

# Xampling—Part I: Practice

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

**Abstract**—We introduce Xampling, a design methodology for sub-Nyquist sampling of continuous-time analog signals. The main principles underlying this framework are the ability to capture a broad signal model, low sampling rate, efficient analog and digital implementation and lowrate baseband processing. The main hypothesis of Xampling is that in order to break through the Nyquist barrier, one has to combine classic methods and results from sampling theory together with recent developments from the literature of compressed sensing. In this paper, we present the Xampling framework and examine several sub-Nyquist approaches in light of the four Xampling principles. It is shown that previous methods suffer from analog implementation issues, large computational loads in the digital domain, and have no baseband processing capabilities. An exception is the recently proposed modulated wideband converter (MWC) which satisfies the model, rate and implementation criteria, though lacking the baseband processing capability. Here, we extend the MWC by proposing a digital algorithm which extracts each band of the signal from the compressed measurements, thus enabling lowrate (baseband) processing. The converter with the proposed algorithm conforms with the Xampling desiderata. In addition, we describe two configurations of the converter for efficient spectrum sensing in wideband cognitive radio receivers. In the second part of this work we study theoretical aspects of rate and stability of sub-Nyquist systems, following the pragmatic theme of the Xampling methodology.

**Index Terms**—Baseband processing, cognitive radio, compressed sensing, modulated wideband converter, sub-Nyquist sampling, Xampling.

## I. INTRODUCTION

**S**IGNAL processing methods have changed substantially over the last several decades. The number of operations that are shifted from analog to digital is constantly increasing, leaving amplifications and fine tunings to the traditional front-end. In the chain of sampling, processing and reconstruction, the conversion to digital has become a serious bottleneck. While technology advances enable mass processing of huge data streams, the acquisition capabilities do not scale sufficiently fast [1]. For some applications, the maximal frequency of the input signals, which dictates the Nyquist rate, already exceeds the possible rates achievable with existing devices. Sampling theory, the gate to the digital world, is needed to break through the rate bottleneck.

Consider the scenario depicted in Fig. 1, which is prevalent in communication systems. A few narrowband transmissions are modulated onto carrier frequencies  $f_i$ , which can take on any value below  $f_{\max}$ . This leads to a *multiband* spectral support that occupies only a small portion of the wideband

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.

The authors are with the Technion—Israel Institute of Technology, Haifa 32000, Israel (email: moshiko@tx.technion.ac.il; yonina@ee.technion.ac.il; elron@tx.technion.ac.il).

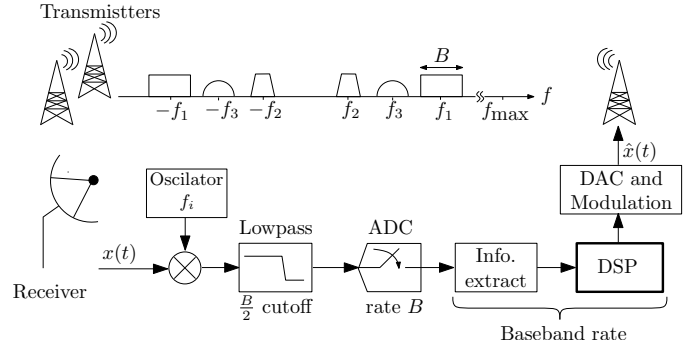


Fig. 1. Three RF transmissions with different carriers  $f_i$ . The receiver demodulates each transmission separately and samples the baseband version.

spectrum defined by  $f_{\max}$ . The receiver converts each transmission to digital by demodulating the carrier frequencies  $f_i$ . Once the transmission contents appear at baseband, that is near the origin, they are lowpass filtered and sampled at a low rate. In the example, the three concurrent transmissions result in a signal  $x(t)$  which is supported on  $N = 6$  frequency intervals, or bands, each of width not greater than  $B$  Hz. This approach leads to sampling at a rate that is proportional to  $NB$ , rather than to the radio-frequency (RF)  $f_{\max}$ , which can be prohibitively large in modern applications. Depending on the modulation technique, the information, either a bit stream or an analog message, is extracted from the samples. Often, this operation involves a matched filter.

Digital signal processing (DSP) is the crowning glory of the chain of blocks in Fig. 1. The prime goal of analog to digital conversion (ADC) is isolating the delicate interaction with the continuous world, so that sophisticated algorithms can be developed in a flexible software environment. Digital filtering, channel equalization, system identification, blind source separation, noise shaping and a rich variety of software algorithms – all lie under the DSP block of Fig. 1. Besides processing, reconstruction of the input  $x(t)$  can be obtained by digital to analog conversion (DAC) and remodulating onto the original carriers  $f_i$ . This option is useful in relay stations that re-transmit the input after local improvements to the signal. All digital computations are carried out at the actual information rate, which is referred to hereafter as *baseband processing*.

Utilizing the scheme of Fig. 1 requires knowing the carrier frequencies  $f_i$ . As explained in Section II, this approach can tolerate only slight deviations from the prespecified carrier values  $f_i$  and cannot extend to arbitrary spectral support. Classic works in sampling theory [2]–[6] study periodic nonuniform sampling as an alternative, though these solutions also rely on knowledge of the carrier frequencies  $f_i$ . The literature describes several sub-Nyquist strategies, other than Fig. 1, that have the potential to treat arbitrary carrier positions: multi-

coset sampling [7], the Nyquist-folding ADC [8], the random demodulator [9] and its parallel version [10], and the modulated wideband converter [11], [12]. The approaches differ in the signal model they assume, in the sampling strategy and in the recovery algorithms. Research on sub-Nyquist sampling has so far focused on perfect recovery of the Nyquist-rate input signal. The ability to process the information at baseband, namely without involving Nyquist-rate computations, was not addressed. In fact, as shown in this paper, none of the systems proposed in [2]–[12] allow for baseband DSP, including the previous works [7], [11] by the authors. Admittedly, whenever  $x(t)$  can be recovered, DSP at the high Nyquist rate is possible. However, this solution wastes the DSP resources and typically results in impractical processing rates.

The main contribution of this two-part series is a design methodology for sub-Nyquist systems, named Xampling. The nomenclature we chose will be explained below. Part one, the present paper, begins with the practical aspects of the Xampling methodology. The framework consists of four criteria: broad signal model, low sampling rate, efficient analog and digital implementation and the baseband processing capability. Our previous publications on multiband sampling incorporated the model and the rate considerations [7], [13], whereas [11] added the aspect of practical implementation. The Xampling methodology has matured from these works and [12]. The present paper capitalizes on the necessity of baseband processing and on the digital input type that standard DSP packages are designed to deal with. Section II elaborates more on these properties, which we consider as ground rules for a successful sub-Nyquist solution. In the same section, we address the relation to compressed sensing (CS), an active research field that was triggered by the works of [14], [15]. Mainstream CS works study the problem of recovering a sparse vector from an underdetermined linear system. In contrast, Xampling is a broad framework for treating analog signals, which takes into account considerations that are not treated in discrete CS.

Part two of Xampling [16] capitalizes on the difference between classic sampling theory and sub-Nyquist from a deeper theoretical perspective. The former relies on a single subspace model [17], whereas analog sparsity, such as multiband with unknown carriers, fits the model of a union of subspaces [12], [18]–[22]. Continuing the pragmatic theme of the present paper, part two introduces additional Xampling criteria for the theoretical study of rate and stability of sub-Nyquist systems.

The first part of the present paper is devoted to support the proposed methodology and its design rules. We revisit [7]–[12] in light of the four Xampling criteria. Our survey is inspective rather than descriptive – each strategy is briefly overviewed, after which we study various practical considerations, which did not appear in the original publications. We begin with pointwise strategies which are shown to suffer from practical analog bandwidth limitations of existing hardware devices. These limitations hold regardless of whether knowledge of the carrier positions is available [2]–[6] or absent [7], and become acute for wideband signals. The Nyquist-folding system [8] also involves possible pointwise limitations. We then proceed to examine in detail two state-of-the-art systems: the random demodulator (RD) [9], [10] and the modulated

wideband converter (MWC), proposed by the authors in [11]. Our analysis relies on the viewpoint of *equivalent systems*, explained in Section IV. The comparison reveals that the RD relies on a sensitive signal model when dealing with analog signals, the time-domain approach boils down to difficult implementation requirements and the computational complexity is severe, as discussed in detail in Section IV. In contrast, the MWC naturally fits analog models, the implementation [23] is based on the standard frequency-domain viewpoint and the required computational complexity is by orders of magnitude smaller compared with the RD. Unfortunately, we conclude that all the methods we survey, including the MWC, do not support baseband processing. Therefore, the impact of [2]–[12] reduces to digital storage of the samples and reconstruction of the input  $x(t)$ . The prominent advantage of DSP at baseband, which is the prime reason for shifting to digital, is not achieved by these systems.

Our second contribution is a digital algorithm that translates the MWC outputs to the required format for baseband processing. We choose to treat the MWC system since it satisfies all the other criteria of Xampling. Two ingredients that are computed by [11] are prerequisites: a coarse estimate of the spectral support and a set of lowrate sequences that together capture the signal information. The algorithm consists of three stages: refining the frequency support estimate, isolating the transmissions, and finally a digital carrier recovery stage. The last step utilizes the balanced quadricorrelator, a reliable frequency detector which is suitable for many data transmission techniques [24]. As a nice feature, we show that once the algorithm is applied and the information is extracted from the samples, the input  $x(t)$  can be reconstructed more efficiently than the method proposed in [11]. Numerical simulations demonstrate the algorithm accuracy in typical noisy wideband scenarios. The consequence is that any existing DSP algorithm of interest can smoothly interface with the MWC.

The third and final contribution of the paper applies the MWC to spectrum sensing, one of the fundamental tasks in cognitive radio systems [25]. The cognitive receiver uses spectrum sensing to decide on available spectrum regions that may be useful for transmissions, until the licensed owner of those bands appear. We present two configurations of the MWC for spectrum sensing which rely on the implementation advantages of the MWC in the wideband regime, and on our digital algorithm. Related works in the field [10], [26]–[30] are discussed and compared within the unified Xampling framework.

The paper is organized as follows. Section II introduces the Xampling methodology, addresses the relation to CS and explains the nomenclature. The three following sections are dedicated to examine the validity of the methodology: lowrate pointwise sampling in Section III; the RD in Section IV; and the MWC in Section V. A mid-term summary is given in Section VI. The digital algorithm that enables baseband processing with the MWC is provided and simulated in Section VII. Finally, in Section VIII, we explain how to apply the MWC and the proposed algorithm to efficient spectrum sensing in sub-Nyquist cognitive radio receivers.

TABLE I  
THE XAMPLING CRITERIA

Criterion	Symbol	Requirement
Signal model	(X1)	broad set of analog signals
Sampling rate	(X2)	approach the minimal
Implementation	(X3)	existing analog devices, light computational loads, (preferably in realtime)
Processing	(X4)	at baseband (preferably in realtime)

## II. XAMPLING

### A. Methodology

This section introduces the Xampling framework and its four design principles. The main theme of Xampling is that a sub-Nyquist system should satisfy all four principles in order to break through the Nyquist barrier. This claim is validated by a thorough study of the practical aspects of a series of sub-Nyquist methods, in the current section and in the ones to follow. The Xampling rules are defined below, and are briefly summarized in Table I for reference.

**Signal model (X1).** The system should be able to handle a broad set of inputs with different spectral contents without altering either the hardware or the digital software. Therefore, when designing the sampling and the reconstruction stages, the goal should be treating analog signals with arbitrary carrier positions  $f_i$  till  $f_{\max}$ , as performed in Fig. 1 for the case of known carrier locations. Extensions to other sparsity models will be treated in future work [22]. Any such model should gracefully approximate the continuous structure of analog sparsity.

**Sampling rate (X2).** The conversion rate is the second criterion, which should be as low as possible in sub-Nyquist systems. Ultimately, one would like to approach the theoretical minimal rate  $2NB$  [7]. In practice, stability considerations require higher rates. In that context, we say that (X2) is satisfied if the theoretical lowest rate due to stability reasons is proportional to  $NB$  (preferably with a small factor), and if the system allows to approach that rate.

**Efficient implementation (X3).** This criterion refers to the ability to realize the sampling strategy with a reasonable number of existing analog devices. It also refers to the computational burden in the digital domain, which should be as light as possible.

