IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

Expected RIP: Conditioning of The Modulated Wideband Converter

Moshe Mishali and Yonina C. Eldar

CCIT Report #736
July 2009

DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 32000, ISRAEL

CCIT REPORT #736 July 2009

Expected RIP: Conditioning of The Modulated Wideband Converter

Moshe Mishali and Yonina C. Eldar

Abstract

The sensing matrix of a compressive system impacts the stability of the associated sparse recovery problem. In this paper, we study the sensing matrix of the modulated wideband converter, a recently proposed system for sub-Nyquist sampling of analog sparse signals. Attempting to quantify the conditioning of the converter sensing matrix with existing approaches leads to unreasonable rate requirements, due to the relatively small size of this matrix. We propose a new conditioning criterion, named the expected restricted isometry property, and derive theoretical guarantees for the converter to satisfy this property. We then show that applying these conditions to popular binary sequences, such as maximal codes or Gold codes, leads to practical rate requirements.

I. Introduction

Signal dimensions in today's applications are growing faster than technology capabilities. The Nyquist rate of analog wideband signals, for example, already exceeds the conversion rate of existing devices. The modulated wideband converter (MWC) is a recent sub-Nyquist sampling system which exploits frequency sparsity to reduce the conversion rate [1]. Figure 1 depicts a block diagram of the converter, which is further described in Section II. The key idea underlying the MWC is that if the signal is periodically-modulated prior to sampling, then the sampling rate can be substantially reduced with respect to the Nyquist rate. The MWC consists of simple mixers and lowpass filters which are easy to implement.

This work was supported in part by the Israel Science Foundation under Grant no. 1081/07 and by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM++ (contract no. 216715).

The authors are with the Technion — Israel Institute of Technology, Haifa Israel. Email: moshiko@tx.technion.ac.il, yonina@ee.technion.ac.il.

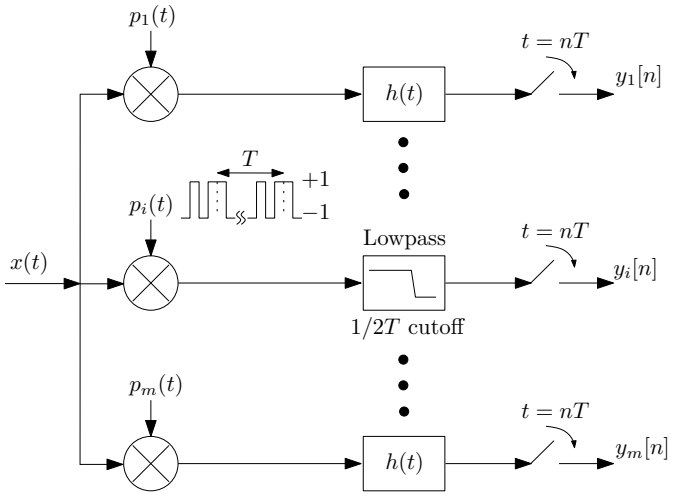


Fig. 1: The modulated wideband converter consists of m parallel channels, which mix the input by periodic waveforms. The mixed signal is then lowpass filtered and sampled at a low rate.

Reconstruction of the analog input from the MWC samples is nonlinear when the frequency support of the signal is unknown; a concrete recovery algorithm is detailed in [1].

In this paper, we study the conditioning of the MWC sampling operator, or equivalently the ability to recover the input in a stable manner. Mathematically, the main nonlinear step boils down to solving for the sparsest solution of a linear underdetermined linear system – a well-studied subject within the compressed sensing (CS) literature. The stability of sparse recovery is dictated by the conditioning of the sensing matrix, which in the MWC configuration depends on the parameters of the system, such as the number of channels and the choice of mixing waveforms. In Section III, we apply known CS results in order to quantify the required number of sampling channels that ensure stability. Unfortunately, these results lead to an unreasonable system size. The relatively small sensing matrix of the MWC is responsible for this behavior, since theoretical constants that are often ignored in large-scale problems have noneligible contribution otherwise.

Going from theory to practice, in Section IV, we aim at practical conditioning guarantees for the MWC, namely for a small number of channels, as the empirical evidence in [1] demonstrates. To achieve this goal, we first introduce a new stability criterion, termed the expected restricted isometry property (ExRIP), which quantifies the stability of a given (deterministic) sensing operator when applied to random sparse signals. The ExRIP extends on two related definitions: the RIP [2] which assumes no randomness and in general cannot be computed in polynomial-time, and the statistical-RIP [3]. The latter uses a partial random model in which the sensing matrix is deterministic while the nonzero locations are random, and

it is limited to matrices with stringent structure, which the MWC, for example, does not posses. The ExRIP relaxes the requirements on the sensing operator by considering a fully random signal model. Our main contribution is in proving that the MWC has the ExRIP when wisely selecting periodic mixing waveforms. Specifically, we show that popular binary sequences, such as maximal codes or Gold codes [4], are adequate candidates that yield reasonable requirements on the system size.

A short version of this technical report appears in [5].

II. The Modulated Wideband Converter

In this section, we begin by describing the sensing mechanism of the MWC. We then formulate our recovery problem and discuss the role of conditioning.

A. Sensing

The MWC consists of an analog front-end with m channels. In the ith channel, the input signal $x(t)$ is multiplied by a periodic waveform $p_i(t)$, lowpass filtered, and then sampled at rate $1/T$. In this paper, we study a simplified version of the converter, as depicted in Fig. 1, in which the sampling interval T equals the period of the waveforms $p_i(t)$. In addition, $p_i(t)$ are chosen as sign alternation waveforms, such that each period T consists of M intervals of length T/M each, and $p_i(t) = \pm 1$ on each such interval. This basic configuration is sufficient for studying the fundamental theoretical trade-off between rate and stability; other configurations with practical advantages are detailed in [1].

The MWC was studied in [1] mainly for multiband analog signals. The support of a multiband signal $x(t)$ resides within N frequency intervals, or bands, such that the width of each band does not exceed B Hz. The band positions are arbitrary and in particular unknown in advance. For example, in communication N represents the number of concurrent transmissions and B is specified by the transmission techniques in use. We note that sub-Nyquist sampling is one of the appealing properties of the MWC, though it can also be used for conventional Nyquist sampling with the proper number of channels.

The MWC sensing relies on the following key observation. The mixing operation scrambles the spectrum of $x(t)$ such that the baseband frequencies that reside below the filter cutoff $1/2T$, contain a mixture of the spectral contents from the entire Nyquist range. The periodicity of each waveform $p_i(t)$ ensures that the mixture has a specific nature – aliases at $1/T$ frequency spacing. Whilst aliasing is often considered as an undesired effect, here it is deliberately utilized to shift various frequency regions to baseband, simultaneously. In the basic configuration, we choose the rate $1/T = B$ and the length M of the sign patterns $p_i(t)$ is set to the compression ratio, namely the integer that is closest to the quotient

of the Nyquist rate by $1/T$. Intuitively, this process is invertible if the total sampling rate is proportional to the information rate NB.

B. Reconstruction

The recovery of $x(t)$ from the digital sequences $y_1[n], \ldots, y_m[n]$ consists of two steps which both exploit the sparse nature of the multiband spectrum. First, the spectral support is determined, and then the signal is recovered from the samples by a closed-form expression. The support recovery involves a series of digital computations, which are grouped together under the Continuous-to-Finite (CTF) block [1], [6]. As the name hints, the CTF allows to treat the resulting continuous recovery problem efficiently, by inferring the support from a small-size finite program. In the noiseless scenario, once the support is found by the CTF block, the input signal $x(t)$ is perfectly recovered. When noise is present, it may impact both the digital support recovery and the actual continuous reconstruction. We refer the reader to [1], [6] for a detailed discussion on the recovery process.

One of the elements of the CTF, which is our main focus here, is solving an underdetermined linear system for the sparsest solution matrix; also known as multiple measurement vectors (MMV) problem in the CS literature. The CTF block generates the MMV system

$$\mathbf{V} = \mathbf{A}\mathbf{U}, \tag{1}$$

where $\mathbf{V}$ is an $m \times r$ matrix that is computed from the given sequences $y_1[n], \ldots, y_m[n]$, with $r > 0$ some positive integer. The goal is to find an $M \times r$ matrix $\mathbf{U}$ with as few nonzero rows as possible. The nonzero rows in the sparsest $\mathbf{U}$ indicate the unknown support of $x(t)$ [1]. The $m \times M$ matrix $\mathbf{A}$ represents the sensing operator

$$\mathbf{A} = \mathbf{S}\mathbf{F}\mathbf{D}, \tag{2}$$

where $\mathbf{S}$ is an $m \times M$ matrix, whose ith row contains the sign pattern of the ith waveform $p_i(t)$. In the basic configuration $\mathbf{F}$ is the M-square DFT matrix (up to a column permutation). The matrix $\mathbf{D}$ is diagonal and accounts for the decay of the Fourier transform of $p_i(t)$ at high frequencies. The decay has to be compensated by analog means but this subject is beyond the current scope. For our purposes $\mathbf{D}$ can be ignored, since the nonzero location set $\mathrm{supp}(\mathbf{u}) = \mathrm{supp}(\mathbf{D}\mathbf{u})$ for any vector $\mathbf{u}$. Therefore, from this point on we focus on the sensing part $\mathbf{SF}$ of the matrix $\mathbf{A}$.

