

BEYOND BAND LIMITED SAMPLING IN SUPER RESOLUTION APPLICATIONS – 60 YEARS AFTER SHANNON

Gamal Fahmy

German University in Cairo, Egypt gamal.fahmy@guc.edu.eg

(Received February 20, 2011 Accepted May 25, 2011)

In this paper we present the topic of sampling of signals that are not band limited 60 years after the well known Shannon-Nyquist sampling theory. We consider the sampling process of several signals that were not exactly considered or treated by Shannon. We can call them non band limited signals. Research in sampling of non band limited signals is divided into two main categories. Signals that are sensed using modern state of the art compressive sensing techniques, and signals with a finite number of degrees of freedom that are sensed through sampling kernels that are physically realizable. These latter types of signals are typically called signals with finite rate of innovation. We present our results with these two main categories with different images and signals. Illustrative examples of the proposed techniques are presented.

1. INTRODUCTION

Driven by the massive development of digital applications over the past few decades with its analog sources and its associated discrete time representations, the sampling theory became a central element in many modern signal processing, sampling rate conversion, coding and decoding and communication systems. To accommodate the high data acquisition rate, while maintaining affordable computational cost, efficient Analog to Digital sampling conversion has to be developed.

Traditionally, sampling conversion has been conducted under the well known Shannon-Nyquist umbrella, which states that data needs to be sampled at twice the signal bandwidth to be able to reconstruct it perfectly. While Shannon's theorem has been one of the greatest contributions during the last century that has had a huge amount of impact on modern electrical engineering, it suffered from several problems such as its assumption that signals are always band limited in terms of their frequencies, which is not usually the case as we will show latter, and its dependence on ideal low pass filters as Sinc functions.

Recently, intense research has investigated the problem of sampling signals beyond the Nyquist limit; these efforts were successful when dealing with non band limited signals such as streams of Diracs or piecewise polynomial signals. The majority of these efforts focused on signals that have a sparse representation [1]. These algorithms managed to perfectly reconstruct signals from non uniform measurements through L_1 norm optimization of nonlinear sampling. These efforts, typically classified as compressive sensing algorithms, considers Shannon's theorem as a special case for the Perfect Reconstruction PR of digital samples.

More recently, non band limited signals have been perfectly reconstructed from uniform samples that are sampled far below the Shannon Nyquist limit. In [2] and

[3] perfect reconstruction was achieved for streams of Diracs that are sampled uniformly with special kernels. These special kernels can reproduce polynomials or exponentials and are also physically realizable. This input signal, that was a stream of Diracs, was sampled uniformly with those special sampling kernels and had a finite number of degrees of freedom per unit of time. Hence, they are typically called signals with a finite rate of innovation (FRI) [2].

The main objective of this paper is to present our analogy of non band limited sampling using compressive sensing (CS) techniques and compare it with the traditional Shannon-Nyquist theorem. We also present and propose techniques for efficient sampling of non band limited signals which can have a wide spectrum of impact on different sensors, data acquisition, rate conversion, and super resolution applications. This is due to the fact that it captures a minimal amount of data samples, that can be stored and transmitted, and can later perfectly reconstruct the original input signal at the receiver side. This reconstruction process is based on a pre-knowledge of the input signal structure/characteristics, or the sampling kernel order/response, and is also physically realizable in a PR manner. We present our simulated results on different low resolution images for super-resolution purposes.

Section 2 gives a brief background of the sampling process of non band limited signals using Compressive Sensing techniques (CS). Section 3 introduces the concept of finite rate of innovation of uniformly sampled signals and shows how it can also be perfectly reconstructed at a data rate that is far below the Shannon Nyquist limit. Section 4 presents a sample of our simulation results for sections 2 and 3, followed by conclusions in section 5.

2. COMPRESSIVE SENSING OF SPARSE SIGNALS

The main idea behind Compressed Sensing, [4-5], is to acquire analog samples and convert them into a digital form, and then compress them simultaneously. This would be like lump summing the two major blocks, sampling and transformation, into one single block. The only important feature that should exist in signals that are compressively sensed, it that they are sparsely in some kind of a plane, or frame. This sparsity feature will guarantee that there is a frame in space that the signal can be projected to and have the minimum number of non-zero coefficients.

