A Fast Reconstruction Algorithm for Deterministic Compressive Sensing using Second Order Reed-Muller Codes

S. D. Howard
Defence Science & Technology Organisation
PO Box 1500
Edinburgh 5111, Australia.

A. R. Calderbank
Electrical Engineering
Princeton University
NJ 08544, USA.

S. J. Searle
Electronic & Electrical Engineering
The University of Melbourne
Victoria 3010 Australia.

***Abstract*—This paper proposes a deterministic compressed sensing matrix that comes by design with a very fast reconstruction algorithm, in the sense that its complexity depends only on the number of measurements n and not on the signal dimension N. The matrix construction is based on the second order Reed-Muller codes and associated functions. This matrix does not have RIP uniformly with respect to all k-sparse vectors, but it acts as a near isometry on k-sparse vectors with very high probability.**

I. Introduction

The goal of compressed sensing is to minimize the number of measurements needed to capture a signal. This is most simply illustrated in the discrete problem where one wants to capture a vector $x \in \mathbb{R}^N$ with N large by making a small number n of linear measurements. The information y extracted from x by such a sensing system is the vector $y = \Phi x \in \mathbb{R}^n$ where Φ is an $n \times N$ matrix. The two main questions in Compressed Sensing are: (i) what are the best matrices Φ and how do we construct them?, (ii) how do we decode the information y to find x (or a good approximation to x) and how to do this in a computationally efficient manner?

A significant contribution to the underlying geometry of good sensing systems was made a few years ago through the work of Candés and Tao [1], [2] and Donoho [3]. One particularly elegant way of framing a sufficient condition on Φ, formulated by Candés and Tao, is known as the *Restricted Isometry Property* (RIP). A matrix is said to have the RIP of order k if Φ acts as a near isometry on *all* k sparse vectors.

Since the introduction of RIP, many approaches have been proposed to the challenging problem of constructing matrices that have the RIP. For small, and asymptotically vanishing values of the ratios n/N and/or k/n, examples can be constructed using standard results in approximation theory [4] or coding theory [5]. Larger size examples are given by probabilistic constructions: matrices Φ in which the $n \times N$ entries are generated by an i. i. d Gaussian or Bernoulli process are know to have the RIP with high probability [6], [7] up to $n \approx N/2$ and $k = O(N/\log N)$. Constructions of similar size random matrices Φ that have the RIP, but with a smaller degree of randomness are also given by several approaches involving filtering [8], [9], or expander graphs [10]; the latter approach may even lead to constructions of Φ of sizes comparable to the random examples.

The decoding of the information $y = \Phi x$ gathered by a sensing system is equally intriguing and dependent on high dimensional geometry. The null space $\mathcal{N}$ of Φ has large dimension ($\geq N - n$). Any vector in $\mathcal{F}(y) = x + \mathcal{N}$ will yield the same information y. It is known that the optimal sensing matrices must have a null space with a democratic structure [11]. This means that the energy of any element in the null space can not be concentrated in a few coordinates.

Finding x from the measured or observed data $y = \Phi x$ requires a search over $\mathcal{F}(y)$. The successful decoders take advantage of the geometry of $\mathcal{F}(y)$. The most prominent example, called Basis Pursuit (BP) (see [3], [2], [1]), decodes y by taking the vector $x^* \in \mathcal{F}(y)$ with smallest ℓ_1 norm.

MacKay [12, Chapter 13] makes the point that in coding theory distance is not everything and he argues that the minimum distance of a code is not of fundamental importance to the goal of achieving reliable communication over noisy channels. He argues that there are codes with blocklength 10,000 and fast decoding algorithms that are known to have many codewords of weight 32, and that can nevertheless correct errors of weight 320 with tiny error probability. Reliable communication is achieved by focusing on typical rather than worst case performance, and this involves constraining the spectrum of all possible distances rather than simply the minimum distance.

