

Construction of a Large Class of Deterministic Sensing Matrices that Satisfy a Statistical Isometry Property

Robert Calderbank, *Fellow, IEEE*, Stephen Howard, *Member, IEEE*, and Sina Jafarpour, *Student Member, IEEE*

Abstract—Compressed Sensing aims to capture attributes of k -sparse signals using very few measurements. In the standard Compressed Sensing paradigm, the $N \times C$ measurement matrix Φ is required to act as a near isometry on the set of all k -sparse signals (Restricted Isometry Property or RIP). If Φ satisfies the RIP, then Basis Pursuit or Matching Pursuit recovery algorithms can be used to recover any k -sparse vector α from the N measurements $\Phi\alpha$. Although it is known that certain probabilistic processes generate $N \times C$ matrices that satisfy RIP with high probability, there is no practical algorithm for verifying whether a given sensing matrix Φ has this property, crucial for the feasibility of the standard recovery algorithms. In contrast this paper provides simple criteria that guarantee that a deterministic sensing matrix satisfying these criteria acts as a near isometry on an overwhelming majority of k -sparse signals; in particular, most such signals have a unique representation in the measurement domain. Probability still plays a critical role, but it enters the signal model rather than the construction of the sensing matrix. An essential element in our construction is that we require the columns of the sensing matrix to form a group under pointwise multiplication. The construction allows recovery methods for which the expected performance is sub-linear in C , and only quadratic in N , as compared to the super-linear complexity in C of the Basis Pursuit or Matching Pursuit algorithms; the focus on expected performance is more typical of mainstream signal processing than the worst-case analysis that prevails in standard Compressed Sensing. Our framework encompasses many families of deterministic sensing matrices, including those formed from discrete chirps, Delsarte-Goethals codes, and extended BCH codes.

Index Terms—Deterministic Compressed Sensing, Statistical Near Isometry, Finite Groups, Martingale Sequences, McDiarmid Inequality, Delsarte-Goethals Codes.

I. INTRODUCTION AND NOTATIONS

The central goal of compressed sensing is to capture attributes of a signal using very few measurements. In most work to date, this broader objective is exemplified by the important special case in which a k -sparse vector $\alpha \in \mathbb{R}^C$ (with C large) is to be reconstructed from a small number N of linear measurements with $k < N < C$. In this problem, the measurement data constitute a vector $f = N^{-1/2} \Phi\alpha$, where Φ is an $N \times C$ matrix called the *sensing matrix*. Throughout this paper we shall use the notation φ_j for the j -th column of the sensing matrix Φ ; its entries will be denoted by $\varphi_j(x)$ (with label x varying from 1 to N). In other words, $\varphi_j(x)$ is the x -th row and j -th column element of Φ .

The work of R. Calderbank and S. Jafarpour is supported in part by NSF under grant DMS 0701226, by ONR under grant N00173-06-1-G006, and by AFOSR under grant FA9550-05-1-0443

The two fundamental questions in compressed sensing are: how to construct suitable sensing matrices Φ , and how to recover α from f efficiently; it is also of practical importance to be resilient to measurement noise and to be able to reconstruct (approximations to) k -compressible signals, i.e. signals that have more than k nonvanishing entries, but where only k entries are significant and the remaining entries are close to zero.

The work of Donoho [9] and of Candès, Romberg and Tao [10], [2], [11] provides fundamental insight into the geometry of sensing matrices. This geometry is expressed by e.g. the Restricted Isometry Property (RIP), formulated by Candès and Tao [10]: a sensing matrix satisfies the k -Restricted Isometry Property if it acts as a near isometry on all k -sparse vectors; to ensure unique and stable reconstruction of k -sparse vectors, it is sufficient that Φ satisfy $2k$ -RIP. When N/C and/or k/N are (very) small, deterministic RIP matrices have been constructed using methods from approximation theory [12] and coding theory [13]. More attention has been paid to probabilistic constructions where the entries of the sensing matrix are generated by an i.i.d Gaussian or Bernoulli process or from random Fourier ensembles, in which larger values of N/C and/or k/N can be considered. These sensing matrices are known to satisfy the k -RIP with high probability [9], [10] and the number N of measurements is $k \log \frac{C}{k}$. This is best possible in the sense that approximation results of Kashin [14] and Glushin [15] imply that $\Omega(k \log \frac{C}{k})$ measurements are required for sparse reconstruction using ℓ_1 -minimization methods. Constructions of random sensing matrices of similar size that have the RIP but require a smaller degree of randomness, are given by several approaches including filtering [16], [17] and expander graphs [18], [1], [6], [5].

The role of random measurement in compressive sensing can be viewed as analogous to the role of random coding in Shannon theory. Both provide worst case performance guarantees in the context of an adversarial signal/error model. Random sensing matrices are easy to construct, and are $2k$ -RIP with high probability. As in coding theory, this randomness has its drawbacks, briefly described as follows:

- First, efficiency in sampling comes at the cost of complexity in reconstruction (see Table 1) and at the cost of error in signal approximation (see Section 5).
- Second, storing the entries of a random sensing matrix may require significant space, in contrast to deterministic matrices where the entries can often be computed on the fly without requiring any storage.

TABLE I

PROPERTIES OF k -SPARSE RECONSTRUCTION ALGORITHMS THAT EMPLOY RANDOM SENSING MATRICES WITH N ROWS AND \mathcal{C} COLUMNS. THE PROPERTY RIP-1 IS THE COUNTERPART OF RIP FOR THE ℓ_1 METRIC AND IT PROVIDES GUARANTEES ON THE PERFORMANCE OF SPARSE RECONSTRUCTION ALGORITHMS THAT EMPLOY LINEAR PROGRAMMING [1]. NOTE THAT EXPLICIT CONSTRUCTION OF THE EXPANDER GRAPHS REQUIRES A LARGE NUMBER OF MEASUREMENTS, AND THAT MORE PRACTICAL ALTERNATIVES ARE RANDOM SPARSE MATRICES WHICH ARE EXPANDERS WITH HIGH PROBABILITY.

Approach	Number of Measurements N	Complexity	Compressible Signals	Noise Resilience	RIP
Basis Pursuit (BP) [2]	$k \log\left(\frac{\mathcal{C}}{k}\right)$	\mathcal{C}^3	Yes	Yes	Yes
Orthogonal Matching Pursuit (OMP) [3]	$k \log^\alpha(\mathcal{C})$	$k^2 \log^\alpha(\mathcal{C})$	Yes	No	Yes
Group Testing [4]	$k \log^\alpha(\mathcal{C})$	$k \log^\alpha(\mathcal{C})$	Yes	No	No
Expanders (Unique Neighborhood) [5]	$k \log\left(\frac{\mathcal{C}}{N}\right)$	$\mathcal{C} \log\left(\frac{\mathcal{C}}{N}\right)$	Yes ^o	Yes ^o	RIP-1
Expanders (BP) [1]	$k \log\left(\frac{\mathcal{C}}{k}\right)$	\mathcal{C}^3	Yes	Yes	RIP-1
Expander Matching Pursuit (EMP) [6]	$k \log\left(\frac{\mathcal{C}}{k}\right)$	$\mathcal{C} \log\left(\frac{\mathcal{C}}{k}\right)$	Yes ^o	Yes ^o	RIP-1
Sparse Matching Pursuit (SMP) [6]	$k \log\left(\frac{\mathcal{C}}{k}\right)$	$\mathcal{C} \log\left(\frac{\mathcal{C}}{k}\right)$	Yes ^o	Yes ^o	RIP-1
CoSaMP [7]	$k \log\left(\frac{\mathcal{C}}{k}\right)$	$\mathcal{C}k \log\left(\frac{\mathcal{C}}{k}\right)$	Yes	Yes	Yes
SSMP [8]	$k \log\left(\frac{\mathcal{C}}{k}\right)$	$\mathcal{C}k \log\left(\frac{\mathcal{C}}{k}\right)$	Yes	Yes	Yes

^o [5] provides an algorithm with smaller constants that is easier to implement and analyze, whereas [6] is able to handle more general noise models.

• Third, there is no algorithm for efficiently verifying whether a sampled sensing matrix satisfies RIP, a condition that is essential for the recovery guarantees of the Basis Pursuit and Matching Pursuit algorithms on *any* sparse signal.

These drawbacks lead us to consider constructions with deterministic sensing matrices, for which the performance is guaranteed in expectation only, for k -sparse signals that are random variables, but which do not suffer from the same drawbacks. The framework presented here provides

- easily checkable conditions on special types of deterministic sensing matrices guaranteeing successful recovery of *all but an exponentially small fraction* of k -sparse signals;
- in many examples, the entries of these matrices can be computed on the fly without requiring any storage, and
- recovery algorithms with lower complexities than Basis Pursuit and Matching Pursuit algorithms.

To make this last point more precise, we note that Basis Pursuit and Matching Pursuit algorithms rely heavily on *matrix-vector* multiplication, and are super-linear with respect to \mathcal{C} , the dimension of the data domain. The reconstruction algorithm for the framework presented here (see Section 5) requires only *vector-vector* multiplication in the measurement domain; as a result, its recovery time is only quadratic in the dimension N of the measurement domain. We suggest that the role of the deterministic measurement matrices presented here for compressive sensing is analogous to the role of structured codes in communications practice: in both cases fast encoding and decoding algorithms are emphasized, and typical rather than worst case performance is optimized. We are not the only ones seeking inspiration in coding theory to construct deterministic matrices for compressed sensing; Table 2 gives

an overview of approaches in the literature that employ deterministic sensing matrices, several of which are based on linear codes (cf. [19] and [21]) and provide expected-case rather than worst-case performance guarantees. It is important to note (see Table 2) that although the use of linear codes makes fast algorithms possible for sparse reconstruction, these are not always resilient to noise. Such non-resilience manifests itself in e.g. Reed-Solomon (RS) constructions [21]; the RS reconstruction algorithm (the roots of which go back to 1795! – see [26], [27]) uses the input data to construct an error-locator polynomial; the roots of this polynomial identify the signals appearing in the sparse superposition. Because the correspondence between the coefficients of a polynomial and its roots is not well conditioned, it is very difficult to deal with compressible signals and noisy measurements in RS-based approaches.

Because we will be interested in expected-case performance only, we need not impose RIP; we shall instead work with the weaker Statistical Restricted Isometry Property. More precisely, we define

Definition 1. ((k, ϵ, δ) -StRIP matrix)

An $N \times \mathcal{C}$ (sensing) matrix Φ is said to be a (k, ϵ, δ) -Statistical Restricted Isometry Property matrix [abbreviated (k, ϵ, δ) -StRIP matrix] if, for k -sparse vectors $\alpha \in \mathbb{R}^{\mathcal{C}}$, the inequalities

$$(1 - \epsilon) \|\alpha\|^2 \leq \left\| \frac{1}{\sqrt{N}} \Phi \alpha \right\|^2 \leq (1 + \epsilon) \|\alpha\|^2, \quad (1)$$

hold with probability exceeding $1 - \delta$ (with respect to a uniform distribution of the vectors α among all k -sparse

TABLE II

PROPERTIES OF k -SPARSE RECONSTRUCTION ALGORITHMS THAT EMPLOY DETERMINISTIC SENSING MATRICES WITH N ROWS AND \mathcal{C} COLUMNS. NOTE THAT FOR LDPC CODES $k \ll \mathcal{C}$. NOTE ALSO THAT RIP HOLDS FOR RANDOM MATRICES WHERE IT IMPLIES EXISTENCE OF A LOW-DISTORTION EMBEDDING FROM ℓ_2 INTO ℓ_1 . GURUSWAMI ET AL. [18] PROVED THAT THIS PROPERTY ALSO HOLDS FOR DETERMINISTIC SENSING MATRICES CONSTRUCTED FROM EXPANDER CODES. IT FOLLOWS FROM THEOREM 8 IN THIS PAPER THAT SENSING MATRICES BASED ON DISCRETE CHIRPS AND DELSARTE-GOETHALS CODES SATISFY THE USTRIP.

Approach	Number of Measurements N	Complexity	Compressible Signals	Noise Resilience	RIP
Low Density Parity Check Codes (LDPC) [19]	$k \log \mathcal{C}$	$\mathcal{C} \log \mathcal{C}$	Yes	Yes	No
Low Density Parity Check Codes (LDPC) [20]	$k \log \left(\frac{\mathcal{C}}{k}\right)$	\mathcal{C}	Yes	Yes	No
Reed-Solomon codes [21]	k	k^2	No	No	No
Explicit Construction of Expander Graphs [22]	\mathcal{C}	\mathcal{C}	Yes	Yes	No
Embedding ℓ_2 into ℓ_1 (BP) [18]	$k(\log \mathcal{C})^{\alpha \log \log \mathcal{C}}$	\mathcal{C}^3	Yes	No	No
Extractors [13]	$k\mathcal{C}^{\alpha(1)}$	$k\mathcal{C}^{\alpha(1)} \log(\mathcal{C})$	No	No	No
Discrete chirps [23]	$\sqrt{\mathcal{C}}$	$kN \log N$	Yes	Yes	UStRIP
Delsarte-Goethals codes This Paper, [24], [25]	$k \log \mathcal{C}$	$k^2 \log^{2+\alpha(1)} \mathcal{C}$	Yes	Yes	UStRIP

vectors in $\mathbb{R}^{\mathcal{C}}$ of the same norm).¹

There is a slight wrinkle in that, unlike the simple RIP case, StRIP does not automatically imply unique reconstruction, not even with high probability. If an $N \times \mathcal{C}$ matrix Φ is $(2k, \epsilon, \delta)$ -StRIP, then, given a k -sparse vector α , it does follow that Φ maps any other randomly picked k -sparse signal β to a different image, i.e. $\Phi \alpha \neq \Phi \beta$, with probability exceeding $1 - \delta$ (with respect to the random choice of β). This does not mean, however, that uniqueness is guaranteed with high probability: requiring that the measure of $\{\alpha \in \mathbb{R}^{\mathcal{C}}; \alpha \text{ is } k\text{-sparse and there is a different } k\text{-sparse } \beta \in \mathbb{R}^{\mathcal{C}} \text{ for which } \Phi \alpha = \Phi \beta\}$ be small, is a more stringent requirement than that the measure of $\{\beta \in \mathbb{R}^{\mathcal{C}}; \beta \neq \alpha \text{ and } \Phi \alpha = \Phi \beta\}$ be small for all k -sparse α . For this reason, we also introduce the following definition:

Definition 2. ((k, ϵ, δ) -UStRIP matrix)

An $N \times \mathcal{C}$ (sensing) matrix Φ is said to be a (k, ϵ, δ) -Uniqueness-guaranteed Statistical Restricted Isometry Property matrix [abbreviated (k, ϵ, δ) -UStRIP matrix] if Φ is a (k, ϵ, δ) -StRIP matrix, and

$$\{\beta \in \mathbb{R}^{\mathcal{C}}; \Phi \alpha = \Phi \beta\} = \{\alpha\}$$

with probability exceeding $1 - \delta$ (with respect to a uniform distribution of the vectors α among all k -sparse vectors in $\mathbb{R}^{\mathcal{C}}$ of the same norm).

