Structured Compressed Sensing: From Theory to Applications

Marco F. Duarte, *Member, IEEE*, and Yonina C. Eldar, *Senior Member, IEEE*

***Abstract*—Compressed sensing (CS) is an emerging field that has attracted considerable research interest over the past few years. Previous review articles in CS limit their scope to standard discrete-to-discrete measurement architectures using matrices of randomized nature and signal models based on standard sparsity. In recent years, CS has worked its way into several new application areas. This, in turn, necessitates a fresh look on many of the basics of CS. The random matrix measurement operator must be replaced by more structured sensing architectures that correspond to the characteristics of feasible acquisition hardware. The standard sparsity prior has to be extended to include a much richer class of signals and to encode broader data models, including continuous-time signals. In our overview, the theme is exploiting signal and measurement structure in compressive sensing. The prime focus is bridging theory and practice; that is, to pinpoint the potential of structured CS strategies to emerge from the math to the hardware. Our summary highlights new directions as well as relations to more traditional CS, with the hope of serving both as a review to practitioners wanting to join this emerging field, and as a reference for researchers that attempts to put some of the existing ideas in perspective of practical applications.**

***Index Terms*—Approximation algorithms, compressed sensing, compression algorithms, data acquisition, data compression, sampling methods.**

I. Introduction and Motivation

COMPRESSED sensing (CS) is an emerging field that has attracted considerable research interest in the signal processing community. Since its introduction only several years ago [1], [2], thousands of papers have appeared in this area, and hundreds of conferences, workshops, and special sessions have been dedicated to this growing research field.

Due to the vast interest in this topic, there exist several excellent review articles on the basics of CS [3]–[5]. These articles focused on the first CS efforts: the use of standard discrete-to-discrete measurement architectures using matrices of randomized nature, where no structure beyond sparsity is assumed on the signal or in its representation. This basic formulation already required the use of sophisticated mathematical tools and rich theory in order to analyze recovery approaches and provide performance guarantees. It was therefore essential to confine attention to this simplified setting in the early stages of development of the CS framework.

Essentially all analog-to-digital converters to date follow the celebrated Shannon–Nyquist theorem which requires the sampling rate to be at least twice the bandwidth of the signal. This basic principle underlies the majority of digital signal processing (DSP) applications such as audio, video, radio receivers, radar applications, medical devices and more. The ever growing demand for data, as well as advances in radio frequency (RF) technology, have promoted the use of high-bandwidth signals, for which the rates dictated by the Shannon–Nyquist theorem impose severe challenges both on the acquisition hardware and on the subsequent storage and DSP processors. CS was motivated in part by the desire to sample wideband signals at rates far lower than the Shannon–Nyquist rate, while still maintaining the essential information encoded in the underlying signal. In practice, however, much of the work to date on CS has focused on acquiring finite-dimensional sparse vectors using random measurements. This precludes the important case of continuous-time (i.e., analog) input signals, as well as practical hardware architectures which inevitably are structured. Achieving the "holy grail" of compressive ADCs and increased resolution requires a broader framework which can treat more general signal models, including continuous-time signals with various types of structure, as well as practical measurement schemes.

In recent years, the area of CS has branched out to many new fronts and has worked its way into several application areas. This, in turn, necessitates a fresh look on many of the basics of CS. The random matrix measurement operator, fundamental in all early presentations of CS, must be replaced by more structured measurement operators that correspond to the application of interest, such as wireless channels, analog sampling hardware, sensor networks and optical imaging. The standard sparsity prior that has characterized early work in CS has to be extended to include a much richer class of signals: signals that have underlying low-dimensional structure, not necessarily represented by standard sparsity, and signals that can have arbitrary dimensions, not only finite-dimensional vectors.

A significant part of the recent work on CS from the signal processing community can be classified into two major contribution areas. The first group consists of theory and applications related to CS matrices that are not completely random and

Manuscript received September 01, 2010; revised February 19, 2011 and June 22, 2011; accepted July 03, 2011. Date of publication July 14, 2011; date of current version August 10, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ivan W. Selesnick. The work of M. F. Duarte was supported in part by NSF Supplemental Funding DMS-0439872 to UCLA-IPAM, P. I. R. Caflisch. The work of Y. C. Eldar was supported in part by the Israel Science Foundation under Grant no. 170/10, and by a Magneton grant from the Israel Ministry of Industry and Trade.

M. F. Duarte is with the Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003 USA (e-mail: mduarte@ecs.umass.edu).

Y. C. Eldar is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: yonina@ee.technion.ac.il).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2011.2161982

that often exhibit considerable structure. This largely follows from efforts to model the way the samples are acquired in practice, which leads to sensing matrices that inherent their structure from the real world. The second group includes signal representations that exhibit structure beyond sparsity and broader classes of signals, such as continuous-time signals with infinite-dimensional representations. For many types of signals, such structure allows for a higher degree of signal compression when leveraged on top of sparsity. Additionally, infinite-dimensional signal representations provide an important example of richer structure which clearly cannot be described using standard sparsity. Since reducing the sampling rate in analog signals was one of the driving forces behind CS, building a theory that can accommodate low-dimensional signals in arbitrary Hilbert spaces is clearly an essential part of the CS framework. Both of these categories are motivated by real-world CS involving actual hardware implementations.

In our review, the theme is exploiting signal and measurement structure in CS. The prime focus is bridging theory and practice—that is, to pinpoint the potential of CS strategies to emerge from the math to the hardware by generalizing the underlying theory where needed. We believe that this is an essential ingredient in taking CS to the next step: incorporating this fast growing field into real-world applications. Considerable efforts have been devoted in recent years by many researchers to adapt the theory of CS to better solve real-world signal acquisition challenges. This has also led to parallel low-rate sampling schemes that combine the principles of CS with the rich theory of sampling such as the finite rate of innovation (FRI) [6]–[8] and Xampling frameworks [9], [10]. There are already dozens of papers dealing with these broad ideas. In this review we have strived to provide a coherent summary, highlighting new directions and relations to more traditional CS. This material can both serve as a review to those wanting to join this emerging field, as well as a reference that attempts to summarize some of the existing results in the framework of practical applications. Our hope is that this presentation will attract the interest of both mathematicians and engineers in the desire to promote the CS premise into practical applications, and encourage further research into this new frontier.

This review paper is organized as follows. Section II provides background motivating the formulation of CS and the layout of the review. A primer on standard CS theory is presented in Section III. This material serves as a basis for the later developments. Section IV reviews alternative constructions for structured CS matrices beyond those generated completely at random. In Section V, we introduce finite-dimensional signal models that exhibit additional signal structure. This leads to the more general union-of-subspaces framework, which will play an important role in the context of structured infinite-dimensional representations as well. Section VI extends the concepts of CS to infinite-dimensional signal models and introduces recent compressive ADCs which have been developed based on the Xampling and FRI frameworks. For each of the matrices and models introduced, we summarize the details of the theoretical and algorithmic frameworks and provide example applications where the structure is applicable.

II. Background

We live in a digital world. Telecommunication, entertainment, medical devices, gadgets, business—all revolve around digital media. Miniature sophisticated black-boxes process streams of bits accurately at high speeds. Nowadays, electronic consumers feel natural that a media player shows their favorite movie, or that their surround system synthesizes pure acoustics, as if sitting in the orchestra instead of the living room. The digital world plays a fundamental role in our everyday routine, to such a point that we almost forget that we cannot "hear" or "watch" these streams of bits, running behind the scenes.

Analog-to-digital conversion (ADC) lies at the heart of this revolution. ADC devices translate the physical information into a stream of numbers, enabling digital processing by sophisticated software algorithms. The ADC task is inherently intricate: its hardware must hold a snapshot of a fast-varying input signal steady while acquiring measurements. Since these measurements are spaced in time, the values between consecutive snapshots are lost. In general, therefore, there is no way to recover the analog signal unless some prior on its structure is incorporated.

After sampling, the numbers or bits retained must be stored and later processed. This requires ample storage devices and sufficient processing power. As technology advances, so does the requirement for ever-increasing amounts of data, imposing unprecedented strains on both the ADC devices and the subsequent DSP and storage media. How then does consumer electronics keep up with these high demands? Fortunately, most of the data we acquire can be discarded without much perceptual loss. This is evident in essentially all compression techniques used to date. However, this paradigm of high-rate sampling followed by compression does not alleviate the large strains on the acquisition device and on the DSP. In his seminal work on CS [1], Donoho posed the ultimate goal of merging compression and sampling: "Why go to so much effort to acquire all the data when most of what we get will be thrown away? Can't we just directly measure the part that won't end up being thrown away?"

A. Shannon–Nyquist Theorem

ADCs provide the interface between an analog signal being recorded and a suitable discrete representation. A common approach in engineering is to assume that the signal is bandlimited, meaning that the spectral contents are confined to a maximal frequency B. Bandlimited signals have limited time variation, and can therefore be perfectly reconstructed from equispaced samples with rate at least $2B$, termed the Nyquist rate. This fundamental result is often attributed in the engineering community to Shannon–Nyquist [11], [12], although it dates back to earlier works by Whittaker [13] and Kotelńikov [14].

Theorem 1 [12]: If a function $x(t)$ contains no frequencies higher than B hertz, then it is completely determined by giving its ordinates at a series of points spaced $\frac{1}{(2B)}$ seconds apart.

A fundamental reason for processing at the Nyquist rate is the clear relation between the spectrum of $x(t)$ and that of its samples $x(nT)$, so that digital operations can be easily substituted for their analog counterparts. Digital filtering is an example

where this relation is successfully exploited. Since the power spectral densities of analog and discrete random processes are associated in a similar manner, estimation and detection of parameters of analog signals can be performed by DSP. In contrast, compression is carried out by a series of algorithmic steps, which, in general, exhibit a nonlinear complicated relationship between the samples $x(nT)$ and the stored data.

While this framework has driven the development of signal acquisition devices for the last half century, the increasing complexity of emerging applications dictates increasingly higher sampling rates that cannot always be met using available hardware. Advances in related fields such as wideband communication and RF technology open a considerable gap with ADC devices. Conversion speeds which are twice the signal's maximal frequency component have become more and more difficult to obtain. Consequently, alternatives to high rate sampling are drawing considerable attention in both academia and industry.

Structured analog signals can often be processed far more efficiently than what is dictated by the Shannon–Nyquist theorem, which does not take any structure into account. For example, many wideband communication signals are comprised of several narrow transmissions modulated at high carrier frequencies. A common practice in engineering is demodulation in which the input signal is multiplied by the carrier frequency of a band of interest, in order to shift the contents of the narrowband transmission from the high frequencies to the origin. Then, commercial ADC devices at low rates are utilized. Demodulation, however, requires knowing the exact carrier frequency. In this review we focus on structured models in which the exact parameters defining the structure are unknown. In the context of multiband communications, for example, the carrier frequencies may not be known, or may be changing over time. The goal then is to build a compressed sampler which does not depend on the carrier frequencies, but can nonetheless acquire and process such signals at rates below Nyquist.

B. Compressed Sensing and Beyond

A "holy grail" of CS is to build acquisition devices that exploit signal structure in order to reduce the sampling rate, and subsequent demands on storage and DSP. In such an approach, the actual information contents dictate the sampling rate, rather than the dimensions of the ambient space in which the signal resides. The challenges in achieving this task both theoretically and in terms of hardware design can be reduced substantially when considering finite-dimensional problems in which the signal to be measured can be represented as a discrete finite-length vector. This has spurred a surge of research on various mathematical and algorithmic aspects of sensing sparse signals, which were mainly studied for discrete finite vectors.

At its core, CS is a mathematical framework that studies accurate recovery of a signal represented by a vector of length N from $M \ll N$ measurements, effectively performing compression during signal acquisition. The measurement paradigm consists of linear projections, or inner products, of the signal vector into a set of carefully chosen projection vectors that act as a multitude of probes on the information contained in the signal. In the first part of this review (Sections III and IV) we survey the fundamentals of CS and show how the ideas can be extended to allow for more elaborate measurement schemes that incorporate structure into the measurement process. When considering real-world acquisition schemes, the choices of possible measurement matrices are dictated by the constraints of the application. Thus, we must deviate from the general randomized constructions and apply structure within the projection vectors that can be easily implemented by the acquisition hardware. Section IV focuses on such alternatives; we survey both existing theory and applications for several classes of structured CS matrices. In certain applications, there exist hardware designs that measure analog signals at a sub-Nyquist rate, obtaining measurements for finite-dimensional signal representations via such structured CS matrices.

In the second part of this review (Sections V and VI), we expand the theory of CS to signal models tailored to express structure beyond standard sparsity. A recent emerging theoretical framework that allows a broader class of signal models to be acquired efficiently is the union of subspaces model [15]–[20]. We introduce this framework and some of its applications in a finite-dimensional context in Section V, which include more general notions of structure and sparsity. Combining the principles and insights from the previous sections, in Section VI we extend the notion of CS to analog signals with infinite-dimensional representations. This new framework, referred to as Xampling [9], [10], relies on more general signal models—union of subspaces and FRI signals—together with guidelines on how to exploit these mathematical structures in order to build sensing devices that can directly acquire analog signals at reduced rates. We then survey several compressive ADCs that result from this broader framework.

III. Compressed Sensing Basics

CS [1]–[5] offers a framework for simultaneous sensing and compression of finite-dimensional vectors, that relies on linear dimensionality reduction. Specifically, in CS we do not acquire x directly but rather acquire $M < N$ linear measurements $y = \Phi x$ using an $M \times N$ *CS matrix* Φ. We refer to y as the *measurement vector*. Ideally, the matrix Φ is designed to reduce the number of measurements M as much as possible while allowing for recovery of a wide class of signals x from their measurement vectors y. However, the fact that $M < N$ renders the matrix Φ *rank-defficient*, meaning that it has a nonempty nullspace; this, in turn, implies that for any particular signal $x_0 \in \mathbb{R}^N$, an infinite number of signals x will yield the same measurements $y_0 = \Phi x_0 = \Phi x$ for the chosen CS matrix Φ.

The motivation behind the design of the matrix Φ is, therefore, to allow for distinct signals x, x' within a class of signals of interest to be uniquely identifiable from their measurements $y = \Phi x$, $y' = \Phi x'$, even though $M \ll N$. We must therefore make a choice on the class of signals that we aim to recover from CS measurements.

A. Sparsity

Sparsity is the signal structure behind many compression algorithms that employ transform coding, and is the most prevalent signal structure used in CS. Sparsity also has a rich history of applications in signal processing problems in the last century

(particularly in imaging), including denoising, deconvolution, restoration, and inpainting [21]–[23].

To introduce the notion of sparsity, we rely on a signal representation in a given basis $\{\psi_i\}_{i=1}^N$ for $\mathbb{R}^N$. Every signal $x \in \mathbb{R}^N$ is representable in terms of N coefficients $\{\theta_i\}_{i=1}^N$ as $x = \sum_{i=1}^N \psi_i\theta_i$; arranging the ψ_i as columns into the $N \times N$ matrix Ψ and the coefficients θ_i into the $N \times 1$ *coefficient vector* θ, we can write succinctly that $x = \Psi\theta$, with $\theta \in \mathbb{R}^N$. Similarly, if we use a frame[1] Ψ containing N unit-norm column vectors of length L with $L < N$ (i.e., $\Psi \in \mathbb{R}^{L\times N}$), then for any vector $x \in \mathbb{R}^L$ there exist infinitely many decompositions $\theta \in \mathbb{R}^N$ such that $x = \Psi\theta$. In a general setting, we refer to Ψ as the sparsifying *dictionary* [24]. While our exposition is restricted to real-valued signals, the concepts are extendable to complex signals as well [25], [26].

We say that a signal x is K-*sparse* in the basis or frame Ψ if there exists a vector $\theta \in \mathbb{R}^N$ with only $K \ll N$ nonzero entries such that $x = \Psi\theta$. We call the set of indices corresponding to the nonzero entries the *support* of θ and denote it by $\mathrm{supp}(\theta)$. We also define the set Σ_K that contains all signals x that are K-sparse.

A K-sparse signal can be efficiently compressed by preserving only the values and locations of its nonzero coefficients, using $\mathcal{O}(K\log_2 N)$ bits: coding each of the K nonzero coefficient's locations takes $\log_2 N$ bits, while coding the magnitudes uses a constant amount of bits that depends on the desired precision, and is independent of N. This process is known as *transform coding*, and relies on the existence of a suitable basis or frame Ψ that renders signals of interest sparse or approximately sparse.

For signals that are not exactly sparse, the amount of compression depends on the number of coefficients of θ that we keep. Consider a signal x whose coefficients θ, when sorted in order of decreasing magnitude, decay according to the power law

$$|\theta(\mathcal{I}(n))| \le Sn^{-\frac{1}{r}}, \quad n = 1, \ldots, N \tag{1}$$

where $\mathcal{I}$ indexes the sorted coefficients. Thanks to the rapid decay of their coefficients, such signals are well-approximated by K-sparse signals. The best K-term approximation error for such a signal obeys

$$\sigma_\Psi(x, K) := \arg\min_{x' \in \Sigma_K} \|x - x'\|_2 \le CSK^{-s} \tag{2}$$

with $s = \frac{1}{r} - \frac{1}{2}$ and C denoting a constant that does not depend on N [27]. That is, the signal's best approximation error (in an ℓ_2-norm sense) has a power-law decay with exponent s as K increases. We dub such a signal *s-compressible*. When Ψ is an orthonormal basis, the best sparse approximation of x is obtained by hard thresholding the signal's coefficients, so that only the K coefficients with largest magnitudes are preserved. The situation is decidedly more complicated when Ψ is a general frame, spurring the development of sparse approximation methods, a subject that we will focus on in Section III-C.

[1]A matrix Ψ is said to be a frame if there exist constants A, B such that $A\|x\|_2 \le \|\Psi x\|_2 \le B\|x\|_2$ for all x.

When sparsity is used as the signal structure enabling CS, we aim to recover x from y by exploiting its sparsity. In contrast with transform coding, we do not operate on the signal x directly, but rather only have access to the CS measurement vector y. Our goal is to push M as close as possible to K in order to perform as much signal "compression" during acquisition as possible. In the sequel, we will assume that Ψ is taken to be the identity basis so that the signal $x = \theta$ is itself sparse. In certain cases we will explicitly define a different basis or frame Ψ that arises in a specific application of CS.

B. Design of CS Matrices

The main design criteria for the CS matrix Φ is to enable the unique identification of a signal of interest x from its measurements $y = \Phi x$. Clearly, when we consider the class of K-sparse signals Σ_K, the number of measurements $M > K$ for any matrix design, since the identification problem has K unknowns even when the support $\Omega = \mathrm{supp}(x)$ of the signal x is provided. In this case, we simply restrict the matrix Φ to its columns corresponding to the indices in Ω, denoted by Φ_Ω, and then use the pseudoinverse to recover the nonzero coefficients of x:

$$x_\Omega = \Phi_\Omega^\dagger y. \tag{3}$$

Here, x_Ω is the restriction of the vector x to the set of indices Ω, and $\mathbf{M}^\dagger = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T$ denotes the pseudoinverse of the matrix $\mathbf{M}$. The implicit assumption in (3) is that Φ_Ω has full column-rank so that there is a unique solution to the equation $y = \Phi_\Omega x_\Omega$.

We begin by determining properties of Φ that guarantee that distinct signals $x, x' \in \Sigma_K$, $x \ne x'$, lead to different measurement vectors $\Phi x \ne \Phi x'$. In other words, we want each vector $y \in \mathbb{R}^M$ to be matched to at most one vector $x \in \Sigma_K$ such that $y = \Phi x$. A key relevant property of the matrix in this context is its *spark*.

Definition 1 [28]: *The spark* $\mathrm{spark}(\Phi)$ *of a given matrix* Φ *is the smallest number of columns of* Φ *that are linearly dependent.*

The spark is related to the *Kruskal rank* from the tensor product literature; the matrix Φ has Kruskal rank $\mathrm{spark}(\Phi) - 1$. This definition allows us to pose the following straightforward guarantee.

Theorem 2 [28]: *If* $\mathrm{spark}(\Phi) > 2K$*, then for each measurement vector* $y \in \mathbb{R}^M$ *there exists at most one signal* $x \in \Sigma_K$ *such that* $y = \Phi x$.

It is easy to see that $\mathrm{spark}(\Phi) \in [2, M+1]$, so that Theorem 2 yields the requirement $M \ge 2K$.

While Theorem 2 guarantees uniqueness of representation for K-sparse signals, computing the spark of a general matrix Φ has combinatorial computational complexity, since one must verify that all sets of columns of a certain size are linearly independent. Thus, it is preferable to use properties of Φ that are easily computable to provide recovery guarantees. The *coherence* of a matrix is one such property.

Definition 2 [28]–[31]: *The coherence* $\mu(\Phi)$ *of a matrix* Φ *is the largest absolute inner product between any two columns of* Φ:

$$\mu(\Phi) = \max_{1 \le i \ne j \le N} \frac{|\langle \phi_i, \phi_j \rangle|}{\|\phi_i\|_2 \|\phi_j\|_2}. \tag{4}$$

It can be shown that $\mu(\Phi) \in [\sqrt{\frac{N-M}{M(N-1)}}, 1]$; the lower bound is known as the Welch bound [32], [33]. Note that when $N \gg M$, the lower bound is approximately $\mu(\Phi) \geq \frac{1}{\sqrt{M}}$. One can tie the coherence and spark of a matrix by employing the Gershgorin circle theorem.

Theorem 3 [34]: *The eigenvalues of an $m \times m$ matrix $\mathbf{M}$ with entries $\mathbf{M}_{i,j}$, $1 \leq i,j \leq m$, lie in the union of m discs $d_i = d_i(c_i, r_i)$, $1 \leq i \leq m$, centered at $c_i = \mathbf{M}_{i,i}$ with radius $r_i = \sum_{j \neq i} |\mathbf{M}_{i,j}|$.*

Applying this theorem on the Gram matrix $\mathbf{G} = \Phi_\Omega^T \Phi_\Omega$ leads to the following result.

Lemma 1: [28] *For any matrix Φ,*

$$\text{spark}(\Phi) \geq 1 + \frac{1}{\mu(\Phi)}. \tag{5}$$

By merging Theorem 2 with Lemma 1, we can pose the following condition on Φ that guarantees uniqueness.

Theorem 4 [28], [30], [31]: *If*

$$K < \frac{1}{2}\left(1 + \frac{1}{\mu(\Phi)}\right) \tag{6}$$

then for each measurement vector $y \in \mathbb{R}^M$ there exists at most one signal $x \in \Sigma_K$ such that $y = \Phi x$.

Theorem 4, together with the Welch bound, provides an upper bound on the level of sparsity K that guarantees uniqueness using coherence: $K = \mathcal{O}(\sqrt{M})$.

The prior properties of the CS matrix provide guarantees of uniqueness when the measurement vector y is obtained without error. Hardware considerations introduce two main sources of inaccuracies in the measurements: inaccuracies due to noise at the sensing stage (in the form of additive noise $y = \Phi x + n$), and inaccuracies due to mismatches between the CS matrix used during recovery, Φ, and that implemented during acquisition, $\Phi' = \Phi + \Delta$ (in the form of multiplicative noise [35], [36]). Under these sources of error, it is no longer possible to guarantee uniqueness; however, it is desirable for the measurement process to be tolerant to both types of error. To be more formal, we would like the distance between the measurement vectors for two sparse signals $y = \Phi x$, $y' = \Phi x'$ to be proportional to the distance between the original signal vectors x and x'. Such a property allows us to guarantee that, for small enough noise, two sparse vectors that are far apart from each other cannot lead to the same (noisy) measurement vector. This behavior has been formalized into the *restricted isometry property* (RIP).

Definition 3 [4]: *A matrix Φ has the (K, δ)-restricted isometry property ((K, δ)-RIP) if, for all $x \in \Sigma_K$,*

$$(1 - \delta)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta)\|x\|_2^2. \tag{7}$$

In words, the (K, δ)-RIP ensures that all submatrices of Φ of size $M \times K$ are close to an isometry, and therefore distance-preserving. We will show later that this property suffices to prove that the recovery is stable to presence of additive noise n. In certain settings, noise is introduced to the signal x prior to measurement. Recovery is also stable for this case; however, there is a degradation in the distortion of the recovery by a factor of $\frac{N}{M}$ [37]–[40]. Furthermore, the RIP also leads to stability with respect to the multiplicative noise introduced by the CS matrix mismatch Δ [35], [36].

The RIP can be connected to the coherence property by using, once again, the Gershgorin circle theorem (Theorem 3).

Lemma 2: [41] *If Φ has unit-norm columns and coherence $\mu = \mu(\Phi)$, then Φ has the (K, δ)-RIP with $\delta \leq (K-1)\mu$.*

One can also easily connect RIP with the spark. For each K-sparse vector to be uniquely identifiable by its measurements, it suffices for the matrix Φ to have the $(2K, \delta)$-RIP with $\delta > 0$, as this implies that all sets of $2K$ columns of Φ are linearly independent, i.e., $\text{spark}(\Phi) > 2K$ (cf. Theorems 2 and 4). We will see later that the RIP enables recovery guarantees that are much stronger than those based on spark and coherence. However, checking whether a CS matrix Φ satisfies the (K, δ)-RIP has combinatorial computational complexity.

Now that we have defined relevant properties of a CS matrix Φ, we discuss specific matrix constructions that are suitable for CS. An $M \times N$ Vandermonde matrix $\mathbf{V}$ constructed from N distinct scalars has $\text{spark}(\mathbf{V}) = M + 1$ [27]. Unfortunately, these matrices are poorly conditioned for large values of N, rendering the recovery problem numerically unstable. Similarly, there are known matrices Φ of size $M \times M^2$ that achieve the coherence lower bound

$$\mu(\Phi) = \frac{1}{\sqrt{M}} \tag{8}$$

such as the Gabor frame generated from the Alltop sequence [42] and more general equiangular tight frames [33]. It is also possible to construct deterministic CS matrices of size $M \times N$ that have the (K, δ)-RIP for $K = \mathcal{O}\left(\frac{\sqrt{M}\log M}{\log(\frac{N}{M})}\right)$ [43]. These constructions restrict the number of measurements needed to recover a K-sparse signal to be $M = \mathcal{O}(K^2 \log N)$, which is undesirable for real-world values of N and K.