**Baseband processing (X4).** A sub-Nyquist system should enable processing of the information contents at a low rate. Baseband processing at the low rate is perhaps the most practical property of sub-Nyquist systems to consider, since the theoretical interest in perfect reconstruction of the original high-rate analog  $x(t)$  is often less useful. To define baseband processing more precisely, we consider the standard quadrature representation of communication signals [24]:

$$s(t) = I(t) \cos(2\pi f_c t) + Q(t) \sin(2\pi f_c t), \quad (1)$$

where  $I(t), Q(t)$  are real-valued narrowband signals, referred to as the information signals, and  $f_c$  is a relatively high

carrier frequency. For example: in analog amplitude modulation (AM),  $Q(t) = 0$ , and the information is carried out in the amplitude of  $I(t)$ . Phase- and frequency-modulation (PM/FM) obey (1) when properly recasting their conventional representation by  $I(t) = \cos(g(t)), Q(t) = -\sin(g(t))$ , such that the analog message is  $g(t) = \arctan(I(t)/Q(t))$  [31]. Various digital techniques, such as frequency- or phase-shift keying (FSK/PSK) also conform with (1). In digital modulation techniques, a symbol encodes one or more information bits. In the basic form, two bitstreams  $I[n], Q[n] \in \{\pm 1\}$  together encode 2 bits/symbol. The analog versions  $I(t), Q(t)$  are pulse-shaped by some narrowband function  $p(t)$  according to

$$I(t) = I[n]p(t - nT_{\text{sym}}), \quad Q(t) = I[n]p(t - nT_{\text{sym}}), \quad (2)$$

with  $T_{\text{sym}}$  being the symbol duration in time. In other digital methods, each  $I[n], Q[n]$  encodes more than a single bit, e.g. in 256-quadrature amplitude modulation (256-QAM),  $I[n], Q[n] \in \{\pm 1, \pm 3, \pm 5, \pm 7\}$ , encoding 8 bits per  $T_{\text{sym}}$ . Therefore, we say that a sub-Nyquist system has the baseband processing capability if  $I(t), Q(t)$  can be extracted from the digital samples, using computation complexity that is proportional to  $NB$ . In particular, there is no need to interpolate the samples to the high Nyquist rate

$$f_{\text{NYQ}} = 2f_{\max}. \quad (3)$$

In the sequel, we refer to the continuous format  $I(t), Q(t)$  as the information signals, under the convention that the DSP actually expects the uniformly-spaced versions of  $I(t), Q(t)$  at rates corresponding to their actual bandwidths. Based on the modulation-technique in use, the DSP device can convert the uniform samples to the bit streams versions  $I[n], Q[n]$  of (2) or the relevant analog message  $g(t)$ . The baseband processing capability is unnecessary when the sole purpose of the system is to store the samples and later recover the analog input. However, in many cases the conversion to digital is carried out with the intention of shifting processing operations from analog to digital, in which case baseband processing is an important factor.

**Realtime processing.** In the border between (X3) and (X4) lies another computational-related property, which is termed realtime processing. This means that the delay, which is introduced due to the computations involved in either reconstruction of  $x(t)$  or when extracting the information contents  $I(t), Q(t)$ , is short. In such a setting, the throughput towards the DSP device is not limited by the acquisition system. We do not consider the realtime property essential for a sub-Nyquist design, since offline applications do not require a short delay. We shortly address the realtime properties of the systems we survey in Section VI.

We point out that the Xampling criteria address the scenario of carriers which are unknown but fixed. When the spectral support changes in time, an intuitive requirement is to track the band locations, a combination of (X1) and (X4). Here, realtime processing is essential so as to return to signal acquisition and processing as fast as possible after a support change happens. Strictly speaking, a signal with time-varying support is not multiband, and in fact not even a bandlimited signal.

Neglecting this mathematical issue, a realtime system should react gracefully to spectral changes.

**Examples.** Before proceeding, we examine two straightforward sampling solutions in light of the Xampling criteria. Uniform sampling at  $f_{\text{NYQ}}$  obviously contradicts **(X2)** and cannot be considered as a sub-Nyquist system. Furthermore, in wideband settings, the implementation may be impractical since the rates of conventional Nyquist ADC devices are still far below the wideband regime [32], contradicting **(X3)**. Baseband processing **(X4)** is also not possible, since by definition the samples arrive at the high Nyquist rate, thus extracting the information signals  $I(t), Q(t)$  must involve computational complexity proportional to  $f_{\text{NYQ}}$ .

The second approach utilizes the scheme of Fig. 1, by searching for the carrier frequencies  $f_i$  prior to sampling, namely using analog components. The motivation for this solution is two-fold. Standard receivers often compensate for slight deviations from the prespecified carrier values  $f_i$ , by performing fine analog tunings to the local oscillator until its frequency is locked to the actual carrier value [33], [34]. The idea is therefore to use the same locking mechanism to search for  $f_i$  over the entire wideband regime. The fact that Fig. 1 satisfies all the four Xampling criteria, once the carriers  $f_i$  are known, is the second motivation.

Unfortunately, this solution is practically infeasible when the carrier frequencies  $f_i$  are unknown a priori and can lie anywhere below  $f_{\text{max}}$ . Locking over the entire wideband spectrum is time consuming; during this time the signal cannot be acquired. To shorten this period, the sampling rate must be increased much above the minimal, contradicting **(X2)**. Furthermore, such a locking stage is both hardware and software excessive. A standard tunable oscillator can cover only a narrow range of frequencies [35], which may require hundreds of devices to cover the range until  $f_{\text{max}}$ . In addition, when initializing the mechanism far away from the true carrier, it may lock to a spurious frequency. Only high-level data-aided algorithms can identify this situation and re-initialize the hardware. This severely burdens the DSP, contradicting also **(X3)**. Furthermore, whenever the band positions change, the locking needs to be reinitiated, and again the signal cannot be acquired until this task is completed.

### B. Xampling = Compressed sensing for analog signals

We now briefly describe the CS framework [14], [15], and distinguish between CS and Xampling.

The majority of compressed sensing publications study variants of the underdetermined sparse recovery problem. The signal model assumes a vector  $\mathbf{x}$  in the finite space  $\mathbb{R}^n$  or  $\mathbb{C}^n$ , which has only a few nonzero entries. Sampling, referred to as sensing in this framework, is carried out by computing the linear projection

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad (4)$$

with  $\mathbf{A}$  having far fewer rows than columns. Results from this field [14], [15] show that under suitable conditions, the linear sensing is stably invertible, even when the length of  $\mathbf{y}$  is proportional to the number of nonzeros in  $\mathbf{x}$ , rather than the ambient dimension  $n$ .

Sensing of sparse vectors is the discrete counterpart of the sub-Nyquist problem illustrated in Fig. 1. However, it is not straightforward to generalize the discrete CS formulation to analog signals. The difficulty can be noticed immediately in the signal model. Sparsity is defined in CS by counting the number of nonzeros in  $\mathbf{x}$ , while analog sparsity of  $x(t)$  involves an uncountable number of zeros and nonzeros. Another difference relates to the sensing matrix  $\mathbf{A}$ . In the analog context,  $\mathbf{A}$  corresponds to generalized sampling, where measurements are inner products with the input  $x(t)$  [17], [36]–[38]. This leads to a structured and deterministic matrix  $\mathbf{A}$ , as it needs to be implemented in hardware. In contrast, mainstream CS results are stated for random unstructured sensing matrices. A final important difference is the issue of recovery complexity. Naïve extensions of CS-type algorithms to the infinite dimensions, such as  $\ell_1$  linear programming

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ s.t. } \|\mathbf{y} - \mathbf{A}\mathbf{x}\| \leq \epsilon, \quad (5)$$

or greedy techniques, lead to undefined or difficult problems. For example, an optimization over the continuous signal  $x(t)$  [39]

$$\min_{x(t)} \text{obj}(x(t)) \text{ s.t. } \|y[n] - \mathcal{A}(x(t))\|_{l_2} \leq \epsilon, \quad (6)$$

where the objective is a sparsity-promoting function  $L_2(\mathbb{R}) \rightarrow \mathbb{R}$  and the constraint involves the infinite sample set  $y[n]$ , the sampling operator  $\mathcal{A} : L_2(\mathbb{R}) \rightarrow l_2(\mathbb{R})$ , and the continuous signal. A program such as (6) is not a well-defined optimization structure [39]. In turn, discretization methods result in very large scale CS systems, which impose a severe burden on the digital processing units. In contrast, continuous reconstruction in sampling theory, though hypothetically involving infinite sequences, is practically performed in realtime over a well localized set of samples.

The baseband processing is another distinct aspect. Lately, there has been growing interest in processing in the compressed domain, e.g. machine learning and classifications [40], or manifold learning [41], [42]. These works exploit the fact that certain properties are approximately invariant under a linear transformation. This allows, to some extent, learning and classification tasks to be carried out directly on the short-length vector  $\mathbf{y}$  of (4). The methods [40]–[42] are specific to discrete vectors. In contrast, sub-Nyquist methods, both classic [3]–[6] and recent [7]–[12], are all based on spreading or aliasing techniques which result in a mixture of the signal content. In particular, the information contents  $I(t), Q(t)$  are not invariant under the mixture, and a DSP algorithm cannot be carried out directly on the samples. To emphasize, baseband processing in Xampling means the ability to translate the seemingly-corrupted samples to the original information bits or analog message. Learning, classification, and other DSP algorithms can then follow so that there is no need to approximate their results in the compressed domain. The crucial requirement of **(X4)** is that the information is to be extracted without interpolating to the Nyquist grid.

The CS paradigm aims at avoiding high-rate redundant sampling. The discrete CS framework [14], [15] initiated a long line of highly influential works. However, it still remains

puzzling from the analog sampling viewpoint; sensing by (4) implicitly assumes that  $\mathbf{x}$  is the Nyquist rate samples of some continuous signal  $x(t)$  on a specific time-interval. Recent works [7]–[12], which are surveyed in this paper, attempt to extend the formulations of CS in various ways, so as to render them applicable to analog sub-Nyquist sampling. As we shall review in the next sections, these approaches are quite different from each other, especially in terms of signal model, rate and implementation. We propose Xampling as a framework within the various solutions can be compared according to the four rules of Table I.

The nomenclature Xampling was chosen to highlight the important aspects of our framework. Sub-Nyquist systems rely on the theory of sampling from a union of subspaces [12], [20]–[22], whereas classic results were stated for single subspace models [17]. The X prefix therefore distinguishes analog sub-Nyquist systems from classic results in the sampling literature, though Xampling still hints that our framework is only a sub-field of generalized sampling theory [17], [36]–[38]. The naming has a symbolic interpretation as well. The X letter is widely used to denote compression, e.g. the DivX format. Here, the compression is carried out during the conversion from analog to digital, conceptually using analog means rather than by software algorithms. The fact that the prefix is integrated into the noun symbolizes the integration between sampling and compression. More importantly, Xampling conveys a guideline for designers. As we shall observe in the next sections, breaking-through the Nyquist barrier necessitates balancing between CS and sampling by combining traditional concepts from sampling theory together with recent developments from the CS literature. When either is missing, one of the Xampling criteria is prone to be violated. Xampling is literally pronounced as CS-Sampling (phonetically /k'sæmpəlɪŋ/), so as to symbolize the necessity of this synergy in practice. Finally, it was recently suggested to us [43] that X can stand for extreme sampling, hinting at the very low rates.

### III. POINTWISE SAMPLING

In this section we describe sampling methods which involve pointwise sampling of the input signal  $x(t)$ .

#### A. Periodic nonuniform sampling

Periodic nonuniform sampling (PNS) utilizes a set of undersampling sequences with relative time-shifts. Specifically, consider a signal  $x(t)$  with Nyquist rate  $f_{\text{NYQ}} = 1/T$ . PNS generates  $m$  lowrate sequences

$$y_i[n] = x(nT_s + \phi_i), \quad 1 \leq i \leq m, \quad (7)$$

with a relatively long sampling interval  $T_s = MT$ . The choice  $m = M$  coincides with Nyquist sampling for which an efficient filter bank reconstruction scheme was proposed in [44]. To reduce the rate,  $m < M$  is used. The shifts  $\phi_i \in [0, T_s]$  are chosen to be different from each other; see Fig. 2(a). PNS was used in [3], [4] for known carrier frequencies  $f_i$  where the sampling rate may approach the minimal value of  $NB$  as derived by Landau [45]. A multicoset version, in which  $\phi_i$  can take on only the discrete values

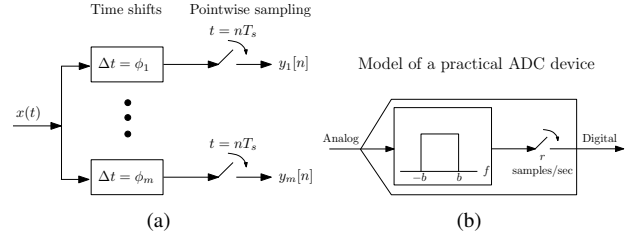


Fig. 2. Schematic implementation of PNS (a) requires no filtering between the time shifts and the actual sampling. However, the front-end of a practical ADC has an inherent bandwidth limitation, which is modeled in (b) as a lowpass preceding the uniform sampling.

$\phi_i \in \{kT | 0 \leq k \leq M - 1\}$ , was studied in [5], [6]. Multicoset has the advantage that the shifts  $\phi_i$  can be set regardless of the band locations, though in [5], [6] this information is required for the reconstruction stage. Spectrum-blind sampling and recovery, namely when  $f_i$  are unknown, was analyzed in detail in [7].

**PNS and Xampling.** In the blind setting, [7] considers the analog multiband model, and proves that the minimal rate for reconstruction in the case of unknown carriers is  $2NB$ . This work also proposes a reconstruction algorithm that theoretically requires no more than the minimal rate of  $2NB$  samples/sec. The results are immediately extended to PNS, that is to arbitrary time-shifts  $\phi_i \in [0, T_s]$ . Therefore, PNS with spectrum-blind recovery satisfies **(X1)**, **(X2)**.

