z_k| \le \lambda$ *and* $\boldsymbol{u}_k$ *be column vectors of a matrix* $\boldsymbol{U}$. *Define a new random variable:* $S = \|\sum_{i=1}^{N} Z_k \boldsymbol{u}_k\|$. *Then for any* $t > 0$:

$$P(S \ge E(S) + t) \le 3\exp(-\frac{t}{cB}\log(1 + \frac{Bt}{\sigma^2 + BE(S)}))$$

where c *is some constant, variance* $\sigma^2 = E(Z_k^2)\|\boldsymbol{U}\|^2$ *and* $B = \lambda \max_{1\le k\le N} \|\boldsymbol{u}_k\|$.

Sourav's Concentration Inequality. *Let* $\{Z_{ij}\}_{1\le i,j\le N}$ *be a collection of numbers from* $[0, 1]$. *Let* $[\omega_1, \omega_2, \ldots, \omega_N]$ *be a uniformly random permutation of* $[1, 2, \ldots, N]$. *Define a new random variable:* $S = \sum_{i=1}^{N} Z_{i\omega_i}$. *Then for any* $t \ge 0$

$$P(|S - E(S)| \ge t) \le 2\exp(-\frac{t^2}{4E(S) + 2t}).$$

REFERENCES

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

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, pp. 1289 – 1306, Apr. 2006.

[3] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction," *to appear in IEEE Journal of Selected Topics in Signal Processing*, 2007.

[4] Elaine T. Hale, Wotao Yin, , and Yin Zhang, "A fixed-point continuation method for '1-regularized minimization with applications to compressed sensing," *Technical Report*, Jul 2007.

[5] Ewout V. D. Berg and Michael P. Friedlander, "Probing the pareto frontier for basis purusit solutions," *Technical Report*, Jan 2008.

[6] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit.," *IEEE Trans. Info. Theory*, vol. 53, pp. 4655–4666, Dec 2007.

[7] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Preprint*, Mar 2008.

[8] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing: Closing the gap between performance and complexity," *Preprint*, Mar 2008.

[9] D. L. Donoho, Y. Tsaig, and Jean-Luc Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *Technical Report*, Mar. 2006.

[10] Thong T. Do, Lu Gan, Nam Nguyen, and Trac D. Tran, "Sparsity adaptive matching pursuit algorithm for practical compressed sensing," *Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, California*, Oct 2008.

[11] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. on Information Theory*, vol. 47(7), Nov. 2001.

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

[13] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Uniform uncertainty principle for bernoulli and subgaussian ensembles.," *Preprint*, Aug. 2006.

[14] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23(3), pp. 969–985, 2007.

[15] R. Coifman, F. Geshwind, and Y. Meyer, "Noiselets," *Appl. Comp. Harmonic Analysis*, vol. 10, pp. 27–44, 2001 2005.

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

[17] J. Tropp, M. Wakin, M. Duarte, D. Baron, and R. Baraniuk, "Random filters for compressive sampling and reconstruction," *Proc. IEEE ICASSP*, vol. 3, pp. 872–875, Toulouse, May 2006.

[18] Waheed Bajwa, Jarvis Haupt, Gil Raz, Stephen Wright, and Robert Nowak, "Toeplitz-structured compressed sensing matrices.," *IEEE Workshop on Statistical Signal Processing (SSP), Madison, Wisconsin.*, Aug 2007.

[19] Justin Romberg, "Compressive sensing by random convolution.," *Preprint*, Jul 2008.

[20] W. Hoeffding, "A combinatorial central limit theorem.," *The Annals of Mathematical Statistics*, vol. 22(4), pp. 558–566, Dec. 1951.

[21] K. Schnass and P. Vandergheynst, "Average performance analysis for thresholding," *IEEE Signal Processing Letters*, vol. 14, Nov 2007.

[22] M. Ledoux, "The concentration of measure phenomenon," *American Mathematical Society*, 2001.

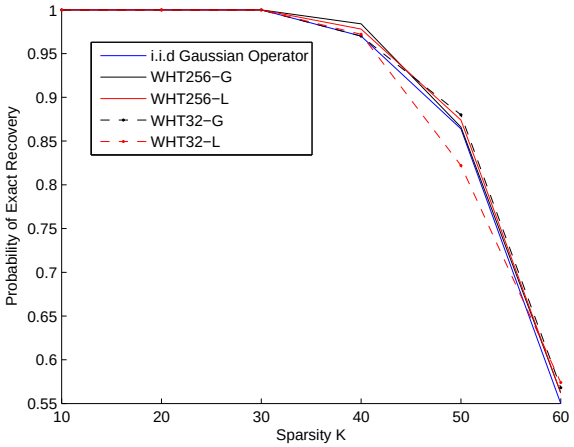
[23] Sourav Chatterjee, "Stein's method for concentration inequalities," *Probab. Theory Related Fields*, vol. 138, pp. 305–321, 2007.

[24] M. Talagrand, "New concentration inequalities in product spaces," *Invent. Math.*, vol. 126, pp. 505–563, 1996.