Data are acquired from an analog image using a linear set of measurements, like big pixels, or lines or sinusoids. Then, an adaptive approximation for the sparse basis in the signal/image is performed through some kind of optimization.

Pre-knowledge of the signal characteristics is the main reason for the Compressive Sensing efficiency in coding and decoding. If we consider a time domain signal $f(t)$, as in fig.1, that consists of a known number of sinusoidal signals that have un known periods and amplitudes. We will only know that they are 15 sinusoids (for example), but the values of α 's and β 's in the Fourier Domain are unknown, eq. 1. Due to the fact that they correspond to the unknown periods and amplitudes, $f(w)$ and $f(t)$ in eq. 1 correspond to the frequency and time representations, with $K=15$.

By taking 30 random measurements of the time domain sequence in fig.1, which correspond to the $f(t)$ values in eq.1, we can determine the unknown values of α 's and β 's, that represent the amplitudes and frequencies of the Fourier domain version of the signal, as in fig.1. These 30 values represent the circles in fig.1, and they

should be randomly distributed over the time domain range. They should also have a structure that is un-similar to the basis in the Fourier domain, as will be shown next

$$f(\omega) = \sum_{i=1}^K \alpha_i \delta(\beta_i - \omega) \quad \Leftrightarrow \quad f(t) = \sum_{i=1}^K \alpha_i e^{i\beta_i t} \quad (1)$$

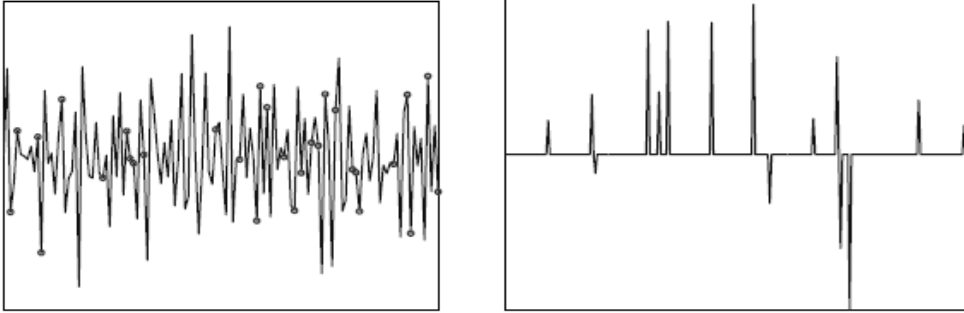


Fig. (1) Time domain (Left) and Frequency domain (Right) signals, where the time domain is reconstructed from circles through CS algorithm

It is well known that the more the values of $f(t)$, the faster is the optimization process and the more accurate are the determined α 's and β 's values. Interestingly to know, the fact that we can get adaptive approximation performance of the basis in the projected domain, which is the Fourier Domain in our case, from a fixed set of measurements. More surprisingly, as stated before, the measurements should not match the projection basis structure, as in eq. 2.

$$\Phi \psi \alpha = y \quad (2)$$

Eq.2 represents the acquisition/transformation process in Compressed Sensing (CS), where Φ represents the measurements matrix, and ψ represents the sparsity matrix, which is the basis matrix of the projected space. α is the input data vector in the un compressed domain, and y is the delivered output data in the compressed form. While the measurements are typically global, and incoherent, the sparsity matrix should be local and coherent.

samples in the sparse domain, as shown in Fig.2. If we assume the original acquired data is the α vector, then with the knowledge of a domain, space or frame, where if the input data vector is transformed into it will be sparse (majority of samples zeros), we can deliver just the few non zero samples out of the acquisition process (non zero samples of output vector y). These delivered y non-zero samples are far beyond the Shannon/Nyquist limit and corresponds to the output samples of the compressed sensor. Perfect reconstruction is guaranteed from these y samples with the knowledge of the projection plane where the input data is sparse. The Φ matrix, which is the measurement matrix, projects the input data vector α into a sparse domain, while the ψ matrix, which represents the sparsity matrix, selects the non zero samples to be delivered as output y samples. Hence for an input vector α with N samples, the size of the measurement matrix should be $N \times N$, while the size of the sparsity matrix is $N \times M$, with M samples as delivered output y , and $N \gg \gg M$, Fig.2.