Fortunately, these bottlenecks can be defeated by randomizing the matrix construction. For example, random matrices Φ of size $M \times N$ whose entries are independent and identically distributed (i.i.d.) with continuous distributions have $\text{spark}(\Phi) = M + 1$ with high probability. It can also be shown that when the distribution used has zero mean and finite variance, then in the asymptotic regime (as M and N grow) the coherence converges to $\mu(\Phi) = 2\sqrt{\frac{\log N}{M}}$ [44], [45]. Similarly, random matrices from Gaussian, Rademacher, or more generally a sub-Gaussian distribution[2] have the (K, δ)-RIP with high probability if [46]

$$M = \mathcal{O}\left(\frac{K \log\left(\frac{N}{K}\right)}{\delta^2}\right). \tag{9}$$

Finally, we point out that while the set of RIP-fulfilling matrices provided above might seem limited, emerging numerical results have shown that a variety of classes of matrices Φ are suitable for CS recovery at regimes similar to those of the matrices advocated in this section, including subsampled Fourier and Hadamard transforms [47], [48].

[2]A Rademacher distribution gives probability $\frac{1}{2}$ to the values ± 1. A random variable X is called sub-Gaussian if there exists $c > 0$ such that $\mathbb{E}(e^{Xt}) \leq e^{\frac{c^2 t^2}{2}}$ for all $t \in \mathbb{R}$. Examples include the Gaussian, Bernoulli, and Rademacher random variables, as well as any bounded random variable.

C. CS Recovery Algorithms

We now focus on solving the CS recovery problem: given y and Φ, find a signal x within the class of interest such that $y = \Phi x$ exactly or approximately.

When we consider sparse signals, the CS recovery process consists of a search for the sparsest signal x that yields the measurements y. By defining the ℓ_0 "norm" of a vector $\|x\|_0$ as the number of nonzero entries in x, the simplest way to pose a recovery algorithm is using the optimization

$$\widehat{x} = \arg\min_{x\in\mathbb{R}^N} \|x\|_0 \text{ subject to } y = \Phi x. \tag{10}$$

Solving (10) relies on an exhaustive search and is successful for all $x \in \Sigma_K$ when the matrix Φ has the sparse solution uniqueness property (i.e., for M as small as $2K$, see Theorems 2 and 4). However, this algorithm has combinatorial computational complexity, since we must check whether the measurement vector y belongs to the span of each set of K columns of Φ, $K = 1, 2, \ldots, N$. Our goal, therefore, is to find computationally feasible algorithms that can successfully recover a sparse vector x from the measurement vector y for the smallest possible number of measurements M.

An alternative to the ℓ_0 "norm" used in (10) is to use the ℓ_1 norm, defined as $\|x\|_1 = \sum_{n=1}^{N} |x(n)|$. The resulting adaptation of (10), known as basis pursuit (BP) [22], is formally defined as

$$\widehat{x} = \arg\min_{x\in\mathbb{R}^N} \|x\|_1 \text{ subject to } y = \Phi x. \tag{11}$$

Since the ℓ_1 norm is convex, (11) can be seen as a convex relaxation of (10). Thanks to the convexity, this algorithm can be implemented as a linear program, making its computational complexity polynomial in the signal length [49].[3]

The optimization (11) can be modified to allow for noise in the measurements $y = \Phi x + n$; we simply change the constraint on the solution to

$$\widehat{x} = \arg\min_{x\in\mathbb{R}^N} \|x\|_1 \text{ subject to } \|y - \Phi x\|_2 \le \epsilon \tag{12}$$

where $\epsilon \ge \|n\|_2$ is an appropriately chosen bound on the noise magnitude. This modified optimization is known as basis pursuit with inequality constraints (BPIC) and is a quadratic program with polynomial complexity solvers [49]. The Lagrangian relaxation of this quadratic program is written as

$$\widehat{x} = \arg\min_{x\in\mathbb{R}^N} \|x\|_1 + \lambda\|y - \Phi x\|_2 \tag{13}$$

and is known as basis pursuit denoising (BPDN). There exist many efficient solvers to find BP, BPIC, and BPDN solutions; for an overview, see [51].

Oftentimes, a bounded-norm noise model is overly pessimistic, and it may be reasonable instead to assume that the noise is random. For example, additive white Gaussian noise $n \sim \mathcal{N}(0, \sigma^2\mathbf{I})$ is a common choice. Approaches designed to address stochastic noise include complexity-based regularization [52] and Bayesian estimation [53]. These methods pose probabilistic or complexity-based priors, respectively, on the set of observable signals. The particular prior is then leveraged together with the noise probability distribution during signal recovery. Optimization-based approaches can also be formulated in this case; one of the most popular techniques is the Dantzig selector [54]:

$$\widehat{x} = \arg\min_{x\in\mathbb{R}^N} \|x\|_1 \text{ s.t. } \left\|\Phi^T(y - \Phi x)\right\|_\infty \le \lambda\sqrt{\log N}\sigma \tag{14}$$

where $\|\cdot\|_\infty$ denotes the ℓ_∞-norm, which provides the largest-magnitude entry in a vector and λ is a constant parameter that controls the probability of successful recovery.

An alternative to optimization-based approaches, are *greedy algorithms* for sparse signal recovery. These methods are iterative in nature and select columns of Φ according to their correlation with the measurements y determined by an appropriate inner product. For example, the matching pursuit and orthogonal matching pursuit algorithms (OMP) [24], [55] proceed by finding the column of Φ most correlated to the signal residual, which is obtained by subtracting the contribution of a partial estimate of the signal from y. The OMP method is formally defined as Algorithm 1, where $\mathcal{T}(x, K)$ denotes a *thresholding* operator on x that sets all but the K entries of x with the largest magnitudes to zero, and $\mathbf{x}|_\Omega$ denotes the restriction of $\mathbf{x}$ to the entries indexed by Ω. The convergence criterion used to find sparse representations consists of checking whether $y = \Phi x$ exactly or approximately; note that due to its design, the algorithm cannot run for more than M iterations, as Φ has M rows. Other greedy techniques that have a similar flavor to OMP include CoSaMP [56], detailed as Algorithm 2, and subspace pursuit (SP) [57]. A simpler variant is known as iterative hard thresholding (IHT) [58]: starting from an initial signal estimate $\widehat{x}_0 = 0$, the algorithm iterates a gradient descent step followed by hard thresholding, i.e.,

$$\widehat{x}_i = \mathcal{T}\left(\widehat{x}_{i-1} + \Phi^T(y - \Phi\widehat{x}_{i-1}), K\right), \tag{15}$$

until a convergence criterion is met.

Algorithm 1: Orthogonal Matching Pursuit

Input: CS matrix Φ, measurement vector y

Output: Sparse representation $\widehat{x}$

Initialize: $\widehat{x}_0 = 0$, $r = y$, $\Omega = \emptyset$, $i = 0$

while halting criterion false **do**

$i \leftarrow i + 1$

$b \leftarrow \Phi^T r$ {form residual signal estimate}

$\Omega \leftarrow \Omega \cup \text{supp}(\mathcal{T}(b, 1))$ {update support with residual}

$\widehat{x}_i|_\Omega \leftarrow \Phi_\Omega^\dagger y, \widehat{x}_i|_{\Omega^C} \leftarrow 0$ {update signal estimate}

$r \leftarrow y - \Phi\widehat{x}_i$ {update measurement residual}

end while

return $\widehat{x} \leftarrow \widehat{x}_i$

[3]A similar set of recovery algorithms, known as total variation minimizers, operate on the gradient of an image, which exhibits sparsity for piecewise smooth images [50].

Algorithm 2: CoSaMP

Input: CS matrix Φ, measurement vector y, sparsity K

Output: K-sparse approximation $\widehat{x}$ to true signal x

Initialize: $\widehat{x}_0 = 0, r = y, i = 0$

while halting criterion false **do**

$i \leftarrow i + 1$

$e \leftarrow \Phi^T r$ {form residual signal estimate}

$\Omega \leftarrow \text{supp}(\mathcal{T}(e, 2K))$ {prune residual}

$T \leftarrow \Omega \cup \text{supp}(\widehat{x}_{i-1})$ {merge supports}

$b|_T \leftarrow \Phi_T^{\dagger} y, b|_{T^C} \leftarrow 0$ {form signal estimate}

$\widehat{x}_i \leftarrow \mathcal{T}(b, K)$ {prune signal using model}

$r \leftarrow y - \Phi\widehat{x}_i$ {update measurement residual}

end while

return $\widehat{x} \leftarrow \widehat{x}_i$

D. CS Recovery Guarantees

Many of the CS recovery algorithms above come with guarantees on their performance. We group these results according to the matrix metric used to obtain the guarantee.

First, we review results that rely on coherence. As a first example, BP and OMP recover a K-sparse vector from noiseless measurements when the matrix Φ satisfies (6) [28], [30], [31]. There also exist coherence-based guarantees designed for measurements corrupted with arbitrary noise.

Theorem 5 [59]: *Let the signal $x \in \Sigma_K$ and write $y = \Phi x + n$. Denote $\gamma = \|n\|_2$. Suppose that $K \leq \frac{\left(\frac{1}{\mu(\Phi)}+1\right)}{4}$ and $\epsilon \geq \gamma$ in (12). Then the output $\widehat{x}$ of (12) has error bounded by*

$$\|x - \widehat{x}\|_2 \leq \frac{\gamma + \epsilon}{\sqrt{1 - \mu(\Phi)(4K-1)}}, \tag{16}$$

while the output $\widehat{x}$ of the OMP algorithm with halting criterion $\|r\|_2 \leq \gamma$ has error bounded by

$$\|x - \widehat{x}\|_2 \leq \frac{\gamma}{\sqrt{1 - \mu(\Phi)(K-1)}} \tag{17}$$

provided that $\gamma \leq \frac{A(1-\mu(\Phi)(2K-1))}{2}$ for OMP, with A being a positive lower bound on the magnitude of the nonzero entries of x.

Note here that BPIC must be aware of the noise magnitude γ to make $\epsilon \geq \gamma$, while OMP must be aware of the noise magnitude γ to set an appropriate convergence criterion. Additionally, the error in Theorem 5 is proportional to the noise magnitude γ. This is because the only assumption on the noise is its magnitude, so that n might be aligned to maximally harm the estimation process.

In the random noise case, bounds on $\|x - \widehat{x}\|_2$ can only be stated in high probability, since there is always a small probability that the noise will be very large and completely overpower the signal. For example, under additive white Gaussian noise (AWGN), the guarantees for BPIC in Theorems 5 hold with high probability when the parameter $\epsilon = \sigma\sqrt{M + \eta\sqrt{2M}}$, with η denoting an adjustable parameter to control the probability of $\|n\|_2$ being too large [60]. A second example gives a related result for the BPDN algorithm.

Theorem 6 [61]: *Let the signal $x \in \Sigma_K$ and write $y = \Phi x + n$, where $n \sim \mathcal{N}(0, \sigma^2\mathbf{I})$. Suppose that $K < \frac{1}{3\mu(\Phi)}$ and consider the BPDN optimization problem (13) with $\lambda = \sqrt{16\sigma^2 \log M}$. Then, with probability on the order of $1 - \frac{1}{M^2}$, the solution $\widehat{x}$ of (13) is unique, its error is bounded by*

$$\|x - \widehat{x}\|_2 \leq C\sigma\sqrt{K \log M}, \tag{18}$$

and its support is a subset of the true K-element support of x.

Under AWGN, the value of ϵ one would need to choose in Theorem 5 is $\mathcal{O}(\sigma\sqrt{M})$, giving a bound much larger than Theorem 6, which is $\mathcal{O}(\sigma\sqrt{K \log M})$. This demonstrates the noise reduction achievable due to the adoption of the random noise model. These guarantees come close to the Cramér–Rao bound, which is given by $\sigma\sqrt{K}$ [62]. We finish the study of coherence-based guarantees for the AWGN setting with a result for OMP.

Theorem 7 [61]: *Let the signal $x \in \Sigma_K$ and write $y = \Phi x + n$, where $n \sim \mathcal{N}(0, \sigma^2\mathbf{I})$. Suppose that $K \leq (1/\mu(\Phi) + 1)/4$* and

$$\min_{1 \leq n \leq N} |x(n)| \geq \frac{2\sigma\sqrt{2(1+\alpha)\log N}}{1 - (2K-1)\mu(\Phi)} \tag{19}$$

for some constant $\alpha > 0$. Then, with probability at least $1 - (N^{\alpha}\sqrt{\pi(1+\alpha)\log N})^{-1}$, the output $\widehat{x}$ of OMP after K iterations has error bounded by

$$\|x - \widehat{x}\|_2 \leq C\sigma\sqrt{(1+\alpha)K \log N}, \tag{20}$$

and its support matches the true K-element support of x.

The greedy nature of OMP poses the requirement on the minimum absolute-valued entry of x in order for the support to be correctly detected, in contrast to BPIC and BPDN.

A second class of guarantees are based on the RIP. The following result for OMP provides an interesting viewpoint of greedy algorithms.

Theorem 8 [63]: *Let the signal $x \in \Sigma_K$ and write $y = \Phi x$. Suppose that Φ has the $(K+1, \delta)$-RIP with $\delta < \frac{1}{3\sqrt{K}}$. Then OMP can recover a K-sparse signal x exactly in K iterations.*

Guarantees also exist for noisy measurement settings, albeit with significantly more stringent RIP conditions on the CS matrix.

Theorem 9 [64]: *Let the signal $x \in \Sigma_K$ and write $y = \Phi x + n$. Suppose that Φ has the $(31K, \delta)$-RIP with $\delta < \frac{1}{3}$. Then the output of OMP after $30K$ iterations has error bounded by*

$$\|x - \widehat{x}\|_2 \leq C\|n\|_2. \tag{21}$$

The next result extends guarantees from sparse to more general signals measured under noise. We collect a set of independent statements in a single theorem.

Theorem 10 [4], [56]–[58]: *Let the signal $x \in \Sigma_K$ and write $y = \Phi x + n$. The outputs $\widehat{x}$ of the CoSaMP, SP, IHT, and BPIC algorithms, with Φ having the (cK, δ)-RIP, obey*

$$\|x - \widehat{x}\|_2 \leq C_1 \|x - x_K\|_2 + C_2 \frac{1}{\sqrt{K}} \|x - x_K\|_1 + C_3 \|n\|_2 \tag{22}$$

where $x_K = \arg\min_{x' \in \Sigma_K} |x - x'|_2$ is the best K-sparse approximation of the vector x when measured in the ℓ_2 norm. The requirements on the parameters c, δ of the RIP and the values of C_1, C_2, and C_3 are specific to each algorithm. For example, for the BPIC algorithm, $c = 2$ and $\delta = \sqrt{2} - 1$ suffice to obtain the guarantee (22).

The type of guarantee given in Theorem 10 is known as *uniform instance optimality*, in the sense that the CS recovery error is proportional to that of the best K-sparse approximation to the signal x for any signal $x \in \mathbb{R}^N$. In fact, the formulation of the CoSaMP, SP and IHT algorithms was driven by the goal of instance optimality, which has not been shown for older greedy algorithms like MP and OMP. Theorem 10 can also be adapted to recovery of exactly sparse signals from noiseless measurements.

Corollary 11: *Let the signal $x \in \Sigma_K$ and write $y = \Phi x$. The CoSaMP, SP, IHT, and BP algorithms can exactly recover x from y if Φ has the (cK, δ)-RIP, where the parameters c, δ of the RIP are specific to each algorithm.*

Similarly to Theorem 5, the error in Theorem 10 is proportional to the noise magnitude $\|n\|_2$, and the bounds can be tailored to random noise with high probability. The Dantzig selector improves the scaling of the error in the AWGN setting.

Theorem 12 [54]: *Let the signal $x \in \Sigma_K$ and write $y = \Phi x + n$, where $n \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$. Suppose that $\lambda = \sqrt{2}\left(1 + \frac{1}{t}\right)$ in (14) and that Φ has the $(2K, \delta_{2K})$ and $(3K, \delta_{3K})$-RIPs with $\delta_{2K} + \delta_{3K} < 1$. Then, with probability at least $1 - \frac{N^t}{\sqrt{\pi \log N}}$, we have*

$$\|\widehat{x} - x\|_2 \leq C \left(1 + \frac{1}{t}\right)^2 K \sigma^2 \log N. \tag{23}$$

Similar results under AWGN have been shown for the OMP and thresholding algorithms [61].

A third class of guarantees relies on metrics additional to coherence and RIP. This class has a *nonuniform* flavor in the sense that the results apply only for a certain subset of sufficiently sparse signals. Such flexibility allows for significant relaxations on the properties required from the matrix Φ. The next example has a probabilistic flavor and relies on the coherence property.

Theorem 13 [65]: *Let the signal $x \in \Sigma_K$ with support drawn uniformly at random from the available $\binom{N}{K}$ possibilities and entries drawn independently from a distribution $P(X)$ so that $P(X > 0) = P(X < 0)$. Write $y = \Phi x$ and fix $s \geq 1$ and $0 < \delta < \frac{1}{2}$. If $K \leq \frac{\log\left(\frac{N}{\delta}\right)}{8\mu^2(\Phi)}$ and*

$$\sqrt{18 \log \frac{N}{\delta} \log\left(\frac{K}{2} + 1\right)} s + \frac{2K}{N} \|\Phi\|_2^2 \leq e^{\frac{-1}{4}} \left(1 - 2^{-\frac{1}{2}}\right),$$

then x is the unique solution to BP (11) with probability at least $1 - 2\delta - \left(\frac{K}{2}\right)^{-s}$.

In words, the theorem says that as long as the coherence $\mu(\Phi)$ and the spectral norm $\|\Phi\|_2$ of the CS matrix are small enough, we will be able to recover the majority of K-sparse signals x from their measurements y. Probabilistic results that rely on coherence can also be obtained for the BPDN algorithm (13) [45].

The main difference between the guarantees that rely solely on coherence and those that rely on the RIP and probabilistic sparse signal models is the scaling of the number of measurements M needed for successful recovery of K-sparse signals. According to the bounds (8) and (9), the sparsity level that allows for recovery with high probability in Theorems 10, 12, and 13 is $K = \mathcal{O}(M)$, compared with $K = \mathcal{O}(\sqrt{M})$ for the deterministic guarantees provided by Theorems 5, 6, and 7. This so-called *square root bottleneck* [65] gives an additional reason for the popularity of randomized CS matrices and sparse signal models.

IV. Structure in CS Matrices

While most initial work in CS has emphasized the use of randomized CS matrices whose entries are obtained independently from a standard probability distribution, such matrices are often not feasible for real-world applications due to the cost of multiplying arbitrary matrices with signal vectors of high dimension. In fact, very often the physics of the sensing modality and the capabilities of sensing devices limit the types of CS matrices that can be implemented in a specific application. Furthermore, in the context of analog sampling, one of the prime motivations for CS is to build analog samplers that lead to sub-Nyquist sampling rates. These involve actual hardware and therefore structured sensing devices. Hardware considerations require more elaborate signal models to reduce the number of measurements needed for recovery as much as possible. In this section, we review available alternatives for structured CS matrices; in each case, we provide known performance guarantees, as well as application areas where the structure arises. In Section VI we extend the CS framework to allow for analog sampling, and introduce further structure into the measurement process. This results in new hardware implementations for reduced rate samplers based on extended CS principles. Note that the survey of CS devices given in this section is by no means exhaustive [66]–[68]; our focus is on CS matrices that have been investigated from both a theoretical and an implementational point of view.

A. Subsampled Incoherent Bases

The key concept of a frame's coherence can be extended to pairs of orthonormal bases. This enables a new choice for CS matrices: one simply selects an orthonormal basis that is incoherent with the sparsity basis, and obtains CS measurements by selecting a subset of the coefficients of the signal in the chosen basis [69]. We note that some degree of randomness remains in this scheme, due to the choice of coefficients selected to represent the signal as CS measurements.

1) Formulation: Formally, we assume that a basis $\Phi \in \mathbb{R}^{N \times N}$ is provided for measurement purposes, where each column of $\Phi = [\phi_1\ \phi_2\ \ldots\ \phi_N]$ corresponds to a different basis element. Let $\overline{\Phi}$ be an $N \times M$ column submatrix of Φ that preserves the basis vectors with indices Γ and set $y = \overline{\Phi}^T x$.

Under this setup, a different metric arises to evaluate the performance of CS.

Definition 4 [28], [69]: *The mutual coherence of the N-dimensional orthonormal bases Φ and Ψ is the maximum absolute value of the inner product between elements of the two* bases:

$$\mu(\Phi, \Psi) = \max_{1 \le i,j \le N} |\langle \phi_i, \psi_j \rangle| \tag{24}$$

where ψ_j denotes the j^{th} column, or element, of the basis Ψ.

The mutual coherence $\mu(\Phi, \Psi)$ has values in the range $\left[N^{-\frac{1}{2}}, 1\right]$. For example, $\mu(\Phi, \Psi) = N^{-\frac{1}{2}}$ when Φ is the discrete Fourier transform basis, or Fourier matrix, and Ψ is the canonical basis, or identity matrix, and $\mu(\Phi, \Psi) = 1$ when both bases share at least one element or column. Note also that the concepts of coherence and mutual coherence are connected by the equality $\mu(\Phi, \Psi) = \mu([\Phi\ \Psi])$. The definition of mutual coherence can also be extended to infinite-dimensional representations of continuous-time (analog) signals [33], [70].

2) Theoretical Guarantees: The following theorem provides a recovery guarantee based on mutual coherence.

Theorem 14 [69]: *Let $x = \Psi\theta$ be a K-sparse signal in Ψ with support $\Omega \subset \{1, \ldots, N\}$, $|\Omega| = K$, and with entries having signs chosen uniformly at random. Choose a subset $\Gamma \subseteq \{1, \ldots, N\}$ uniformly at random for the set of observed measurements, with $M = |\Gamma|$. Suppose that $M \ge CKN\mu^2(\Phi, \Psi) \log\left(\frac{N}{\delta}\right)$ and $M \ge C' \log^2\left(\frac{N}{\delta}\right)$ for fixed values of $\delta < 1, C, C'$. Then with probability at least $1 - \delta$, θ is the solution to (11).*

The number of measurements required by Theorem 14 ranges from $O(K \log N)$ to $O(N)$. It is possible to expand the guarantee of Theorem 14 to compressible signals by adapting an argument of Rudelson and Vershynin in [71] to link coherence and restricted isometry constants.

Theorem 15 [71 Remark 3.5.3]: *Choose a subset $\Gamma \subseteq \{1, \ldots, N\}$ for the set of observed measurements, with $M = |\Gamma|$. Suppose that*

$$M \ge CK\sqrt{N}t\mu(\Phi, \Psi) \log(tK \log N) \log^2 K \tag{25}$$

for a fixed value of C. Then with probability at least $1 - 5e^{-t}$ the matrix $\Phi^T\Psi$ has the RIP with constant $\delta_{2K} \le \frac{1}{2}$.

Using this theorem, we obtain the guarantee of Theorem 10 for compressible signals with the number of measurements M dictated by the coherence value $\mu(\Phi, \Psi)$.

3) Applications: There are two main categories of applications where subsampled incoherent bases are used. In the first category, the acquisition hardware is limited by construction to measure directly in a transform domain. The most relevant examples are magnetic resonance imaging (MRI) [72] and tomographic imaging [73], as well as optical microscopy [74], [75]; in all of these cases, the measurements obtained from the hardware correspond to coefficients of the image's 2-D continuous Fourier transform, albeit not typically selected in a randomized fashion. Since the Fourier functions, corresponding to sinusoids, will be incoherent with functions that have localized support, this imaging approach works well in practice for sparsity/compressibility transforms such as wavelets [69], total variation [73], and the standard canonical representation [74].

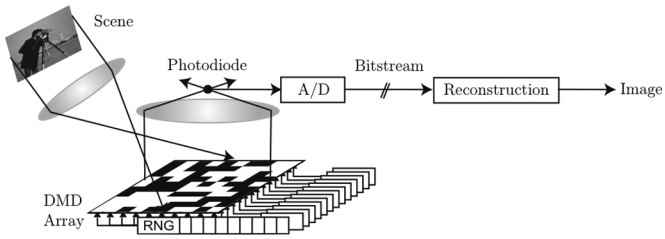


Fig. 1. Diagram of the single pixel camera. The incident lightfield (corresponding to the desired image x) is reflected off a digital micro-mirror device (DMD) array whose mirror orientations are modulated in the pseudorandom pattern supplied by the random number generator (RNG). Each different mirror pattern produces a voltage at the single photodiode that corresponds to one measurement $y(m)$. The process is repeated with different patterns M times to obtain a full measurement vector y (taken from [76]).

In the case of optical microscopy, the Fourier coefficients that are measured correspond to the lowpass regime. The highpass values are completely lost. When the underlying signal x can change sign, standard sparse recovery algorithms such as BP do not typically succeed in recovering the true underlying vector. To treat the case of recovery from lowpass coefficients, a special purpose sparse recovery method was developed under the name of Nonlocal Hard Thresholding (NLHT) [74]. This technique attempts to allocate the off-support of the sparse signal in an iterative fashion by performing a thresholding step that depends on the values of the neighboring locations.

The second category involves the design of new acquisition hardware that can obtain projections of the signal against a class of vectors. The goal of the matrix design step is to find a basis whose elements belong to the class of vectors that can be implemented on the hardware. For example, a class of single pixel imagers based on optical modulators [60], [76] (see an example in Fig. 1) can obtain projections of an image against vectors that have binary entries. Example bases that meet this criterion include the Walsh–Hadamard and noiselet bases [77]. The latter is particularly interesting for imaging applications, as it is known to be maximally incoherent with the Haar wavelet basis. In contrast, certain elements of the Walsh-Hadamard basis are highly coherent with wavelet functions at coarse scales, due to their large supports. Permuting the entries of the basis vectors (in a random or pseudorandom fashion) helps reduce the coherence between the measurement basis and a wavelet basis. Because the single pixel camera modulates the light field through optical aggregation, it improves the signal-to-noise ratio of the measurements obtained when compared to standard multisensor schemes [78]. Similar imaging hardware architectures have been developed in [79], [80].