Again, we are not the first to propose a weaker version of RIP that permits the construction of deterministic sensing matrices. The construction by Guruswami et al. in [18] can

be viewed as another instance of a weakening of RIP, in the following different direction. RIP implies that Φ defines a low-distortion ℓ_2 - ℓ_1 -embedding that plays a crucial role in the proofs of [9], [10], [2], [11]. In [18], Guruswami et al. prove that this ℓ_2 - ℓ_1 -embedding property also holds for deterministic sensing matrices constructed from expander codes. These matrices satisfy an ‘‘almost Euclidean null space property’’ property, that is for any α in the null space of Φ , $\frac{\sqrt{N}\|\alpha\|_2}{\|\alpha\|_1}$ is bounded by a constant; this is their main tool to obtain the results reported in Table 2.

In this paper we formulate simple design rules, imposing that the columns of the sensing matrix form a group under pointwise multiplication, that all row sums vanish, that different rows are orthogonal, and requiring a simple upper bound on the absolute value of any column sum (other than the multiplicative identity). The properties we require are satisfied by a large class of matrices constructed by exponentiating codewords from a linear code; several examples are given in Section 2. In Sections 3, we show that our relatively weak design rules are sufficient to guarantee that Φ is UStRIP, provided the parameters satisfy certain constraints. The group property makes it possible to avoid intricate combinatorial reasoning about coherence of collections of mutually unbiased bases (cf. [28]). Section 4 applies our results to the case where the sensing matrix is formed by taking random rows of the FFT matrix. In Section 5 we emphasize a particular family of constructions involving subcodes of the second order Reed-Muller code; in this case codewords correspond to multivariable quadratic functions defined over the binary field or the integers modulo 4. Section VI provides a discussion regarding the noise resilience.

¹Throughout the paper norms without subscript denote ℓ_2 -norms

II. STRIP-ABLE: BASIC DEFINITIONS, WITH SEVERAL EXAMPLES

In this section we formulate three basic conditions and give examples of deterministic sensing matrices Φ with N rows and \mathcal{C} columns that satisfy these conditions. Note that throughout the paper, we shall assume (without stating this again explicitly) that Φ has no repeated columns.

Definition 3. An $N \times \mathcal{C}$ -matrix Φ is said to be η -**StRIP-able**, where η satisfies $0 < \eta \leq 1$, if the following three conditions are satisfied:

- **(St1)** The rows of Φ are orthogonal, and all the row sums are zero. i.e.

$$\sum_{j=1}^{\mathcal{C}} \varphi_j(x) \overline{\varphi_j(y)} = 0 \text{ if } x \neq y \quad (2)$$

$$\sum_{j=1}^{\mathcal{C}} \varphi_j(x) = 0, \text{ for all } x. \quad (3)$$

- **(St2)** The columns of Φ form a group under ‘‘pointwise multiplication’’, defined as follows

$$\begin{aligned} &\text{for all } j, j' \in \{1, \dots, \mathcal{C}\}, \\ &\text{there exists a } j'' \in \{1, \dots, \mathcal{C}\} \text{ such that} \\ &\text{for all } x : \varphi_j(x) \varphi_{j'}(x) = \varphi_{j''}(x). \end{aligned} \quad (4)$$

In particular, there is one column of Φ for which all the entries are 1, and that acts as a unit for this group operation; this column will be denoted by $\mathbf{1}$. Without loss of generality, we will assume the columns of Φ are ordered so that $\varphi_1 = \mathbf{1}$, i.e. $\varphi_1(x) = 1$ for all x .

- **(St3)** For all $j \in \{2, \dots, \mathcal{C}\}$,

$$\left| \sum_x \varphi_j(x) \right|^2 \leq N^{2-\eta}. \quad (5)$$

Remarks

1. Condition (5) applies to all columns *except* the first column (i.e. the column which consists of all ones).
2. The justification of the name *StRIP-able* will be given in the next section.
3. When the value of η in (5) does not play a special role, we just don't spell it out explicitly, and simply call Φ StRIP-able.

The conditions (2-5) have the following immediate consequences:

Lemma 4. *If the matrix Φ satisfies (4), then $|\varphi_j(x)| = 1$, for all j and all x .*

Proof: For every x , $(\varphi_j(x))_{j \in \{1, \dots, \mathcal{C}\}}$ is a group of complex numbers under multiplication; all finite groups of this type consist of unimodular numbers. ■

Lemma 5. *If the matrix Φ satisfies (4), then the collection of columns of Φ is closed under complex conjugation, i.e. for all $j \in \{1, \dots, \mathcal{C}\}$, there exists a $j' \in \{1, \dots, \mathcal{C}\}$*

$$\text{such that, for all } x, \quad \varphi_{j'}(x) = \overline{\varphi_j(x)}. \quad (6)$$

Proof: Pick $j \in \{1, \dots, \mathcal{C}\}$. Since the columns of Φ form a group under pointwise multiplication, there is some $j' \in$

$\{1, \dots, \mathcal{C}\}$ such that $\varphi_{j'}$ is the inverse of φ_j for this group operation. Using Lemma 4, we have then, for all x , $\varphi_{j'}(x) = [\varphi_j(x)]^{-1} = \overline{\varphi_j(x)}$. ■

Lemma 6. *If the matrix Φ satisfies (2), (3) and (4), then the normalized columns $(N^{-1/2} \varphi_j)_{j \in \{1, \dots, \mathcal{C}\}}$ form a tight frame in \mathbb{C}^N , with redundancy \mathcal{C}/N .*

Proof: By Lemma 4 and (2), we have

$$(\Phi \Phi^\dagger)_{x,y} = \sum_{j=1}^{\mathcal{C}} \varphi_j(x) \overline{\varphi_j(y)} = \mathcal{C} \delta_{x,y}$$

i.e. $\Phi \Phi^\dagger = \mathcal{C} \mathbf{I}_N$, so that, for any vector $v \in \mathbb{C}^N$,

$$\sum_{j=1}^{\mathcal{C}} |\langle v, \varphi_j \rangle|^2 = v \Phi \Phi^\dagger v^\dagger = \mathcal{C} \|v\|^2. \quad \blacksquare$$

Lemma 7. *If the matrix Φ satisfies (4), then the inner product of two columns φ_j and $\varphi_{j'}$, defined as $\varphi_j \cdot \varphi_{j'} := \sum_x \varphi_j(x) \overline{\varphi_{j'}(x)}$, equals N if and only if $j = j'$.*

Proof:

If $j = j'$, we obviously have $\varphi_j \cdot \varphi_{j'} = N$, by Lemma 4. If $\varphi_j \cdot \varphi_{j'} = N$, then we have, by Cauchy-Schwarz,

$$N = \varphi_j \cdot \varphi_{j'} \leq |\varphi_j \cdot \varphi_{j'}| \leq \|\varphi_j\| \|\varphi_{j'}\| = N,$$

implying that in this instance the Cauchy-Schwarz inequality must be an equality, so that $\varphi_{j'}$ must be some multiple of φ_j . Since $N = \varphi_j \cdot \varphi_{j'}$, the multiplication factor must equal 1, so that $\varphi_j = \varphi_{j'}$. Since Φ has no repeated columns, $j = j'$ follows. ■

We shall prove that StRIP-able matrices have (as their name already announces) a Restricted Isometry Property in a Statistical sense, provided the different parameters satisfy certain constraints, which will be made clear and explicit in the next section. Before we embark on that mathematical analysis, we show that there are many examples of StRIP-able matrices.

A. Discrete Chirp Sensing Matrices

Let p be a prime and let ω be a primitive (complex) p^{th} root of unity. A length p chirp signal takes the form

$$\varphi_{mp+r}(x) = \omega^r \omega^{mx+rx^2} \text{ where } x = 0, 1, \dots, p-1.$$

Here m is the *base frequency* and r is the *chirp rate*. Consider now the family of chirp signals (φ_{mp+r}) where $r, m = 0, 1, \dots, p-1$; the ‘‘extra’’ phase factor (usually not present in chirps) ensures that the row sums $\sum_{\ell=0}^{p^2-1} \varphi_\ell(x)$ vanish for all x . It is easy to check that this family satisfies (St1), (St2), and (St3) [23]. For the corresponding sensing matrix Φ , Applebaum et al. [23] have analyzed an algorithm for sparse reconstruction that exploits the efficiency of the FFT in each of two steps: the first to recover the chirp rate and the second to recover the base frequency. The Gerschgorin Circle Theorem [29] is used to prove that the RIP holds for sets of $\frac{\sqrt{p+1}}{2}$ columns. Numerical experiments reported in [23] compare the eigenvalues of deterministic chirp sensing

matrices with those of random Gaussian sensing matrices. The singular values of restrictions to k -dimensional subspaces of $N \times \mathcal{C}$ random Gaussian sensing matrices have a gaussian distribution, with mean $\mu_{N,\mathcal{C},k}$ and standard deviation $\sigma_{N,\mathcal{C},k}$; the experiments show that, for the same values of N , \mathcal{C} and k , the singular values of restrictions of deterministic chirp sensing matrices have a similar spread around a central value $\mu \in (\mu_{N,\mathcal{C},k}, 1)$ that is closer to 1; in fact, the experiments suggest that $\mu - \mu_{N,\mathcal{C},k} > \sigma_{N,\mathcal{C},k}$.

B. Kerdock, Delsarte-Goethals and Second Order Reed Muller Sensing Matrices

In our construction of deterministic sensing matrices based on Kerdock, Delsarte-Goethals and second order Reed Muller codes, we start by picking an odd number m . The 2^m rows of the sensing matrix Φ are indexed by the binary m -tuples x , and the $2^{(r+2)m}$ columns are indexed by the pairs P, b , where P is an $m \times m$ binary symmetric matrix in the Delsarte-Goethals set $DG(m, r)$, and b is a binary m -tuple. The entry $\varphi_{P,b}(x)$ is given by

$$\varphi_{P,b}(x) = i^{wt(d_P)+2wt(b)} i^{xPx^\top + 2bx^\top} \quad (7)$$

where d_p denotes the main diagonal of P , and wt denotes the *Hamming weight* (the number of 1s in the binary vector). Note that all arithmetic in the expressions $xPx^\top + 2bx^\top$ and $wt(d_P) + 2wt(b)$ takes place in the ring of integers modulo 4, since they appear only as exponents for i . Given P, b the vector $xPx^\top + 2bx^\top$ is a codeword in the Delsarte-Goethals code (defined over the ring of integers modulo 4) For a fixed matrix P , the 2^m columns $\varphi_{P,b}$, $b \in \mathbb{F}_2^m$ form an orthonormal basis Γ_P that can also be obtained by postmultiplying the Walsh-Hadamard basis by the unitary transformation $\text{diag}[i^{xPx^\top}]$.

The Delsarte-Goethals set $DG(m, r)$ is a binary vector space containing $2^{(r+1)m}$ binary symmetric matrices with the property that the difference of any two distinct matrices has rank at least $m - 2r$ (See [30]). The Delsarte-Goethals sets are nested

$$DG(m, 0) \subset DG(m, 1) \subset \dots \subset DG(m, (m-1)/2).$$

The first set $DG(m, 0)$ is the classical Kerdock set, and the last set $DG(m, (m-1)/2)$ is the set of all binary symmetric matrices. The r^{th} Delsarte-Goethals sensing matrix is determined by $DG(m, r)$ and has $N = 2^m$ rows and $\mathcal{C} = 2^{(r+2)m}$ columns. The initial phase in (7) is chosen so that the Delsarte-Goethals sensing matrices satisfy (St1) and (St2). (See Appendix A).

Coherence between orthonormal bases Γ_P and Γ_Q indexed by binary symmetric matrices P and Q is determined by the rank R of the binary matrix $P \oplus Q$ (See Appendix A). Any vector in one of the orthonormal bases has inner product of absolute value $2^{-R/2}$ with 2^R vectors in the other basis and is orthogonal to the remaining basis vectors. The column sums in this r^{th} Delsarte-Goethals sensing matrix satisfy

$$\left| \sum_x \varphi_{P,b}(x) \right|^2 = 0 \text{ or } N^{2-r/m},$$

so that condition (St3) is trivially satisfied. Details are provided in Appendix A; we refer the interested reader to [31], [32], [30] and Chapter 15 of [33] for more information about subcodes of the second order Reed-Muller code.

C. BCH Sensing Matrices

The Carlitz- Uchiyama Bounds (See Chapter 9 of [33]) imply that the interval

$$\left[2^{m-1} - (t-1)2^{m/2}, 2^{m-1} + (t-1)2^{m/2} \right]$$

contains all non-zero weights in the dual of the extended binary BCH code $BCH(m, t)$ of length $N = 2^m$ and designed distance $e = 2t + 1$, with the exception of $wt(\mathbf{1}) = N$. Setting $BCH(m, t)^\perp = \langle \mathbf{1} \rangle \oplus C_{m,t}$, the columns of the t^{th} BCH sensing matrix are obtained by exponentiating the codewords in $C_{m,t}$. The column determined by the codeword $c = (c_j)$ is given by

$$\varphi_c(j) = (-1)^{bc^\top} (-1)^{c_j}, \text{ where } j = 0, 1, \dots, 2^m - 1,$$

and where b is any vector not orthogonal to $C_{m,t}$. Conditions (St1) and (St2) hold by construction and

$$\begin{aligned} \left| (-1)^{bc^\top} \sum_{j=0}^{2^m-1} (-1)^{c_j} \right|^2 &= |N - 2wt_H(c)|^2 \\ &\leq \left[2(t-1)2^{m/2} \right]^2 \end{aligned}$$

so that (St3) holds. These sensing matrices have been analyzed by Ailon and Liberty [34].