Extending to compressed sensing MacKay's point with respect to coding, we propose a deterministic compressed sensing matrix that comes by design with a very fast reconstruction algorithm, in the sense that its complexity depends only on the number of measurements n and not on the signal dimension N. The matrix construction is based on the second order Reed-Muller codes and associated functions. This matrix does not have RIP uniformly with respect to all k-sparse vectors, but it acts as a near isometry on k-sparse vectors with very high probability.

978-1-4244-2247-0/08/$25.00 ©2008 IEEE.

II. Second Order Reed-Muller Functions

There is a duality between time and frequency in the continuous world where sinusoids are eigenfunctions of time shifts. This section describes binary Hamming space, where there are discrete counterparts of time and frequency shifts, where Walsh functions are the counterparts of sinusoids, and where second order Reed-Muller functions are the counterparts of chirps.

Hamming space is the vector space $\mathbb{Z}_2^m$ of binary vectors of length m. If $a = (a_0, \cdots, a_{m-1})^T$ and $b = (b_0, \cdots, b_{m-1})^T \in \mathbb{Z}_2^m$ are two binary vectors, then

$$a + b = ((a_0 + b_0), \cdots, (a_{m-1} + b_{m-1})) \mod 2, \quad (1)$$

and we can define an inner product

$$a^T b = \sum_{j=0}^{m-1} a_j b_j \mod 2. \quad (2)$$

As a short hand we will often just write $a = a_0 a_1 \dots a_{m-1}$.

The *first-order Reed-Muller functions*, also called the *Walsh functions*, are functions $\mathbb{Z}_2^m \to \mathbb{R}$ defined by

$$\phi_{0,b}(a) = \frac{1}{\sqrt{2^m}} (-1)^{b^T a}. \quad (3)$$

There are 2^m Walsh functions on $\mathbb{Z}_m^2$, one for each $b \in \mathbb{Z}_2^m$, and they form the rows of a Hadamard matrix H_m; for $m = 2$

$$H_2 = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

The Walsh functions form an orthonormal basis for $\mathbb{R}^{2^m}$ and so H_m is unitary. The Walsh functions can be considered the "sinusoids" of the binary world, with b defining a multidimensional binary frequency. The value of its j^{th} bit indicates whether the function changes sign when the j^{th} bit of the vector a is flipped. We view translation by the binary vector e as a discrete multidimensional time shift and observe that Walsh functions are eigenfunctions of such shifts:

$$\phi_{0,b}(a + e) = (-1)^{b^T e} \phi_{0,b}(a). \quad (4)$$

We can also define the binary world equivalent of "chirps". A sinusoid is linearly chirped when it frequency changes linearly with time. In the binary case we construct functions

$$\phi_{\mathsf{P},b}(a) = \frac{(-1)^{\text{wt}(b)}}{\sqrt{2^m}} i^{(2b + \mathsf{P}a)^T a} \quad (5)$$

where P is a binary symmetric matrix. These functions, which are parameterized by all binary symmetric matrices P and binary vectors b, are called the *second order Reed-Muller functions*. The Walsh functions are associated with $\mathsf{P} = 0$. We note here that we have had to move from working in $\mathbb{Z}_2$ to $\mathbb{Z}_4$ to accommodate all possible "chirps". Thus, although all quantities in the exponent in (5) are defined on $\mathbb{Z}_2$, the exponent itself is to be computed modulo 4. In what follows we will also restrict to the set of functions where the diagonal of the matrix P is constrained to be identically zero. These functions are real valued. The factor $(-1)^{\text{wt}(b)}$, where $\text{wt}(b)$ denotes the weight of the b, i.e. the number of 1s the vector contains, is included to make the distribution of inner products symmetric.

The second-order Reed-Muller functions have a number important properties. For a fixed binary symmetric matrix P the set

$$\mathcal{F}_{\mathsf{P}} = \{\phi_{\mathsf{P},b} | b \in \mathbb{Z}_2^m\}, \quad (6)$$

forms an orthonormal basis for $\mathbb{R}^{2^m}$. Each function in this orthonormal basis is an eigenfunction of a commutative group of time-frequency shift operators determined by P [13].