In terms of analog implementation, both the time shifts and the ideal samplers are difficult, and may be even impossible, to realize [11]. To understand these issues, we concentrate first on the ideal sampler and ignore the time shifts for the moment. Each PNS sequence (7) requires sampling a wideband input, whose spectral contents reach  $f_{\text{max}}$ , at the low rate  $1/T_s$ . Existing ADCs devices, however, limit the input bandwidth far below  $f_{\text{max}}$ . To account for the bandwidth limitation, [11] proposed the model that is depicted in Fig. 2(b), in which an ideal sampler of rate  $r$  samples/sec is preceded by a lowpass filter with cutoff  $b$ . When adding the missing filter to the PNS scheme of Fig. 7(a), we conclude that  $b = f_{\text{max}}$  is required in order to obtain the sequences (7) without distortion. Unfortunately, the achievable front-end of existing ADC devices [32] has bandwidth which is still far behind the possible frequency ranges of communication transmissions. Consequently, pointwise sampling of a wideband input, even at a low rate, cannot be implemented by standard ADCs.

Another limitation of PNS is maintaining accurate time-shifts  $\phi_i$ , which may be difficult to implement in high-speed systems [44], [46], [47]. Evidently, time-shifts at the RF resolution  $f_{\text{NYQ}}$  are another kind of bandwidth limitation. These hardware issues stand against the requirement for efficient implementation **(X3)**. In addition, spectrum-blind reconstruction from PNS sequences [7], [48] requires interpolating the samples to the RF rate before reconstruction or any kind of processing can occur. Thus, **(X4)** is also not satisfied.

#### B. The Nyquist-Folding System

The Nyquist-folding system [8] is presented in Fig. 3. In this method, the zero crossings of a time-varying sine

waveform  $\sin(\phi(t))$  define a set of time instances  $t_i$ , where  $\phi(t) = \omega_1 t + \theta(t)$  is monotone increasing in time. The signal  $x(t)$  is multiplied by a pulse train  $p(t - t_i)$  generated by all zero crossings  $t_i$ . The product is filtered using an interpolation kernel  $F(\omega)$  and then sampled at an angular frequency  $\omega_2$ . The relation between the samples and the analog input was approximated in [8] for a single FM transmission,  $x(t) = \cos(\omega_c t + \psi(t))$  with angular carrier  $\omega_c$  and narrowband information contents  $\psi(t)$ . The pulse shape  $p(t)$  is required to satisfy mild smoothness properties for the approximation to hold. In addition,  $\omega_1 \gg \max |\theta'(t)|$  is needed. Since  $\omega_2 \ll \omega_c$  the sampling rate is reduced below Nyquist.

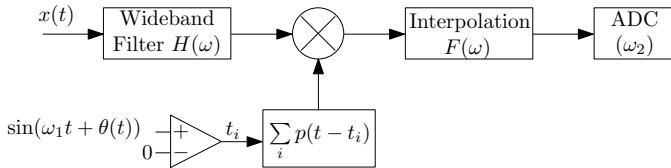


Fig. 3. Block diagram of the Nyquist-folding system.

**The Nyquist-folding and Xampling.** The experiments in [8] report the folding effect for pure sinusoidal signals, that is when the information  $\psi(t) = 0$ . In this case the width  $B \rightarrow 0$ , and there is no overlap between the sinusoids in baseband. However, in practice, transmissions have  $B > 0$  and the baseband contains aliasing from the entire spectrum, in which case the folding effect is more complicated even for a small set of FM signals. It appears that the typical setting of  $B > 0$ , when the aliases at baseband are likely to overlap, was not studied yet by [8]. Therefore, at this stage, it is not clear whether **(X1)** is satisfied.

Mixing by the pulse train  $p(t - t_i)$  is tantamount to pointwise sampling of  $x(t)$  according to the zero crossings of  $\sin(\phi(t))$ , and then filtering the samples with the pulse shape  $p(t)$ . Consequently, the bandwidth limitations of standard ADCs apply here as well. Aligning the pulses  $p(t - t_i)$  to the asynchronous time instances  $t_i$  may be as difficult to implement as realizing the time-shifts of PNS. In addition, the pulses  $p(t - t_i)$  may need to overlap in time, since otherwise  $x(t)$  is multiplied by zero between consecutive pulses. On the other hand, when the pulses overlap the summation requires generating several pulses in parallel, which may require additional hardware. In addition, the results of [8] do not provide a reconstruction algorithm. It is not clear how the analysis of the folding effect for a single FM signal  $x(t)$  extends to a concrete recovery algorithm for an arbitrary number of transmissions, with other modulation techniques, so that **(X3)** is currently not satisfied. Baseband processing **(X4)** is also not discussed.

So far, we examined systems which are designed based on sampling theory principles, that is pointwise sampling (up to pulse shaping in [8]) and analysis in the frequency-domain. The works [3]–[6], [8] do not incorporate CS ideas, such as sparse representations from underdetermined systems. As it appears, when CS is missing the Xampling criteria are not satisfied. An exception is our earlier work [7], which does incorporate both CS and sampling. However, since it relies on PNS, it suffers from the same practical limitations.

## IV. THE RANDOM DEMODULATOR

We proceed to survey two methods that are based on CS: the RD [9] (and its parallel version [10]) in the current section, and the MWC [11] in Section V. To investigate these systems, we propose to view their *equivalent system*, namely the one which acquires the input at the high Nyquist rate, and then invokes digital operations to yield the original sub-Nyquist samples. The equivalent system viewpoint is an approach which was not considered explicitly in [9]–[11], but turns out to reveal various practical issues.

### A. System overview

Fig. 4(a) presents the random demodulator of [9]. The input signal  $f(t)$  is mixed by a pseudorandom sign waveform 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  $N_R$ . For brevity, we study the single channel RD, but address the extension to a bank of RD channels [10] where relevant.

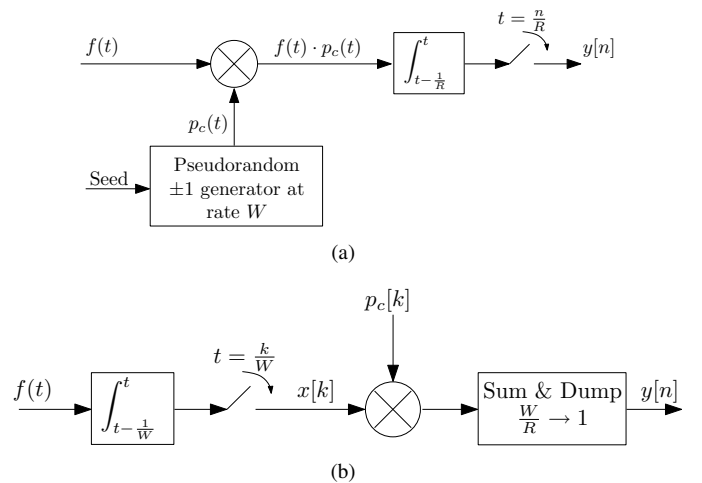


Fig. 4. Block diagram of the random demodulator (a), and the equivalent system (b) for an integer ratio  $W/R$ .

The authors [9] describe the analog RD system of Fig. 4(a), and then analyze a discrete CS system, eq. (13) below, which requires the equivalence to Fig. 4(b) and certain time-domain assumptions. The equivalent system integrates and dumps the input at rate  $W$ , producing the sequence  $x[k]$ . Then, the samples  $x[k]$  are digitally multiplied by a discrete pseudorandom sign stream  $p_c[k]$ . It can be easily verified from the figure that if  $R$  divides  $W$ , then every measurement  $y[n]$  corresponds to the sum over  $W/R$  consecutive products  $x[k]p_c[k]$ .

To connect the input signal  $f(t)$  to  $x[k]$ , a multitone model is assumed:

$$f(t) = \sum_{\omega \in \Omega} a_{\omega} e^{j2\pi\omega t}, \quad (8)$$

where  $\Omega$  is a finite set of  $K$  out of an even number  $Q$  of possible harmonics

$$\Omega \subset \{0, \pm\Delta, \pm2\Delta, \dots, \pm(0.5Q - 1)\Delta, 0.5Q\Delta\}. \quad (9)$$

We denote by  $\mathcal{N}$  the set of signals obeying this model, whose parameters are  $K$ ,  $Q$  and the tone spacing  $\Delta$ . For multitone signals, the rate  $W$  sequence  $x[k]$  is given by

$$x[k] = \int_{(k-1)W}^{k/W} f(t) dt \quad (10)$$

$$= \sum_{\omega \in \Omega} \underbrace{a_\omega \left[ \frac{e^{-j2\pi\omega/W} - 1}{j2\pi\omega} \right]}_{s_\omega} e^{-j2\pi\omega(k-1)/W}. \quad (11)$$

Under an additional assumption that

$$W = Q\Delta, \quad (12)$$

time-domain analysis in [9] shows that the vector  $\mathbf{x} = [x[1], \dots, x[Q]]^T$  is the discrete Fourier transform (DFT) of the coefficients  $s_\omega$ , which relates to the unknowns  $a_\omega$  by (11). The input-output relation can then be expressed as

$$\mathbf{y} = \Phi \mathbf{s} = \Phi_{\mathbf{H}} \Phi_{\mathbf{D}} \Phi_{\mathbf{F}} \mathbf{s}, \quad (13)$$

where  $\mathbf{y} = [y[1], \dots, y[N_R]]^T$  is the vector of samples and the unknown vector  $\mathbf{s}$ , which collects the coefficients  $s_\omega$ , is  $K$ -sparse due to the model assumptions. Therefore, the sampling matrix  $\Phi$  has dimensions  $N_R \times Q$ . It consists of a  $Q$ -square reordered DFT matrix  $\Phi_{\mathbf{F}}$ , which accounts for the relation (11), and a diagonal sign matrix  $\Phi_{\mathbf{D}}$  that performs the multiplication by  $p_c[k]$ . The sum-and-dump is represented by  $\Phi_{\mathbf{H}}$ , having  $N_R$  rows and  $N_R W/R$  columns. For the dimensions to match, (13) holds only if

$$Q = N_R \frac{W}{R} = N_R \frac{Q\Delta}{R} \rightarrow \Delta = \frac{R}{N_R}. \quad (14)$$

The normalization  $\Delta = 1$  Hz is used in [9]. For this choice,  $W = Q$  and  $N_R = R$ , which means observing  $f(t)$  over the time interval  $0 \leq t < 1$ .

Once the parameters are set properly, the system (13) is an underdetermined system whose sparsest solution  $\mathbf{s}$  defines the original multitone signal  $f(t)$ . Recovering  $\mathbf{s}$  from the linear system (13) is the fundamental problem studied in the CS literature, see eq. (4). Although the sparse recovery problem is NP-hard in general, there are many sub-optimal algorithms which yield the sparsest  $\mathbf{s}$  at the expense of a slight increase in the number of measurements  $N_R$  [14], [15]. To guarantee recovery, the sampling rate should be on the order of [9]

$$R \approx 1.7K \log(W/K + 1). \quad (15)$$

### B. The RD and Sampling

**Signal model.** The multitone model (8) is discrete by nature;  $f(t)$  is defined by  $2K$  parameters, the active tones and their amplitudes. The highest harmonic  $Q$  Hz (or the rate  $W$  under  $\Delta = 1$ ) conceptually stands for the Nyquist rate. However, there is no meaning to a sampling rate when a finite number of parameters define the signal  $f(t)$ . Indeed, the Landau theorem [45] for multitone signals implies that the minimal required density for a stable sampling set is of zero-measure – a direct consequence of the finite model. The essential point is that multitones are parameterized by a finite set of numbers to begin with. In contrast, the Fourier transform

of analog signals, such as the multiband model, consists of an uncountable number of harmonics; analog signals necessitate strictly positive sampling rates [45].

It was noted in [9] that a multiband analog signal with  $N$  bands of width  $B$  each can be approximated by a multitone signal with about  $K = NB$  tones in (8). More accurately, the representation yields a reasonable error only if convolving the input  $x(t)$  by a window function, prior to the RD system. However, it is not stated which window to choose, and how to recover the analog input  $x(t)$  from the windowed version, if at all possible. Therefore, the finite parametrization of signals  $f(t) \in \mathcal{N}$  may adequately capture only a limited type of signals – those that contain finitely many unknowns. In contrast, a broad analog signal model, as implied by **(X1)**, captures signals with infinitely many unknowns. As we show later, the issue of approximation with  $K = NB$  results in a large computational load on the digital reconstruction algorithm.

**Sampling rate.** The required rate, eq. (15), predicts the number of samples per second that are required to determine  $f(t)$  with time-varying support. Here  $W = Q$  counts the number of degrees of freedom per time interval. The RD system, however, does not allow to approach (15) in general. For example: consider  $f(t)$  with  $Q = 10$  GHz being the highest harmonic, and  $K$  tones such that (15) evaluates to  $R \approx 1.5$  GHz. The analysis of [9] relies on the equivalence between the systems of Fig. 4, which in turn requires that  $R$  divides  $W$ . In the example,  $R = 1.25$  GHz and  $R = 2.5$  GHz are the closest allowed rates, and the system is forced to sample 66% faster than the minimal rate. Therefore, in practice,  $R$  should be set by

$$R = W/r \geq 1.7K \log(W/K + 1) \quad (16)$$

for the smallest possible integer  $r$ . The rate gap in (16) is positive in general, and may be quite large as in the example, thus **(X2)** is not satisfied.