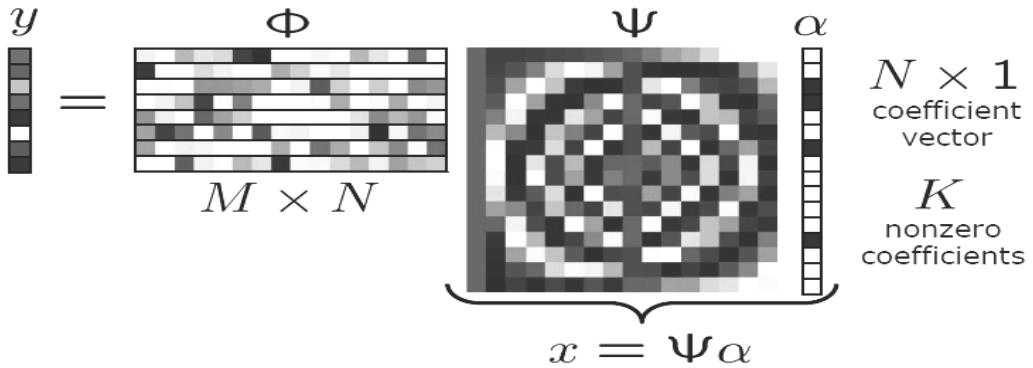


Fig. (2) a graphic illustration for the measurement and sparsity

It is worth mentioning that there may be a small amount of error in the reconstruction quality due to the non optimal optimization process. However, as the number of measurements increases, the error decreases (and also get determined faster) at a near –optimal rate. The sparsity projected domain matrix can be any well known domain such as DCT, wavelet, Hadamard, or it can be an application specific domain that is pre-determined.

We note here that while the sensing algorithm (measurements) is non-adaptive, the projection basis optimization is adaptive and requires some kind of post sampling computing power.

The optimization process of the projection space basis, is calculated through the L_1 norm equation, eq.3. This L_1 norm equation would require that the ψ should be unstructured with respect to ϕ and would be the only type of norm that can work (not L_2 norm or any type of norm), as will be shown in the next equation.

$$\min_{\alpha} \|\alpha\|_{L_1} \text{ using the } L_1 \text{ so that } \Phi \psi \alpha = y \quad (3)$$

The L_1 norm is used, since it guarantees that during the optimization process, the projected basis will converge to zeros and non zeros, while the L_2 norm will result in small amounts (values) to even smaller (may be negligible) values, but still can't be classified into zeros or non zeros, which is against the sparsity concept.

From a different point of view and analysis, in this optimization process we also try to look to a common point between two *orthogonal* planes, the original data plane (which has the horizontal axis x), and the projection plane (which must have the point of intersection of the planes). The optimization process is like the point of origin and then we increase it in different direction until it meets the 2 planes.

With the L_1 norm this increasing size origin point would be like the diamond point (as diamond corresponds to L_1 optimization), and the intersection point is a diamond tangent with the 2 planes simultaneously, while the L_2 norm would be like the circle point (as the circle point corresponds to L_2 optimization), and the intersection point is a circle tangent with the 2 planes, which can never happen except in perpendicular planes (Shannon case). Fig.3 shows the intersection in both cases (diamond which corresponds to L_1 and circle which corresponds to L_2).

In quick comparison with the well known Shannon-Nyquist sampling limit, Compressive Sensing (CS) can sense data samples at a sampling rate that is far beyond

the Shannon limit. In our time domain example, fig.1, the number of required samples to recover $f(t)$ is far higher than 30 samples, which is required with CS. With Shannon, the reconstruction happens from equally spaced samples in a known space and the reconstruction is done with Sinc interpolation in a linear manner, while with CS the reconstruction happens from unequal spaced (arbitrarily spaced) samples in an unknown space, and the reconstruction is done with convex programming in a non linear manner.

