RADS Converter: An Approach to Analog to Information Conversion

Javier Haboba
ARCES
Università di Bologna
via Toffano 2/2, 40125 Bologna, ITALY
Email: jhaboba@arces.unibo.it

Riccardo Rovatti
ARCES
Università di Bologna
via Toffano 2/2, 40125 Bologna, ITALY
Email: rrovatti@arces.unibo.it

Gianluca Setti
ENDIF
Università di Ferrara
via Saragat 1, 44100 Ferrara, ITALY
Email: gianluca.setti@unife.it

Abstract—While classical compressive sensing aims to reduce the number of measurements with respect to Nyquist-based sampling methods, it is usually important to consider the total number of bits needed to represent this small amount of measurements in order to maintain the same signal quality. In this paper we study an architecture for signal acquisition that produces a stream of 1-bit measurements, and we show that a proper exploitation of its features, as well as of the sparse representation of the input signal can result in an extremely high reconstruction accuracy with a total number of bits much smaller than classical sampling.

I. INTRODUCTION

Analog to Digital conversion is one of the most important operations in signal processing. It maps a continuous-time and real-value signal into a discrete sequence of discrete values. Classical sampling methods rely on the hypothesis that the analog signal to be acquired is band-limited, and the Nyquist-Shannon theorem states the minimum distance between samples (or Nyquist rate) needed to uniquely describe the analog signal by its samples.

The newly introduced paradigm of Compressive Sensing (CS) [1], [2] uses an extra hypothesis on the signal called “sparsity”. Under this extra hypothesis, CS theory states that the number of samples (measurements in the general case) can be reduced well below the one stated by the Nyquist-Shannon theorem.

Formally, we may focus on a normalized acquisition time of one second and sample the signal $x(t)$ at the Nyquist rate $N$ by collecting $x[n]=x\left(\frac{n}{N}\right)$ for $n=0, \ldots, N-1$. Clearly $x \in \mathbb{R}^N$ and $x$ is sparse if there is an $N \times N$ matrix $\Psi$ such that $x=\Psi s$ for some vector $s \in \mathbb{R}^N$ in which at most $K << N$ components are non-zero. CS relies on taking $M$ linear measurements of the form $y = \Phi x$, where $y \in \mathbb{R}^M$ and $\Phi \in \mathbb{R}^{M \times N}$, to recover the vector $s$ and thus $x$. Clearly, the “sparsity” hypothesis is used to find the sparsest $s$ that is coherent with the measurements.

So far we have used real valued measurements and neglected the quantization process always present in practical implementations. Quantization can be modeled as noise in the measurements $y = \Phi x + e$, and the most robust reconstruction algorithms rely on the Restricted Isometry Property (RIP) of the measurement matrix $\Theta = \Phi \Psi$. Roughly speaking, though $\Theta$ cannot be a bijection, it should be as close as possible to an isometry (length preserving mapping) for vectors that are sparse as the $s$ we are willing to recover. This property guarantees signal recovery with a reconstruction error proportional to the noise level, that in the case of quantization noise it will depend on the quantization step.

Yet, the quantization step and the signal range determine the number of bits needed to encode each sample, and a careful design of a CS system should take into account not only the number of measurements but the overall number of bits generated by the acquisition mechanism. This is particularly important when considering measurements of the kind $y = \Phi x$ since each component $y[j] = \sum_{k=0}^{N-1} \Phi_{jk} x[k]$ has a range (e.g., $\pm 3\sigma$ around the signal average as it would be common if $y[j]$ can be considered Gaussian) that is potentially $\sqrt{N}$-times larger than that of $x[n]$.

The classical approach to CS is to take $M << N$ to compensate for the burden of finely quantized measurements with a large dynamic range. This is most natural when the quantization noise must be assumed white and thus can be reduced only by lowering its variance that is proportional to the quantization step.

What we propose here is to exploit the noise shaping capabilities of Delta Sigma ($\Delta/\Sigma$) structures and produce a large number of measurements ($M \geq N$) each coarsely quantized (actually with only 1 bit). We show that, also in this case, a proper exploitation of sparsity allows for substantial compression in the number of acquired bits. Random Delta Sigma (RADS) converters, first introduced in [3], are the result of this method. Although 1-bit encoding for compressive measurements was first studied in [4] and [5], we here present a complete architecture for Analog to Information conversion.

In this paper, we evaluate the performance of the RADS converter and we propose a new reconstruction algorithm that exploits the peculiarities of its acquisition strategy to produce an estimate of the signal with an accuracy that depends on its information content. The algorithm also achieves a very high probability of successful reconstruction over different sparsity conditions. We also give an analytical expression for the estimated Reconstruction Signal to Noise Ratio (RSNR), that is validated with simulation results. The conversion architecture in conjunction with the reconstruction algorithm is shown to outperform other approaches for analog-to-information conversion.
Fig. 1. RADS Converter schematic diagram.

