A Nonuniform Quantization Scheme for High Speed SAR ADC Architecture

Youngchun Kim
Electrical and Computer Engineering
The University of Texas
Austin, Texas, USA

Wenjuan Guo
Intel Corporation
Austin, Texas, USA

Ahmed H. Tewfik
Electrical and Computer Engineering
The University of Texas
Austin, Texas, USA

***Abstract*—We introduce a new signal sampling scheme which allows high quality signal conversion to overcome the constraint of effective number of bits in high speed signal acquisition. The proposed scheme is based on the popular successive approximation register (SAR) and employs compressive sensing technique to increase the resolution of a SAR analog-to-digital converter (ADC) architecture. We present signal acquisition and recovery model which provides better performance in signal acquisition. The sampled signal shows higher resolution after recovery than conventional compressive sensing based sampling schemes. Circuit level architecture is discussed to implement the proposed scheme using the SAR ADC architecture. Simulation result shows that the proposed nonuniform quantization strategy can be a way to overcome the sampling rate-resolution limitation which is a challenging problem in SAR ADC design even with the most advanced technology.**

***Index Terms*—Compressive sensing, SAR ADC, nonuniform quantization, signal recovery**

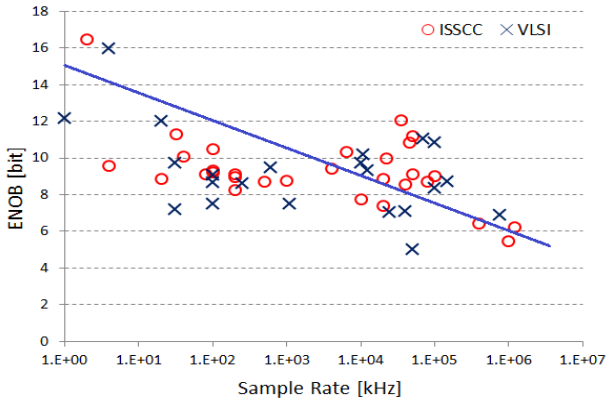


Fig. 1. Sampling rate and ENOB survey in SAR ADC [1].

I. Introduction

Analog-to-digital converters are one of the essential blocks in modern mixed-signal systems. In the market there are many different ADC architectures including sigma-delta, successive approximation register, pipeline, time-interleaving and flash. With the improvement in scaling technology, SAR ADC is becoming more and more popular due to its high power-efficiency and easy scaling with the technology since its most circuitry is in digital domain. However, for a SAR ADC, each conversion cycle only gives one more bit at a time and SAR ADCs require longer conversion time to obtain more bits. Therefore, the applications of SAR ADC is limited to medium resolution and medium speed. When the sampling speed reaches around 1GHz, the resolution of SAR ADC is limited to about 6-bit using the most advanced circuit technologies which is displayed in Fig.1. To overcome this limitation, we propose a sampling and signal conversion scheme based on the SAR ADC architecture which takes advantage of random signal sampling to increase the resolution of SAR ADC in the high-speed region.

Authors in [2], [3] proposed a SAR ADC architecture which is capable of sampling and reconstructing multi-channel information, and and they reported the circuit level implementation and measurement results in [4]. Yet, it does not provide enhanced signal quality: better resolution. In recent results, 1-bit compressive sensing (CS) [5] acquires coarsely quantized measurements to detect the active frequency components. Another approach [6] also investigates the trade-off of bit-depth versus measurement-rate in compressive sensing when you have limited store memory or data transmission rate. However, all these works are still doing uniform quantization, which is different from the nonuniform resolution case we are investigating in. In [7], the paper proposes an idea similar to our work and it works quite well for high SNR signals with very small number of measurements. However, the conversion scheme of [7] includes some limitations as well. The work is focused on the low speed sampling cases ($\leq$ 1kHz) with few number of measurements rather than high speed sampling application. The highly compressed CS strategy limits the allowable signal sparseness of interest in the sparse domain. Furthermore, the work does not fully utilize the SAR ADC architecture by varying quantization depth from 1-bit to 16-bit which possibly generate a wide range of variance in the sampled resolution and it leads to sub-optimal efficiency utilizing system resources. Lastly, the study only gives reconstruction error performance with the nonuniform quantization strategy while we include more concrete analysis on ADC architecture, which is more practical in circuit design. In terms of real circuit implementation, we think our architecture is highly feasible. We propose a different reconstruction algorithm by using element-wise constraint which is proved to be more efficient than the weighted constraint in [7]. Our simulation results show that we can get a better reconstruction performance than

ISBN 978-0-9928626-7-1

only doing coarse quantization.