Given a vector $\psi \in \mathcal{F}_{\mathsf{P}}$ the inner product

$$|(\psi, \chi)| = \begin{cases} 1/\sqrt{2^\ell}, & 2^\ell \text{ times}, \\ 0, & 2^m - 2^\ell \text{ times}. \end{cases} \quad (7)$$

as χ ranges over $\mathcal{F}_{\mathsf{Q}}$, where

$$\ell = \text{rank}(\mathsf{P} - \mathsf{Q}). \quad (8)$$

Note that this depends only on ℓ and not on the choice of P and Q.

Classical results in the theory of alternating bilinear forms on $\mathbb{Z}_2^m$ [14, Page 43], can be used to deduce the distribution of mutual inner products of second order Reed-Muller functions. The number of zero diagonal $m \times m$ binary symmetric matrices with rank r is given by

$$v_{2h} = 2^{h(h-1)} \frac{(2^m - 1)(2^{m-1} - 1) \dots (2^{m-2h+1} - 1)}{(2^{2h} - 1)(2^{2h-2} - 1) \dots (2^2 - 1)}, \quad (9)$$

and $v_{2h+1} = 0$, for $h = 0, \cdots, [m/2]$. Together, (9) and (7) lead to the distribution of inner product values

Value	Proportion
$-1/2^h$	$v_{2h} 2^{2h-1} / 2^{m(m+1)/2}$
0	$1 - 2\sum_h v_{2h} 2^{2h-1} / 2^{m(m+1)/2}$
$1/2^h$	$v_{2h} 2^{2h-1} / 2^{m(m+1)/2}$

for $h = 1, \cdots, [m/2]$. This distribution is intimately related to the weight distribution of the second order Reed-Muller codes [14], [15].

The possible inner product values and proportions are shown in Figure 1 for the case $m = 10$. We note that that the proportion drops off rapidly with the magnitude of the inner product.

III. The Compressed Sensing Matrix

In this paper we will restrict consideration to the case of real second-order Reed-Muller functions, and so all matrices P will have zero main diagonals. We note, however, that it is straightforward to generalize the results given below to the complex case. Let U_{P} be the unitary matrix corresponding

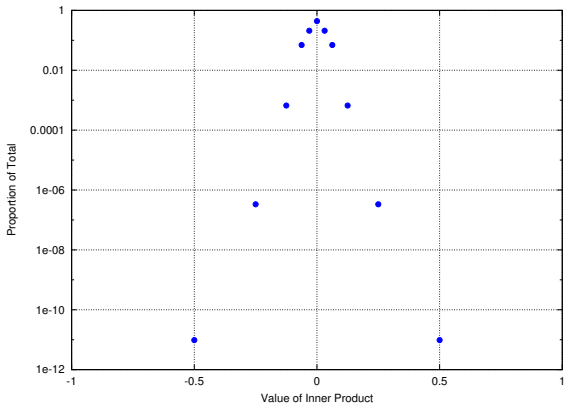


Fig. 1. Distribution of mutual inner products for second order Reed-Muller functions for $m = 10$.

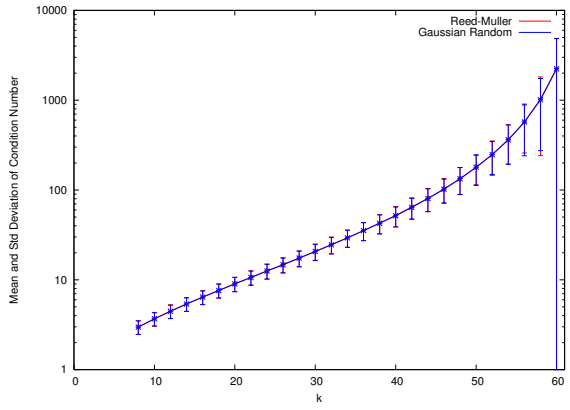


Fig. 2. Mean and standard deviation for the condition number of k-Gram matrices for Φ_{RM}, with $m = 6$, compared to that of a Gaussian random matrix of the same size.