An additional example of a configurable acquisition device is the random sampling ADC [81], which is designed for acquisition of periodic, multitone analog signals whose frequency components belong to a uniform grid (similarly to the random demodulator of Section IV-B-3). The sparsity basis for the signal once again is chosen to be the discrete Fourier transform basis. The measurement basis is chosen to be the identity basis, so that the measurements correspond to standard signal samples taken at randomized times. The hardware implementation employs a randomized clock that drives a traditional low-rate ADC to sample the input signal at specific times. As shown in Fig. 2,

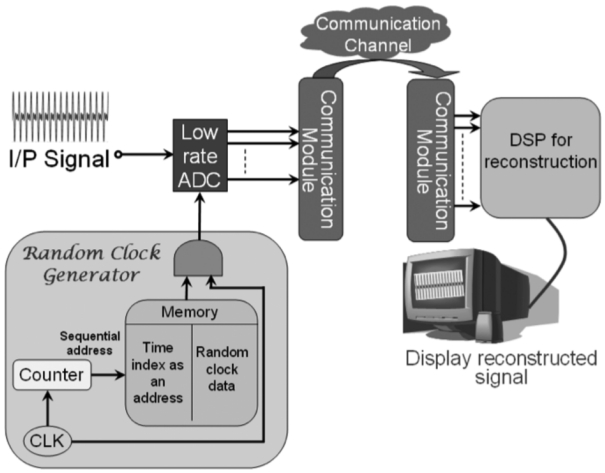


Fig. 2. Diagram of the random sampling ADC. A pseudorandom clock generated by an FPGA drives a low-rate standard ADC to sample the input signal at predetermined pseudorandom times. The samples obtained are then fed into a CS recovery algorithm or a sketching algorithm for Fourier sparse signals to estimate the frequencies and amplitudes of the tones present in the signal (taken from [81]).

the random clock is implemented using an FPGA that outputs a predetermined pseudorandom pulse sequence indicating the times at which the signal is sampled. The patterns are timed according to a set of pseudorandom arithmetic progressions. This process is repeated cyclically, establishing a windowed acquisition for signals of arbitrary length. The measurement and sparsity framework used by this randomized ADC is also compatible with sketching algorithms designed for signals that are sparse in the frequency domain [81], [82].

B. Structurally Subsampled Matrices

In certain applications, the measurements obtained by the acquisition hardware do not directly correspond to the sensed signal's coefficients in a particular transform. Rather, the observations are a linear combination of multiple coefficients of the signal. The resulting CS matrix has been termed a structurally subsampled matrix [83].

1) Formulation: Consider a matrix of available measurement vectors that can be described as the product $\Phi = \mathbf{RU}$, where $\mathbf{R}$ is a $P \times N$ mixing matrix and $\mathbf{U}$ is a basis. The CS matrix $\overline{\Phi}$ is obtained by selecting M out of P rows at random, and normalizing the columns of the resulting subsampled matrix. There are two possible downsampling stages: first, $\mathbf{R}$ might offer only $P < N$ mixtures to be available as measurements; second, we only preserve $M < P$ of the mixtures available to represent the signal. This formulation includes the use of subsampled incoherent bases simply by letting $P = N$ and $\mathbf{R} = \mathbf{I}$, i.e., no coefficient mixing is performed.

2) Theoretical Guarantees: To provide theoretical guarantees we place some additional constraints on the mixing matrix $\mathbf{R}$.

Definition 5: *The (P, N) integrator matrix $\mathbf{S}$, for P a divisor of N, is defined as $\mathbf{S} = [\mathbf{s}_1^T\ \mathbf{s}_2^T\ \ldots\ \mathbf{s}_P^T]^T$, where the p^{th} row of $\mathbf{S}$ is defined as $\mathbf{s}_p = [\mathbf{0}_{1\times(p-1)L}\ \mathbf{1}_{1\times L}\ \mathbf{0}_{1\times(N-pL)}]$, $1 \le p \le P$, with $L = \frac{P}{N}$.*

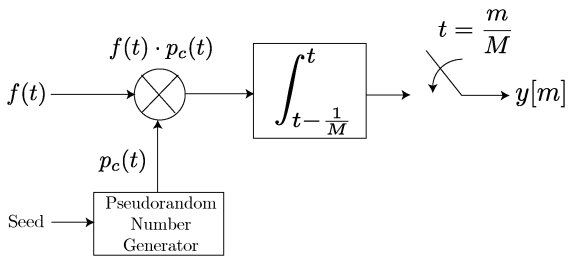


Fig. 3. Block diagram of the random demodulator (taken from [84]).

In words, using $\mathbf{S}$ as a mixing matrix sums together intervals of L adjacent coefficients of the signal under the transform $\mathbf{U}$. We also use a diagonal modulation matrix $\mathbf{M}$ whose nonzero entries are independently drawn from a Rademacher distribution, and formulate our mixing matrix as $\mathbf{R} = \mathbf{SM}$.

Theorem 16 [83, Theorem 1]: *Let $\overline{\Phi}$ be a structurally subsampled matrix of size $M \times N$ obtained from the basis $\mathbf{U}$ and the $P \times N$ mixing matrix $\mathbf{R} = \mathbf{SM}$ via randomized subsampling. Then for each integer $K > 2$, any $z > 1$ and any $\delta \in (0, 1)$, there exist absolute positive constants c_1 and c_2 such that if*

$$M \ge c_1 z K N \mu^2(\mathbf{U}, \Psi) \log^3 N \log^2 K, \tag{26}$$

then the matrix $\overline{\Phi}$ has the (K, δ)-RIP with probability at least $1 - 20\max\{\exp(-c_2\delta^2 z), N^{-1}\}$.

Similarly to the subsampled incoherent bases case, the possible values of $\mu(\mathbf{U}, \Psi)$ provide us with a required number of measurements M ranging from $O(K \log^3 N)$ to $O(N)$.

3) Applications: Compressive ADCs are one promising application of CS, which we discuss in detail in Section VI-B-6 after introducing the infinite-dimensional CS framework. A first step in this direction is the architecture known as the random demodulator (RD) [84]. The RD employs structurally subsampled matrices for the acquisition of periodic, multitone analog signals whose frequency components belong in a uniform grid. Such signals have a finite parametrization and therefore fit the finite-dimensional CS setting.

To be more specific, our aim is to acquire a discrete uniform sampling $x \in \mathbb{R}^N$ of a continuous-time signal $f(t)$ observed during an acquisition interval of 1 s, where it is assumed that $f(t)$ is of the form

$$f(t) = \sum_{k=1}^{N} x_k e^{j2\pi kt}. \tag{27}$$

The vector x is sparse so that only K of its values are nonzero. As shown in Fig. 3, the sampling hardware consists of a mixer element that multiplies the signal $f(t)$ with a chipping sequence $p_c(t)$ at a rate N chirps/s. This chipping sequence is the output of a pseudorandom number generator. The combination of these two units effectively performs the product of the Nyquist-sampled discrete representation of the signal x with the matrix $\mathbf{M}$ in the analog domain. The output from the modulator is sampled by an "integrate-and-dump" sampler—a combination of an accumulator unit and a uniform sampler synchronized with each other—at a rate of M samples per interval. Such sampling effectively performs a multiplication of the output of the modulator

by the (M, N) integrator matrix $\mathbf{S}$. To complete the setup, we set $\mathbf{U} = \mathbf{I}$ and Ψ to be the discrete Fourier transform basis; it is easy to see that these two bases are maximally incoherent. In the RD architecture, all subsampling is performed at this stage; there is no randomized subsampling of the output of the integrator.

A prototype implementation of this architecture is reported in [85]; similar designs are proposed in [5] and [86]. Note that the number of columns of the resulting CS matrix scales with the maximal frequency in the representation of $f(t)$. Therefore, in practice, this maximal frequency cannot be very large. For example, the implementations reported above reach maximum frequency rates of 1 MHz; the corresponding signal representation therefore has one million coefficients. The matrix multiplications can be implemented using algorithmic tools such as the fast Fourier transform (FFT). In conjunction with certain optimization solvers and greedy algorithms, this approach significantly reduces the complexity of signal recovery. While the original formulation sets the sparsity basis Ψ to be the Fourier basis, limiting the set of recoverable signals to periodic multitone signals, we can move beyond structurally subsampled matrices by using redundant Fourier domain frames for sparsity. These frames, together with modified recovery algorithms, can enable acquisition of a wider class of frequency-sparse signals [87], [88].

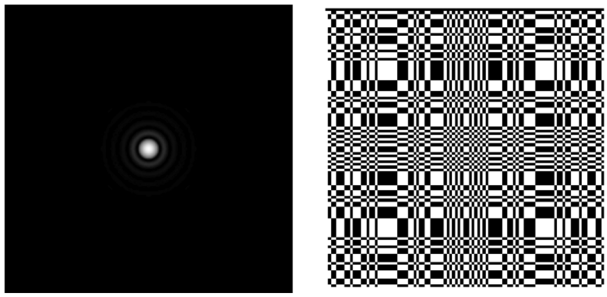
While the RD is capable of implementing CS acquisition for a specific class of analog signals having a finite parametrization, the CS framework can be adapted to infinite-dimensional signal models, enabling more efficient analog acquisition and digital processing architectures. We defer the details to Section VI.

C. Subsampled Circulant Matrices

The use of Toeplitz and circulant structures [89]–[91] as CS matrices was first inspired by applications in communications—including channel estimation and multi-user detection—where a sparse prior is placed on the signal to be estimated, such as a channel response or a multiuser activity pattern. When compared with generic CS matrices, subsampled circulant matrices have a significantly smaller number of degrees of freedom due to the repetition of the matrix entries along the rows and columns.

1) Formulation: A circulant matrix $\mathbf{U}$ is a square matrix where the entries in each diagonal are all equal, and where the first entry of the second and subsequent rows is equal to the last entry of the previous row. Since this matrix is square, we perform random subsampling of the rows (in a fashion similar to that described in Section IV-B) to obtain a CS matrix $\Phi = \mathbf{R}\mathbf{U}$, where $\mathbf{R}$ is an $M \times N$ subsampling matrix, i.e., a submatrix of the identity matrix. We dub Φ a *subsampled circulant matrix*.

2) Theoretical Guarantees: Even when the sequence defining $\mathbf{U}$ is drawn at random from the distributions described in Section III, the particular structure of the subsampled circulant matrix $\Phi = \mathbf{R}\mathbf{U}$ prevents the use of the proof techniques used in standard CS, which require all entries of the matrix to be independent. However, it is possible to employ different probabilistic tools to provide guarantees for subsampled circulant matrices. The results still require randomness in the selection of the entries of the circulant matrix.



(a) (b)

Fig. 4. Compressive imaging via coded aperture. (a) Example of an ideal point spread function for a camera with a pinhole aperture. The size of the support of the function is dependent on the amount of blur caused by the imager. (b) Coded aperture used for compressive imaging (taken from [94]).

Theorem 17 [91]: *Let Φ be a subsampled circulant matrix whose distinct entries are independent random variables following a Rademacher distribution, and $\mathbf{R}$ is an arbitrary $M \times N$ identity submatrix. Furthermore, let δ be the smallest value for which (7) holds for all $x \in \Sigma_K$. Then for $\delta_0 \in (0,1)$ we have $\mathbb{E}[\delta] \leq \delta_0$ provided that*

$$M \geq \mathcal{C} \max\left\{\delta_0^{-1} K^{\frac{3}{2}} \log^{\frac{3}{2}} N, \delta_0^{-2} K \log^2 N \log^2 K\right\}$$

where $\mathcal{C} > 0$ is a universal constant. Furthermore, for $0 \leq \lambda \leq 1$,

$$\mathbb{P}\left(\delta_K \geq \mathbb{E}[\delta] + \lambda\right) \leq e^{\frac{\lambda^2}{\sigma^2}}, \text{ where } \sigma^2 = C' \frac{K}{m} \log^2 K \log N,$$

for a universal constant $C' > 0$.

In words, if we have $M = \mathcal{O}(K^{1.5} \log^{1.5} N)$, then the RIP required for Φ by many algorithms for successful recovery is achieved with high probability.

3) Applications: There are several sensing applications where the signal to be acquired is convolved with the sampling hardware's impulse response before it is measured. Additionally, because convolution is equivalent to a product operator in the Fourier domain, it is possible to speed up the CS recovery process by performing multiplications by the matrices Φ and Φ^T via the FFT. In fact, such an FFT-based procedure can also be exploited to generate good CS matrices [92]. First, we design a matrix Φ in the Fourier domain to be diagonal and have entries drawn from a suitable probability distribution. Then, we obtain the measurement matrix $\overline{\Phi}$ by subsampling the matrix Φ, similarly to the incoherent basis case. While this formulation assumes that the convolution during signal acquisition is circulant, this gap between theory and practice has been studied and shown to be controllable [93].

Our first example concerns imaging architectures. The impulse response of the imaging hardware is known as the point spread function (PSF), and it represents imaging artifacts such as blurring, distortion, and other aberrations; an example is shown in Fig. 4(a). It is possible then to design a compressive imaging architecture by employing an imaging device that has a dense PSF: an impulse response having a support that includes all of the imaging field. This dense PSF is coupled with a

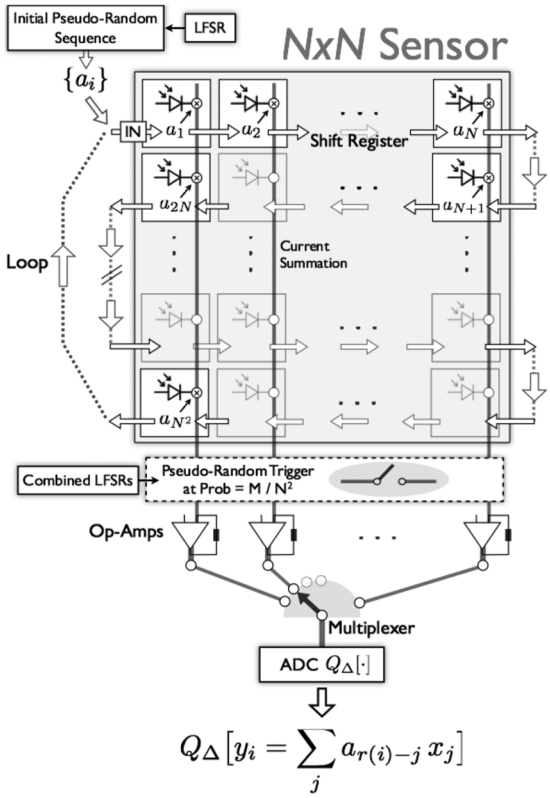


Fig. 5. Schematic for the microelectronic architecture of the convolution-based compressive imager. The acquisition device effectively implements a subsampled circulant CS matrix. Here, Q_Δ denotes the quantization operation with resolution Δ (taken from [95]).

downsampling step in the pixel domain to achieve compressive data acquisition [93], [94]. This is in contrast to the random sampling advocated by Theorem 17. A popular way to achieve such a dense PSF is by employing so-called coded apertures, which change the pinhole aperture used for image formation in most cameras to a more complex design. Fig. 4(b) shows an example coded aperture that has been used successfully in compressive imaging [93], [94].

Our second example uses special-purpose light sensor arrays that perform the multiplication with a Toeplitz matrix using a custom microelectronic architecture [95], which is shown in Fig. 5. In this architecture, an $N \times N$ pixel light sensor array is coupled with a linear feedback shift register (LFSR) of length N^2 whose input is connected to a pseudorandom number generator. The bits in the LFSR control multipliers that are tied to the outputs from the N^2 light sensors, thus performing additions and subtractions according to the pseudorandom pattern generated. The weighted outputs are summed in two stages: column-wise sums are obtained at an operational amplifier before quantization, whose outputs are then added together in an accumulator. In this way, the microelectronic architecture calculates the inner product of the image's pixel values with the sequence contained in the LFSR. The output of the LFSR is fed back to its input after the register is full, so that subsequent measurements correspond to inner products with shifts of previous sequences. The output of the accumulator is sampled in a pseudorandom fashion, thus performing the subsampling required for CS. This results in an effective subsampled circulant CS matrix.

D. Separable Matrices

Separable matrices [96], [97] provide computationally efficient alternatives to measure very large signals, such as hypercube samplings from multidimensional data. These matrices have a succinct mathematical formulation as Kronecker products, and feature significant structure present as correlations among matrix subblocks.

1) Formulation: Kronecker product bases are well suited for CS applications concerning multidimensional signals, that is, signals that capture information from an event that spans multiple dimensions, such as spatial, temporal, spectral, etc. These bases can be used both to obtain sparse or compressible representations or as CS measurement matrices.

The *Kronecker product* of two matrices A and B of sizes $P \times Q$ and $R \times S$, respectively, is defined as

$$A \otimes B := \begin{bmatrix} A(1,1)B & A(1,2)B & \dots & A(1,Q)B \\ A(2,1)B & A(2,2)B & \dots & A(2,Q)B \\ \vdots & \vdots & \ddots & \vdots \\ A(P,1)B & A(P,2)B & \dots & A(P,Q)B \end{bmatrix}.$$

Thus, $A \otimes B$ is a matrix of size $PR \times QS$. Let Ψ_1 and Ψ_2 be bases for $\mathbb{R}^{N_1}$ and $\mathbb{R}^{N_2}$, respectively. Then one can find a basis for $\mathbb{R}^{N_1} \otimes \mathbb{R}^{N_2} = \mathbb{R}^{N_1 N_2}$ as $\widetilde{\Psi} = \Psi_1 \otimes \Psi_2$. We focus now on CS matrices that can be expressed as the Kronecker product of D matrices:

$$\overline{\Phi} = \Phi_1 \otimes \Phi_2 \otimes \dots \otimes \Phi_D. \tag{28}$$

If we denote the size of Φ_d as $M_d \times N_d$, then the matrix Φ has size $M \times N$, with $M = \prod_{d=1}^{D} M_d$ and $N = \prod_{d=1}^{D} N_d$.

Next, we describe the use of Kronecker product matrices in CS of multidimensional signals. We denote the entries of a D-dimensional signal x by $x(n_1, \dots, n_d)$. We call the restriction of a multidimensional signal to fixed indices for all but its d^{th} dimension a d-section of the signal. For example, for a 3-D signal $x \in \mathbb{R}^{N_1 \times N_2 \times N_3}$, the vector $x_{i,j,\cdot} := [x(i,j,1)\, x(i,j,2)\, \dots\, x(i,j,N_3)]$ is a 3-section of x.

Kronecker product sparsity bases make it possible to simultaneously exploit the sparsity properties of a multidimensional signal along each of its dimensions. The new basis is simply the Kronecker product of the bases used for each of its d-sections. More formally, we let $x \in \mathbb{R}^{N_1} \otimes \dots \otimes \mathbb{R}^{N_D} = \mathbb{R}^{N_1 \times \dots \times N_D} = \mathbb{R}^{\prod_{d=1}^{D} N_d}$ and assume that each d-section is sparse or compressible in a basis Ψ_d. We then define a sparsity/compressibility basis for x as $\overline{\Psi} = \Psi_1 \otimes \dots \otimes \Psi_D$, and obtain a coefficient vector $\overline{\theta}$ for the signal ensemble so that $\overline{x} = \overline{\Psi}\overline{\theta}$, where $\overline{x}$ is a column vector-reshaped representation of x. We then have $y = \overline{\Phi}\overline{x} = \overline{\Phi\Psi}\theta$.

2) Theoretical Guarantees: Due to the similarity between blocks of Kronecker product CS matrices, it is easy to obtain bounds for their performance metrics. Our first result concerns the RIP.

Lemma 3 [96], [97]: *Let Φ_d, $1 \leq d \leq D$, be matrices that have the (K, δ_d)-RIP, $1 \leq d \leq D$, respectively. Then $\overline{\Phi}$, defined in (28), has the (K, δ)-RIP, with*

$$\delta \leq \prod_{d=1}^{D}(1+\delta_d) - 1. \tag{29}$$

When Φ_d is an orthonormal basis, it has the (K, δ_d)-RIP with $\delta_d = 0$ for all $K \leq N$. Therefore the RIP constant of the Kronecker product of an orthonormal basis and a CS matrix is equal to the RIP constant of the CS matrix.

It is also possible to obtain results on mutual coherence (described in Section III) for cases in which the basis used for sparsity or compressibility can also be expressed as a Kronecker product.

Lemma 4 [96], [97]: *Let Φ_d, Ψ_d be bases for $\mathbb{R}^{N_d}$ for $d = 1, \ldots, D$. Then*

$$\mu(\overline{\Phi}, \overline{\Psi}) = \prod_{d=1}^{D} \mu(\Phi_d, \Psi_d). \tag{30}$$

Lemma 4 provides a *conservation of mutual coherence* across Kronecker products. Since the mutual coherence of each d-section's sparsity and measurement bases is upper bounded by one, the number of Kronecker product-based measurements necessary for successful recovery of the multidimensional signal (following Theorems 14 and 15) is always lower than or equal to the corresponding number of necessary *partitioned measurements* that process only a portion of the multidimensional signal along its d^{th} dimension at a time, for some $d \in \{1, \ldots, D\}$. This reduction is maximized when the d-section measurement basis Φ is maximally incoherent with the d-section sparsity basis Ψ.

3) Applications: Most applications of separable CS matrices involve multidimensional signals such as video sequences and hyperspectral datacubes. Our first example is an extension of the single pixel camera (see Fig. 1) to hyperspectral imaging [98]. The aim here is to record the reflectivity of a scene at different wavelengths; each wavelength has a corresponding spectral frame, so that the hyperspectral datacube can be essentially interpreted as the stacking of a sequence of images corresponding to the different spectra. The single pixel hyperspectral camera is obtained simply by replacing the single photodiode element by a spectrometer, which records the intensity of the light reaching the sensor for a variety of different wavelenghts. Because the digital micromirror array reflects all the wavelengths of interest, the spectrometer records measurements that are each dependent only on a single spectral band. Since the same patterns are acting as a modulation sequence to all the spectral bands in the datacube, the resulting measurement matrix is expressible as $\overline{\Phi} = \mathbf{I} \otimes \Phi$, where Φ is the matrix that contains the patterns programmed into the mirrors. Furthermore, it is possible to compress a hyperspectral datacube by using a hyperbolic wavelet basis, which itself is obtained as the Kronecker product of one-dimensional wavelet bases [96].

Our second example concerns the transform imager [99], an imaging hardware architecture that implements a separable CS matrix. A sketch of the transform imager is shown in Fig. 6. The

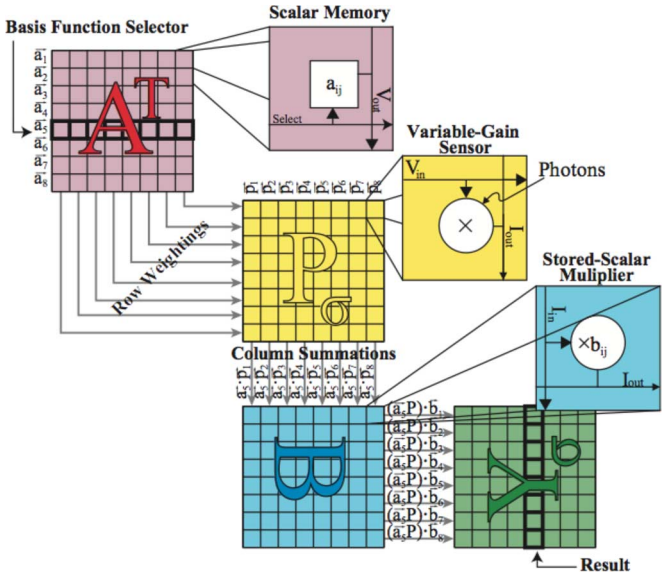


Fig. 6. Schematic for the compressive analog CMOS imager. The acquisition device effectively implements a separable CS matrix (taken from [99]).

image x is partitioned into blocks P_σ of size 16×16 to form a set of tiles; here σ indexes the block locations. The transform imager is designed to perform the multiplication $y_\sigma = A^T P_\sigma B$, where A and B are fixed matrices. This product of three matrices can be equivalently rewritten as $y_\sigma = (B \otimes A)^T \overline{P_\sigma}$, where $\overline{P_\sigma}$ denotes a column vector reshaping of the block P_σ [100]. The CS measurements are then obtained by randomly subsampling the vector y_σ. The compressive transform imager sets A and B to be noiselet bases and uses a 2-D undecimated wavelet transform to sparsify the image.

V. Structure in Finite-Dimensional Models

Until now we focused on structure in the measurement matrix Φ, and considered signals x with finite-dimensional sparse representations. We now turn to discuss how to exploit structure beyond sparsity in the input signal in order to reduce the number of measurements needed to faithfully represent it. Generalizing the notion of sparsity will allow us to move away from finite-dimensional models extending the ideas of CS to reduce sampling rates for infinite-dimensional continuous-time signals, which is the main goal of sampling theory.

We begin our discussion in this section within the finite-dimensional context by adding structure to the nonzero values of x. We will then turn, in Section VI, to more general, infinite-dimensional notions of structure, that can be applied to a broader class of analog signals.

A. Multiple Measurement Vectors

Historically, the first class of structure that has been considered within the CS framework has been that of multiple measurement vectors (MMVs) [101] which, similarly to sparse approximation, has been an area of interest in signal processing for more than a decade [21]. In this setting, rather than trying to recover a single sparse vector x, the goal is to jointly recover a set of vectors $\{x_i\}_{i=1}^{L}$ that share a common support. Stacking these vectors into the columns of a matrix $\mathbf{X}$, there will be at most K

nonzero rows in $\mathbf{X}$. That is, not only is each vector K-sparse, but the nonzero values occur on a common location set. We use the notation $\Omega = \mathrm{supp}(\mathbf{X})$ to denote the set of indices of the nonzero rows.