In the binary case, the column sums take the form $N - 2w$ where w is the Hamming weight of the exponentiated codeword, and a similar interpretation is possible for codes that are linear over the ring of integers modulo 4 (see [30]). Property (St3) connects the Hamming geometry of the code domain, as captured by the weight enumerator of the code, with the geometry of the complex domain.

III. IMPLICATIONS FOR DETERMINISTIC STRIP-ABLE SENSING MATRICES: MAIN RESULT

In this section we prove our main result, namely that if Φ satisfies (St1), (St2) and (St3), then Φ is UStRIP, under certain fairly weak conditions on the parameters. More precisely,

Theorem 8. *Suppose the $N \times \mathcal{C}$ matrix Φ is η -StRIP-able, and suppose $k < 1 + (\mathcal{C} - 1)\epsilon$ and $\eta > 1/2$. Then there exists a constant c such that, if $N \geq \left(c \frac{k \log \mathcal{C}}{\epsilon^2} \right)^{\frac{1}{\eta}}$, then Φ is (k, ϵ, δ) -UStRIP with $\delta := 2 \exp \left[- \frac{[\epsilon - (k-1)/(\mathcal{C}-1)]^2 N^\eta}{8k} \right]$.*

The proof of Theorem 8 has two parts: we shall first, in Section 3.1, prove that Φ is StRIP; when this is established we turn our attention to proving UStRIP in Section 3.2.

A. Proving StRIP

3.1.1 Setting up the Framework

It will be convenient to decompose the random process generating the vectors α as follows: first pick (randomly) the

indices of the nonzero entries of α , and then the values of those entries. For the first step, we pick a random permutation $\pi \doteq (\pi_j)_{j \in \{1, \dots, \mathcal{C}\}}$ of $\{1, \dots, \mathcal{C}\}$; the k numbers π_1, \dots, π_k will then be the indices of the non-vanishing entries of α . Next, we pick k random values $\alpha_1, \dots, \alpha_k$; these will be the non-zero values of the entries of the vector α . Computing expectations with respect to α can be decomposed likewise; when we average over all possible choices of π , but not yet over the values of the random variables $\alpha_1, \dots, \alpha_k$, we shall denote such expectations by \mathbb{E}_π , adding a subscript. We start by proving the following

Lemma 9. *For π , Φ , α as described above and $f := N^{-1/2} \Phi \alpha$, we have*

$$\begin{aligned} \left(1 - \frac{k-1}{\mathcal{C}-1}\right) \|\alpha\|^2 &\leq \mathbb{E}_\pi [\|f\|^2] \\ &\leq \left(1 + \frac{1}{\mathcal{C}-1}\right) \|\alpha\|^2. \end{aligned}$$

Proof: With the notations introduced above, the entries of $f := N^{-1/2} \Phi \alpha$ are given by $f(x) := N^{-1/2} \sum_{j=1}^k \alpha_j \varphi_{\pi_j}(x)$. We have then

$$\begin{aligned} \|f\|^2 &= \sum_{x=1}^N |f(x)|^2 \\ &= \frac{1}{N} \sum_{x=1}^N \left(\sum_{j=1}^k |\alpha_j|^2 + \Psi(x) \right) \end{aligned} \quad (8)$$

where $\Psi(x) = \sum_{i,j,j \neq i} \alpha_j \overline{\alpha_i} \varphi_{\pi_j}(x) \overline{\varphi_{\pi_i}(x)}$.

The first term in (8) is independent of π ; it just equals $\sum_{j=1}^k |\alpha_j|^2 = \|\alpha\|^2$.

For the second term, we have

$$\begin{aligned} &\mathbb{E}_\pi \left[\sum_x \sum_{i,j \text{ with } j \neq i} \alpha_j \overline{\alpha_i} \varphi_{\pi_j}(x) \overline{\varphi_{\pi_i}(x)} \right] \\ &= \sum_{i,j \text{ with } j \neq i} \alpha_j \overline{\alpha_i} \sum_x \mathbb{E}_\pi \left[\sum_x \varphi_{\pi_j}(x) \overline{\varphi_{\pi_i}(x)} \right]. \end{aligned} \quad (9)$$

By (4) and Lemma 6, we have $\sum_x \varphi_\ell(x) \overline{\varphi_{\ell'}(x)} = \sum_x \varphi_m(x)$ for some appropriate $m := m(\ell, \ell')$; if $\ell \neq \ell'$, then $\overline{\varphi_{\ell'}} = (\varphi_{\ell'})^{-1} \neq (\varphi_\ell)^{-1}$, so that $m(\ell, \ell') \neq 1$.

As π ranges over all possible permutations of $\{1, \dots, \mathcal{C}\}$, the index $m(\pi_i, \pi_j)$ (with $j \neq i$) will range (uniformly) over all possible values $2, \dots, \mathcal{C}$ (i.e. excluding 1). It follows that, for $j \neq i$,

$$\begin{aligned} &\mathbb{E}_\pi \left[\sum_x \varphi_{\pi_j}(x) \overline{\varphi_{\pi_i}(x)} \right] \\ &= (\mathcal{C}-1)^{-1} \sum_{\ell \neq 1} \sum_x \varphi_\ell(x) \\ &= (\mathcal{C}-1)^{-1} \sum_x (-1) = -\frac{N}{\mathcal{C}-1}, \end{aligned} \quad (10)$$

where we have made use of a counting argument in the first

equality, and of (3) in the second. It then follows that

$$\begin{aligned} &\mathbb{E}_\pi \left[\sum_x \sum_{i,j \text{ with } j \neq i} \alpha_j \overline{\alpha_i} \varphi_{\pi_j}(x) \overline{\varphi_{\pi_i}(x)} \right] \\ &= -\frac{N}{\mathcal{C}-1} \sum_{i,j; \text{ with } j \neq i}^k \alpha_j \overline{\alpha_i}. \end{aligned}$$

Applying the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} 0 &\leq \sum_{i,j \text{ with } j \neq i} \alpha_j \overline{\alpha_i} + \sum_{j=1}^k |\alpha_j|^2 \\ &= \left| \sum_{j=1}^k \alpha_j \right|^2 \leq k \sum_{j=1}^k |\alpha_j|^2. \end{aligned}$$

Combining this with the previous equality gives

$$\begin{aligned} &-\frac{N(k-1)}{\mathcal{C}-1} \|\alpha\|^2 \\ &\leq \mathbb{E}_\pi \left[\sum_x \sum_{i,j \text{ with } j \neq i} \alpha_j \overline{\alpha_i} \varphi_{\pi_j}(x) \overline{\varphi_{\pi_i}(x)} \right] \\ &\leq \frac{N}{\mathcal{C}-1} \|\alpha\|^2 \end{aligned}$$

It then suffices to substitute this into (8) to prove the Lemma. \blacksquare

Remark 10. By using the Cauchy-Schwarz inequality in the last step of the proof of 9 we may have sacrificed quite a bit, especially if the non vanishing entries in α differ appreciably in order of magnitude. Without this step, the final inequality would be

$$\begin{aligned} &-\frac{N}{\mathcal{C}-1} (\|\alpha\|_{\ell_1}^2 - \|\alpha\|^2) \\ &\leq \mathbb{E}_\pi \left[\sum_x \sum_{i,j \text{ with } j \neq i} \alpha_j \overline{\alpha_i} \varphi_{\pi_j}(x) \overline{\varphi_{\pi_i}(x)} \right] \\ &\leq \frac{N}{\mathcal{C}-1} \|\alpha\|^2 \end{aligned} \quad (11)$$

To prove the concentration of $\|f\|^2$ around its expected value, we will make use of a version of the McDiarmid inequality [35] based on concentration of martingale difference random variables with *distinct* values (as opposed to *independent* values for the standard McDiarmid inequality). In what follows, upper case letters denote random variables, lower case letters denote values taken on by these random variables.

Theorem 11 (Self-Avoiding McDiarmid inequality). *Let $\mathcal{X}_1, \dots, \mathcal{X}_m$ be probability spaces and define \mathcal{X} as the probability space of all distinct m -tuples². In other words, the set \mathcal{X} is the subset of the product set $\mathbb{X} \doteq \mathcal{X}_1 \times \dots \times \mathcal{X}_m$ given*

²We follow a widespread custom, and denote by the same letter both the set carrying the probability measure, and the probability space [i.e. the triplet (set, σ -algebra of measurable sets, measure)]. We shall specify which is meant when confusion could be possible.

by

$$\mathcal{X} \doteq \{(t_1, \dots, t_m) \in \Pi_{i=1}^m \mathcal{X}_i \text{ s.th. } \forall i \neq j : t_i \neq t_j\}; \quad (12)$$

the probability measure on \mathcal{X} is just the renormalization (so as to be a probability measure) of the restriction to \mathcal{X} of the standard product measure on \mathbb{X} .

Let $h(t_1, \dots, t_m)$ be a function from the set \mathcal{X} to \mathbb{R} , such that for any coordinate i , given t_1, \dots, t_{i-1} :

$$\left| \begin{aligned} & \sup_{u \in \mathcal{X}_i; u \neq t_n, n=1 \rightarrow i} \mathbb{E}[h(t_1, \dots, t_{i-1}, u, T_{i+1}, \dots, T_m)] \\ & - \inf_{l \in \mathcal{X}_i; l \neq t_n, n=1 \rightarrow i} \mathbb{E}[h(t_1, \dots, t_{i-1}, l, T_{i+1}, \dots, T_m)] \end{aligned} \right| \leq c_i \quad (13)$$

where the expectations are taken over the random variables T_{i+1}, \dots, T_m (conditioned on taking values that are all different from each other and from t_1, \dots, t_{i-1} as well as u (first expectation) or l (second expectation)). Then for any positive γ ,

$$\begin{aligned} & \Pr[|h(T_1, \dots, T_m) - \mathbb{E}[h(T_1, \dots, T_m)]| \geq \gamma] \\ & \leq 2 \exp\left(\frac{-2\gamma^2}{\sum c_i^2}\right). \end{aligned} \quad (14)$$

Proof: See Appendix B. ■

3.1.2 Proof of StRIP

We are now ready to start the

Proof: (of the (k, ϵ, δ) -StRIP property, claimed in Theorem 8)

Let \mathcal{P}_k denote the set of all k -tuples (π_1, \dots, π_k) where $(\pi_i, \dots, \pi_{\mathcal{C}})$ is a permutation of $\{1, 2, \dots, \mathcal{C}\}$. It follows from the definition that all entries of each element of \mathcal{P}_k are distinct. The set \mathcal{P}_k is finite; equipped with the counting measure, renormalized so as to have total mass 1, \mathcal{P}_k is the probability space of the k non-zero entries of the random signal α : the (π_1, \dots, π_k) , corresponding to (uniformly) randomly picked permutations π of $\{1, \dots, \mathcal{C}\}$, are random variables distributed uniformly in \mathcal{P}_k . For $1 \leq i < j \leq k$, we denote by $\pi_{i \rightarrow j}$ the $(j - i + 1)$ -tuple of random variables $(\pi_i, \pi_{i+1}, \dots, \pi_j)$.

Given values $\alpha_1, \alpha_2, \dots, \alpha_k$, let $f : \mathcal{P}_k \rightarrow \mathbb{C}^N$ be defined by $f(\pi_1, \dots, \pi_k) = \frac{1}{\sqrt{N}} \sum_{i=1}^k \alpha_i \varphi_{\pi_i}$, and $h : \mathcal{P}_k \rightarrow \mathbb{R}$ by $h(\pi_1, \dots, \pi_k) = \|f(\pi_1, \dots, \pi_k)\|^2$. Clearly

$$h(\pi_1, \dots, \pi_k) = \frac{1}{N} \sum_{i,j=1}^k \alpha_i \bar{\alpha}_j (\varphi_{\pi_i})^\top \overline{\varphi_{\pi_j}}. \quad (15)$$

Our strategy of proof will be the following. We want to upper bound $\Pr_\pi[\|f\|^2 - \|\alpha\|^2 \geq \epsilon \|\alpha\|^2]$. From Lemma 9 we know that $\mathbb{E}_\pi[\|f\|^2]$ is close to $\|\alpha\|^2$. This suggests that we investigate, for $\beta > 0$, the function $G(\beta)$ defined by $G(\beta) \doteq \Pr_\pi[\|f\|^2 - \mathbb{E}_\pi[\|f\|^2] \geq \beta \|\alpha\|^2] = \Pr_\pi[h - \mathbb{E}_\pi[h] \geq \beta \|\alpha\|^2]$. This last expression is exactly of the type for which the Self-Avoiding McDiarmid Inequality gives upper bounds, provided we can establish first that h satisfies the required conditions of the Self-Avoiding McDiarmid inequality. Deriving such a bound is thus our first step.