A large body of CS literature studies sparse recovery problems, such as (2), with either $r \geq 1$. It is well-known that finding the sparsest $\mathbf{U}$ is NP-hard in general. Fortunately, there are many sub-optimal polynomial-time algorithms that yield the exact solution under different conditions on the sensing matrix.

Typical recovery conditions of CS algorithms require the number m of rows to be proportional to the cardinality K of the nonzeros support. A logarithmic dependency on the number M of columns is also necessary for stable recovery [7]. In our setting, these conditions translate to a requirement on the number of parallel channels in the MWC. Obviously, we would like to have theoretical recovery guarantees with a reasonable number of channels, since those are implemented in hardware.

The simulations in [1] show 97% accurate support recovery rate on extensive sets of band locations. In the simulations, 3 concurrent transmissions, each of width $B = 50$ MHz, were generated with additive noise, where a Nyquist rate of 10 GHz defined the wideband input range. The MWC system used $m = 40$ channels with $M = 195$ length sign-patterns and rate $1/T \approx B$, which implies 80% rate saving with respect to the Nyquist rate. In this setting, there are 12 nonzero rows at most in $\mathbf{U}$ but due to the conjugate symmetry of the Fourier transform, it amounts to sparse recovery with $K = 6$ nonzeros. These numbers will serve us as a gold standard. As mentioned earlier, in practice the number of channels m can be substantially reduced when using other configurations of the MWC.

In the next section, we study existing conditions from the CS literature and show that they require a prohibitively large number of channels. Then, in Section IV we show how a wise selection of the sign patterns in $p_i(t)$ leads to conditioning guarantees with a small number of channels.

III. Applying Known Recovery Guarantees

In what follows, $\mathbf{\Phi}$ denotes an arbitrary matrix of size $m \times M$ with $m < M$, and $\mathbf{u}$ is an unknown K-sparse vector, with no more than K nonzeros. The goal is to ensure that the recovery of $\mathbf{u}$ from the underdetermined measurement $\mathbf{v} = \mathbf{\Phi u}$ is well-conditioned. We study the scenario (1) with $r = 1$ but comment that for $r > 1$, which is the typical MMV dimensions that the CTF generates, slightly better results can be obtained. In addition, we relate all conditions to the convex recovery method

$$\min_{\mathbf{u}} \|\mathbf{u}\|_1 \text{ s.t. } \|\mathbf{v} - \mathbf{A}\mathbf{u}\|_2^2 \leq \epsilon. \tag{3}$$

Program (3) is known as basis pursuit (BP) for $\epsilon = 0$. BP denoising refers to the case $\epsilon > 0$. The quadratic constraint is sometimes regularized and merged into the objective.

A. Coherence-based Guarantees

The coherence of a matrix $\mathbf{\Phi}$ is defined as

$$\mu = \max_{i \neq j} \frac{|\langle \mathbf{\Phi}_i, \mathbf{\Phi}_j \rangle|}{\|\mathbf{\Phi}_i\| \, \|\mathbf{\Phi}_j\|}, \tag{4}$$

where the subscript $\mathbf{\Phi}_i$ denotes the ith column. The coherence μ can be efficiently computed for any given matrix. A well known CS result [8, Th. 7] is that if

$$K \leq \frac{1}{2}\left(1+\frac{1}{\mu}\right), \tag{5}$$

then any K-sparse vector $\mathbf{u}$ is perfectly recovered by BP.

Another result by Tropp [9, App. IV-A] shows that if $K \leq 1/3\mu$, and the measurement $\mathbf{\Phi u}$ is contaminated by Gaussian white noise with covariance $\sigma^2\mathbf{I}$, then the support of $\mathbf{u}$ can be recovered with high probability using the BP denoising program.

Candès and Plan [10, Th. 1.2] considered the noisy model in which the support $\mathrm{supp}(\mathbf{u})$ is drawn uniformly at random among all possible choices, and the nonzero values have amplitudes $|u_i| > (6+\sqrt{2})\sigma\sqrt{2\log M}$ and random signs. Under the conditions $\mu < c/\log M$ and $K \leq cM/\|\mathbf{\Phi}\|^2 \log M$, the BP denoising is proved to recover the support with high probability.

In Table I, we numerically evaluate the recovery guarantees based on these bounds. The "best-case" column indicates the smallest dimensions m, M of $\mathbf{S}$, such that the theoretical requirements are satisfied for the given K. For example, for $K = 2$ the bound (5) is satisfied with $\mathbf{S}$ of dimensions 300×2000 at least. We selected comparable dimensions for $\mathbf{S}$ and slightly adjusted K where required. The best matrix $\mathbf{S}$, in the sense of lowest $\mu(\mathbf{SF})$, was chosen out of 100 realizations with ± 1 entries that were drawn independently with equal probability. We could not verify the exact value of the constant c from the statement or the proof of [10]. For coherence results the probability $p = 1$ in the table means no randomness in both $\mathbf{\Phi}$ and $\mathbf{u}$. The MWC column assumes $M = 195, K = 12$ and the value of m that is required to satisfy the relevant bounds is displayed. As before, the instance $\mathbf{S}$ which gives the lowest value for m is used.

As the table shows, in small-scale problems the coherence-based guarantees require an unreasonable number of channels, leading to a sampling rate that exceeds Nyquist by orders of magnitude. Evidently, the constants can be very significant.

We next examine conditioning guarantees that measure the correlation between large subsets of columns, rather than pairs as in (4). Obviously, these conditions can better predict the conditioning of $\mathbf{\Phi}$ on sparse vectors, but as it turns out they are harder to compute for a given matrix.

B. RIP Guarantees

Candès et. al. [2] introduced the restricted isometry property (RIP) as a standard tool for analyzing sensing matrices. A matrix $\mathbf{\Phi}$ is said to have the RIP with isometry constant δ_K, if $0 \leq \delta_K < 1$ is the

smallest number such that

$$(1-\delta_K)\|\mathbf{u}\|^2 \leq \|\mathbf{A}\mathbf{u}\|^2 \leq (1+\delta_K)\|\mathbf{u}\|^2 \tag{6}$$

holds for *all* K-sparse vectors $\mathbf{u}$ [2]. The RIP quantifies how well $\mathbf{\Phi}$ preserves the norm of sparse vectors. If $\delta_{2K} < 1$ then every K-sparse vector $\mathbf{u}$ is uniquely determined by $\mathbf{\Phi}\mathbf{u}$. Furthermore, if $\delta_{2K} < \sqrt{2}-1$ then BP recovers $\mathbf{u}$ exactly [11]. In the presence of noise, the same condition ensures that BP denoising recovers the signal up to a small bounded error [11].

The main drawback of RIP conditions is that the isometry constant δ_{2K} cannot be computed in polynomial time for an arbitrary deterministic matrix. The common workaround is to consider random matrix ensembles, e.g. when the entries of $\mathbf{\Phi}$ are drawn from the Gaussian, or the Bernoulli distributions. A relevant result appears in [7] based on [12]:

Theorem 1: Let $\mathbf{\Phi}$ be an $m \times M$ matrix generated by drawing entries from an appropriately scaled sub-Gaussian distribution. If

$$m \geq \frac{2}{c\delta_K}\left(\ln(2L) + K\ln\left(\frac{12}{\delta_K}\right) + t\right), \tag{7}$$

where $L = \binom{M}{K}$ and c is a distribution-dependent constant, then, with probability at least $p = 1 - e^{-t}$, $\mathbf{\Phi}$ has the RIP.

In our setting, if $\mathbf{S}$ is random with equali-likely ± 1 entries, and (7) is satisfied, then it has the RIP. The constant $c = 7/18$ in this setting [12]. Since $\mathbf{F}$ is a unitary matrix, the compounded sensing matrix $\mathbf{SF}$ has the same RIP constant as $\mathbf{S}$ [12]. In Table I we calculate the typical setting in which the bound (7) is satisfied for $\delta_{2K} = \sqrt{2}-1$ and $p = 0.97$. As before, the theoretical requirements seem pessimistic.