to the orthonormal basis $\mathcal{F}_{\mathsf{P}}$. We propose a $2^m \times 2^{m(m+1)/2}$ compressed sensing matrix of the form

$$\Phi_{RM} = \begin{pmatrix} U_{\mathsf{P}_1} & U_{\mathsf{P}_2} & \cdots & U_{\mathsf{P}_{2^{m(m-1)/2}}} \end{pmatrix}. \qquad (10)$$

Apart from an overall normalization factor of $1/\sqrt{2^m}$ the entries in this matrix are ± 1. The columns of Φ_{RM} form a tight frame for $\mathbb{R}^{2^m}$ with redundancy $2^{m(m-1)/2}$, that is,

$$\Phi_{RM}\Phi_{RM}^{\dagger} = 2^{m(m-1)/2} I_{2^m \times 2^m}. \qquad (11)$$

In the case m even, it will prove useful to further restrict the set of P matrices to sets DG_{2h} of zero-diagonal symmetric binary $m \times m$ matrices with the property that for any distinct pair of matrices $\mathsf{P}, \mathsf{Q} \in DG_{2h}$, the rank of $\mathsf{P}+\mathsf{Q}$ is at least $m-2h$. The sets DG_{2h} are nested, so that $DG_0 \subseteq DG_2 \subseteq \cdots$, and are associated with subcodes of the second order Reed-Muller code discovered by Delsarte and Goethals [16], [14, Chapter 15]. The first subcode in this hierarchy, which is associated with DG_0, is the nonlinear binary code first discovered by Kerdock [17].

In compressed sensing, the *Restricted Isometry Property* (RIP) is the condition that that the eigenvalues of all Gram matrices determined by sets of k columns from the sensing matrix lie in an interval $[1-\delta, 1+\delta]$. This property is typical of random matrices [1]. Our compressed sensing matrix almost certainly does not have the RIP property. We know of linear dependencies between the columns when m is small and we do not expect these to disappear as m increases. However, we do know from experimentation that any linear dependencies become increasing difficult to find at random. For example, Figure 2 shows the result of a large simulation comparing the distribution of condition numbers for k-Gram matrices of Φ_{RM} for $m = 6$, with that for a Gaussian random matrix of the same size. The simulation was based on $10,000$ randomly selected k-Gram matrices for each value of k. Note that due to the size of the Gaussian random matrix involved the results would not be significantly different if a new random matrix were chosen at each stage. In fact, subsequent runs of the simulation produced almost identical plots to Figure 2. It can be seen that the results for our matrix are essentially identical to those for the Gaussian random matrix. Furthermore, direct inspection showed the histograms for each k to be statistically indistinguishable. We obtained similar results for $m = 7$. For $m = 5$, the above mentioned linear dependencies cause singular k-Gram matrices to appear quite suddenly at $k = 18$ in runs of $20,000$ simulations. We note again that no dependencies were evident for $m > 5$.

IV. Fast Reconstruction Algorithm

Our aim in constructing the deterministic compressed sensing matrix (10) was to exploit the structure of the second-order Reed-Muller functions to create computationally efficient reconstruction algorithms. In this section we describe such an algorithm. We begin with an important property of the second order Reed-Muller functions. Take any $e \in \mathbb{Z}_2^m$, then

$$\phi_{\mathsf{P},b}(a+e)\overline{\phi_{\mathsf{P},b}(a)} = \frac{1}{2^m}(-1)^{b^T e}(-1)^{e^T \mathsf{P} a} \qquad (12)$$

Note that this operation produces a Walsh function with frequency $\mathsf{P}e$. If we were presented with a unknown second order Reed-Muller function ψ (even in noise) a fast algorithm for recovering its parameters is as follows. For each unit weight $(e_j)_k = \delta_{j,k}$, compute the product in (12) and apply the fast Hadamard transform (projection onto the Walsh basis). The largest magnitude component of the Hadamard transform will be at $\mathsf{P}e_j$, which is the j^{th} column of P. One then "dechirps" ψ, by multiplying it by $\phi_{\mathsf{P},0}$, and another Hadamard transform recovers b. Thus, the parameters of an unknown second order Reed-Muller function can be recovered at the cost of $(m+1)2^m$ multiplications and the cost of searching the 2^m terms in each of the Hadamard transforms (actually, not all bins in the Hadamard transform need to be searched since the P matrices are constrained).