The remainder of this paper is organized as follows: Section II introduces fundamental backgrounds in SAR ADC architecture and CS based sampling. Section III presents our signal acquisition and recovery strategy which enables high resolution sampling in high frequency signals. Section IV discusses details about implementation of the proposed nonuniform quantization scheme. Simulation results are demonstrated in Section V. Finally, we summarize the advantages and future works in Section VI.

II. Background

A. Data conversion in SAR ADCs

A data converter or signal acquisition system requires a quantization unit to convert input voltage to a finite number of bits. The SAR ADC architecture is getting more popular in practice for its high power-efficiency, and easy scaling with current technology since most of circuitry is designed in digital domain. SAR ADCs produce one more bit at a time in each conversion cycle, and a block diagram of the architecture is displayed in Fig.2.

The analog-to-digital conversion starts with enabling sample and hold (S/H) circuit to latch the input voltage V_{in} during the conversion cycle. The SAR logic directs the DAC to generate outputs through binary outputs which range the most significant bit (MSB) to the least significant bit (LSB). With the latched the input and DAC output, a comparator decides whether DAC output DAC_{out} is greater the latched V_{in} or not. The DAC starts producing DAC_{out} which is initialized by $1/2\ V_{ref}$. If DAC_{out} is larger than V_{in}, then the most significant bit (MSB) is set to zero which reduces DAC_{out} by half ($1/4\ V_{ref}$) in the next iteration. If DAC_{out} is smaller than V_{in}, then the MSB is set to one and the MSB remains to the rest of process. This binary search process is continuing iteratively from MSB to LSB. The CTRL logic controls time instances when the conversion starts and completes, and the iteration interval. Because of the iterative binary search process, to obtain more bits, SAR ADCs require longer conversion time and it causes sample rate vs. quantization depth constraint in high speed sampling. Therefore, the applications of SAR ADC

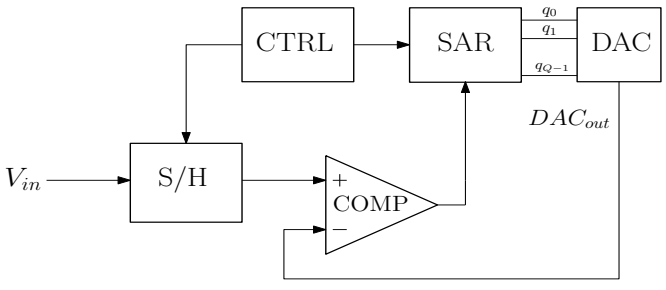


Fig. 2. A typical SAR ADC architecture.

is limited to medium resolution and medium speed. We hope to overcome this hardware constraint by making use of sub-Nyquist sampling strategy which randomly samples a sparse signal of interest in the frequency or DFT domain.

B. Compressive sensing

Recent study in the field of CS shows the other way to take fewer number of samples than those from Nyquist criteria. For the class of sparse signal, CS pursuits acquiring $M < N$ samples rather than N samples of a signal at Nyquist rate with linear measurement. Let $N \times 1$ input signal vector $\vec{x} \in \Re^N$ is K-sparse in a transform basis Ψ. One can rewrite it as $\vec{x} = \Psi\vec{\alpha}$, where $||\vec{\alpha}|\,|_0 = K$. In CS, a $M \times 1$ measurement $\vec{y}$ which is sufficient to represent the input signal $\vec{x}$ with measurement operator Φ:

$$\vec{y} = \mathbf{\Phi}\vec{x} = \mathbf{\Phi}\mathbf{\Psi}\vec{\alpha}. \tag{1}$$

Let us define a matrix $\mathbf{A} = \mathbf{\Phi}\mathbf{\Psi}$, and one can recover the input vector $\vec{x}$ without loss of information by solving a minimization problem:

$$\begin{aligned} &\text{minimize} \quad ||\vec{\alpha}||_1 \\ &\text{subject to} \quad \vec{y} = \mathbf{A}\vec{\alpha} \end{aligned} \tag{2}$$

where $\vec{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_N]^T$ and $||\vec{\alpha}||_1 = \sum_{r=1}^{N} |\alpha_r|$.