Besides the requirement for a large number of sampling channels, there is a delicate logical issue in the above inference. Theorem 1 predicts the RIP property of a random matrix. In practice, the entries of $\mathbf{S}$ are fixed to the specific sign patterns that are realized in the system. Notice also the different meaning of the probability; in Theorem 1 it refers to instances of $\mathbf{S}$, while in empirical simulations the recovery rate refers to signal realizations with fixed $\mathbf{S}$.

A recent approach in CS considers deterministic matrices by switching the role of randomness from the sensing operator to the signal model. This framework conforms with the conventional Bayesian approach in signal processing, and naturally fits simulation methodology. In the next section we quote results of this kind.

C. Statistical-RIP Guarantees

Calderbank et. al. [3] proposed the Statistical RIP (StRIP) as an alternative tool for quantifying sensing matrices. A matrix $\mathbf{\Phi}$ has the StRIP(K, δ_K, p) if (6) is satisfied with probability at least p for a K-sparse vectors $\mathbf{u}$, whose support is uniformly drawn from all possible choices (the nonzero values are arbitrary). Calculating the StRIP for an arbitrary matrix $\mathbf{\Phi}$ is not easier than RIP computations. However, structured matrices can greatly simplify the calculations. In [3] the authors consider matrices, whose columns form a closed-group under element-wise multiplication. In addition, it is assumed that the rows of $\mathbf{\Phi}$ are orthogonal and each sums to zero. Under these hypothesis they prove that the StRIP is satisfied with

$$\frac{K-1}{M-1} < \delta_K < 1 \qquad\qquad p \geq 1 - \frac{\frac{2K}{m} + \frac{2K+7}{M-3}}{\left(\delta_K - \frac{K-1}{M-1}\right)^2}. \tag{8}$$

This result is however not useful for the MWC, since the columns of $\mathbf{SF}$ do not form the required group, and this property is essential for the arguments in [3]. Even when ignoring this issue, the required dimensions m, M are high. In the MWC setting $m = 150$ channels give probability $p = 0$ in (8).

More recently, Gan et. al. [13] studied StRIP guarantees for $\mathbf{\Phi}$ with unit-norm columns and zero-sum rows. They showed that the StRIP is satisfied with

$$\frac{1}{M-1} < \delta_K \qquad\qquad p \geq 1 - 2\exp\left(-\frac{\left(\delta_K - \frac{1}{M-1}\right)^2}{16\mu^2 K}\right), \tag{9}$$

where as before μ is the coherence of $\mathbf{\Phi}$. See Table I for evaluation of this result.

Tropp [14, Th. 12] also provides guarantees on the conditioning of a deterministic matrix. We bring this result using StRIP terminology. A matrix $\mathbf{\Phi}$ with unit-norm columns satisfies the StRIP with probability at least $p = 1 - (K/2)^{-t}$, for some $t \geq 1$, if

$$\sqrt{144\mu^2 K t \log(K/2+1)} + \frac{2K}{M}\|\mathbf{\Phi}\|^2 \leq e^{-1/4}\delta_K. \tag{10}$$

Additional conditions are used to bound the probability that BP recovers $\mathbf{u}$ exactly [14, Th. 14]. Observe the table for the applicability of this bound. Note that both [13], [14] require $\mathbf{\Phi}$ to have unit-norm columns. Therefore we applied them on $\mathbf{\Phi} = \mathbf{SF}/\sqrt{mM}$ which approximately satisfies this requirement.

The attempt to derive practical recovery guarantees for the MWC, which are based on existing results for general structured $\mathbf{\Phi}$, is perhaps sentenced to fail; It does not take into account the specific structure of the sensing matrix $\mathbf{SF}$. The next section capitalizes on this structure in order to reduce the requirements on the number of sampling channels.

TABLE I: Recovery guarantees for the MWC

	Conditioning measure	"Best-case" setting m	M	K	p	MWC Setting (see text)	Support	Values	Noise	Remarks/Issues
Coherence	Donoho-Elad [8]	300	2000	2	1	$m \geq 4230$	D	D	-	
	Tropp [9]	300	2000	1	1	$m \geq 9540$	D	D	+	
	Candès-Plan [10]	n/a				n/a	R	Random signs	+	unspecified constant
RIP	Blumensath+ [7], [12]	700	5000	3	0.95	$m \geq 950$	D	D	-	random $\mathbf{S}$
StRIP	Calderbank+ [3]	700	5000	3	0.93	$m = 150$, $p = 0$	R	D	-	columns form a group
	Gan+ [13]	500	3000	1	0	n/a	R	D	-	unit-norm columns
	Tropp [14]	1000	2000	3	0.01	n/a	R	D	-	unit-norm columns
ExRIP	This paper	80	511	12	0.94	$m \geq 40$	R	R	-	$\mathbf{\Phi} = \mathbf{SF}$, theory
	Simulations [1]	40	195	12	0.96	$m \geq 40$	R	R	+	$\mathbf{\Phi} = \mathbf{SF}$, simulations

D=deterministic, R=random, +/-=with/without noise, n/a=not applicable

IV. Conditioning of the MWC

A. The ExRIP

We begin with defining a StRIP-like property, which accounts for randomness in the nonzero values.

Definition 1: A matrix $\mathbf{\Phi}$ has the expected restricted isometry property (ExRIP), if (6) holds with probability at least p for K-sparse random vectors $\mathbf{u}$ whose support is uniformly distributed and whose nonzeros are i.i.d random variables.

The ExRIP involves several constants: the sparsity level K, the isometry constant δ_K, the probability p, and finally the distribution from which the nonzeros are drawn. Both ExRIP and StRIP assume random support. However, the ExRIP adds another layer of randomness in the nonzero values. On the one hand, the ExRIP is mathematically weaker since worst-case signal values that are unlikely to encounter are averaged out. On the other hand, since the random support assumption anyway excludes worst-case scenarios it makes sense to consider random values as well. Moreover, random signal values conform with the conventional Bayesian framework. Another advantage of the ExRIP is in simplifying complicated expressions that otherwise require $\mathbf{\Phi}$ to have a stringent structure as in [3]. For instance, compare between the StRIP proof [3] and the one provided in the sequel for the ExRIP. Besides these reasons, applying the StRIP results in high system dimensions, whereas as Table I shows the ExRIP is the only measure that leads to a reasonable number of channels.

Our goal is to prove that the sensing matrix $\mathbf{SF}$ has the ExRIP with high probability, for the MWC dimensions. To achieve this goal, we characterize the quality of a given set of sign patterns as follows.

The correlation of the rows is captured by

$$\boldsymbol{\alpha}(\mathbf{S}) = \frac{1}{(mM)^2} \sum_{i,k=1}^{m} (\mathbf{S}_i^T \mathbf{S}_k)^2, \tag{11}$$

where the subscripts indicate the relevant rows of $\mathbf{S}$. Note that $\boldsymbol{\alpha}(\mathbf{S})$ resembles the coherence μ with two distinguishing properties: the coherence is computed over the columns of $\mathbf{SF}$, while $\boldsymbol{\alpha}(\mathbf{S})$ involves the rows of $\mathbf{S}$ only. In addition, μ involves maximization while $\boldsymbol{\alpha}(\mathbf{S})$ computes sums of squares. Another quality of interest is the total power of all auto- and cross-correlation functions, as measured by

$$\boldsymbol{\beta}(\mathbf{S}) = \frac{1}{m^2 M^3} \sum_{i,k=1}^{m} \|\mathbf{S}_i \odot \mathbf{S}_k\|^2. \tag{12}$$

Here $\mathbf{S}_i \odot \mathbf{S}_k$ stands for cyclic convolution. We will also need

$$\boldsymbol{\gamma}(\mathbf{S}) = \frac{1}{(mM)^2} \sum_{i,k=1}^{m} (\mathbf{S}_i^T \mathbf{S}_k^-)^2, \tag{13}$$

where for any vector $\mathbf{a}^-[n] = \mathbf{a}[-n], n = 0, \ldots, M-1$ under the convention that the indices are module M.