The noise free reconstruction problem for our compressed sensing matrix Φ_{RM} is given by

$$y = \Phi_{RM} x, \qquad (13)$$

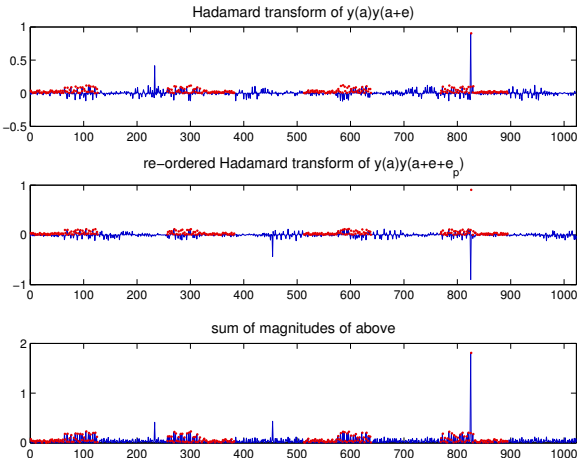


Fig. 3. An example of the determination of the second column of the P matrix of a component.

where y is the data from which we need to reconstruct the k-sparse vector x. The data vector takes the form

$$y = \sum_{\ell=1}^{k} c_\ell \phi_{\mathsf{P}_\ell, b_\ell} \tag{14}$$

Where the c_j are the k non-zero components of the vector x.

Our fast reconstruction algorithm is based on application of shift-and-multiply to the data with result

$$y(a+e)\overline{y(a)} = \frac{1}{2^m}\sum_{\ell=1}^{k} c_\ell^2 (-1)^{b_\ell^T e}(-1)^{e^T \mathsf{P}_\ell a} + \text{“chirps”}. \tag{15}$$

The right hand side of (15) is a linear combination of Walsh functions plus a term “chirps” which consists of cross terms which are all second order Reed Muller functions with non-zero Ps. Applying the Hadamard transform to (15) results in peaks at frequencies $\mathsf{P}_\ell e$ for each P_ℓ in the data. Since the second order Reed Muller functions with non-zero P have Hadamard transforms which are generally distributed in frequency, the result is a number of peaks with a general background of interference. A plot of one such Hadamard transform for data generated for $m = 10$ is shown in the first graph of Figure (3). Here peaks corresponding to the two strongest components can be seen as well as the background due to the cross terms. If we have deduced the P_ℓ of one of the components of the signal we can then robustly estimate the corresponding b_ℓ, by dechirping the data to obtain

$$y(a)\overline{\phi_{\mathsf{P},0}}(a) = c_\ell (-1)^{\mathrm{wt}(b)}(-1)^{b^T a} + \text{“chirps”} \tag{16}$$

and applying the Hadamard transform to the result.

The basic idea of our algorithm is:

1) Extract the next P_j the largest energy ($|c_j|^2$) component of the reduce data y'.
2) Determine the frequency b_j.
3) Determine the c_ℓ which minimize $\|y - \sum_{\ell=0}^{j} c_\ell \phi_{\mathsf{P}_\ell, b_\ell}\|_2$ and substact to obtain reduced data

$$y' = y - \sum_{\ell=0}^{j} \hat{c}_\ell \phi_{\mathsf{P}_\ell, b_\ell}. \tag{17}$$

4) Repeat until $\|y'\| < \epsilon$.

A difficulty encountered with this basic approach occurs when two components have Ps with a common column. When this happens two Walsh functions with the same frequency occur in (15). They may either reinforce or cancel. Cancellation is problematic and may lead to a column being missed. This problem may be averted to a large degree by noting that were we to use a vector e of weight 2 in (15) we would them obtain a combination of Walsh functions with frequencies that are the sums of pair of columns of the P matrices.