From (15) and Lemma 2.2 we get

$$h(\pi_1, \dots, \pi_k) = \sum_{j=1}^k |\alpha_j|^2 + \frac{1}{N} \sum_{i,j \text{ with } i \neq j} \alpha_j \bar{\alpha}_i \varphi_{\pi_j}^\top \overline{\varphi_{\pi_i}}.$$

We have then

$$\begin{aligned} & h(\pi_1, \dots, \pi_\ell, \dots, \pi_k) - h(\pi_1, \dots, \pi'_\ell, \dots, \pi_k) \\ &= \frac{1}{N} \sum_{j \text{ with } j \neq \ell} \left[\alpha_\ell \bar{\alpha}_j [\varphi_{\pi_\ell} - \varphi_{\pi'_\ell}]^\top \overline{\varphi_{\pi_j}} \right] \\ & \quad + \frac{1}{N} \sum_{j \text{ with } j \neq \ell} \left[\alpha_j \bar{\alpha}_\ell \varphi_{\pi_j}^\top [\varphi_{\pi_\ell} - \varphi_{\pi'_\ell}] \right] \\ &= \frac{1}{N} \sum \left[\alpha_\ell \bar{\alpha}_j \sum_x \left(\varphi_{m(\pi_\ell, \pi_j)}(x) - \varphi_{m(\pi'_\ell, \pi_j)}(x) \right) \right] \\ & \quad + \frac{1}{N} \sum \left[\alpha_j \bar{\alpha}_\ell \sum_x \left(\varphi_{m(\pi_j, \pi_\ell)}(x) - \varphi_{m(\pi_j, \pi'_\ell)}(x) \right) \right], \end{aligned}$$

where we have used the same notation as in the proof of Lemma 9, i.e. $\varphi_{m(i,j)}(x) := \varphi_i(x) \overline{\varphi_j(x)}$.

Because $(\pi_1, \dots, \pi_\ell, \dots, \pi_k)$ and $(\pi_1, \dots, \pi'_\ell, \dots, \pi_k)$ are both in \mathcal{P}_k , the indices $\pi_1, \dots, \pi_\ell, \dots, \pi_k$ and π'_ℓ are all different. It then follows from (5) that

$$\begin{aligned} & |h(\pi_1, \dots, \pi_\ell, \dots, \pi_k) - h(\pi_1, \dots, \pi'_\ell, \dots, \pi_k)| \\ & \leq \frac{2}{N} |\alpha_\ell| \sum |\alpha_j| \left| \sum_x \varphi_{\pi_m(\ell,j)}(x) - \varphi_{\pi_m(\ell',j)}(x) \right| \\ & \leq \frac{2}{N} |\alpha_\ell| \sum_{j \text{ with } j \neq \ell} |\alpha_j| 2N^{1-\eta/2} \\ & = \frac{4}{N^{\eta/2}} |\alpha_\ell| \sum_{j \text{ with } j \neq \ell} |\alpha_j| \end{aligned} \quad (16)$$

where we have used that $m(\pi_\ell, \pi_j) \neq 1$ if $\pi_\ell \neq \pi_j$, i.e. if $\ell \neq j$. Because this bound is uniform over the π_ℓ in \mathcal{X}_ℓ , it is now clear that this implies the sufficient condition of the Self-Avoiding McDiarmid inequality, with c_ℓ given by the right hand side of (16). We can thus conclude from Theorem 11 that

$$\begin{aligned} & \Pr_\pi[|h - \mathbb{E}[h]| \geq \beta \|\alpha\|^2] \\ & \leq 2 \exp\left(-\frac{2\beta^2 N^\eta \|\alpha\|^4}{16 \sum_{\ell=1}^k |\alpha_\ell|^2 \left[\sum_{j \text{ with } j \neq \ell} |\alpha_j|\right]^2}\right). \end{aligned} \quad (17)$$

Since

$$\begin{aligned} & \sum_{\ell=1}^k |\alpha_\ell|^2 \left[\sum_{j \text{ with } j \neq \ell} |\alpha_j| \right]^2 \\ & \leq \sum_{\ell=1}^k |\alpha_\ell|^2 \left[\sum_{j=1}^k |\alpha_j| \right]^2 \\ & \leq \|\alpha\|^2 k \|\alpha\|^2 = k \|\alpha\|^4, \end{aligned}$$

where we have used the Cauchy-Schwarz inequality in the

penultimate step, it follows that

$$\Pr_{\pi} [|h - \mathbb{E}[h]| \geq \beta \|\alpha\|^2] \leq 2 \exp\left(-\frac{\beta^2 N^{\eta}}{8k}\right).$$

After substituting $\|f\|^2$ for h , and applying Lemma 9, we finally obtain that

$$\begin{aligned} \Pr_{\pi} \left[\left| \|f\|^2 - \|\alpha\|^2 \right| \geq \left(\beta + \frac{k-1}{\mathcal{C}-1} \right) \|\alpha\|^2 \right] \\ \leq 2 \exp\left(-\frac{\beta^2 N^{\eta}}{8k}\right). \end{aligned}$$

For $\epsilon > (k-1)/(\mathcal{C}-1)$, we can set $\beta = \epsilon - (k-1)/(\mathcal{C}-1)$, thus recovering the StRIP-bound claimed in the statement of Theorem 8 for this case: with probability at least $1 - 2 \exp\left(-\frac{[\epsilon - (k-1)/(\mathcal{C}-1)]^2 N^{\eta}}{8k}\right)$, we have the following near-isometry for k -sparse vectors α :

$$(1 - \epsilon)\|\alpha\|^2 \leq \|f\|^2 \leq (1 + \epsilon)\|\alpha\|^2. \quad (18)$$

Remark 12. Equation (18) implies that as long as $\sqrt{\frac{k}{N^{\eta}}} + \frac{k-1}{\mathcal{C}-1} < \epsilon$, the probability of failure (i.e. the probability that the near-isometry inequality fails to hold) drops to zero as $\mathcal{C} \rightarrow \infty$. In particular, if η equals 1, $k \leq \mu(\mathcal{C}-1)\epsilon + 1$ for some constant μ less than one, and $N = O\left(\frac{k \log \mathcal{C}}{\epsilon^2}\right)$ then the probability of failure approaches zero at the rate \mathcal{C}^{-1} . ■

Remark 13. Figure 1 shows the distribution of condition numbers for the singular values of restrictions of the sensing matrix to sets of K columns. Two cases are considered; the Reed Muller matrices constructed in Section 2.2 and random Gaussian matrices of the same size. The figure suggests that the decay of

$$\Pr \left[\left| \|f\|^2 - \|\alpha\|^2 \right| \geq \epsilon \|\alpha\|^2 \right]$$

is similar for both types of compressive sensing matrices.

Remark 14. Note that similar to the case of random and expander matrices, the number of measurements N grows as the inverse square of the distortion parameter ϵ , $N \propto \frac{1}{\epsilon^2}$, as $\epsilon \rightarrow 0$.

Remark 15. By avoiding the use of the Cauchy-Schwarz inequality at the end of the proof, and making use of Remark 10, one can sharpen the bounds. From (17) it follows that with probability at least $1 - 2 \exp\left(-\frac{\beta^2 N^{\eta} \|\alpha\|^2}{8\|\alpha\|_{\ell_1}^2}\right)$

$$-\left(\beta - \frac{1}{\mathcal{C}-1}\right) \|\alpha\|^2 - \frac{1}{\mathcal{C}-1} \|\alpha\|_{\ell_1}^2 \leq \|f\|^2 - \|\alpha\|^2$$

and

$$\|f\|^2 - \|\alpha\|^2 \leq \left(\beta + \frac{1}{\mathcal{C}-1}\right) \|\alpha\|^2$$

This implies (set $\beta = \gamma\rho$, $\rho = \|\alpha\|_{\ell_1} \|\alpha\|^{-1}$)

$$\begin{aligned} \Pr_{\pi} \left[\left| \|f\|^2 - \|\alpha\|^2 \right| \geq \left(\gamma\rho + \frac{1}{\mathcal{C}-1} \max(1, \rho^2 - 1) \right) \|\alpha\|^2 \right] \\ \leq 2e^{-\gamma^2 N^{\eta}/8}, \end{aligned}$$

or, equivalently,

$$\begin{aligned} \Pr_{\pi} \left[\left| \|f\|^2 - \|\alpha\|^2 \right| \geq \epsilon \|\alpha\|^2 \right] \\ \leq 2 \exp\left(-\frac{1}{8} N^{\eta} \left[\frac{\epsilon - (\mathcal{C}-1)^{-1}}{\rho} - \chi_{(\rho > \sqrt{2})} \frac{\rho^2 - 2}{\rho(\mathcal{C}-1)} \right]^2 \right), \end{aligned}$$

with $\rho = \|\alpha\|_{\ell_1} \|\alpha\|^{-1}$, as above, and $\chi_{(a>0)} = 1$ if $a > 0$, $\chi_{(a>0)} = 0$ otherwise. The worst case for this bound is when $\rho = \sqrt{k}$, in which case we recover the bound in Theorem 8; if one is restricted, for whatever reason, to k -sparse vectors that are known to have some entries that are much larger than other non vanishing entries, then the more complicated bound given here is tighter.

Remark 16. If the sparsity level k is greater than $\sqrt{\mathcal{C}}$, then $\mathcal{C} \leq k^2 \leq N^2$. However, since some deterministic sensing matrices of section II structurally require the condition $N^2 \leq \mathcal{C}$, a deterministic matrix with $N' = O\left(\frac{k \log \mathcal{C}}{\epsilon^2}\right)$ rows and $N'^2 \leq \mathcal{C}'$ columns is required. In this case, the $N' \times \mathcal{C}$ sensing matrix Φ is constructed by choosing \mathcal{C} random columns from the $N' \times \mathcal{C}'$ deterministic matrix.

B. Proving UStRIP: Uniqueness of Sparse Representation

Although we have established the desired near-isometry bounds, we still have to address the Uniqueness guarantee; unlike the standard RIP case, this does not follow automatically from a StRIP bound, as pointed out in the Introduction. More precisely, we need to estimate the probability that a randomly picked k -sparse vector α has an “evil twin” $\alpha' \neq \alpha$ that maps to the same image under Φ , i.e. $\Phi\alpha = \Phi\alpha'$, and prove that this probability is very small.

If $S \subset \{1, \dots, \mathcal{C}\}$ is the union of possible support sets of a two k -sparse vectors, that is, if $s \doteq |S| \leq 2k$, then we define Φ_S to be the $N \times s$ matrix obtained by picking up only the columns indexed by labels in S . In other words, the matrix elements of Φ_S are those $\varphi_j(x)$ for which $j \in S$, with x varying over its full range. There will be two different k -sparse vectors $\alpha' \neq \alpha$, the supports of which are both contained in S , if and only if the $s \times s$ matrix $\Phi_S^\dagger \Phi_S$ is rank-deficient (where Φ_S^\dagger denotes conjugate transpose of Φ_S). Note that this property concerns the support set S only – the values of the entries of α are not important. This is similar to the discussion of sparse reconstruction when Φ satisfies a deterministic Null Space Property [12]. Once uniqueness is found to be overwhelmingly likely, we can derive from it the probability that decoding algorithms (such as the quadratic decoding algorithms described in Section V) succeed in constructing, from $\Phi\alpha$, a faithfully exact or close copy (depending on the application) of the k -sparse source vector α .

In fact, it turns out that we won't even have to consider matrices Φ_S with $|S| = 2k$; as we shall see below, it suffices to consider Φ_S for sets S of cardinality up to k .

Once again, condition (St3) will play a crucial role. For the StRIP analysis, in the previous subsection, it sufficed to take $\eta > 0$, where η is the parameter that measures the closeness of column sums in (St3). In this subsection, we will impose a non-zero lower bound on η ; we shall see that $\eta > 0.5$ suffices for our analysis.

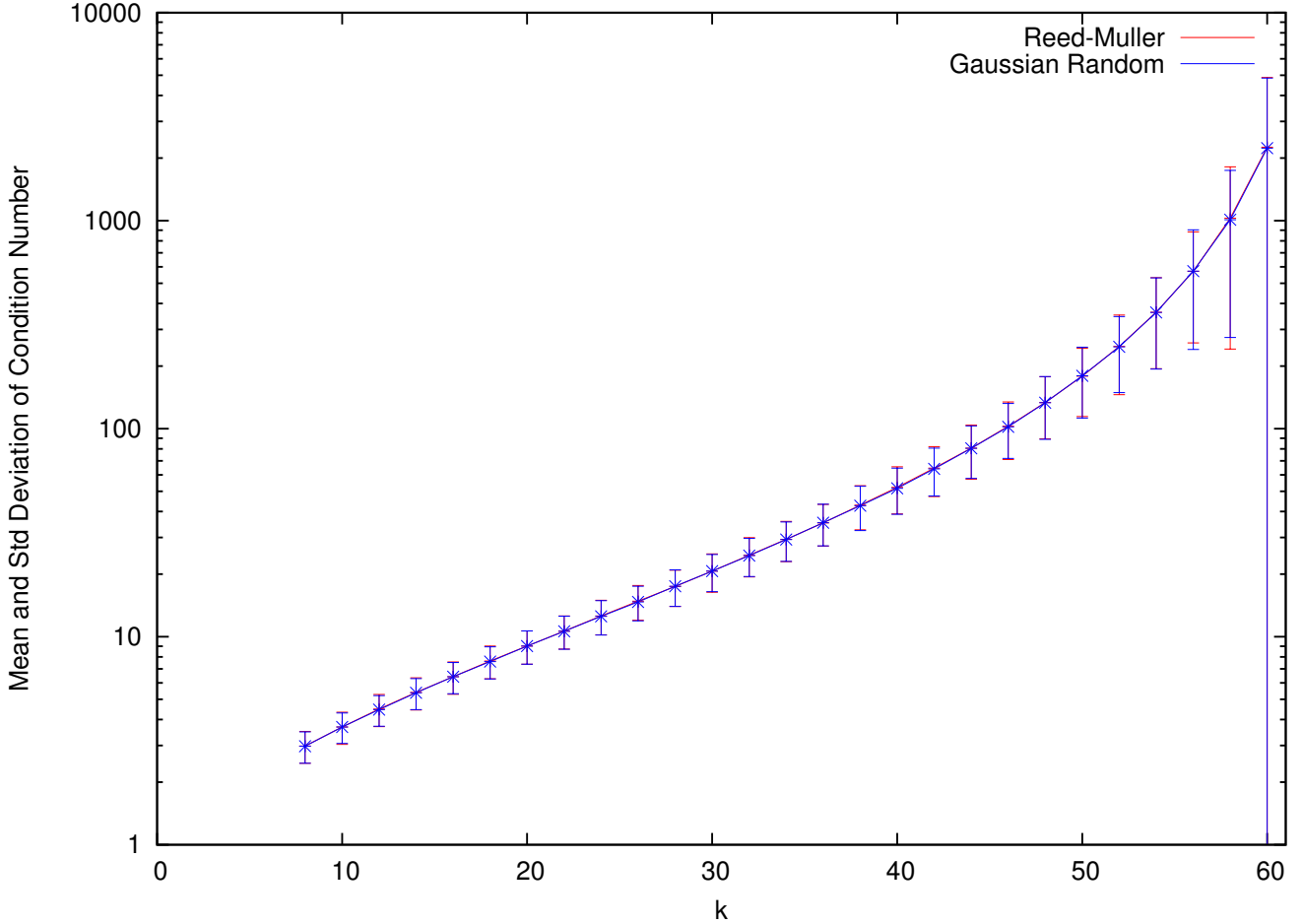


Fig. 1. Mean and standard deviation for the condition number of k -Gram matrices for Φ_{RM} , with $m = 6$, compared to that of a Gaussian random matrix of the same size.

