

From Theory to Practice: Sub-Nyquist Sampling of Sparse Wideband Analog Signals

Moshe Mishali, *Student Member, IEEE*, and Yonina C. Eldar, *Senior Member, IEEE*

Abstract—Conventional sub-Nyquist sampling methods for analog signals exploit prior information about the spectral support. In this paper, we consider the challenging problem of spectrum-blind sub-Nyquist sampling of multiband signals. The Fourier transform of such signals occupy only a small portion of a wide spectrum, with unknown frequency support. Our primary design goals are efficient hardware implementation and low computational load on the supporting digital processing. We suggest a system, named the modulated wideband converter, which first multiplies the analog signal by a bank of periodic waveforms. The product is then lowpass filtered and sampled uniformly at low rate. We derive necessary and sufficient conditions ensuring perfect recovery from the proposed system. In particular, the waveform period and the uniform rate can be made as low as the expected width of each band, which is orders of magnitude smaller than the Nyquist rate. Reconstruction relies on recent ideas developed in the context of analog compressed sensing, and is comprised of a digital step which recovers the spectral support. Our approach enables baseband processing, namely generating a low rate sequence corresponding to any information band of interest from the given samples, without going through the high Nyquist rate. Numerical simulations demonstrate robustness as well as several further hardware simplifications. In particular, the ideas can be applied to a single channel with a higher sampling rate. We compare our system with two previous approaches: periodic nonuniform sampling, which is bandwidth limited by existing hardware devices, and the random demodulator, which is sensitive to parameter choice, has a high computational load, and is restricted to multitone signals. In addition, both these methods do not allow baseband processing. In the broader context of Nyquist sampling, our scheme has the potential to break through the bandwidth barrier of state-of-the-art analog conversion technologies such as interleaved converters.

Index Terms—Analog to digital conversion, compressive sampling, infinite measurement vectors (IMV), multiband sampling, spectrum-blind reconstruction, sub-Nyquist sampling.

I. INTRODUCTION

Radio frequency (RF) technology enables the modulation of narrowband signals by high carrier frequencies. Consequently, manmade radio signals are often sparse. That is, they consist of a relatively small number of narrowband transmissions spread across a wide spectrum range. A convenient way to describe this class of signals is through a multiband model. The frequency support of a multiband signal resides within several continuous intervals spread over a wide spectrum. Figure 1 depicts a typical communication application, the wideband receiver, in which the received signal follows the

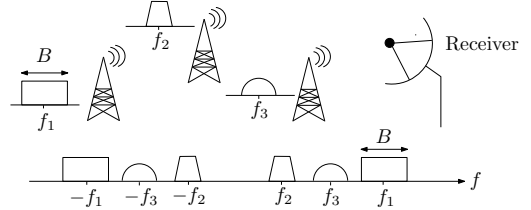


Fig. 1. Three RF transmissions with different carriers f_i . The receiver sees a multiband signal (bottom drawing).

multiband model. The basic operations in such an application are conversion of the incoming signal to digital, and low-rate processing of some or all of the individual transmissions. Ultimately, the digital product is transformed back to the analog domain for further transmission.

Due to the wide spectral range of multiband signals, their Nyquist rates may exceed the specifications of the best analog-to-digital converters (ADCs) by orders of magnitude. Any attempt to acquire a multiband signal must therefore exploit its structure in an intelligent way. When the carrier frequencies are known, a common practical engineering approach is to employ baseband processing, in which an analog front-end modulates the signal by a carrier frequency such that the spectral contents of a band of interest are centered around the origin. A lowpass filter follows in order to reject frequencies due to the other bands. Conversion to digital is then performed at a rate matching the actual information width of the band of interest. Repeating the process for each band separately results in a sampling rate which is the sum of the band widths. This method achieves the minimal sampling rate, as derived by Landau [1], which is equal to twice the actual frequency occupancy. An alternative sampling approach that does not require analog preprocessing was proposed in [2]. In this strategy, periodic nonuniform sampling is used to directly sample a multiband signal at an average rate approaching that derived by Landau. Both baseband processing and the method of [2] rely on knowledge of the carrier frequencies.

In scenarios in which the carrier frequencies are unknown to the receiver, or change with time, a challenging task is to design a *spectrum-blind* receiver at a sub-Nyquist rate. In [3], [4] a multicore sampling strategy was developed, independent of the signal support, to acquire multiband signals at low rates. Although the sampling method is blind, in order to recover the original signal from the samples, knowledge of the frequency support is needed. Recently [5], we proposed a fully spectrum-blind system based on multicore sampling. Our system does not require knowledge of the frequency support in both the sampling and the recovery stages. To reconstruct

This work has been submitted to the IEEE for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible.

Moshe Mishali and Yonina C. Eldar are with the Technion—Israel Institute of Technology, Haifa Israel. Emails: moshiko@tx.technion.ac.il, yonina@ee.technion.ac.il.

the signal blindly, we developed digital algorithms that process the samples and identify the unknown spectral support. Once the support is found, the continuous signal is reconstructed using closed-form expressions.

Periodic nonuniform sampling is a popular approach in the broader context of wideband analog conversion when the spectrum is fully occupied. Instead of implementing a single ADC at a high-rate R , interleaved ADCs use M devices at rate R/M with appropriate time shifts [6]–[8]. However, time interleaving has two fundamental limitations. First, the M lowrate samplers have to share an analog front-end which must tolerate the input bandwidth R . With today’s technology the possible front-ends are still far below the wideband regime. Second, maintaining accurate time shifts, on the order of $1/RM$, is difficult to implement. Multicoset sampling, is a special case of interleaved ADC, in which only some of the M branches are used. Consequently, the same limitations apply. In Section II-B we discuss in more detail the difficulty in implementing interleaved ADCs and multicoset sampling. In practice, such systems are limited to intermediate input frequencies and cannot deal with wideband inputs.

Recently, a new architecture to acquire analog signals, called the random demodulator, was studied in the literature of compressed sensing (CS) [9], [10]. In this approach, the signal is modulated by a high-rate pseudorandom number generator, integrated, and sampled at a low rate. This scheme can be implemented using standard hardware devices. The random demodulator was studied for a multitone model, which contains signals with a discrete and finite set of harmonics. Using time-domain techniques and under additional assumptions, polynomial-time algorithms can recover such a multitone signal from the samples of the random demodulator [10]. However, as elaborated on in Section VI, the discrete approach leads to a solution which is sensitive to design imperfections or even slight model mismatches. Furthermore, the time-domain analysis in [10] aims at the recovery of the Nyquist rate samples, which results in a huge-scale recovery problem. This leads to severe computational loads even with pure multitone signals. The time-domain approach also precludes baseband processing since interpolation to the Nyquist rate is an essential ingredient in the reconstruction. Finally, the model cannot properly treat true analog inputs.

A. Main contributions

In this work we aim to combine the advantages of all previous approaches: The ability to treat analog multiband models, a sampling stage with a practical implementation, and a spectrum-blind recovery stage which involves efficient digital processing. In addition, we would like a method that will allow for baseband processing, namely the ability to process any one of the transmitted bands without first requiring interpolation to the high Nyquist rate.

Our first contribution is an analog system, referred to as the modulated wideband converter, which is comprised of a bank of modulators and lowpass filters. Each channel resembles the random demodulator of [9] with a few essential differences. The signal is multiplied by a periodic waveform, whose period

corresponds to the multiband model parameters. A square-wave alternating at the Nyquist rate is one choice; other possibilities and their advantages are also discussed. The goal of the modulator is to alias the spectrum into baseband. The modulated output is then lowpass filtered, and sampled at a low rate. The rate can be as low as the expected width of an individual transmission. We prove that an appropriate choice of the parameters (waveform period, sampling rate) guarantees that our system uniquely determines a multiband input signal. In addition, we describe how to trade the number of channels by a higher rate in each branch, at the expense of additional processing. Theoretically, this method allows to collapse the entire system to a single channel operating at a rate lower than Nyquist. Some essential differences and advantages with respect to the random demodulator are detailed in Section VI.

Our second contribution is a recovery stage complementing the sampling. Our algorithm relies on the machinery of blind reconstruction, which we originally developed for multicoset samples [5], and later extended to a broader setting in [11]. In contrast to [10], we use frequency domain analysis, a standard tool in sampling theory, to express the relation between the sample sequences and the unknown signal. As we show, the frequency domain viewpoint allows to control the dimensions of the recovery problem. A representative wideband scenario is simulated in Section V-A and demonstrates recovery from a low scale problem (up to 9 orders of magnitude below the counterpart recovery of [10]). In addition, our method separates the support recovery from the actual reconstruction of the analog input. This enables baseband processing, namely generating a low rate sequence for any specific band of interest, without going through recovery of the Nyquist rate sequence. In contrast, the recovery formulation in [9], [10] couples the support and signal recovery. This results in CS problems of large dimensions, and precludes low rate processing. Section VI highlights other advantages of our method in terms of digital processing.

The current paper focuses on the engineering aspects of our system. Numerical simulations are used to prove the concept and the stability of the solution in the presence of noise. Ongoing work studies the theoretical aspects in more depth [12].

B. Outline

The paper is organized as follows. Section II describes the multiband model and defines the goals for a practical sampling stage. We also detail limitations of multicoset in the wideband regime. In Section III, we describe the modulated wideband sampling system and provide a frequency-domain analysis of the resulting samples. This leads to a concrete parameter selection which guarantees a unique signal matching the digital samples. We conclude the section with a discussion of the tradeoff between number of channels, rate, and complexity. Recovery is discussed in Section IV. A design example and additional implementation aspects are presented in Section V. In Section VI we compare our system with multicoset sampling and the random demodulator.

II. FORMULATION AND BACKGROUND

A. Problem formulation

Let $x(t)$ be a real-valued continuous-time signal in L_2 , with bounded energy. Throughout the paper, continuous signals are assumed to be bandlimited to $\mathcal{F} = [-1/2T, +1/2T)$. Formally, the Fourier transform of $x(t)$, which is defined by

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt, \quad (1)$$

is zero for every $f \notin \mathcal{F}$. We denote by $f_{\text{NYQ}} = 1/2T$ the Nyquist rate of $x(t)$. For technical reasons, it is also assumed that $X(f)$ is piecewise continuous in f .

We treat signals from the multiband model \mathcal{M} defined below.

Definition 1: The set \mathcal{M} contains all signals $x(t)$, such that the support of the Fourier transform $X(f)$ is contained within a union of N disjoint intervals (bands) in \mathcal{F} , and each of the band widths does not exceed B .

Signals in \mathcal{M} have an even number N of bands due to the conjugate symmetry of $X(f)$. The band positions are arbitrary, and in particular, unknown in advance. A typical spectral support of a signal from the multiband model \mathcal{M} is illustrated in the example of Fig. 1, in which $N = 6$ and B, f_{NYQ} are dictated according to the specifications of the possible transmitters.

We wish to design a sampling system for signals from the model \mathcal{M} that satisfies the following properties:

- 1) The sampling rate should be as low as possible;
- 2) the system has no prior knowledge of the band locations; and
- 3) the system can be implemented with existing analog devices and (preferably low rate) ADCs.

The design of the sampling stage should match a corresponding reconstruction stage, which converts the discrete samples back to the continuous-time domain. This stage may involve digital processing prior to reconstruction. An implicit (but crucial) requirement is that recovery involves a reasonable amount of computations. Realtime applications may also necessitate short latency from input to output and a constant throughput. Therefore, two main factors dictate the input spectrum range that the overall system can handle: analog hardware at the required frequency that can convert the signal to digital, and a digital stage that can accommodate the computation load.

In our previous work [5], we proved that the minimal sampling rate for \mathcal{M} to allow perfect blind reconstruction is $2NB$, provided that $2NB$ is lower than the Nyquist rate. The case $2NB \geq f_{\text{NYQ}}$ represents signals which occupy more than half of the Nyquist range. No rate improvement is possible in that case (for arbitrary signal), and thus we assume $2NB < f_{\text{NYQ}}$ in the sequel. Concrete algorithms for blind recovery, achieving the minimal rate, were developed in [5] based on a multicoset sampling strategy. The next section briefly describes this method, which achieves the goals of minimal rate and blindness. We also explain the limitations of practical ADCs which influence the applicability of multicoset to sampling wideband signals. As described

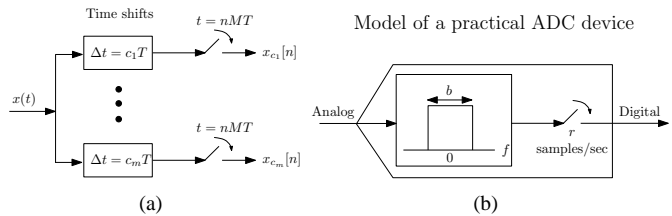


Fig. 2. Schematic implementation of multicoset sampling (a) requires no filtering between the time shifts and the actual sampling. However, the front-end of a practical ADC has an inherent bandwidth limitation, which is modeled in (b) as a lowpass preceding the uniform sampling.

later in Section III-A, the sampling scheme proposed in this paper circumvents these limitations and has other advantages in terms of practical implementation.