The quality measures are bounded below and above by:

$$\frac{1}{m} \leq \boldsymbol{\alpha}(\mathbf{S}) \leq 1 \qquad \frac{1}{M} \approx \frac{2m-1}{2mM-1} \leq \boldsymbol{\beta}(\mathbf{S}), \boldsymbol{\gamma}(\mathbf{S}) \leq 1. \tag{14}$$

To prove the bounds, recall that the entries of $\mathbf{S}$ are binary. Noting the sum of squares in (11) and the fact that $\|\mathbf{S}_i\|^2 = M$ give the lower bound on $\boldsymbol{\alpha}(\mathbf{S})$. Orthogonal rows achieve this bound. To prove (14) for $\boldsymbol{\beta}(\mathbf{S}), \boldsymbol{\gamma}(\mathbf{S})$, we recall the well-known result by Welch [15]. For any set of l binary sequences, of length L, it holds that the cross-correlation function

$$R_c[k] = \sum_{\tau=0}^{L-1} \mathbf{a}[\tau]\mathbf{b}[\tau+k] \geq L\sqrt{\frac{l-1}{lL-1}} \approx \sqrt{L}, \tag{15}$$

for every pair of sequences $\mathbf{a}, \mathbf{b}$ from the set, and indices modulo L. In (13), the sum involves the correlations between the rows $\mathbf{S}_i$ and their flipped versions $\mathbf{S}_i^-$. Considering the set of $2m$ sequences $\{\mathbf{S}_0, \ldots, \mathbf{S}_{m-1}, \mathbf{S}_0^-, \ldots, S_{m-1}^-\}$ proves the lower bound on $\boldsymbol{\gamma}(\mathbf{S})$. Similar claims are used to show the left-inequality for $\boldsymbol{\beta}(\mathbf{S})$. On the other extreme, identical rows give $\boldsymbol{\alpha}(\mathbf{S}) = 1$ and if in addition all the entries of $\mathbf{S}$ are equal then also $\boldsymbol{\beta}(\mathbf{S}) = \boldsymbol{\gamma}(\mathbf{S}) = 1$. Clearly, identical sign patterns are not adequate for the MWC since the channels produce the same measurements. Intuitively, the mixing patterns $p_i(t)$ should differ from each other as much as possible in order to provide additional information on the signal. For these reasons, the best quality is attained when $\boldsymbol{\alpha}(\mathbf{S}), \boldsymbol{\beta}(\mathbf{S}), \boldsymbol{\gamma}(\mathbf{S})$ are low. Notice again the analogous requirement for low coherence in standard CS.

The following theorem is our main contribution. It relates the quantities $\boldsymbol{\alpha}(\mathbf{S}), \boldsymbol{\beta}(\mathbf{S}), \boldsymbol{\gamma}(\mathbf{S})$ to the probability of satisfying the ExRIP.

Theorem 2: Let $\mathbf{\Phi} = \mathbf{SF}/\sqrt{mM}$ be the scaled sensing matrix of the MWC. If the nonzeros are drawn from a symmetric distribution, then $\mathbf{\Phi}$ has the ExRIP with probability at least

$$p = 1 - \frac{(1 - C_K)\rho_M \left(1 + \boldsymbol{\alpha}(S) - 2\boldsymbol{\beta}(S)\right)}{\delta_K^2} - \frac{(B_K - C_K)\rho_M \left(\boldsymbol{\gamma}(S) - \boldsymbol{\beta}(S)\right) + C_K M \boldsymbol{\beta}(S) - 1}{\delta_K^2} \tag{16}$$

where

$$C_K = \mathbb{E}\left\{\frac{\sum_{i=1}^K |u_i|^4}{\|\mathbf{u}\|^4}\right\}, \qquad B_K = \mathbb{E}\left\{\frac{|\sum_{i=1}^K u_i^2|^2}{\|\mathbf{u}\|^4}\right\} \tag{17}$$

are distribution-dependent constants, u_i are the K random variables representing the nonzeros of $\mathbf{u}$, and $\rho_M = M/(M-1)$.

The symmetric distribution in the theorem refers to symmetry of the probability density function around the origin. For example, Gaussian, uniform and Bernoulli distributions satisfy this condition. Note that $B_K = 1$ whenever the nonzeros are real-valued. It can also be verified that $C_K = 3K/(2K + K^2)$ when u_i are standard normal variables. Explicit formulas for other distributions exist, but we find it more convenient to sample the distribution and approximate B_K, C_K by averaging. We note that the bound (16) may be negative when the ExRIP is not satisfied.

Theorem 2 is proved by calculating the second and the forth moments of $Z = \|\mathbf{\Phi u}\|/\|\mathbf{u}\|$, a somewhat tedious procedure. Point out that the $\sqrt{mM}$ scaling in Theorem 2 stems from the default scaling $1 \pm \delta_K$ in (6). In practice, the recovery conditioning depends on the ratios between the largest singular value to the smallest one of all column subsets of $\mathbf{\Phi}$. These ratios are insensitive to constant scaling. The proof of the theorem appears in the Appendix.

To realize the advantage of Theorem 2, we now evaluate the ExRIP for several choices of sign patterns. Not surprisingly, popular binary sequences that are used in communication channels appear adequate due to their low correlation-energy.

B. Choosing the Sign Patterns

The literature describes several families of binary sequences with proven correlation guarantees. The Maximal, Gold [4], Kasami [16], and Hadamard codes are families of sequences with different bounds on the self and the mutual correlations. The first three families require $M = 2^n - 1$ with different limitations

on the possible order n and on the number of sequences within the code. The M-square Hadamard matrix provides up to $M = 2^n$ sequences.

In Table II we calculate the probability p of satisfying the ExRIP for these codes using (16). The binary families were generated according to the description of [17]. For the maximal sequences (a.k.a. m-sequences), we used the order $n = 9$ so as to have comparable dimensions to the Gold codes, which do not exist for $n = 4k$ for any integer k. The first eight characteristic polynomials for the order $n = 9$ generate 8 m-sequences of length $M = 2^n - 1 = 511$. To generate more m-sequences, we repeatedly cyclic shifted each of these sequences by 10 positions (to the right), then by 20 positions, by 30 positions and so on, until we collected the required $m = 80$ sequences. A Gold code is defined by a preferred-pair of two m-sequences and their shifts. We used the characteristic polynomial $f(D) = D^4 + D^9$ to generate a single m-sequence, which was then decimated by a factor of 3. It can be verified that this is a preferred-pair [17]. The Gold code in this example contains 513 sequences, of which we used the first $m = 80$. The small set of the Kasami family is also generated by an m-sequence pair. Here the order is $n = 8$ since this family requires even n. The characteristic polynomial $f(D) = D + D^6 + D^7 + D^8$ was used to generate a single m-sequence, which was then decimated by a factor of 17. This pair and the procedure from [17] gave the 16 possible Kasami sequences. Finally, for Hadamard we took the first 80 rows of the canonical Hadamard matrix. In the table we assume $\delta_{2K} = \sqrt{2} - 1$ to evaluate the bound (16).

Empirically, we noticed that random instances of $\mathbf{S}$ perform well as demonstrated in the table. This is reasoned by the cumulative nature of the measures $\boldsymbol{\alpha}(\mathbf{S}), \boldsymbol{\beta}(\mathbf{S}), \boldsymbol{\gamma}(\mathbf{S})$, which forgive local peaks and averages the total power of the correlation functions. Combining the fact that random instances are not limited to pre-specified lengths together with the ability to compute $\boldsymbol{\alpha}(\mathbf{S}), \boldsymbol{\beta}(\mathbf{S}), \boldsymbol{\gamma}(\mathbf{S})$ efficiently allows for a wise selection of the patterns. The probabilities in Table II are computed for complex variables u_i, for which both the real and the imaginary parts are either normal or uniform on $[-0.5, 0.5]$. Complex-valued MMVs result naturally in the MWC system [1].

Maximal, Gold and even random sequences all seem adequate. In contrast, Hadamard sequences yield a poor probability estimate since their cross-correlation functions have many peaks, as implied by $\boldsymbol{\beta}(\mathbf{S}), \boldsymbol{\gamma}(\mathbf{S})$, which are one order of magnitude above the other codes. The small set of Kasami sequences, which offers only 16 patterns, is considered optimal in communication, since their auto- and cross-correlations achieve the Welch bound [15]. In our quality measures Kasami sequences achieve poor scores since the aggregated energy is high due to many peaks (at the Welch bound level). In contrast, maximal and Gold sequences have a few high local peaks but their average energy is low.