II. RADS CONVERTER ARCHITECTURE

A schematic diagram of the RADS converter is shown in Figure 1. The input signal \( x(t) \) is first multiplied by the antipodal signal \( p(t) \) that alternates between the levels \( \pm 1 \) at the Nyquist frequency \( N \) of \( x(t) \). After modulation, the resulting signal is sampled at rate \( M \geq N \) to produce the discrete sequence \( y[n] \) that feeds a discrete-time \( \Delta/\Sigma \) modulator of order \( L \). At the output of the system we get a single bit sequence \( z[n] \). Under certain sparsity conditions on the input signal \( x(t) \), the circuit is able to collect the necessary information for signal recovery, with a total number of bits much smaller than classical sampling methods for a given \( RSNR \). The simplicity of the architecture allows it to operate at very high frequencies, making possible for example to acquire with a high resolution frequency sparse signal, that are spread over a large bandwidth.

A. Encoding

To achieve the expression of the vector \( z \in \mathbb{R}^M \) corresponding to the output sequence \( z[n] \), let us start by writing \( x = \Psi s \), with \( x \in \mathbb{R}^N \), \( s \in \mathbb{R}^N \), and \( \Psi \in \mathbb{R}^{N \times N} \) and where \( s \) has at most \( K \) non-zero elements. The analog waveform \( x(t) \) corresponding to the samples in \( x \) is sampled at frequency \( M \). To model this we use the \( M \times N \) sinc-interpolation matrix \( A \)

\[
A_{jk} = \text{sinc} \left( \frac{N-1}{M-1} (j-1) - (k-1) \right) \quad j = 1, \ldots, M, \quad k = 1, \ldots, N
\]

yielding the new samples \( Ax = A \Psi s \).

The signal \( p(t) \) alternates randomly between \( \pm 1 \) every \( 1/N \) seconds. The multiplication by \( p(t) \) ensures that for any signal \( x(t) \), the resulting signal \( y(t) \) will contain in any frequency sub-band a contribution for every non-zero component in \( s \).

If \( p(t) \) is generated by the sequence \( p_1, p_2, \ldots, p_N \), then its samples at the sampling rate \( M \) can be aligned in the matrix

\[
P_{jk} = \begin{cases} p_j/M & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases}
\]

Then, the output binary sequence is \( z = PAx + e = \Theta s + e \) where \( \Theta = PA \Psi \) and \( e \) is the quantization noise due to the \( \Delta/\Sigma \). The spectrum of the quantization noise, is defined by the Noise Transfer Function (NTF) of the modulator. If we consider a small band (compared with the sampling frequency) around DC, the \( \text{rms} \) noise is given approximately by [6]

\[
e_{\text{rms}} = e_{\text{rms}} \frac{\pi^t}{\sqrt{2L+1}} \left( \frac{M}{2B} \right)^{-\left(L+\frac{1}{2}\right)}
\]

where \( B \) is the considered band and \( e_{\text{rms}} = 1/\sqrt{2} \) for 1-bit mid-riser quantizer and \([-2,2]\) (uniform distributed) input range.

B. Decoding and reconstruction

In general, signal reconstruction for a CS acquisition scheme can be split into two parts: support recovery, i.e. the identification of the location of the nonzero components, and amplitude estimation over that support. Consider first the situation in which the support is already known. If the columns of the measurement matrix \( \Theta \) indexed by the location of the nonzero components form a full-rank matrix, the natural approach is to reconstruct the signal by least squares and the approximation error will be only limited by the power magnitude of the noise introduced by the measurement process.

On the other hand, in the general case where the support is not known, most algorithms can be ensured to work based on the RIP of the measurement matrix as stated on the introduction. For some matrix construction with entries that are Gaussian or sub-Gaussian, the RIP is satisfied with overwhelming probability if the number of measurements is bigger than a multiple of the signal sparsity \( M \geq CK \log(N/K) \). A common phenomena in compressive sensing is the so called phase transition effect i.e. if the number of measurements falls below a certain minimum number, the probability of successful reconstruction change from a very high to a very poor one.

Thanks to spreading, every non-zero entry in \( s \) implies a waveform whose energy can be detected at practically any frequency including those where the quantization error is reduced by the \( \Delta/\Sigma \). Hence, to remove the quantization noise it is desirable to take only a small bandwidth around zero. On the other hand, the considered signal should contain enough information for the recovery of the original sparse signal. The correct choice of the bandwidth will determine the system performance. The nature of the architecture allows us to develop an iterative algorithm that recovers the support with a very high probability, and reduces the quantization noise to the minimum possible depending on the sparsity level.