MMV problems appear quite naturally in many different applications areas. Early work on MMV algorithms focused on magnetoencephalography, which is a modality for imaging the brain [21], [102]. Similar ideas were also developed in the context of array processing [21], [103], [104], equalization of sparse communication channels [105], [106], and more recently cognitive radio and multiband communications [9], [86], [107]–[109].

1) Conditions on Measurement Matrices: As in standard CS, we assume that we are given measurements $\{y_i\}_{i=1}^{L}$, where each vector is of length $M < N$. Letting $\mathbf{Y}$ be the $M \times L$ matrix with columns y_i, our problem is to recover $\mathbf{X}$ assuming a known measurement matrix Φ so that $\mathbf{Y} = \Phi\mathbf{X}$. Clearly, we can apply any CS method to recover x_i from y_i as before. However, since the vectors $\{x_i\}$ all have a common support, we expect intuitively to improve the recovery ability by exploiting this joint information. In other words, we should in general be able to reduce the number of measurements ML needed to represent $\mathbf{X}$ below SL, where S is the number of measurements required to recover one vector x_i for a given matrix Φ.

Since $|\Omega| = K$, the rank of $\mathbf{X}$ satisfies $\mathrm{rank}(\mathbf{X}) \leq K$. When $\mathrm{rank}(\mathbf{X}) = 1$, all the sparse vectors x_i are multiples of each other, so that there is no advantage to their joint processing. However, when $\mathrm{rank}(\mathbf{X})$ is large, we expect to be able to exploit the diversity in its columns in order to benefit from joint recovery. This essential result is captured by the following necessary and sufficient uniqueness condition:

Theorem 18 [110]: *A necessary and sufficient condition for the measurements $\mathbf{Y} = \Phi\mathbf{X}$ to uniquely determine the jointly sparse matrix $\mathbf{X}$ is that*

$$|\mathrm{supp}(\mathbf{X})| < \frac{\mathrm{spark}(\Phi) - 1 + \mathrm{rank}(\mathbf{X})}{2}. \tag{31}$$

The sufficiency result was initially shown for the case $\mathrm{rank}(\mathbf{X}) = |\mathrm{supp}(\mathbf{X})|$ [111]. As shown in [110], [112], we can replace $\mathrm{rank}(\mathbf{X})$ by $\mathrm{rank}(\mathbf{Y})$ in (31). The sufficient direction of this condition was shown in [113] to hold even in the case where there are infinitely many vectors x_i. A direct consequence of Theorem 18 is that matrices $\mathbf{X}$ with larger rank can be recovered from fewer measurements. Alternatively, matrices $\mathbf{X}$ with larger support can be recovered from the same number of measurements. When $\mathrm{rank}(\mathbf{X}) = K$ and $\mathrm{spark}(\Phi)$ takes on its largest possible value equal to $M + 1$, condition (31) becomes $M \geq K + 1$. Therefore, in this best-case scenario, only $K + 1$ measurements per signal are needed to ensure uniqueness. This is much lower than the value of $2K$ obtained in standard CS via the spark (cf. Theorem 4), which we refer to here as the single measurement vector (SMV) setting. Furthermore, as we now show, in the noiseless setting $\mathbf{X}$ can be recovered by a simple algorithm, in contrast to the combinatorial complexity needed to solve the SMV problem from $2K$ measurements for general matrices Φ.

2) Recovery Algorithms: A variety of algorithms have been proposed that exploit the joint sparsity in different ways. As in the SMV setting, two main approaches to solving MMV problems are based on convex optimization and greedy methods. The analogue of (10) in the MMV case is

$$\widehat{\mathbf{X}} = \arg \min_{\mathbf{X} \in \mathbb{R}^{N \times L}} ||\mathbf{X}||_{0,q} \text{ subject to } \mathbf{Y} = \Phi\mathbf{X} \tag{32}$$

where we define the $\ell_{p,q}$ norms for matrices as

$$||\mathbf{X}||_{p,q} = \left(\sum_i ||x^i||_p^q \right)^{\frac{1}{q}}, \tag{33}$$

with x^i denoting the ith row of $\mathbf{X}$. With a slight abuse of notation, we also consider the quasi-norms with $p = 0$ such that $||\mathbf{X}||_{0,q} = |\mathrm{supp}(\mathbf{X})|$ for any q. Optimization based algorithms relax the ℓ_0 norm in (32) and attempt to recover $\mathbf{X}$ by mixed norm minimization [17], [112], [114]–[117]:

$$\widehat{\mathbf{X}} = \arg \min_{\mathbf{X} \in \mathbb{R}^{N \times L}} ||\mathbf{X}||_{p,q} \text{ subject to } \mathbf{Y} = \Phi\mathbf{X} \tag{34}$$

for some $p, q \geq 1$; values of $p, q = 1, 2$ and ∞ have been advocated.

The standard greedy approaches in the SMV setting have also been extended to the MMV case [101], [110], [114], [118]–[120]. The basic idea is to replace the residual vector r by a residual matrix $\mathbf{R}$, which contains the residuals with respect to each of the measurements, and to replace the surrogate vector $\Phi^T r$ by the q-norms of the rows of $\Phi^T\mathbf{R}$. For example, making these changes to OMP (Algorithm 1) yields a variant known as simultaneous orthogonal matching pursuit, shown as Algorithm 3, where $\mathbf{X}|_\Omega$ denotes the restriction of $\mathbf{X}$ to the rows indexed by Ω.

Algorithm 3: Simultaneous Orthogonal Matching Pursuit

Input: CS matrix Φ, MMV matrix $\mathbf{Y}$

Output: Row-sparse matrix $\widehat{\mathbf{X}}$

Initialize: $\widehat{\mathbf{X}}_0 = 0, \mathbf{R} = \mathbf{Y}, \Omega = \emptyset, i = 0$

while halting criterion false **do**

$i \leftarrow i + 1$

$b(n) \leftarrow ||\phi_n^T\mathbf{R}||_q, 1 \leq n \leq N$ {form residual matrix ℓ_q-norm vector}

$\Omega \leftarrow \Omega \cup \mathrm{supp}(\mathcal{T}(b, 1))$ {update row support with index of residual's row with largest magnitude}

$\widehat{\mathbf{X}}_i|_\Omega \leftarrow \Phi_\Omega^\dagger\mathbf{Y}, \widehat{\mathbf{X}}_i|_{\Omega^C} \leftarrow 0$ {update signal estimate}

$\mathbf{R} \leftarrow \mathbf{Y}$ {update measurement residual}

end while

return $\widehat{\mathbf{X}} \leftarrow \widehat{\mathbf{X}}_i$

An alternative MMV strategy is the ReMBo (reduce MMV and boost) algorithm [113]. ReMBo first reduces the problem to an SMV that preserves the sparsity pattern, and then recovers the signal support set; given the support, the measurements can be inverted to recover the input. The reduction is performed by merging the measurement columns with random coefficients.

The details of the approach together with a recovery guarantee are given in the following theorem.

Theorem 19: *Let $\overline{\mathbf{X}}$ be the unique K-sparse solution matrix of $\mathbf{Y} = \Phi\mathbf{X}$ and let Φ have spark greater than $2K$. Let $\mathbf{a}$ be a random vector with an absolutely continuous distribution and define the random vectors $\mathbf{y} = \mathbf{Y}\mathbf{a}$ and $\overline{\mathbf{x}} = \overline{\mathbf{X}}\mathbf{a}$. Consider the random SMV system $\mathbf{y} = \Phi\overline{\mathbf{x}}$. Then:*

1) *for every realization of $\mathbf{a}$, the vector $\overline{\mathbf{x}}$ is the unique K-sparse solution of the SMV;*
2) $\Omega(\overline{\mathbf{x}}) = \Omega(\overline{\mathbf{X}})$ *with probability one.*

According to Theorem 19, the MMV problem is first reduced to an SMV counterpart, for which the optimal solution $\overline{\mathbf{x}}$ is found. We then choose the support of $\overline{\mathbf{X}}$ to be equal to that of $\overline{\mathbf{x}}$, and invert the measurement vectors $\mathbf{Y}$ over this support. In practice, computationally efficient methods are used to solve the SMV counterpart, which can lead to recovery errors in the presence of noise, or when insufficient measurements are taken. By repeating the procedure with different choices of $\mathbf{a}$, the empirical recovery rate can be boosted significantly [113], and lead to superior performance over alternative MMV methods.

The MMV techniques discussed so far are *rank blind*, namely, they do not explicitly take the rank of $\mathbf{X}$, or that of $\mathbf{Y}$, into account. Theorem 18 highlights the role of the rank of $\mathbf{X}$ in the recovery guarantees. If $\mathrm{rank}(\mathbf{X}) = K$ and (31) is satisfied, then every K columns of Φ are linearly independent. This, in turn, means that $\mathcal{R}(\mathbf{Y}) = \mathcal{R}(\Phi_\Omega)$, where $\mathcal{R}(\mathbf{Y})$ denotes the column range of the matrix $\mathbf{Y}$. We can therefore identify the support of $\mathbf{X}$ by determining the columns ϕ_n that lie in $\mathcal{R}(\mathbf{Y})$. One way to accomplish this is by minimizing the norm of the projections onto the orthogonal complement of $\mathcal{R}(\mathbf{Y})$:

$$\min_n \|(\mathbf{I} - P_{\mathcal{R}}(\mathbf{Y}))\,\phi_n\|_2 \tag{35}$$

where $P_{\mathcal{R}}(\mathbf{Y})$ is the orthogonal projection onto the range of $\mathbf{Y}$. The objective in (35) is equal to zero if and only if $n \in \Omega$. Since, by assumption, the columns of Φ_Ω are linearly independent, once we find the support we can determine $\mathbf{X}$ as $\mathbf{X}_\Omega = \Phi_\Omega^\dagger\mathbf{Y}$. We can therefore formalize the following guarantee.

Theorem 20 [110]: *If $\mathrm{rank}(\mathbf{X}) = K$ and (31) holds, then the algorithm (35) is guaranteed to recover $\mathbf{X}$ from $\mathbf{Y}$ exactly.*

In the presence of noise, we choose the K values of n for which the expression (35) is minimized. Since (35) leverages the rank to achieve recovery, we say that this method is *rank aware*. More generally, any method whose performance improves with increasing rank will be termed rank aware. It turns out that it is surprisingly simple to modify existing greedy methods, such as OMP and thresholding, to be rank aware: instead of taking inner products with respect to $\mathbf{Y}$ or the residual $\mathbf{R}$, at each stage the inner products are computed with respect to an orthonormal basis $\mathbf{U}$ for the range of $\mathbf{Y}$ or $\mathbf{R}$ [110].

The criterion in (35) is similar in spirit to the MUSIC algorithm [121], popular in array signal processing, which also exploits the signal subspace properties. As we will see below in Section VI, array processing algorithms can be used to treat a variety of other structured analog sampling problems.

The MMV model can be further extended to include the case in which there are possibly infinitely many measurement vectors

$$\mathbf{y}(\lambda) = \Phi\mathbf{x}(\lambda), \quad \lambda \in \Lambda, \tag{36}$$

where Λ indicates an appropriate index set (which can be countable or uncountable). Here again the assumption is that the set of vectors $\{\mathbf{x}(\lambda), \lambda \in \Lambda\}$ all have common support Ω of size K. Such an infinite-set of equations is referred to as an infinite measurement vector (IMV) system. The common, finite, support set can be exploited in order to recover $\mathbf{x}(\lambda)$ efficiently by solving an MMV problem [113] Reduction to a finite MMV counterpart is performed via the continuous to finite (CTF) block, which aims at robust detection of Ω. The CTF builds a frame (or a basis) from the measurements using

$$\begin{aligned} \mathbf{y}(\lambda) \quad &\xrightarrow{\text{Frame construct}} \quad \mathbf{Q} = \sum_{\lambda\in\Lambda} \mathbf{y}(\lambda)\mathbf{y}^H(\lambda) \\ &\xrightarrow{\text{Decompose}} \quad \mathbf{Q} = \mathbf{V}\mathbf{V}^H. \end{aligned} \tag{37}$$

Typically, $\mathbf{Q}$ is constructed from roughly $2K$ snapshots $\mathbf{y}(\lambda)$, and the (optional) decomposition allows removal of the noise space [109]. Once the basis $\mathbf{V}$ is determined, the CTF solves an MMV system $\mathbf{V} = \Phi\mathbf{U}$ with $\mathrm{supp}(\mathbf{U}) \le K$. An alternative approach based on the MUSIC algorithm was suggested in [111] and [122].

The crux of the CTF is that the indices of the nonidentically zero rows of the matrix $\mathbf{U}$ that solves the finite underdetermined system $\mathbf{V} = \Phi\mathbf{U}$ coincide with the index set Ω that is associated with the infinite signal-set $\mathbf{x}(\lambda)$ [113], as incorporated in the following theorem.

Theorem 21 [109]: *Suppose that the system of (36) has a unique K-sparse solution set with support Ω, and that the matrix Φ satisfies (31). Let $\mathbf{V}$ be a matrix whose column span is equal to the span of $\{\mathbf{y}(\lambda), \lambda \in \Lambda\}$. Then, the linear system $\mathbf{V} = \Phi\mathbf{U}$ has a unique K-sparse solution with row support equal Ω.*

Once Ω is found, the IMV system reduces to $\mathbf{y}(\lambda) = \Phi_\Omega\mathbf{x}_\Omega(\lambda)$, which can be solved simply by computing the pseudo-inverse $\mathbf{x}_\Omega(\lambda) = \Phi_\Omega^\dagger\mathbf{y}(\lambda)$.

3) Performance Guarantees: In terms of theoretical guarantees, it can be shown that MMV extensions of SMV algorithms will recover $\mathbf{X}$ under similar conditions to the SMV setting in the worst-case scenario [17], [112], [119] so that theoretical equivalence results for arbitrary values of $\mathbf{X}$ do not predict any performance gain with joint sparsity. In practice, however, multichannel reconstruction techniques perform much better than recovering each channel individually. The reason for this discrepancy is that these results apply to all possible input signals, and are therefore worst-case measures. Clearly, if we input the same signal to each channel, namely when $\mathrm{rank}(\mathbf{X}) = 1$, no additional information on the joint support is provided from multiple measurements. However, as we have seen in Theorem 18, higher ranks of the input $\mathbf{X}$ improve the recovery ability. In particular, when $\mathrm{rank}(\mathbf{X}) = K$, rank-aware algorithms such as (35) recover the true value of $\mathbf{X}$ from the minimal number of measurements given in Theorem 18. This property is not shared by the other MMV methods.

Another way to improve performance guarantees is by posing a probability distribution on $\mathbf{X}$ and developing conditions under which $\mathbf{X}$ is recovered with high probability [101], [118], [119], [123]. Average case analysis can be used to show that fewer measurements are needed in order to recover $\mathbf{X}$ exactly [119].

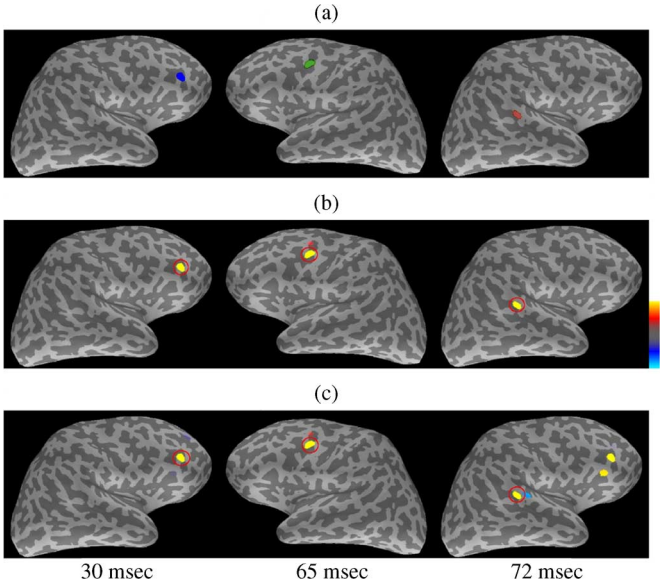


Fig. 7. Example performance of the MMV model for EEG inversion with temporal regularization. The setup consists of a configuration of 306 sensors recording simulated brain activity consisting of three vertices peaking at distinct times over a 120-ms period. The figures show (a) the ground truth EEG activity, (b) the inversion obtained by applying MMV recovery on all the data recorded by the sensors, and (c) the inversion obtained by independently applying sparse recovery on the data measured at each particular time instance. In each case, we show three representative time instances (30, 65, and 72 ms, respectively). The results show that MMV is able to exclude spurious activity detections successfully through implicit temporal regularization (taken from [124]).

Theorem 22 [119]: *Let* $\mathbf{X} \in \mathbb{R}^{N\times L}$ *be drawn from a probabilistic model in which the indices for its* K *nonzero rows are drawn uniformly at random from the* $\binom{N}{K}$ *possibilities and its nonzero rows (when concatenated) are given by* $\Sigma\Delta$*, where* Σ *is an arbitrary diagonal matrix and each entry of* Δ *is an i.i.d. standard Gaussian random variable. If* $K \leq \min\left(\frac{C_1}{\mu^2(\Phi)}, \frac{C_2 N}{\|\Phi\|^2}\right)$ *then* $\mathbf{X}$ *can be recovered exactly from* $\mathbf{Y} = \Phi\mathbf{X}$ *via (34), with* $p = 2$ *and* $q = 1$*, with high probability.*

In a similar fashion to the SMV case (cf. Theorem 13), while worst-case results limit the sparsity level to $K = \mathcal{O}(\sqrt{M})$, average-case analysis shows that sparsity up to order $K = \mathcal{O}(M)$ may enable recovery with high probability. Moreover, under a mild condition on the sparsity and on the matrix Φ, the failure probability decays exponentially in the number of channels L [119].

4) Applications: The MMV model has found several applications in the applied CS literature. One example is in the context of electroencephalography and magnetoencephalography (EEG/MEG) [21], [102]. As mentioned earlier, sparsity-promoting inversion algorithms have been popular in EEG/MEG literature due to their ability to accurately localize electromagnetic source signals. It is also possible to further improve estimation performance by introducing temporal regularization when a sequence of EEG/MEG is available. For example, one may apply the MMV model on the measurements obtained over a coherent time period, effectively enforcing temporal regularity on the brain activity [124]. Such temporal regularization can correct estimation errors that appear as temporal "spikes" in EEG/MEG activity. The example in Fig. 7 shows a test MEG activation signal with three active vertices peaking at separate time instances. A 306-sensor acquisition configuration was simulated with $\mathrm{SNR} = 3$ dB. The performance of MEG inversion with independent recovery of each time instance exhibits spurious activity detections that are removed by the temporal regularization enforced by the MMV model. Additionally, the accuracy of the temporal behavior for each vertex is improved; see [124] for details.

MMV models are also used during CS recovery for certain infinite-dimensional signal models [18]. We will discuss this application in more detail in Section VI-B.

B. Unions of Subspaces

To introduce more general models of structure on the input signal, we turn now to extend the notion of sparsity to a much broader class of signals which can incorporate both finite-dimensional and infinite-dimensional signal representations. The enabling property that allows recovery of a sparse vector $x \in \Sigma_K$ from $M < N$ measurements is the fact that the set Σ_K corresponds to a union of K-dimensional subspaces within $\mathbb{R}^N$. More specifically, if we know the K nonzero locations of x, then we can represent x by a K-dimensional vector of coefficients and therefore only K measurements are needed in order to recover it. Therefore, for each *fixed* location set, x is restricted to a K-dimensional subspace, which we denote by $\mathcal{U}_i$. Since the location set is unknown, there are $\binom{N}{K}$ possible subspaces $\mathcal{U}_i$ in which x may reside. Mathematically, this means that sparse signals can be represented by a *union* of subspaces [15]:

$$x \in \mathcal{U} = \bigcup_{i=1}^{m} \mathcal{U}_i \tag{38}$$

where each subspace $\mathcal{U}_i$, $1 \leq i \leq m$, is a K-dimensional subspace associated with a specific set of K nonzero values.

For canonically sparse signals, the union Σ_K is composed of canonical subspaces $\mathcal{U}_i$ that are aligned with K out of the N coordinate axes of $\mathbb{R}^N$. Allowing for more general choices of $\mathcal{U}_i$ leads to powerful representations that accommodate many interesting signal priors. It is important to note that union models are not closed under linear operations: The sum of two signals from a union $\mathcal{U}$ is generally no longer in $\mathcal{U}$. This nonlinear behavior of the signal set renders sampling and recovery more intricate. To date, there is no general methodology to treat all unions in a unified manner. Therefore, we focus our attention on some specific classes of union models, in order of complexity.

The simplest class of unions result when the number of subspaces comprising the union is finite, and each subspace has finite dimensions. We call this setup a finite union of subspaces (FUS) model. Within this class, we consider below two special cases:

- *Structured sparse supports*: This class consists of sparse vectors that meet additional restrictions on the support (i.e., the set of indices for the vector's nonzero entries). This

corresponds to only certain subspaces $\mathcal{U}_i$ out of the $\binom{N}{K}$ subspaces present in Σ_K being allowed [20].[4]

- *Sparse sums of subspaces* where each subspace $\mathcal{U}_i$ comprising the union is a direct sum of K low-dimensional subspaces [17]

$$\mathcal{U}_i = \bigoplus_{|j|=K} \mathcal{A}_j. \tag{39}$$

Here, $\{\mathcal{A}_j, 1 \leq j \leq m\}$ are a given set of subspaces with dimensions $\dim(\mathcal{A}_j) = d_j$, and $|j| = K$ denotes a sum over K indices. Thus, each subspace $\mathcal{U}_i$ corresponds to a different choice of K subspaces $\mathcal{A}_j$ that comprise the sum. The dimensionality of the signal representation in this case will be $N = \sum_{j=1}^{m} d_j$; for simplicity, we will often let $d_j = d$ for all j so that $N = dm$. As we show, this model leads to block sparsity in which certain blocks in a vector are zero, and others are not [120], [133], [134]. This framework can model standard sparsity by letting $\mathcal{A}_j$, $j = 1, \ldots, N$ be the one-dimensional subspace spanned by the j^{th} canonical vector.

The two FUS cases above can be combined to allow only certain sums of K subspaces to be part of the union $\mathcal{U}$.

More complicated is the setting in which the number of possibilities is still finite while each underlying subspace has infinite dimensions. Finally, there may be infinitely many possibilities of finite or infinite subspaces. These last two classes allow us to treat different families of analog signals, and will be considered in more detail in Section VI.

1) Conditions on Measurement Matrices: Guarantees for signal recovery using a FUS model can be developed by extending tools used in the standard sparse setting. For example, the $(\mathcal{U}, \delta)$-RIP for FUS models [15]–[17], [20] is similar to the standard RIP where instead of the inequalities in (7) having to be satisfied for all sparse vectors $x \in \Sigma_K$, they have to hold only for vectors $x \in \mathcal{U}$. If the constant δ is small enough, then it can be shown that recovery algorithms tailored to the FUS model will recover the true underlying vector x.

An important question is how many samples are needed roughly in order to guarantee stable recovery. This question is addressed in the following proposition [16], [17], [20], [46].

Proposition 1 [46, Theorem 3.3]: *Consider a matrix Φ of size $M \times N$ with entries independently drawn from a sub-Gaussian distribution, and let $\mathcal{U}$ be composed of L subspaces of dimension D. Let $t > 0$ and $0 < \delta < 1$ be constant numbers. If*

$$M \geq \frac{36}{7\delta}\left(\ln(L) + D\ln\left(\frac{12}{\delta}\right) + t\right), \tag{40}$$

then Φ satisfies the $(\mathcal{U}, \delta)$-RIP with probability at least $1 - e^{-t}$.

As observed in [46], the first term in (40) has the dominant impact on the required number of measurements for sparse signals in an asymptotic sense. This term quantifies the amount of measurements that are needed to code the exact subspace where the sparse signal resides. We now specialize this result to the two FUS classes given earlier.

- In the structured sparse supports case, L corresponds to the number of distinct K-sparse supports allowed by the constraints, with $L \leq \binom{N}{K}$; this implies a reduction in the number of measurements needed as compared to the traditional sparsity model. Additionally, $D = K$, since each subspace has dimension K.
- For the sparse sum of subspaces setting, we focus on the case where each $\mathcal{A}_j$ is of dimension d; we then have $L = \binom{\frac{N}{d}}{K}$. Using the approximation

$$\left(\frac{N}{dK}\right)^K \leq L = \binom{\frac{N}{d}}{K} \leq \left(\frac{eN}{dK}\right)^K \tag{41}$$

we conclude that for a given fraction of nonzeros $r = \frac{dK}{N}$, roughly $M \approx K\log\left(\frac{N}{dK}\right) = -K\log(r)$ measurements are needed. For comparison, to satisfy the standard RIP a larger number $n \approx -Kd\log(r)$ is required. The FUS model reduces the total number of subspaces and therefore requires d times less measurements to code the signal subspace. The subspace dimension $D = Kd$ equals the number of degrees of freedom in x.

Since the number of nonzeros is the same regardless of the sparsity structure, the term $D = \mathcal{O}(K)$ is not reduced in either setting.

There also exists an extension of the coherence property to the sparse sum of subspaces case [120]. We must first elaborate on the notation for this setup. Given a signal $x \in \mathcal{U}$ our goal is to recover it from measurements $y = Ax$ where A is an appropriate measurement matrix of size $M \times N$. To recover x from y we exploit the fact that any $x \in \mathcal{U}$ can be represented in an appropriate basis as a block-sparse vector [135]. This follows by first choosing a basis Ψ_j for each of the subspaces $\mathcal{A}_j$. Then, any $x \in \mathcal{U}$ can be expressed as

$$x = \sum_{|j|=K} \Psi_j \theta[j], \tag{42}$$

following the notation of (39), where $\theta[j] \in \mathbb{R}^{d_j}$ are the representation coefficients in Ψ_j. Let Ψ be the column concatenation of Ψ_j, and denote by $\theta[j]$ the jth subblock of a length-N vector θ over $\mathcal{I} = \{d_1, \ldots, d_m\}$. The jth subblock is of length d_j, and the blocks are formed sequentially so that

$$\theta^T = [\underbrace{\theta_1 \ \ldots \ \theta_{d_1}}_{\theta[1]} \ldots \underbrace{\theta_{N-d_m+1} \ \ldots \ \theta_N}_{\theta[m]}]^T. \tag{43}$$

If for a given x the jth subspace $\mathcal{A}_j$ does not appear in the sum (39), then $\theta[j] = 0$. Finally, we can write $x = \Psi\theta$, where there are at most K nonzero blocks $\theta[i]$. Consequently, our union model is equivalent to the model in which x is represented by a block-sparse vector θ. The CS measurements can then be written as $y = Ax = A\Psi\theta = \Phi\theta$, where we denoted $\Phi = A\Psi$. An example of a block-sparse vector with $k = 2$ is depicted in Fig. 8. When $d_i = 1$ for all i, block sparsity reduces to conventional sparsity in the dictionary Ψ. We will say that a vector $\theta \in \mathbb{R}^N$ is block K-sparse if $\theta[i]$ is nonzero for at most K indices i.