**Implementation.** Before studying the implementation feasibility of Fig. 4 and the complexity of the recovery algorithm, we first point out an inherent sensitivity which is independent of the actual chosen devices. The rates  $R, W$  in the RD are presumably triggered by some clock signals. In practice, however, a clock circuitry may vary its basis frequency with voltage, temperature, humidity, aging and other factors. Recall that (12) and (14) require  $R, W$  to satisfy a strict relation with the tone spacing of the input. Consequently, for the RD to work, one has to manually match the rates  $R, W$  for the tone spacing  $\Delta$  of the signal  $f(t)$  at hand. Such delicate tunings require additional analog hardware which is not described in [9]. The normalization  $\Delta = 1$  Hz and  $R = N_R$ , which greatly simplifies the presentation in [9], implicitly assume the existence of such a hardware mechanism.

The authors propose to treat any mismodeling from  $\mathcal{N}$  as an additive noise. However, the following toy-example demonstrates that the strict system-signal dependency translates to a large recovery error. Let  $W = 1$  kHz,  $N_R = R = 100$  Hz and consider the signal

$$f(t) = 3 \cos(2\pi 120t) + 4 \cos(2\pi 350t), \quad t \in [0, 1). \quad (17)$$

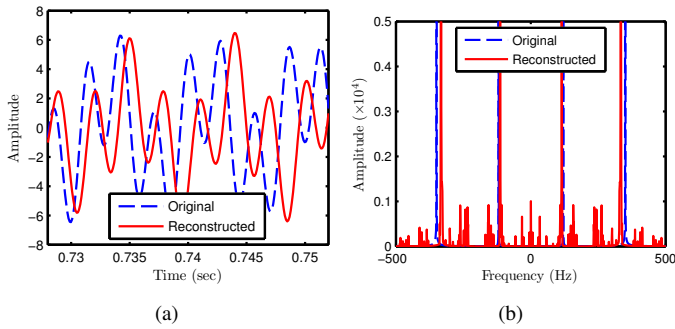


Fig. 5. Recovery of a multitone signal from random demodulator samples under design imperfections. The original and reconstructed signal are plotted in (a) on a short time interval. The frequency transforms (b) reveal many spurious tones due to the clock shift.

Applying the random demodulator to  $f(t)$  gives  $N_R = R = 100$  samples. The  $\ell_1$  minimization (5) with  $\epsilon = 0$  reconstructs  $\hat{f}(t) = f(t)$  exactly. Now, suppose that the clock network suffers from 0.5% inaccuracy, so that  $R' = 100.5$  Hz and  $W' = 1.005$  kHz. In this case, when allowing some error  $\epsilon = 0.5\% \|y\|$  in (5) we obtain the reconstructed  $\hat{f}(t)$  of Fig. 5(a), with normalized squared-error  $\|f - \hat{f}\|^2 / \|f\|^2 = 1.7$ . The frequency contents are compared in Fig. 5(b). Similar results are encountered if  $R, W$  are unchanged but the input obeys tone spacing  $\Delta$  with 0.5% deviation from the system spacing  $R/N_R$ . The scenario above considers changes in  $R, W$  by the same proportion and direction, such that  $R/N_R = W/Q$  and (12) and (14) are partially satisfied. In practice,  $R, W$  may change independently and manual tunings of their frequencies should be carried out separately.

We now proceed to study the actual implementation of Fig. 4, and for that purpose  $\Delta = R/N_R = W/Q = 1$  is assumed hereafter. The RD does not suffer from either of the practical limitations of PNS. There are no time delays, and the sampling occurs at rate  $R$  where the integrator acts as a first-order lowpass filter to mitigate bandwidth limitations. An accurate analog integrate-and-dump, however, may be difficult to implement. An ideal integrator is realized by a single capacitor. In practice, a capacitor has an effective series resistivity and a parallel inductivity, which cannot be avoided. Consequently, the exact correspondence between the systems in Fig. 4 is breached. As noted in [9], when the integration is nonideal,  $\Phi_H$  becomes signal-dependent. In other words, it requires another tuning mechanism to calibrate  $\Phi_H$  every second, in parallel to signal acquisition. Another source for breaching the equivalence is the pseudorandom waveform  $p_c(t)$ . If the alternations do not occur exactly on the Nyquist grid  $n/W$ , then  $\Phi_H$  becomes again signal-dependent and another synchronization is required.

Moving on to the recovery steps in the digital domain, we recall that  $K = NB$  tones are required to represent an analog signal within  $\mathcal{N}$  (ignoring the pre-windowing issue). We now examine the recovery complexity under this approximation. Consider a wideband scenario with  $N = 6$  bands of width  $B = 50$  MHz and  $f_{\text{NYQ}} = 10$  GHz, which boils down to  $K = NB = 300 \cdot 10^6$  tones. In this setting  $W = 10^{10}$  and

$\Phi$  has about  $R = N_R = 2.6 \cdot 10^9$  rows, resulting in a huge-scale CS system. The system is idle until  $N_R = 2.6 \cdot 10^9$  samples are collected, that is 1 second for  $\Delta = 1$  Hz. Memory storage of that volume is against efficient analog implementation (X3). Solving a CS system with huge-scale dimensions, namely  $\Phi$  of size  $2.6 \cdot 10^9 \times 10^{10}$ , is the other side of the coin. It imposes severe computational loads even when using polynomial-time methods and exploiting fast matrix-vector multiplications, contradicting again (X3).

Note that in huge-scale CS matrices, the finite-precision of the matrix entries leads to undesired high correlations between the columns of  $\Phi$ . The rate requirement (15) and the properties of  $\Phi$  were studied in [9] assuming ideal infinite-precision setting. We noticed in simulations that the finite precision dramatically degrades the performance of CS-algorithms even when the sensing matrix is of moderate dimensions. Overall, the RD system contradicts (X3) in many aspects. The bank of RD channels [10] duplicates the analog issues and the computational complexity is not improved by much.

**Baseband processing.** Solving the system (13) aims at recovering the sparsest vector  $\mathbf{s}$ , which lives in the ambient space  $\mathbb{C}^Q$ . Under the convention that  $\Delta = R/N_R = 1$ ,  $W = Q$ , the length of  $\mathbf{s}$  is proportional to the Nyquist rate. The recovery yields at once both the tone locations and their amplitudes, however the output  $\mathbf{s}$  is already at the Nyquist rate. Therefore, (X4), which requires the ability to process any band of interest without involving Nyquist rate computations, cannot be satisfied. Besides, there is no clear connection between  $\mathbf{s}$  and the information  $I(t), Q(t)$  of a band of interest.

## V. THE MODULATED WIDEBAND CONVERTER

In this section, we examine the MWC system, which overcomes the previous limitations by sticking to the traditional frequency-domain analysis, and in the same time employing CS algorithms where beneficial. This balanced combination of CS and sampling is shown to satisfy (X1)-(X3). Unfortunately, there is one fly in the ointment; the MWC generates baseband sequences that are incompatible with the required formats  $I(t), Q(t)$  for standard DSP packages. This limitation is solved in Section VII.

### A. System overview

The MWC consists of an analog front-end with  $m$  channels. In the  $i$ th channel, the input signal  $x(t)$  is multiplied by a periodic waveform  $p_i(t)$ , lowpass filtered, and then sampled at rate  $1/T$ . In this paper, we study a simplified version of the converter, as depicted in Fig. 6, in which the sampling interval  $T$  equals the period of the waveforms  $p_i(t)$ . This basic configuration has three parameters: number of channels  $m$ , periodic waveforms  $p_i(t)$  and sampling rate  $1/T$ . This scheme is sufficient for studying the applicability of the system; other configurations with practical advantages are detailed in [11].

The MWC sensing relies on the following key observation. The mixing operation scrambles the spectrum of  $x(t)$  such that the baseband frequencies that reside below the filter cutoff  $1/2T$ , contain a mixture of the spectral contents from the entire Nyquist range. The periodicity of each waveform  $p_i(t)$



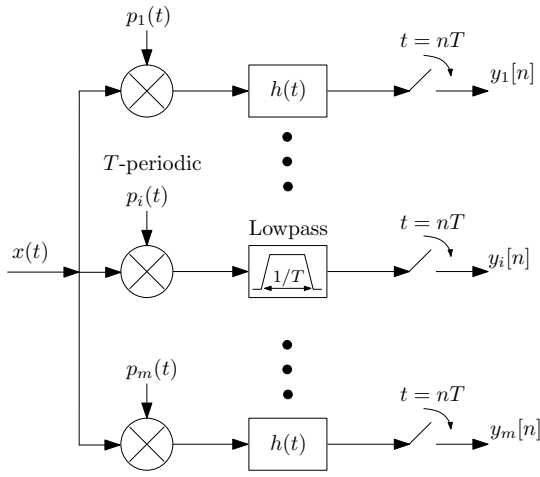


Fig. 6. Block diagram of the modulated wideband converter.

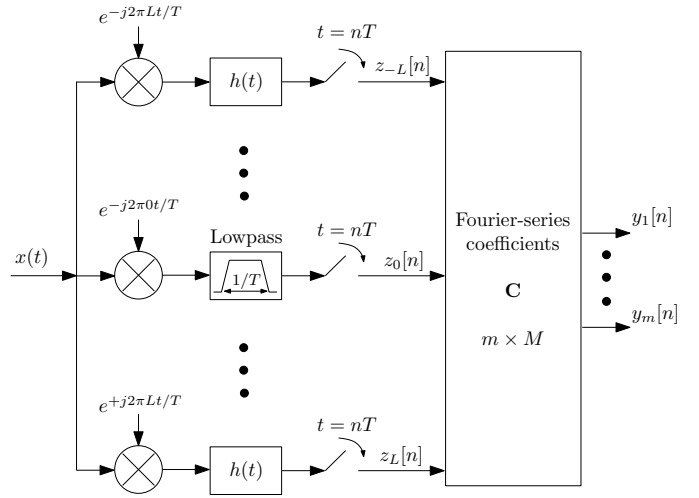


Fig. 7. The equivalent system of the modulated wideband converter.

ensures that the mixture has a specific nature – aliases at  $1/T$  frequency spacing. Whilst aliasing is often considered as an undesired effect, here it is deliberately utilized to shift various frequency regions to baseband, simultaneously. In the basic configuration, we choose the rate  $1/T \geq B$ .

To understand the act of the MWC on multiband signals, we consider the equivalent system that is depicted in Fig. 7. The signal  $x(t)$  enters  $M = 2L + 1$  channels, where  $L$  is the smallest integer such that  $M \geq T f_{\text{NYQ}}$ . Since  $1/T \approx B$ ,  $M$  represents the compression ratio, that is the quotient of the Nyquist rate  $f_{\text{NYQ}}$  by the rate  $1/T$  of a single channel. In the  $l$ th channel,  $x(t)$  is frequency-shifted (hence modulated in time) by  $l/T$  Hz,  $-L \leq l \leq L$ . Then, the baseband region  $[-1/2T, 1/2T]$  is filtered and sampled every  $T$  seconds, giving the (complex-valued) sequence  $z_l[n]$ ; see the illustration in Fig. 8. Clearly, if  $x(t)$  contains only a few transmissions, then most of the time sequences  $z_l[n]$  will be identically zero. This is where CS comes into play. For each time-point, the vector  $\mathbf{z}[n] = [z_{-L}[n], \dots, z_L[n]]^T$  is compressed into the output vector  $\mathbf{y} = [y_1[n], \dots, y_m[n]]^T$  using the linear projection

$$\mathbf{y}[n] = \mathbf{C}\mathbf{z}[n]. \quad (18)$$

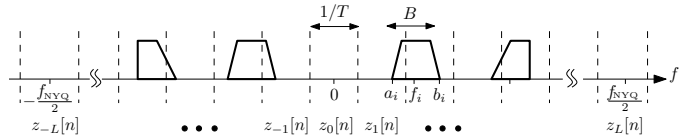
The equivalence to Fig. 6 holds due to the periodicity of the waveforms  $p_i(t)$ . Since  $p_i(t) = p_i(t + T)$  for all  $t \in \mathbb{R}$ , we have the Fourier expansion

$$p_i(t) = \sum_{l=-\infty}^{\infty} c_{il} e^{j\frac{2\pi}{T}lt}. \quad (19)$$

Choosing the matrix  $\mathbf{C}$  of Fig. 7 such that its  $il$ th entry is equal to the Fourier coefficient  $c_{il}$  results in the desired equivalence

$$y_i[n] \stackrel{\text{Fig. 6}}{=} (x(t)p_i(t)) \star h(t) \Big|_{t=nT} \\ \stackrel{\text{Fig. 7}}{=} \sum_{l=-L}^L c_{il} \left( (x(t)e^{-j2\pi lt/T}) \star h(t) \Big|_{t=nT} \right). \quad (20)$$

Conceptually, the MWC shifts the mixing matrix  $\mathbf{C} = \{c_{il}\}$  into the analog domain, such that each channel realizes a single row of  $\mathbf{C}$  in analog hardware.