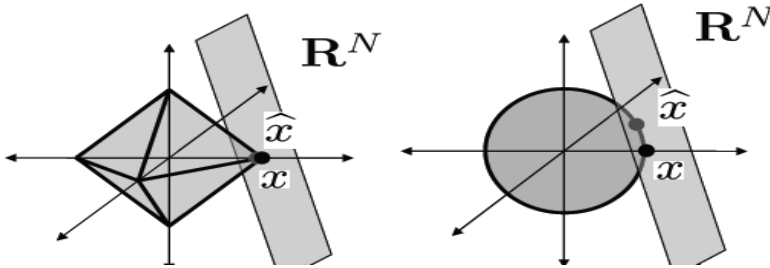


Fig. (3) a graphic illustration of the L_1 and L_2 norms for basis optimization. L_1 norm is the only one that can work

The beauty and out performance of CS lies not only in the fact that compression is built in measurements, but also in the fact that all measurements are equally important or unimportant (democratic), and losing some of them due to any noise or communication problem is negligible. The recovery is also robust and flexible and independent of the acquisition process.

CS can also be applied not only in sensing applications for higher sampling rates at lower bandwidths for non band limited signals, but it can be exploited in different sensor network or super-resolution applications that can afford to measure data in separate modalities, but reconstruct in a joint manner.

Several applications has been reported in the recent literature that exploit the CS theory such as, single pixel camera [1], CS analog to digital converters, and other classification applications

3. PERFECT RECONSTRUCTION OF UNIFORMLY SAMPLED SIGNALS THAT HAS A FINITE RATE OF INNOVATION

The main idea behind perfect reconstruction (PR) of uniformly sampled signals which have a finite number of degrees of freedom per unit of time, which is known as finite rate of innovation, is that these signals result from acquiring a stream of Diracs through a kernel that can reproduce polynomials. The resulted signal will have a signature of both the input Dirac and the sampling kernel, and would be like the graph in fig.6. Digital samples are obtained after sampling the time domain output signal (that is the result of convolution between the input Diracs and the sampling kernel) with period T . This period T can also go beyond the Shannon limit and will still allow PR of the input Diracs given some restrictions on the relationship between the number of Diracs and the kernel order. In [4] the B-spline family was used as a sampling kernel and perfect reconstruction was achieved for K input Diracs given the kernel order is $2K+1$, as will be shown next.

We first calculate $N+1$ moments of the input signal $x(t)$, with the following equation:

$$\tau_m = \sum_n c_{m,n} y_n = \left\langle x(t), \sum_n c_{m,n} \varphi(t-n) \right\rangle = \left\langle \sum_{k=0}^{K-1} a_k \delta(t-t_k), \sum_n c_{m,n} \varphi(t-n) \right\rangle = \sum_{k=0}^{K-1} a_k t_k^m, m=0,1,\dots,N \quad (4)$$

Where $c_{m,n}$ are the basis from which we can construct a polynomial of the sampling kernel with max degree N , y_n are the sampled digital samples, as in fig.6, $\varphi(t)$ is the sampling kernel, a_k are the amplitudes of the input Diracs, and t_k are the locations of the input Diracs at the input time domain range

From the spectral estimation literature, we can assume the existence of a filter $H(z)$, with h_m 's as its coefficients, and where its roots are the t_k 's and the following equation would be applied

$$h_m * \tau_m = \sum_{i=0}^K h_i \tau_{m-i} = \sum_{i=0}^K \sum_{k=0}^{K-1} a_k h_i t_k^{m-i} = \sum_{k=0}^{K-1} a_k t_k^m \sum_{i=0}^K h_i t_k^{-i} = 0 \quad (5)$$

This $H(z)$ filter, is typically called an annihilating filter and by calculating the h_m coefficients, we can determine the t_k values, as the t_k values are the roots of this annihilating filter. The only restriction that we have is that the number of input Diracs K , should be according to $N \geq 2K - 1$, where N is the order of the sampling kernel $\varphi(t)$

We can calculate the h_m values, from the following Yule-Walker coefficients:

$$\begin{bmatrix} \tau_{K-1} & \tau_{K-2} & \dots & \tau_0 \\ \tau_K & \tau_{K-1} & \dots & \tau_1 \\ \cdot & \cdot & \dots & \cdot \\ \tau_{N-1} & \tau_{N-2} & \dots & \tau_{N-K} \end{bmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \cdot \\ h_K \end{pmatrix} = - \begin{pmatrix} \tau_K \\ \tau_{K+1} \\ \dots \\ \tau_N \end{pmatrix} \quad (6)$$

And then we can calculate the amplitudes of the input Diracs, according to the following equation, which is a Vandermonde optimization equation.