TABLE II: ExRIP guarantees for different sign patterns

	Dimensions			Quality $\times 100$			ExRIP prob. p	
Family	m	M	$2K$	$\boldsymbol{\alpha}(\mathbf{S})$	$\boldsymbol{\beta}(\mathbf{S})$	$\boldsymbol{\gamma}(\mathbf{S})$	Normal	Uniform
Maximal	80	511	24	1.438	0.196	0.408	0.932	0.931
Gold	80	511	24	1.255	0.198	0.199	0.939	0.939
Hadamard	80	512	24	1.250	1.094	1.238	0.000	0.000
Random1	80	511	24	1.439	0.198	0.202	0.927	0.927
Kasami	16	255	12	6.667	0.392	0.294	0.689	0.675
Random2	40	195	24	3.025	0.526	0.537	0.856	0.858

Based on the table, we suggest a simple approximation to (16). First, we note that $B_K \approx C_K$ for complex-valued nonzeros; this can be observed numerically for a wide range of K. Therefore, we ignore the term containing $(B_K - C_K)$. In addition, note that $\boldsymbol{\gamma}(\mathbf{S}) \approx \boldsymbol{\beta}(\mathbf{S})$ for the binary families of interest, as Table II shows. This is another reason to ignore this term in (16). For the remaining terms, the probability p depends linearly on C_K with scaling

$$\frac{\partial p}{\partial C_K} = \frac{1}{\delta_K^2}\Big(\rho_M(1 + \boldsymbol{\alpha}(\mathbf{S}) - 2\boldsymbol{\beta}(\mathbf{S})) - M\boldsymbol{\beta}(\mathbf{S})\Big). \tag{18}$$

For the Maximal and Gold codes, the scaling (18) is strictly positive implying p is monotone increasing in C_K. Substituting $C_K = 0$ in (16) results in

$$p \geq 1 - \frac{\rho_M(1 + \boldsymbol{\alpha}(\mathbf{S}) - 2\boldsymbol{\beta}(\mathbf{S})) - 1}{\delta_K^2} \approx 1 - \frac{\rho_M(1 + \boldsymbol{\alpha}(\mathbf{S})) - 1}{\delta_K^2} \approx 1 - \frac{\boldsymbol{\alpha}(\mathbf{S})}{\delta_K^2}, \tag{19}$$

since $2\boldsymbol{\beta}(S)$ is one order below $1 + \boldsymbol{\alpha}(\mathbf{S})$ and $\rho_M = M/(M-1) \approx 1$. Next, we decompose

$$\boldsymbol{\alpha}(\mathbf{S}) = \frac{1}{(mM)^2} \underbrace{\sum_i (\mathbf{S}_i^T \mathbf{S}_i)^2}_{mM^2} + \underbrace{\sum_{i \neq k} (\mathbf{S}_i^T \mathbf{S}_k)^2}_{\leq (m^2 - m)\varepsilon}. \tag{20}$$

The first term dominates since ε, which upper bounds the inner product between different rows, is typically negligible. Gold and Kasami sequences have explicit guarantees on ε. For example, the Gold family of Table II has $\varepsilon \leq 33$, which renders the first term in (20) at least two orders of magnitude higher than the second term. Consequently, (16) reduces to the shorter form

$$p \approx 1 - \frac{1}{m\delta_K^2}. \tag{21}$$

Fig. 2 plots the probability of satisfying ExRIP, as predicted by Theorem 2, for increasing number of sampling channels and (complex-valued) normally distributed nonzeros. We added the approximated bound (21) for comparison. Evidently, the conditioning of the MWC has two dominant factors: the number of sampling channels m and the required conditioning level δ_K.

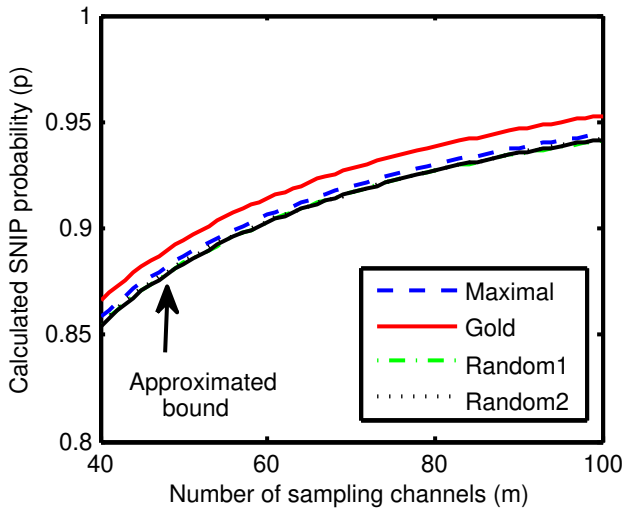


Fig. 2: The probability of satisfying the ExRIP depends mainly on the number of sampling channels. The black solid curve shows that the approximated bound (21) coincides with the theoretical curve of the *Random2* sequence family, whose dimensions match the MWC setting of [1].

To further support the approximated bound (21), we conducted the following experiment to estimate the value of the ExRIP probability p, statistically. For m varying from 10 to 50, and M varying from 120 to 195, we constructed the matrix $\mathbf{S}$ from the first m rows and the first M columns of the *Random2* code of Table II. Then, for each K between 4 to 15, we generated 5000 instances of random K-sparse vector $\mathbf{u}$. The support was chosen uniformly at random, and for the nonzero values we used the normal distribution for both real and imaginary parts. For each trial, the condition

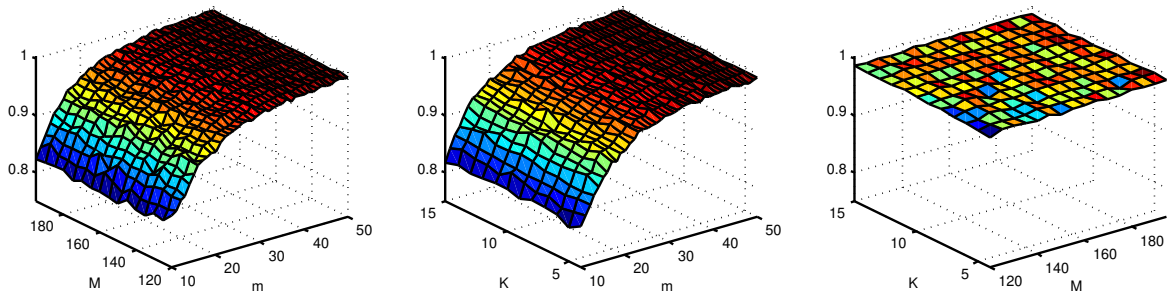
$$\left| \frac{\|\mathbf{SFu}\|^2}{mM\|\mathbf{u}\|^2} - 1 \right| \leq \delta_{2K} = \sqrt{2} - 1$$

was checked. Fig. 3 reports the ratio of trials that satisfied the condition, for the various dimensions m, M, K. Observe that for a fixed m, the ExRIP probability is practically constant, whereas when m varies, the tendency (21) is apparent.

To conclude we comment that Theorem 2 proves that the MWC has the ExRIP. Relating the conditioning probability to success recovery by basis pursuit requires additional steps as carried out in [14] or as proposed in [3].

REFERENCES

[1] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *arXiv.org 0902.4291*, 2009.



(a) $K = 15$ (b) $M = 175$ (c) $m = 39$

Fig. 3: Statistical evaluation of the ExRIP probability p for various dimensions.

[2] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[3] R. Calderbank, S. Howard, and S. Jafarpour, "Deterministic compressive sensing with groups of random variables," [Online]. Available: http://www.dsp.ece.rice.edu/files/cs/strip-more.pdf, 2009.

[4] R. Gold, "Maximal recursive sequences with 3-valued recursive cross-correlation functions," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 154–156, 1968.

[5] M. Mishali and Y. C. Eldar, "Conditioning of the modulated wideband converter," in *ITW'09 to appear.*, 2009.

[6] ——, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4692–4702, Oct. 2008.

[7] T. Blumensath and M. E. Davies, "Sampling theorems for signals from the union of finite-dimensional linear subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1872–1882, April 2009.

[8] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell 1$ minimization," *Proc. Natl. Acad. Sci.*, vol. 100, pp. 2197–2202, Mar. 2003.

[9] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.

[10] E. Candes and Y. Plan, "Near-ideal model selection by ℓ_1 minimization," *Ann. Statist.(to appear)*, 2008.

[11] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *C. R. Acad. Sci. Paris, Ser. I*, vol. 346, pp. 589–592, 2008.

[12] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Const. Approx.*, 2007.

[13] L. Gan, C. Ling, T. T. Do, and T. T. D., "Analysis of the Statistical Restricted Isometry Property for Deterministic Sensing Matrices Using Steins Method," [Online]. Available: http://www.dsp.ece.rice.edu/files/cs/Gan_StatRIP.pdf, 2009.

[14] J. A. Tropp, "On the conditioning of random subdictionaries," *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, pp. 1–24, 2008.

[15] L. Welch, "Lower bounds on the maximum cross correlation of signals (corresp.)," *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 397–399, May 1974.