[4]While the description of structured sparsity given via unions of subspaces is deterministic, there exists many different CS recovery approaches that leverage probabilistic models designed to promote structured sparsity, such as Markov random fields, hidden Markov Trees, etc. [125]–[131], as well as deterministic approaches that rely on structured sparsity-inducing norm minimization [17], [132].

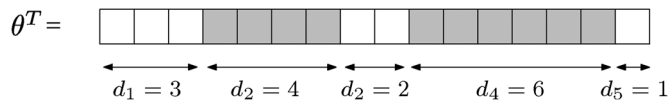


Fig. 8. A block-sparse vector θ over $\mathcal{I} = \{d_1, \ldots, d_5\}$. The gray areas represent ten nonzero entries that occupy two blocks.

The block-sparse model we present here has also been studied in the statistical literature, where the objective is often quite different. Examples include group selection consistency [136], [137], asymptotic prediction properties [137], [138], and block sparsity for logistic regression [139]. Block sparsity models are also interesting in their own right (namely, not only as an equivalence with an underlying union), and appear naturally in several problems. Examples include DNA microarray analysis [140], [141], equalization of sparse communication channels [105], and source localization [104].

We now introduce the adaptation of coherence to sparse sums of subspaces: the *block-coherence* of a matrix Φ is defined as

$$\mu_{\mathrm{B}}(\Phi) = \max_{\ell, r \neq \ell} \frac{1}{d} \rho\left(\Phi^H[\ell]\Phi[r]\right) \tag{44}$$

with $\rho(A)$ denoting the spectral norm of the matrix A. Here Φ is represented as a concatenation of column-blocks $\Phi[\ell]$ of size $M \times d$:

$$\Phi = [\underbrace{\phi_1 \ \ldots \ \phi_d}_{\Phi[1]} \underbrace{\phi_{d+1} \ \ldots \ \phi_{2d}}_{\Phi[2]} \cdots \underbrace{\phi_{N-d+1} \ \ldots \ \phi_N}_{\Phi[m]}]. \tag{45}$$

When $d = 1$, as expected, $\mu_{\mathrm{B}}(\Phi) = \mu(\Phi)$. More generally, $\mu_{\mathrm{B}}(\Phi) \leq \mu(\Phi)$.

While $\mu_{\mathrm{B}}(\Phi)$ quantifies global properties of the matrix Φ, local properties are characterized by the *subcoherence* of Φ, defined as

$$\nu(\Phi) = \max_{\ell} \max_{i,j \neq i} \left|\phi_i^H \phi_j\right|, \quad \phi_i, \phi_j \in \Phi[\ell]. \tag{46}$$

We define $\nu(\Phi) = 0$ for $d = 1$. In addition, if the columns of $\Phi[\ell]$ are orthogonal for each ℓ, then $\nu(\Phi) = 0$.

2) Recovery Algorithms: Like in the MMV setting, it is possible to extend standard sparsity-based signal recovery algorithms to the FUS model. For example, greedy algorithms may be modified easily by changing the thresholding $\mathcal{T}(x, K)$ (which finds the best approximation of x in the union of subspaces Σ_K) to a structured sparse approximation step:

$$\mathbb{M}_{\mathcal{U}}(x) = \arg\min_{x' \in \mathcal{U}} \|x - x'\|_2. \tag{47}$$

For example, the CoSaMP algorithm (see Algorithm 2) is modified according to the FUS model [20] by changing the following two steps:

1) prune residual: $\Omega \leftarrow \mathrm{supp}(\mathbb{M}_{\mathcal{U}_2}(e))$;
2) prune signal: $\widehat{x}_i \leftarrow \mathbb{M}_{\mathcal{U}}(b)$.

A similar change can be made to the IHT algorithm (15) to obtain a model-based IHT variant:

$$\widehat{x}_i = \mathbb{M}_{\mathcal{U}}\left(\widehat{x}_{i-1} + \Phi^T(y - \Phi\widehat{x}_{i-1})\right).$$

Structured sparse approximation algorithms of the form (47) are feasible and computationally efficient for a variety of structured sparse support models [20], [142]–[144]. For example, the approximation algorithm $\mathbb{M}_{\mathcal{U}}$ under block sparsity is equivalent to block thresholding, with the K blocks with the largest energies (or ℓ_2 norms) being preserved; we will show another example in greater detail later in this section.

For the sparse sum of subspaces setting, it is possible to formulate optimization-based algorithms for signal recovery. A convex recovery method can be obtained by minimizing the sum of the energy of the blocks $\theta[i]$. To write down the problem explicitly, we define the mixed $\frac{\ell_2}{\ell_1}$ norm as

$$\|\theta\|_{2,\mathcal{I}} = \sum_{i=1}^{m} \|\theta[i]\|_2. \tag{48}$$

We may then recover θ by solving [17], [134], [136], [137]

$$\widehat{\theta} = \arg\min_{\theta \in \mathbb{R}^N} \|\theta\|_{2,\mathcal{I}} \text{ subject to } y = \Phi\theta. \tag{49}$$

The optimization constraints can be relaxed to address the case of noisy measurements, in a way similar to the standard BPIC algorithm. Generalizations of greedy algorithms to the block sparse setting have been developed in [101] and [120].

3) Recovery Guarantees: Many recovery methods for the FUS model inherit the guarantees of their standard counterparts. Our first example deals with the model-based CoSaMP algorithm. Since CoSaMP requires RIP of order $4K$, here we must rely on enlarged unions of subspaces.

Definition 6: *For an integer J and a FUS model $\mathcal{U}$, denote the J-sum of $\mathcal{U}$ as the sum of subspaces*

$$S_J(\mathcal{U}) = \bigoplus_{|j|=J} \mathcal{U}_j, \tag{50}$$

following the notation of (39), which contains all sums of J signals belonging in U.

We can then pose the following guarantee.

Theorem 23 [20]: *Let $x \in \mathcal{U}$ and let $y = \Phi x + n$ be a set of noisy CS measurements. If Φ has the $(S_4(\mathcal{U}), \delta)$-RIP with $\delta \leq 0.1$, then the signal estimate $\widehat{x}_i$ obtained from iteration i of the model-based CoSaMP algorithm satisfies*

$$\|x - \widehat{x}_i\|_2 \leq 2^{-i}\|x\|_2 + 15\|n\|_2. \tag{51}$$

One can also show that under an additional condition on Φ the algorithm is stable to signal mismodeling [20]. Similar guarantees exist for the model-based IHT algorithm [20].

Guarantees are also available for the optimization-based approach used in the sparse sum of subspaces setting.

Theorem 24 [17]: *Let $x \in \mathbb{R}^N$ and let $y = \Phi x + n$ be a set of noisy CS measurements, with $\|n\|_2 \leq \epsilon$. If Φ has the $(S_2(\mathcal{U}), \delta)$-RIP with $\delta \leq \sqrt{2} - 1$, then the signal estimate $\widehat{x}$ obtained from (49) with relaxed constraints $\|y - \Phi x\|_2 \leq \epsilon$ satisfies*

$$\|x - \widehat{x}\|_2 \leq C_1 K^{-\frac{1}{2}} \|x - \mathbb{M}_{\mathcal{U}}(x)\|_2 + C_2\epsilon \tag{52}$$

where $\mathbb{M}_{\mathcal{U}}(x)$ is defined in (47), and $\mathcal{U}$ is a sparse sum of subspaces.

Finally, we point out that recovery guarantees can also be obtained based on block coherence.

Fig. 9. The class of tree-sparse signals is an example of a FUS that enforces structure in the sparse signal support. (a) Example of a tree-structured sparse support for a 1-D piecewise smooth signal's wavelet coefficients (taken from [20]). (b) Example recovery of the $N = 512 \times 512 = 262\,144$-pixel *Peppers* test image from $M = 40\,000$ random measurements (15%) using standard CoSaMP (SNR =17.45 dB). (c) Example recovery of the same image from the same measurements using model-based CoSaMP for the tree-structured FUS model (SNR = 22.6 dB). In both cases, the Daubechies-8 wavelet basis was used as the sparsity/compressibility transform.

Theorem 25 [120]: *Both the greedy methods and the optimization-based approach of (49) recover a block-sparse vector θ from measurements $y = \Phi\theta$ if the block-coherence satisfies*

$$Kd < \frac{1}{2}\left(\frac{1}{\mu_{\mathrm{B}}(\Phi)} + d - (d-1)\frac{\nu(\Phi)}{\mu_{\mathrm{B}}(\Phi)}\right). \tag{53}$$

In the special case in which the columns of $\Phi[\ell]$ are orthonormal for each ℓ, we have $\nu(\Phi) = 0$ and therefore the recovery condition becomes $Kd < \frac{(\mu_{\mathrm{B}}^{-1}(\Phi)+d)}{2}$. Comparing to Theorem 4 for recovery of a conventional Kd-sparse signal we can see that, thanks to $\mu_{\mathrm{B}}(\Phi) \leq \mu(\Phi)$, making explicit use of block-sparsity leads to guaranteed recovery for a potentially higher sparsity level. These results can also be extended to both adversarial and random additive noise in the measurements [145].

4) Applications: A particularly interesting example of structured sparse supports corresponds to tree-structured supports [20]. For signals that are smooth or piecewise smooth, including natural images, sufficiently smooth wavelet bases provide sparse or compressible representations θ. Additionally, because each wavelet basis element acts as a discontinuity detector in a local region of the signal at a certain scale, it is possible to link together wavelet basis elements corresponding to a given location at neighboring scales, forming what are known as branches that connect at the coarsest wavelet scale. The corresponding full graph is known as a wavelet tree. For locations where there exist discontinuities, the corresponding branches of wavelet coefficients tend to be large, forming a connected subtree inside the wavelet tree. Thus, the restriction of $\mathcal{U} \subseteq \Sigma_K$ includes only the subspaces corresponding to this type of structure in the signal's support. Fig. 9(a) shows an example of a tree-sparse 1-D signal support for a wavelet coefficient vector. Since the number of possible supports containing this structure is limited to $L \leq \frac{C^K}{K}$ for a constant C, we obtain that $M = \mathcal{O}(K)$ random measurements are needed to recover these signals using model-based recovery algorithms. Additionally, there exist approximation algorithms to implement $\mathbb{M}$ for this FUS model based on both greedy and optimization-based approaches; see [20] for more details. Fig. 9(b) shows an example of improved signal recovery from random measurements leveraging the FUS model.

In our development so far, we did not consider any structure within the subspaces comprising the unions. In certain applications it may be beneficial to add internal structure. For example, the coefficient vectors $\theta[j]$ may themselves be sparse. Such scenarios can be accounted for by adding an ℓ_1 penalty in (49) on the individual blocks [146], an approach that is dubbed C-HiLasso in [147]. This allows to combine the sparsity-inducing property of ℓ_1 optimization at the individual feature level, with the block-sparsity property of (49) on the group level, obtaining a hierarchically structured sparsity pattern. An example FUS model featuring sparse sums of subspaces that exploits the additional structure described above is shown in Fig. 10. This example is based on identification of digits; a separate subspace is trained for each digit 0, ..., 9, and the task addressed is separation of a mixture of digits from subsampled information. We collect bases that span each of the 10 subspaces $\mathcal{A}_0, \ldots, \mathcal{A}_9$ into a dictionary Φ, and apply the FUS model where the subspaces $\mathcal{U}_i$ considered correspond to mixtures of pairs of digits. The FUS model allows for successful source identification and separation.

VI. Structure in Infinite-Dimensional Models

One of the prime goals of CS is to allow for reduced-rate sampling of analog signals. In fact, many of the original papers in the field state this as a motivating drive for the theory of CS. In this section we focus on union models that include a degree of infiniteness: This can come into play either by allowing for infinitely many subspaces, by letting the subspaces have infinite dimension, or both. As we will show, the resulting models may be used to describe a broad class of structured continuous-time signals. We will then demonstrate how such priors can be translated into concrete hardware solutions that allow sampling and recovery of analog signals at rates far below that dictated by Nyquist.

The approach we present here to reduced-rate sampling is based on viewing analog signals in unions of subspaces, and is therefore fundamentally different than previous attempts to

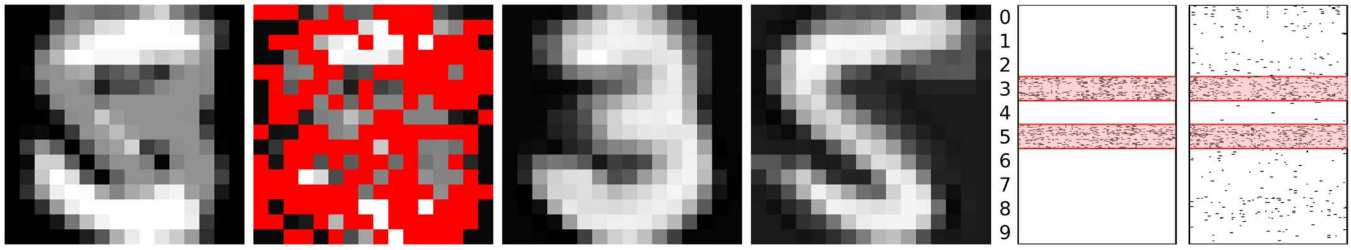


Fig. 10. Example of recovered digits (3 and 5) from a mixture with 60% of missing components. From left to right: noiseless mixture, observed mixture with missing pixels highlighted in red, recovered digits 3 and 5, and active set recovered for 180 different mixtures using the C-HiLasso and (13) respectively. The last two figures show the active sets of the recovered coefficients vectors θ as a matrix (one column per mixture), where black dots indicate nonzero coefficients. The coefficients corresponding to the subspace bases for digits 3 and 5 are marked as pink bands. Notice that C-HiLasso efficiently exploits the FUS model, succeeding in recovering the correct active groups in all the samples. The standard approach (13), which lacks this stronger signal model, clearly is not capable of doing so, and active sets spread all over the groups (taken from [147]).

treat similar problems [42], [84], [104], [148], [149]. The latter typically rely on discretization of the analog input and are often not concerned with the actual hardware. Furthermore, in some cases the reduced rate analog stage results in high rate DSP. In contrast, in all of the examples below we treat the analog formulation directly, the DSP can be performed in real time at the low rate, and the analog front end can be implemented in hardware using standard analog design components such as modulators, low rate ADCs and lowpass filters (LPFs). A more detailed discussion on the comparison to discretized approaches can be found in [10] and [150].

In our review, we consider three main cases:

1) finite unions of infinite dimensional spaces;
2) infinite unions of finite dimensional spaces;
3) infinite unions of infinite dimensional spaces.

In each one of the three settings above there is an element that can take on infinite values. We present general theory and results behind each of these cases, and focus in additional detail on a representative example application for each class.

Before describing the three cases, we first briefly introduce the notion of sampling in shift-invariant (SI) subspaces, which plays a key role in the development of standard (subspace) sampling theory [135], [151]. We then discuss how to incorporate structure into SI settings, leading to the union classes outlined above.

A. Shift-Invariant Spaces for Analog Signals

A signal class that plays an important role in sampling theory are signals in SI spaces [151]–[154]. Such signals are characterized by a set of generators $\{h_\ell(t), 1 \le \ell \le N\}$ where in principle N can be finite or infinite (as is the case in Gabor or wavelet expansions of L_2). Here, we focus on the case in which N is finite. Any signal in such a SI space can be written as

$$x(t) = \sum_{\ell=1}^{N} \sum_{n \in \mathbb{Z}} d_\ell[n] h_\ell(t - nT), \tag{54}$$

for some set of sequences $\{d_\ell[n] \in \ell_2, 1 \le \ell \le N\}$ and period T. This model encompasses many signals used in communication and signal processing including bandlimited functions, splines [151], multiband signals [108], [109], [155], [156], and pulse amplitude modulation signals.

The subspace of signals described by (54) has infinite dimensions, since every signal is associated with infinitely many coefficients $\{d_\ell[n], 1 \le \ell \le N\}$. Any such signal can be recovered from samples at a rate of $\frac{N}{T}$; one possible sampling paradigm at the minimal rate is given in Fig. 11. Here, $x(t)$ is filtered with a bank of N filters, each with impulse response $s_\ell(t)$ which can be almost arbitrary, and the outputs are uniformly sampled with period T. Denote by $\mathbf{c}(\omega)$ a vector collecting the frequency responses of $c_\ell[n]$, $1 \le \ell \le N$. The signal is then recovered by first processing the samples with a filter bank with frequency response $\mathbf{G}(e^{j\omega})$, which depends on the sampling filters and the generators $h_\ell(t)$ (see [18] for details). In this way we obtain the vectors

$$\mathbf{d}(\omega) = \mathbf{G}(e^{j\omega})\mathbf{c}(\omega) \tag{55}$$

containing the frequency responses of the sequences $d_\ell[n]$, $1 \le \ell \le N$. Each output sequence is then modulated by a periodic impulse train $\sum_{n \in \mathbb{Z}} \delta(t - nT)$ with period T, followed by filtering with the corresponding analog filter $h_\ell(t)$.

In the ensuing subsections we consider settings in which further structure is incorporated into the generic SI model (54). In Sections VI-B and VI-D, we treat signals of the form (54) involving a small number K of generators, chosen from a finite or infinite set, respectively, while in Section VI-C, we consider a finite-dimensional counterpart of (54) in which the generators are chosen from an infinite set. All of these examples lead to union of subspaces models for analog signals. Our goal is to exploit the available structure in order to reduce the sampling rate.

Before presenting the more detailed applications we point out that in all the examples below the philosophy is similar: we develop an analog sensing stage that consists of simple hardware devices designed to spread out (alias) the signal prior to sampling, in such a way that the samples contain energy from all subspace components. The first step in the digital reconstruction stage identifies the underlying subspace structure. Recovery is then performed in a subspace once the parameters defining the subspace are determined. The difference between the examples is in how the aliasing is obtained and in the digital recovery step that identifies the subspace. This framework has been dubbed Xampling in [9] and [10], which combines CS and sampling, emphasizing that this is an analog counterpart to discrete CS theory and methods.

B. Finite Union of Infinite-Dimensional Subspaces

In this first case, we follow the model (38), (39) where $\mathcal{U}$ is composed of a finite number m of subspaces, and each subspace has infinite dimension (i.e., $d_j = \infty$ for $1 \leq j \leq m$).

1) Analog Signal Model: To incorporate structure into (54) we proceed in two ways (see Section VI-D for the complement). In the first, we assume that only K of the N generators are active, leading to the model

$$x(t) = \sum_{|\ell|=K} \sum_{n \in \mathbb{Z}} d_\ell[n] h_\ell(t - nT) \tag{56}$$

where the notation $|\ell| = K$ means a union (or sum) over at most K elements. If the K active generators are known, then it suffices to sample at a rate of $\frac{K}{T}$ corresponding to uniform samples with period T at the output of K appropriate filters. However, a more difficult question is whether the rate can be reduced if we know that only K of the generators are active but do not know in advance which ones. This is a special case of the model (39) where now each of the subspaces $\mathcal{A}_\ell$ is a single-generator SI subspace spanned by $h_\ell(t)$. For this model, it is possible to reduce the sampling rate to as low as $\frac{2K}{T}$ [18]. Such rate reduction is achieved by using a bank of appropriate sampling filters that replaces the filters $s_\ell(t)$ in Fig. 11, followed by postprocessing via subspace reduction and solving an MMV problem (cf. Section V-A). We now describe the main ideas underlying this approach.

2) Compressive Signal Acquisition Scheme: A block diagram of the basic architecture is very similar to the one given in Fig. 11. We simply change the N sampling filters $s_1(-t), \ldots, s_N(-t)$ to $p < N$ sampling filters $w_1(t), \ldots, w_p(t)$ and replace the filter bank $\mathbf{G}(e^{j\omega})$ with the CTF block with real-time recovery (cf. Section V-A-2). The design of the filters $w_\ell(t)$, $1 \leq \ell \leq p$ relies on two main ingredients:

1) a $p \times N$ matrix $\mathbf{\Phi}$ chosen such that it solves a discrete MMV problem with sparsity K;
2) a set of functions $\{s_\ell(t), 1 \leq \ell \leq N\}$ which can be used to sample and reconstruct the entire set of generators $\{h_\ell(t), 1 \leq \ell \leq N\}$ according to the Nyquist-rate scheme of Fig. 11.

Based on these ingredients, the compressive sampling filters $w_\ell(t)$ consist of linear combinations of $s_\ell(t)$, with coefficients that depend on the matrix $\mathbf{\Phi}$ through [18]

$$\mathbf{w}(\omega) = \mathbf{M}^*(e^{j\omega T}) \mathbf{\Phi}^* \mathbf{G}^*(e^{j\omega T}) \mathbf{h}(\omega) \tag{57}$$

where $(\cdot)^*$ denotes the conjugate, $\mathbf{w}(\omega), \mathbf{h}(\omega)$ concatenate the frequency responses of $w_\ell(t)$ $(1 \leq \ell \leq p)$ and $h_\ell(t)$ $(1 \leq \ell \leq N)$, respectively, and $\mathbf{M}(e^{j\omega T})$ is a $p \times p$ arbitrary invertible matrix representing the discrete-time Fourier transform (DTFT) of a bank of filters. Since this matrix can be chosen arbitrarily, it allows for freedom in selecting the sampling filters.

3) Reconstruction Algorithm: Directly manipulating the expression for the sampled sequences leads to a simple relationship between the measurements $y_\ell[n]$ and the unknown sequences $d_\ell[n]$ [18]:

$$\mathbf{y}[n] = \mathbf{\Phi} \mathbf{d}[n], \quad ||\mathbf{d}[n]||_0 \leq K \tag{58}$$

where the vector $\mathbf{y}[n] = [y_1[n], \cdots, y_p[n]]^T$ collects the measurements at $t = nT$ and the vector $\mathbf{d}[n] = [d_1[n], \cdots, d_N[n]]^T$ collects the unknown generator coefficients for time period n. Since only K out of the N sequences $d_\ell[n]$ are identically nonzero by assumption, the vectors $\{\mathbf{d}[n]\}$ are jointly K-sparse. Equation (58) is valid when $\mathbf{M}(e^{j\omega T}) = \mathbf{I}$ for all ω; otherwise, the samples must first be pre-filtered with the inverse filter bank to obtain (58). Therefore, by properly choosing the sampling filters, we have reduced the recovery problem to an IMV problem, as studied in Section V-A-2. To recover $\mathbf{d}[n]$, we therefore rely on the CTF and then reconstruct $x(t)$ by interpolating $d_\ell[n]$ with their generators $h_\ell(t)$.

4) Recovery Guarantees: From Theorem 21, it follows that $p = 2K$ filters are needed in order to ensure recovery for all possible input signals. In practice, since polynomial-time algorithms will be used to solve the equivalent MMV, we will need to increase p slightly beyond $2K$.

Theorem 26 [18]: *Consider a signal of the form (56). Let $w_\ell(t)$, $1 \leq \ell \leq p$ be a set of sampling filters defined by (57) where $\mathbf{\Phi}$ is a size $p \times N$ matrix. Then, the signal $x(t)$ can be recovered exactly using the scheme of Fig. 11 as long as $\mathbf{\Phi}$ allows the solution of an MMV system of size $p \times N$ with sparsity K.*

Since our approach to recovering $x(t)$ relies on solving an MMV system, the quality of the reconstruction depends directly on the properties of the underlying MMV problem. Therefore, results regarding noise, mismodeling, and suboptimal recovery using polynomial techniques developed in the MMV context can immediately be adapted to this setting.

5) Example Application: We now describe an application of the general analog CS architecture of Fig. 11: sampling of multiband signals at sub-Nyquist rates. We also expand on practical alternatives to this system, which reduce the hardware complexity.

The class of multiband signals models a scenario in which $x(t)$ consists of several concurrent radio-frequency (RF) transmissions. A receiver that intercepts a multiband $x(t)$ sees the typical spectral support that is depicted in Fig. 12. We assume that the signal contains at most N (symmetric) frequency bands with carriers f_i, each of maximal width B. The carriers are limited to a maximal frequency of $f_{\max}$. When the carrier frequencies f_i are fixed, the resulting signal model can be described as a subspace, and standard demodulation techniques may be used to sample each of the bands at low rate. A more challenging scenario is when the carriers f_i are unknown. This situation arises, for example, in spectrum sensing for mobile cognitive radio (CR) receivers [157], which aim at utilizing unused frequency regions on an opportunistic basis.