B. Multicoset using practical ADCs

In multicoset sampling, samples of $x(t)$ are obtained on a periodic and nonuniform grid which is a subset of the Nyquist grid. Formally, denote $x(nT)$ as the sequence of samples taken at the Nyquist rate. Let M be a positive integer, and $C = \{c_i\}_{i=1}^m$ be a set of m distinct integers with $0 \leq c_i \leq M - 1$. The multicoset grid consists of m uniform sequences, called cosets, such that the i th coset is defined by

$$x_{c_i}[n] = x(nMT + c_i T), \quad n \in \mathbb{Z}. \quad (2)$$

Only $m < M$ cosets are used, and thus the average sampling rate is m/MT , which is lower than the Nyquist rate $1/T$.

A possible implementation of the sampling sequences (2) is depicted in Fig. 2(a). The building blocks are m uniform samplers at rate $1/MT$, where the i th sampler is shifted by $c_i T$ from the origin. Although this scheme seems intuitive and straightforward, practical ADCs introduce inherent bandwidth limitation, which is modeled as a preceding lowpass filter in Fig. 2(b). This additional constraint becomes crucial for high rate inputs. To understand this limitation, consider the uniform sampler block, which should output pointwise samples of the input at rate $1/MT$. An appropriate choice for this task is an ADC device with sampling rate $r = 1/MT$. Practical ADCs have an additional property, termed analog (full-power) bandwidth [13], which determines the maximal frequency b that the device can handle, as dictated by technology. Any spectral content beyond b Hz is attenuated and distorted. Manufacturers thus usually recommend adding a preceding anti-aliasing lowpass filter, with cutoff b to further reject higher frequencies. The ratio b/r implies the complexity of the ADC circuit design, and is typically in the range $1.5r \leq b \leq 7r$; see the online table in [14]. A simplified model based on the specifications b, r is depicted in Fig. 2(b).

The practical ADC model raises two issues. First, RF technology allows transmissions at rates which exceed the analog bandwidth b of state-of-the-art devices, typically by orders of magnitude. For example, ADC devices manufactured by Analog Devices Corp. have front-end bandwidths which reach up to $b = 780$ MHz [14]. Therefore, any attempt to acquire a wideband signal with a practical ADC results in a loss of the spectral contents beyond b Hz. The sample

sequences (2) are attenuated and distorted and are no longer pointwise values of $x(t)$. This limitation is fundamental and holds even in other architectures of multicaset (*e.g.*, a single ADC triggered by a nonuniform clock). The second issue is a waste of resources which is less severe but applies also when the Nyquist rate $f_{\text{NYQ}} = b$ for some available device. For a signal with a sparse spectrum, multicaset reduces the average sampling rate by using only m out of M possible cosets, where $M \gg 1$ is commonly used. Each coset in Fig. 2 samples at rate f_{NYQ}/M . Therefore, the ADC samples at rate b/M , which is far less than standard ratios b/r [14]. This implies sampling at a rate which is much lower than the maximal capability of the ADC.

As a consequence, implementing multicaset for wideband signals requires the design of a specialized fine-tuned ADC circuit, in order to meet the wide analog bandwidth, and still exploit the nonstandard ratio b/r that is expected. Though this may be an interesting task for experts, it contradicts the basic goal of our design - that is, using standard and available devices. In [15] a nonconventional ADC is designed by means of high-rate optical devices. The hybrid optic-electronic system introduces a front-end whose bandwidth reaches the wideband regime, at the expense and size of an optical system. Unfortunately, at present, this performance cannot be achieved with purely electronic technology.

Another practical issue of multicaset sampling, which also exists in the optical implementation, arises from the time shift elements. Maintaining accurate time delays between the ADCs in the order of the Nyquist interval T is difficult. Any uncertainty in these delays influences the recovery from the sampled sequences [16]. A variety of different algorithms have been proposed in the literature in order to compensate for timing mismatches. However, this adds substantial complexity to the receiver [17], [18].

III. SAMPLING

We now present an alternative sampling scheme that uses available devices and does not require non-zero time synchronization. The system is schematically drawn in Fig. 3 with its various parameters. In the next subsections, the scheme is described and analyzed for arbitrary sets of parameters. Later on, we specify a parameter choice, independent of the band locations, that approaches the minimal rate. The resulting system satisfies all requirements of our problem formulation.

A. System description

Our system exploits spread-spectrum techniques from communication theory [19], [20], and is partially inspired by the use of these techniques in [9], [10]. An analog mixing front-end aliases the spectrum, such that a portion of energy from each band appears in baseband. The system comprises several channels implementing different mixtures, so that, in principle, a sufficiently large number of mixtures allows to recover a relatively sparse multiband signal.

More specifically, the signal $x(t)$ enters m channels simultaneously. In the i th channel, $x(t)$ is multiplied by a mixing function $p_i(t)$, which is T_p -periodic. After mixing, the signal

spectrum is truncated by a lowpass filter with cutoff $1/(2T_s)$ and the filtered signal is sampled at rate $1/T_s$. Standard ADCs can be used for that task. The design parameters are therefore the number of channels m , the period T_p , the sampling rate $1/T_s$ and the mixing functions $p_i(t)$ for $1 \leq i \leq m$.

In the sequel, $p_i(t)$ is chosen as a piecewise constant function that alternates between the levels ± 1 for each of M equal time intervals. Formally,

$$p_i(t) = \alpha_{ik}, \quad k \frac{T_p}{M} \leq t \leq (k+1) \frac{T_p}{M}, \quad 0 \leq k \leq M-1, \quad (3)$$

with $\alpha_{ik} \in \{+1, -1\}$, and $p_i(t + nT_p) = p_i(t)$ for every $n \in \mathbb{Z}$. Other choices for $p_i(t)$ are possible; Some examples are described in Section VI.

The system proposed in Fig. 3 has several advantages for practical implementation:

- (A1) Analog mixers are a provable technology in the wideband regime [21], [22]. In fact, since transmitters use mixers to modulate the information by a high-carrier frequency, the mixer bandwidth defines the input bandwidth.
- (A2) Sign alternating functions can be implemented by a standard (high rate) shift register. Today's technology allows to reach alternation rates of 23 GHz [23] and even 80 GHz [24].
- (A3) Analog filters are accurate and typically do not require more than a few passive elements (*e.g.*, capacitors and coils) [25].
- (A4) The sampling rate $1/T_s$ matches the cutoff of $H(f)$. Therefore, an ADC with a conversion rate $r = 1/T_s$, and any bandwidth $b \geq 0.5r$ can be used to implement this block, where $H(f)$ serves as a preceding anti-aliasing filter. In the sequel, we choose $1/T_s$ on the order of B , which is the width of a single band of $x(t) \in \mathcal{M}$. In practice, this sampling rate allows flexible choice of an ADC from a variety of commercial devices in the low rate regime.
- (A5) Sampling is synchronized in all channels, that is there are no time shifts. This is beneficial since the trigger for all ADCs can be generated accurately (*e.g.*, with a zero-delay synchronization device [26]). Moreover, the sampling sequences can be fed directly to a digital processor which operates at rate $1/T_s$. In contrast, the multicaset samples (2) arrive at different time points, and additional hardware is required to buffer the sequences before entering the digital part.

B. Frequency domain analysis

We now derive the relation between the sample sequences $y_i[n]$ and the unknown signal $x(t)$. This analysis is used for several purposes in the following sections. First, for specifying a choice of parameters ensuring a unique mapping between $x(t)$ and the sequences $y_i[n]$. Second, we use this analysis to explain the reconstruction scheme. Finally, stability and implementation issues will also be based on this development.

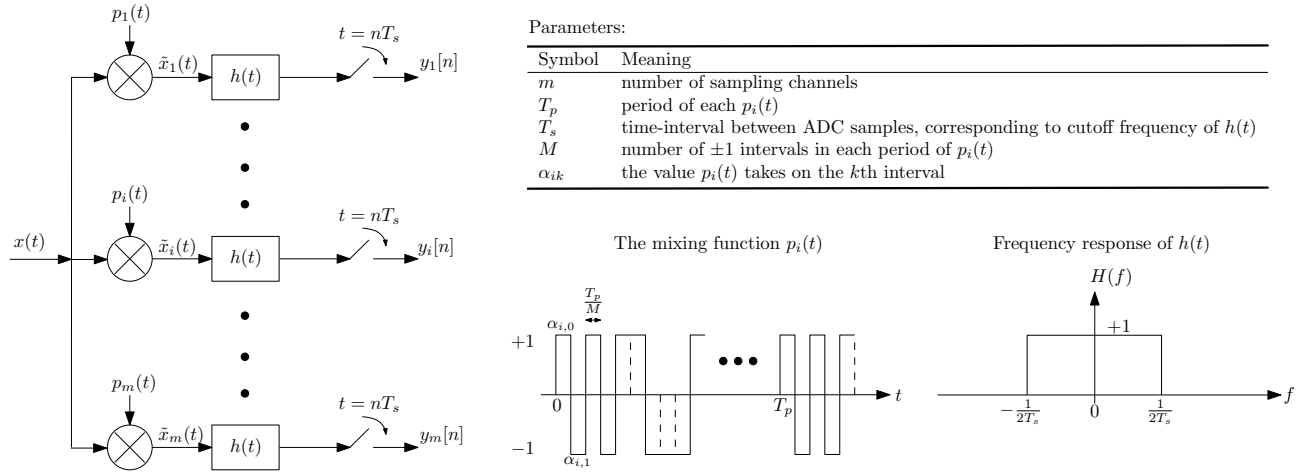


Fig. 3. The modulated wideband converter - a practical sampling stage for multiband signals.

To this end, we introduce the definitions

$$f_s = 1/T_s, \quad \mathcal{F}_s = [-f_s/2, +f_s/2] \quad (4a)$$

$$f_p = 1/T_p, \quad \mathcal{F}_p = [-f_p/2, +f_p/2]. \quad (4b)$$

Consider the i th channel. Since $p_i(t)$ is T_p -periodic, it has a Fourier expansion

$$p_i(t) = \sum_{l=-\infty}^{\infty} c_{il} e^{j\frac{2\pi}{T_p}lt}, \quad (5)$$

where

$$c_{il} = \frac{1}{T_p} \int_0^{T_p} p_i(t) e^{-j\frac{2\pi}{T_p}lt} dt. \quad (6)$$

Expressing the Fourier transform $P_i(f)$ in terms of the Fourier series coefficients c_{il} leads to

$$P_i(f) = \int_{-\infty}^{\infty} p_i(t) e^{-j2\pi ft} dt = \sum_{l=-\infty}^{\infty} c_{il} \delta(f - lf_p), \quad (7)$$

with $\delta(t)$ denoting the Dirac delta function. The analog multiplication $x(t)p_i(t)$ translates to a convolution in the frequency domain,

$$X(f) * P_i(f) = \sum_{l=-\infty}^{\infty} c_{il} X(f - lf_p). \quad (8)$$

Therefore, the input to $H(f)$ is a linear combination of f_p -shifted copies of $X(f)$. Since $X(f) = 0$ for $f \notin \mathcal{F}$, the above sum contains (at most) $\lceil f_{\text{NYQ}}/f_p \rceil$ nonzero terms¹, for every f .