Fig. 8. The lowrate sequences  $z_l[n]$  correspond to equal-width spectrum slices from the Fourier transform of the input multiband signal.

Note that, the system of Fig. 7 samples the signal at the high Nyquist rate  $f_{\text{NYQ}}$ , thus clearly does not satisfy **(X2)**. Moreover, most channels are likely to provide identically zero sequences, which is a waste of mixing resources; having  $M \approx f_{\text{NYQ}}/B$  oscillators, mixers and filters is hardware excessive. However, it nicely captures the intuition behind the MWC. The periodicity of  $p_i(t)$  is the only essential requirement for the equivalence to hold.

Recovery of  $x(t)$  from the sample sequences  $\mathbf{y}[n]$  starts from (18). Under appropriate conditions on  $m$  and  $\mathbf{C}$ , we have that  $m \ll M$  and that (18) determines a unique  $\mathbf{z}[n]$  for every  $n$  [11]. Theoretically, we can solve for the sparsest solution  $\mathbf{z}[n]$  of (18), for every  $n$ , and then reconstruct  $x(t)$  by properly re-positioning the nonzero sequences  $z_l[n]$ . However, this approach is inefficient, since the sparsest solution  $\mathbf{z}[n]$ , even if obtained by polynomial-time CS algorithms, is computed separately for every  $n$ . Instead, [7], [39] suggest a more efficient way which exploits the fact that the bands occupy continuous intervals in the spectrum. This fact implies that the vectors  $\mathbf{z}[n]$  for different time instances share a common nonzero location set [7]. Thus, instead of solving separately for every  $n$ , we construct a finite frame (or a basis)  $\mathbf{V}$  from a set of consecutive sample sets  $\mathbf{y}[n]$ . It is shown in [39] that any such frame has the same support as the joint sparsity of  $\mathbf{z}[n]$ . The Continuous-to-Finite block (CTF), which is depicted in Fig. 9, implements this principle. The joint support is inferred from a multiple measurement vector (MMV) CS system

$$\mathbf{V} = \mathbf{C}\mathbf{U}, \quad (21)$$

which extends standard vector sparsity to matrices with a few rows that are not identically zero. In the figure, the support  $S$  is determined by merging the supports of all the columns  $\bar{\mathbf{U}}_i$

of the sparsest matrix  $\bar{\mathbf{U}}$ . For the CTF to function, the required rate is increased to  $4NB$  [7], [39], which is still proportional to the actual bandwidths and not to  $f_{\text{NYQ}}$ . Once the support  $S$  is found, the pseudo-inverse  $\mathbf{C}_S^\dagger$  is computed and is then used to recover  $\mathbf{z}_S[n]$ . The notation  $\mathbf{C}_S$  means the column subset of  $\mathbf{C}$  indicated by  $S$ , and similarly  $\mathbf{z}_S[n]$  is the relevant vector subset. In contrast to the huge dimensions of the RD matrix  $\Phi$ , the MWC recovery involves the matrix  $\mathbf{C}$  of size  $m \times M$  which is typically small, and thus the digital computations, such as the sparse recovery of (21) or the inversion  $\mathbf{C}_S^\dagger$ , are not an issue.

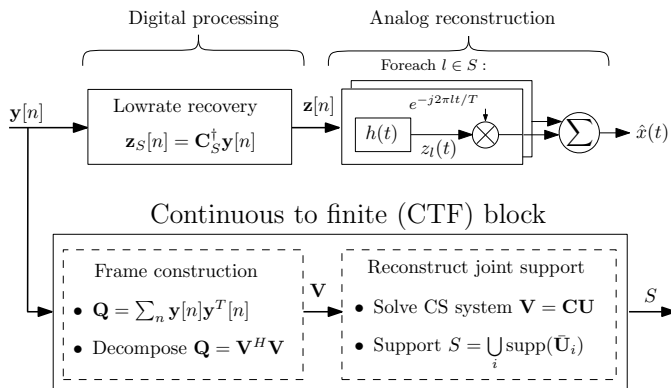


Fig. 9. The Continuous-to-finite (CTF) block recovers the spectral support from a set of consecutive vector samples  $\mathbf{y}[n]$ . Then, the nonzero lowrate sequences of  $\mathbf{z}[n]$  are generated at realtime, and the signal  $x(t)$  is reconstructed.

The choice of the waveforms  $p_i(t)$  directly dictates certain CS-related properties of  $\mathbf{C}$ . In [11], an MWC system with  $p_i(t)$  consisting of  $M$  sign alternations was analyzed and simulated, though the specific choice of the sign patterns was not studied. The second part of this work [16] proposes a theoretical framework for the required properties of the sensing matrix in sub-Nyquist systems. In that context, we complete the MWC design by proving that certain binary sequences, such as Maximal or Gold codes, are suitable choices for the MWC.

### B. The MWC and Xampling

**Signal model.** The analysis of [11] is based on the analog multiband model with  $N$  bands of width  $B$ , and the bands are allowed to reside anywhere below  $f_{\text{max}}$ . As in conventional sampling, the analog signal is converted into sequences of digital samples. No finite parametrization is used.

**Sampling rate.** The basic configuration has  $m$  channels, each sampling at rate  $1/T$ . Therefore, the rate can be adjusted in steps of  $1/T$ . The step size is independent of the Nyquist rate  $f_{\text{NYQ}}$  and the required rate  $4NB$  can be approached. For example, for  $N = 6$  and  $B = 50$  MHz, the sampling rate  $4NB = 600$  MHz can be achieved by  $m = 4N = 24$  channels. Additional options to control the rate using fewer analog channels are detailed in [11].

**Implementation.** We start with the analog implementation of Fig. 6. The parameter choice of the MWC is summarized by

$$1/T \geq B, \quad M \geq T f_{\text{NYQ}}. \quad (22)$$

In contrast to the RD, there is no need to synchronize or manually tune the system to the input signal, since (22) involves only inequalities. Setting  $T, M$  with small safeguards allows to apply the MWC to multiband signals, even if the actual width  $B$  or the frequency  $f_{\text{max}}$  are higher to some extent than what assumed in design.

The design has two additional flexibilities. The lowpass filter  $h(t)$  can be realized by standard analog methods. In practice, ripples and non-smooth transitions in the frequency response are compensated by a digital filter, whose coefficients are computed once after manufacturing by a signal-independent optimization program [49]. In addition, there is no need to maintain accurate sign alterations on the Nyquist grid as required for the RD. Only the periodicity of  $p_i(t)$  matters. Calibrating the Fourier coefficients  $c_{il}$  is performed de-facto. The calibration is also signal-independent and is performed once after manufacturing [23].

These implementation advantages were utilized in a board-level prototype of the MWC, which we report in [23]. The board consists of commercial devices only. An elliptic filter of order 7 realizes  $h(t)$  by only a few coils and capacitors. The waveforms  $p_i(t)$  are derived from a cyclic shift-register running at a 2.4 GHz clock rate. The periodicity of  $p_i(t)$  is guaranteed by no more than a voltage-controlled-oscillator and a standard phased-lock-loop synthesizer driven by an accurate crystal source. In the design process, we made no effort to improve the time-domain appearance of  $p_i(t)$ , and those are far from nice rectangular shapes. Verifying periodicity is carried out by observing  $p_i(t)$  in a spectrum analyzer, a customary equipment of RF engineers. The point we would like to emphasize here is that the frequency-domain approach is what renders the MWC immune to practical issues.

Moving on to recovery complexity, we first recall that the numerical simulations in [11] demonstrated accurate recovery for signals with  $N = 6, B = 50$  MHz and  $f_{\text{NYQ}} = 10$  GHz by an MWC system with  $m = 35$  channels and sign waveforms of  $M = 195$  alternations per period. The first task in the CTF constructs a frame for  $\mathbf{y}[n]$ , which involves computing

$$\mathbf{Q} = \sum_n \mathbf{y}[n] \mathbf{y}^T[n], \quad (23)$$

that is  $m^2$  multiplication and summations per input vector  $\mathbf{y}[n]$ . This may seem quite challenging, as theoretically it involves the samples  $\mathbf{y}[n]$  for  $-\infty < n < \infty$ . However, since  $\mathbf{y}[n]$  is a length  $m$  vector for every  $n$ ,  $\text{rank}(\mathbf{Q}) \leq m$ , which means that a set of  $m$  linearly independent vectors suffice. In fact due to the sparsity of  $\mathbf{z}[n]$  only  $K = 2N$  independent vectors are needed [11]. In the presence of noise, a slight rate increase is required to identify the signal space and reject the noise influence. In practice, the simulations in [11] demonstrated exact support recovery from noisy samples when the frame is constructed from only 40 time instances  $n$ , for  $K = 2N = 12$ . The memory overhead is minor. Theoretically, there are pathological examples [7] which may require a dramatically larger set of samples, however in practice these are unlikely to occur.

The decomposition

$$\mathbf{Q} = \mathbf{V} \mathbf{V}^H \quad (24)$$

is more expensive than (23) but for matrices of size  $35 \times 35$  does not represent great difficulty. In fact, (24) can be avoided since  $\mathbf{Q}$  is also a frame for  $\mathbf{y}[n]$ . In the MMV system (21) of the CTF, we can substitute the right-hand-side by  $\mathbf{Q}$ . This method was not suggested previously in [7], [11], [39]. The advantages of performing the decomposition (24) are two-fold:  $\mathbf{V}$  typically has less columns than  $\mathbf{Q}$  which reduces complexity, and in noisy settings, the decomposition can be used to reject the noise space, as done in the simulations of [7], [11]. Nonetheless, these advantages are negligible:  $m$  is anyway small and the noise space can be identified by other methods, e.g. by identifying the dominant nonzero rows of  $\mathbf{U}$  (in  $\ell_2$ -norm), or by observing the order a greedy-type CS algorithm constructs  $\mathbf{U}$ .

The last computation in the CTF block is solving the MMV system (21) for the sparsest matrix  $\bar{\mathbf{U}}$ . Since the dimensions  $m, M$  are relatively small (by orders of magnitude compared with the RD sensing matrix  $\Phi$ ), solving this CS system can be performed quite fast.

It is worth emphasizing that sub-Nyquist sampling is one of the appealing properties of the MWC, though it can also be used for conventional Nyquist sampling of wideband signals with the proper number of channels. In contrast, the RD cannot be used for Nyquist sampling, since the single ADC runs at the high-rate  $f_{\text{NYQ}}$ .

**Baseband processing.** The MWC does not utilize carrier knowledge prior to sampling and consequently the sequences  $z_i[n]$  do not relate directly to the information signals  $I(t), Q(t)$  of a band of interest. For example: in Fig. 8, the energy of the  $i$ th band splits between the two consecutive spectrum slices, and typically  $z_i[n]$  contains simultaneous contributions of several bands, as illustrated in Fig. 8. Even when  $z_i[n]$  contains a single band, the carrier frequency is an unknown parameter within a wide range of possible frequencies: for example, searching for the carrier in a spectrum width of  $1/T = B = 50$  MHz is still a demanding processing task, which is prone to errors. Therefore, the baseband processing capability (X4) is not satisfied. Section VII bridges this gap by estimating the band edges  $[a_i, b_i]$ , recovering the carrier frequencies  $f_i$  and providing a single lowrate sequence  $s_i[n]$  per band. The information contents  $I(t), Q(t)$  of any band of interest are then immediately obtained.

## VI. MID-TERM SUMMARY

Table II summarizes the systems we examined in light of the Xampling requirements. The Whittaker, Kotelnikov, and Shannon (WKS) theorem, namely uniform sampling at the Nyquist rate, is added for reference.

The Xampling criteria can be divided into pairs. Model (X1) and rate (X2) which quantify the theory underlying the approach, and implementation (X3) and baseband processing (X4) which capitalize on practical aspects. Table III summarizes in more detail the comparison between [7], [9], [11]. The numbers in the table refer to the wideband scenario that was considered throughout, namely  $N = 6, B = 50$  MHz and  $f_{\text{NYQ}} = 10$  GHz. The rate gap is zero for methods that can achieve their minimal rate requirement. We compare a

TABLE II  
SUB-NYQUIST DESIGNS AND XAMPLING

	Model	Rate	Implementation		Processing	
			Analog	Digital	BB	RT
WKS-theorem	+	-	+/-	-	-	-
PNS [7]	+	+	-	-	-	-
Nyquist-folding [8]	-	+	-	?	?	?
RD [9], [10]	-	-	-	-	-	-
MWC [11]	+	+	+	+	-	+

BB=Baseband, RT=Realtime.

few properties regarding the sensing matrix. The dimensions row highlights 9 orders of magnitude difference between the MWC and the RD. During the execution of CS algorithms, the sensing matrix is often repeatedly applied on the estimated sparse signal. The complexity of this operation, which can utilize the structure of the matrix, is denoted in the next row. In addition, the storage requirements of the sensing matrix are also compared. Note that for the MWC, we stated that the system comprises of  $m \geq 4N$  channels [11], though advanced configurations of the MWC remove this requirement. In the prototype [23], only 4 channels were successfully used for  $N = 6$  bands.