The algorithm is based on iterative filtering, decimation, and estimation. We can model the filter process as the application of an \( l \)-order FIR filter \((l \leq m) \) and arrange its impulse response coefficients \( h_1, h_2, \ldots, h_l \) as the rows of a matrix \( H^{(m)} \) of \( m \times m \) elements, where \( m \) is the length of the sequence to be filtered.

\[
H_{jk}^{(m)} = \begin{cases} h_i & \text{if } j = k + i - 1 \\ 0 & \text{if } j \neq k + i - 1 \end{cases}
\]

Once filtered, the sequence can be decimated by a factor \( r \) multiplying it by the matrix \( D^{(r,m)} \) of \( \left[ m/r \right] \times m \) elements

\[
D_{jk}^{(r,m)} = \begin{cases} 1 & \text{if } j = \left[ \frac{k}{r} \right] \\ 0 & \text{if } j \neq \left[ \frac{k}{r} \right] \end{cases}
\]
In Algorithm 1 we show the pseudocode of a possible implementation of the reconstruction algorithm. It is divided into an inner and an outer loop. At every iteration the outer loop filters and decimates the signal in order to remove the quantization noise. At the end of the algorithm, we should remain with at least the 2K values needed to differentiate two different K-sparse vectors x. To simplify the implementation, we fix the decimation factor to two, and we limited the total number of iterations.

The inner loop is a greedy pursuit algorithm based on CoSaMP [7], that produces a new signal estimate on every iteration. The amplitude estimation is made by least square we fix the decimation factor to two, and we limited the total number of iterations plus the support of the largest components of the residuals of the previous iteration.

### Algorithm 1: Reconstruct x from 1-bit vector z

<table>
<thead>
<tr>
<th>Require: Sampling matrix Θ, 1-bit vector z, sparsity level K.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I ← \frac{\text{total # of iterations}}{\log_2(M/2K)}</td>
</tr>
<tr>
<td>m ← M</td>
</tr>
<tr>
<td>s ← (0,...,0)^T</td>
</tr>
<tr>
<td>while m &gt; 4K do</td>
</tr>
<tr>
<td>z ← D(2,m)H(m)z</td>
</tr>
<tr>
<td>Θ ← D(2,m)H^{(m)}Θ</td>
</tr>
<tr>
<td>v ← z − Θs</td>
</tr>
<tr>
<td>m ← \left\lfloor m/2 \right\rfloor</td>
</tr>
<tr>
<td>for \ i = 1,2,...,I do</td>
</tr>
<tr>
<td>w ← Θv</td>
</tr>
<tr>
<td>T ← {\text{supp}(w_{K}) \cup \text{supp}(s_{K})}</td>
</tr>
<tr>
<td>b(T) ← Θ(\cdot,T)\dagger s</td>
</tr>
<tr>
<td>b({1,...,N}\setminus T) ← (0,...,0)^T</td>
</tr>
<tr>
<td>s ← b_{K}</td>
</tr>
<tr>
<td>v ← z − Θs</td>
</tr>
<tr>
<td>end for</td>
</tr>
<tr>
<td>T ← \text{supp}(s_{K})</td>
</tr>
<tr>
<td>b(T) ← Θ(\cdot,T)\dagger s</td>
</tr>
<tr>
<td>b({1,...,N}\setminus T) ← (0,...,0)^T</td>
</tr>
<tr>
<td>s ← b_{K}</td>
</tr>
<tr>
<td>end while</td>
</tr>
<tr>
<td>\hat{x} ← \Psi s</td>
</tr>
</tbody>
</table>

Intuitively, the high probability of correct support recovery comes from the fact that we estimate it under large noise condition, but large number of measurements. Once the support is identified (at every iteration the support estimation is improved), the bandwidth is decreased in order to reduce the quantization noise. The key fact is to note that the signal energy decreases linearly when the frequency decreases, while noise energy decreases polynomially thanks to the ΔΣ noise shaping properties [6]. This combination of filtering and estimation, has the benefit of recovering the signal with a very high probability of success, while reducing the quantization noise to the minimum.

### III. Reconstruction Signal to Noise Ratio Estimation

It is important to analyze the noise introduced by the complete acquisition system, i.e. quantization noise and errors produced in the reconstruction procedure. The reconstruction quality is measured by the (RSNR) as

\[
\text{RSNR} = 20\log_{10}\left(\frac{||x||_2}{||x−\hat{x}||_2}\right)
\]

where \(\hat{x}\) is the reconstructed signal.