Our algorithm uses both weight 1 and 2 vectors in step 1 above to assemble the columns of the P matrix of a particular component. Figure 3 shows an example of the determination of the second column of the P matrix of a component. The first graph shows the result for $e = 010\ldots$, while the second graph shows the result for $e = 110\ldots$ but with a shift to align the peaks (since we already now the first column of P). The magnitudes of the two graphs are then added to form a detector for the second row of the matrix (third graph). The search over the Hadamard transform is limited to the area marked in red due to the symmetry of the P matrix and our knowledge of its first column.

A number of reconstruction algorithms have been proposed [1, 2, 13], all are computationally expensive for long signals, with complexity typically polynomial in the signal length N (See [18, Figure 1]). The complexity of the proposed algorithm does not depend on the length N of the signal to be reconstructed, but rather on the number of measurements n. The overall complexity of the algorithm is $O(n \log^2 n)$, where $n = 2^m$.

We have applied the algorithm to simulated data generated by applying the full matrix Φ_{RM} to k-sparse vectors. We give results here for the cases $m = 8$ and $m = 10$. The matrix Φ_{RM} has dimension 256×2^{32} for $m = 8$ and 1024×2^{55} for $m = 10$. Now we are not advocating that one apply such a matrix to data, but merely demonstrating that our algorithm makes reconstruction practical even in such an extreme situation. Figures 4 and 5 show the number of successful reconstructions as a function of the sparsity factor k. Note that in the simulation, failure was reported if there was still significant residual energy remaining after a fixed number (200) iterations of the algorithm. We note that for $m = 8$ the proportion of successful reconstructions falls away at about $k = 7$, while for $m = 10$ this occurs at $k = 20$. Baron et al. [19] have proposed a *rule of thumb* based on the analysis of Donoho and Tanner [20], [21] of the probability of successful reconstruction using basis pursuit. The rule is that the number of measurements n necessary for reliable Reconstruction satisfies $n > k \log_2(1 + N/k)$. For $m = 8, k = 7$ and $m = 10, k = 20$ this gives approximately

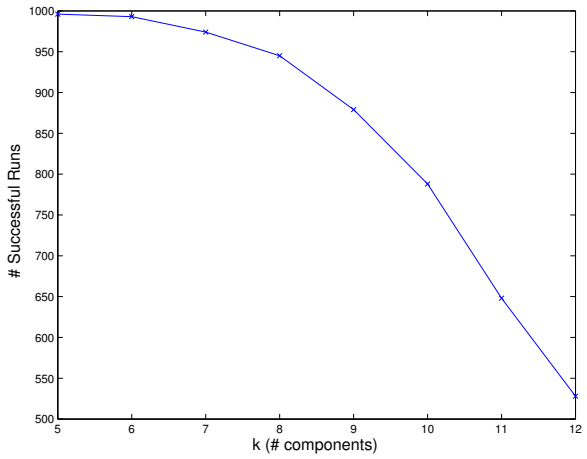


Fig. 4. The number of successful reconstructions in 1000 trials vs. sparsity factor k for $m = 8$

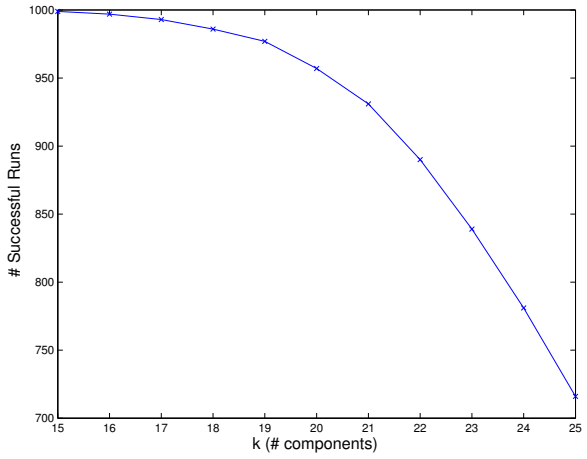


Fig. 5. The number of successful reconstructions in 1000 trials vs. sparsity factor k for $m = 10$

$n > 232$ and $n > 1013$, compared with the actual number of measurements 256 and 1024. Thus, experimentally our algorithm gives results consistent with the theoretical limit for basis pursuit.