After sampling, the rest problem is to solve $\vec{\alpha}$ and reconstruct $\vec{x}$ via nonlinear optimization. One popular way to get $\vec{y}$ is randomly sampling $\vec{x}$. In this case, the $M \times N$ sampling matrix Φ is a diagonal matrix which consists of the random canonical basis (delta functions). For frequency sparse signal, if Ψ is a IDFT matrix with $N \times N$ dimension, $\mathbf{A}$ is a matrix consisting of randomly subsampled rows of the IDFT matrix. Such a matrix has a high probability of satisfying the Restricted Isometry Property (RIP), and thus is suitable for compressive sensing operations. Applying to frequency sparse signal, the matrix $\mathbf{A}$ is sampling randomly in time domain, and the best theoretic result of Fourier sampling shows that $M = O(K \log^4 N)$ is sufficient to satisfy the RIP condition [8]. To find the sparsest solution, the problem becomes a ℓ_1 minimization problem. Considering ADC's quantization noise, the minimization problem can be modified as,

$$\begin{aligned} &\text{minimize} \quad ||\vec{\alpha}||_1 \\ &\text{subject to} \quad ||\vec{y} - \mathbf{A}\vec{\alpha}||_2 \leq \epsilon, \end{aligned} \tag{3}$$

where the constant ϵ which bounds the measurement distortion. Without considering other noise effects, ϵ can be treated as the quantization noise sigma value.

III. Signal acquisition and recovery

In this section, we introduce our sampling and recovery schemes which enable high resolution signal acquisition in high speed sampling. First, we employ random sampling operator which is the same as CS and it is applied to the SAR ADC architecture, but the quantization depth is varied by the consecutive non-sampling period. If the sampling operator is one (*1*) at corresponding sampling time instance, the input signal is sampled and latched until the SAR unit finds the finite bit description: quantization stage. The quantization depth is decided by the number of following zeros (*0*s) which corresponds to non-sampling period in the sampling stage, so that the SAR unit makes use of the time budget to convert the input signal into high resolution bit description.

ISBN 978-0-9928626-7-1 © EURASIP 2017

Thus, the more consecutive zeros (*0*s) allow more precise signal description with higher quantization depth. With the sampling and quantization strategies, every sampled signal can be translated into different precision which is illustrated in Fig.3.

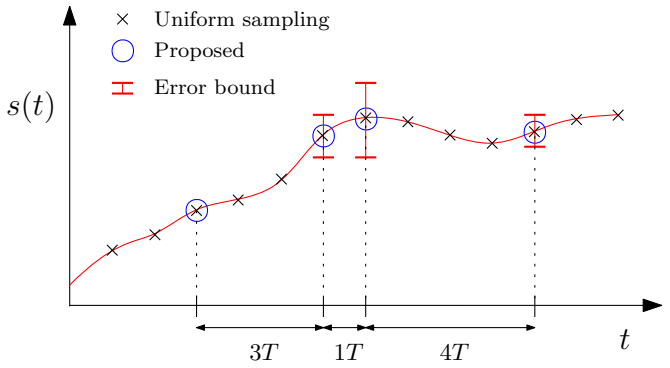


Fig. 3. Comparison of sampling schemes.