[16] T. Kasami, "Weight distribution formula for some class of cyclic codes," *University of Illinois, Urbana, Rept. R-265, April*

1966.

[17] E. H. Dinan and B. Jabbari, "Spreading codes for direct sequence CDMA and wideband CDMA cellular networks," *IEEE Commun. Mag.*, vol. 36, no. 9, pp. 48–54, 1998.

APPENDIX

Proof of Theorem 2

Let $\mathbf{\Phi} = \mathbf{SF}/\sqrt{mM}$ be the scaled sensing matrix of the MWC and denote by $\mathbf{u}$ a random K-sparse vector with random support set $\Lambda = \mathrm{supp}(\mathbf{u}) = \{\Lambda_0, \ldots, \Lambda_{K-1}\}$, such that Λ_i indicates the location of the ith nonzero u_i. Define also the random variable $Z = \|\mathbf{\Phi u}\|/\|\mathbf{u}\|$. The proof consists of three steps:

1) Calculating $\mathbb{E}\{Z^2\}$;
2) Calculating $\mathbb{E}\{Z^4\}$; and
3) Applying the Chebyshev inequality to bound the probability $\mathrm{Prob}\{|Z^2 - 1| \leq \delta_K\}$ from below.

To simplify the presentation, we enumerate the columns of $\mathbf{\Phi}$ from 0 to $M-1$, and the rows from 0 to $m-1$. We shall also use the symbols $\imath = \sqrt{-1}$ and

$$\omega = e^{\frac{\imath 2\pi}{M}}.$$

The notation $\mathbb{E}_\Lambda\{\cdot\}$ means taking the expectation over the random support Λ, when fixing (or conditioning on) the nonzero values $u_0, \ldots, u_{K-1}$. The following conditional expectation will be used repeatedly in our derivations

$$\mathbb{E}_\Lambda\{\omega^{i\Lambda_j}\} = \frac{1}{M}\sum_{a=0}^{M-1}\omega^{ia} = \begin{cases} 1 & \omega^i = 1 \\ 0 & \text{otherwise} \end{cases}, \qquad 0 \leq j \leq K-1. \tag{22}$$

Defining the indicator function

$$\mathbf{I}_x = \begin{cases} 1 & \omega^x = 1 \\ 0 & \text{otherwise} \end{cases}, \tag{23}$$

we can also write (22) as

$$\mathbb{E}_\Lambda\{\omega^{i\Lambda_j}\} = \mathbf{I}_i, \qquad 0 \leq j \leq K-1. \tag{24}$$

In the sequel, we will sometimes use $\mathbf{I}(x)$ instead of $\mathbf{I}_x$.

The following lemma develops a recursive formula for a conditional expectation, similar to (22), when more than a single support index from Λ is involved. We provide the general result, though we will use the lemma only for sizes $t = 1, 2$.

Lemma 1: Suppose $K = |\Lambda|$ is the cardinality of the support set Λ, and let $\bar{\Lambda} = \{\bar{\Lambda}_1, \ldots, \bar{\Lambda}_K\}$ be a fixed permutation of Λ, such that there is a one-to-one mapping between $\Lambda, \bar{\Lambda}$. For every $1 \leq t \leq K$, and

$$f_t(a_1, \cdots, a_t) = \mathbb{E}\left\{\omega^{\sum_{i=1}^{t} a_i \bar{\Lambda}_i}\right\}, \tag{25}$$

we have that

$$f_{t+1}(a_1, \cdots, a_{t+1}) = \frac{M\mathbf{I}(a_{t+1}) f_t(a_1, \cdots, a_t) - \sum_{i=1}^{t} f_t(\cdots, a_i + a_t, \cdots)}{M - t}. \tag{26}$$

for every $1 \leq t \leq K - 1$.

Proof: Expanding the expression for the conditional expectation gives

$$f_{t+1}(a_1, \cdots, a_{t+1}) = \frac{1}{M(M-1)\cdots(M-t)} \sum_{\substack{b_1, \cdots, b_{t+1} = 0 \\ \text{pairwise different}}}^{M-1} \omega^{\sum_{i=1}^{t+1} a_i b_i} \tag{27}$$

$$= \frac{1}{M(M-1)\cdots(M-t)} \sum_{\substack{b_1, \cdots, b_t = 0 \\ \text{pairwise different}}}^{M-1} \omega^{\sum_{i=1}^{t} a_i b_i} \sum_{\substack{b_{t+1}=0 \\ \forall j \neq t+1,\, b_{t+1} \neq b_j}}^{M-1} \omega^{a_{t+1} b_{t+1}} \tag{28}$$

$$= \frac{1}{M(M-1)\cdots(M-t)} \sum_{\substack{b_1, \cdots, b_t = 0 \\ \text{pairwise different}}}^{M-1} \omega^{\sum_{i=1}^{t} a_i b_i} \left(M\mathbf{I}(a_{t+1}) - \sum_{l=1}^{t} \omega^{a_{t+1} b_l} \right). \tag{29}$$

Rearranging and using the definition of f_t leads to

$$f_{t+1}(a_1, \cdots, a_{t+1}) = \frac{M\mathbf{I}(a_{t+1}) f_t(a_1, \cdots, a_t)}{M - t} - \frac{\sum_{l=1}^{t} \sum_{\substack{b_1, \cdots, b_t = 0 \\ \text{pairwise different}}}^{M-1} \omega^{a_{t+1} b_l} \omega^{\sum_{i=1}^{t} a_i b_i}}{M(M-1)\cdots(M-t)}. \tag{30}$$

Invoking the definition of f_t again completes the proof. ■

From now on, we assume that the first column of $\mathbf{S}$ contains only $+1$ entries, since otherwise we can multiply $\mathbf{S}$ from the left by an appropriate signs diagonal matrix, without affecting Z.

Step 1 - Calculating the second moment of Z:

We begin by calculating the moment

$$\mathbb{E}_\Lambda\{\|\mathbf{SFu}\|^2\} = \mathbb{E}_\Lambda \left\{ \sum_{r=0}^{m-1} |\mathbf{S}_r^T \mathbf{Fu}|^2 \right\}, \tag{31}$$

$$= \sum_{r=0}^{m-1} \mathbb{E}_\Lambda\{|\mathbf{S}_r^T \mathbf{Fu}|^2\} \tag{32}$$

where $\mathbf{S}_r$ is the rth row of $\mathbf{S}$, written as a column vector. Consider the term corresponding to $r = 0$, and for brevity denote $\mathbf{s} = \mathbf{S}_0 = [s_0, \ldots, s_{M-1}]^T$ as the signs of the first row of $\mathbf{S}$. Taking the conditional expectation over Λ gives

$$\mathbb{E}_\Lambda\{|\mathbf{s}^T\mathbf{F}\mathbf{u}|^2\} = \mathbb{E}_\Lambda\left\{\left|\sum_{i=0}^{M-1}\sum_{l=0}^{K-1} s_i u_l \omega^{i\Lambda_l}\right|^2\right\} \tag{33}$$

$$= \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} s_i s_j \sum_{l=0}^{K-1}\sum_{m=0}^{K-1} u_l\overline{u_m}\mathbb{E}_\Lambda\left\{\omega^{i\Lambda_l - j\Lambda_m}\right\}, \tag{34}$$

where $\overline{u_m}$ is the conjugate of the mth nonzero value u_m. Whenever $l = m$, the expectation becomes

$$\mathbb{E}_\Lambda\left\{\omega^{(i-j)\Lambda_l}\right\} = \mathbf{I}_{i-j}. \tag{35}$$

For $l \neq m$, Lemma 1 implies that

$$\mathbb{E}_\Lambda\left\{\omega^{i\Lambda_l - j\Lambda_m}\right\} = \frac{M\mathbf{I}_i\mathbf{I}_j - \mathbf{I}_{i-j}}{M-1}. \tag{36}$$

Substituting back into (34) leads to

$$\mathbb{E}_\Lambda\{|\mathbf{s}^T\mathbf{F}\mathbf{u}|^2\} = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} s_i s_j \left(\underbrace{\sum_{l=0}^{M-1}|u_l|^2}_{\|\mathbf{u}\|^2}\mathbf{I}_{i-j} + \sum_{l=0}^{K-1}\sum_{m=0,m\neq l}^{K-1}\frac{u_l\overline{u_m}}{M-1}\left(M\mathbf{I}_i\mathbf{I}_j - \mathbf{I}_{i-j}\right)\right). \tag{37}$$