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ t_0 & t_1 & \dots & t_{K-1} \\ \cdot & \cdot & \dots & \cdot \\ t_0^{K-1} & t_1^{K-1} & \dots & t_{K-1}^{K-1} \end{bmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \cdot \\ a_{K-1} \end{pmatrix} = - \begin{pmatrix} \tau_0 \\ \tau_1 \\ \dots \\ \tau_{K-1} \end{pmatrix} \quad (7)$$

From all above, we can show that it is possible to perfectly reconstruct uniformly sampled signals, that have a finite number of degrees of freedom per unit of time, and hence they are called signals with a finite rate of innovation, like Diracs. We mean by non band limited signals, signals that are sampled below the Shannon-Nyquist limit. The only restriction we have is that the sampling kernel, that the signal is filtered though before sampling, has to be able to reproduce polynomials, like the B-spline family, and its order should be around twice the number of input Diracs over a time domain range.

This reconstruction idea can be applied to any analysis or decomposition approach to analyze input uniformly sampled data, such as the wavelet decomposition. From different wavelet bands we can select data samples that can be used for reconstruction through the proposed nonlinear reconstruction eq. (4-7).

Figure 9, illustrated the reconstruction idea for signals with a finite rate of innovation (FRI), from the uniformly sampled digital samples through a known kernel (that can reproduce polynomials) like a B-spline kernel. We can perfectly reconstruct the input analog signal in discrete form with a number of samples much less than what is needed with the regular Shannon/Nyquist approach.

4. SIMULATION RESULTS

We applied the Compressed Sensing theorem in section 2 on many time domain signals and projected them into the sparse domain, which was the Fourier Domain in our case; Fig.1 shows the original and reconstructed signal.

We also applied the same Compressed Sensing (CS) approach on a low resolution image, and projected it in the wavelet domain with 5 layers, similar to many well known wavelet based compression techniques like JPEG2000, or SPIHT). Fig. 5 shows the sensed low resolution image that was delivered as y output in eq.2. Fig. 4 shows the original input high resolution image that can also be reconstructed from the y samples (which is the low resolution image in fig.5 in this example).

We also applied the perfect reconstruction of Finite rate of innovation signals theorem in section 3 on a stream of Diracs, shown in Fig.7. The discrete filtered samples are in Fig.9, the impulse response of the sampling kernel is in Fig. 8, while the output reconstructed Diracs in time domain spectrum are displayed in fig.10.

Figure 11, shows the wavelet decomposition samples of a piecewise linear signal. Again it is possible to construct the signal in Fig 11, from uniform samples in Fig.12. Fig.13 shows the reconstructed Lina image where each 8 pixels are treated as 8 Diracs (1-dimensional), and has been filtered through the Bspline kernel, then got reconstructed according to the technique in section 3.



Fig. (4) original High resolution image that can be Perfectly reconstructed from the delivered image in Fig. 5