We recall here the formulation of (St3): for any column φ_j of the sensing matrix, with $j \geq 2$,

$$\left| \sum_x \varphi_j(x) \right| \leq N^{1-\eta/2}.$$

We introduced the notation Φ_S at the start of this subsection. We shall also use the special case where we wish to restrict the sensing matrix Φ to a single column indexed by w ; in that case, we denote the restriction by φ_w . Finally we denote the conjugate transpose of a matrix Φ_S by Φ_S^\dagger . We shall use Tropp's argument (see Section 7 of [Tro08b]) to prove uniqueness of sparse representation; to apply this argument we first need to prove that a random submatrix φ_κ has small coherence with the remaining columns of the sensing matrix.

Lemma 17. *Let Φ be η -StRIP-able with $\eta > 1/2$, and assume that the conditions $k < \epsilon(C-1) + 1$, and $N = O\left((k \log C/\epsilon^2)^{1/\eta}\right)$ hold, and δ is as defined in Theorem 8, i.e*

$$\delta := 2 \exp \left[- \frac{[\epsilon - (k-1)/(C-1)]^2 N^\eta}{8k} \right].$$

Let w be a fixed column of Φ , and let $\kappa = \{\kappa_1, \dots, \kappa_k\}$ be

the positions of the first k elements of a random permutation of $\{1, \dots, C\} \setminus \{w\}$. Then

$$\mathbb{E} \left[\left\| \frac{1}{\sqrt{N}} \Phi_\kappa^\dagger \frac{1}{\sqrt{N}} \varphi_w \right\|^2 \right] = \frac{k}{N} \frac{C-N}{C-1}, \quad (19)$$

where the expectation is with respect to the choice of the set κ .

Proof: By linearity of expectation we have

$$\mathbb{E}_\kappa \left[\left\| \frac{1}{\sqrt{N}} \Phi_\kappa^\dagger \frac{1}{\sqrt{N}} \varphi_w \right\|^2 \right] = \frac{1}{N^2} \sum_{i=1}^k \mathbb{E}_{\kappa_i} \left[\left| (\overline{\varphi_{\kappa_i}})^\top \varphi_w \right|^2 \right]. \quad (20)$$

Since the set of columns of Φ is invariant under complex conjugation, and forms a group under pointwise multiplication, we have

$$(\overline{\varphi_{\kappa_i}})^\top \varphi_w = \sum_x \overline{\varphi_{\kappa_i}(x)} \varphi_w(x) = \sum_x \varphi_{m(w, \kappa_i)}(x),$$

where we use again the notation introduced just below (9): $\varphi_\ell(x) \overline{\varphi_{\ell'}(x)} \doteq \varphi_{m(\ell, \ell')}(x)$. As κ ranges over all the possible permutations that do not move w , κ_i ranges uniformly over $\{1, \dots, C\} \setminus \{w\}$, and the different $z_i := m(w, \kappa_i)$ range uniformly over $\{2, \dots, C\}$.

Hence:

$$\begin{aligned}
& \sum_{i=1}^k \frac{1}{N^2} \mathbb{E}_{\kappa_i} \left[\left| (\overline{\varphi_{\kappa_i}})^\top \varphi_w \right|^2 \right] \\
&= \sum_{i=1}^k \frac{1}{N^2} \mathbb{E}_{z_i} \left[\left| \sum_x \varphi_{z_i}(x) \right|^2 \right] \\
&= \sum_{i=1}^k \frac{1}{N^2} \mathbb{E}_{z_i} \left[\sum_{x,y=1}^N \varphi_{z_i}(x) \overline{\varphi_{z_i}(y)} \right] \\
&= \frac{k}{N^2} \frac{1}{(C-1)} \sum_{j=2}^C \sum_{x,y=1}^N \varphi_j(x) \overline{\varphi_j(y)} \quad (21) \\
&= \frac{k}{N^2} \frac{1}{(C-1)} \left(\sum_{x,y=1}^N [C\delta_{x,y} - 1] \right) \\
&= \frac{k}{N^2} \frac{1}{(C-1)} (NC - N^2) = \frac{k}{N} \frac{C-N}{(C-1)},
\end{aligned}$$

where we have used (St1). \blacksquare

Next, we use the Self-Avoiding McDiarmid inequality, together with property (St3) to derive a uniform bound for the random variable $\|\Phi_{\kappa}^\dagger \varphi_w\|^2$:

Theorem 18. *Let Φ be η -StRIP-able with $\eta > 1/2$, and assume that the conditions $k < \epsilon(C-1) + 1$, and $N = O\left(\left(\frac{k \log C}{\epsilon^2}\right)^{1/\eta}\right)$ hold, define δ as in Theorem 8, and let λ be a set of k random columns of Φ . Then with probability at least $1 - \delta$, there exists no w such that*

$$\left\| \frac{1}{\sqrt{N}} \Phi_\lambda^\dagger \frac{1}{\sqrt{N}} \varphi_w \right\|^2 \geq \frac{k}{N} \frac{C-N}{C-1} + \frac{\sqrt{2k \log C/\delta}}{N^\eta} \quad (22)$$

Proof: The proof is in several steps. In the first step, we pick any $w \in \{1, \dots, C\}$, and keep it fixed (for the time being). Let

$$f(t_1, \dots, t_k) = \frac{1}{N^2} \sum_{i=1}^k \left| (\overline{\varphi_{t_i}})^\top \varphi_w \right|^2,$$

where we assume that t_1, \dots, t_k are k different elements of $\{1, \dots, C\} \setminus \{w\}$, picked at random. Note that if λ is a random permutation of $\{1, \dots, C\} \setminus \{w\}$, then $f(\lambda_1, \dots, \lambda_k) = \left\| \frac{1}{\sqrt{N}} \Phi_\lambda^\dagger \frac{1}{\sqrt{N}} \varphi_w \right\|^2$. The function f , as defined above from $\{(t_1, t_2, \dots, t_k); t_i \in \{1, \dots, C\} \setminus \{w\} \forall i, t_i \neq t_j, \forall i \neq j\}$

to \mathbb{R} , is information-theoretically indistinguishable from the function F from the permutations of $\{1, \dots, C\} \setminus \{w\}$ to \mathbb{R} defined by

$$F(\lambda) = \left\| \frac{1}{\sqrt{N}} \Phi_\lambda^\dagger \frac{1}{\sqrt{N}} \varphi_w \right\|^2.$$

We have computed $\mathbb{E}[f] = \mathbb{E}[F]$ in Lemma 17; in order to apply the Self-Avoiding McDiarmid Inequality to f , we need verify only that a necessary condition of the Self-Avoiding McDiarmid inequality holds.

When we subtract $f(t_1, \dots, t_{i-1}, t'_i, t_{i+1}, \dots, t_k)$ from $f(t_1, \dots, t_{i-1}, t_i, t_{i+1}, \dots, t_k)$, only the i -th term survives;

we have

$$\begin{aligned}
& \mathbb{E}[f(t_1, \dots, t_{i-1}, t_i, t_{i+1}, \dots, t_k)] \\
& - \mathbb{E}[f(t_1, \dots, t_{i-1}, t'_i, t_{i+1}, \dots, t_k)] \\
&= \left| \frac{1}{N^2} \left| (\overline{\varphi_{t_i}})^\top \varphi_w \right|^2 - \frac{1}{N^2} \left| (\overline{\varphi_{t'_i}})^\top \varphi_w \right|^2 \right| \\
&= \frac{1}{N^2} \left| \left| \sum_{x=1}^N \varphi_{m(w, t_i)}(x) \right|^2 - \left| \sum_{x=1}^N \varphi_{m(w, t'_i)}(x) \right|^2 \right| \\
&\leq 2N^{-\eta}, \quad (23)
\end{aligned}$$

by (St3), since $m(w, t_i) \neq 1 \neq m(w, t'_i)$. It immediately follows that the concentration condition holds for f , with $c_i = N^{-\eta}$. Therefore the Self-Avoiding McDiarmid Inequality holds for f , which means it also holds for F : for any positive γ ,

$$\begin{aligned}
& \Pr_\lambda \left[\left\| \frac{1}{\sqrt{N}} \Phi_\lambda^\dagger \frac{1}{\sqrt{N}} \varphi_w \right\|^2 \geq \frac{k}{N} + \gamma \right] \\
&\leq \Pr_\lambda \left[\left\| \frac{1}{\sqrt{N}} \Phi_\lambda^\dagger \frac{1}{\sqrt{N}} \varphi_w \right\|^2 \geq \frac{k}{N} \frac{C-N}{C-1} + \gamma \right] \\
&\leq \exp\left(\frac{-\gamma^2 N^{2\eta}}{2k}\right).
\end{aligned}$$

All this was for one fixed choice of w ; note that the bound does not depend on the identity of w . This implies that by applying union bounds over the C possible choices for the column w of Φ , we get that the probability that there exists a w such that

$$\left\| \frac{1}{\sqrt{N}} \Phi_\lambda^\dagger \frac{1}{\sqrt{N}} \varphi_w \right\|^2 \geq \frac{k}{N} + \gamma,$$

is at most $C \exp\left(\frac{-\gamma^2 N^{2\eta}}{2k}\right)$. Writing γ in terms of δ completes the proof. \blacksquare

If $N = O\left(\left(\frac{k \log C}{\epsilon^2}\right)^{1/\eta}\right)$, the right hand side of (22) reduces to

$$O\left(\frac{\epsilon^2}{N}\right) + O\left(\epsilon^{2\eta} \left[1 + \frac{|\log \delta|}{\log C}\right]^{1/2} (k \log C)^{-(\eta-1/2)}\right)$$

Thus, if $\eta > 1/2$, then (for sufficiently small ϵ , and sufficiently large C) a choice of k random columns of Φ has a very high probability of having small coherence with *any* other column of the matrix; in particular, we have, with probability exceeding $1 - \delta$, that

$$\left\| \frac{1}{\sqrt{N}} \Phi_\lambda^\dagger \frac{1}{\sqrt{N}} \varphi_w \right\|^2 < (1 - \epsilon)^2. \quad (24)$$

This establishes incoherence between the random submatrix Φ_λ and the remaining columns of the sensing matrix.

We can now complete the UStRIP proof by following an argument of Tropp [36]; for completeness we include the argument here:

Lemma 19. *Let $\lambda = \{\lambda_1, \dots, \lambda_k\}$ be a set of k indices sampled uniformly from $\{1, \dots, C\}$. Assume that Φ is (k, ϵ, δ) -StRIP. Let S be any other subset of $\{1, \dots, C\}$ of size less*

than or equal to k . Then, with probability at least $(1 - \delta)$ (with respect to the randomness in the choice of λ)

$$\dim(\text{range}(\Phi_\lambda) \cap \text{range}(\Phi_S)) < k. \quad (25)$$

Proof: First, note that we need check only the case $\dim(\text{range}(\Phi_S)) = k$, since otherwise (25) is immediate. Note also that, because Φ is (k, ϵ, δ) -StRIP, the probability that the randomly picked set $\lambda = \{\lambda_1, \dots, \lambda_k\}$ satisfies

$$(1 - \epsilon)\text{Id}_\lambda \leq \frac{1}{N}\Phi_\lambda^\dagger \Phi_\lambda \leq (1 + \epsilon)\text{Id}_\lambda$$

is at least $1 - \delta$. (The notation Id_λ stands for the identity matrix on λ ; this just amounts to restating the (k, ϵ, δ) -StRIP condition in matrix form.) It follows that, with probability at least $1 - \delta$,

$$\sigma_{\min}(\Phi_\lambda) \geq \sqrt{(1 - \epsilon)N}, \quad (26)$$

where $\sigma_{\min}(\Phi_\lambda)$ is the smallest singular value of Φ_λ .

Since $S \neq \lambda$, S has at least one index not in λ . Denote that index by s . Since the entries of the matrix are all unimodular, we have

$$\|\varphi_s\|^2 = \sum_x |\varphi_s(x)|^2 = N. \quad (27)$$

Let \mathbb{P}_λ be the orthogonal projection operator on the range \mathcal{R}_λ of Φ_λ . We shall prove (25) by showing that $\|\mathbb{P}_\lambda \varphi_s\|^2 < \|\varphi_s\|^2$, which implies that there exists a vector in the range of Φ_S that is outside the range of Φ_λ . Note that

$$\mathbb{P}_\lambda = \Phi_\lambda \left(\Phi_\lambda^\dagger \Phi_\lambda \right)^{-1} \Phi_\lambda^\dagger. \quad (28)$$

Since Φ_λ is (k, ϵ, δ) -StRIP, we have, still with probability at least $1 - \delta$,

$$\begin{aligned} \|\mathbb{P}_\lambda \varphi_s\|^2 &= \left(\Phi_\lambda^\dagger \varphi_s \right)^\dagger \left(\Phi_\lambda^\dagger \Phi_\lambda \right)^{-1} \left(\Phi_\lambda^\dagger \varphi_s \right) \\ &\leq \frac{\left\| \Phi_\lambda^\dagger \varphi_s \right\|^2}{(\sigma_{\min}(\Phi_\lambda))^2} \leq \frac{\left\| \Phi_\lambda^\dagger \varphi_s \right\|^2}{N(1 - \epsilon)} \\ &\leq (1 - \epsilon)N < N, \end{aligned}$$

where the penultimate inequality is by Equation (24). \blacksquare

Theorem 20. *Let Φ be η -StRIP-able with $\eta > 1/2$, and assume that the conditions $k < \epsilon(C - 1) + 1$, and $N = O\left(\left(\frac{k \log C}{\epsilon^2}\right)^{1/\eta}\right)$ hold, define δ as in Theorem 8, and let α be a randomly picked k -sparse signal. Then with probability at least $1 - \delta$ (with respect to the random choice of α), α is the only k -sparse vector that satisfies the equation $f = \frac{1}{\sqrt{N}}\Phi\alpha$.*

Proof: We have already proved in Section 3.1.2 that Φ is (k, ϵ, δ) -StRIP. We start by recalling that the random choice of α can be viewed as first choosing its support, a uniformly distributed subset of size k within $\{1, \dots, C\}$, and then, once the support is fixed, choosing a random vector within the corresponding k -dimensional vector space. For this last choice no distribution has been specified; we shall just assume that it is absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^k or \mathbb{C}^k .