In terms of realtime processing, PNS reconstruction necessitates digital interpolation to the high Nyquist rate, and thus we indicated “huge” in the last rows of the table. For the RD and the MWC, the specific CS-algorithm in use determines the delay. The size of the sensing matrix has a major effect again, but quantifying this delay is beyond the current scope. To simplify, in the table, we consider the scenario of known support – tone locations  $\Lambda$  or support set  $S$ , respectively. For the RD, the nonzeros of the discrete vector  $\mathbf{s}$  are recovered by

$$\mathbf{s}_\Lambda = \Phi_\Lambda^\dagger \mathbf{y}. \quad (25)$$

Since (25) requires the entire vector  $\mathbf{y}$ , the memory requirements remain the same. Applying the matrix-vector multiplication in (25) involves  $KN_R = 780 \cdot 10^9$  million instructions per second (MIPS), a severe computational cost. In contrast, the computations in the MWC suit a realtime environment – the memory (of about  $2N$  values of  $\mathbf{y}[n]$ ) introduces a short delay of 780 nanoseconds. The matrix-vector multiplication in Fig. 9 translates to only  $2Nm = 420$  multiplication and summations per input vector  $\mathbf{y}[n]$ . Extrapolating to a period of a second, gives  $2Nm/T = 22 \cdot 10^3$  MIPS. The memory size, the delay and the MIPS are all at least 6 orders of magnitude lower than [9].

The technology barrier of each approach is highlighted at the bottom of Table III. The front-end of a practical ADC limits the applicability of multicaset or PNS strategies as explained in Section III. 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 processors 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

TABLE III  
METHODS FOR WIDEBAND SPECTRUM-BLIND RECOVERY.

	WKS theorem $y[n] = x(nT_s)$	PNS Eq. (7)	Random demodulator Fig. 4	Modulated wideband converter Fig. 6	
Theory	Model	$\beta$ -Bandlimited	Multiband	Multiband	
	Type	Continuous	Continuous	Continuous	
	Model parameters	$\beta$	$N, B, f_{\text{NYQ}}$	$K, Q, \Delta$	$N, B, f_{\text{NYQ}}$
	Sampling parameters	$T_s$	$m, M, \phi_i$	$R, W, N_R, p_c(t)$	$m, M, T, p_i(t)$
	Setup	$1/T_s > 2\beta$	see [7]	$\Delta = R/N_R$	$1/T \geq B, M \geq T f_{\text{NYQ}}$
	Sensitivity (mismatching)	Low	Low	High	Low
	Sensitivity (inaccuracy)	Low	High	High	Low
	Minimal rate gap	0	0	(16)	0
Analysis domain	Frequency	Frequency	Time	Frequency	
Practice	Number of channels	1	$m \geq 4N \approx 35$	1	$m \geq 4N \approx 35$ <sup>§</sup>
	Devices (per channel)	ADC	ADC Time-shifter	mixer, sign waveform integrator	mixer, periodic waveform, filter
	Sensing matrix (dimensions)	None	$m \times M$ $35 \times 195$	$R \times W$ $2.6 \cdot 10^9 \times 10^{10}$	$m \times M$ $35 \times 195$
	Sensing matrix (apply)	None	$\mathcal{O}(M \log M)$	$\mathcal{O}(W \log W)$	$\mathcal{O}(mM + M \log M)$
	Sensing matrix (storage)	None	$\mathcal{O}(m)$	$\mathcal{O}(W)$	$\mathcal{O}(mM)$
	Support recovery (mem. size, time)		$\geq 2N \approx 40, 1\mu\text{sec}$	$N_R = 2.6 \cdot 10^9, 1 \text{ sec}$	$\geq 2N \approx 40, 1\mu\text{sec}$
	Realtime	Memory size Delay MIPS	huge	$N_R$	None
			huge	$N_R/R = 1 \text{ sec}$	real time
			huge	$KN_R = 780 \cdot 10^9$	$2Nm/T = 22 \cdot 10^3$
	Technology barrier	ADC's front-end bandwidth ( $\sim 1 \text{ GHz}$ )	CS algorithms <sup>‡</sup> ( $\sim 10 \text{ MHz}$ )	Waveform generator ( $\sim 23 \text{ GHz}$ )	

<sup>§</sup> see text    <sup>‡</sup> Our estimate

seems to be preferred in this setting. The MWC is limited by the technology for generating the periodic waveforms  $p_i(t)$  [11], which depends on the specific choice of waveform. The estimated barrier of 23 GHz refers to periodic sign waveforms [50], [51].

## VII. BASEBAND PROCESSING WITH THE MWC

The MWC has many advantages for standard analog applications in the wideband regime, as summarized in Tables II and III. However, baseband processing – the prominent reason for shifting to the digital domain – is missing. In this section, we propose a three-step algorithm which overcomes this deficiency. We assume the multiband model with  $N$  bands, such that each of the  $N/2$  transmissions is of the standard quadrature form (1). The MWC with the baseband processing capability we propose here results in a powerful system, which is capable of sampling, processing the information content and recovering the input, and all operations involve only lowrate computations. Even when  $f_{\text{NYQ}}$  is small, such that a Nyquist ADC is available, the baseband and realtime processing of the MWC becomes a significant advantage which may reduce the power-consumption and cost of the DSP device.

Fig. 10 depicts the three main steps of our algorithm:

- 1) refinement of the support estimate  $S$  to the actual band edges  $[a_i, b_i]$ . Here, we rely on two additional model parameters: the minimal width of a single band  $B_{\min}$  and the smallest spacing between bands  $\Delta_{\min}$ . These quantities are often implied by the application specification, though uncertainty in the values  $B_{\min}, \Delta_{\min}$  has little effect on the performance, as described later on;

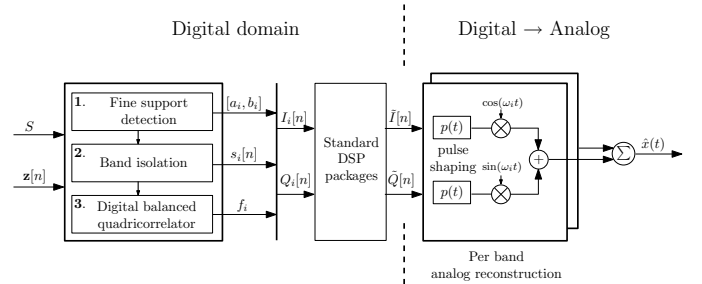


Fig. 10. Digital algorithm for baseband processing with the MWC.

- 2) generating a lowrate sequence  $s_i[n]$  per band  $1 \leq i \leq N/2$ . This step processes  $z_i[n]$  and incorporates the edges  $[a_i, b_i]$ ;
- 3) a digital version of the balanced quadricorrelator [24], an accurate carrier detector, is used to estimate  $f_i$ .

The desired information signals  $I(t), Q(t)$  are obtained upon completion at no additional cost. DSP software can alter the information as desired and generate new baseband information signals  $\tilde{I}[n], \tilde{Q}[n]$ .

As a nice feature, using the proposed algorithm, the conversion to analog of Fig. 9 is replaced by a more efficient method. In [11],  $x(t)$  is reconstructed directly from  $z_i[n]$  by interpolation to  $z_i(t)$  and properly positioning of the spectrum slices. Since the scenario of band splitting is common, it can be verified that this procedure requires  $2N$  mixers and filters at the most. The present approach requires only  $N$  mixers and filters. The reconstruction reduces to standard modulation of the narrowband bit stream  $I[n], Q[n]$  (or the analog message).

(1.1) Complex to real (1.2) PSD+Threshold (1.3) Edge detection

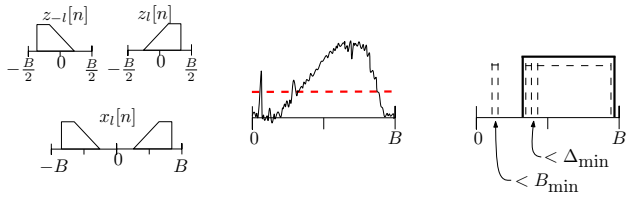


Fig. 11. Fine support (band edges) detection (Step 1).

In the sequel, we mention the relevant MATLAB commands (in verbatim style) that are used in our implementation. Numerical simulations are used for demonstration. The algorithm does not assume any specific modulation technique; the only essential assumption is the quadrature form (1). As a consequence, the carrier-frequency-offset (CFO) is not expected to be zero. In commercial receivers, such as Fig. 1, perfect lock on the input carrier is achieved only when employing modulation-specific and data-aided algorithms. The purpose of our algorithm is not to reach zero CFO, but rather to reach a CFO within the specifications of commercial standards. The MWC without the proposed algorithm has CFO intermediacy of  $1/T = B$ , e.g. 50 MHz in the examples we considered so far, which is far above standard CFO specifications.

### A. Algorithm description

**Step 1.** For convenience, we start with converting the complex-valued  $z_l[n]$  to real-valued counterparts. Recall that the input  $x(t)$  is real-valued with a conjugate symmetric Fourier transform. Therefore,  $l \in S$  implies  $-l \in S$  and  $z_{-l}[n] = z_l^*[n]$ . In step (1.1) of Fig. 11, a real-valued sequence  $x_l[n]$  at rate  $2B$  for each  $l \in S, l > 0$  is obtained by re-positioning  $z_l[n], z_{-l}[n]$  on both sides of the origin. Mathematically,  $x_l[n] = I_{2,0.5B}\{z_{\pm l}[n]\}$ , where the operator

$$I_{r,F}\{z_{\pm l}[n]\} \triangleq (z_l[n] \uparrow r)e^{-j2\pi Fn} + (z_{-l}[n] \uparrow r)e^{j2\pi Fn}, \quad (26)$$

and  $\uparrow r$  denotes rate conversion by a factor of  $r$ , with the appropriate post-filtering. By abuse of notation, here and in the sequel the same index  $n$  is used before and after the rate conversion, where the context resolves the ambiguity. The case  $l = 0 \in S$ , has  $x_0[n] = z_0[n]$ . We used `interpft` to carry out the interpolations in (26). The information rate is not changed;  $z_l[n]$  is complex-valued at rate  $B$ , while  $x_l[n]$  is real-valued at rate  $2B$ .

Power spectral density (PSD) estimation of  $x_l[n]$  is invoked in (1.2) in order to locate the energy concentration within each spectrum slice. In our simulations, we used the Welch PSD estimation method [52], implemented by `pwelch`, which divides the input to overlapping sections with overlap ratio 50%, filters each section by a Hamming window, performs a discrete Fourier transform (DFT) on each section, and finally averages the results. The frequency resolution and the window size are determined by:

$$f_{\text{res}} = \min(B_{\min}, \Delta_{\min}), \quad W_{\text{size}} \geq \frac{2B}{f_{\text{res}}}. \quad (27)$$

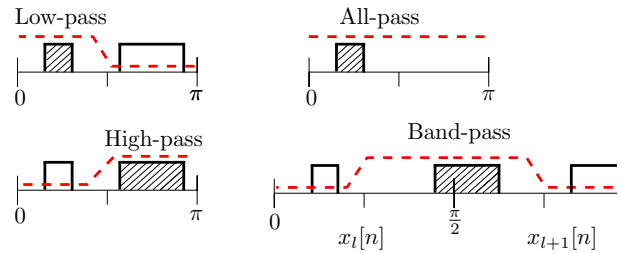


Fig. 12. Band isolation (Step 2). Merging occurred in the right-bottom drawing.

The PSD estimation produces  $P_{xx}^{(l)}[k]$  for  $1 \leq k \leq K \approx W_{\text{size}}/2$ , where the accuracy of the estimation increases with the number of samples in  $x_l[n]$ . The window size introduces an inherent trade-off, where a short one gives better averaging of additive noise, while a longer one allows higher DFT orders and thus improves the number of frequency bins  $K$ . In (27), the shortest possible window is used. A logarithmic threshold

$$\log_{10}(\text{Threshold}) = \frac{1}{K} \sum_{k=1}^K \log_{10} P_{xx}^{(l)}[k], \quad (28)$$

translates  $P_{xx}^{(l)}[k]$  to a binary decision on the energy concentration.

Finally, in step (1.3) we mitigate undesired noise effects that were encountered in simulations; support regions that are closer than  $\Delta_{\min}$  are united, and isolated regions with width smaller than  $B_{\min}$  are pruned. The operations (1.1)-(1.3) are carried out for each  $l \in S, l \geq 0$ . To conclude this step, only the  $N/2$  most powerful bands, according to the PSD values, are retained to further mitigate noise effects. The output of step 1 consists of  $N/2$  pairs  $[a_i, b_i]$  roughly indicating the start and the stop edges of the transmission bandwidths. To this end, the pairs are ordered such that  $a_i < b_i < a_{i+1}$ , and by convention  $a_0 = b_0 = 0, a_{(N/2)+1} = b_{(N/2)+1} = f_{\text{NYQ}}/2$ .