The sampling and quantization schemes require a more efficient recovery algorithm to fully facilitate the sampled signals with nonuniform quantization levels. Since each sample has different approximation error bound, one can set tighter constraints for the recovery. Leveraging the error bounds, we propose an element-wise constraint which is modeled as,

$$\begin{aligned} &\text{minimize} \quad ||\vec{\alpha}||_1 \\ &\text{subject to} \quad |y_r - (\mathbf{A}\vec{\alpha})_r| \leq \epsilon_r, \end{aligned} \tag{4}$$

where $r = 1, 2, \cdots, M$, and the subscript r means the r-th entry of the vector. The error vector is defined as $\vec{\epsilon} = [\epsilon_1, \epsilon_2, \cdots, \epsilon_M]^T$ which allows to bound the quantization error. For example, in uniform quantization, the error constraint vector becomes $\epsilon_1 = \epsilon_2 = \cdots = \epsilon_M$. The constraint will give the same result as the ℓ_2 norm case. For nonuniform quantization, ϵ_r depends on the instantaneous resolution, which produces more accurate results than the ℓ_2 norm constraint by solving the minimization problem with the element-wise constraint vector.

IV. Implementation Consideration

We describe more details about the proposed scheme considering real circuit implementation, and we present the simulation results in the following section.

The proposed scheme can be implemented in two ways: 1) using two or multiple ADCs (low and high resolution ADCs), and 2) single ADC with a high-bit DAC architecture. We choose the later approach due to its simpler architecture and lower power design capability.

A. Sampling sequences

We employ pseudo random sequences with the same probability for one and zero which imply $M = N/2$. Assume that the SAR ADC resolution is limited to 6-bit in high-speed sampling. With conventional uniform sampling and traditional CS quantization schemes, there are 50% samples quantized to be 6-bit. With nonuniform quantization scheme, the quantization depth will be varying depending on the following *1*s and *0*s. Pseudo random sequences are useful selection for the proposed SAR ADC architecture because the sequences can be pre-defined so that they can be stored in non-volatile memory space. The stored sequences are not only used in the sampling phase to determine the corresponding quantization depth, but also used to represent the compressively sampled signals in the recovery afterwords. We note here that the proposed scheme can be realized in a fully-passive CS framework. The proposed CS SAR ADC operates in discrete time rather than continuous time using switch capacitor circuit. In real implementation, the switch capacitor circuit in the SAR ADC operates with the pre-defined pseudo random sequences which allows to achieve high efficient sampling. Readers may refer to the architecture of switch capacitor based sample-and-hold architecture in [3], [4] which differs from previous random modulator based CS ADC architecture.

B. Quantization depth

The sampling period is equivalent to the symbol status of the measurement matrix Φ which can be easily implemented by turning on and off the switch capacitor circuit. The quantization happens whenever sampling takes place (*1* in the matrix Φ), but just the quantization bit depends on the number of following *0*s. Before sampling the next sample, the input signal is quantized in successive manner the same as the conventional SAR ADC does.

To be more specific, the proposed quantization scheme starts quantizing with b_0-bit and keeps increasing the quantization depth depending on the number of following zero. Defining quantization operator $\mathbb{Q}$, the input signal x_r is to be converted with quantization operator as $y(r) = \mathbb{Q}(\mathbf{\Phi}x_r, b_r)$, where $b_r = b_0 + step \times z_{c,r}$ bit (e.g., $b_0 = 6$ bit in our simulation), $step$ is the step increase in quantization bits, and $z_{c,r}$ is the zero count after the sampling instance of *1* at r. Since the sampling power is usually negligible in a SAR ADC design, the quantization phase takes far less time compared to a conventional SAR ADC which quantizes every sample with the maximum quantization bits meaning that the proposed architecture can save the total power consumption in sampling and quantizing phases, and allows savings in memory space after conversion.

In Table I, for the sake of better understanding, we list the expected resolution of the uniform randomly generated measurement matrix $\mathbf{\Phi}$ in the case of $step$ = 1, b_0 = 6-bit, and the maximum quantization bit b_{max} = 12-bit. We choose b_0 = 6-bit from the trend graph in Fig. 1 which shows that 6-bit precision is the maximum resolution at gigahertz sampling with SAR ADC architecture using the latest technology.

Ideally, the possible range of $step$ can be set $step = b_0$ in theory, but it would be better to keep it more conservative manner such as $step < b_0$ considering delay factors in circuit realization. Most of samples will be translated into low-bit depth (e.g., 6-8 bits), but quantization with high-bit depth will be taken by selecting lengthy random sequences.