The filter $H(f)$ has a frequency response which is an ideal rectangular function, as depicted in Fig. 3. Consequently, only frequencies in the interval \mathcal{F}_s pass to the uniform sequence $y_i[n]$. Thus the discrete-time Fourier transform (DTFT) of the

i th sequence $y_i[n]$ is expressed as

$$\begin{aligned} Y_i(e^{j2\pi f T_s}) &= \sum_{n=-\infty}^{\infty} y_i[n] e^{-j2\pi f n T_s} \\ &= \sum_{l=-L_0}^{+L_0} c_{il} X(f - lf_p), \quad f \in \mathcal{F}_s, \end{aligned} \quad (9)$$

where \mathcal{F}_s is defined in (4b), and L_0 is chosen as the smallest integer such that the above sum contains all possible nonzero contributions of $X(f)$ over \mathcal{F}_s . The exact value of L_0 is calculated by

$$-\frac{f_s}{2} + (L_0 + 1)f_p \geq \frac{f_{\text{NYQ}}}{2} \rightarrow L_0 = \left\lceil \frac{f_{\text{NYQ}} + f_s}{2f_p} \right\rceil - 1. \quad (10)$$

The relation (9) ties the known DTFTs of $y_i[n]$ to the unknown $X(f)$. This equation is the key to recovery of $x(t)$. For our purposes, it is convenient to write (9) in matrix form

$$\mathbf{y}(f) = \mathbf{A}\mathbf{z}(f), \quad f \in \mathcal{F}_s, \quad (11)$$

where $\mathbf{y}(f)$ is a vector of length m with i th element $y_i(f) = Y_i(e^{j2\pi f T_s})$. The unknown vector $\mathbf{z}(f) = [z_1(f), \dots, z_L(f)]^T$ is of length

$$L = 2L_0 + 1 \quad (12)$$

and

$$z_i(f) = X(f + (i - L_0 - 1)f_p), \quad 1 \leq i \leq L, \quad f \in \mathcal{F}_s. \quad (13)$$

The $m \times L$ matrix \mathbf{A} contains the coefficients c_{il} (6). Fig. 4 depicts the vector $\mathbf{z}(f)$ and the effect of aliasing $X(f)$ in f_p -shifted copies for $N = 4$ bands, aliasing rate $f_p = 1/T_p \geq B$ and two sampling rates. Each entry of $\mathbf{z}(f)$ is a windowed shifted interval of $X(f)$ whose length is f_s . Thus, in order to recover $x(t)$, it is sufficient to determine $\mathbf{z}(f)$ in the interval $f \in \mathcal{F}_p$.

The analysis so far holds for every choice of T_p -periodic functions $p_i(t)$. Before proceeding, we discuss the role of each parameter. The period T_p determines the aliasing of $X(f)$ by setting the shift intervals to $f_p = 1/T_p$. Equivalently,

¹The ceiling operator $\lceil a \rceil$ returns the greater (or equal) integer which is closest to a .

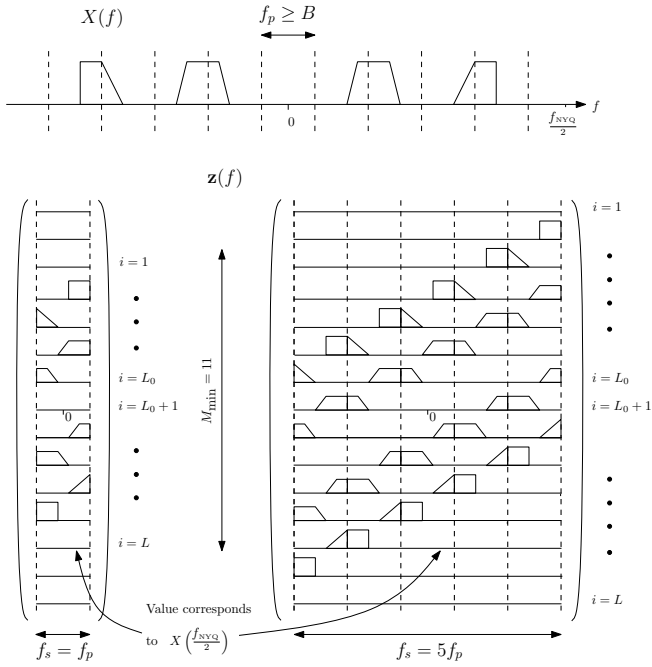


Fig. 4. The relation between the Fourier transform $X(f)$ and the vector set $\mathbf{z}(f)$, of (13). In the left pane, $f_s = f_p$ so that the length of $\mathbf{z}(f)$ is $L = 11$. The right pane demonstrates $f_s = 5f_p$ which gives $L = 15$. Entries in locations $i \leq L_0$ ($i > L_0 + 1$) contain shifted and windowed copies of $X(f)$ to the right (left) of the frequency axis. No shift occurs for the middle entry, $i = L_0 + 1$.

the aliasing rate f_p controls the way the bands are arranged in the spectrum slices $\mathbf{z}(f)$, as Fig. 4 depicts. We choose $f_p \geq B$ so that each band contributes only a single nonzero element to $\mathbf{z}(f)$, and consequently $\mathbf{z}(f)$ has at most N nonzeros. In practice f_p is chosen slightly more than B to avoid edge effects. Thus, the parameter T_p is used to translate the multiband prior $x(t) \in \mathcal{M}$ to a bound on the sparsity level of $\mathbf{z}(f)$. The sampling rate f_s of a single channel sets the frequency range \mathcal{F}_s in which (11) holds. It is clear from Fig. 4 that as long as $f_s \geq f_p$, recovering $x(t)$ from the sample sequences $y_i[n]$ amounts to recovery of $\mathbf{z}(f)$ from $\mathbf{y}(f)$, for every $f \in \mathcal{F}_p$. The number of channels m determines the overall sampling rate mf_s of the system. The simplest choice $f_s = f_p \simeq B$, which is presented on the left pane of Fig. 4, allows to control the sampling rate at a resolution of f_p . Later on, we explain how to trade the number of channels m by a higher rate f_s in each channel. Observe that setting f_p, f_s determines L by (10) and (12), which is the number of spectrum slices in $\mathbf{z}(f)$ that may contain energy for some $x(t) \in \mathcal{M}$.

The role of the mixing functions appears implicitly in (11) through the coefficients c_{il} . Each $p_i(t)$ provides a single row in the matrix \mathbf{A} . Roughly speaking, $p_i(t)$ should have many transients within the time period T_p so that its Fourier expansion (5) contains about L dominant terms. Then, the channel output $y_i[n]$ is a mixture of all (nonidentically zero) spectrum slices in $\mathbf{z}(f)$. The functions $p_i(t)$ should differ from each other to yield linearly independent rows in \mathbf{A} . We next make these requirements precise for a specific choice of $p_i(t)$ - the sign waveforms.

Consider the sign alternating function $p_i(t)$ which is depicted in Fig. 3. Calculating the coefficients c_{il} in this setting gives

$$c_{il} = \frac{1}{T_p} \int_0^{T_p} \sum_{k=0}^{M-1} \alpha_{ik} e^{-j\frac{2\pi}{T_p}l(t+k\frac{T_p}{M})} dt \quad (14)$$

$$= \frac{1}{T_p} \sum_{k=0}^{M-1} \alpha_{ik} e^{-j\frac{2\pi}{M}lk} \int_0^{T_p} e^{-j\frac{2\pi}{T_p}lt} dt. \quad (15)$$

Evaluating the integral we have

$$d_l = \frac{1}{T_p} \int_0^{T_p} e^{-j\frac{2\pi}{T_p}lt} dt = \begin{cases} \frac{1}{M} & l = 0 \\ \frac{1-\theta^l}{2j\pi l} & l \neq 0 \end{cases} \quad (16)$$

where $\theta = e^{-j2\pi/M}$, and thus

$$c_{il} = d_l \sum_{k=0}^{M-1} \alpha_{ik} \theta^{lk}. \quad (17)$$

Note that $c_{il} = c_{i,-l}^*$. Let $\bar{\mathbf{F}}$ be the $M \times M$ discrete Fourier transform matrix (DFT) whose i th column is

$$\bar{\mathbf{F}}_i = \left[\theta^{0 \cdot (i-1)}, \theta^{1 \cdot (i-1)}, \dots, \theta^{(M-1) \cdot (i-1)} \right]^T, \quad 1 \leq i \leq L, \quad (18)$$

where as before $\theta = e^{-j2\pi/M}$. Define \mathbf{F} to be the $M \times L$ matrix with $\mathbf{F}_i = \bar{\mathbf{F}}_k$ for $k = i - L_0$ (where $\bar{\mathbf{F}}_{k+M}$ is used if $k < 1$). Note that \mathbf{F} is a re-ordered column subset of $\bar{\mathbf{F}}$, and for $M = L$, \mathbf{F} is unitary. Then,

$$\mathbf{y}(f) = \mathbf{SFD} \mathbf{z}(f), \quad f \in \mathcal{F}_s, \quad (19)$$

reformulates (11) in terms of the $m \times M$ sign matrix \mathbf{S} , with $\mathbf{S}_{ik} = \alpha_{ik}$, and the $L \times L$ diagonal matrix $\mathbf{D} = \text{diag}(d_l)$, where d_l are defined in (16). The dependency on the sign patterns $\{\alpha_{ik}\}$ is further expanded in (20).

Intuitively, a sign alternating function $p_i(t)$ is implemented by a shift register, where M determines the number of flops. The clock rate of the register $(T_p/M)^{-1}$ is also dictated by M . The next section shows that $M \geq L$, where L is defined in (12), is one of the conditions for blind recovery. Since L is roughly f_{NYQ}/B for $f_p = B$, this implies a large value for M . In practice this is not an obstacle, since standard logic gates and feedback can be used to generate a sign pattern of length M (a.k.a, m-sequence) with just a few components [19], [20]. Nonetheless, to reduce the clock rate we choose $M = L$. The sign patterns $\{\alpha_{ik}\}$ initialize the shift register.

Note that the magnitude of d_l decays as l move away from $l = 0$. This is a consequence of the specific choice of sign alternating waveforms for the mixing functions $p_i(t)$. Under this selection, spectrum regions of $X(f)$ are weighted according to their proximity to the origin. In the presence of noise, the signal to noise ratio depends on the band locations due to this asymmetry.

C. Choice of parameters

An essential property of a sampling system is that the sample sequences match a unique analog input $x(t)$, since otherwise recovery is impossible. The following theorems address that issue. The first theorem states necessary conditions on

$$\underbrace{\begin{pmatrix} Y_1(e^{j2\pi f T_s}) \\ Y_2(e^{j2\pi f T_s}) \\ \vdots \\ Y_m(e^{j2\pi f T_s}) \end{pmatrix}}_{\mathbf{y}(f)} = \underbrace{\begin{bmatrix} \alpha_{1,0} & \cdots & \alpha_{1,M-1} \\ \vdots & \ddots & \vdots \\ \alpha_{m,0} & \cdots & \alpha_{m,M-1} \end{bmatrix}}_{\mathbf{S}} \underbrace{\begin{bmatrix} | & \cdots & | & \cdots & | \\ \bar{\mathbf{F}}_{-L_0} & \cdots & \bar{\mathbf{F}}_0 & \cdots & \bar{\mathbf{F}}_{L_0} \\ | & \cdots & | & \cdots & | \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} d_{-L_0} & & \\ & \ddots & \\ & & d_{L_0} \end{bmatrix}}_{\mathbf{D}} \underbrace{\begin{pmatrix} X(f - L_0 f_p) \\ \vdots \\ X(f) \\ \vdots \\ X(f + L_0 f_p) \end{pmatrix}}_{\mathbf{z}(f)} \quad (20)$$

the system parameters to allow a unique mapping. A concrete parameter selection which is sufficient for uniqueness, is provided in the second theorem. As a nice feature, the same selection works with half as many sampling channels, when the band locations are known. Thus, the system appearing in Fig. 3 can also replace baseband processing in the non-blind scenario. This may be beneficial for a receiver that switches between blind and non-blind modes according to availability of the transmitter carriers. More importantly, Fig. 3 suggests a possible architecture in the broader context of ADC design. The analog bandwidth of the frontend, which is dictated by the mixers, breaks the conventional bandwidth limitation in interleaved ADCs.

For brevity, we use sparsity notations in the statements. A vector \mathbf{u} is called K -sparse if \mathbf{u} contains no more than K nonzero entries. The set $\text{supp}(\mathbf{u})$ denotes the indices of the nonzeros in \mathbf{u} . The support of a collection of vectors over a continuous interval, such as $\mathbf{z}(\mathcal{F}_p) = \{\mathbf{z}(f) : f \in \mathcal{F}_p\}$, is defined by

$$\text{supp}(\mathbf{z}(\mathcal{F}_p)) = \bigcup_{f \in \mathcal{F}_p} \text{supp}(\mathbf{z}(f)). \quad (21)$$

A vector collection is called jointly K -sparse if its support contains no more than K indices.

Theorem 1 (Necessary conditions): Let $x(t)$ be an arbitrary signal within the multiband model \mathcal{M} , which is sampled according to Fig. 3 with $f_p = B$. Necessary conditions to allow exact spectrum-blind recovery (of an arbitrary $x(t) \in \mathcal{M}$) are $f_s \geq f_p$, $m \geq 2N$. For mixing with sign waveforms an additional necessary requirement is

$$M \geq M_{\min} \triangleq 2 \left\lceil \frac{f_{\text{NYQ}}}{2f_p} + \frac{1}{2} \right\rceil + 1. \quad (22)$$

Note that for $f_s = f_p$, $M_{\min} = L$ of (12); see also Fig. 4.