Since Φ is (k, ϵ, δ) -StRIP, Φ_λ is non-singular with probability

exceeding $(1 - \delta)$, so that

$$\dim(\text{range}(\Phi_\lambda)) = k$$

with probability exceeding $1 - \delta$. The near-isometry property of Φ_λ implies that no two signals with support λ can have the same value in the measurement domain. If there nevertheless were a vector α' such that $\Phi\alpha' = \Phi\alpha$, the support S of α' would therefore necessarily be different from λ . By Lemma 19, we know that $V \doteq \text{range}(A_\lambda) \cap \text{range}(A_S)$ is at most $(k - 1)$ -dimensional. It follows that in order to possibly have an ‘‘evil twin’’ α' , the vector α must itself lie in the at most $(k - 1)$ -dimensional space that is the inverse image of V under Φ_λ . This set, however, has measure zero with respect to any measure that is absolutely continuous with respect to the k -dimensional Lebesgue measure. Thus, for each k -set λ for which Φ_λ is a near-isometry, the vectors that are not uniquely determined by their image $\Phi\alpha$, constitute a set of measure zero. Since randomly chosen k -sets λ produce restrictions Φ_λ that are near-isometric with probability exceeding $1 - \delta$, the theorem is proved. \blacksquare

Combining Remark 12 with Theorem 20 completes the proof of Theorem 8.

IV. PARTIAL FOURIER ENSEMBLES

In Partial Fourier ensembles the matrix Φ is formed by uniform random selection of N rows from the $\mathcal{C} \times \mathcal{C}$ discrete Fourier Transform matrix. The resulting random sensing matrices are widely used in compressed sensing, because the corresponding memory cost is only $O(N \log \mathcal{C})$, in contrast to the $O(N\mathcal{C})$ cost of storing Gaussian and Bernoulli matrices. Moreover, it is known [10], [7] that if $N \geq k \log^5 \mathcal{C}$, then with overwhelming probability, the partial Fourier matrix satisfies the RIP property. It is easy to verify that such Φ satisfies the Conditions (St1), and (St2). We now show that it also satisfies Condition (St3) almost surely.

Note that here in contrast to the proof of Theorem 8, the randomness is with respect to the choice of the N rows from the Discrete Fourier Transform matrix. We show that with overwhelming probability, the condition (St3) is satisfied for every column of this randomly sampled matrix. First fix a column φ_i other than the identity column, and define the random variable Z_x to be the value of the entry $\varphi_i(x)$, where the randomness is with respect to the choice of the rows of Φ (that is with respect to the choice of x). Since the rows are chosen uniformly at random, and the column sums (for all but the first column) in the discrete Fourier transform are zero, we have

$$\mathbb{E} \left[\frac{\sum_x Z_x}{N} \right] = \frac{\sum_x \mathbb{E}[Z_x]}{N} = 0. \quad (29)$$

Since all entries are unimodular, we may apply Hoeffding’s inequality to both the real and the imaginary part of the random variable $\frac{\sum_x Z_x}{N}$, then apply union bounds to conclude that for all $\epsilon > 0$

$$\text{Pr} \left[\left| \frac{\sum_x Z_x}{N} \right| \geq \epsilon \right] \leq 4 \exp \{-2N\epsilon^2\}.$$

Applying union bounds to all $\mathcal{C} - 1$ admissible columns we

get

$$\Pr[\text{there exists a column average greater than } \epsilon] \quad (30)$$

is at most $4\mathcal{C} \exp\{-2N\epsilon^2\}$. Hence, with probability at least $1 - \delta$ all column averages are $O\left(\sqrt{\frac{\log \mathcal{C}}{N}}\right)$, and all column sums are less than $\sqrt{N \log \mathcal{C}}$, so that condition (St3) is indeed satisfied. Applying Theorem 8 we see that a partial Fourier matrix satisfies StRIP with only $k \log \mathcal{C}$ measurements. This improves upon the best previous upper bound of $k \log^5 \mathcal{C}$ obtained in [10] and helps explain why partial Fourier matrices work well in practice.

V. QUADRATIC RECONSTRUCTION ALGORITHM

Algorithm 1 Quadratic Reconstruction Algorithm

Input: N dimensional vector $f = \frac{1}{\sqrt{N}}\Phi\alpha + \nu$

Output: An approximation $\hat{\alpha}$ to the signal α

- 1: Set $f_1 = f$, $\Theta = \{\}$, $\hat{\alpha} = 0_N$.
 - 2: **for** $t = 1, \dots, k$ or while $\|f_t\|_2 \geq \epsilon$ **do**
 - 3: **for** each entry $x = 1$ to N **do**
 - 4: pointwise multiply f_t with a shift (offset) of itself as in (31).
 - 5: **end for**
 - 6: Compute the fast Walsh-Hadamard transform of the pointwise product: Equation (32)
 - 7: Find the position p_t of the next peak in the Hadamard domain: Equation (33) implies that the chirp-like cross terms appear as a constant background signal.
 - 8: **if** $p_t \in \text{Keys}(\Theta)$ **then**
 - 9: Restore $f_t \leftarrow f_t + \Theta(p_t)\varphi_{p_t}$.
 - 10: **end if**
 - 11: Update $\beta_t \doteq \frac{1}{\sqrt{N}}f_t^\top \varphi_{p_t}$ which minimizes $\|f_t - \frac{1}{\sqrt{N}}\beta_t \varphi_{p_t}\|_2$.
 - 12: Add β^t to entry p_t of $\hat{\alpha}$.
 - 13: Set $\Theta(p_t) = \beta_t$.
 - 14: Set $f_{t+1} \leftarrow f_t - \beta_t \varphi_{p_t}$.
 - 15: **end for**
-

The Quadratic Reconstruction Algorithm [23], [24], [25], described in detail above, takes advantage of the multivariable quadratic functions that appear as exponents in Delsarte-Goethals sensing matrices. It is this structure that enables the algorithm to avoid the matrix-vector multiplication required when Basis and Matching Pursuit algorithms are applied to random sensing matrices. Because our algorithm requires only vector-vector multiplication in the measurement domain, the reconstruction complexity is sublinear in the dimension of the data domain. The Delsarte-Goethals sensing matrix was introduced in Section 2.2: there are 2^m rows indexed by binary m -tuples x , and $2^{(r+2)m}$ columns φ_{P_i, b_i} indexed by pairs P_i, b_i where P_i is a binary symmetric matrix and b_i is a binary m -tuple. The first step in our algorithm is pointwise

multiplication of a sparse superposition

$$f(x) = \frac{1}{\sqrt{N}} \sum_{i=1}^k \alpha_i \varphi_{P_i, b_i}(x)$$

with a shifted copy of itself. The sensing matrix is obtained by exponentiating multivariable quadratic functions so the first step produces a sparse superposition of pure frequencies (in the example below, these are Walsh functions in the binary domain) against a background of chirp-like cross terms.

$$f(x+a)\overline{f(x)} = \frac{1}{N} \sum_{j=1}^k |\alpha_j|^2 (-1)^{a^\top P_j x} \quad (31)$$

$$+ \frac{1}{N} \sum_{j \neq t} \alpha_j \overline{\alpha_t} \varphi_{P_j, b_j}(x+a) \overline{\varphi_{P_t, b_t}(x)}.$$

Then the (fast) Hadamard transform concentrates the energy of the first term $\frac{1}{N} \sum_{j=1}^k |\alpha_j|^2 (-1)^{a^\top P_j x}$ at (no more than) k Walsh-Hadamard tones, while the second term distributes energy uniformly across all N tones. The l^{th} Fourier coefficient is

$$\Gamma_a^l = \frac{1}{N^{3/2}} \sum_{j \neq t} \alpha_j \overline{\alpha_t} \sum_x (-1)^{l^\top x} \varphi_{P_j, b_j}(x+a) \overline{\varphi_{P_t, b_t}(x)}, \quad (32)$$

and it can be shown (see [25]) that the energy of the chirp-like cross terms is distributed uniformly in the Walsh-Hadamard domain. That is for any coefficient l

$$\lim_{N \rightarrow \infty} \mathbb{E} \left[N^2 |\Gamma_a^l|^2 \right] = \sum_{j \neq t} |\alpha_j|^2 |\alpha_t|^2. \quad (33)$$

Equation (33) is related to the variance of f and may be viewed as a fine-grained concentration estimate. In fact the proof of (33) mirrors the proof of the UStRIP property given in Section 3; first we show that the expected value of any Walsh-Hadamard coefficient is zero, and then we use the Self-Avoiding McDiarmid Inequality to prove concentration about this expected value. The Walsh-Hadamard tones appear as spikes above a constant background signal and the quadratic algorithm learns the terms in the sparse superposition by varying the offset a . These terms can be peeled off in decreasing order of signal strength or processed in a list. The quadratic algorithm is a repurposing of the chirp detection algorithm commonly used in navigation radars which is known to work extremely well in the presence of noise. Experimental results show close approach to the information theoretic lower bound on the required number of measurements. For example, numerical experiments show that the quadratic decoding algorithm is able to reconstruct greater than 40-sparse superpositions when applied to deterministic Kerdock sensing matrices with $N = 2^9$ and $\mathcal{C} = 2^{18}$. In this case, the information theoretic lower bound is $k \log_2(1 + \mathcal{C}/k) = 507$ [24].

We now explain how the StRIP property provides performance guarantees for the Quadratic Reconstruction Algorithm. At each iteration the algorithm returns the location of one of the k significant entries and an estimate for the value of that entry. The StRIP property guarantees that the estimate is within ϵ of the true value. These errors compound as the

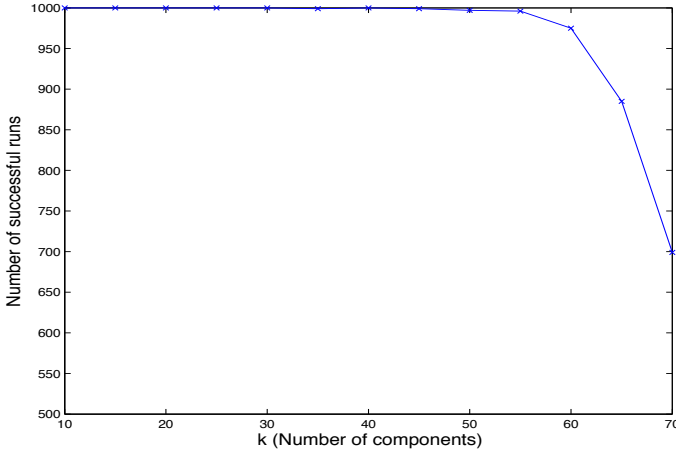


Fig. 2. The number of successful reconstructions in 1000 trials versus the sparsity factor k for the deterministic Kerdock sensing matrix corresponding to $m = 9$

algorithm iterates, but since the chirp cross-terms and noise are uniformly distributed in the Walsh-Hadamard domain, the error in recovery is bounded by the difference between the true signal α and its best k -term approximation α_k . More precisely, if Φ is (k, ϵ, δ) -StRIP, if the position of the k significant entries are chosen uniformly at random, if the near-zero entries and the measurement noise ν come from a Gaussian distribution, and if the Quadratic Recovery Algorithm is used to recover an approximation $\hat{\alpha}$ for α , then

$$\|\alpha - \hat{\alpha}\|_2 \leq \frac{5 + \epsilon}{1 - \epsilon} \|\alpha - \alpha_k\|_2 + \frac{2}{1 - \epsilon} \|\nu\|_2. \quad (34)$$

The role of the StRIP property is to bound the error of approximation in Step 11 of the Quadratic Reconstruction Algorithm. Note that if it were somehow possible to identify the support of α beforehand, then the UStRIP property would guarantee that we would be able to recover the signal values by linear regression. However identifying the support of a k -sparse signal is known to be almost as hard as full reconstruction, and that is why our algorithm finds location and estimates signal value simultaneously, and does so one location at a time.

Note that the error bound is of the form ℓ_2/ℓ_2 :

$$\|\alpha - \hat{\alpha}\|_2 \leq C \|\alpha - \alpha_k\|_2. \quad (35)$$

This bound is tighter than ℓ_2/ℓ_1 bounds of random ensembles [2], and ℓ_1/ℓ_1 of expander-based methods [6].

VI. RESILIENCE TO NOISE

A. Noisy Measurements

In this Section, we consider deterministic sensing matrices satisfying the hypothesis of Theorem 8, and show resilience to independent identically distributed (iid) Gaussian noise that is uncorrelated with the measured signal. Note we have introduced the square of $(1 \pm \epsilon')$ in (36) merely to simplify the notation in the proof. (This ϵ' could be, for instance, picked so that $\epsilon'(2 - \epsilon') \geq \epsilon$, where ϵ has the same meaning as in Theorem 8.)