V. Discussion

In terms of accuracy of reconstruction the number of common columns of the P matrices should be kept to a minimum. As the Φ_{RM} matrices have an abundance of columns, we can afford to place conditions of the P matrices. In particular, by placing restriction on the rank of the differences of the P matrices, that is, taking all $P \in DG_{2h}$, the performance of the algorithm is increases as the number of columns in Φ decreases.

Although we have not discussed the noise performance of the algorithm in this paper, the algorithm has been designed to perform well in noise, in fact, the noise would just add to the already present "chirp" cross terms. The only real modification the algorithm would need is to stopping criterion.

A detailed analysis of the algorithm and the proof that the matrix Φ_{RM} acts as a near isometry on k-sparse vectors with very high probability will be given elsewhere.

References

[1] E. Candés and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, pp. 4203–4215, 2005.

[2] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[3] D. Donoho. and M. Elad, "Optimally sparse representation from overcomplete dictionaries via ℓ_1 norm minimization," *Proc. Natl. Acad. Sci. USA*, pp. 2197–2002, 2003.

[4] R. A. DeVore, "Deterministic constructions of compressed sensing matrices," *preprint*, 2007.

[5] P. Indyk, "Explicit constructions for compressed sensing of sparse signals," in *19th Symposium on Discrete Algorithms*, 2008.

[6] E. Candes and T. Tao, "Near optimal signal recovery from random projections: universal encoding strategies," *IEEE Transactions on Information Theory*, vol. 52, pp. 5406–5425, 2006.

[7] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, 2006.

[8] W. Bajwa, J. Haupt, G. Raz, S. Wright, and R. Nowak, "Toeplitz-structured compressed sensing matrices," in *IEEE Workshop on Statistical Signal Processing (SSP), Madison, Wisconsin, August*, 2007.

[9] J. Tropp, M. Wakin, M. Duarte, D. Baron, and R. Baraniuk, "Random filters for compressive sampling and reconstruction," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Toulouse, France*, may 2006.

[10] R. Berinde and P. Indyk, "Sparse recovery using sparse random matrices," *preprint*, 2008.

[11] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation.," *Preprint*, 2007.

[12] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithm*. Cambridge University Press, 2003.

[13] S. D. Howard, A. R. Calderbank, and W. Moran, "The finite Heisenberg-Weyl groups in radar and communications," *EURASIP Journal on Advances in Signal Processing, Article ID 85685*, vol. 2006, pp. 1–12, 2006.

[14] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error Correcting Codes*. North-Holland Elsevier, 1983.

[15] N. Sloane and E. Berlekamp, "Weight enumerator for second-order reed-muller codes," *IEEE Transactions on Information Theory*, vol. 16, no. 6, pp. 745–751, Nov 1970.

[16] P. Delsarte and J. Goethals, "Alternating bilinear forms over $GF(q)$," *Journal of Combinatorial Theory*, vol. 19, pp. 26–50, 1975.

[17] A. M. Kerdock, "A class of low rate nonlinear binary codes," *Information and Control*, vol. 20, pp. 182–187, 1972.

[18] R. Berinde and P. Indyk, "Sparse recovery using sparse random matrices.," *Preprint*, 2008.

[19] D. Baron, M. F. Duarte, S. Sarvotham, M. B. Wakin, and R. G. Baraniuk, "An information-theoretic approach to distributed compressed sensing," in *Allerton Conf. Comm., Control, Comput*, 2005.

[20] D. L. Donoho and J. Tanner, "Neighborliness of randomly-projected simplices in high dimensions," *Proc. National Academy of Sciences*, vol. 102, no. 27, pp. 9452–9457, 2005.

[21] D. L. Donoho, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension," *Discrete & Computational Geometry*, vol. 35, no. 4, pp. 617–652, 2006.