Proof: Observe that according to (9) and Fig. 4, the frequency transform of the i th entry of $\mathbf{z}(f)$ sums f_p -shifted copies of $X(f)$. If $f_s < f_p$, then the sum lacks contributions from $X(f)$ for some $f \in \mathcal{F}$. An arbitrary multiband signal may contain an information band within those frequencies. Thus, $f_s \geq f_p$ is necessary.

The other conditions are necessary to allow enough linearly independent equations in (11) for arbitrary $x(t) \in \mathcal{M}$. To prove the argument on m , first consider the linear system $\mathbf{v} = \mathbf{A}\mathbf{u}$ for the $m \times L$ matrix \mathbf{A} of (11). In addition, assume $f_s = f_p = B$. Substituting these values into (10),(12) and using $f_{\text{NYQ}} \geq 2NB$ gives $L > 2N$, namely \mathbf{A} has more than $2N$ columns.

If $m < 2N$, then since $\text{rank}(\mathbf{A}) \leq m$ there exist two N -sparse vectors $\bar{\mathbf{u}}_1 \neq \bar{\mathbf{u}}_2$ such that $\mathbf{A}\bar{\mathbf{u}}_1 = \mathbf{A}\bar{\mathbf{u}}_2$. The proof now follows from the following construction. For a given N -sparse vector \mathbf{u} , choose a frequency interval $\Delta \subset \mathcal{F}_p$ of length $B/2$. Construct a vector $\mathbf{z}(f)$ of spectrum slices, by letting $\mathbf{z}(f) = \mathbf{u}$ for every $f \in \Delta$, and $\mathbf{z}(f) = \mathbf{0}$ otherwise. Clearly, that $\mathbf{z}(f)$ corresponds to some $x(t) \in \mathcal{M}$ (see below an argument that treats the case that this construction results in a complex-valued $x(t)$). Follow this argument for $\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_2$ to provide $\bar{x}_1(t) \neq \bar{x}_2(t)$ within \mathcal{M} . Since $\mathbf{A}\bar{\mathbf{u}}_1 = \mathbf{A}\bar{\mathbf{u}}_2$, both $\bar{x}_1(t), \bar{x}_2(t)$ are mapped to the same samples. It can be verified that since $c_{il} = c_{i,-l}^*$, the existence of complex-valued $\bar{x}_1(t) \neq \bar{x}_2(t)$ implies the existence of a corresponding real-valued pair of signals within \mathcal{M} , which have the same samples.

The condition (22) comes from the structure of \mathbf{F} . For $M < M_{\min}$, $\mathbf{F}_1 = \mathbf{F}_{M+1}$. Now, set $\hat{\mathbf{u}}_1$ to be the zero vector except the value $1/d_1$ on the first entry. Similarly, let $\hat{\mathbf{u}}_2$ have zeros except for $1/d_{M+1}$ on the $(M+1)$ th entry. We can then use the arguments above to construct the signals $\hat{x}_1(t), \hat{x}_2(t)$ from $\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2$. It is easy to see that the signals (or their real-valued counterparts) are mapped to the same samples although they are different.

The proof on the necessity of $m \geq 2N$, $M \geq M_{\min}$ for $f_s > f_p$ follows from the same arguments. ■

We point out that the necessary conditions on m, M may change with other choices of f_p . However, $f_p = B$ is sufficient for our purposes, and allows to reduce the total sampling rate as low as possible. In addition, note that it is recommended (though not necessary) to have $M \leq 2^{m-1}$. This requirement stems from the fact that \mathbf{S} is defined over a finite alphabet $\{+1, -1\}$ and thus cannot have more than 2^{m-1} linearly independent columns. Therefore, in a sense, the degrees of freedom in $\mathbf{A} = \mathbf{S}\mathbf{F}\mathbf{D}$ are decreased² for $M > 2^{m-1}$. We next show that the conditions of Theorem 1 are also sufficient for blind recovery, under additional conditions.

Theorem 2 (Sufficient conditions): Let $x(t)$ be an arbitrary signal within the multiband model \mathcal{M} , which is sampled according to Fig. 3 with sign waveforms $p_i(t)$. If

1. $f_s \geq f_p \geq B$,
2. $M \geq M_{\min}$, where M_{\min} is defined in (22),
3. $m \geq N$ for non-blind reconstruction or $m \geq 2N$ for blind,
4. Every $2N$ columns of $\mathbf{S}\mathbf{F}$ are linearly independent,

²Note that repeating the arguments of the proof for $M > 2^{m-1}$ allows to construct spectrum slices $\mathbf{z}(f)$ in the null space of $\mathbf{S}\mathbf{F}$. However, these do not necessarily correspond to $x(t) \in \mathcal{M}$ and thus this requirement is only a recommendation.

TABLE I
POSSIBLE PARAMETER CHOICES FOR MULTIBAND SAMPLING.

Model	
$N = 6$ $B = 50$ MHz $f_{\text{NYQ}} = 10$ GHz	
Sampling parameters	
Option A	Option B
$f_p = \frac{f_{\text{NYQ}}}{195} \approx 51.3$ MHz	$f_p = \frac{f_{\text{NYQ}}}{195} \approx 51.3$ MHz
$f_s = f_p \approx 51.3$ MHz	$f_s = 5f_p \approx 256.4$ MHz
$m \geq 2N = 12$	$m \geq \lceil \frac{2N}{5} \rceil = 3$
$M = M_{\min} = 195$	$M = 199$
$M_{\min} = L = 195$	$M_{\min} = 195, L = 199$
Rate $mf_s \geq 615$ MHz	Rate $mf_s \geq 770$ MHz

then, for every $f \in \mathcal{F}_s$, the vector $\mathbf{z}(f)$ is the unique N -sparse solution of (19).

Proof: The choice $f_p \geq B$ ensures that every band can contribute only a single non-zero value to $\mathbf{z}(f)$. Fig. 4 and the earlier explanations provide a proof of this statement. As a consequence, $\mathbf{z}(f)$ is N -sparse for every $f \in \mathcal{F}_s$.

Recall that $M \geq L$ is necessary. Under this condition, \mathbf{D} contains nonzero diagonal entries, since $d_l = 0$ only for $l = \pm kM$ for some $k > 1$. This implies that \mathbf{D} is nonsingular and $\text{rank}(\mathbf{SFD}) = \text{rank}(\mathbf{SF})$. Thus linear independence of any column subset of \mathbf{SF} implies corresponding linear independence for \mathbf{SFD} .

In the non-blind setting, the band locations imply the support $\text{supp}(\mathbf{z}(f))$ for every $f \in \mathcal{F}_s$. The other two conditions (on m , \mathbf{SF}) ensure that (19) can be inverted on the proper column subset, thus providing the uniqueness claim. A closed-form expression is given in (33) below.

In blind recovery, the nonzero locations of $\mathbf{z}(f)$ are unknown. We therefore rely on the following result from the CS literature: A K -sparse vector \mathbf{u} is the unique solution of $\mathbf{v} = \mathbf{A}\mathbf{u}$ if every $2K$ columns of \mathbf{A} are linearly independent [27]. This condition translates into $m \geq 2N$ and the condition on \mathbf{SF} of the theorem. ■

To reduce the sampling rate to minimal we may choose $f_s = f_p = B$ and $m = 2N$ (for the blind scenario). This translates to an average sampling rate of $2NB$, which is the lowest possible for $x(t) \in \mathcal{M}$ [5]. Table I presents two parameter choices for a representative signal model. Option A in the table uses $f_s = f_p$ and leads to a sampling rate as low as 615 MHz, which is slightly above the minimal rate $2NB = 600$ MHz. Option B is discussed in the next section.

Recall the proof of Theorem 1, which shows that \mathbf{A} has $L > 2N$ columns. Therefore, if $m = 2N$ is sufficiently small, the requirement $M \geq L$ may contradict the recommendation $M \leq 2^{m-1}$. This situation is rare due to the exponential nature of the upper bound; it does not happen in the examples of Table I. Nonetheless, if it happens, then we may view $x(t) \in \mathcal{M}$ as conceptually having ρN bands, each of width B/ρ , and set $f_p = B/\rho$. The upper bound on M grows exponentially with ρ while the lower bound grows only linearly, thus for some integer $\rho \geq 1$ we may have a valid selection for M . This approach requires $m = 2\rho N$ branches which corresponds to a large number of sampling channels. Fortunately, this situation

can be solved by trading the number of sampling channels for a higher sampling rate f_s .

To complete the sampling design, we need to specify how to select the matrix \mathbf{S} , namely the sign patterns $\{\alpha_{ik}\}$. This choice effects the reconstruction stability. In Section IV, we describe the mechanism for recovering $x(t)$ from the given samples $y_i[n]$ and address the choice of $\{\alpha_{ik}\}$.

D. Trading channels for sampling rate

The burden on hardware implementation is highly impacted by the total number of hardware devices, which includes the mixers, the lowpass filters and the ADCs. Clearly, it would be beneficial to reduce the number of channels as low as possible. We now examine a method which reduces the number of channels at the expense of a higher sampling rate f_s in each channel and additional digital processing.

Suppose $f_s = qf_p$, with odd $q = 2q' + 1$. To analyze this choice, consider the i th channel of (11) for $f \in \mathcal{F}_p$:

$$y_i(f + kf_p) = \sum_{l=-\infty}^{\infty} c_{il} X(f + kf_p - lf_p) \quad (23)$$

$$= \sum_{l=-L_0-k}^{+L_0-k} c_{i,(l+k)} X(f - lf_p) \quad (24)$$

$$= \sum_{l=-L_0}^{+L_0} c_{i,(l+k)} X(f - lf_p) \quad (25)$$

where $-q' \leq k \leq q'$. The first equality follows from a change of variable, and the second from the definition of L_0 in (10), which implies that $X(f - lf_p) = 0$ over $f \in \mathcal{F}_p$ for every $|l| > L_0 - q'$. Now, according to (25), a system with $f_s = qf_p$ provides q equations on \mathcal{F}_p for each physical channel. Equivalently, m hardware branches (including all components) amounts to $m q$ channels having $f_s = f_p$. Eq. (28) expands this relation. Note that the sign patterns do not appear explicitly as in (19).

Theorem 2 ensures that $\mathbf{z}(f)$ has N nonzero elements for every $f \in \mathcal{F}_s$. Nonetheless, as detailed in the next section, for efficient recovery it is more interesting to determine the joint sparsity level of $\mathbf{z}(f)$ over \mathcal{F}_s . As Fig. 4 depicts, over $f \in \mathcal{F}_p$, $\mathbf{z}(f)$ is $2N$ -jointly sparse, whereas over the wider range $f \in \mathcal{F}_s$, $\mathbf{z}(f)$ may have a larger joint support set. It is therefore beneficial to truncate the sequences appearing in (25) to the interval \mathcal{F}_p , prior to reconstruction. In terms of digital processing, the left-hand-side of (28) is obtained from the input sequence $y_i[n]$ as follows. For every $-q' \leq k \leq q'$, the frequency shift $y_i(f + kf_p)$ is carried out by time modulation. Then, the sequence is lowpass filtered by $h_D[n]$ and decimated by q . The filter $h_D[n]$ is an ideal lowpass filter with digital cutoff π/q , where π corresponds to half of the input sampling rate f_s . This processing yields the rate $f_p = f_s/q$ sequences

$$\tilde{y}_{i,k}[\tilde{n}] = (y_i[n] e^{-j2\pi k f_p n T_s}) * h_D[n] \Big|_{n=\tilde{n}q} \quad (26)$$

$$= (y_i[n] e^{-j\frac{2\pi}{q} kn}) * h_D[n] \Big|_{n=\tilde{n}q} \quad (27)$$

Conceptually, the sampling system consists of $m q$ channels which generate the sequences (27) with $f_s = f_p$.

$$\begin{pmatrix} y_i(f - q'f_p) \\ \vdots \\ y_i(f) \\ \vdots \\ y_i(f + q'f_p) \end{pmatrix} = \begin{bmatrix} c_{i,-L_0-q'} & & \cdots & & c_{i,L_0-q'} \\ & \vdots & \ddots & & \vdots \\ & c_{i,-L_0} & \cdots & c_{i,-1} & c_{i,0} & c_{i,1} & \cdots & c_{i,L_0} \\ & \vdots & & & & & \ddots & \vdots \\ c_{i,-L_0+q'} & & \cdots & & & & & c_{i,L_0+q'} \end{bmatrix} \begin{pmatrix} | \\ \mathbf{z}(f) \\ | \end{pmatrix}, \quad f \in \mathcal{F}_p. \quad (28)$$

Table I presents a parameter choice, entitled Option B, which makes use of this strategy. Thus, instead of the proposed setting of Theorem 2 with $m \geq 12$ channels, uniqueness can be guaranteed from only 3 channels. Observe that the lowest sampling rate in this setting is higher than the minimal $2NB$, since the strategy expands each channel to an integer number q of sequences. In the example, 3 channels are digitally expanded to $3q = 15$ channels. In Section V-C we demonstrate empirical evidence for this approach, when using a finite impulse response (non-ideal) filter to approximate $h_D[n]$.