Theorem 21. Let Φ and α be such that

$$(1 - \epsilon')^2 \|\alpha\|^2 \leq \left\| \frac{1}{\sqrt{N}} \Phi \alpha \right\|^2 \leq (1 + \epsilon')^2 \|\alpha\|^2, \quad (36)$$

with probability exceeding $\delta > 0$, and let $f = \frac{1}{\sqrt{N}} \Phi \alpha + \nu$, where the noise samples $\nu(x)$ are iid complex Gaussian random variables with zero mean and variance $2\sigma^2$. Then, for $\gamma \geq 0$,

$$(1 - \epsilon' - \gamma)^2 \|\alpha\|^2 \leq \|f\|^2 \leq (1 + \epsilon' + \gamma)^2 \|\alpha\|^2, \quad (37)$$

with probability greater than $1 - 2 \left(\delta + \mathcal{S} \left[\frac{\gamma \|\alpha\|}{\sigma} \right] \right)$, where

$$\mathcal{S}(r) \doteq \left(\int_r^\infty e^{-y^2/2} y^{N-1} dy \right) \left(\int_0^\infty e^{-y^2/2} y^{N-1} dy \right)^{-1}$$

Proof: First consider the probability that $\|f\|$ exceeds the upper bound in (37). Setting $g = \frac{1}{\sqrt{N}} \Phi \alpha$, we have

$$\begin{aligned} & \Pr[\|f\| \geq (1 + \epsilon' + \gamma) \|\alpha\|] \\ & \leq \Pr[\|g\| + \|\nu\| \geq (1 + \epsilon' + \gamma) \|\alpha\|] \\ & \leq \Pr[\|g\| \geq (1 + \epsilon') \|\alpha\|] + \Pr[\|\nu\| \geq \gamma \|\alpha\|] \\ & \leq \delta + \frac{1}{(2\pi\sigma^2)^{N/2}} \int_{\|y\| > \gamma \|\alpha\|} \exp\left(-\frac{1}{2\sigma^2} \|y\|^2\right) d^N y \\ & = \delta + \frac{1}{(2\pi)^{N/2}} \int_{\|u\| > \gamma \|\alpha\|/\sigma} e^{-\|u\|^2/2} d^N u \end{aligned}$$

The estimate for $\Pr[\|f\| \leq (1 - \epsilon' - \gamma) \|\alpha\|]$ is similar, and the desired bound then follows from the union bounds. ■

B. Noisy Signals

If the signal α is contaminated by white gaussian noise then the measurements are given by

$$y = \frac{1}{\sqrt{N}} (\Phi \alpha + \Phi \mu), \quad (38)$$

where μ is complex multivariate Gaussian distributed, with zero mean and covariance

$$E(\mu \mu^\dagger) = 2\sigma^2 I_{C \times C}. \quad (39)$$

The reconstruction algorithm thus needs to recover the signal from the noisy measurements

$$y = f + \nu, \quad (40)$$

where $\nu = \frac{1}{\sqrt{N}} \Phi \mu$ is complex multivariate Gaussian distributed with mean zero and covariance

$$\mathbb{E}(\nu \nu^\dagger) = \frac{2\sigma^2}{N} \Phi \Phi^\dagger. \quad (41)$$

The deterministic compressive sensing schemes considered in this paper have some advantage over random compressive sensing schemes in that $\frac{1}{\sqrt{N}} \Phi \left(\frac{1}{\sqrt{N}} \Phi^\dagger \right) = \frac{C}{N} I_{N \times N}$ and consequently $E(\nu(x) \overline{\nu(x')}) = \frac{2\sigma^2 C}{N} \delta_{x,x'}$, i.e., the noise samples on distinct measurements are independent. One can thus use the estimates of the previous subsection again. Noise of this type is of course harder to deal with; this is illustrated here by the measurement variance being a (possibly huge) factor C/N larger than the source noise variance σ^2 .

VII. CONCLUSIONS

We have provided simple criteria, that when satisfied by a deterministic sensing matrix, guarantee successful recovery of all but an exponentially small fraction of k -sparse signals. These criteria are satisfied by many families of deterministic sensing matrices including those formed from subcodes of the second order binary Reed Muller codes. The criteria also apply to random Fourier ensembles, where they improve known bounds on the number of measurements required for sparse reconstruction. Our proof of unique reconstruction uses a version of the classical McDiarmid Inequality that may be of independent interest.

We have described a reconstruction algorithm for Reed Muller sensing matrices that takes special advantage of the code structure. Our algorithm requires only vector-vector multiplication in the measurement domain, and as a result, reconstruction complexity is only quadratic in the number of measurements. This improves upon standard reconstruction algorithms such as Basis and Matching Pursuit that require matrix-vector multiplication and have complexity that is super-linear in the dimension of the data domain.

ACKNOWLEDGMENT

The authors would like to thank Lorne Applebaum, Richard Baraniuk, Doug Cochran, Ingrid Daubechies, Anna Gilbert, Shamgar Gurevich, Ronnie Hadani, Piotr Indyk, Robert Schapiro, Joel Tropp, Rachel Ward, and the anonymous reviewers for their insights and helpful suggestions.

REFERENCES

- [1] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss, "Combining geometry and combinatorics: a unified approach to sparse signal recovery.," *46th Annual Allerton Conference on Communication, Control, and Computing*, pp. 798-805, September 2008.
- [2] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, Vol. 59 (8) , pp. 1207-1223., 2006.
- [3] A. Gilbert, M. Strauss, J. Tropp, and R. Vershynin, "One sketch for all: fast algorithms for compressed sensing," *Proceedings of the thirty-ninth annual ACM Symposium on Theory of Computing (STOC)*, pp. 237-246, 2007.
- [4] G. Cormode and S. Muthukrishnan, "Combinatorial algorithms for Compressed Sensing," *In Proceedings of 40th Annual Conference on Information Sciences and Systems (CISS)*, Princeton, 2006.
- [5] S. Jafarpour, W. Xu, B. Hassibi, and R. Calderbank, "Efficient compressed Sensing using Optimized Expander Graphs," *to appear in IEEE Transactions on Information Theory*, 2009.
- [6] P. Indyk and M. Ruzic, "Near-optimal sparse recovery in the ℓ_1 norm," *49th Annual IEEE Symposium on Foundations of Computer Science, 2008 (FOCS '08)*, pp. 199-207, 2008.
- [7] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples.," *Applied and Computational Harmonic Analysis*, Vol. 26 (3), pp. 301-321, May 2009.
- [8] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing: Closing the gap between performance and complexity," *to appear in IEEE Transactions on Information Theory*, 2009.
- [9] D. Donoho, "Compressed Sensing," *IEEE transactions on Information Theory*, Vol. 52 (4), pp. 1289-1306, April 2006.
- [10] E. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies," *IEEE Transactions on Information Theory*, Vol. 52 (12), pp. 5406-5425, December 2006.
- [11] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE transactions on Information Theory*, Vol. 52 (2), pp. 489-509, 2006.
- [12] R. A. DeVore, "Deterministic constructions of compressed sensing matrices," *Journal of Complexity*, Vol. 23 (4-6), pp. 918-925, August-December 2007.
- [13] P. Indyk, "Explicit constructions for compressed sensing of sparse signals," *Proceedings of the 19th annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 30-33, January 2008.
- [14] B. Kashin, "The widths of certain finite dimensional sets and classes of smooth functions," *Izvestia*, vol. 41, pp. 334-351, 1977.
- [15] E. D. Gluskin, "Norms of random matrices and widths of finite dimensional sets," *Math USSR Sbornik*, vol. 48, pp. 173-182, 1984.
- [16] W. Bajwa, J. Haupt, G. Raz, S. Wright, and R. Nowak, "Toeplitz-structured compressed sensing matrices," *Statistical Signal Processing. IEEE/SP 14th Workshop on Publication*, pp. 294-298, August 2007.
- [17] J. Tropp, M. Wakin, M. Duarte, D. Baron, and R. Baraniuk, "Random filters for compressive sampling and reconstruction," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. III, pp. 872-875, May 2006.
- [18] V. Guruswami, J. Lee, and A. Razborov, "Almost euclidean subspaces of ℓ_1 via expander codes," *Proceedings of the 19th annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 353-362, January 2008.
- [19] D. Baron, Sh. Sarvotham, and R. Baraniuk, "Bayesian Compressed Sensing via Belief Propagation," *Rice ECE Department Technical Report TREE 0601*, 2006.
- [20] M. Akakaya, J. Park, and V. Tarokh, "Compressive Sensing Using Low Density Frames.," *Preprint*, 2009.
- [21] M. Akcakaya and V. Tarokh, "A frame construction and a universal distortion bound for sparse representations," *IEEE Transactions on Signal Processing*, Vol. 56 (6) , pp. 2443-2450, June 2008.
- [22] W. Xu and B. Hassibi, "Efficient compressive sensing with deterministic guarantees using expander graphs.," *IEEE Information Theory Workshop*, Vol. III, Lake Tahoe, pp. 414-419, September 2007.
- [23] L. Applebaum, S. Howard, S. Searle, and R. Calderbank, "Chirp sensing codes: Deterministic compressed sensing measurements for fast recovery," *Applied and Computational Harmonic Analysis*, Vol. 26 (2), pp. 283-290, March 2009.
- [24] S. Howard, R. Calderbank, and S. Searle, "A fast reconstruction algorithm for deterministic compressive sensing using second order Reed-Muller codes," *Conference on Information Sciences and Systems (CISS)*, Princeton, ISBN: 978-1-4244-2246-3, pp: 11 - 15, March 2008.
- [25] R. Calderbank, S. Howard, and S. Jafarpour, "A sub-linear algorithm for Sparse Reconstruction with ℓ_2/ℓ_2 Recovery Guarantees," *Preprint*, 2009.
- [26] M. R. de Prony, "Essai experimentalle et analytique," *J. Ecole Polytech.Paris*, Vol. 1, pp. 24-76, 1795.
- [27] J. Wolf, "Decoding of Bose-Chaudhuri-Hocquenghem Codes and Prony's method for curve fitting," *IEEE Trans. Inform. Theory*, Vol. 3, p. 608, 1967.
- [28] Sh. Gurevich and R. Hadani, "Incoherent dictionaries and the statistical restricted isometry property," *Preprint*, 2008.
- [29] R. Horn and C. Johnson, "Matrix Analysis," *Cambridge University Press*, 1985.
- [30] A. R. Hammons, P. V. Kumar, A. R. Calderbank, N. J. A. Sloane, and P. Sole, "The \mathbb{Z}_4 -linearity of Kerdock Codes, Preparata, Goethals, and related codes," *IEEE Transactions on Information Theory*, Vol. 40 (2), pp. 301-319, March 1994.
- [31] A.M. Kerdock, "A class of low-rate nonlinear binary codes," *Information and Control*, vol. 20, pp.182-187, 1972.
- [32] P. Delsarte and J. M. Goethals, "Alternating bilinear forms over $\text{GF}(q)$," *Journal of Combinatorial Theory*, vol. 19, pp. 26-50, 1975.
- [33] F.J. MacWilliams and N.J.A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland: Amsterdam, 1977.
- [34] N. Ailon and E. Liberty, "Fast dimension reduction using Rademacher series on dual BCH codes," *Proceedings of the nineteenth annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 215-224, January 2008.
- [35] C. McDiarmid, "On the method of bounded differences (norwich, 1989)," *Surveys in combinatorics*, PP. 148-188, Cambridge Univ. Press, Cambridge, 1989.
- [36] J. Tropp, "On The Conditioning of Random Subdictionaries," *Applied and Computational Harmonic Analysis*, Vol. 25, pp. 124, 2008.

APPENDIX A
PROPERTIES OF DELSARTE-GOETHALS SENSING
MATRICES

First we prove that the columns of the r^{th} Delsarte-Goethals sensing matrix form a group under pointwise multiplication.

Proposition A.1. *Let $\mathcal{G} = \mathcal{G}(m, r)$ be the set of column vectors $\varphi_{P,b}$ where*

$$\varphi_{P,b}(x) = i^{wt(d_P)+2wt(b)} i^{xPx^\top+2bx^\top}, \text{ for } x \in \mathbb{F}_2^m$$

where $b \in \mathbb{F}_2^m$ and where the binary symmetric matrix P varies over the Delsarte-Goethals set $DG(m, r)$. Then \mathcal{G} is a group of order $2^{(r+2)m}$ under pointwise multiplication.

Proof: We have

$$\begin{aligned} & \varphi_{P,b}(x)\varphi_{P',b'}(x) \\ &= i^{wt(d_P)+wt(d_{P'})+2wt(b\oplus b')} i^{x(P+P')x^\top+2(b\oplus b')x^\top} \end{aligned}$$

where \oplus is used to emphasize addition in \mathbb{F}_2^m . Write $P+P' = (P\oplus P')+2Q \pmod{4}$ where Q is a binary symmetric matrix. Observe that $2xQx^\top = 2d_Qx^\top \pmod{4}$, where the diagonal $d_Q = d_P * d_{P'}$ is a pointwise product of d_P and $d_{P'}$.

Thus

$$\begin{aligned} & \varphi_{P,b}(x)\varphi_{P',b'}(x) \\ &= i^{([wt(d_P)+wt(d_{P'})+2wt(d_P*d_{P'}])+2wt(b\oplus b'\oplus d_P*d_{P'}))} \\ & \quad i^{x(P+P')x^\top+2(b\oplus b'\oplus d_P*d_{P'})x^\top} \\ &= \varphi_{P\oplus P', b\oplus b'\oplus d_P*d_{P'}}(x), \end{aligned}$$

and \mathcal{G} is closed under pointwise multiplication. Hence the possible inner products of columns $\varphi_{P,d}, \varphi_{P',d'}$ are exactly the possible column sums for columns $\varphi_{Q,b}$ where $Q = P\oplus P'$. ■

Next we verify property (St3).