The Nyquist rate associated with $x(t)$ is $f_{\mathrm{NYQ}} = 2f_{\max}$, which can be quite high in modern applications—on the order of several GHz. On the other hand, by exploiting the multiband structure, it can be shown that a lower bound on the sampling rate with unknown carriers is $2NB$, as incorporated in the fol-

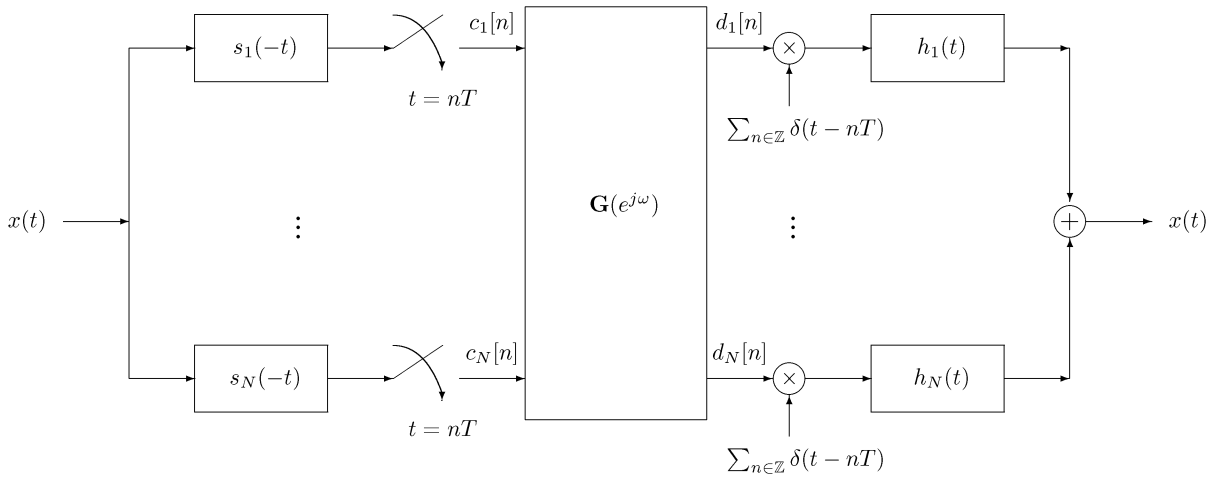


Fig. 11. Sampling and reconstruction in shift-invariant spaces. A compressive signal acquisition scheme can be obtained by reducing the number of sampling filters from N to $p < N$ and replacing the filter bank $\mathbf{G}(e^{j\omega})$ with a CTF block with real-time recovery (taken from [18]).

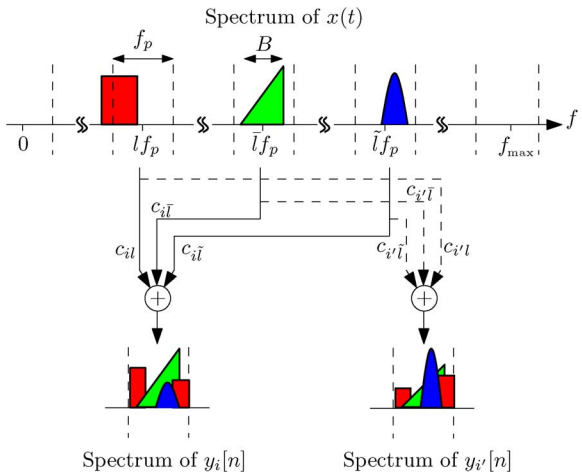


Fig. 12. Spectrum slices of $x(t)$ are overlayed in the spectrum of the output sequences $y_i[n]$. In the example, channels i and i' realize different linear combinations of the spectrum slices centered around $lf_p, \tilde{l}f_p, \tilde{l}f_p$. For simplicity, the aliasing of the negative frequencies is not drawn (taken from [9]).

lowing theorem. We refer to a sampling scheme that does not exploit the carrier frequencies as blind.

Theorem 27 [108]: *Consider a multiband model with maximal frequency $f_{\max}$ and total support Ω. Let $D(R)$ denote the sampling density of the blind sampling set*[5] *R. Then, $D(R) \geq \min\{\Omega, 2f_{\max}\}$.*

In our case the support Ω is equal to $2NB$. Therefore, as long as $NB < f_{\max}$ we can potentially reduce the sampling rate below Nyquist.

Our goal now is to use the union of subspaces framework in order to develop a sampling scheme which achieves the lower bound of Theorem 27. To describe a multiband signal as a union of subspaces, we divide the Nyquist range $[-f_{\max}, f_{\max}]$ into $M = 2L + 1$ consecutive, nonoverlapping, slices of individual widths f_p as depicted in Fig. 12, such that $\frac{L}{T} \geq f_{\max}$. Each spectrum slice represents a single bandpass subspace $\mathcal{U}_i$. By choosing $f_p \geq B$, we ensure that no more than $2N$ spectrum slices are active, i.e., contain signal energy [18]. The conceptual division to spectrum slices does not restrict the band positions; a single band can split between adjacent slices.

One way to realize the sampling scheme of Fig. 11 is through periodic nonuniform sampling (PNS) [108]. This strategy corresponds to choosing

$$w_i(t) = \delta(t - c_i T_{\mathrm{NYQ}}), \quad 1 \leq i \leq p \tag{59}$$

where $T_{\mathrm{NYQ}} = \frac{1}{f_{\mathrm{NYQ}}}$ is the Nyquist period, and using a sampling period of $T = MT_{\mathrm{NYQ}}$ with $M > p$. Here c_i are integers which select part of the uniform sampling grid, resulting in p uniform sequences

$$y_i[n] = x\left((nM + c_i)T_{\mathrm{NYQ}}\right). \tag{60}$$

The IMV model (58) that results from PNS has sequences $d_\ell[n]$ representing the contents of the ℓth bandpass subspace of the relevant spectrum slice [108]. The sensing matrix Φ is a partial discrete Fourier transform (DFT), obtained by taking only the row indices c_i from the full $M \times M$ DFT matrix.

We note that PNS was utilized for multiband sampling already in classic studies, though the traditional goal was to approach a rate of NB samples/s. This rate is optimal according to the Landau theorem [158], though achieving it for all input signals is possible only when the spectral support is known and fixed. When the carrier frequencies are unknown, the optimal rate is $2NB$ [108]. Indeed, [155] and [159] utilized knowledge of the band positions to design a PNS grid and the required interpolation filters for reconstruction. The approaches in [156], [160] were semi-blind: a sampler design independent of band positions combined with the reconstruction algorithm of [155] which requires exact support knowledge. Other techniques targeted the rate NB by imposing alternative constraints on the input spectrum [111], [122]. Here, we demonstrate how analog CS tools [18], and [113] can lead to a fully-blind sampling system of multiband inputs with unknown spectra at the appropriate optimal rate [108]. A more thorough discussion in [108] studies the differences between the analog CS method presented here based on [18], [108], [113] and earlier approaches.

[5]For a formal definition of these parameters, see [108]. Intuitively, $D(R)$ describes the average sampling rate where $R = \{r_n\}$ are the sampling points.

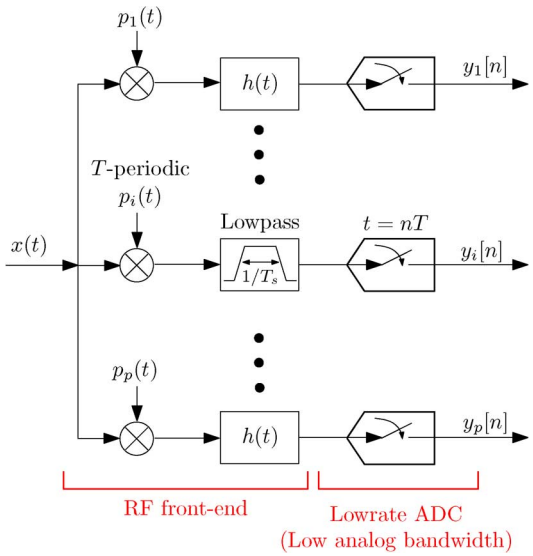


Fig. 13. Block diagram of the modulated wideband converter. The input passes through p parallel branches, where it is mixed with a set of periodic functions $p_i(t)$, lowpass filtered and sampled at a low rate (taken from [9]).

6) Example Hardware: As shown in [108], the PNS strategy can approach the minimal rate of Theorem 27. However, it requires pointwise samplers that are capable of accommodating the wide bandwidth of the input. This necessitates a high bandwidth track and hold device, which is difficult to build at high frequencies. An alternative architecture referred to as the modulated wideband converter (MWC) was developed in [109]. The MWC replaces the need for a high bandwidth track and hold device by using high bandwidth modulators, which are easier to manufacture. Indeed, off-the-shelf modulators at rates of tens of gigahertz are readily available.

The MWC combines the spectrum slices $d_\ell[n]$ according to the scheme depicted in Fig. 13. Its design is modular so that when the carrier frequencies are known the same receiver can be used with fewer channels or lower sampling rate. Furthermore, by increasing the number of channels or the rate on each channel the same realization can be used for sampling full band signals at the Nyquist rate.

The MWC consists of an analog front-end with p channels. In the ith channel, the input signal $x(t)$ is multiplied by a periodic waveform $p_i(t)$ with period T, lowpass filtered by a generic filter with impulse response $h(t)$ with cutoff $\frac{1}{2T}$, and then sampled at rate $f_s = \frac{1}{T}$. The mixing operation scrambles the spectrum of $x(t)$, so that as before, the spectrum is conceptually divided into slices of width $\frac{1}{T}$, and a weighted-sum of these slices is shifted to the origin [109]. The lowpass filter $h(t)$ transfers only the narrowband frequencies up to $\frac{f_s}{2}$ from that mixture to the output sequence $y_i[n]$. The output has the aliasing pattern illustrated in Fig. 12. Sensing with the MWC leads to a matrix $\mathbf{\Phi}$ whose entries are the Fourier expansion coefficients c_{il} of the periodic sequences $p_i(t)$.

The MWC can operate with as few as $p = 2N$ channels and with a sampling rate $f_s = \frac{1}{T} \geq B$ on each channel, so that it approaches the minimal rate of $2NB$. Advanced configurations enable additional hardware savings by collapsing the number of branches p by a factor of q at the expense of increasing the sampling rate of each channel by the same factor [109]. The choice of periodic functions $p_i(t)$ is flexible: The highest Dirac frequency needs to exceed f_{max}. In principle, any periodic function with high-speed transitions within the period T can satisfy this requirement. One possible choice for $p_i(t)$ is a sign-alternating function, with $M = 2L + 1$ sign intervals within the period T [109], [161]. Imperfect sign alternations are allowed as long as periodicity is maintained [9]. This property is crucial since precise sign alternations at high speeds are extremely difficult to maintain, whereas simple hardware wirings ensure that $p_i(t) = p_i(t + T_p)$ for every $t \in \mathbb{R}$. The waveforms $p_i(t)$ need low mutual correlation in order to capture different mixtures of the spectrum. Popular binary patterns, e.g., the Gold or Kasami sequences, are especially suitable for the MWC [161]. Another important practical design aspect is that the lowpass filter $h(t)$ does not have to be ideal. A nonflat frequency response can be compensated for in the digital domain, using the algorithm developed in [162].

The MWC has been implemented as a board-level hardware prototype [9].[6] The hardware specifications cover inputs with a 2-GHz Nyquist rate with spectrum occupation $NB = 120$ MHz. The total sampling rate is 280 MHz, far below the 2-GHz Nyquist rate. In order to save analog components, the hardware realization incorporates the advanced configuration of the MWC [109] with a collapsing factor $q = 3$. In addition, a single shift-register provides a basic periodic pattern, from which p periodic waveforms are derived using delays, that is, by tapping p different locations of the register. A nice feature of the recovery stage is that it interfaces seamlessly with standard DSPs by providing (samples of) the narrowband information signals. This capability is provided by a digital algorithm that is developed in [10].[7]

The MWC board is a first hardware example of the use of ideas borrowed from CS for sub-Nyquist sampling and low-rate recovery of wideband signals where the sampling rate is directly proportional to the actual bandwidth occupation and not the highest frequency. Existing implementations of the random demodulator (RD) (cf. Section IV-B-3) recover signals at effective sampling rates below 1 MHz, falling outside of the class of wideband samplers. Additionally, the signal representations used by the RD have size proportional to the Nyquist frequency, leading to recovery problems that are much larger than those posed by the MWC. See [9] and [10] for additional information on the similarities and differences between the MWC, the RD, and other comparable architectures.

C. Infinite Union of Finite-Dimensional Subspaces

The second union class we consider is when $\mathcal{U}$ is composed of an infinite number m of subspaces, and each subspace has finite dimension.

[6]A video of experiments and additional documentation for the MWC hardware are available at http://webee.technion.ac.il/Sites/People/YoninaEldar/hardware.html. A graphical package demonstrating the MWC numerically is available at http://webee.technion.ac.il/Sites/People/YoninaEldar/software_det3.html.

[7]The algorithm is available online at http://webee.technion.ac.il/Sites/People/YoninaEldar/software_det4.html.

1) Analog Signal Model: As we have seen in Section VI-A, the SI model (54) is a convenient way to describe analog signals in infinite-dimensional spaces. We can use a similar approach to describe analog signals that lie within finite-dimensional spaces by restricting the number of unknown gains $a_\ell[n]$ to be finite. In order to incorporate structure into this model, we assume that each generator $h_\ell(t)$ has an unknown parameter θ_ℓ associated with it, which can take on values in a continuous interval, resulting in the model

$$x(t)=\sum_{\ell=1}^{L} a_\ell h_\ell(t,\theta_\ell). \tag{61}$$

Each possible choice of the set $\{\theta_\ell\}$ leads to a different L-dimensional subspace of signals $\mathcal{U}_i$, spanned by the functions $\{h(t,\theta_\ell)\}$. Since θ_ℓ can take on any value in a given interval, the model (61) corresponds to an infinite union of finite dimensional subspaces (i.e., $m=\infty$).

An important example of (61) is when $h_\ell(t,\theta_\ell)=h(t-t_\ell)$ for some unknown time delay $\theta_l=t_\ell$, leading to a stream of pulses

$$x(t)=\sum_{\ell=1}^{L} a_\ell h(t-t_\ell). \tag{62}$$

Here, $h(t)$ is a known pulse shape and $\{t_\ell,a_\ell\}_{\ell=1}^{L}$, $t_\ell\in[0,\tau)$, $a_\ell\in\mathbb{C}$ are unknown delays and amplitudes. This model was first introduced and studied by Vetterli *et al.* [6]–[8] as a special case of signals having a finite number of degrees of freedom per unit time, termed finite rate of innovation (FRI) signals.

Our goal is to sample $x(t)$ and reconstruct it from a minimal number of samples. The primary interest is in pulses which have small time-support, and therefore the required Nyquist rate would be very high. However, since the pulse shape is known, the signal has only $2L$ degrees of freedom, and therefore, we expect the minimal number of samples to be $2L$, much lower than the number of samples resulting from Nyquist rate sampling.

A simpler version of the problem is when the signal $x(t)$ of (62) is repeated periodically leading to the model

$$x(t)=\sum_{m\in\mathbb{Z}}\sum_{\ell=1}^{L} a_\ell h(t-t_\ell-m\tau), \tag{63}$$

where τ is the known period. This periodic setup is easier to treat because we can exploit the properties of the Fourier series representation of $x(t)$ due to the periodicity. The dimensionality and number of subspaces included in the model (38) remain unchanged.

2) Compressive Signal Acquisition: To date, there are no general acquisition methods for signals of the form (61). Instead, we focus on the special case of (63).

Our sampling scheme follows the ideas of [6]–[8] and consists of a filter $s(t)$ followed by uniform sampling of the output with period $T=\frac{\tau}{N}$, where N is the number of samples in one period, and $N\geq 2L$. The resulting samples can be written as inner products $c[n]=\langle s(t-nT),x(t)\rangle$. The following theorem establishes properties of the filter $s(t)$ that allow recovery from $2L$ samples.

Theorem 28 [163]: *Consider the τ-periodic stream of pulses of order L given by (63). Choose a set $\mathcal{K}$ of consecutive indices for which $H\left(\frac{2\pi k}{\tau}\right)\neq 0$ where $H(\omega)$ is the Fourier transform of the pulse $h(t)$. Then the samples $c[n]$ for $n=0,\ldots,N-1$, uniquely determine the signal $x(t)$ as long as $N\geq|\mathcal{K}|\geq 2L$ for any $s(t)$ satisfying*

$$S(\omega)=\begin{cases}0, & \omega=\frac{2\pi k}{\tau}, k\notin\mathcal{K}\\ \text{nonzero}, & \omega=\frac{2\pi k}{\tau}, k\in\mathcal{K}\\ \text{arbitrary}, & \text{otherwise.}\end{cases} \tag{64}$$

Theorem 28 was initially focused on low pass filters in [6], and was later extended to arbitrary filters in [163].

To see how we can recover $x(t)$ from the samples of Theorem 28, we note that the Fourier series coefficients $X[k]$ of the periodic pulse stream $x(t)$ are given by [6]

$$\begin{aligned}X[k]&=\frac{1}{\tau}H\left(\frac{2\pi k}{\tau}\right)\sum_{\ell=1}^{L} a_\ell e^{-\frac{j2\pi k t_\ell}{\tau}}\\ &=\frac{1}{\tau}H\left(\frac{2\pi k}{\tau}\right)\sum_{\ell=1}^{L} a_\ell u_\ell^k\end{aligned} \tag{65}$$

where $u_\ell=e^{-\frac{j2\pi t_\ell}{\tau}}$. Given $X[k]$, the problem of retrieving a_ℓ and t_ℓ in (65) is a standard problem in array processing [6], [164] and can be solved using methods developed in that context such as the matrix pencil [165], subspace-based estimators [166], [167], and the annihilating filter [8]. These methods require $2L$ Fourier coefficients to determine a_ℓ and u_ℓ. In the next subsection, we show that the vector $\mathbf{X}$ of Fourier coefficients $X[k]$ can be computed from the N samples $c[n]$ of Theorem 28.

Since the LPF has infinite time support, the approach of (64) cannot work with time-limited signals, such as those of the form (62). A class of filters satisfying (64) that have finite time support are Sum of Sincs (SoS) [163], which are given in the Fourier domain by

$$G(\omega)=\frac{\tau}{\sqrt{2\pi}}\sum_{k\in\mathcal{K}} b_k\operatorname{sinc}\left(\frac{\omega}{\frac{2\pi}{\tau}}-k\right) \tag{66}$$

where $b_k\neq 0$, $k\in\mathcal{K}$. Switching to the time domain

$$g(t)=\operatorname{rect}\left(\frac{t}{\tau}\right)\sum_{k\in\mathcal{K}} b_k e^{\frac{j2\pi k t}{\tau}} \tag{67}$$

which is clearly a time compact filter with support τ. For the special case in which $\mathcal{K}=\{-p,\ldots,p\}$ and $b_k=1$,

$$g(t)=\operatorname{rect}\left(\frac{t}{\tau}\right)\sum_{k=-p}^{p} e^{\frac{j2\pi k t}{\tau}}=\operatorname{rect}\left(\frac{t}{\tau}\right)D_p\left(\frac{2\pi t}{\tau}\right)$$

where $D_p(t)$ denotes the Dirichlet kernel.

Alternative approaches to sample finite pulse streams of the form (63) rely on the use of splines [7]; this enables obtaining moments of the signal rather than its Fourier coefficients. The moments are then processed in a similar fashion (see the next subsection for details). However, this approach is unstable for high values of L [7]. In contrast, the SoS class can be used for stable reconstruction even for very high values of L, e.g., $L=100$.

Multichannel schemes can also be used to sample pulse streams. This approach was first considered for Dirac streams, where a successive chain of integrators allows obtaining moments of the signal [168]. Unfortunately, the method is highly sensitive to noise. A simple sampling and reconstruction

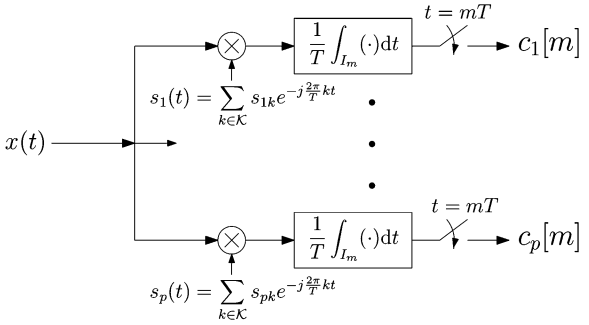


Fig. 14. Extended sampling scheme using modulating waveforms (taken from [170]).

scheme consisting of two channels, each with an RC circuit, was presented in [169] for the special case where there is no more than one Dirac per sampling period. A more general multichannel architecture that can treat a broader class of pulses, while being much more stable, is depicted in Fig. 14 [170]. The system is very similar to the MWC presented in the previous section. By correct choice of the mixing coefficients, the Fourier coefficients $X[k]$ may be extracted from the samples by a simple matrix inversion.

3) Recovery Algorithms: In both the single-channel and multichannel approaches, recovery of the unknown delays and amplitudes proceeds in two steps. First, the vector of samples $\mathbf{c}$ is related to the Fourier coefficient vector $\mathbf{x}$ through a $p \times |\mathcal{K}|$ mixing matrix $\mathbf{Q}$, as $\mathbf{c} = \mathbf{Q}\mathbf{x}$. Here, $p \geq 2L$ represents the number of samples. When using the SoS approach with a filter $S(\omega)$, $\mathbf{Q} = \mathbf{V}\mathbf{S}$ where $\mathbf{S}$ is a $p\times p$ diagonal matrix with diagonal elements $S^*\left(-\frac{2\pi\ell}{\tau}\right)$, $1 \leq \ell \leq p$, and $\mathbf{V}$ is a $p \times |\mathcal{K}|$ Vandermonde matrix with ℓ^{th} element given by $e^{\frac{j2\pi\ell T}{\tau}}$, $1 \leq \ell \leq p$, where T denotes the sampling period. For the multichannel architecture of Fig. 14, $\mathbf{Q}$ consists of the modulation coefficients $s_{\ell k}$. The Fourier coefficient vector $\mathbf{x}$ can be obtained from the samples as $\mathbf{x} = \mathbf{Q}^\dagger\mathbf{c}$.

The unknown parameters $\{t_l, a_l\}_{l=1}^L$ are recovered from $\mathbf{x}$ using, e.g., the annihilating filter method. The annihilating filter $\{r[k]\}_{k=0}^L$ is defined by its z-transform

$$R(z) = \sum_{k=0}^{L} r[k]z^{-k} = \prod_{\ell=1}^{L}(1 - u_\ell z^{-1}). \tag{68}$$

That is, the roots of $R(z)$ equal the values u_ℓ through which the delays t_ℓ can be found. It then follows that

$$\begin{aligned} r[k] * X[k] &= \sum_{l=0}^{L} r[l]X[k-l] = \sum_{l=0}^{L}\sum_{\ell=1}^{L} r[l]a_\ell u_\ell^{k-l} \\ &= \sum_{\ell=1}^{L} a_\ell u_\ell^k \sum_{l=0}^{L} r[l]u_\ell^{-l} = 0 \end{aligned} \tag{69}$$

where the last equality is due to $R(u_\ell) = 0$. Assuming without loss of generality that $R[0] = 1$, the identity in (69) can be written in matrix/vector form as

$$\begin{pmatrix} X[-1] & \dots & X[-L] \\ X[0] & \dots & X[-L+1] \\ \vdots & \ddots & \vdots \\ X[L-2] & \dots & X[-1] \end{pmatrix} \begin{pmatrix} r[1] \\ r[2] \\ \vdots \\ r[L] \end{pmatrix} = \begin{pmatrix} X[0] \\ X[1] \\ \vdots \\ X[L-1] \end{pmatrix}.$$

Thus, we only need $2L$ consecutive values of $X[k]$ to determine the annihilating filter. Once the filter is found, the values t_ℓ are retrieved from the zeros u_ℓ of the z-transform in (68). Finally, the Fourier coefficients a_ℓ are computed using (65). For example, if we have the coefficients $X[k]$, $0 \leq k \leq 2L-1$, then (65) may be written as

$$\frac{1}{\tau}\begin{pmatrix} 1 & 1 & \dots & 1 \\ u_0 & u_1 & \dots & u_L \\ \vdots & \vdots & \ddots & \vdots \\ u_0^{2L} & u_1^{2L} & \dots & u_L^{2L} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_L \end{pmatrix} = \begin{pmatrix} \frac{X[0]}{H(0)} \\ \frac{X[1]}{H\left(\frac{2\pi}{\tau}\right)} \\ \vdots \\ \frac{X[2L-1]}{H\left(\frac{2\pi(2L-1)}{\tau}\right)} \end{pmatrix}.$$

Since this Vandermonde matrix is left-invertible, the values a_ℓ can be computed by matrix inversion.

Reconstruction results for the sampling scheme based on the SoS filter with $b_k = 1$ are depicted in Fig. 15. The original signal consists of $L = 5$ Gaussian pulses, and $N = 11$ samples were used for reconstruction. The reconstruction is exact to numerical precision. A comparison of the performance of various methods in the presence of noise is depicted in Fig. 15 for a finite stream consisting of three and five pulses. The pulse-shape is a Dirac delta, and white Gaussian noise is added to the samples with a proper level in order to reach the desired SNR for all methods. All approaches operate using $2L + 1$ samples.

4) Applications: As an example application of FRI, we consider multiple image registration for superresolution imaging. A superresolution algorithm aims at creating a single detailed image, called a super-resolved image (SR), from a set of low-resolution input images of the same scene. If different images from the same scene are taken such that their relative shifts are not integer multiples of the pixel size, then subpixel information exists among the set. This allows to obtain a higher resolution accuracy of the scene once the images are properly registered.

Image registration involves any group of transformations that removes the disparity between two low resolution (LR) images. This is followed by image fusion, which blends the properly aligned LR images into a higher resolution output, possibly removing blur and noise introduced by the system [171]. The registration step is crucial is order to obtain a good quality SR image. The theory of FRI can be extended to provide superresolution imaging. The key idea of this approach is that, using a proper model for the point spread function (PSF) of the scene acquisition system, it is possible to retrieve the underlying continuous geometric moments of the irradiance light-field. From this information, and assuming the disparity between any two images can be characterized by a global affine transformation, the set of images may be registered. The parameters obtained via FRI correspond to the shift vectors that register the different images before image fusion. Fig. 16 shows a real-world superresolution example in which 40 low-resolution images allow an improvement in image resolution by a factor of 8.