Theoretically, this strategy allows to collapse a system with m channels to a single channel with sampling rate $f_s = mf_p$. However, each channel requires q digital filters to reduce the rate back to f_p , which increases the computational load. In addition, as q grows, approximating a digital filter with cutoff is π/q requires more taps.

The next section discusses the reconstruction stage, which takes the m sample sequences $y_i[n]$ (or the mq decimated sequences $\tilde{y}_{i,k}[\tilde{n}]$) and recovers the Nyquist rate sequence $x(nT)$ or the analog version $x(t)$. As we explain, the reconstruction stage also allows to output a digital sequence corresponding to any specific band of interest. This sequence is generated at low rate, namely without going through the recovery of the Nyquist rate sequence $x(nT)$.

IV. RECONSTRUCTION

In this section, $f_s = f_p$ is assumed under the interpretation that the input sequences were expanded and decimated earlier if needed. Recovery of $x(t)$ from the sequences $y_i[n]$ boils down to recovery of the sparsest $\mathbf{z}(f)$ of (11) for every $f \in \mathcal{F}_s$. The system (11) falls into a broader framework of sparse solutions to a parameterized set of linear systems [11]. The relevant results are quoted and are then specified to the multiband scenario. The recovery stage is developed for arbitrary mixing functions. Stability aspects are described for alternating sign functions $p_i(t)$.

A. IMV model

Let \mathbf{A} be an $m \times M$ matrix with $m < M$. The constants m, M are arbitrary in this section, but later on we apply the results to \mathbf{A} of (11) which explains this notation. Consider a parameterized family of linear systems

$$\mathbf{v}(\lambda) = \mathbf{A}\mathbf{u}(\lambda), \quad \lambda \in \Lambda, \quad (29)$$

indexed by a fixed set Λ that may be infinite. Let $\mathbf{u}(\Lambda) = \{\mathbf{u}(\lambda) : \lambda \in \Lambda\}$ be a collection of M -dimensional vectors that solves the entire family of systems. We will assume that the vectors in $\mathbf{u}(\Lambda)$ are jointly K -sparse in the sense that

Continuous to finite (CTF) block

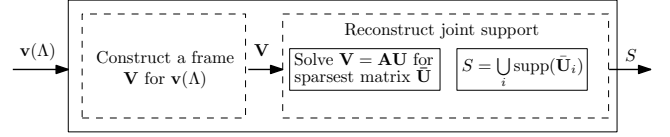


Fig. 5. Recovery of the joint support $S = \text{supp}(\mathbf{u}(\Lambda))$.

$|\text{supp}(\mathbf{u}(\lambda))| \leq K$. In words, the nonzero entries of each vector $\mathbf{u}(\lambda)$ lie within a set of at most K indices.

When the support $S = \text{supp}(\mathbf{u}(\Lambda))$ is known, recovering $\mathbf{u}(\Lambda)$ from the known vector set $\mathbf{v}(\Lambda) = \{\mathbf{v}(\lambda) : \lambda \in \Lambda\}$ is possible if the submatrix \mathbf{A}_S , which contains the columns of \mathbf{A} indexed by S , has full column rank. In this case,

$$\mathbf{u}_S(\lambda) = \mathbf{A}_S^\dagger \mathbf{v}(\lambda) \quad (30a)$$

$$\mathbf{u}_i(\lambda) = 0, \quad i \notin S \quad (30b)$$

where $\mathbf{u}_S(\lambda)$ contains only the entries of $\mathbf{u}(\lambda)$ indexed by S and $\mathbf{A}_S^\dagger = (\mathbf{A}_S^H \mathbf{A}_S)^{-1} \mathbf{A}_S^H$ is the (Moore-Penrose) pseudoinverse of \mathbf{A}_S . For unknown support S , (29) is still invertible if $K = |S|$ is known, and every set of $2K$ columns from \mathbf{A} is linearly independent [11], [27], [28]. In general, finding the support of $\mathbf{u}(\Lambda)$ is NP-hard because it may require a combinatorial search. Nevertheless, recent advances in compressive sampling and sparse approximation delineate situations where polynomial-time recovery algorithms correctly identify S for finite Λ . This challenge is referred to as a multiple measurement vectors (MMV) problem [28]–[33].

The sparsest solution of a linear system, for unknown support S , has no closed-form solution. Thus, when Λ has infinite cardinality, referred to as the infinite measurement vectors (IMV) problem [11], solving for $\mathbf{u}(\Lambda)$ conceptually requires an independent treatment for infinitely many systems [11]. To avoid this difficulty of IMV, we proposed in [5], [11] a two step flow which recovers the support set S from a finite-dimensional system, and then uses (30) to recover $\mathbf{u}(\Lambda)$. The algorithm begins with the construction of a (finite) frame \mathbf{V} for $\mathbf{v}(\Lambda)$. Then, it finds the (unique) solution $\tilde{\mathbf{U}}$ to the MMV system $\mathbf{V} = \mathbf{A}\mathbf{U}$ that has the fewest nonzero rows. The main result is that $S = \text{supp}(\mathbf{u}(\Lambda))$ equals $\text{supp}(\tilde{\mathbf{U}})$, namely the index set of the nonidentically zero rows of $\tilde{\mathbf{U}}$. These operations are grouped in a block entitled continuous to finite (CTF), depicted in Fig. 5. Therefore, the support recovery can be accomplished by solving only a finite dimensional problem.

In the next section, we specify the CTF block for multiband reconstruction. Some additional insights into the CTF block are given in the specific context of our problem.

B. Multiband reconstruction

The system (11) clearly obeys the IMV model with $\Lambda = \mathcal{F}_s$. In order to use the CTF block, we need to construct a frame \mathbf{V} for the measurement set $\mathbf{y}(\Lambda)$. Such a frame can be obtained by computing [11]

$$\mathbf{Q} = \int_{f \in \mathcal{F}_s} \mathbf{y}(f) \mathbf{y}^H(f) df \quad (31)$$

$$= \sum_{n=-\infty}^{+\infty} \mathbf{y}[n] \mathbf{y}^T[n], \quad (32)$$

where $\mathbf{y}[n] = [y_1[n], \dots, y_m[n]]^T$ is the vector of samples at time instances nT_s . Then, any matrix \mathbf{V} , for which $\mathbf{Q} = \mathbf{V}\mathbf{V}^H$, is a frame for $\mathbf{y}(\mathcal{F}_s)$ [11]. The CTF block, Fig. 5, can then be used to recover the support $S = \text{supp}(\mathbf{z}(\mathcal{F}_p))$. It follows that,

$$\mathbf{z}_S[n] = \mathbf{A}_S^\dagger \mathbf{y}[n] \quad (33a)$$

$$\mathbf{z}_i[n] = 0, \quad i \notin S, \quad (33b)$$

where $\mathbf{z}[n] = [z_1[n], \dots, z_L[n]]^T$ and $z_i[n]$ is the inverse-DTFT of $z_i(f)$. Therefore, once the support is recovered, the sequences $z_i[n]$ are generated at the input rate f_s . At this point, we may recover $x(t)$ by either of the two following options. If f_{NYQ} is not prohibitively large, then we can generate the Nyquist rate sequences $x(nT)$ digitally and then use an analog lowpass (with cutoff $1/2T$) to recover $x(t)$. The digital sequence $x(nT)$ is generated by shifting each spectrum slice $z_i(f)$ to the proper position in the spectrum, and then summing up the contributions. In terms of digital processing, the sequences $z_i[n]$ are first zero padded:

$$\tilde{z}_i[\tilde{n}] = \begin{cases} z_i(n) & \tilde{n} = nL, n \in \mathbb{Z} \\ 0 & \text{otherwise.} \end{cases} \quad (34)$$

Then, $\tilde{z}_i[\tilde{n}]$ is interpolated to the Nyquist rate, using an ideal (digital) filter. Finally, the interpolated sequences are modulated in time and summed:

$$x[n] = x(nT) = \sum_{i \in S} (\tilde{z}_i[n] * h_I[n]) e^{2\pi i f_p n T}. \quad (35)$$

The alternative option is to handle the sequences $z_i[n]$ directly by analog hardware. Every $z_i[n]$ passes through an analog lowpass filter with cutoff $f_s/2$ and gives (the complex-valued) $z_i(t)$. Then,

$$x(t) = \sum_{i \in S, i > L_0} \mathcal{R}\{z_i(t)\} \cos(2\pi i f_p t) + \mathcal{I}\{z_i(t)\} \sin(2\pi i f_p t), \quad (36)$$

where $\mathcal{R}(\cdot), \mathcal{I}(\cdot)$ denote the real and imaginary part of their argument, respectively. We emphasize that although the analysis of Section III-B was carried out in the frequency domain, the recovery of $x(t)$ is done completely in the time-domain, via (32)-(36).

The next section summarizes the recovery flow and its advantages from a high-level viewpoint.

C. Architecture and advantages

Fig. 6 depicts a high-level architecture of the entire recovery flow. The sample sequences entering the digital domain are expanded by the factor $q = f_s/f_p$ (if needed) and are bundled together. The application layer triggers the CTF block on initialization and when identifying that the spectral support has changed. The digital signal processor (DSP) treats the samples, based on the recovered support, and outputs baseband sequences for each spectrum slice. An analog back-end interpolates the sequences and sums them up according to (36). The controller has the ability to selectively activate the digital recovery of any specific band of interest, and in particular to produce an analog counterpart (at baseband) by overriding the relevant carrier frequencies.

CTF and sampling rate. The frame construction step of the CTF conceptually merges the infinite collection $\mathbf{z}(\mathcal{F}_s)$ to a finite basis or frame, which preserves the original support. For the CTF to work in the multiband reconstruction, the sampling rate must be doubled due to a specific property that this scenario exhibits. Observe that under the choices of Theorem 2, $\mathbf{z}(\mathcal{F}_p)$ is jointly $2N$ -sparse, while each $\mathbf{z}(f)$ is N -sparse. This stems from the continuity of the bands which permits each band to have energy in (at most) two spectrum pieces within \mathcal{F}_p . Therefore, when aggregating the frequencies the support $\text{supp}(\mathbf{z}(\mathcal{F}_p))$ cannot contain more than $2N$ indices. An algorithm which makes use of several CTF instances and gains back this factor was proposed in [5]. Although the same algorithm applies here as well, we do not pursue this direction so as to avoid additional digital computations.

MMV recovery complexity. The CTF block requires to solve an MMV system, which is a known NP-hard problem. In practice, sub-optimal polynomial-time CS algorithms may be used for this computation [11], [28], [32], [33]. The price for tractability is an increase in the sampling rate. In the next section, we quantify this effect for a specific recovery approach. We refer the reader to [28], [33], [34] for theoretical guarantees regarding MMV recovery algorithms.

Realtime processing. Standard CS algorithms, for the finite Λ scenario, couple the tasks of support recovery and the construction of the entire solution. In the infinite scenario, however, the separation between the two tasks has a significant advantage. The support recovery step yields an MMV system, whose dimensions are $m \times L$. Thus, we can control the recovery problem size by setting the number of channels m , and setting L via f_p, f_s in (12). Once the support is known, the actual recovery has a closed form (33), and can be carried out in realtime. Indeed, even the recovery of the Nyquist rate sequence (34)-(36), can be done at a constant rate. Had these tasks been coupled, the reconstruction stage would have to recover the Nyquist rate signal directly. In turn, the CS algorithm would have to run on a huge-scale system, dictated by the ambient Nyquist dimension, which is time and memory consuming. This point is discussed in further detail in Section VI.

Spectrum-blind baseband processing. The frame construction step in the CTF (31) is theoretically noncausal.