ISBN 978-0-9928626-7-1 © EURASIP 2017

TABLE I
EXPECTED RESOLUTION

Sequence	Probability	Resolution
0	1/2	0
11	1/4	6
101	1/8	7
1001	1/16	8
⋮	⋮	⋮
1000001	1/128	11
10000001	1/256	$b_{max} = 12$
Other	1/256	$b_{max} = 12$

For simple presentation, without taking into account other additive noise, the error constraint is equivalent to the quantization noise which is not uniform. The recovery model of input signal is formulated in (4), and it makes use of the error constraint which bounds the representation error solving the minimization problem. The quantization noise is a function of bit depth, thus the error constraint can be replaced by $\epsilon_r = V_{ref}/2^{b_r}$.

C. Clock generator

In Fig. 4, we plot important clock timing which explains the relation among random sequence clock ($diag(\mathbf{\Phi})$), master clock(ϕ_{clk}), sampling clock (ϕ_{sample}), and conversion clock (ϕ_{conv}) cycles which operate the SAR ADC architecture appearing in Fig.2. Other system clocks are operated as the same as the conventional SAR ADC, but the conversion clock ϕ_{conv} is triggered with different operation scheme to produce extra quantization bits.

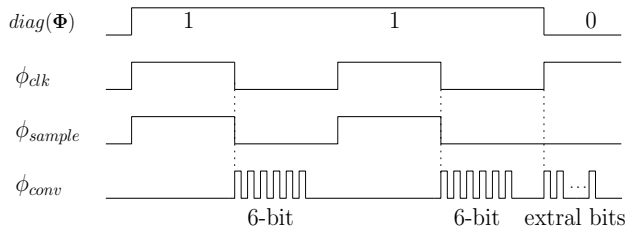


Fig. 4. Clocking for sampling and conversion.

If sampling time instance has *1* in the diagonal element of the measurement matrix $\mathbf{\Phi}$, ϕ_{sample} triggers ϕ_{conv} to start conversion. Next, the conversion clock ϕ_c is triggered simultaneously to operate the comparator for the SAR process, and the DAC starts producing digital bits of the sampled signal. If sampling time instance has *0* in the measurement matrix, the sampling clock ϕ_{sample} stays low and prepares the ADC for the next conversion after producing conversion outputs to the DAC. The delay between the end of conversion bit and the next conversion clock ϕ_{conv} is to reset and prepare for the next conversion. There is the main difference between the conventional SAR ADC and the proposed architecture in clocking ϕ_{conv}. Although the the sampling clock ϕ_{sample} stays low, ϕ_{conv} is being triggered to produce extra bit(s) if there is following zero(s).

The proposed architecture is flexible to operate in a Nyquist-rate SAR mode for non-sparse signals since the new architecture requires minimum modification of the conventional SAR ADC and mostly compatible to the architecture.

V. EXPERIMENTS

We compare the performance of the proposed sensing scheme to the uniform sampling (us) and traditional CS (cs) through MATLAB simulations. In the simulation, we explore the quantization scenarios with from 6 to 12-bit depth for the proposed quantization scheme. At each simulation case, 100 trials of simulation are performed and the results are averaged. Effective Number of Bits (ENOB) is investigated for comparison purpose and ENOB is defined as ENOB = (SQNR - 1.76) / 6.02 bit, where SQNR stands for signal-to-quantization noise ratio, the divisor 6.02 is for dB conversion, and the subtraction term of 1.76 is to compensate quantization error in an ideal ADC.

Fig. 5 shows the comparison of ENOB (y-axis) vs. quantization level of SAR ADC (x-axis) for input signals with $K \in \{3, 5, 10, 20\}$ sparseness in discrete Fourier transform domain. We compare ENOBs among uniform sampling (us), compressive sampling (cs), and the proposed quantization scheme with $step = \{1, 2, 4\}$. The x-axis indicates the quantization bit levels. For the result of us, it means ADC is uniformly quantized to this resolution. For cs, it performs random sampling with the fixed quantization levels. Lastly, the proposed scheme randomly samples the input with initial quantization bit from b_0 to b_r-bit which is described in Section IV-B. As displayed, in Fig. 5 (a) the proposed scheme outperforms us and cs schemes converting sparse input signals; (b) for inputs with low sparseness, the proposed scheme at least equivalent to us scheme; (c) for the least sparse signals in our simulation, us shows the best ENOB, but the proposed scheme works better than conventional cs based sampling results.

