RAND PPM : A LOW POWER COMPRESSIVE SAMPLING ANALOG TO DIGITAL CONVERTER

Praveen K. Yenduri[1], *Anna C. Gilbert*[2], *Michael P. Flynn*[3] *and Shahrzad Naraghi*[4]

Departments of EECS[1−3−4] and Mathematics[2], University of Michigan, Ann Arbor, MI 48109

ABSTRACT

Analog-to-digital converters that digitize time information are known for their low power consumption. Coupling them with sampling techniques that take advantage of the signal compressibility leads to a further efficient data conversion. In this direction, we propose a new analog-to-digital converter, **rand PPM**, that employs compressive sampling techniques to efficiently sample at sub-Nyquist rates. The PPM (pulse-position-modulation) architecture uses a periodic ramp signal as reference and compares it with the analog input signal to eventually measure the ramp crossing times which determine the signal amplitudes. By appropriately introducing randomness into the reference ramp signal the PPM ADC architecture can be modified into a compressive sampling analog-to-digital converter. The sub-sampled signal is reconstructed using the developed algorithms tailored for practical hardware implementation. We have developed a theoretical analysis of the hardware system and reconstruction algorithm along with a suite of numerical experiments that support the theory.

***Index Terms*—** Analog to digital conversion, time to digital converters, compressive sampling, low power ADC

1. INTRODUCTION

Low power analog to digital converters find a number of applications in areas ranging from power constrained wireless environmental sensing to biomedical monitoring devices. One way to reduce the power consumption of an ADC is to employ time domain signal processing. The analog voltage input amplitude is converted into time delay which is in turn digitized with an efficient time-to-digital converter. Exploiting these facts, a low power ADC called the PPM (pulse-position-modulation) ADC was proposed in [1]. Other time encoded designs have been proposed in [2]. However they process the signal in power hungry analog domain and a conversion to digital domain leaves no advantage over the PPM design. Another way to improve the power efficiency of an ADC is to decrease its sampling frequency ([3]) since to first order power is proportional to sampling frequency. This can be achieved by employing random sampling techniques that exploit the sparsity or compressibility of the input signal to reduce the sampling rates to sub-Nyquist. We present a new low power compressive sampling analog to digital converter which we call a random PPM ADC, that lies at the intersection of these two trends. The rand PPM is a modification of the existing PPM ADC architecture. The reconstruction algorithms developed for the rand PPM are tailored to reduce computational cost and thus are viable for practical hardware implementation of the reconstruction unit.

This work is supported by NSF Award 0910765. Gilbert is an Alfred P. Sloan Fellow and partially supported by NSF DMS 0547744 and DARPA ONR N66001-06-1-2011.

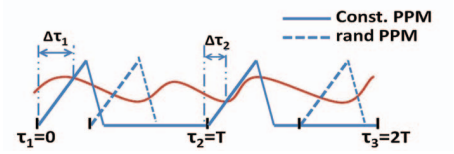

Fig. 1. Sampling scheme in the regular and the rand PPM

2. HARDWARE DESIGN

The PPM ADC sampling technique [1] is shown in the Figure 1. The input signal is continuously compared with a reference ramp voltage. Let $F = 1/T$ denote the sampling frequency (sampling period T) of the ADC. The time elapsed between the starting point of the ramp and the instant the input signal crosses the ramp (i.e. $\Delta\tau_1, \Delta\tau_2, ..$) is measured and quantized by a time to digital converter. Keeping the slope of the ramp signal constant, these $\Delta\tau_i$ are proportional to the signal amplitudes y_i at the ramp crossing times given by $t_i = \tau_i + \Delta\tau_i$, where $\tau_i = (i-1)T, \forall i$. The reconstruction technique used in [1] requires the sampling frequency to be above Nyquist rate.

By using the reconstruction algorithms in section 3.2 (or section 3.4), the sampling frequency of the PPM ADC can be reduced to a rate below Nyquist. A straight forward way to operate PPM ADC as a compressive sampling ADC is to sub-sample the input signal by simply reducing the sampling frequency F to $F < F_N$, where F_N is the Nyquist frequency of the input signal. We refer to this sampling architecture as the regular PPM. In the rand PPM design some randomness is introduced into the system. Specifically let $\tau_i - (i-1)T \sim \text{Uniform}[0, T], \forall i$. (with appropriate boundary conditions so that the ramps do not intersect). The number of measurements $K = NF/F_N$, where N is the number of samples obtained if the input signal was sampled at Nyquist rate. We will see in section 4 that rand PPM is much more successful in reconstructing the sub-sampled input signal compared to the regular PPM.