We can calculate the total power of quantization noise as in (1) considering the remaining bandwidth at the end of the algorithm, that in the best case is only twice the sparsity level.

The price we pay to have a contribution of every possible component of the signal in the low pass portion of the band is that we have to spread its energy in the whole bandwidth of the original signal. Since the energy is the same before and after the random multiplication (we have multiplied by a ±1 sequence), the magnitude of the signal that remains at the end of the reconstruction algorithm will be inversely proportional to the original signal bandwidth.

Finally, we can estimate the RSNR as

\[
\text{RSNR} = 20\log_{10}\left(\frac{||x||_2}{N\epsilon_{rms}\sqrt{\frac{\pi}{2L+1}}\left(\frac{M}{4\pi}\right)^{−(L+\frac{1}{2})}}\right)
\]

### IV. Numerical results

In this section, we evaluate the performance of the converter plus the reconstruction algorithm by a series of numerical experiments. All the simulations were run 1000 times for which we drew a new signal with a random support in every trial. All the plots shown the mean value over all the trials. The 1-bit encoding was made using third and fifth order ΔΣ modulators designed with delsig [8]. The filters used in the reconstruction process have all sinc^3 frequency response, and the total number of iterations was fixed to 40.

For the first experiment we consider an input signal that is K-sparse in the Fourier domain. \(x = Fs\), where

\[
F_{j,k} = e^{-\frac{2\pi}{M}(j−1)(k−1)} \quad j,k = 1,\ldots,N
\]

The power of the input signal was kept constant along the experiment, but since the phases combine randomly, we kept the amplitude of those components different from zero to \(C = \frac{1}{\sqrt{2}}\) in order to maintain 99% of the signal amplitude values in the range \([-0.5,0.5]\). Higher input amplitude could produce instability in the modulator loop resulting in an extra signal distortion.

In Figure 2 and 3 we plot the RSNR as a function of the oversampling ratio \(M/N\) fixing the number of 1-bit compressive measurements to \(M = 2048\) and for different sparsity
levels $K$. The simulated $\text{RSNR}$ is shown as a solid line while the theoretical estimation in (2) is depicted as a dashed line. The estimated $\text{RSNR}$ is shown to follow the simulated system behavior, in terms of variation with parameters $K$ and $N$ (the same occurs with $M$ but is not shown here). The differences are due to the linear model used in the approximation of the in-band noise of the $\Delta/\Sigma$ converter (that is a very non-linear system), and the non-ideal behavior of the filters used. Clearly, the expression (2) can be used as a design guideline. It is also important to note the large $\text{RSNR}$ that is achieved especially for very sparse signals. Note that for $M/N = 1$ we are just taking 1-bit measurement at Nyquist rate.

Another interesting fact, not shown explicitly in the plots for space reasons, is that the support recovery was always correct for $K < 40$, while it was substantially less than 100% only for $K \geq 40$ when low oversampling ratios are considered. This is mainly due to the large bandwidth that remains at the end of the algorithm, i.e., to the residual noise energy that is large with respect to the signal energy.

In the second experiment we compare the performance achieved by our system with a state of the art 1-bit compressive sensing algorithm, i.e., the Restricted Step Shrinkage (RSS) [5], that is a generic scheme working on measurement matrices with nice theoretical properties rather than on a suitably designed signals. To avoid possible biases due to the choice of a particular sparsity basis, every trial simulates the acquisition of a signal that is $(K = 10)$-sparse along a random basis obtained by orthonormalizing a matrix with Gaussian independent entries with zero average. Independently of the architecture, the same amount of 1-bit samples $\mathbf{z}$ are considered as input for the reconstruction algorithm.

In all cases, the $\text{RADS}$ scheme was able to perfectly reconstruct the support of the signal to acquire and, as shown in Figure 4, achieve an $\text{RSNR}$ largely superior to that of the reference one.

V. CONCLUSION

In this paper we propose a new reconstruction algorithm for the $\text{RADS}$ converter, and we evaluate its performance by means of theoretical results and numerical simulations. The architecture allows a “simple” hardware implementation for the acquisition of large bandwidth signals that are sparse over a variety of supports, obtaining a very high resolution after reconstruction. This is in contrast with classical sampling methods, where the resolution drastically decreases with the sampling frequency. We also evaluate numerically the quality of the algorithm to retrieve a correct support under different input signal condition, obtaining a very high probability over a wide range of sparsity levels.

Finally, we compare our proposal with the RSS algorithm for 1-bit compressive sensing, showing that a careful design of the measurement scheme may lead to substantial improvement.

REFERENCES