Proposition A.2. *Let Q be a binary symmetric $m \times m$ matrix with rank r and let $b \in \mathbb{F}_2^m$. If*

$$S = \sum_x i^{xQx^\top+2bx^\top}$$

then either $S = 0$ or

$$S^2 = i^{z_1Qz_1^\top+2bz_1^\top} 2^{2m-r}, \quad \text{where } z_1Q = d_Q.$$

Proof: We have

$$\begin{aligned} S^2 &= \sum_{x,y} i^{xQx^\top+yQy^\top+2b(x+y)^\top} \\ &= \sum_{x,y} i^{(x+y)Q(x+y)^\top+2*Qy^\top+2b(x+y)^\top} \end{aligned}$$

Changing variables to $z = x \oplus y$ and y gives

$$S^2 = \sum_z i^{zQz^\top} \sum_y (-1)^{(d_Q+zQ)y^\top}.$$

Since the diagonal d_Q of a binary symmetric matrix Q is contained in the row space of Q there exists a solution $zQ = d_Q$. The solutions to the equation $zQ = 0$ form a vector space E of dimension $m - r$, and for all $e, f \in E$

$$eQe^\top + fQf^\top = (e+f)Q(e+f)^\top \pmod{4}.$$

Hence

$$\begin{aligned} S^2 &= 2^m \sum_{e \in E} i^{(z_1+e)Q(z_1+e)^\top+2(z_1+e)b^\top} \\ &= 2^m i^{z_1Qz_1^\top+2z_1b^\top} \sum_{e \in E} i^{eQe^\top+2eb^\top}. \end{aligned}$$

The map $e \rightarrow eQe^\top$ is a linear map from E to \mathbb{Z}_2 , so the numerator $eQe^\top + 2eb^\top$ also determines a linear map from E to \mathbb{Z}_2 (here we identify \mathbb{Z}_2 and $2\mathbb{Z}_4$). If this linear map is the zero map then

$$S^2 = 2^{2m-r} i^{z_1Qz_1^\top+2bz_1^\top},$$

and if it is not zero then $S = 0$. Note that given $e \rightarrow eQe^\top$, there are 2^n ways to choose b so that $e \rightarrow eQe^\top + 2eb^\top$ is the zero map. ■

The 0^{th} Delsarte-Goethals sensing matrix is a matrix with $N = 2^m$ rows and N^2 columns. These columns are the union of N mutually unbiased bases, where vectors in one orthogonal basis look like noise to all other orthogonal bases.

APPENDIX B
GENERALIZED MCDIARMID'S INEQUALITY

The method of ‘‘independent bounded differences’’ ([35]) gives large-deviation concentration bounds for multivariate functions in terms of the maximum effect on the function value of changing just one coordinate. This method has been widely used in combinatorial applications, and in learning theory. In this appendix, we prove that a modification of McDiarmid’s inequality also holds for *distinct* (in contrast to *independent*) random variables; our proof consists again in forming martingale sequences.

We first introduce some notation. Let $\mathcal{X}_1, \dots, \mathcal{X}_m$ be probability spaces and define \mathcal{X} as the probability space of all distinct m -tuples, that is,

$$\mathcal{X} \doteq \{(x_1, \dots, x_m) \in \prod_{i=1}^m \mathcal{X}_i \text{ such that } \forall i \neq j : x_i \neq x_j\}. \quad (42)$$

(This definition is spelled out in more detail at the end of subsection 3.1.2.) Let $f(x_1, \dots, x_m)$ be a function from \mathcal{X} to \mathbb{R} , and let $f(X_1, \dots, X_m)$ be the corresponding random variable on \mathcal{X} . Denote by $X_{1 \rightarrow i}$ the i -tuple of random variables (X_1, \dots, X_i) on the probability space \mathcal{X} . (The ‘‘complete’’ m -tuple (X_1, \dots, X_m) will also be denoted by just X .) Analogously, define $X_{(i+1) \rightarrow m}$ to be the $(m-i)$ -tuple of random variables $(X_{(i+1)}, \dots, X_m)$. We shall also use the notations $x_{1 \rightarrow i} \doteq (x_1, \dots, x_i) \in \prod_{\ell=1}^i \mathcal{X}_\ell$, $\mathcal{X}_{1 \rightarrow i} \doteq \{x_{1 \rightarrow i} \in \prod_{n=1}^i \mathcal{X}_n; x_\ell \neq x_n \text{ if } \ell \neq n\}$; $x_{(i+1) \rightarrow m} \in \prod_{\ell=(i+1)}^m \mathcal{X}_\ell$ and $\mathcal{X}_{(i+1) \rightarrow m} \subset \prod_{n=(i+1)}^m \mathcal{X}_n$ are defined analogously.

Theorem B.1 (Self-avoiding McDiarmid inequality). *Let \mathcal{X} be the probability space defined in Equation (42), and let $f : \mathcal{X} \rightarrow \mathbb{R}$ be a function such that for any index i , and any $x_{1 \rightarrow (i-1)} \in \mathcal{X}_{1 \rightarrow (i-1)}$,*

$$\begin{aligned} & \sup_{u \in \mathcal{X}_i; u \neq x_n, n=1 \rightarrow i} \mathbb{E}[f(x_1, \dots, x_{i-1}, u, X_{i+1}, \dots, X_m)] \\ & - \inf_{l \in \mathcal{X}_i; l \neq x_n, n=1 \rightarrow i} \mathbb{E}[f(x_1, \dots, x_{i-1}, l, X_{i+1}, \dots, X_m)] \leq c_i. \end{aligned} \quad (43)$$

Then for any positive ϵ ,

$$\Pr [|f(X_1, \dots, X_m) - \mathbb{E}[f(X_1, \dots, X_m)]| \geq \epsilon] \leq 2 \exp\left(\frac{-2\epsilon^2}{\sum_{i=1}^m c_i^2}\right).$$

Our proof will invoke Hoeffding's Lemma [[35]]

Proposition B.2 (Hoeffding's Lemma). *Let X be a random variable with $\mathbb{E}[X] = 0$ and $a \leq X \leq b$ then for $t > 0$*

$$\mathbb{E}[e^{tX}] \leq \exp\left\{\frac{t^2(b-a)^2}{8}\right\}.$$

In our proof we will also make use of the functions

$$Z_i(x_{1 \rightarrow i}) \doteq \mathbb{E}[f(X) | X_{1 \rightarrow i} = x_{1 \rightarrow i}] \quad \text{where } x_{1 \rightarrow i} \in \mathcal{X}_{1 \rightarrow i}$$

As a result, for all $x_{1 \rightarrow (i-1)}$ in $\mathcal{X}_{1 \rightarrow (i-1)}$

$$\left| \sup_{u \in \mathcal{X}_i; u \neq x_n, n=1 \rightarrow i} Z_i(x_{1 \rightarrow (i-1)}, u) - \inf_{l \in \mathcal{X}_i; l \neq x_n, n=1 \rightarrow i} Z_i(x_{1 \rightarrow (i-1)}, l) \right|$$

is less than c_i . This implies, for all $x_{1 \rightarrow i} \in \mathcal{X}_{1 \rightarrow i}$,

$$\begin{aligned} -c_i &\leq \inf_{l \in \mathcal{X}_i; l \neq x_n, n=1 \rightarrow i-1} Z_i(x_{1 \rightarrow (i-1)}, l) \\ &\quad - \sup_{u \in \mathcal{X}_i; u \neq x_n, n=1 \rightarrow i-1} Z_i(x_{1 \rightarrow (i-1)}, u) \\ &\leq Z(x_{1 \rightarrow i}) \end{aligned} \quad (46)$$

$$\begin{aligned} &- \mathbb{E}[f(X_{1 \rightarrow (i-1)}, X_i, X_{(i+1) \rightarrow m}) | x_{1 \rightarrow (i-1)}] \\ &= Z_i(x_{1 \rightarrow i}) - Z_{i-1}(x_{1 \rightarrow (i-1)}) \\ &\leq Z_i(x_{1 \rightarrow i}) \end{aligned} \quad (47)$$

$$\begin{aligned} &- \inf_{l \in \mathcal{X}_i; l \neq x_n, n=1 \rightarrow i-1} Z_i(x_{1 \rightarrow (i-1)}, l) \\ &\leq \sup_{u \in \mathcal{X}_i; u \neq x_n, n=1 \rightarrow i-1} Z_i(x_{1 \rightarrow (i-1)}, u) \end{aligned} \quad (48)$$

$$\begin{aligned} &- \inf_{l \in \mathcal{X}_i; l \neq x_n, n=1 \rightarrow i-1} Z_i(x_{1 \rightarrow (i-1)}, l) \\ &\leq c_i, \end{aligned}$$

or

$$|Z_i(x_{1 \rightarrow i}) - Z_{i-1}(x_{1 \rightarrow (i-1)})| \leq c_i \quad (49)$$

Until now, we have viewed each Z_i as a function on the subset $\mathcal{X}_{1 \rightarrow i}$ of $\prod_{\ell=1}^i \mathcal{X}_\ell$; it is straightforward to lift the Z_i to functions on all of \mathcal{X} . The $Z_i(X_{1 \rightarrow i}) = Z_i(X)$ can also be considered as random variables on \mathcal{X} , depending only on the first i components of X ,

$$Z_i(X_{1 \rightarrow i}) = \mathbb{E}_{X_{(i+1) \rightarrow m}}[f(X) | X_{1 \rightarrow i}]$$

(The subscript $X_{(i+1) \rightarrow m}$ on the expectation indicates that one averages only with respect to the variables listed in the subscript, in this case the last $m - i$ variables. We adopt this subscript convention in what follows; only expectations without subscript are with respect to the whole probability space \mathcal{X} .)

Viewing the Z_i as random variables, we observe that $Z_0 = \mathbb{E}[f(X_1, \dots, X_m)]$, and that $Z_m = f(X_1, \dots, X_m)$. Because of the restriction to \mathcal{X} , the random variables X_ℓ, Z_ℓ are not independent. However, with respect to averaging over X_i , the $Z_i, i = 1, \dots, m$ constitute a martingale in the following sense:

$$\mathbb{E}_{X_i}[Z_i(X) | X_{1 \rightarrow (i-1)}] = Z_{i-1}(X), \quad (50)$$

Proof: Using Markov's inequality, we see that for any positive t

$$\begin{aligned} \Pr[f - \mathbb{E}[f] \geq \epsilon] &= \Pr[e^{t(f - \mathbb{E}[f])} \geq e^{t\epsilon}] \\ &\leq e^{-t\epsilon} \mathbb{E}[e^{t(f - \mathbb{E}[f])}] \end{aligned} \quad (51)$$

Since $f - \mathbb{E}[f] = Z_m - Z_0$, we can rewrite this as

$$\mathbb{E}[e^{t(f - \mathbb{E}[f])}] = \mathbb{E}\left[\exp\left(t \sum_{i=1}^m (Z_i - Z_{i-1})\right)\right]$$

By marginalization of the expectation,

$$\begin{aligned} &\mathbb{E}\left[\exp\left(t \sum_{i=1}^m (Z_i - Z_{i-1})\right)\right] \\ &= \mathbb{E}_{X_{1 \rightarrow (m-1)}}\left[\mathbb{E}_{X_m}\left[\exp\left(t \sum_{i=1}^m (Z_i - Z_{i-1})\right) | X_{1 \rightarrow (m-1)}\right]\right] \\ &= \mathbb{E}\left[\exp\left(\sum_{i=1}^{m-1} (Z_i - Z_{i-1})\right) \mathbb{E}_{X_m}\left[e^{t(Z_m - Z_{m-1})} | X_{1 \rightarrow (m-1)}\right]\right], \end{aligned} \quad (45)$$

where we have used that each Z_i depends on only the first i components of X , so that only $(Z_{m-1} - Z_m)$ is affected by the averaging over X_m .

By (49), we have, for all $x_{1 \rightarrow i} \in \mathcal{X}_{1 \rightarrow i}$, $|Z_i(x_{1 \rightarrow i}) - Z_{i-1}(x_{1 \rightarrow (i-1)})| \leq c_i$, which can also be rewritten as $-c_i \leq Z_i(X) - Z_{i-1}(X) \leq c_i$.

Because of the martingale property (50) we have $\mathbb{E}[Z_i - Z_{i-1} | X_1^{i-1}] = \mathbb{E}_{X_{(i+1) \rightarrow m}}[Z_i - Z_{i-1} | X_1^{i-1}] = \mathbb{E}_{X_i}[Z_i - Z_{i-1} | X_1^{i-1}] = 0$.

Combining these last two observations with Hoeffding's Lemma [35] we conclude

$$\begin{aligned} &\mathbb{E}_{X_{1 \rightarrow m}}\left[\exp\left(\sum_{i=1}^m (Z_i - Z_{i-1})\right)\right] \\ &= \mathbb{E}_{X_{1 \rightarrow (m-1)}}\left[e^{t \sum_{i=1}^{m-1} (Z_i - Z_{i-1})}\right] \mathbb{E}_{X_m}\left[e^{(Z_m - Z_{m-1})} | X_{1 \rightarrow (m-1)}\right] \\ &\leq e^{t^2 c_m^2 / 8} \mathbb{E}_{X_{1 \rightarrow (m-1)}}\left[\exp\left(\sum_{i=1}^{m-1} (Z_i - Z_{i-1})\right)\right] \\ &\leq \dots \leq \exp\left(\frac{1}{8} t^2 \sum_{i=1}^m c_i^2\right) \end{aligned}$$

Substituting this into (51) we obtain

$$\Pr[f - \mathbb{E}[f] \geq \epsilon] \leq \exp\left(-t\epsilon + \frac{1}{8} t^2 \sum_{i=1}^m c_i^2\right) \quad (52)$$

Since equation (52) is valid for any $t > 0$, we can optimize over t . By substituting the value $t = 4\epsilon (\sum c_i^2)^{-1}$ we get

$$\Pr[f - \mathbb{E}[f] \geq \epsilon] \leq \exp\left(\frac{-2\epsilon^2}{\sum c_i^2}\right).$$

Replacing the function f by $\mathbb{E}[f] - f$, it follows that $\Pr[f - \mathbb{E}[f] \leq -\epsilon] \leq \exp\left(\frac{-2\epsilon^2}{\sum c_i^2}\right)$; union bounds therefore imply that

$$\Pr[|f - \mathbb{E}[f]| \geq \epsilon] \leq 2 \exp\left(\frac{-2\epsilon^2}{\sum c_i^2}\right). \quad \blacksquare$$