Changing the sum order results in

$$\mathbb{E}_\Lambda\{|\mathbf{s}^T\mathbf{F}\mathbf{u}|^2\} = \|\mathbf{u}\|^2\sum_{i=0}^{M-1}s_i^2 + \left(\sum_{m\neq l}\frac{u_l\overline{u_m}}{M-1}\right)\left(Ms_0^2 - \sum_{i=0}^{M-1}s_i^2\right). \tag{38}$$

Since $s_i \in \{+1, -1\}$, $s_i^2 = 1$ for all i. Therefore,

$$\mathbb{E}_\Lambda\{|\mathbf{s}^T\mathbf{F}\mathbf{u}|^2\} = M\|\mathbf{u}\|^2. \tag{39}$$

This expression does not depend on the specific signs of $\mathbf{s}$. In other words, the expectation is independent of the row index r, so that

$$\mathbb{E}_\Lambda\{\|\mathbf{S}\mathbf{F}\mathbf{u}\|^2\} = mM\|\mathbf{u}\|^2. \tag{40}$$

Consequently,

$$\mathbb{E}\{Z^2\} = \mathbb{E}\left\{\mathbb{E}_\Lambda\{Z^2\}\right\} = \mathbb{E}\left\{\mathbb{E}_\Lambda\left\{\frac{\|\mathbf{\Phi}\mathbf{u}\|^2}{\|\mathbf{u}\|^2}\right\}\right\} = \mathbb{E}\left\{\frac{\mathbb{E}_\Lambda\left\{\|\mathbf{\Phi}\mathbf{u}\|^2\right\}}{\|\mathbf{u}\|^2}\right\} = 1. \tag{41}$$

The power of the fully random model is to be realized in the next step.

Step 2 - Calculating the forth moment of Z:

Conditioning on the nonzero values we have

$$\mathbb{E}_\Lambda\left\{\|\mathbf{SFu}\|^4\right\} = \mathbb{E}_\Lambda\left\{\left(\sum_{r=0}^{m-1}|\mathbf{s}_r^T\mathbf{Fu}|^2\right)^2\right\} \tag{42}$$

$$= \mathbb{E}_\Lambda \sum_{r,q=0}^{m-1}|\mathbf{s}_r^T\mathbf{Fu}|^2|\mathbf{s}_q^T\mathbf{Fu}|^2, \tag{43}$$

where as before $\mathbf{s}_r$ represents the rth row of $\mathbf{S}$, denoted as a column vector. To simplify the writing we treat a single term from the last sum, and denote $\mathbf{r} = \mathbf{S}_r = [r_0, \ldots, r_{M-1}]^T$ and $\mathbf{q} = \mathbf{S}_q = [q_0, \ldots, q_{M-1}]^T$. Switching the order of expectation and sum due to linearity gives (for the r, qth term):

$$\sum_{\alpha,\beta,\gamma,\delta=0}^{M-1} r_\alpha r_\beta q_\gamma q_\delta \sum_{a,b,c,d=0}^{K-1} u_a\overline{u_b}u_c\overline{u_d}\mathbb{E}_\Lambda\left\{\omega^{\alpha\Lambda_a-\beta\Lambda_b+\gamma\Lambda_c-\delta\Lambda_d}\right\}. \tag{44}$$

The computation of the expectation becomes dependent on the number of distinct values among $\{a, b, c, d\}$. In what follows, we separate the computation to different cases. Each case leads to a different expression for the inner sum in (44).

Case 1: Equal indices

When $a = b = c = d$, the inner sum becomes

$$\sum_{a=0}^{K-1}|u_a|^4\mathbb{E}_\Lambda\left\{\omega^{\Lambda_a(\alpha-\beta+\gamma-\delta)}\right\} = \mathbf{I}(\alpha-\beta+\gamma-\delta)\left(\sum_i|u_i|^4\right). \tag{45}$$

Case 2: Three indices equal

When 3 indices out of $\{a, b, c, d\}$ are equal but different from the last one, we obtain

- $a = b = c \neq d$: $\left(\sum_{i\neq j}|u_i|^2u_i\overline{u_j}\right) f_2(\alpha-\beta+\gamma,-\delta) \xrightarrow{\text{see below}} 0$
- $a = b = d \neq c$: $\left(\sum_{i\neq j}|u_i|^2\overline{u_i}u_j\right) f_2(\alpha-\beta-\delta,\gamma) \to 0$
- $a = c = d \neq b$: $\left(\sum_{i\neq j}|u_i|^2u_i\overline{u_j}\right) f_2(\alpha+\gamma-\delta,-\beta) \to 0$
- $b = c = d \neq a$: $\left(\sum_{i\neq j}|u_i|^2\overline{u_i}u_j\right) f_2(-\beta+\gamma-\delta,\alpha) \to 0$

Each indices triplet yields the relevant expression when taking the conditional expectation $\mathbb{E}_\Lambda\{\cdot\}$. Then, the right arrow is used to abbreviate the following argument. Recall that our goal is to compute the unconditional expectation $\mathbb{E}\{Z^4\}$, which involves the denominator $\|\mathbf{u}\|^4$, similarly to the final step (41) of calculating $\mathbb{E}\{Z^2\}$. Now, whenever a sum over the nonzero values contains a unit-power term, such as $\overline{u_j}$ in the first sum above, the expectation evaluates to zero. More specifically, when calculating the expectation $\mathbb{E}\left\{\frac{\cdot}{\|\mathbf{u}\|^4}\right\}$, we can condition on (i.e. fix) the other nonzero values and compute the expectation

on $\overline{u_j}$ only. Since u_j has unit power, the expectation integral consists of asymmetric nominator, symmetric denominator $\|\mathbf{u}\|^4$ and symmetric probability density function, which computes to zero. Baring in mind that our goal is to compute $\mathbb{E}\{Z^4\}$, we can ignore expressions which will eventually cancel out.

We point out that the cancelation occurs due to the fully random model of the ExRIP. Had the nonzero values been deterministic, as in StRIP, neither of the expressions would have been canceled. The power of utilizing the ExRIP become more significant in cases 4 and 5.

Case 3: Two pairs

There are 3 scenarios of two pairs of equal indices, which are different from each other:

- $a=b\neq c=d\rightarrow \left(\sum_{i\neq j}|u_i|^2|u_j|^2\right) f_2(\alpha-\beta,\gamma-\delta)$
- $a=c\neq b=d\rightarrow \left(\sum_{i\neq j}u_i^2\overline{u_j}^2\right) f_2(\alpha+\gamma,-\beta-\delta)$
- $a=d\neq b=c\rightarrow \left(\sum_{i\neq j}|u_i|^2|u_j|^2\right) f_2(\alpha-\delta,-\beta+\gamma)$

In this case, no cancelation occurs since we did not assume anything on the second order statistics of the nonzeros distribution.

Case 4: Three different indices

When there is a pair of equal indices and the rest are different we have

- $a=b\neq c\neq d:\quad \left(\sum_{i\neq j\neq k}|u_i|^2u_j\overline{u_k}\right) f_3(\alpha-\beta,\gamma,-\delta)\rightarrow 0$
- $a=c\neq b\neq d:\quad \left(\sum_{i\neq j\neq k}u_i^2\overline{u_ju_k}\right) f_3(\alpha+\gamma,-\beta,-\delta)\rightarrow 0$
- $a=d\neq b\neq c:\quad \left(\sum_{i\neq j\neq k}|u_i|^2u_j\overline{u_k}\right) f_3(\alpha-\delta,-\beta,\gamma)\rightarrow 0$
- $b=c\neq a\neq d:\quad \left(\sum_{i\neq j\neq k}|u_i|^2u_j\overline{u_k}\right) f_3(\alpha,-\beta+\gamma,-\delta)\rightarrow 0$
- $b=d\neq a\neq c:\quad \left(\sum_{i\neq j\neq k}\overline{u_i}^2u_ju_k\right) f_3(\alpha,-\beta-\delta,\gamma)\rightarrow 0$
- $c=d\neq a\neq b:\quad \left(\sum_{i\neq j\neq k}|u_i|^2u_j\overline{u_k}\right) f_3(\alpha,-\beta,\gamma-\delta)\rightarrow 0$

The notation, $a\neq b\neq c$ implies pairwise difference. No need to expand $f_3(\cdot,\cdot,\cdot)$ since all these expressions cancel out due to the ExRIP following the previous arguments.