3. RECONSTRUCTION ALGORITHMS

3.1. The measurement matrix

Let the N-length vector X represent the input signal in the Fourier domain. We say that X is s-sparse if it has only s dominant terms. Let $(t_i, y_i), i = 1, .., K$ denote the measurements obtained from the ADC. We relate the input signal X with the measurement vector y through the equation $BX = y$, where B is the measurement matrix which is a function of $t_i, i = 1, .., K < N$. The matrix is shown in Figure 2. It is to be noted that B is not a random sub-matrix of the N-point IDFT Matrix, since t_i are non-uniform and do not lie on any Nyquist grid. The matrix B can be intuitively constructed by making the observation that if f is the only frequency in the signal X i.e, $x = exp(j2\pi ft)$ then the samples from x at time points $t_1, .., t_K$ are given by $\{exp(j2\pi ft_i), i = 1, .., K\}$. Putting $f = (n/N)F_N$,

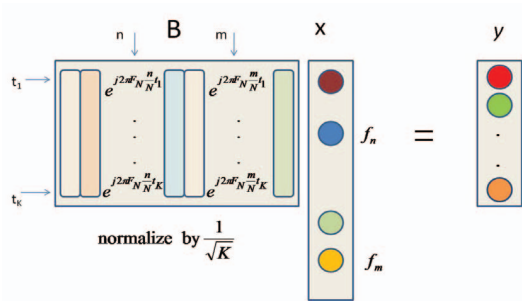

Fig. 2. Measurement matrix

the samples form the n^{th} column of the measurement matrix, for $n = [-N/2 : N/2 - 1]$ (if N even) or $[(-N-1)/2 : (N-1)/2]$ (if N odd). The random PPM ADC samples the signal at a rate proportional to its finite rate of innovation. However, the algorithms proposed in [4] cannot be used as they need the signal to be sampled at Nyquist rate and uniformly. B is a signal-dependent non-uniform random Fourier matrix, and as such, does not necessarily satisfy the Restricted Isometry Property (RIP) assumed in [5] or the conditions assumed in [6]. This leads to different algorithms with different theoretical analysis. We present a probabilistic approach. To aid the analysis we assume that the phase of the input signal is a random variable which also contributes to the randomness of B.

3.2. ALGORITHM 1

INPUT: N (Block length), s (sparsity), (t_i, y_i),$i = 1, 2, .., K$.

OUTPUT: X (Signal in frequency domain, length N but s-sparse)

$X_0 = \underline{0}$, residual $r_0 = y$
T =supp$\{[B^H y]_{2s}\}$, where $[z]_s$ refers to s largest terms in $|z|$
$X_{2s} = \left(B_T^H B_T\right)^{-1} B_T^H y$ (Least Squares)
$[X_0]_T = X_{2s}$, $r_0 = r_0 - B_T X_{2s}$

for $i = 0, 1, 2, ..$
$X_{i+1} = [X_i + B^H r_i]_s$
$r_{i+1} = y - BX_{i+1}$.
until $||r_{i+1}||_2$ does not vary within a tolerance θ.

Theorem 1 *Let $y = BX + \xi$ be the time domain samples of signal X obtained through the rand PPM ADC sampling block, where ξ is an arbitrary noise contamination in the measurements and B is the resultant measurement matrix of size KxN. Suppose $|_sX|^2 \geq 2\alpha||X||_2^2/s + |_{(s+1)}X|^2$ for some constant α and a given sparsity parameter s, where $|_iX|$ is the magnitude of the i^{th} largest element of X. Given the error tolerance in reconstruction θ and $K = \mathrm{O}(s/\epsilon^2)$, with probability $> 1-\mathrm{O}(\epsilon^2)$ the algorithm produces an s-term estimate $\tilde{X}$ of signal with the following property,*

$$||X - \tilde{X}||_2^2 \leq 2.\max\left\{\theta^2, \frac{||X - X_{(s)}||_2^2}{1-\alpha} + \frac{c(B)||\xi||_2^2}{1-\alpha}\right\} \quad (1)$$

where $X_{(s)}$ is the best s-term approximation of X. The runtime is $\mathrm{O}(IN\log N)$ where I = no. of iterations, which has a gross upper bound of I $<\log(||X||_2/\theta)$. The net storage requirement is $\mathrm{O}(N)$+$\mathrm{O}(sK)$. $c(B)$ depends on the measurement matrix B.