A second example application of the FRI model is ultrasonic imaging. In this imaging modality, an ultrasonic pulse is transmitted into a tissue, e.g., the heart, and a map of the underlying tissues is created by locating the echoes of the pulse. Correct

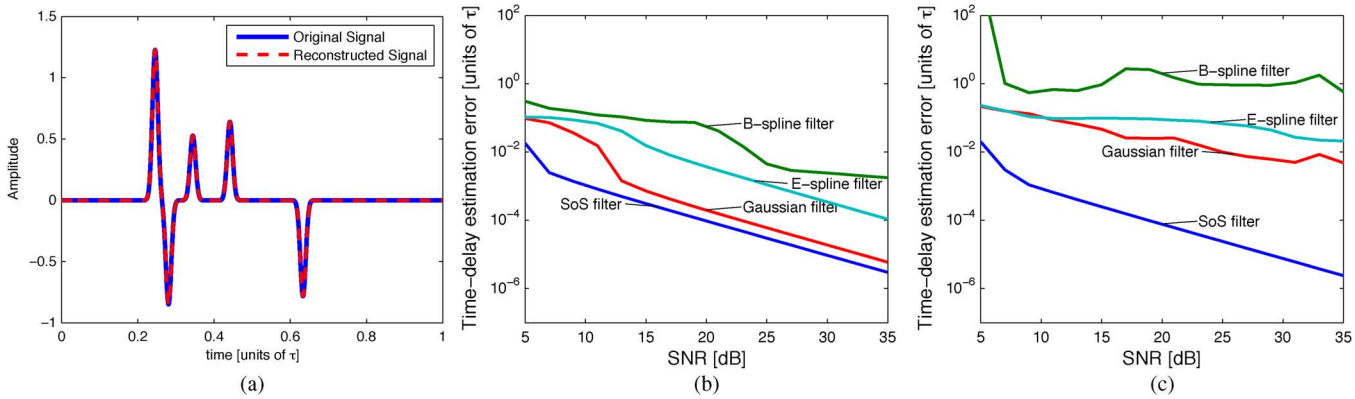


Fig. 15. Performance comparison of finite pulse stream recovery using Gaussian [6], B-spline, E-spline [7], and SoS sampling kernels. (a) Reconstructed signal using SoS filters versus original one. The reconstruction is exact to numerical precision. (b) $L = 3$ Dirac pulses are present, (c) $L = 5$ pulses (taken from [163]).

location of the tissues and their edges is crucial for medical diagnosis. Current technology uses high rate sampling and processing in order to construct the image, demanding high computational complexity. Noting that the received signal is a stream of delayed and weighted versions of the known transmitted pulse shape, FRI sampling schemes can be exploited in order to reduce the sampling rate and the subsequent processing rates by several orders of magnitude while still locating the echoes. The received ultrasonic signal is modeled as a finite FRI problem, with a Gaussian pulse-shape.

In Fig. 17, we consider a signal obtained using a phantom consisting of uniformly spaced pins, mimicking point scatterers, which is scanned by GE Healthcare's Vivid-i portable ultrasound imaging system. The data recorded by a single element in the probe is modeled as a 1D stream of pulses. The recorded signal is depicted in Fig. 17(a). The reconstruction results using the SoS filter with $b_k = 1$ are depicted in Fig. 17(b)–(c). The algorithm looked for the $L = 4$ strongest echoes, using $N = 17$ and $N = 33$ samples. In both simulations, the estimation error in the location of the pulses is around 0.1 mm. These ideas have also been extended recently to allow for 2-D imaging with multiple received elements [172].

D. Infinite Union of Infinite-Dimensional Subspaces

Extending the finite-dimensional model of Section VI-C to the SI setting (54), we now incorporate structure by assuming that each generator $h_\ell(t)$ has an unknown parameter θ_ℓ associated with it, leading to an infinite union of infinite-dimensional spaces. As with its finite counterpart, there is currently no general sampling framework available to treat such signals. Instead, we focus on the case in which $h_\ell(t) = h(t - \tau_\ell)$.

1) Analog Signal Model: Suppose that

$$x(t) = \sum_{l \in \mathbb{Z}} a_\ell h(t - t_\ell), \quad t_\ell, a_\ell \in \mathbb{R} \tag{70}$$

where there are no more than L pulses in an interval of length T, and that the pulses do not overlap interval boundaries. In this case, the signal parameters in each interval can be treated separately, using the schemes of the previous section. In particular, we can use the sampling system of Fig. 14 where now the integral is obtained over intervals of length T [170]. This requires obtaining a sample from each of the channels once every T seconds, and using $p \geq 2K$ channels, resulting in samples taken at the minimal possible rate.

We next turn to treat the more complicated scenario in which $h(t)$ may have support larger than T. This setting can no longer be treated by considering independent problems over periods of T. To simplify, we consider the special case in which the time delays in (70) repeat periodically (but not the amplitudes) [19], [150]. As we will show in this special case, efficient sampling and recovery is possible even using a single filter, and without requiring the pulse $h(t)$ to be time limited. Under our assumptions, the input signal can be written as

$$x(t) = \sum_{n \in \mathbb{Z}} \sum_{\ell=1}^{L} a_\ell[n] h(t - t_\ell - nT) \tag{71}$$

where $\tau = \{t_\ell\}_{\ell=1}^L$ is a set of unknown time delays contained in the time interval $[0, T)$, $\{a_\ell[n]\}$ are arbitrary bounded energy sequences, and $h(t)$ is a known pulse shape.

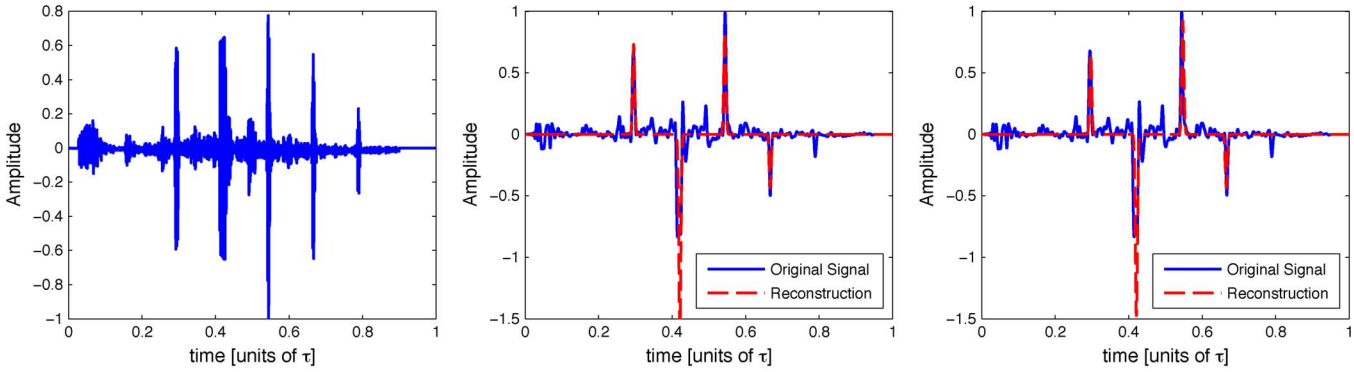
2) Compressive Signal Acquisition: We follow a similar approach to that in Section VI-B, which treats a structured SI setting where there are N possible generators. The difference is that in this current case there are infinitely many possibilities. Therefore, we replace the CTF in Fig. 11 with a block that supports this continuity: we will see that the ESPRIT method essentially replaces the CTF block [173].

A sampling and reconstruction scheme for signals of the form (71) is depicted in Fig. 18 [19]. The analog sampling stage is comprised of p parallel sampling channels. In each channel, the input signal $x(t)$ is filtered by a band-limited sampling kernel $s_\ell^*(-t)$ with frequency support contained in an interval of width $\frac{2\pi p}{T}$, followed by a uniform sampler operating at a rate of $\frac{1}{T}$, thus providing the sampling sequence $c_\ell[n]$. Note that just as in the MWC (Section VI-B-6), the sampling filters can be collapsed to a single filter whose output is sampled at p times the rate of a single channel. The role of the sampling kernels is to spread out the energy of the signal in time, prior to low rate sampling.

3) Recovery Algorithms: To recover the signal from the samples, a properly designed digital filter correction bank, whose frequency response in the DTFT domain is given by $\mathbf{W}^{-1}(e^{j\omega T})$, is applied to the sampling sequences in a manner similar to (55). The matrix $\mathbf{W}(e^{j\omega T})$ depends on the choice

(a) (b) (c) (d)

Fig. 16. Example of FRI-based image superresolution. 40 images of a target scene were acquired with a digital camera. (a) Example acquired image. (b) Region of interest (128 × 128 pixels) used for superresolution. (c) Superresolved image of size 1024 × 1024 pixels (SR factor = 8). The PSF in this case is modeled by a B-spline of order 7 (scale 1). (d) Superresolved image of size 1024 × 1024 pixels (SR factor = 8). The PSF in this case is modeled by a B-spline of order 3 (scale 2) (taken from [171]).



(a) (b) (c)

Fig. 17. Applying the SoS sampling scheme with $b_k = 1$ on real ultrasound imaging data. Results are shown versus original signal which uses 4160 samples. (a) Recorded ultrasound imaging signal. The data was acquired by GE healthcare's Vivid-i ultrasound imaging system. Reconstructed signal (b) using $N = 17$ samples and (c) using $N = 33$ samples (taken from [163]).

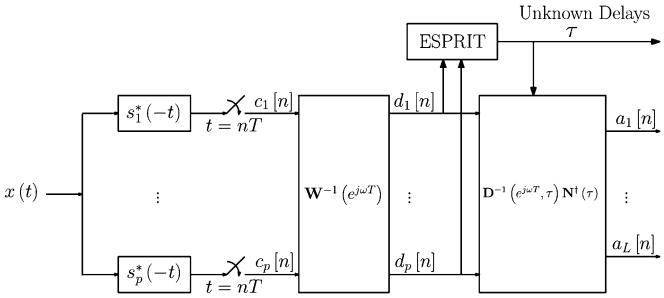


Fig. 18. Sampling and reconstruction scheme for signals of the form (71).

of the sampling kernels $s_\ell^*(-t)$ and the pulse shape $h(t)$. Its entries are defined for $1 \le \ell, m \le p$ as

$$\mathbf{W}(e^{j\omega T})_{\ell,m} = \frac{1}{T} S_\ell^* \left(\omega + \frac{2\pi m}{T}\right) H\left(\omega + \frac{2\pi m}{T}\right). \quad (72)$$

After the digital correction stage, it can be shown that the corrected sample vector $\mathbf{d}[n]$ is related to the unknown amplitude vector $\mathbf{a}[n]$ by a Vandermonde matrix which depends on the unknown delays [19]. Therefore, we can now exploit known tools taken from the direction of arrival [174] and spectral estimation [164] literature to recover the delays $\tau = \{t_1, \ldots, t_L\}$, such as the well-known ESPRIT algorithm [173]. Once the delays are determined, additional filtering operations are applied on the samples to recover the sequences $a_\ell[n]$. In particular, referring to Fig. 18, the matrix $\mathbf{D}$ is a diagonal matrix with diagonal elements equal to $e^{-j\omega t_k}$, and $\mathbf{N}(\tau)$ is a Vandermonde matrix with elements $e^{-\frac{j2\pi m t_k}{T}}$. The ESPRIT algorithm is summarized for our setting as Algorithm 4, where $\mathbf{E}_{s\downarrow}$ and $\mathbf{E}_{s\uparrow}$ denote the sub matrices extracted from $\mathbf{E}_s$ by deleting its last/first row, respectively.

Algorithm 4: ESPRIT Algorithm

Input: Signal d, number of parameters L.

Output: Delays $\tau = \{t_1, \ldots, t_L\}$.

$\mathbf{R}_{dd} = \sum_{n\in\mathbb{Z}} \mathbf{d}[n]\mathbf{d}^H[n]$ {construct correlation matrix}

$(\mathbf{E}, \sigma) \leftarrow \mathrm{SVD}(\mathbf{R}_{dd})$ {calculate SVD}

$\mathbf{E}_s \leftarrow L$ singular vectors of $\mathbf{R}_{dd}$ associated with nonzero singular values.

$\Phi = \mathbf{E}_{s\downarrow}^{\dagger}\mathbf{E}_{s\uparrow}$ {compute ESPRIT matrix}

$\{\lambda_1, \ldots, \lambda_L\} \leftarrow \mathrm{eig}(\Phi)$ {obtain eigenvalues}

$t_i \leftarrow -\frac{T}{2\pi}\arg(\lambda_i)$, $1 \le i \le L$ {map eigenvalues to delays}

return $\tau = \{t_1, \ldots, t_L\}$

4) Recovery Guarantees: The proposed algorithm is guaranteed to recover the unknown delays and amplitudes as long as p is large enough [19].

Theorem 29 [19]: *The sampling scheme of Fig. 18 is guaranteed to recover any signal of the form (71) as long as* $p \ge 2L - \eta + 1$, *where* η *is the dimension of the minimal subspace*

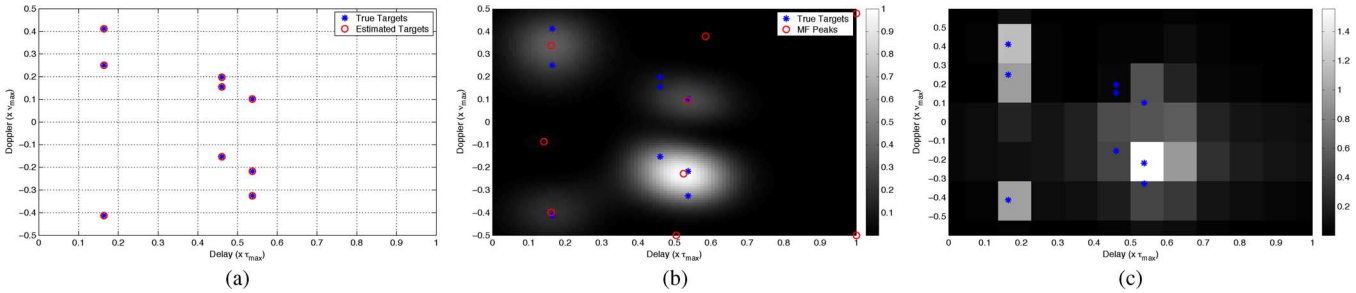


Fig. 19. Comparison between the target-detection performance for the case of nine targets (represented by $*$) in the delay—Doppler space with $\tau_{max} = 10\,\mu$s, $\nu_{max} = 10$ kHz, $W = 1.2$ MHz, and $T = 0.48$ ms. The probing sequence $\{x_n\}$ corresponds to a random binary (± 1) sequence with $N = 48$, the pulse $p(t)$ is designed to have a nearly flat frequency response and the pulse repetition interval is $T = 10\,\mu$s. Recovery of the Doppler-delay plane using (a) a union of subspaces approach, (b) a standard matched filter, and (c) a discretized delay-Doppler plane (taken from [150]).

containing the vector set $\{\mathbf{d}[n], n \in \mathbb{Z}\}$. *In addition, the filters* $s_\ell(t)$ *are supported on* $\frac{2\pi p}{T}$ *and must be chosen such that* $\mathbf{W}(e^{j\omega T})$ *of (72) is invertible.*

The sampling rate resulting from Theorem 29 is no larger than $\frac{2L}{T}$ since $0 \leq \eta \leq L$. For certain signals, the rate can be reduced to $\frac{(L+1)}{T}$. This is due to the fact that the outputs are processed jointly, similar to the MMV model. Evidently, the minimal sampling rate is not related to the Nyquist rate of the pulse $h(t)$. Therefore, for wideband pulse shapes, the reduction in rate can be quite substantial. As an example, consider the setup in [175], used for characterization of ultra-wide band wireless indoor channels. Under this setup, pulses with bandwidth of $W = 1$ GHz are transmitted at a rate of $\frac{1}{T} = 2$ MHz. Assuming that there are 10 significant multipath components, we can reduce the sampling rate down to 40 MHz compared with the 2-GHz Nyquist rate.

5) Applications: Problems of the form (71) appear in a variety of different settings. For example, the model (71) can describe multipath medium identification problems, which arise in applications such as radar [176], underwater acoustics [177], wireless communications, and more. In this context, pulses with known shape are transmitted through a multipath medium, which consists of several propagation paths, at a constant rate. As a result the received signal is composed of delayed and weighted replicas of the transmitted pulses. The delays t_ℓ represent the propagation delays of each path, while the sequences $a_\ell[n]$ describe the time-varying gain coefficient of each multipath component.

Another important application of (71) is in the context of radar. In this example, we translate the rate reduction to increased resolution with a fixed time–bandwidth product (TBP), thus enabling super-resolution radar from low rate samples. In radar, the goal is to identify the range and velocity of K targets. The delay in this case captures the range while the time varying coefficients are a result of the Doppler delay related to the target velocity [150]. More specifically, we assume that several targets can have the same delays but possibly different Doppler shifts so that $\{t_\ell\}_{\ell=1}^{L}$ denote the set of distinct delays. For each delay value t_ℓ there are K_ℓ values of associated Doppler shifts $\nu_{\ell k}$ and reflection coefficients $\alpha_{\ell k}$. We also assume that the system is highly underspread, namely $\nu_{\max} T \ll 1$, where $\nu_{\max}$ and T denote the maximal Doppler shift and delay. To identify the targets we transmit the signal

$$x_T = \sum_{n=0}^{N-1} x_n h(t - nT) \tag{73}$$

where x_n is a known N-length probing sequence, and $h(t)$ is a known pulse shape. The received signal can then be described in the form (71), where the sequences $a_\ell[n]$ satisfy

$$a_\ell[n] = x_n \sum_{k=1}^{K_\ell} \alpha_{\ell k} e^{j2\pi\nu_{\ell k} nT}. \tag{74}$$

The delays and the sequences $a_\ell[n]$ can be recovered using the general scheme for time delay recovery. The Doppler shifts and reflection coefficients are then determined from the sequences $a_\ell[n]$ using standard spectral estimation tools [164]. The targets can be exactly identified as long as the bandwidth $\mathcal{W}$ of the transmitted pulse satisfies $\mathcal{W} \geq \frac{4\pi L}{T}$, and the length of the probing sequence satisfies $N \geq 2 \max K_\ell$ [150]. This leads to a minimal TBP of the input signal of $\mathcal{W}\mathcal{T} \geq 8\pi L \max K_\ell$, which is much lower than that obtained using standard radar processing techniques, such as matched-filtering (MF).

An example of the identification of nine close targets is illustrated in Fig. 19(a). The sampling filter used is a simple LPF. The original and recovered targets are shown on the Doppler-delay plane. Evidently all the targets were correctly identified. The result obtained by MF, with the same TBP, is shown in Fig. 19(b). Clearly, the compressive method has superior resolution than the standard MF in this low noise setting. Thus, the union of subspaces viewpoint not only offers a reduced-rate sampling method, but allows to increase the resolution in target identification for a fixed low TBP when the SNR is high enough, which is of paramount importance in many practical radar problems.

Previous approaches for identifying the unknown delays and gains involve sampling the received signal at the Nyquist rate of the pulse $h(t)$ [178]–[180]. However, prior knowledge of the pulse shape results in a parametric viewpoint, and we would expect that the rate should be proportional to the number of degrees of freedom, i.e., the number of paths L.

Another approach is to quantize the delay-Doppler plane by assuming the delays and Doppler shifts lie on a grid [42], [104], [148], [149]. After discretization, CS tools for finite representations can be used to capture the sparsity on the discretized grid. Clearly, this approach has an inherent resolution limitation. Moreover, in real world scenarios, the targets do not lie exactly on the grid points. This causes a leakage of their energies in the quantized space into adjacent grid points [36], [87]. In contrast, the union of subspaces model avoids the discretization issue and offers concrete sampling methods that can be implemented efficiently in hardware (when the SNR is not too poor).

E. Discussion

Before concluding, we point out that much of the work in this section (Sections VI-B, VI-C and VI-D) has been focused on infinite unions. In each case, the approach we introduced is based on a sampling mechanism using analog filters, followed by standard array processing tools for recovery in the digital domain. These techniques allow perfect recovery when no noise is present, and generally degrade gracefully in the presence of noise, as has been analyzed extensively in the array processing literature. Nonetheless, when the SNR is poor, alternative digital recovery methods based on CS may be preferable. Thus, we may increase robustness by using the analog sampling techniques proposed here in combination with standard CS algorithms for recovery in the digital domain. To apply this approach, once we have obtained the desired samples, we can view them as the measurement vector y in a standard CS system; the CS matrix Φ is now given by a discretized Vandermonde matrix that captures all possible frequencies to a desired accuracy; and the vector x is a sparse vector with nonzero elements only in those indices corresponding to actual frequencies [42], [84], [104], [148], [149]. In this way, we combine the benefits of CS with those of analog sampling, without requiring discretization in the sampling stage. The discretization now only appears in the digital domain during recovery and not in the sampling mechanism. If we follow this strategy, then the results developed in Section III regarding recovery in the presence of noise are relevant here as well.

VII. Conclusion

In this review, our aim was to summarize applications of the basic CS framework that integrate specific constraints of the problem at hand into the theoretical and algorithmic formulation of signal recovery, going beyond the original "random measurement/sparsity model" paradigm. Due to constraints given by the relevant sensing devices, the classes of measurement matrices available to us are limited. Similarly, some applications focus on signals that exhibit structure which cannot be captured by sparsity alone and provide additional information that can be leveraged during signal recovery. We also considered the transition between continuous and discrete representations bridged by analog-to-digital converters. Analog signals are continuous-time by nature; in many cases, the application of compressed sensing to this larger signal class requires the formulation of new devices, signal models, and recovery algorithms. It also necessitates a deeper understanding of hardware considerations that must be taken into account in theoretical developments. This acquisition framework, in turn, motivates the design of new sampling schemes and devices that provide the information required for signal recovery in the smallest possible representation.

While it is difficult to cover all of the developments in compressive sensing theory and hardware [66]–[68], [181]–[184], our aim here was to select few examples that were representative of wider classes of problems and that offered a balance between a useful theoretical background and varied applications. We hope that this summary will be useful to practitioners in signal acquisition and processing that are interested in leveraging the features of compressive sensing in their specific applications. We also hope that this review will inspire further developments in the theory and practice underlying CS: we particularly envision extending the existing framework to broader signals sets and inspiring new implementation and design paradigms. With an eye to the future, more advanced configurations of CS can play a key role in many new frontiers. Some examples already mentioned throughout are cognitive radio, optical systems, medical devices such as MRI, ultrasound and more. These techniques hold promise for a complete rethinking of many acquisition systems and stretch the limit of current sensing capabilities.

As we have also demonstrated, CS holds promise for increasing resolution by exploiting signal structure. This can revolutionize many applications such as radar and microscopy by making efficient use of the available degrees of freedom in these settings. Consumer electronics, microscopy, civilian and military surveillance, medical imaging, radar and many other rely on ADCs and are resolution-limited. Removing the Nyquist barrier in these applications and increasing resolution can improve the user experience, increase data transfer, improve imaging quality and reduce exposure time—in other words, make a prominent impact on the analog-digital world surrounding us.

Acknowledgment

The authors would like to thank their colleagues for many useful comments and for their collaboration on many topics related to this review. In particular, they are grateful to M. Wakin for many discussions on the structure of the review, to M. Mishali, H. Rauhut, J. Romberg, Y. Xie, and the anonymous reviewers for offering many helpful comments on an early draft of the manuscript, and to P. L. Dragotti, P. Golland, L. Jacques, Y. Massoud, J. Romberg, and R. Willett for authorizing the use of their figures.

References

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Sep. 2006.

[2] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[3] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–120, 124, Jul. 2007.

[4] E. J. Candès, "Compressive sampling," in *Proc. Int. Congr. Math.*, Madrid, Spain, 2006, vol. 3, pp. 1433–1452.

[5] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[6] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, 2002.

[7] P. L. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang-fix," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1741–1757, May 2007.

[8] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 31–40, Mar. 2008.

[9] M. Mishali, Y. C. Eldar, O. Dounaevsky, and E. Shoshan, "Xampling: Analog to digital at sub-Nyquist rates," *IET Circuits, Devices, Syst.*, vol. 5, no. 1, pp. 8–20, Jan. 2011.

[10] M. Mishali and Y. C. Eldar, "Xampling: Compressed sensing for analog signals," in *Compressed Sensing: Theory and Applications*, Y. C. Eldar and G. Kutyniok, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[11] H. Nyquist, "Certain topics in telegraph transmission theory," *Trans. Amer. Inst. Electr. Eng.*, vol. 47, pp. 617–644, Apr. 1928.

[12] C. E. Shannon, "Communications in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, Jan. 1949.

[13] E. T. Whittaker, "On the functions which are represented by the expansions of the interpolation theory," *Proc. R. Soc. Edinb.*, vol. 35, pp. 181–194, 1915.

[14] V. A. Kotelnikov, "On the transmission capacity of the 'ether' and of cables in electrical communications," in *All-Union Conf. Technol. Reconstruction Commun. Sector and Develop. of Low-Current Eng.*, Moscow, Russia, 1933.

[15] Y. M. Lu and M. N. Do, "A theory for sampling signals from a union of subspaces," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2334–2345, 2008.

[16] 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, Apr. 2009.

[17] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.

[18] Y. C. Eldar, "Compressed sensing of analog signals in shift-invariant spaces," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 2986–2997, Aug. 2009.

[19] K. Gedalyahu and Y. C. Eldar, "Time delay estimation from low rate samples: A union of subspaces approach," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3017–3031, Jun. 2010.

[20] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[21] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[22] S. Chen, D. L. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[23] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, Feb. 2009.

[24] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[25] S. Foucart, "A note on guaranteed sparse recovery via ℓ_1-minimization," *Appl. Comput. Harmon. Anal.*, vol. 29, no. 1, pp. 97–103, 2010.

[26] S. Foucart and R. Gribonval, "Real versus complex null space properties for sparse vector recovery," *C. R. Acad. Sci. Paris, Ser. I*, vol. 348, no. 15–16, pp. 863–865, 2010.

[27] A. Cohen, W. Dahmen, and R. A. DeVore, "Compressed sensing and best k-term approximation," *J. Amer. Math. Soc.*, vol. 22, no. 1, pp. 211–231, Jan. 2009.