**Step 2.** The purpose of this step is to isolate a sequence  $s_i[n]$  for each  $1 \leq i \leq N/2$ , such that  $s_i[n]$  contains the entire contribution of exactly one band of information. Using the edges  $[a_i, b_i]$  we identify cases in which the information resides in adjacent spectrum slices  $x_l[n], x_{l+1}[n]$  for some  $0 \leq l \in S$ ; see Fig. 8 for example. In such cases, merging occurs via

$$\tilde{s}_i[n] = I_{4,0.5B}\{z_{\pm l}[n]\} + I_{4,B}\{z_{\pm(l+1)}[n]\}, \quad (29)$$

whereas  $\tilde{s}_i[n] = x_l[n]$  when both  $a_i, b_i$  lie within the same spectrum slice. As a result,  $\tilde{s}_i[n]$  contains the entire energy of the  $i$ th band with possible contributions from other bands. The information  $[a_i, b_i]$  from step 1 is utilized again to decide on the next actions.

Consider the  $i$ th band, and for brevity assume no merging step was required, so that  $[a_i, b_i] \subseteq [lB - B/2, lB + B/2]$  for some  $0 \leq l \in S$ . Let  $[\omega_{p,L}, \omega_{p,H}]$  be the normalized angular frequencies of  $x_l[n]$  corresponding to  $[a_i, b_i]$ , and set  $\omega_{s,L} = 0, \omega_{s,H} = \pi/2$ . If either  $b_{i-1}, a_{i+1}$  resides within the same spectrum slices  $z_l[n]$ , update the normalized angular frequencies  $\omega_{s,L}, \omega_{s,H}$ , respectively. Next, design a digital

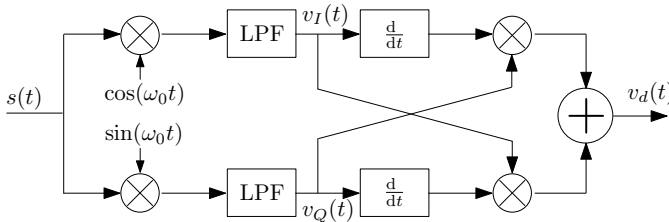


Fig. 13. The analog balanced-quadrice correlator.

filter  $D(e^{j\omega})$  satisfying

$$|D(e^{j\omega})| \leq \begin{cases} A_p & \omega_{p,L} \leq \omega \leq \omega_{p,H} \\ A_s & \omega \in [0, \omega_{s,L}] \text{ or } \omega \in [\omega_{s,H}, \pi] \end{cases}, \quad (30)$$

where  $A_p, A_s$  are the allowed ripples in the pass- and stopbands, respectively. As Fig. 12 shows, the resulting filter may be low-, high-, band- or all-pass, depending on the specific values of  $\omega_{s,L}, \omega_{s,H}$ . We used `firpmord` and `firpm` to determine the filter. The ripple amplitudes are set as  $A_p = 10^{-6}, A_s = 10^{-2}$ . The filter order is often small, since the actual spacing between the bands relaxes the cutoff constraints. At last,  $s_i[n]$  is obtained by filtering  $\tilde{s}_i[n]$  with the designed finite impulse response (FIR). A similar procedure is performed when merging occurred in  $\tilde{s}_i[n]$ .

At this point, we have a sequence  $s_i[n]$  for each band  $1 \leq i \leq N/2$  at a uniform rate of either  $2B$  or  $4B$ , depending on whether the merging (29) was required. In either case, the middle frequency  $f_i = (a_i + b_i)/2$  can serve as a rough estimate of the unknown carrier of (1). In fact, in simulations we observed that  $f_i$  is not far from the true carrier, as long as the PSD estimation is sufficiently accurate, a situation which occurs for a high signal to noise ratio (SNR) and many samples from  $x_l[n]$ . In such setting, the number of PSD points  $K = W_{\text{size}}/2$  is large, which allows to average out the noise and also to mitigate undesired effects due to the windowing. The next step incorporates an accurate carrier frequency detector which relies on (1) to better predict the carriers even in cases in which the PSD curve is noisy and inaccurate.

**Step 3.** We start with describing the balanced quadrice correlator (BQ), which was analyzed in [24] and whose circuit appears in Fig. 13. The BQ receives an input  $s(t)$  of the form (1), assumes a certain carrier frequency  $f_0 = \omega_0/2\pi$ , and outputs  $v_d(t)$  whose expected value is proportional to the carrier offset

$$\mathbb{E}[v_d(t)] = -K_G(f_c - f_0)(\mathbb{E}[I^2(t)] + \mathbb{E}[Q^2(t)]). \quad (31)$$

The signals  $I(t), Q(t)$  that build  $s(t)$  are assumed random with zero cross-correlation,  $\mathbb{E}[I(t_1)Q(t_2)] = 0$  for all  $t_1, t_2$ . In practice, time averaging replaces the expectation. The constant  $K_G$  in (31) captures the analog gains along the way: the mixers, the filters, and the differentiators. Note that zero cross-correlation holds for AM, and also for FSK/PSK with a preceding source coding stage [24].

In the proposed algorithm, we implement a digital version of the BQ. A fundamental requirement for the BQ operation, either in analog or digital, is that the first mixing yields non-overlapping copies of  $s(t)$ . To ensure this

property, each  $s_i[n]$  is interpolated by a factor of three, and the positive and negative frequencies are re-positioned in  $[\pi/3, 2\pi/3], [-2\pi/3, -\pi/3]$ , respectively. For example, when no merging occurs this operation boils down to  $I_{6,1.5B}\{z_{\pm l}[n]\}$  with the relevant  $l$ . The digital BQ is applied on the outcome.

Our digital implementation consists of FIR lowpass filters, and the continuous derivatives are approximated by the finite difference – a filter with the discrete impulse response  $[1, -1]$ . Note that a wide family of filters can substitute the true differentiators [24]. The BQ is initialized with a normalized angular frequency  $\omega_0 = \pi/2$  and is repeatedly applied. At each iteration,  $\omega_0$  is updated by

$$\omega_0^{\text{new}} = \omega_0^{\text{old}} + G \frac{\sum_n v_d[n]}{\sum_n |s_i[n]|^2}, \quad (32)$$

where the loop gain  $G = 5 \cdot 10^6$ . The procedure monitors  $\omega_0 \in [\pi/3, 2\pi/3]$  and terminates upon convergence or if a pre-defined number of iterations is reached.

**Properties.** Upon completion, the desired information contents  $I(t), Q(t)$  of a band of interest are instantly available – these are the signals  $v_I(t), v_Q(t)$  from the last BQ iteration, see Fig. 13. In fact, due to the digital implementation, the information signals are already given in a uniformly-spaced sampled version which the DSP expects. The rate of  $I_i[n], Q_i[n]$  is either  $6B$  or  $12B$ , depending on the rate of  $s_i[n]$ . The recovered carrier  $f_i$  and the detected band edges  $[a_i, b_i]$  allow to reduce the rate of  $I_i[n], Q_i[n]$  to the minimal rate  $2(b_i - a_i)$ .

Besides the information signals  $I(t), Q(t)$ , the algorithm outputs additional useful information per band: the edges  $[a_i, b_i]$ , the isolated sequence  $s_i[n]$  and the carrier  $f_i$ . The latter is computed from the normalized angular frequency  $\omega_0$  that the BQ converged to by

$$f_i = B \left( l + c \frac{\omega_0 - \pi/3}{\pi/3} \right), \quad (33)$$

where  $c = 1$  when merging was not required, and  $c = 2$  otherwise. These products are utilized in Section VIII for spectrum sensing in cognitive radio applications.

For applications in which the exact  $B_{\min}, \Delta_{\max}$  are unknown, an approximate value can be set. The uncertainty with respect to the true values may yield many possible support regions in steps (1.1)-(1.2). Nonetheless, the effect on the overall performance is little, since only the  $N/2$  powerful regions are selected in step 1. Furthermore, the exact band locations have only minor effect on the filter design in step 2, as Fig. 12 depicts. The BQ iterations in step 3 are also insensitive to inaccuracies in  $[a_i, b_i]$ . Therefore, approximate values for  $B_{\min}, \Delta_{\max}$  are sufficient in practice.

## B. Simulations

To evaluate the proposed algorithm, we considered an example of a multiband model  $N = 6, B = 50$  MHz. Quadrature phase-shift keying (QPSK) modulation was used to generate

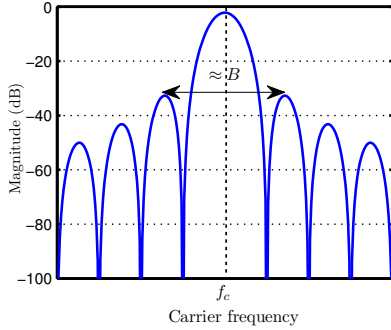


Fig. 14. The theoretical PSD of a QPSK signal.

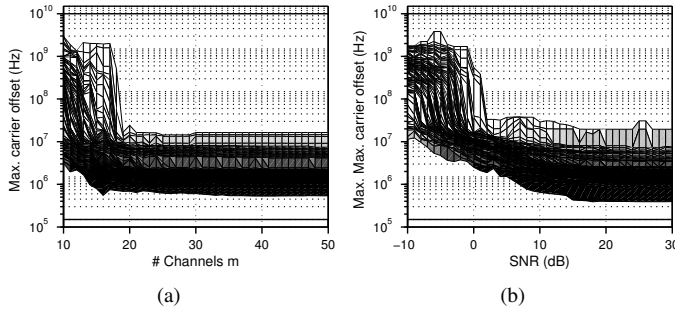


Fig. 15. Maximal carrier offset vs. number of channels and SNR level. In (a) SNR=20 dB and in (b)  $m = 30$  channels.

$$x(t) = \sum_{i=1}^3 x_i(t) \text{ via}$$

$$x_i(t) = \sqrt{\frac{2E_i}{T_{\text{sym}}}} \left( \sum_n I_i[n] p(t - nT_{\text{sym}}) \right) \cos(2\pi f_i t) \quad (34)$$

$$+ \left( \sum_n Q_i[n] p(t - nT_{\text{sym}}) \right) \sin(2\pi f_i t) + n(t),$$

where  $E_i = \{1, 2, 3\}$ ,  $1/T_{\text{sym}} = 30$  MHz,  $p(t) = \text{sinc}(t/T_{\text{sym}})$  are the symbol energy, rate and pulse shape. The carriers  $f_i \in [0, 5]$  GHz, the bit streams  $I_i[n] = \pm 1$ ,  $Q_i[n] = \pm 1$ , and the additive white Gaussian noise  $n(t)$  were all drawn independently at random. The theoretical PSD of a single QPSK transmission is illustrated in Fig. 14.

An MWC with the basic configuration was used with sign alternating waveforms  $p_i(t)$ ,  $M = 195$  alternations points per period  $T = 1/B$ . The CTF and the matrix inversion of Fig. 6 were carried out as suggested in [11]. The outputs,  $z_i[n]$ , were processed by our algorithm. For 100 test signals, we measure the carrier frequency offset (CFO) of each  $f_i$ . Fig. 15 reports the maximal CFO (for  $i = 1, 2, 3$ ) for various number of channels  $m$  and for several SNR levels. In the figures, gray intensity is used to report the empirical CFO distribution between the maximal and minimal values. As evident, for  $m \geq 20$  and SNR greater than 10 dB, most trials resulted in CFOs smaller than 350 kHz. For comparison, the 40 ppm CFO specifications of IEEE 802.11 standards, e.g. [53], tolerate this error for transmissions located around 10 GHz. The CFO can be further reduced in the DSP, based on QPSK-specific synchronization techniques.

## VIII. APPLICATION – SPECTRUM SENSING

The most intensive task of a sub-Nyquist system consists of acquiring an analog wideband input  $x(t)$  at a low rate, translating the samples to the format that standard DSP packages can deal with, and possible reconstruction of the analog  $x(t)$ . In this section, we consider a lighter application – the cognitive radio (CR) transceiver [25]. The prime goal of a CR is to identify spectrum holes, namely frequency intervals with no transmissions. The CR uses the spectrum holes, which are often licensed to a single primary user, in order to transmit secondary signals. The spectrum is repeatedly monitored, or sensed, and in case the primary user appears, the secondary transmissions must be stopped immediately and relocated to available spectrum holes.

Formulating the problem mathematically, the CR has an input multiband signal  $x(t)$ , and the goal is to find the complement of the spectral support. Clearly, this task is lighter than sampling the analog signal and providing the information contents  $I(t), Q(t)$ . Nonetheless, the implementation is in the wideband regime, and the computational cost should be minimal in order to adapt quickly to changes in the spectral support. Scanning the spectrum for the holes is not much easier than searching for an unknown carrier  $f_i$  in the wideband spectrum, for the reasons explained in Section II. Below we propose two degenerate configurations of the MWC that provide the spectrum sensing functionality.

**Spectrum sensing – option 1.** Assume a multiband model with an expected number of concurrent transmissions equal to  $N/2$ , and maximal expected bandwidth of a single transmission given by  $B$ . Design a standard MWC system. Then, execute only step 1 of the digital algorithm of Section VII. The output is a list of band positions  $[a_i, b_i]$ .