Case 5: Four different indices

The last case is when all four indices a,b,c,d are different from each other.

$$a\neq b\neq c\neq d:\quad \left(\sum_{i\neq j\neq k\neq l}u_i\overline{u_j}u_k\overline{u_l}\right) f_4(\alpha,-\beta,\gamma,-\delta)\rightarrow 0. \tag{46}$$

Substituting the remaining expressions from cases 1 and 3 into (44) and using (36) gives the following for the inner sum in (44):

$$\frac{\left(\sum_{i\neq j}|u_i|^2|u_j|^2\right)}{M-1}\left[M\left(\mathbf{I}_{-\beta+\gamma}\mathbf{I}_{\alpha-\delta}+\mathbf{I}_{\alpha-\beta}\mathbf{I}_{\gamma-\delta}\right)-2\mathbf{I}_{\alpha-\beta+\gamma-\delta}\right] \tag{47}$$
$$+\frac{\left(\sum_{i\neq j}u_i^2\overline{u_j}^2\right)}{M-1}\left[M\mathbf{I}_{\alpha+\gamma}\mathbf{I}_{-\beta-\delta}-\mathbf{I}_{\alpha-\beta+\gamma-\delta}\right]+\left(\sum_i|u_i|^4\right)\mathbf{I}_{\alpha-\beta+\gamma-\delta}.$$

We are now in the position to change the sum order in (44) and exploit the fact the $\mathbf{r},\mathbf{q}$ are binary vectors. Before proceeding we recall the definition of the cyclic convolution between two vectors $\mathbf{a},\mathbf{b}$ of length M:

$$(\mathbf{a}\odot\mathbf{b})[n]=\sum_{m=0}^{M-1}\mathbf{a}[m]\mathbf{b}\left[(n-m)\bmod M\right]. \tag{48}$$

The cyclic convolution is commutative and associative. Define the flipping operation $\tilde{\mathbf{q}}[n]=\mathbf{q}[-n]$ with indices modulo M. The following consequences are now in force:

$$\sum_{\alpha,\beta,\gamma,\delta=0}^{M-1}r_\alpha r_\beta q_\gamma q_\delta\mathbf{I}_{-\beta+\gamma}\mathbf{I}_{\alpha-\delta}=\sum_{\alpha,\delta=0}^{M-1}r_\alpha q_\delta\mathbf{I}_{\alpha-\delta}\sum_{\beta,\gamma=0}^{M-1}r_\beta q_\gamma\mathbf{I}_{-\beta+\gamma}=(\mathbf{r}^T\mathbf{q})^2, \tag{49a}$$

$$\sum_{\alpha,\beta,\gamma,\delta=0}^{M-1}r_\alpha r_\beta q_\gamma q_\delta\mathbf{I}_{\alpha-\beta}\mathbf{I}_{\gamma-\delta}=\sum_{\alpha,\beta=0}^{M-1}r_\alpha r_\beta\mathbf{I}_{\alpha-\beta}\sum_{\gamma,\delta=0}^{M-1}q_\gamma q_\delta\mathbf{I}_{\gamma-\delta}=M^2, \tag{49b}$$

$$\sum_{\alpha,\beta,\gamma,\delta=0}^{M-1}r_\alpha r_\beta q_\gamma q_\delta\mathbf{I}_{\alpha+\gamma}\mathbf{I}_{-\beta-\delta}=\sum_{\alpha,\gamma=0}^{M-1}r_\alpha q_\gamma\mathbf{I}_{\alpha+\gamma}\sum_{\beta,\delta=0}^{M-1}r_\beta q_\delta\mathbf{I}_{-\beta-\delta}=(\mathbf{r}^T\tilde{\mathbf{q}})^2, \tag{49c}$$

$$\begin{aligned}\sum_{\alpha,\beta,\gamma,\delta=0}^{M-1}r_\alpha r_\beta q_\gamma q_\delta\mathbf{I}_{\alpha-\beta+\gamma-\delta}&=\sum_{\alpha,\gamma,\delta=0}^{M-1}r_\alpha q_\gamma q_\delta r_{\alpha+\gamma-\delta}=\sum_{\alpha,\gamma=0}^{M-1}r_\alpha q_\gamma(\mathbf{r}\odot\mathbf{q})_{\alpha+\gamma}\\&=\sum_{\alpha,\tau=0}^{M-1}r_\alpha q_{\tau-\alpha}(\mathbf{r}\odot\mathbf{q})_\tau=\sum_{\tau=0}^{M-1}(\mathbf{r}\odot\mathbf{q})_\tau(\mathbf{r}\odot\mathbf{q})_\tau=\|\mathbf{r}\odot\mathbf{q}\|^2.\end{aligned} \tag{49d}$$

Therefore, (44) is transformed into

$$\frac{\left(\sum_{i\neq j}|u_i|^2|u_j|^2\right)}{M-1}\left[M\left((\mathbf{r}^T\mathbf{q})^2+M^2\right)-2\|\mathbf{r}\odot\mathbf{q}\|^2\right] \tag{50}$$
$$+\frac{\left(\sum_{i\neq j}u_i^2\overline{u_j}^2\right)}{M-1}\left[M(\mathbf{r}^T\tilde{\mathbf{q}})^2-\|\mathbf{r}\odot\mathbf{q}\|^2\right]+\left(\sum_i|u_i|^4\right)\|\mathbf{r}\odot\mathbf{q}\|^2.$$

After some algebraic manipulations (50) becomes

$$\frac{\left(\|\mathbf{u}\|^4 - \sum |u_i|^4\right)}{M-1}\left[M\left((\mathbf{r}^T\mathbf{q})^2 + M^2\right) - 2\|\mathbf{r}\odot\mathbf{q}\|^2\right] \tag{51}$$
$$+ \frac{\left(|\sum u_i^2|^2 - \sum |u_i|^4\right)}{M-1}\left[M(\mathbf{r}^T\tilde{\mathbf{q}})^2 - \|\mathbf{r}\odot\mathbf{q}\|^2\right] + \left(\sum |u_i|^4\right)\|\mathbf{r}\odot\mathbf{q}\|^2.$$

So far we have treated the r,qth term of (43). Taking the sum over all row pairs from $\mathbf{S}$, and using the definitions of $\boldsymbol{\alpha}(S), \boldsymbol{\beta}(S), \boldsymbol{\gamma}(S)$ results in

$$\mathbb{E}_\Lambda\{\|\mathbf{SFu}\|^4\} = \frac{\left(\|\mathbf{u}\|^4 - \sum |u_i|^4\right)}{M-1}\left[M\left(m^2M^2\boldsymbol{\alpha}(S) + m^2M^2\right) - 2m^2M^3\boldsymbol{\beta}(S)\right] \tag{52}$$
$$+ \frac{\left(|\sum u_i^2|^2 - \sum |u_i|^4\right)}{M-1}\left[Mm^2M^2\boldsymbol{\gamma}(S) - m^2M^3\boldsymbol{\beta}(S)\right] + \left(\sum |u_i|^4\right)m^2M^3\boldsymbol{\beta}(S).$$

Consequently,

$$\mathbb{E}\{Z^4\} = \mathbb{E}\left\{\mathbb{E}_\Lambda\left\{\frac{\|\mathbf{SFu}\|^4}{m^2M^2\|\mathbf{u}\|^4}\right\}\right\} \tag{53}$$
$$= \frac{(1-C_K)}{M-1}\left[M\left(\boldsymbol{\alpha}(S)+1\right) - 2M\boldsymbol{\beta}(S)\right] + \frac{(B_K-C_K)}{M-1}\left[M\boldsymbol{\gamma}(S) - M\boldsymbol{\beta}(S)\right] + C_KM\boldsymbol{\beta}(S)$$
$$= (1-C_K)\rho_M\left(1+\boldsymbol{\alpha}(S) - 2\boldsymbol{\beta}(S)\right) + (B_K-C_K)\rho_M\left(\boldsymbol{\gamma}(S) - \boldsymbol{\beta}(S)\right) + C_KM\boldsymbol{\beta}(S).$$

Step 3 - Calculating the ExRIP probability:

To calculate the ExRIP probability we note that for any vector $\mathbf{u}$

$$(1-\delta_K)\|\mathbf{u}\|^2 \le \|\mathbf{\Phi u}\|^2 \le (1+\delta_K)\|\mathbf{u}\|^2 \quad \longleftrightarrow \quad \left|\underbrace{\frac{\|\mathbf{\Phi u}\|^2}{\|\mathbf{u}\|^2}}_{Z^2} - 1\right| \le \delta_K. \tag{54}$$

With $\mathbb{E}\{Z^2\} = 1$, we can now apply the Chebyshev inequality

$$\text{Prob}\left\{|Z^2 - \mathbb{E}\{Z^2\}| \le \delta_K\right\} \ge 1 - \frac{\text{Var}\{Z^2\}}{\delta_K^2}. \tag{55}$$

Substituting the variance completes the proof. ■