3.3. Analysis of ALGORITHM 1

The following lemma bounds the correlation between different columns of the rand PPM measurement matrix.

Lemma 2 *Let $I_f = [-N/2 : N/2 - 1]$ (if N even) or $[-(N-1)/2 : (N-1)/2]$ (if N odd). For $n, m \in I_f$ and $n \neq m$, define $C_{n,m} = B_m^H B_n$ where B_i is the i^{th} column of B. We have,*

$$|\mathbb{E}(C_{n,m})| = |\mathbb{E}(C_{1,|n-m|})| \leq \mathrm{O}\left(\frac{1}{N}\right) \quad (2)$$

Lemma 3 *If number of measurements $K = \mathrm{O}(s/\epsilon^2)$ then each estimate of the form $\tilde{X}_m = B_m^H BX$ satisfies*

$$\mathbb{E}(\tilde{X}_m) = X_m + \mathrm{O}\left(\frac{1}{N}\right)||X||_1 \quad (3)$$

$$\mathrm{Var}(\tilde{X}_m) \leq \frac{\epsilon^2}{s}||X||_2^2 \quad (4)$$

Lemma 3 says that the estimators produce correct values and their second moments are bounded. Now we present a brief sketch for the proof of theorem 1.

Proof: Let us begin with signal X exactly s-sparse. For simplicity lets assume that $X_i, i = 1, .., s$ are the non-zeros. There exists a $\beta < 1$ such that $|X_i|^2 \geq \beta||X||^2/s$ for these i. Let $0 < 2\alpha \leq \beta$. It can be proved using Chebyshev inequalities that by the end of k^{th} iteration the number of good coefficient estimators among $i = 1, .., s$ is $\geq (1 - (\frac{1}{4})^k)s$, with probability $\geq (1 - \frac{4\epsilon^2}{\alpha})^k$. Where the probability for an estimator to be good, i.e $\mathbb{P}\mathrm{r}\left(\tilde{X}_i^{(k)}\text{good}\right)$ is

$$\mathbb{P}\mathrm{r}\left(|\tilde{X}_i^{(k)} - X_i|^2 \leq \frac{\alpha||X - \tilde{X}^{(k-1)}||^2}{s}\right) \geq 1 - \frac{\epsilon^2}{\alpha}$$

where $\tilde{X}^{(k)}$ is the estimate produced by the algorithm after k^{th} iteration. Similar statements can be made about coefficients for which $i > s$. Hence after sufficient number of iterations (say I),all the estimators are good which implies that all the non-zero terms will be identified by the algorithm with

$$\mathbb{P}\mathrm{r}(\text{Success}) \geq (1 - \frac{4\epsilon^2}{\alpha})^{2I} \approx (1 - \frac{8I\epsilon^2}{\alpha}) = 1 - O(\epsilon^2)$$

after absorbing some constants into the number of measurements. Note that in reality since all the estimators are highly dependent, the probability that they will be good together is higher than the product of the individual success probabilities, which is the gross lower bound produced by the above theory. Now let the signal X be not exactly s-sparse. For $i > s$ we assume that $|X_i|^2 \leq \gamma||X||^2/s$ for some $\gamma < 1$. The above arguments can be repeated with $0 < 2\alpha \leq \beta - \gamma$. Now, for any $k > I$,at iteration $k+1$, $|(\tilde{X}_i^{(k+1)} - X_i)|^2 < \alpha||X - \tilde{X}^{(k)}||^2/s$ for $i = 1, .., s$. Summing up the s inequalities we get,

$$||\tilde{X}^{(k+1)} - X_{(s)}||^2 \leq \alpha||X - \tilde{X}^{(k)}||^2$$

where $X_{(s)}$ is the best s-term approximation to X. Now, $||X - \tilde{X}^{(k+1)}||^2 \leq ||X - X_{(s)}||^2 + ||X_{(s)} - \tilde{X}^{(k+1)}||^2 \leq ||X - X_{(s)}||^2 + \alpha||X - \tilde{X}^{(k)}||^2$. This implies, $||X - \tilde{X}^{(k+1)}||^2 \leq \frac{1-\alpha^{k-I}}{1-\alpha}||X - X_{(s)}||^2 + \alpha^{k-I}||X - \tilde{X}^{(k-I)}||^2$. For k large enough we have,

$$||X - \tilde{X}||^2 \leq \frac{1}{1-\alpha}||X - X_{(s)}||^2$$

This is consistent with Equation 1. For $y = Bx + \xi$ a similar expression can be derived. ∎