Fig. (5) delivered sensed Low resolution image from CS

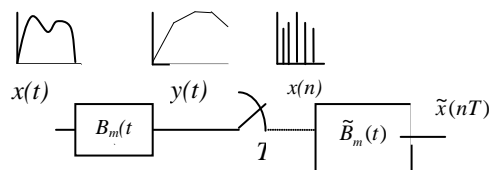


Fig.6 B-spline Reconstruction of samples

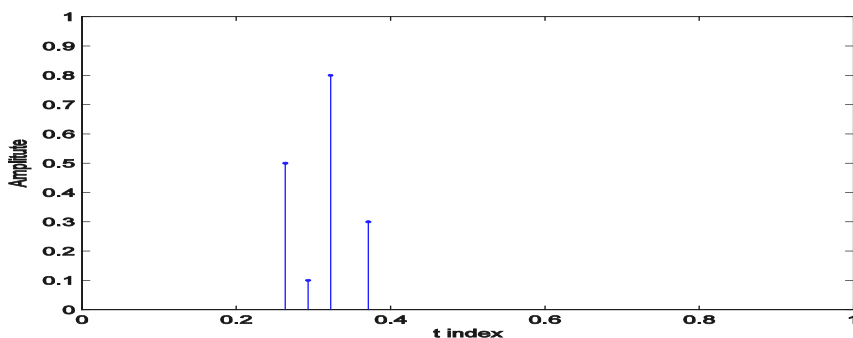


Fig.(7) Input Stream of Diracs

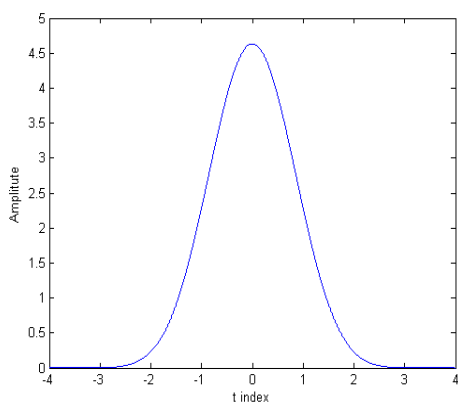


Fig.(8) Impulse Response of a B-spline sampling Kernel, $x(t)$.

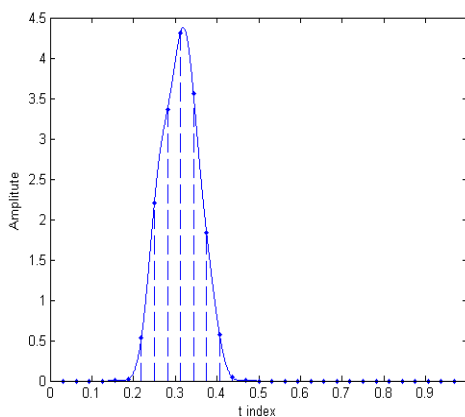


Fig.(9) Low pass filtered output of a B-spline sampling kernel, $y_0(t)$

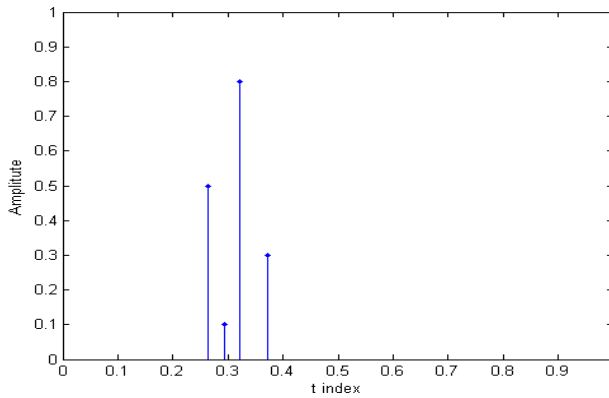


Fig.(10) Final PR samples after F0 and F1 convolution and addition, $x(nT)$

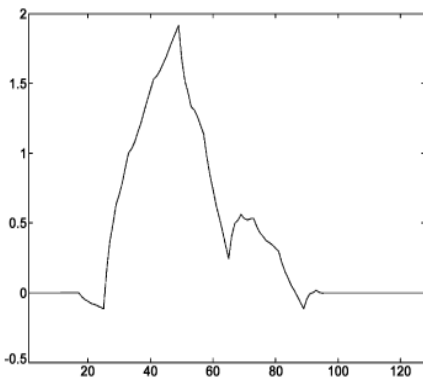


Fig. (11) wavelet decomposition samples of a piecewise linear signal

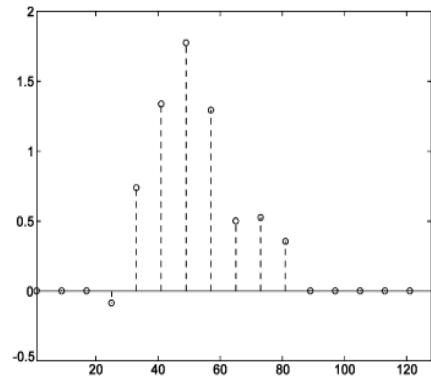


Fig. (12) Reconstructed samples of a signal in Fig. 11



Fig. (13) Reconstructed Lina image, from taking each 8 pixels as a stream of diracs and filtering them through a B-spline kernel and then reconstructing them

5. DISCUSSION

We note here that the presented algorithm for the perfect reconstruction for non band limited signals has different restrictions; it is our main belief that they are useful extensions to the Shannon Nyquist umbrella, but they don't replace them. They only deal with a special type of signals that were not investigated by Shannon.