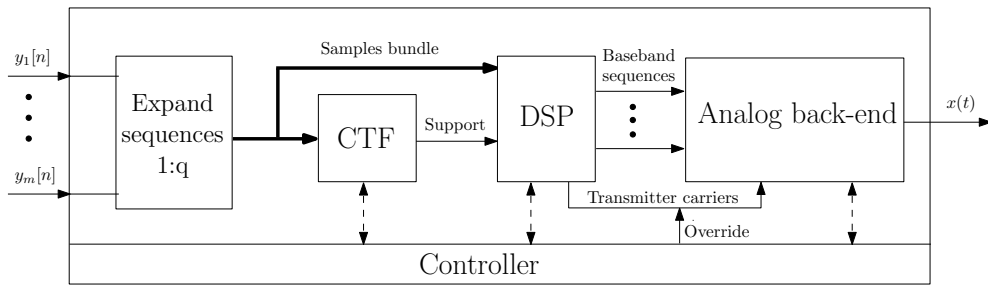


Fig. 6. High-level architecture for efficient multiband reconstruction.

Nonetheless, empirical evidence shows that a few samples suffice to approximate \mathbf{Q} ; We discuss this issue further in Section V-A. The CTF is executed only when the spectral support changes, and thus the short delay introduced by its execution is negligible on average. This practically allows realtime processing (at low rate) of any specific information band of interest. In particular, there is no need to recover the Nyquist rate signal before a higher application layer can access the digital information. This aspect is another prominent advantage over the system of [9], [10].

Robustness and sensitivity. The entire system, sampling and reconstruction, is robust against inaccuracies in the parameters f_s, f_p . This is a consequence of setting the parameters according to Theorem 2, with only the inequalities $f_s \geq f_p \geq B$. In particular, f_p is chosen above the minimal to ensure safety guard regions against hardware inaccuracies. Furthermore, observe that the exact values of f_s, f_p do not appear anywhere in the recovery flow: the expanding equations (27), the frame construction (32), the CTF block - Fig. 5, and the recovery equations (33). Only the ratio $q = f_s/f_p$ is used, which remains unchanged if the a single clock circuitry is used in the design. In addition, in the recovery of the Nyquist rate sequence (35), only the ratio T/T_p is used, which remains fixed for the same reasons. If using the analog recovery of $x(t)$ with (36), then f_p is provided to the back-end from the same clock triggering the sampling stage. The recovery is also stable in the presence of noise as numerically demonstrated in Section V-A. The next subsection briefly discusses the choice of the sign patterns which impacts stability in some sense. A forthcoming publication [12] studies sensitivity, robustness and stability aspects in more detail.

Finally we point out an advantage with respect to the reconstruction of a multicoset based receiver. The IMV formulation holds for this strategy with a different sampling matrix \mathbf{A} [5]. However, the IMV system holds only with respect to a (Nyquist rate) zero padded version of (2). Consequently, constructing a frame \mathbf{V} from the multicoset low rate sequences (2) requires interpolating the lowrate sequences (2) to the Nyquist rate. Only then can \mathbf{Q} be computed (see Eq. (61)-(62) in [5]). In contrast, the current mixing stage has the advantage that the IMV is expressed directly in terms of the lowrate sequences $y_i[n]$, and the computation of \mathbf{Q} in (32) is carried out directly on the input sequences. In fact, one may implement an adaptive frame construction at the input rate f_s . Digital processing at rate f_s is obviously preferred over a

processor running at the Nyquist rate.

D. Choosing the sign patterns

Theorem 2 requires that for uniqueness, every $2N$ columns of \mathbf{SF} must be linearly independent. To apply the CTF block the requirement is strengthened to every $4N$ columns, which also implies the minimal number of rows in \mathbf{S} [5]. Verifying that a set of sign patterns $\{\alpha_{ik}\}$ satisfies such a condition is computationally difficult because one must check the rank of every set of $4N$ columns from \mathbf{SF} . In practice, when noise is present or when solving the MMV by sub-optimal polynomial-time CS algorithms, the number of rows in \mathbf{S} should be increased beyond $m = 4N$. A preliminary discussion on how to choose \mathbf{S} , which appears in the conference version of this work [35], is summarized below. A more rigorous treatment is beyond the current scope, and will be investigated in detail in [12].

Consider the system $\mathbf{v} = \mathbf{A}\mathbf{u}$, where \mathbf{u} is an unknown sparse vector, \mathbf{v} is the measurement vector, and \mathbf{A} is of size $m \times M$. A matrix \mathbf{A} is said to have the restricted isometry property (RIP) [29] of order K , if there exists $0 \leq \delta_K < 1$ such that

$$(1 - \delta_K)\|\mathbf{u}\|^2 \leq \|\mathbf{A}\mathbf{u}\|^2 \leq (1 + \delta_K)\|\mathbf{u}\|^2 \quad (37)$$

for every K -sparse vector \mathbf{u} [29]. The requirement of Theorem 2 thus translates to $\delta_{2N} < 1$. The RIP requirement is also hard to verify for a given matrix. Instead, it can be easier to prove that a random \mathbf{A} , chosen from some distribution, has the RIP with high probability. In particular, it is known that a random sign matrix, whose entries are drawn independently with equal probability, has the RIP of order K if $m \geq CK \log(M/K)$, where C is a positive constant independent of everything [36]. The log factor is necessary [37]. The RIP of matrices with random signs remains unchanged under any fixed unitary transform of the rows [36]. This implies that if \mathbf{S} is a random sign matrix, then \mathbf{SF} has the RIP of order $2N$ for the above dimension selection. Note that \mathbf{D} is ignored in this analysis, since the diagonal has nonzero entries and thus $\text{supp}(\mathbf{D}\mathbf{u}) = \text{supp}(\mathbf{u})$ for any vector \mathbf{u} .

To advance, observe that solving for \mathbf{u} would require the combinatorial search implied by

$$\min_{\mathbf{u}} \|\mathbf{u}\|_0 \text{ s.t. } \mathbf{v} = \mathbf{A}\mathbf{u}. \quad (38)$$

A popular approach is to approximate the sparsest solution by

$$\min_{\mathbf{u}} \|\mathbf{u}\|_1 \text{ s.t. } \mathbf{v} = \mathbf{A}\mathbf{u}. \quad (39)$$

The relaxed program, named basis pursuit (BP) [38], is convex and can be tackled with polynomial-time solvers [29]. Many works have analyzed the basis pursuit method and its ability to recover the sparsest vector \mathbf{u} . For example, if $\delta_{2K} \leq \sqrt{2}-1$ then (39) recovers the sparsest \mathbf{u} [39]. The squared error of the recovery in the presence of noise or model mismatch was also shown to be bounded under the same condition [39]. Similar conditions were shown to hold for other recovery algorithms. In particular, [34] proved a similar argument for a mixed ℓ_2/ℓ_1 program in the MMV setting (which incorporates the joint sparsity prior).

In practice, the matrix \mathbf{S} is not random once the sampling stage is implemented, and its RIP constant cannot be calculated efficiently. Extensive simulations on synthesized data are a common tool to evaluate the performance and the stability (e.g., see [11], [28], [31]). Clearly, the numerical results do not ensure a desired RIP constant. Nonetheless, for practical applications, the behavior observed in simulations may be sufficient. The discussion above implies that stable recovery of the MMV of Fig. 5 requires roughly

$$m \approx 4N \log(M/2N) \quad (40)$$

channels to estimate the correct support, using polynomial-time algorithms.

V. NUMERICAL SIMULATIONS

A. Design example

To evaluate the performance of the proposed system (see Fig. 3) we simulate the act of the system on test signals contaminated with white Gaussian noise.

More precisely, we evaluate the performance on 500 noisy test signals of the form $x(t) + w(t)$, where x is a multiband signal and w is a white Gaussian noise process. The multiband model of Table I is used hereafter. The signal consists of 3 pairs of bands (total $N = 6$), each of width $B = 50$ MHz, constructed using the formula

$$x(t) = \sum_{i=1}^3 \sqrt{E_i B} \operatorname{sinc}(B(t - \tau_i)) \cos(2\pi f_i(t - \tau_i)), \quad (41)$$

where $\operatorname{sinc}(x) = \sin(\pi x)/(\pi x)$. The energy coefficients are $E_i = \{1, 2, 3\}$ and the time offsets are $\tau_i = \{0.4, 0.7, 0.2\}$ μsecs . The exact values $X(f)$ takes on the support do not affect the results and thus E_i, τ_i are fixed in all our simulations. For every signal the carriers f_i are chosen uniformly at random in $[-f_{\text{NYQ}}/2, f_{\text{NYQ}}/2]$ with $f_{\text{NYQ}} = 10$ GHz.

We design the sampling stage according to "Option A" of Table I. Specifically, $f_s = f_p = f_{\text{NYQ}}/195 \simeq 51.3$ MHz. The number of channels is set to $m = 100$, where each mixing function $p_i(t)$ alternates sign at most $M = M_{\min} = 195$ times. Each sign α_{ik} is chosen uniformly at random and fixed for the duration of the experiment. To represent continuous signals in simulation, we place a dense grid of 50001 equispaced points in the time interval $[0, 1\mu\text{secs}]$. The time resolution under this choice, $T/5$, is used for accurate representation of the signal after mixing, which is not bandlimited. The Gaussian noise is added and scaled so that the test signal has the desired signal-to-noise ratio (SNR), where the SNR is defined to be

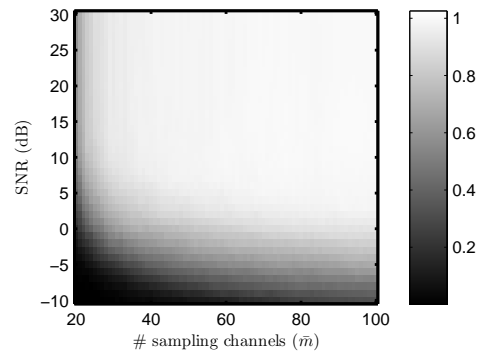


Fig. 7. Image intensity represents percentage of correct support set recovery $\hat{S} = S$, for reconstruction from different number of sampling sequences \bar{m} and under several SNR levels.

$10 \log(\|x\|^2/\|w\|^2)$, with the standard l_2 norms. To imitate the analog filtering and sampling, we use a lengthy digital FIR filter followed by decimation at the appropriate factor. After removing the delay caused by this filter, we end up with 40 samples per channel at rate f_s , which corresponds to observing the signal for 780 nsecs. We emphasize that these steps are required only when simulating an analog hardware numerically. In practice, the continuous signals pass through an analog filter (e.g., a 3rd order Chebyshev type-I), and there is no need for decimation or a dense time grid.

The support of the input signal is reconstructed from $\bar{m} \leq m$ channels. (More precisely, $S = \text{supp}(\mathbf{z}(\mathcal{F}_p))$ is recovered.) We follow the procedure described in Fig. 5 to reduce the IMV system (19) to an MMV system. Due to Theorem 2, \mathbf{Q} is expected to have (at most) $2N = 12$ dominant eigenvectors. The noise space, which is associated with the remaining negligible eigenvalues is discarded by simple thresholding (10^{-9} is used in the simulations). Then, the frame \mathbf{V} is constructed and the MMV is solved using simultaneous orthogonal matching pursuit [31], [32]. We slightly modified the algorithm to select a symmetric pair of support indices in every iteration, based on the conjugate symmetry of $X(f)$. Success recovery is declared when the estimated support set is equal the true support, $\hat{S} = S$. As explained, recovery of the Nyquist rate signal can be carried out by (35)-(36). Fig. 7 reports the percentage of correct support recoveries for various numbers \bar{m} of channels and several SNRs.

The results show that in the high SNR regime, correct recovery is accomplished when using $\bar{m} \geq 35$ channels, which amounts to less than 18% of the Nyquist rate. This rate conforms with (40) which predicts an order of $4N \log(M/2N) \simeq 30$ channels for stable recovery. A saving factor 2 is possible if using more than a single CTF block and a complicated processing (see [5] for details) or by brute-force MMV solvers with exponential recovery time. An obvious trend which appears in the results is that the recovery rate is inversely proportional to the SNR level and to the number of channels \bar{m} used for reconstruction.

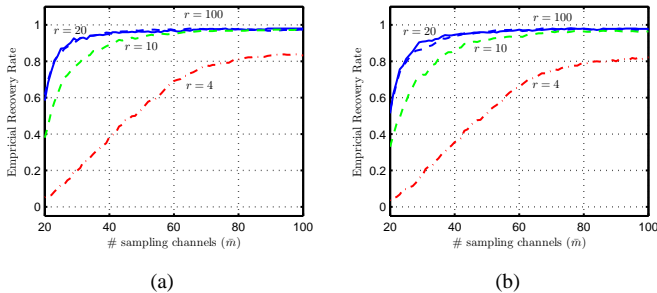


Fig. 8. Percentage of correct support recovery, when drawing the sign patterns randomly only for the first r channels. Results are presented for (a) SNR=25 dB and (b) SNR=10 dB.