3.4. ALGORITHM 2: Median Of Estimators (MOE)

Note that in ALGO 1 the input signal was sampled for a total time of $t = N/F_N$ to get K samples. By collecting the samples from m consecutive blocks of time t, we get mK measurements. Assume that the set of top s frequencies in the signal remains the same in all the m blocks of time (their coefficients can change). Then we

can take a median over the estimators from different blocks to improve the identification of the top s frequencies. The idea of taking a median instead of mean was used in the count sketch algorithm [7] which estimates the most frequent items in a data stream. We propose to use the following algorithm to identify and estimate the top s frequencies in the signal. Let $B(i)$ be the measurement matrix formed (as shown in Section 3.1)from the time points in the i^{th} block of time, for $i = 1, .., m$. Similarly let $y(i)$ be the vector of measurements obtained from the i^{th} block.

INPUT: N (Block length),m (No. of Blocks),s (sparsity) $(t_i, y_i), i = 1, 2, .., m * K$.
OUTPUT: $X(i)$ for $i = 1, .., m$ (signal in each block)
Identification: For $j = 1, .., N$, $\hat{X}_j = \text{median}\{\lvert B(1)_j^H y(1)\rvert, .., \lvert B(m)_j^H y(m)\rvert\}$ $T = \text{supp}([\hat{X}]_s)$ **Estimation:** For $i = 1, .., m$, $[\tilde{X}(i)]_T = \left(B(i)_T^H B(i)_T\right)^{-1} B(i)_T^H y(i)$

Theorem 4 *For $m = \text{O}(\ln(\frac{N}{\delta}))$, the MOE algorithm correctly identifies the set T of top s frequencies in the signal with* $\mathbb{P}$r(Success) $\geq 1 - \delta$.

Proof: Note that the arguments in proof of Theorem 1 hold for all the m blocks of time. Let $\tilde{X}_{ij} = B(i)_j^H y(i)$ for $i = 1, .., m$ and $j = 1, .., N$. Hence $\hat{X}_j$ =median($|\tilde{X}_{ij}|, i = 1, .., m$) for $j = 1, .., N$. From the proof before, $\mathbb{P}$r($|\tilde{X}_{ij}|$ good) $\geq 1 - \frac{\epsilon^2}{\alpha} = p$(say), for $i = 1, .., m$. Assuming $p > 0.5$, from Chernoff bound we have $\mathbb{P}$r($\hat{X}_j$ good) $\geq 1 - e^{-2m(p-0.5)^2} \geq 1 - \delta'$ for m =O($\ln(\frac{1}{\delta'})$). Hence $\mathbb{P}$r(Success) = $\mathbb{P}$r($\hat{X}_j$ good for $j = 1, .., N$) $\geq (1 - \delta')^N \approx 1 - \delta$ for $\delta' = \delta/N$. ∎

Note that the computationally intensive step of least squares is performed only once (per block of the signal) in both the algorithms. The least squares was implemented using the accelerated Richardson iteration [8] with runtime of O($sK\log(2/e_t)$) where e_t is a tolerance parameter. The structure of the measurement matrix lends us to use the inverse NUFFT [9] with cardinal B-spline interpolation for forming the products of the form $B^H r$, in a runtime of O($N\log N$). Hence the total runtime of ALGO 1 is dominated by O($IN\log N$) where I is the number of iterations. The per block runtime of ALGO 2 which has only one iteration is O($sK\log(2/e_t)$)+O($N\log N$), which is much less than that of ALGO 1. The sampling percentage ($K/N = mK/mN$) is the same for both the algorithms. (100% sampling implies sampling at Nyquist rate.)

4. RESULTS AND DISCUSSION

The regular PPM (also referred to as const. PPM in the figures) and the rand PPM sampling architectures (described in section 2) and the reconstruction algorithms were simulated in MATLAB. The time resolution t_r of the TDC block in the ADC induces some quantization into the measurements, which is kept about 7 bits (= $\log_2$(ramp on-time/t_r), with ramp on-time = 0.25 μsec and t_r = 2 nsec) for most of the experiments to follow (corresponds to a signal to quantization noise ratio of about 44 dB for an input sinusoid). The series of experiments compares the performance of the algorithms with each of the sampling methods. The experiments use the s-term Nyquist as the benchmark performance, which is the signal sampled at Nyquist rate at the same quantization level of the ADC and then truncated in frequency domain to keep only s terms. Note that the output SNR can be higher than the input SNR, as the algorithms only calculate the coefficients of the top s frequencies in the signal and thus inherently filter out the noise at other frequencies.