The pre-knowledge of the sparse domain space, that sensed data will be projected to, is a restriction in the Compressed Sensing acquisition theorem, while the existence of a stream of Diracs signal sampled with a special type of kernel that can reproduce polynomials with the number of input Diracs around half the kernel order is a restriction in the theorem presented in section 3.

However, both algorithms were able to deliver output time domain signals that are perfectly error free and through a number of samples that are much less than what is required by Shannon-Nyquist in any regular sampling case.

In spite of these restrictions, the presented algorithms are considered as a significant breakthrough during the last decade of the sampling and reconstruction theory literature, as research was at a real dead end when dealing with sampling reconstruction for 50 years [7], until the presented techniques were released. There is a general belief/ conscious that we are currently in the non band-limited sampling era.

Several imaging [5], sensing [1], and super resolution applications [6] have tried to exploit the advanced non band limited signal sampling theorems, and they were reported successful to a certain extent, and in some specific applications.

6. CONCLUSIONS

In this paper we presented the topic of sampling of signals that are not band limited. Two algorithms were presented that represent the state of the art sampling reconstruction advances during the last decade

Sensing data with compressed sensors (CS), that depend on the sparsity of the signal for perfect reconstruction. And uniformly sampled signals with a finite number of degrees of freedom that are sensed through sampling kernels that are physically realizable. Further research in this area, would include investigating what other types of signals that can be reconstructed using the proposed theorems, along with exploring other applications that would benefit from the proposed research. This work is funded from the Alexander von Humboldt foundation, Germany and from the ministry of Communications and Information Technology, ITIDA, Egypt.

7. REFERENCES

- 1.M. Vetterli, P. Marziliano and T. Blu, Sampling signals with finite rate of innovation, *IEEE Transactions on Signal Processing*, Vol. 50, Nr. 6, pp. 1417-1428, 2002.
- 2.R. G. Baraniuk, "Compressive Sensing," Lecture Notes in *IEEE Signal Processing Magazine*, Vol. 24, No. 4, pp. 118-120, July 2007.
- 3.P.L. Dragotti, M. Vetterli and T. Blu, Sampling Moments and Reconstructing Signals of Finite Rate of Innovation: Shannon meets Strang-Fix, *IEEE Trans. on Signal Processing*, vol.55 (5), pp. 1741-1757, May 2007.

4. Y. C. Eldar and T. Michaeli, "Beyond Bandlimited Sampling," IEEE Signal Proc. Magazine, vol. 26, no. 3, pp. 48-68, May 2009
5. David Donoho, Compressed Sensing, IEEE Transaction on Information Theory, vol. 52, pp. 1289-1306, April 2006
6. L. Baboulaz and P.L. Dragotti, Exact Feature Extraction using Finite Rate of Innovation Principles with an Application to Image Super-Resolution, IEEE Trans. on Image Processing, Vo. 18(2), pp. 281-298, February 2009
7. M. Unser, "Sampling—50 Years After Shannon," Proc. IEEE, vol. 88, no. 4, pp. 569-587, 2000

الأشارات الرقمية ذات المجال الترددي الغير محدد بعد ستون عاما على نظرية شانون

في هذه المقالة نعرض أحدث الطرق لتحويل الأشارات التناظرية لأشارات رقمية بعد أكثر من ستون عاما على نظرية شانون الشهيرة ذات التطبيقات اللامحدودة في نظرية الاتصالات. حيث نقوم بتعريف الأشارات الرقمية ذات المجال الترددي الغير محدد و تقسم بشكل عام لأشارات تم ألتقاطها بأستخدام أسلوب الألتقاط المضغوط الحديث الشهير و أشارات ذات معدل أستحداث محدد في المجال الترددي و تم ألتقاطها بأستخدام مرشحات مجال ترددي عملية التطبيق. وهذه الأشارات الثانية الأخيرة أسمها أشارات ذات معدل أستحداث محدد في المجال الترددي. و لقد تم إجراء الكثير من التجارب و التطبيقات لأختبار الطريقة و فاعليتها كما هو موضح بالمقالة.