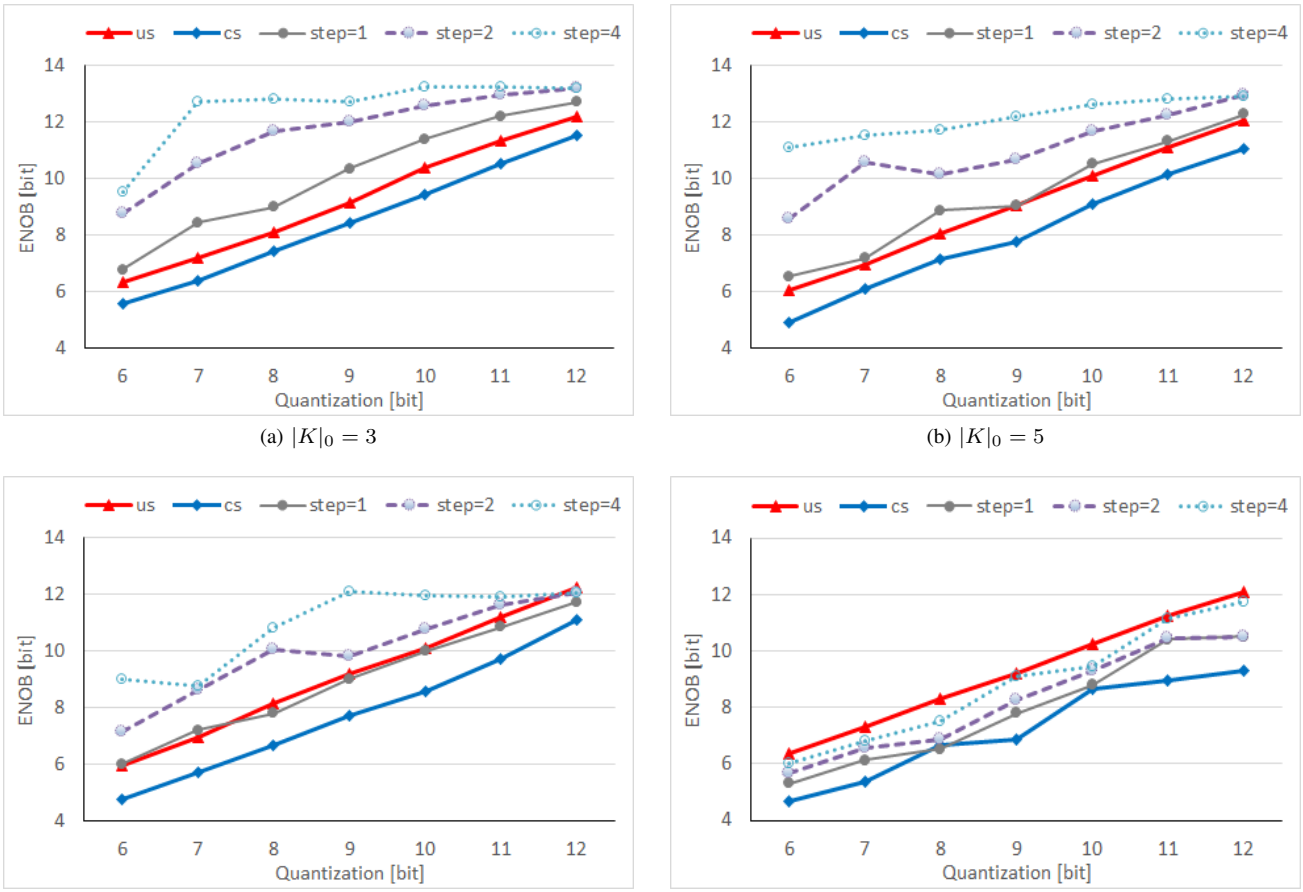
As a result, we can conclude that the performance of the proposed architecture is ruled by input signal sparseness with respect to the total number of measurements. The proposed system becomes more effective if the target input is sparser in the frequency domain. It is a expected result since the overall system is ruled by RIP condition in compressive sensing which is noted in Subsection II-B.