B. Simplifying the mixing stage

Each channel needs a mixing function $p_i(t)$, which supposedly requires a shift register of M flip flops. In the setting of Fig. 7, every channel requires $M = 195$ flip flops with a clock operating at $(T_p/M)^{-1} = 10$ GHz. Without delving into the circuit level, the high clock frequency may require to implement each register as a chain of discrete devices.

We propose a simple method to reduce the total number of flip flops by sharing the same register for a few channels, and using consecutive taps to produce several mixing functions simultaneously. This strategy however reduces the degrees of freedom in \mathbf{S} and may affect the recovery performance. To qualitatively evaluate this approach, we generated sign matrices \mathbf{S} whose first r rows are drawn randomly as before. Then, the i th row, $r < i \leq m$, is five cyclic shifts (to the right) of the $(i-r)$ th row. Fig. 8 reports the recovery success for several choices of r and two SNR levels. As evident, this strategy enables a saving of 80% of the total number of flip flops, with no empirical degradation in performance.

C. Collapsing analog channels

Section III-D introduced a method to collapse q sampling channels to a single channel with a higher sampling rate $f_s = qf_p$. To evaluate this strategy, we choose the parameter set "Option B" of Table I. Specifically, the system design of Section V-A is now changed to $f_s = 5f_p$, with $m = 20$ physical channels.

In the simulation, the time interval in which the signal is observed is extended to $[0, 4\mu\text{secs}]$, such that every channel records (after filtering and sampling) about 500 samples. The extended window enables accurate digital filtering in order to separate each sequence to $q = 5$ different equations. We design a 100-tap digital FIR filter with the MATLAB command `h=fir1(100,1/q)` to approximate the optimal filter $h_D[n]$ of Section III-D. Then, for the i th sample sequence $y_i[n]$, h is convolved with each of the modulated versions $y_i[n]e^{j2\pi/qln}$, where $-q' \leq l \leq q' = 2$. Fig. 9 reports the recovery performance for different SNR levels and versus the number of sampling channels. The performance trend remains as in Fig. 7. In particular, $35/q = 7$ channels achieve an acceptable recovery rate. This implies a significant saving in hardware components.

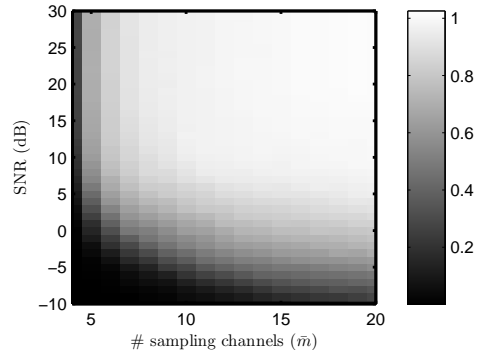


Fig. 9. The recovery rate for different SNR levels when using $f_s = qf_p$.

VI. RELATED WORK

We now compare the modulated wideband converter, Figs. 3 and 6, to multicaset sampling [5] and the random demodulator [9], [10]. Multicaset sampling was described in Section II-B; reconstruction from these samples shares the architecture of Fig. 6. The random demodulator is briefly described next.

A. The random demodulator

Fig. 10 presents the random demodulator of [9], [10]. The input signal $f(t)$ is first mixed by a sign waveform with a long period, produced by a pseudorandom sign generator which alternates at rate W . The mixed output is then integrated and dumped at a constant rate R , resulting in the sequence $y[n]$, $1 \leq n \leq N_R$. The design parameters are the rates W , R and the number of samples N_R .

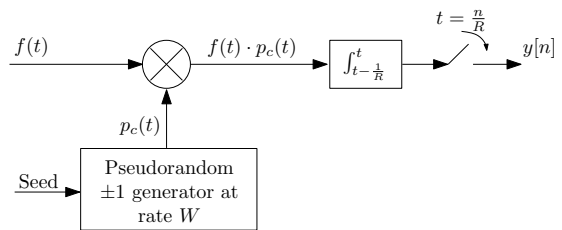


Fig. 10. Block diagram of the random demodulator.

The act of the random demodulator was studied for multi-tone signals:

$$f(t) = \sum_{w \in \Omega} a_w e^{j2\pi w t}, \quad (42)$$

where Ω is a finite set of K out of Q harmonics

$$\Omega \subset \{0, \pm\Delta, \pm2\Delta, \dots, \pm(Q-1)\Delta, Q\Delta\}. \quad (43)$$

We denote by \mathcal{N} the set of signals obeying this model, whose parameters are K , Q and the frequency resolution Δ . The normalization $\Delta = 1$ Hz is used in [10]. For this choice, the parameters are set to $W = Q$, and R is an integer dividing W . In addition $N_R = R$, which means observing $f(t)$ over the time interval $0 \leq t < 1$. Time-domain analysis for $f(t) \in \mathcal{N}$ shows that R samples on the interval $t \in [0, 1)$ correspond to

W samples taken by an integrate-and-dump block on Nyquist rate intervals. The relation is expressed as

$$\mathbf{y} = \Phi \mathbf{s} = \Phi_{\mathbf{H}} \Phi_{\mathbf{D}} \Phi_{\mathbf{F}} \mathbf{s}, \quad (44)$$

where $\mathbf{y} \in \mathbb{C}^R$ collects the samples. The unknown vector \mathbf{s} of length W contains the coefficients a_w , up to constant factors. Therefore, \mathbf{s} is K -sparse. In (44), the sampling matrix Φ of size $R \times W$ consists of a W -square reordered DFT matrix $\Phi_{\mathbf{F}}$, and $\Phi_{\mathbf{H}}$, $\Phi_{\mathbf{D}}$ represent the act of the integrator-and-dump and the sign waveform, respectively. To recover \mathbf{s} from \mathbf{y} , any CS algorithm for sparse recovery may be used. Roughly speaking, R should be on the order of $1.7K \log(W/K + 1)$ to allow stable recovery [10]. Then, $f(t)$ is constructed using \mathbf{s} and (42).

Among the assumptions made in [10], the following are essential for the random demodulator to function, at least from a theoretical viewpoint:

- (R1) The input $f(t)$ lies (or can be well approximated) on the grid (43);
- (R2) The grid resolution is $\Delta = (N_R/R)^{-1}$, which explains the time interval of 1 second for a model with 1 Hz frequency resolution;
- (R3) W/R is an integer;
- (R4) Integrate-and-dump architecture (rather than an accurate lowpass); and
- (R5) the waveform $p_c(t)$ is constant over time-intervals of length $1/W$.

To relax some of these constraints, the authors of [10] regard any deviation from the multitone model \mathcal{N} as an additive noise. In addition, for R which does not divide W , it is proposed to modify $\Phi_{\mathbf{H}}$. However this modification is signal dependent.

B. Detailed comparison

Table II highlights the differences between the spectrum-blind methods under consideration. The Whittaker, Kotelnikov, and Shannon (WKS) theorem, namely uniform sampling at the Nyquist rate, is added for reference. To simplify the discussion we assume that $f_s = f_p$ for the modulated wideband converter.

Before considering the table in detail, we present a simple toy-example that highlights the difficulty in implementing the random demodulator. Let $W = 1$ kHz, $R = 100$ Hz and observe the signal $f(t) = 3 \cos(2\pi 120t) + 4 \cos(2\pi 350t)$ for $t \in [0, 1)$. Applying the random demodulator to $f(t)$ gives $N_R = R = 100$ samples. Using basis pursuit (39) we can reconstruct $\hat{f}(t) = f(t)$ exactly. In practice, however, the rates R, W are triggered by a clock signal, whose frequency varies with temperature, humidity and other factors. We considered 1% frequency inaccuracy, namely $R' = 101$ Hz and $W' = 1.01$ kHz. In this case, simple reconstruction via basis pursuit leads to a complex-valued signal $\hat{f}(t)$. By adding constraints that enforce a real-valued solution, we obtain the reconstructed $\hat{f}(t)$ of Fig. 11(a), with the normalized squared-error $\|f - \hat{f}\|^2 / \|f\|^2 = 1.7$. The frequency contents are compared in Fig. 11(b). This example demonstrates that practical implementation of the random demodulator (even for

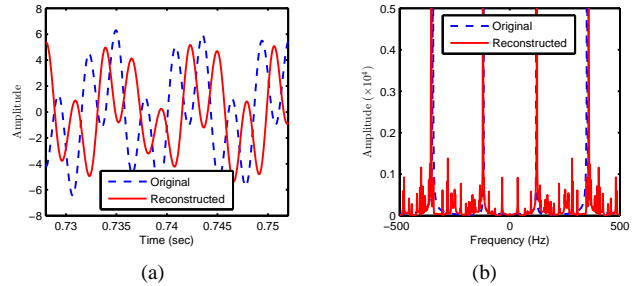


Fig. 11. Recovery of a multitone signal from random demodulator samples under design imperfections. The original and reconstructed signal are plotted in (a) on a short time interval. The frequency transforms (b) reveal many spurious tones due to the clock shift.

pure multitone signals and no noise) may be extremely difficult, since the slightest model mismatch (which is inevitable in real life) has a significant impact on the recovery error. Similar sensitivity appears in the requirement for accurate time delays in multicore sampling. In contrast, the WKS theorem is insensitive to the exact sampling rate, as long as it is above twice the maximal frequency. In Section IV-B we discussed robustness of our recovery algorithm in the presence of hardware inaccuracies.

We now consider the wideband scenario that was simulated in Section V-A, namely $N = 6$, $B = 50$ MHz and $f_{\text{NYQ}} = 10$ GHz. The dimension of the sensing matrix $\mathbf{A} = \text{SFD}$ in (19) is $m \times M$, and in particular Fig. 7 shows that stable recovery is achieved (empirically) with $m = 35$, $M = 195$. In contrast, the matrix Φ has dimensions $R \times W$, where W is the Nyquist rate. An analog multiband signal requires about $K = NB$ tones³ to approximate it within \mathcal{N} [10]. Consequently, in the wideband example, Φ has huge dimensions: $W = 10^{10}$ columns and $R = 2.6 \cdot 10^9$ rows (even in the toy-example above, Φ is of moderate size). Our approach enjoys several advantages due to the large difference in the dimensions of \mathbf{A}, Φ :

- Section V-A demonstrates that the CTF requires recording about $40 \cdot 35 = 1400$ samples, corresponding to observing the signal for 780 nsecs. To solve (44), $N_R = R = 2.6 \cdot 10^9$ samples are needed.
- CS recovery algorithms require (sometimes implicitly) to compute $\mathbf{A}_S^\dagger, \Phi_S^\dagger$. This is a highly demanding task with the dimensions of Φ .
- Now, suppose the band (or the tone) positions are unchanged for several seconds. In this scenario, there is no need to execute any sparse recovery algorithm, but only to apply $\mathbf{A}_S^\dagger, \Phi_S^\dagger$ repeatedly on the new samples. One should again record R samples (= 1 second delay) before applying Φ_S^\dagger to the samples. Then, about $KR = 780 \cdot 10^9$ million instructions per second (MIPS) are performed when calculating the nonzero coefficients in \mathbf{s} . In contrast, (33) is applied at the input rate and

³In fact, [10] show that analog multiband signals are not well approximated within \mathcal{N} , unless convolving them by a window function prior to sampling. Theoretically, the convolved version can be stably reconstructed, although it is not clear which window function to choose and how to extract the original analog signal from the windowed solution.

TABLE II
METHODS FOR WIDEBAND SPECTRUM-BLIND RECOVERY.

	WKS theorem $y[n] = x(nT_s)$	Multicoset Eq. (2)	Random demodulator Fig. 10	Modulated wideband converter Fig. 3	
Theory	Model	β -Bandlimited	Multiband \mathcal{M}	Multitone \mathcal{N}	
	Type	Continuous	Continuous	Discrete parameterization	
	Model parameters	β	N, B, T	K, Q, Δ	N, B, T
	Sampling parameters	T_s	m, M, C	$R, W, N_R, p_c(t)$	$m, M, f_p, f_s, \alpha_{ik}$
	Restrictions	$1/T_s > 2\beta$	see [5]	(R1)-(R5)	Theorem 2
	Rate granularity	Any T_s	m/MT	$R = W/r, r \in \mathbb{Z}$	mf_s
	Analysis domain	Frequency	Frequency	Time	Frequency
Practice	Design sensitivity	Low	High	High	
	Number of channels	1	m	1	
	Sensing matrix dimensions	None	$m \times M$ (small)	$R \times W$ (huge)	$(mf_s/f_p) \times M$ (small)
	Memory req. (support recovery)		Constant $\times m$	R	Constant $\times m$
	Memory req.*		None	R	None
	Latency*	real time	real time	$N_R/R = 1$ sec	real time
	MIPS*		$2Nm/T$	RK	$2Nm f_s$
	Baseband processing		Yes*		Yes
Technology barrier [14], [23] ADC's front-end bandwidth (~ 1 GHz) CS algorithms [‡] (~ 10 MHz) Waveform generator (~ 23 GHz)					

* After support is recovered.