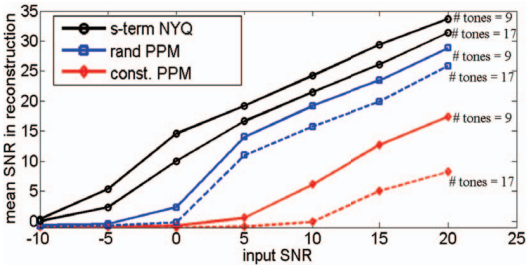

Fig. 3. Output Vs Input SNR for a multitone signal

Multitone signals In the first experiment we reconstruct multitone signals, which are a linear combination of sinusoids. Each sinusoid has a random phase, comparable amplitude and its frequency is chosen randomly from the Nyquist grid. The Nyquist frequency is 3 MHz whereas the sampling frequency of the ADC is chosen to be 1 MHz. The input signal is corrupted by additive white Gaussian noise, sampled by the two sampling schemes and reconstructed using the ALGO 1. In Figure 3 we plot the mean reconstruction output SNRs for two different number of tones (9 and 17). The experiment demonstrates the better performance of the rand PPM in two ways. First, it produces higher output SNR compared to the regular PPM and is also closer and approaching the benchmark as the SNR increases, owing to the better correlation properties of the measurement matrix (Lemma 2). Secondly as the number of tones increases (making the signal less sparse) the rand PPM output SNR is unaffected relative to the benchmark while the output of regular PPM degrades. This indicates that with same number of measurements, rand PPM design can handle less sparse signals much better than the regular PPM.

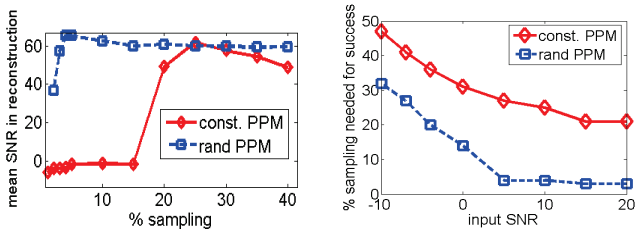

Fig. 4. Reconstruction of a single tone signal with varying no. of measurements (% sampling = $100K/N$) (a) with only quantization noise (b) with additional noise

Sampling percentage The next experiment reconstructs a single tone (random on-grid frequency) signal with varying number of measurements and noise levels using ALGO 1. The sampling percentage is given by $100K/N$. For demonstration purposes, as an example we choose to call the reconstruction a success when the output SNR $>$ rect(input SNR) +10 dB, where rect(a) = $(a + |a|)/2$. This definition guarantees that the reconstruction is at least 10dB greater than a positive input SNR and is appropriate even for a negative input SNR. The sampling percentage needed for success is empirically calculated for each input SNR level and is plotted in Figure 4(b). Observe that the rand PPM succeeds with a lot fewer measurements

than the regular PPM at all SNR levels and the gap increases with the input SNR. Further when the input SNR is high enough the % sampling for success in the rand PPM quickly drops down to about 3% and stays about the same. This can also be seen in the no noise (only quantization noise) case (Figure 4(a)), where the regular PPM scheme breaks down when the sampling goes below 20%, whereas the random scheme perform well enough for sampling as low as 2%.

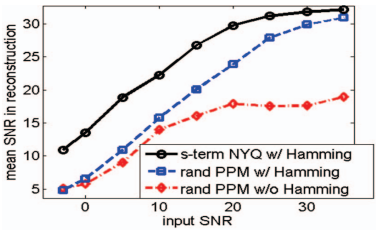

Fig. 5. Output SNR vs input SNR for a demodulated FM signal

FM signal with off-grid frequencies If a frequency falls in between two Nyquist grid points, it causes spectral spread or leakage, thus adversely affecting the sparsity of the signal. To counter this we propose to multiply the measurements from the ADC (before reconstruction) with a window function, like the Hamming (which is non-zero at all times, hence its effect can be reversed after the reconstruction). A frequency modulated (FM) signal with single tone message, where both the carrier and message frequencies are appropriately chosen to be off-grid acts as the input signal for this experiment. At 33% sampling, the noisy FM signal is windowed, reconstructed (using ALGO 1), demodulated and the resultant output SNR is plotted in Figure 5. Some output SNR is lost as the amplitude of the message is smaller than the dynamic range of the ADC. The use of Hamming window improves the performance of the algorithm at all SNR levels and approaches the benchmark as SNR increases. Similar observations have been made for amplitude modulated (AM) signals with different message signals. The plots exhibit similar qualitative behavior when the sampling rate is increased or decreased. Note that for an FM signal, windowing need not be reversed as the message is in the frequency of the signal.