**Spectrum sensing – option 2.** Consider a multiband model with  $B = B_{\min}$ , the expected bandwidth of the narrowest primary transmission. The number of bands  $N$  is set such that  $NB \geq \Omega$ , where  $\Omega$  is the expected occupied bandwidth till  $f_{\max}$ . Both specifications are typically known in CR environments. Next, design an MWC system according to the flow described in [11] for this multiband model. The CTF will output the support set  $S$ , such that  $l \in S$  indicates the presence of signal energy in the length  $B_{\min}$  spectrum slice, centered around  $lB_{\min}$ . The union over all  $l \notin S$  indicates the spectrum holes. A note about analog implementation is in order. Since  $B = B_{\min}$  is typically small and  $N$  is large in this approach, the basic configuration with  $m \geq 4N$  channels may be impractical. To reduce the number of channels to a reasonable size, we may use one of the advanced MWC configurations [11], in which the sampling rate per channel  $f_s$  is set to  $q$  times  $f_p$ , where  $f_p = 1/T$ , and  $T$  remains the period of  $p_i(t)$ . This option conceptually collapses every  $q$  channels of the basic configuration to a single branch with  $q$  times higher sampling rate. The price for this solution is additional digital computations in the form of  $q$  digital filters per channel [11].

**Comparison.** In both solutions, the CTF block is used to recover the set  $S$ . The decomposition (24) can be avoided as suggested. For option 1, the digital processing ends after

step 1 of the baseband algorithm. Steps 2 and 3 are skipped. In option 2, the digital processing ends even earlier right after the CTF recovers the support estimate  $S$ . In either case, there is no need to extract the information  $I(t), Q(t)$  or perform any other computation.

The second approach has a clear advantage that there is no need to compute the pseudo-inverse  $\mathbf{C}_S^\dagger$  or to generate the sequences  $\mathbf{z}[n]$ . On the other hand, the first solution uses the basic configuration with  $q = 1$ , thus avoiding the additional digital filters required by the advanced configuration. In addition, the matrix  $\mathbf{C}$  in this setting is typically smaller, since the size of  $\mathbf{C}$  is inversely proportional to the aliasing rate  $1/T$  and since  $1/T = B$  is larger than  $1/T = B_{\min}$ . The price is the set of operations of step 1, as depicted in Fig. 11, and especially the PSD estimation. To decide on a solution, one must compare the computational costs for the specifications  $N, B, B_{\min}, \Omega$  at hand.

**Related works.** The relation between cognitive radio and compressed sensing was noted already in several publications [10], [26]–[30]. The parallel RD scheme [10] has the limitations mentioned in Section IV. The work [26] is based on a finite parametrization of analog signals, similar to the RD. The transition between analog signals and discrete CS, eq. (4) and (5) in [26] is not detailed. Pointwise sampling is used in [27], see eq. (22). In [28] the authors list implementation issues in the RD, though the approach taken eventually also uses a finite parametrization [28, eq. (4)] and the computational burden is noticed in the experiments. Finally, [29], [30] isolate all analog issues to an analog-to-information (AIC) device which is assumed to exist. They also use a finite representation of analog signals, and focus their developments on auto-correlations in the compressed domain. For the sake of decency, the authors point out that “*the scheme results in a somewhat paradoxical architecture since sub-Nyquist sampling is achieved by first sampling the wide-band analog signal at Nyquist rate and then applying CS*” [29], [30].

## IX. CONCLUSIONS

We have proposed the Xampling methodology as a fundamental tool for the design of sub-Nyquist system: broad signal model, low rate, efficient implementation and baseband processing are the necessary ingredients. This framework was supported by a comprehensive and careful examination of leading approaches from both classic sampling papers and recent CS publications.

We also described a digital algorithm that provides the MWC with the baseband processing capability. This turns the MWC to a powerful system, which is capable of sampling, processing the information contents and recovering the input. All operations involve only lowrate computations. We then considered an application to cognitive radio receivers. The MWC has prominent advantages in sub-Nyquist wideband scenarios. The ability of baseband processing makes the MWC strategy useful even when Nyquist sampling is possible, but processing at a low rate is needed.

A famous question by Donoho [14] triggered many works in the CS literature: “*Can we not just directly measure the part*

*that will not end up being thrown away ?*”. Interpreting this question for analog signals reduces to whether one can sample at a rate below Nyquist. The Xampling framework highlights additional considerations in designing of sub-Nyquist sampling, besides reducing the rate. Indeed, one *can directly measure* the part that will not end up being thrown away using the MWC. However, a breakthrough of the Nyquist barrier, in theory and in practice, must incorporate truly analog models and efficient hardware and software implementations together with the ability to process the information captured by the samples, without interpolating to the high Nyquist rate.

## REFERENCES

- [1] L. Bin, T. W. Rondeau, J. H. Reed, and C. W. Bostian, “Analog-to-digital converters,” *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 69–77, Nov. 2005.
- [2] R. G. Vaughan, N. L. Scott, and D. R. White, “The theory of bandpass sampling,” *IEEE Trans. Signal Process.*, vol. 39, no. 9, pp. 1973–1984, Sep. 1991.
- [3] A. Kohlenberg, “Exact interpolation of band-limited functions,” *J. Appl. Phys.*, vol. 24, pp. 1432–1435, Dec. 1953.
- [4] 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.
- [5] C. Herley and P. W. Wong, “Minimum rate sampling and reconstruction of signals with arbitrary frequency support,” *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1555–1564, Jul. 1999.
- [6] R. Venkataramani and Y. Bresler, “Perfect reconstruction formulas and bounds on aliasing error in sub-Nyquist nonuniform sampling of multiband signals,” *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 2173–2183, Sep. 2000.
- [7] M. Mishali and Y. C. Eldar, “Blind multiband signal reconstruction: Compressed sensing for analog signals,” *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.
- [8] G. L. Fudge, R. E. Bland, M. A. Chivers, S. Ravindran, J. Haupt, and P. E. Pace, “A Nyquist folding analog-to-information receiver,” in *Proc. 42nd Asilomar Conf. on Signals, Systems and Computers*, Oct. 2008, pp. 541–545.
- [9] 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.
- [10] Z. Yu, S. Hoyos, and B. M. Sadler, “Mixed-signal parallel compressed sensing and reception for cognitive radio,” in *ICASSP’08*, 2008, pp. 3861–3864.
- [11] M. Mishali and Y. C. Eldar, “From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals,” *arXiv.org 0902.4291; to appear IEEE J. Sel. Topics Signal Process.*
- [12] Y. C. Eldar, “Compressed sensing of analog signals in shift-invariant spaces,” *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 2986–2997, Aug. 2009.
- [13] M. Mishali and Y. C. Eldar, “Spectrum-blind reconstruction of multiband signals,” in *ICASSP’08*, pp. 3365–3368.
- [14] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.
- [15] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [16] M. Mishali and Y. C. Eldar, “Xampling. Part II: Theory,” *in preparation*.
- [17] Y. C. Eldar and T. Michaeli, “Beyond bandlimited sampling,” *IEEE Signal Process. Mag.*, vol. 26, no. 3, pp. 48–68, May 2009.
- [18] Y. M. Lu and M. N. Do, “A theory for sampling signals from a union of subspaces,” *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2334–2345, Jun. 2008.
- [19] T. Blumensath and M. E. Davies, “Sampling theorems for signals from the union of finite-dimensional linear subspaces,” *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1872–1882, April 2009.
- [20] Y. C. Eldar and M. Mishali, “Robust recovery of signals from a structured union of subspaces,” *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.
- [21] Y. C. Eldar, “Uncertainty relations for shift-invariant analog signals,” *arXiv.org 0809.3731; to appear IEEE Trans. Inf. Theory*, Sep. 2008.



- [22] K. Gedalyahu and Y. C. Eldar, "Low rate sampling schemes for time delay estimation," *arXiv.org 0905.2429*, May 2009.
- [23] M. Mishali, Y. C. Eldar, O. Dounaevsky, and E. Shoshan, "A 2.3 GHz board prototype of the modulated wideband converter," *in preparation*.
- [24] F. Gardner, "Properties of frequency difference detectors," *IEEE Trans. Commun.*, vol. 33, no. 2, pp. 131–138, Feb. 1985.
- [25] Mitola III, J., "Cognitive radio for flexible mobile multimedia communications," *Mobile Networks and Applications*, vol. 6, no. 5, pp. 435–441, 2001.
- [26] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *ICASSP'07*, vol. 4, Apr. 2007, pp. IV–1357–IV–1360.
- [27] H. Li, C. Li, and H. Dai, "Quickest spectrum sensing in cognitive radio," in *CISS'08*, Mar. 2008, pp. 203–208.
- [28] P. Zhang, P. Hu, R. C. Qiu, and B. M. Sadler, "A Compressed Sensing Based Ultra-Wideband Communication System," 2008, [Online]. Available: <http://dsp.rice.edu/files/cs/CS.Based.UWB.pdf>.
- [29] Y. Polo, "Compressive Wideband Spectrum Sensing for Cognitive Radio Applications," Master's thesis, Delft University of Technology, 2009.
- [30] Y. L. Polo, Y. Wang, A. Pandharipande, and G. Leus, "Compressive wide-band spectrum sensing," in *ICASSP'09*, Apr. 2009, pp. 2337–2340.
- [31] N. Boutin and H. Kallel, "An arctangent type wideband PM/FM demodulator with improved performances," in *Circuits and Systems, 1990., Proceedings of the 33rd Midwest Symposium on*, 1990, pp. 460–463.
- [32] "A/D Converters," *Analog Devices Corp.*, 2009, [Online]. Available: <http://www.analog.com/en/analog-to-digital-converters/ad-converters/products/index.html>.
- [33] R. Dorf, *The electrical engineering handbook*. CRC Pr I Llc, 1997.
- [34] R. Best, *Phase Locked Loops: Design, Simulation, and Applications*. McGraw-Hill Professional, 2007.
- [35] "Voltage controlled oscillators," *Mini-Circuits Corp.*, 2009, [Online]. Available: [http://www.minicircuits.com/products/vcos\\_main.html](http://www.minicircuits.com/products/vcos_main.html).
- [36] M. Unser, "Sampling – 50 years after Shannon," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.
- [37] P. P. Vaidyanathan, "Generalizations of the sampling theorem: Seven decades after Nyquist," *IEEE Trans. Circuit Syst. I*, vol. 48, no. 9, pp. 1094–1109, Sep. 2001.
- [38] Y. C. Eldar and T. G. Dvorkind, "A minimum squared-error framework for generalized sampling," *IEEE Trans. Signal Processing*, vol. 54, no. 6, pp. 2155–2167, Jun. 2006.
- [39] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4692–4702, Oct. 2008.
- [40] R. Calderbank, S. Jafarpour, and R. Schapire, "Compressed learning: Universal sparse dimensionality reduction and learning in the measurement domain," [Online]. Available: <http://dsp.rice.edu/files/cs/cl.pdf>, 2009.
- [41] C. Hegde, M. B. Wakin, and R. G. Baraniuk, "Random projections for manifold learning," in *Advances in Neural Information Processing Systems 20*, 2008, pp. 641–648.
- [42] M. B. Wakin, "A manifold lifting algorithm for multi-view compressive imaging," in *Picture Coding Symposium (PCS), Chicago, Illinois*, May 2009, [Online]. Available: <http://dsp.rice.edu/>.
- [43] R. Nowak, personal communications, 2009.
- [44] Y. C. Eldar and A. V. Oppenheim, "Filter bank reconstruction of bandlimited signals from nonuniform and generalized samples," *IEEE Trans. Signal Process.*, vol. 48, no. 10, pp. 2864–2875, Oct. 2000.
- [45] H. J. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Math.*, vol. 117, pp. 37–52, Feb. 1967.
- [46] H. Johansson and P. Lowenborg, "Reconstruction of nonuniformly sampled bandlimited signals by means of digital fractional delay filters," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2757–2767, 2002.
- [47] J. Elbornsson, F. Gustafsson, and J.-E. Eklund, "Blind equalization of time errors in a time-interleaved ADC system," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1413–1424, 2005.
- [48] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," in *Proc. IEEE Int. Conf. ASSP*, May. 1996, pp. 1688–1691 vol. 3.
- [49] Y. Chen, M. Mishali, Y. C. Eldar, and A. O. Hero III, "Modulated wideband converter with non-ideal lowpass filters," preprint, 2009.
- [50] E. Laskin and S. P. Voinigesu, "A 60 mW per Lane,  $4 \times 23$ -Gb/s  $2^7 - 1$  PRBS Generator," *IEEE J. Solid-State Circuits*, vol. 41, no. 10, pp. 2198–2208, Oct. 2006.
- [51] T. O. Dickson, E. Laskin, I. Khalid, R. Beerkens, X. Jingqiong, B. Karajica, and S. P. Voinigesu, "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.
- [52] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.
- [53] "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: High-speed physical layer in the 5 GHz band," *IEEE Std. 802.11a-1999*.