[28] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

[29] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decompositions," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[30] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[31] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

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

[33] T. Strohmer and R. Heath, "Grassmanian frames with applications to coding and communication," *Appl. Comput. Harmon. Anal.*, vol. 14, no. 3, pp. 257–275, Nov. 2003.

[34] S. A. Geršgorin, "Über die abgrenzung der eigenwerte einer matrix," *Izv. Akad. Nauk SSSR Ser. Fiz.-Mat.*, vol. 6, pp. 749–754, 1931.

[35] M. A. Herman and T. Strohmer, "General deviants: An analysis of perturbations in compressed sensing," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 342–349, Apr. 2010.

[36] Y. Chi, L. L. Scharf, A. Pezeshki, and R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, May 2011.

[37] J. Treichler, M. A. Davenport, and R. G. Baraniuk, "Application of compressive sensing to the design of wideband signal acquisition receivers," presented at the U.S./Australia Joint Workshop Defense Appl. Signal Process. (DASP), Lihue, HI, Sep. 2009.

[38] S. Aeron, V. Saligrama, and M. Zhao, "Information theoretic bounds for compressed sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5111–5130, Oct. 2010.

[39] Z. Ben-Haim, T. Michaeli, and Y. C. Eldar, "Performance bounds and design criteria for estimating finite rate of innovation signals," Sep. 2010 [Online]. Available: http://arxiv.org/pdf/1009.2221, preprint.

[40] E. Arias-Castro and Y. C. Eldar, "Noise folding in compressed sensing," *IEEE Signal Process. Lett.*, vol. 18, no. 8, pp. 478–481, 2011.

[41] T. T. Cai, G. Xu, and J. Zhang, "On recovery of sparse signals via ℓ_1 minimization," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3388–3397, July 2009.

[42] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2275–2284, Jun. 2009.

[43] R. A. DeVore, "Deterministic constructions of compressed sensing matrices," *J. Complex.*, vol. 23, no. 4, pp. 918–925, Aug. 2007.

[44] D. Donoho, "For most large underdetermined systems of linear equations, the minimal ℓ_1-norm solution is also the sparsest solution," *Comm. Pure Appl. Math.*, vol. 59, no. 6, pp. 797–829, 2006.

[45] E. J. Candès and Y. Plan, "Near-ideal model selection by ℓ_1 minimization," *Ann. Statist.*, vol. 37, no. 5A, pp. 2145–2177, Oct. 2009.

[46] R. G. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Construct. Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[47] D. L. Donoho and J. Tanner, "Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing," *Philos. Trans. Roy. Soc. London A, Math. Phys. Sci.*, vol. 367, no. 1906, pp. 4273–4293, Nov. 2009.

[48] C. Dossal, G. Peyré, and J. Fadili, "A numerical exploration of compressed sampling recovery," *Linear Algebra Its Appl.*, vol. 432, no. 7, pp. 1663–1679, Mar. 2010.

[49] S. Boyd and L. Vanderberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[50] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, pp. 259–268, Nov. 1992.

[51] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, Jun. 2010.

[52] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4036–4048, Sep. 2006.

[53] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, June 2008.

[54] E. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, vol. 35, no. 6, pp. 2313–2351, 2005.

[55] Y. Pati, R. Rezaifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," presented at the Asilomar Conf. Signals, Syst., Comput., Pacific Grove, CA, Nov. 1993.

[56] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, May 2008.

[57] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[58] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, Nov. 2008.

[59] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[60] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[61] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5030–5043, Oct. 2010.

[62] Z. Ben-Haim and Y. C. Eldar, "The Cramér-Rao bound for estimating a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3384–3389, June 2010.

[63] M. B. Wakin and M. A. Davenport, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4395–4401, Sep. 2010.

[64] T. Zhang, "Sparse recovery with orthogonal matching pursuit under RIP," May 2010 [Online]. Available: http://arxiv.org/pdf/1005.2249, preprint.

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

[66] M. E. Gehm, R. John, D. Brady, R. Willett, and T. J. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Opt. Exp.*, vol. 15, no. 21, pp. 14013–14027, Oct. 2007.

[67] M. A. Neifeld and J. Ke, "Optical architectures for compressive imaging," *Appl. Opt.*, vol. 46, no. 22, pp. 5293–5303, Jul. 2007.

[68] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Appl. Opt.*, vol. 47, no. 10, pp. B44–B51, Apr. 2008.

[69] E. J. Candès and J. K. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, 2007.

[70] Y. C. Eldar, "Uncertainty relations for shift-invariant analog signals," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5742–5757, Dec. 2009.

[71] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Comm. Pure Appl. Math.*, vol. 61, no. 8, pp. 1025–1171, Aug. 2008.

[72] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, Mar. 2008.

[73] E. J. Candès, J. K. 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, 2006.

[74] S. Gazit, A. Szameit, Y. C. Eldar, and M. Segev, "Super-resolution and reconstruction of sparse sub-wavelength images," *Opt. Exp.*, vol. 17, pp. 23920–23946, 2009.

[75] Y. Shechtman, S. Gazit, A. Szameit, Y. C. Eldar, and M. Segev, "Super-resolution and reconstruction of sparse images carried by incoherent light," *Opt. Lett.*, vol. 35, pp. 23920–23946, Apr. 2010.

[76] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, Mar. 2008.

[77] R. Coifman, F. Geshwind, and Y. Meyer, "Noiselets," *Appl. Comp. Harmon. Anal.*, vol. 10, pp. 27–44, 2001.

[78] R. A. DeVerse, R. R. Coifman, A. C. Coppi, W. G. Fateley, F. Geshwind, R. M. Hammaker, S. Valenti, F. J. Warner, and G. L. Davis, "Application of spatial light modulators for new modalities in spectrometry and imaging," in *Spectral Imaging: Instrumentation, Applications, and Analysis II*, San Jose, CA, Jan. 2003, vol. 4959, Proc. SPIE, pp. 12–22.

[79] W. L. Chan, K. Charan, D. Takhar, K. F. Kelly, R. G. Baraniuk, and D. M. Mittleman, "A single-pixel terahertz imaging system based on compressed sensing," *Appl. Phy. Lett.*, vol. 93, no. 12, Sep. 2008.

[80] P. Ye, J. L. Paredes, G. R. Arce, Y. Wu, C. Chen, and D. W. Prather, "Compressive confocal microscopy," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 429–432.

[81] S. Pfetsch, T. Ragheb, J. Laska, H. Nejati, A. Gilbert, M. Strauss, R. Baraniuk, and Y. Massoud, "On the feasibility of hardware implementation of sub-Nyquist random-sampling based analog-to-information conversion," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Seattle, WA, May 2008, pp. 1480–1483.

[82] A. C. Gilbert, M. J. Strauss, and J. A. Tropp, "A tutorial on fast Fourier sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 57–66, Mar. 2008.

[83] W. U. Bajwa, A. Sayeed, and R. Nowak, "A restricted isometry property for structurally subsampled unitary matrices," in *Proc. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, Sep. 2009, pp. 1005–1012.

[84] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 520–544, Jan. 2010.

[85] T. Ragheb, J. N. Laska, H. Nejati, S. Kirolos, R. G. Baraniuk, and Y. Massoud, "A prototype hardware for random demodulation based compressive analog-to-digital conversion," in *Proc. IEEE Midwest Symp. Circuits Syst.*, Knoxville, TN, Aug. 2008, pp. 37–40.

[86] Z. Yu, S. Hoyos, and B. M. Sadler, "Mixed-signal parallel compressed sensing and reception for cognitive radio," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Las Vegas, NV, Apr. 2008, pp. 3861–3864.

[87] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," Feb. 2010 [Online]. Available: http://www.math.princeton.edu/~mduarte/images/SCS-TSP.pdf, preprint.

[88] G. C. Valley and T. J. Shaw, "Compressive sensing recovery of sparse signals with arbitrary frequencies," *IEEE Trans. Signal Process.*, 2010, submitted for publication.

[89] J. D. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5862–5875, Jun. 2010.

[90] H. Rauhut, "Compressive sensing and structured random matrices," in *Theoretical Foundations and Numerical Methods for Sparse Recovery*, ser. Radon Series on Computational and Applied Mathematics, H. Rauhut, Ed. Berlin, Germany: De Gruyter, 2010, vol. 9, .

[91] H. Rauhut, J. K. Romberg, and J. A. Tropp, "Restricted isometries for partial random circulant matrices," *Appl. Comput. Harmon. Anal.*, May 2011, to be published.

[92] J. K. Romberg, "Compressive sensing by random convolution," *SIAM J. Imag. Sci.*, vol. 2, no. 4, pp. 1098–1128, Dec. 2009.

[93] R. F. Marcia and R. M. Willett, "Compressive coded aperture superresolution image reconstruction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Las Vegas, NV, March 2008, pp. 833–836.

[94] R. Marcia, Z. Harmany, and R. Willett, "Compressive coded aperture imaging," in *Computational Imaging VII*, San Jose, CA, Jan. 2009, vol. 7246, Proc. SPIE.

[95] L. Jacques, P. Vandergheynst, A. Bibet, V. Majidzadeh, A. Schmid, and Y. Leblebici, "CMOS compressed imaging by random convolution," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 1113–1116.

[96] M. F. Duarte and R. G. Baraniuk, "Kronecker compressive sensing," 2009 [Online]. Available: http://www.math.princeton.edu/~mduarte/images/KCS-TIP11.pdf, preprint.

[97] Y. Rivenson and A. Stern, "Compressed imaging with a separable sensing operator," *IEEE Signal Process. Lett.*, vol. 16, no. 6, pp. 449–452, Jun. 2009.

[98] T. Sun and K. F. Kelly, "Compressive sensing hyperspectral imager," presented at the Comput. Optical Sens. Imag. (COSI), San Jose, CA, Oct. 2009.

[99] R. Robucci, J. Gray, L. K. Chiu, J. K. Romberg, and P. Hasler, "Compressive sensing on a CMOS separable-transform image sensor," *Proc. IEEE*, vol. 98, no. 6, pp. 1089–1101, Jun. 2010.

[100] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

[101] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," Electr. Comput. Eng. Dept., Rice Univ., Houston, TX, Tech. Rep. TREE-0612, Nov. 2006.

[102] J. W. Phillips, R. M. Leahy, and J. C. Mosher, "MEG-based imaging of focal neuronal current sources," *IEEE Trans. Med. Imag.*, vol. 16, no. 3, pp. 338–348, June 1997.

[103] R. Gribonval, "Sparse decomposition of stereo signals with matching pursuit and application to blind separation of more than two sources from a stereo mixture," presented at the IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), Minneapolis, MN, Apr. 1993.

[104] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[105] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374–377, Mar. 2002.

[106] I. J. Fevrier, S. B. Gelfand, and M. P. Fitz, "Reduced complexity decision feedback equalization for multipath channels with large delay spreads," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 927–937, Jun. 1999.

[107] J. A. Bazerque and G. B. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1847–1862, Mar. 2010.

[108] M. Mishali and Y. C. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Process.*, vol. 57, pp. 993–1009, Mar. 2009.

[109] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 375–391, Apr. 2010.

[110] M. E. Davies and Y. C. Eldar, "Rank awareness in joint sparse recovery,"
Apr. 2010 [Online]. Available: http://arxiv.org/pdf/1004.4529, preprint.

[111] P. Feng, "Universal Minimum-rate sampling and spectrum-blind reconstruction for multiband signals," Ph.D. dissertation, Univ. of Illinois, Urbana, IL, 1998.

[112] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.

[113] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4692–4702, Oct. 2008.

[114] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, pp. 572–588, Apr. 2006.

[115] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Process.*, vol. 86, pp. 589–602, Apr. 2006.

[116] M. Fornasier and H. Rauhut, "Recovery algorithms for vector valued data with joint sparsity constraints," *SIAM J. Numer. Anal.*, vol. 46, no. 2, pp. 577–613, 2008.

[117] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.

[118] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, "Atoms of all channels, unite! Average case analysis of multi-channel sparse recovery using greedy algorithms," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 655–687, 2008.

[119] Y. C. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Trans. Inf. Theory*, vol. 6, no. 1, pp. 505–519, Jan. 2010.

[120] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, pp. 3042–3054, June 2010.

[121] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. RADC Spectral Estimation Workshop*, Rome, NY, Oct. 1979, pp. 243–258.

[122] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Atlanta, GA, May 1996, vol. 3, pp. 1688–1691.

[123] K. Schnass and P. Vandergheynst, "Average performance analysis for thresholding," *IEEE Signal Process. Lett.*, vol. 14, no. 11, pp. 828–831, Nov. 2007.

[124] W. Ou, M. S. Hämäläinen, and P. Golland, "A distributed spatio-temporal EEG/MEG inverse solver," *Neuroimage*, vol. 44, no. 3, pp. 932–946, Feb. 2009.

[125] V. Cevher, M. F. Duarte, C. Hegde, and R. G. Baraniuk, "Sparse signal recovery using Markov random fields," presented at the Workshop Neural Info. Process. Syst. (NIPS), Vancouver, Canada, Dec. 2008.

[126] P. Schniter, "Turbo reconstruction of structured sparse signals," presented at the Conf. Inf. Sci. Syst. (CISS), Princeton, NJ, Mar. 2010.

[127] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 269–280, Jan. 2010.

[128] L. He and L. Carin, "Exploiting structure in wavelet-based Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3488–3497, Sep. 2009.

[129] S. Ji, D. Dunson, and L. Carin, "Multi-task compressive sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 92–106, Jan. 2009.

[130] L. He, H. Chen, and L. Carin, "Tree-structured compressive sensing with variational Bayesian analysis," *IEEE Signal Process. Lett.*, vol. 17, no. 3, pp. 233–236, Mar. 2010.

[131] T. Faktor, Y. C. Eldar, and M. Elad, "Exploiting statistical dependencies in sparse representations for signal recovery," Oct. 2010 [Online]. Available: http://arxiv.org/pdf/1010.5734, preprint.

[132] R. Jenatton, J. Y. Audibert, and F. Bach, "Structured variable selection with sparsity inducing norms," Apr. 2009 [Online]. Available: http://arxiv.org/pdf/0904.3523, preprint.

[133] L. Peotta and P. Vandergheynst, "Matching pursuit with block incoherent dictionaries," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4549–4557, Sep. 2007.

[134] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Stat. Soc. Ser. B Stat. Methodol.*, vol. 68, no. 1, pp. 49–67, 2006.

[135] Y. C. Eldar and T. Michaeli, "Beyond bandlimited sampling," *IEEE Signal Process. Mag.*, vol. 26, no. 3, pp. 48–68, May 2009.

[136] F. R. Bach, "Consistency of the group lasso and multiple kernel learning," *J. Mach. Learn. Res.*, vol. 9, pp. 1179–1225, 2008.

[137] Y. Nardi and A. Rinaldo, "On the asymptotic properties of the group lasso estimator for linear models," *Electron. J. Statist.*, vol. 2, pp. 605–633, 2008.

[138] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3075–3085, Aug. 2009.

[139] L. Meier, S. van de Geer, and P. Bühlmann, "The group lasso for logistic regression," *J. R. Statist. Soc. B*, vol. 70, no. 1, pp. 53–77, 2008.

[140] S. Erickson and C. Sabatti, "Empirical Bayes estimation of a sparse vector of gene expression changes," *Statist. Appl. Genetics Mol. Biol.*, vol. 4, no. 1, p. 22, 2005.

[141] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi, "Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 3, pp. 275–285, Jun. 2008.

[142] C. Hegde, M. F. Duarte, and V. Cevher, "Compressive sensing recovery of spike trains using structured sparsity," presented at the Workshop Signal Process. Adapt.. Sparse Structured Representations (SPARS), Saint Malo, France, Apr. 2009.

[143] V. Cevher, P. Indyk, C. Hegde, and R. G. Baraniuk, "Recovery of clustered sparse signals from compressive measurements," presented at the Int. Conf. Sampling Theory Appl. (SAMPTA), Marseille, France, May 2009.

[144] M. F. Duarte, V. Cevher, and R. G. Baraniuk, "Model-based compressive sensing for signal ensembles," presented at the Allerton Conf. Commun., Control, Comput., Monticello, IL, Sep. 2009.

[145] Z. Ben-Haim and Y. C. Eldar, "Near-oracle performance of greedy block-sparse estimation techniques from noisy measurements," *IEEE J. Sel. Topics Signal Process.*, 2010 [Online]. Available: http://arxiv.org/pdf/1009.0906, to be published

[146] J. Friedman, T. Hastie, and R. Tibshirani, "A note on the group Lasso and a sparse group Lasso," Jan. 2010 [Online]. Available: http://arxiv.org/pdf/1001.0736, preprint.

[147] P. Sprechmann, I. Ramirez, G. Sapiro, and Y. C. Eldar, "C-HiLasso: A collaborative hierarchical sparse modeling framework," *IEEE Trans. Signal Process.*, Jun. 2010 [Online]. Available: http://arxiv.org/abs/1006.1346, preprint, to be published

[148] A. W. Habboosh, R. J. Vaccaro, and S. Kay, "An algorithm for detecting closely spaced delay/Doppler components," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Munich, Germany, Apr. 1997, pp. 535–538.

[149] W. U. Bajwa, A. M. Sayeed, and R. Nowak, "Learning sparse doubly-selective channels," in *Proc. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, Sep. 2008, pp. 575–582.

[150] W. U. Bajwa, K. Gedalyahu, and Y. C. Eldar, "Identification of parametric underspread linear systems with application to super-resolution radar," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2548–2561, Aug. 2011.

[151] M. Unser, "Sampling—50 years after Shannon," *Proc. IEEE*, vol. 88, pp. 569–587, Apr. 2000.

[152] C. de Boor, R. DeVore, and A. Ron, "The structure of finitely generated shift-invariant spaces in $L_2(\mathbb{R}^d)$," *J. Funct. Anal.*, vol. 119, no. 1, pp. 37–78, 1994.

[153] A. Ron and Z. Shen, "Frames and stable bases for shift-invariant subspaces of $\mathrm{L}_2(R^d)$," *Can. J. Math.*, vol. 47, no. 5, pp. 1051–1094, 1995.

[154] A. Aldroubi and K. Gröchenig, "Non-uniform sampling and reconstruction in shift-invariant spaces," *SIAM Rev.*, vol. 43, pp. 585–620, 2001.

[155] Y.-P. Lin and P. P. Vaidyanathan, "Periodically nonuniform sampling of bandpass signals," *IEEE Trans. Circuits Syst. II*, vol. 45, no. 3, pp. 340–351, Mar. 1998.

[156] C. Herley and P. W. Wong, "Minimum rate sampling and reconstruction of signals with arbitrary frequency support," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1555–1564, July 1999.

[157] J. Mitola, III, "Cognitive radio for flexible mobile multimedia communications," *Mobile Netw. Appl.*, vol. 6, no. 5, pp. 435–441, 2001.

[158] H. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Mathematica*, vol. 117, no. 1, pp. 37–52, Jul. 1967.
[159] A. Kohlenberg, "Exact interpolation of band-limited functions," *J. Appl. Phys.*, vol. 24, no. 12, pp. 1432–1436, Dec. 1953.
[160] R. Venkataramani and Y. Bresler, "Perfect reconstruction formulas and bounds on aliasing error in sub-Nyquist nonuniform sampling of multiband signals," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 2173–2183, Sep. 2000.
[161] M. Mishali and Y. C. Eldar, "Expected-RIP: Conditioning of the modulated wideband converter," in *IEEE Inf. Theory Workshop (ITW)*, Taormina, Italy, Oct. 2009, pp. 343–347.
[162] Y. Chen, M. Mishali, Y. C. Eldar, and A. O. Hero, III, "Modulated wideband converter with non-ideal lowpass filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Dallas, TX, Apr. 2010, pp. 3630–3633.
[163] R. Tur, Y. C. Eldar, and Z. Friedman, "Innovation rate sampling of pulse streams with application to ultrasound imaging," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1827–1842, Apr. 2011.
[164] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1997.
[165] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 70, pp. 1272–1281, May 1990.
[166] S. Y. Kung, K. S. Arun, and B. D. Rao, "State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem," *J. Opt. Soc. Amer.*, vol. 73, no. 122, pp. 1799–1811, Dec. 1983.
[167] B. D. Rao and K. S. Arun, "Model based processing of signals: A state space approach," *Proc. IEEE*, vol. 80, no. 2, pp. 283–309, 1992.
[168] J. Kusuma and V. K. Goyal, "Multichannel sampling of parametric signals with a successive approximation property," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2006, pp. 1265–1268.
[169] C. S. Seelamantula and M. Unser, "A generalized sampling method for finite-rate-of-innovation-signal reconstruction," *IEEE Signal Process. Lett.*, vol. 15, pp. 813–816, 2008.
[170] K. Gedalyahu, R. Tur, and Y. C. Eldar, "Multichannel sampling of pulse streams at the rate of innovation," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1491–1504, Apr. 2011.
[171] L. Baboulaz and P. L. Dragotti, "Exact feature extraction using finite rate of innovation principles with an application to image super-resolution," *IEEE Trans. Image Process.*, vol. 18, no. 2, pp. 281–298, Feb. 2009.
[172] N. Wagner, Y. C. Eldar, A. Feuer, G. Danin, and Z. Friedman, "Xampling in ultrasound imaging," in *Medical Imaging: Ultrasonic Imaging, Tomography, and Therapy*, Lake Buena Vista, FL, 2011, vol. 7968, Proc. SPIE.
[173] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.
[174] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, July 1996.
[175] M. Z. Win and R. A. Scholtz, "Characterization of ultra-wide bandwidth wireless indoor channels: A communication-theoretic view," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 9, pp. 1613–1627, Dec. 2002.
[176] A. Quazi, "An overview on the time delay estimate in active and passive systems for target localization," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 3, pp. 527–533, 1981.
[177] R. J. Urick, *Principles of Underwater Sound*. New York: McGraw-Hill, 1983.
[178] A. Bruckstein, T. J. Shan, and T. Kailath, "The resolution of overlapping echos," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 6, pp. 1357–1367, 1985.
[179] M. A. Pallas and G. Jourdain, "Active high resolution time delay estimation for large BT signals," *IEEE Trans. Signal Process.*, vol. 39, no. 4, pp. 781–788, Apr. 1991.
[180] F.-X. Ge, D. Shen, Y. Peng, and V. O. K. Li, "Super-resolution time delay estimation in multipath environments," *IEEE Trans. Circuits Syst. I*, vol. 54, no. 9, pp. 1977–1986, Sep. 2007.
[181] L. Borup, R. Gribonval, and M. Nielsen, "Beyond coherence: Recovering structured time-frequency representations," *Appl. Comp. Harmon. Anal.*, vol. 24, no. 1, pp. 120–128, Jan. 2008.
[182] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," *Found. Comput. Math.*, vol. 9, no. 1, pp. 51–77, Jan. 2009.
[183] E. J. Candès and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, Jun. 2010.
[184] A. Hansen, "Generalized sampling and infinite dimensional compressed sensing," Feb. 2011 [Online]. Available: http://www.damtp.cam.ac.uk/user/na/people/Anders/Inf_CS43.pdf, preprint.

Marco F. Duarte (S'99–M'09) received the B.Sc. degree in computer engineering (with distinction) and the M.Sc. degree in electrical engineering from the University of Wisconsin-Madison in 2002 and 2004, respectively, and the Ph.D. degree in electrical engineering from Rice University, Houston, TX, in 2009.

He was an NSF/IPAM Mathematical Sciences Postdoctoral Research Fellow in the Program of Applied and Computational Mathematics at Princeton University, Princeton, NJ, from 2009 to 2010, and in the Department of Computer Science at Duke University, Durham, NC, from 2010 to 2011. He is currently an Assistant Professor in the Department of Electrical and Computer Engineering at the University of Massachusetts, Amherst, MA. His research interests include machine learning, compressed sensing, sensor networks, and optical image coding.

Dr. Duarte received the Presidential Fellowship and the Texas Instruments Distinguished Fellowship in 2004 and the Hershel M. Rich Invention Award in 2007, all from Rice University. He coauthored (with C. Hegde and V. Cevher) the Best Student Paper at the 2009 International Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS). He is also a member of Tau Beta Pi.

Yonina C. Eldar (S'98–M'02–SM'07) received the B.Sc. degree in physics and the B.Sc. degree in electrical engineering both from Tel-Aviv University (TAU), Tel-Aviv, Israel, in 1995 and 1996, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 2002.

From January 2002 to July 2002, she was a Postdoctoral Fellow at the Digital Signal Processing Group at MIT. She is currently a Professor in the Department of Electrical Engineering at the Technion—Israel Institute of Technology, Haifa. She is also a Research Affiliate with the Research Laboratory of Electronics at MIT and a Visiting Professor at Stanford University, Stanford, CA. Her research interests are in the broad areas of statistical signal processing, sampling theory and compressed sensing, optimization methods, and their applications to biology and optics.

Dr. Eldar was in the program for outstanding students at TAU from 1992 to 1996. In 1998, she held the Rosenblith Fellowship for study in electrical engineering at MIT, and in 2000, she held an IBM Research Fellowship. From 2002 to 2005, she was a Horev Fellow of the Leaders in Science and Technology program at the Technion and an Alon Fellow. In 2004, she was awarded the Wolf Foundation Krill Prize for Excellence in Scientific Research, in 2005 the Andre and Bella Meyer Lectureship, in 2007 the Henry Taub Prize for Excellence in Research, in 2008 the Hershel Rich Innovation Award, the Award for Women with Distinguished Contributions, the Muriel & David Jacknow Award for Excellence in Teaching, and the Technion Outstanding Lecture Award, in 2009 the Technion's Award for Excellence in Teaching, in 2010 the Michael Bruno Memorial Award from the Rothschild Foundation, and in 2011 the Weizmann Prize for Exact Sciences. She is a member of the IEEE Signal Processing Theory and Methods technical committee and the Bio Imaging Signal Processing technical committee, an Associate Editor for the SIAM *Journal on Imaging Sciences*, and on the Editorial Board of *Foundations and Trends in Signal Processing*. In the past, she served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the EURASIP *Journal of Signal Processing*, and the SIAM *Journal on Matrix Analysis and Applications*.