VI. CONCLUSION AND DISCUSSION

We propose a new sampling scheme, in this paper, combining a nonuniform quantization scheme with random sampling strategy. The proposed scheme is to increase the resolution of a SAR ADC making use of pre-defined pseudo random sequences which allow to acquire high frequency signal with above 6-bit resolution which is challenging with conventional SAR ADC architecture. With the nonuniform quantization scheme, the average quantization noise is decreased and it provides higher quality signal recovery than traditional compressive sensing based sampling schemes.

We also discuss details for circuit level implementation to realize the scheme with SAR ADC architecture. The per-

ISBN 978-0-9928626-7-1 © EURASIP 2017



(a) $|K|_0 = 3$ (b) $|K|_0 = 5$

(c) $|K|_0 = 10$ (d) $|K|_0 = 20$

Fig. 5. Simulation results.

formance of CS based sampling is mainly affected by the sparseness of signals and the proposed scheme as well. In addition, the noise folding effect from CS strategy alleviates the performance of CS based sampling schemes which is amplified by 3dB per octave of compression. Thus, the proposed architecture may not be efficient for every type of signal, but be ideal for bandlimited sparse signal acquisitions with moderate noise level. For applications, we are investigating around gigahertz sensing and communication signals which meet those requirements. In practice, there exist multiple noise sources which can contribute the total sum of noise level acquiring the signal of interest, e.g. quantization and thermal noise. The proposed scheme reduces the total quantization noise with respect to conventional CS sample strategy, but the effect of other noise sources needs to be investigated in the following study. Also, we plan to explore the power consumption of the proposed architecture which is not covered in this work. Nevertheless, the proposed nonuniform quantization strategy can be a possible method to overcome the sampling rate-and-precision limit which is a challenging problem in SAR ADC design even with the recent advanced technology.

ACKNOWLEDGMENT

We would like to express our gratitude to Prof. Rachel A. Ward, Department of Mathematics, and Prof. Nan Sun, Department of Electrical and Computer Engineering, for their guidance and suggestion while completing this work.

REFERENCES

[1] B. Murmann, "Adc performance survey 1997-2016," [Online]. Available: http://web.stanford.edu/~murmann/adcsurvey.html.

[2] Y. C. Kim, A. H. Tewfik, and B. V. Gowreesunker, "Multi-channel analog-to-digital conversion using a single-channel quantizer," in *Proc. of the 20th EUSIPCO*, Aug. 2012, pp. 1044 – 1048.

[3] Y. C. Kim, W. Guo, B. V. Gowreesunker, N. Sun, and A. H. Tewfik, "Multi-channel sparse data conversion with a single analog-to-digital converter," *IEEE J. Emerg. Sel. Topic Circuits Syst*, vol. 2, no. 3, pp. 470–481, Sept. 2012.

[4] W. Guo, Y. C. Kim, A. H. Tewfik, and N. Sun, "Ultra-low power multi-channel data conversion with a single sar adc for mobile sensing applications," in *Custom Integrated Circuit Conference (CICC)*, Sept. 2015, pp. 1–4.

[5] P. T. Boufounos and R. G. Baraniuk, "1-bit compressive sensing," in *in 42nd Annual Conference on Information Sciences and Systems*, 2008, pp. 16–21.

[6] J. N. Laska and R. G. Baraniuk, "Regime change: Bit-depth versus measurement-rate in compressive sensing," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3496 – 3505, July 2012.

[7] C. Luo and J. H. McClellan, "Compressive sampling with a successive approximation adc architecture," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 22-27 May 2011.

[8] M. Rudelson and R. Vershynin, "On sparse reconstruction from fourier and gaussian measurements," *Comm. Pure Appl. Math.*, vol. 8, no. 8, pp. 1025–1045, Aug. 2008.