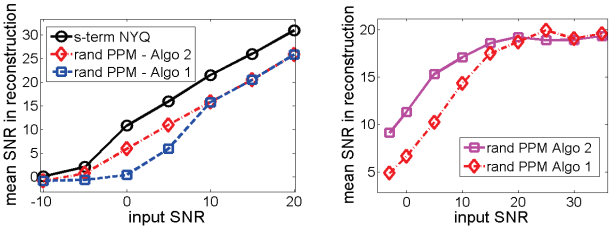

Fig. 6. Output Vs Input SNR for a (a) multitone signal (b) demodulated AM signal with a sawtooth message

Comparison of algorithms We now repeat the multitone (with 13 on-grid frequencies) signal reconstruction and the reconstruction of AM signal with sawtooth message (off-grid frequencies) experiments with ALGO 2, choosing the number of blocks $m = 7$. From Figure 6 we see that at low SNR conditions ALGO 2 gives a better performance than ALGO 1. This is because the identification stage in ALGO 2 is more successful as it nullifies the effect of noise to some extent by taking the median over a set of m blocks. At high SNR, both the methods give comparable performance even though ALGO 2 has only 1 iteration. Hence ALGO 2 can be used to reduce computations whenever the input signal satisfies the additional conditions (in section 3.4), particularly if it is also known that the input SNR levels are low. We also observed that when Hamming window is employed the performance of ALGO 1 improves whereas the ALGO 2 shows little to no improvement. This is because, upon application of the Hamming window the input signal does not strictly satisfy the assumptions made in section 3.4 and hence the improvement in sparsity of the signal is balanced by the error amplification due to Hamming window reversal.

5. CONCLUSION

We propose a new low power compressive sampling analog to digital converter called the rand PPM, which inherits the advantages of time to digital conversion as well as exploits the merits of compressive sampling techniques, for improving the power efficiency of the data conversion. An existing PPM ADC design can be modified to obtain the rand PPM, through random clocking and other changes in its modes of operation. The new random design enables the reduction of sampling rates to sub-Nyquist levels. The rand PPM performs much better than the non-random PPM operated at a sub-Nyquist sampling rate, in terms of obtaining closer-to-benchmark output SNR, handling signals that are less sparse and signals with off-grid frequencies. It also enables the usage of reconstruction algorithms such as those presented in the above sections, which are not only faster but also feasible for a hardware implementation. With on-the-chip reconstruction and a low power front-end, the rand PPM ADC is attractive for power constrained applications such as remote wireless sensor networks as it reduces the power consumption as well as the amount of data that needs to be communicated at each sensor location.

6. REFERENCES

[1] S. Naraghi, M. Courcy, and M.P. Flynn, "A 9b 14μW 0.06mm^2 PPM ADC in 90nm digital CMOS," in *International SolidState Circuits Conference*. IEEE, 2009, vol. 54, pp. 168–169.

[2] M. Kurchuk and Y. Tsividis, "Signal-Dependent Variable-Resolution clockless A/D conversion with application to CT-DSP," *IEEE Trans. Circuits and Systems*, vol. 57(5), pp. 982–991, May 2010.

[3] B. Murmann, "A/D converter trends: Power dissipation, scaling and digitally assisted architectures," in *Custom Integrated Circuits Conference (CICC)*. IEEE, 2008, pp. 105–112.

[4] J. Berent, P. Dragotti, and T. Blu, "Sampling piecewise sinusoidal signals with Finite Rate of Innovation methods," *IEEE Trans. Sig. Proc.*, vol. 58(2), pp. 613–625, February 2010.

[5] D. Needell and J.A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, pp. 301–321, April 2008.

[6] T. Blumensath and M.E. Davis, "Iterative thresholding for sparse approximations," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 629–654, September 2008.

[7] M. Charikar, K. Chen, and M. Farach-colto, "Finding frequent items in data streams," *Theoretical Computer Science*, vol. 312, pp. 3–15, January 2004.

[8] A. Bjorck, *Numerical Methods for Least Squares Problems*, Philadelphia: SIAM, 17 edition, 1996.

[9] G. Steidl, "A note on fast Fourier transforms for nonequispaced grids," *Advances in Computational Mathematics*, vol. 9, pp. 337–352, November 1998.