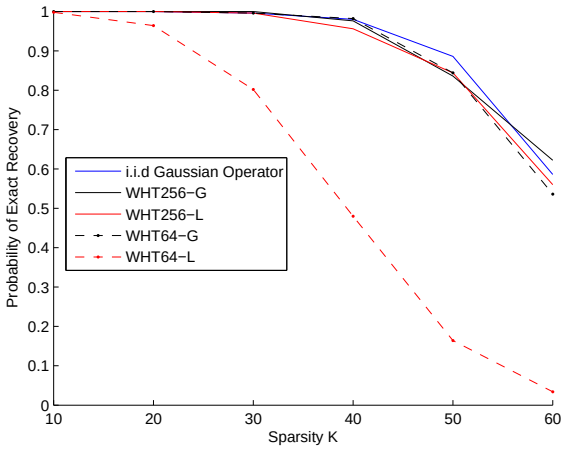
[25] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, pp. 1207–1223, Aug. 2006.

[26] M. F. Duarte, M. B. Wakin, and R. G. Baraniuk, "Fast reconstruction of piecewise smooth signals from incoherent projections," *SPARS'05*, Rennes, France, Nov 2005.

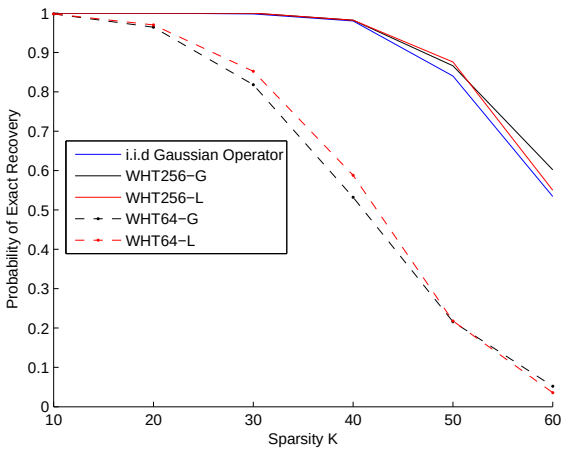
[27] Nir Ailon and Bernard Chazelle, "Approximate nearest neighbors and the fast johnsonlindenstrauss transform," *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, vol. 66, pp. 557 – 563, 2006.



(a)

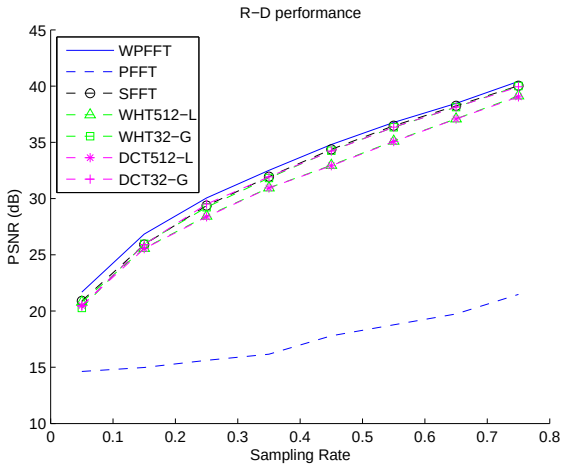


(b)

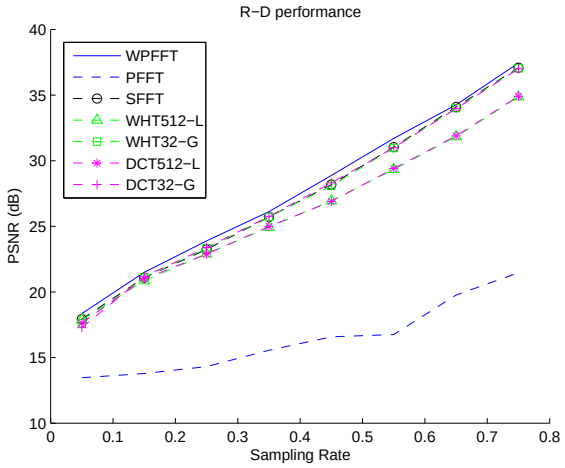


(c)

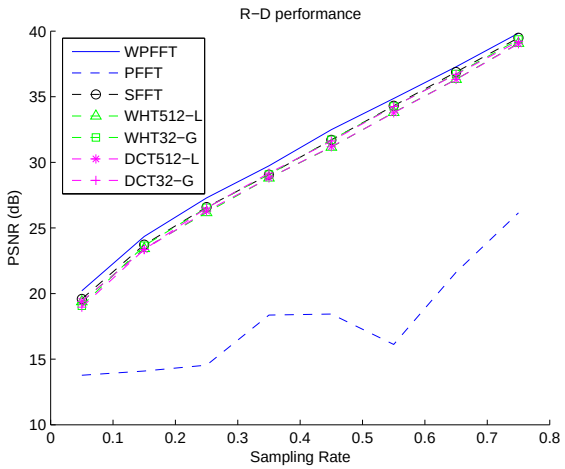
Fig. 2. Performance curves: probability of exact recovery vs. Sparsity K. (a) when $\mathbf{\Psi}$ is IDCT basis. (b) when $\mathbf{\Psi}$ is Daubechies-8 wavlet basis. (c) when $\mathbf{\Psi}$ is the identity basis



(a)



(b)



(c)

Fig. 3. Performance curves: Quality of signal reconstruction vs. sampling rate M/N. (a) the 512×512 Lena image. (b) the 512×512 Barbara image. (c) the 512×512 Boat image

(a) (b)

(c) (d)

Fig. 4. Reconstructed 512×512 *Lena* images from $M/N = 25\%$ sampling rate. (a) The original Lena image; (b) using the WPFFT ensemble: 30.1dB; (c) using the WHT32-G ensemble: 29.3dB; (d) using the WHT512-L: 28.5dB