‡ Our estimate.

requires no recording of samples. The computation rate is $2Nm = 420$ multiplications per sample interval T_s , or $2Nm f_s = 22 \cdot 10^3$ MIPS, which is 6 orders of magnitude lower than [10].

We are now in a position to explain the spectrum-blind baseband processing ability of our approach, which is a direct consequence of the above discussion. Suppose the analog input has a time-varying spectral content. In practice, a high-level application layer triggers the CTF when the previous recovered support becomes invalid. The CTF introduces a short latency ($\sim 1 \mu\text{sec}$ in our simulations). Then, each $z_i[n]$ is a digital sequence at (the low) rate f_s , where for $i \in S$, $z_i[n]$ corresponds to a specific band⁴. The recovery architecture of Fig. 6 allows to select a specific band of interest for further processing (at rate f_s) by generating only the appropriate sequence $z_i[n]$. In contrast, [10] aim at the recovery of the Nyquist rate sequence. The latency and computational loads of this approach do not allow baseband processing. As explained in Section IV-B, for multicoset sampling, the frame construction in the CTF requires interpolating the sequences to the Nyquist rate, thus baseband processing applies in this setting only after the support is recovered.

We now compare several other aspects between the systems, starting with rate granularity. To design a signal independent random demodulator R must divide W , where W is fixed to the Nyquist rate of $f(t)$. Thus, for example, for $W = 10$ GHz, there is a significant rate increase between $R = 1.25$ GHz and $R = 2.5$ GHz. In contrast, Fig. 3 allows rate granularity at steps $f_p = B$. The experiment in Section V-A demonstrates recovery from sampling at rate $35f_s \simeq 1.8$ GHz, which saves 30% of the sampling rate with respect $R = 2.5$ GHz.

The choice of analog devices is also interesting to compare. The time-domain analysis in [10] requires $p_c(t)$ to be constant

⁴If the band splits between two spectrum slices in $\mathbf{z}(f)$, then a simple merging operation generates a sequence at rate $2f_s$.

over $1/W$ time intervals, for (44) to hold. In contrast, the mixing functions $p_i(t)$ are only required to be T_p -periodic, such that their Fourier series contains enough dominant coefficients c_{il} . For example, a sawtooth waveform with varying amplitudes, where T_p/M is the duration of a single teeth, would also be appropriate for $p_i(t)$. This choice avoids the discontinuities in the sign waveform, which are a major source for analog noise. The random demodulator, however, cannot incorporate such a waveform for $p_c(t)$ unless compromising on a signal dependent matrix $\Phi_{\mathbf{H}}$. Such dependence is not desired of course when the signal is unknown. The integrate-and-dump, a first-order lowpass filter, has also a main role in the time-domain analysis, while arbitrary high-order accurate lowpass filters are allowed in Fig. 3.

To conclude, we point out the technology barrier of each approach. The front-end of a practical ADC limits the applicability of the multicoset as explained in Section II-B. Uniform sampling at the Nyquist rate shares the same barrier. The above discussion shows that the computational load and memory requirements in the digital domain are the bottleneck of the random demodulator approach. Therefore the size of CS problems that can be solved with available DSPs limits the recovery. We estimate that $W \approx 1$ MHz may be already quite demanding using convex solvers, whereas $W \approx 10$ MHz is probably the barrier using greedy methods. In fact, uniform sampling at 10 MHz seems to be preferred in this setting. Our system is limited by the technology for generating the periodic waveforms $p_i(t)$, which depends on the specific choice of waveform. The applicability range of each method appears in Table II. Our approach, Figs. 3 and 6, provides many advantages for standard analog applications in the wideband regime. Furthermore, even at low rates, the realtime processing becomes a significant benefit over both Nyquist rate sampling and the random demodulator [10].

VII. CONCLUSIONS

We presented a sub-Nyquist sampling system, the modulated wideband converter, which is designed independently of the spectral support of the input signal. The analog front-end supports wideband applications and can also be used to sample wideband inputs occupying the entire spectral support. A spectrum-blind recovery stage was also developed consisting of digital support recovery and an analog back-end. The digital operations required for the support recovery require only a small number of observations, thus introducing a short delay. Once the support is known, various realtime computations are possible. Recovery of the original signal at the Nyquist rate is only one possible application. Perhaps more important is the ability to digitally process any information band at a low rate, which we refer to as spectrum-blind baseband processing.

Our approach is compared to other sub-Nyquist strategies, such as nonuniform sampling and the random demodulator. The detailed discussion highlights the main advantages of our method both in terms of hardware implementation and light computational loads.

In this paper we focused on the engineering aspects of the system. In [12] we will examine various tradeoffs regarding the mixing waveforms used in our front-end, including guarantees on stable recovery. We are currently also implementing our system on board to demonstrate real-time applicability.

The current work embeds theorems and algorithms from compressed sensing (CS), an emerging research field which exploits sparsity for dimension reduction. The mainstream line of CS papers studies sparsity for discrete and finite vectors. The random demodulator expands this approach by parameterizing continuous signals in a finite setting. In contrast, this work continues the line of [5], [11] and belongs to a recently-developed framework within CS [34], [40], [41], which studies signals from a truly continuous domain. Within this analog framework, we propose selecting a practical implementation among the various possible sampling stages covered by [40].

ACKNOWLEDGMENT

The authors would like to thank Prof. Joel A. Tropp for fruitful discussions and for helpful comments on the first draft of this manuscript.

REFERENCES

- [1] H. J. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Math.*, vol. 117, pp. 37–52, Feb. 1967.
- [2] 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.
- [3] C. Herley and P. W. Wong, "Minimum rate sampling and reconstruction of signals with arbitrary frequency support," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1555–1564, Jul. 1999.
- [4] R. Venkataramani and Y. Bresler, "Perfect reconstruction formulas and bounds on aliasing error in sub-Nyquist nonuniform sampling of multiband signals," *IEEE Trans. Inform. Theory*, vol. 46, no. 6, pp. 2173–2183, Sep. 2000.
- [5] M. Mishali and Y. C. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Processing*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.
- [6] W. Black and D. Hodges, "Time interleaved converter arrays," in *Solid-State Circuits Conference. Digest of Technical Papers. 1980 IEEE International*, vol. XXIII, Feb., pp. 14–15.
- [7] Y.-C. Jenq, "Digital spectra of nonuniformly sampled signals: A robust sampling time offset estimation algorithm for ultra high-speed waveform digitizers using interleaving," *IEEE Trans. Instrum. Meas.*, vol. 39, no. 1, pp. 71–75, 1990.
- [8] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Select. Areas Commun.*, vol. 17, no. 4, pp. 539–550, 1999.
- [9] J. N. Laska, S. Kirolos, M. F. Duarte, T. S. Ragheb, R. G. Baraniuk, and Y. Massoud, "Theory and implementation of an analog-to-information converter using random demodulation," in *Proc. IEEE Int. Sym. Circuits and System*, May 2007, pp. 1959–1962.
- [10] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. Baraniuk, "Beyond nyquist: Efficient sampling of sparse bandlimited signals," *arXiv.org 0902.0026*, Jan. 2009.
- [11] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. Signal Processing*, vol. 56, no. 10, pp. 4692–4702, Oct. 2008.
- [12] M. Mishali, J. A. Tropp, and Y. C. Eldar, "Theoretical study of sub-Nyquist sampling for sparse wideband analog signals," *in preparation*, 2009.
- [13] "A/D Converter - Definition of terms," *App. notes, National Semiconductors Corp.*, Jan. 2000, [Online]. Available: http://www.national.com/appinfo/adc/files/definition_of_terms.pdf.
- [14] "A/D Converters," *Analog Devices Corp.*, Feb. 2009, [Online]. Available: <http://www.analog.com/en/analog-to-digital-converters/ad-converters/products/index.html>.
- [15] M. Fleyer, A. Rosenthal, A. Linden, and M. Horowitz, "Multirate synchronous sampling of sparse multiband signals," *Arxiv preprint arXiv:0806.0579*, 2008.
- [16] Y. C. Eldar and A. V. Oppenheim, "Filter bank reconstruction of bandlimited signals from nonuniform and generalized samples," *IEEE Trans. Signal Processing*, vol. 48, no. 10, pp. 2864–2875, 2000.
- [17] H. Johansson and P. Lowenborg, "Reconstruction of nonuniformly sampled bandlimited signals by means of digital fractional delay filters," *IEEE Trans. Signal Processing*, vol. 50, no. 11, pp. 2757–2767, 2002.
- [18] J. Elbornsson, F. Gustafsson, and J.-E. Eklund, "Blind equalization of time errors in a time-interleaved ADC system," *IEEE Trans. Signal Processing*, vol. 53, no. 4, pp. 1413–1424.
- [19] A. J. Viterbi, "CDMA principles of spread spectrum communication," *Addison-Wesley Wireless Communications Series*, 1995.
- [20] R. Pickholtz, D. Schilling, and L. Milstein, "Theory of Spread-Spectrum Communications – A Tutorial," *IEEE Trans. Commun.*, vol. 30, no. 5, pp. 855–884, 1982.
- [21] C. Kienmayer, M. Tiebout, W. Simburger, and A. L. Scholtz, "A low-power low-voltage NMOS bulk-mixer with 20 GHz bandwidth in 90 nm CMOS," *Circuits and Systems, 2004. ISCAS '04. Proceedings of the 2004 International Symposium on*, vol. 4, pp. IV–385–8 Vol.4, May 2004.
- [22] B. Razavi, "A 60-GHz CMOS receiver front-end," *IEEE J. Solid-State Circuits*, vol. 41, no. 1, pp. 17–22, Jan. 2006.
- [23] E. Laskin and S. P. Voinigescu, "A 60 mW per Lane, 4×23 -Gb/s 2^7-1 PRBS Generator," *IEEE J. Solid-State Circuits*, vol. 41, no. 10, pp. 2198–2208, Oct. 2006.
- [24] T. O. Dickson, E. Laskin, I. Khalid, R. Beerkens, X. Jingqiong, B. Karajica, and S. P. Voinigescu, "An 80-Gb/s $2^{31}-1$ pseudorandom binary sequence generator in SiGe BiCMOS technology," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2735–2745, Dec. 2005.
- [25] A. I. Zverev, *Handbook of Filter Synthesis*. Wiley-Interscience, 1967.
- [26] K. Gentile, "Introduction to zero-delay clock timing techniques," *App. notes AN-0983, Analog Devices Corp.* [Online]. Available: http://www.analog.com/static/imported-files/application_notes/AN-0983.pdf.
- [27] D. L. Donoho and M. Elad, "Maximal sparsity representation via ℓ_1 minimization," *Proc. Natl. Acad. Sci.*, vol. 100, pp. 2197–2202, Mar. 2003.
- [28] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Processing*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.
- [29] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [30] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.
- [31] 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 Processing*, vol. 53, no. 7, pp. 2477–2488, July 2005.

- [32] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process. (Special Issue on Sparse Approximations in Signal and Image Processing)*, vol. 86, pp. 572–588, Apr. 2006.
- [33] —, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Process. (Special Issue on Sparse Approximations in Signal and Image Processing)*, vol. 86, pp. 589–602, Apr. 2006.
- [34] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a union of subspaces," *arXiv.org 0807.4581*; submitted to *IEEE Trans. Inform. Theory*, Jul. 2008.
- [35] M. Mishali, Y. C. Eldar, and J. A. Tropp, "Efficient sampling of sparse wideband analog signals," in *Proc. of IEEEI, 25th convention*, Dec. 2008, pp. 290–294.
- [36] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Const. Approx.*, 2007.
- [37] T. Blumensath and M. E. Davies, "Sampling theorems for signals from the union of linear subspaces," *accepted to IEEE Trans. Inform. Theory*, 2008.
- [38] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1999.
- [39] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *C. R. Acad. Sci. Paris, Ser. I*, vol. 346, pp. 589–592, 2008.
- [40] Y. C. Eldar, "Compressed sensing of analog signals," *arXiv.org 0806.3332*; submitted to *IEEE Trans. Signal Processing*, Jun. 2008.
- [41] —, "Uncertainty relations for analog signals," *arXiv.org 0809.3731*; submitted to *IEEE Trans. Inform. Theory*, Sep. 2008.