DESIGN AND IMPLEMENTATION OF A FULLY INTEGRATED COMPRESSED-SENSING SIGNAL ACQUISITION SYSTEM

Juhwan Yoo[1]*, Stephen Becker*[2]*, Manuel Monge*[1]*, Matthew Loh*[1]*, Emmanuel Candès*[3]*, and Azita Emami-Neyestanak*[1]

[1]Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125
[2]Department of Applied and Computational Mathematics, California Institute of Technology, Pasadena, CA 91125
[3]Department of Statistics, Stanford University, Stanford, CA 94305

ABSTRACT

Compressed sensing (CS) is a topic of tremendous interest because it provides theoretical guarantees and computationally tractable algorithms to fully recover signals sampled at a rate close to its information content. This paper presents the design of the first physically realized fully-integrated CS based Analog-to-Information (A2I) pre-processor known as the Random-Modulation Pre-Integrator (RMPI) [1]. The RMPI achieves 2GHz bandwidth while digitizing samples at a rate 12.5x lower than the Nyquist rate. The success of this implementation is due to a coherent theory/algorithm/hardware co-design approach. This paper addresses key aspects of the design, presents simulation and *hardware measurements*, and discusses limiting factors in performance.

Index Terms— Compressed-Sensing, Random Modulation Pre-Integration, Analog-to-Information, CMOS

Dedicated to the memory of Dennis M. Healy Jr.

1. INTRODUCTION

1.1. Motivation

To date, the design of most information acquisition systems is based on the requirements of the Shannon-Nyquist sampling theorem. The principal difficulty in continuing to meet the enduring demand for higher bandwidth systems is realizing the required analog-to-digital converter (ADC).

Several surveys [2, 3] of ADC technology note that while power efficiency continues to improve rapidly as a result of technology scaling, increases in speed-resolution product do not. Based on empirical evidence, it has been suggested that aggressively pushing the sampling speed of ADCs in a given process technology sacrifices enormous power efficiency [2]. Indeed, while converters with impressive speeds have been reported [2], their required power dissipation and relatively low resolution have rendered them unsuitable in many applications.

Given the slowing of process technology scaling and the historically large disparity between rates of converter and digital system performance improvements, it appears that an increase in system bandwidths and resolutions will have to be achieved by means beyond advances in ADC technology.

This paper presents one such advance, a successful realization of a high bandwidth system that digitizes samples at a sub-Nyquist rate. Specifically, we present the results of feasibility studies as well as the design strategies used in the development of our system. The architecture of the prototype chip (implemented in 90nm CMOS) is named the Random-Modulation Pre-Integrator (RMPI), first introduced in [4, 5]. The RMPI achieves 2GHz of bandwidth while digitizing samples at an aggregate rate of just 320MSPS. The paper focuses on solutions that were used to relax the ideal models of previous paper designs [6] into implementable blocks which are robust to physical non-idealities, thus completing the journey from theory to practice. This is a challenging task as evidenced by the recent multitude of sub-Nyquist proposals yet the paucity of physical implementations. As a capstone to the program, the paper concludes with results from our physical implementation.

1.2. Compressed sensing background

Compressed sensing (CS) theory [7,8] states that a signal can be sampled non-adaptively and without information loss at a rate close to its information content. Recent advances in CS has provided alternative means of realizing systems with increased bandwidth and resolution without utilizing superior ADC.

The theory is complete in that it provides stable recovery guarantees in the presence of noise and corruption. In the context of RF systems, CS asserts that all signals with spectral occupancy $< B$ can be acquired without information loss by non-adaptive random sampling in an incoherent domain at a rate proportional to the occupied bandwidth[1]. This is remarkable since the bands locations need not be known.

After its introduction, CS spurred much work aimed at implementing *s*ub-Nyquist signal acquisition systems. These efforts have resulted in CS-based architectures such as the random demodulator (RD) [6], the modulated-wideband converter (MWC) [9], and others [10, 11]; see [12] for a more complete review. The present work does not compare these various architectures, but rather discusses the issues encountered while implementing the RMPI in CMOS technology.

The proposed RMPI design is a 'universal' encoder which,

[1]The frequency domain is not crucial to the theory; consider a signal that can be represented as a superposition of a few elements taken from a structured basis. CS theory says that the number of required measurements is proportional to the number of terms in the sparse expansion.

978-1-4673-0046-9/12/$26.00 ©2012 IEEE

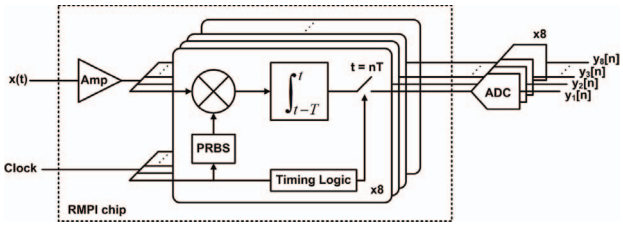


Fig. 1: Block Diagram of the RMPI

unlike other architectures, works with signals that are sparse in *any* fixed domain. However, for the purpose of concreteness, this paper specializes the signal domain to trains of short radar pulses embedded in an ultra-wideband.

It is also important to note that while CS systems implement a novel *abstract* form of sampling, it is misleading to classify them as ADC. The primary function of an ADC is to *digitize physical* voltage levels, which represent desired information, a function that all current A2I systems perform using standard ADCs. Most proposed CS samplers bear a much stronger resemblance to conventional RF/base-band architectures and are more accurately classified as analog pre-processors.

2. RMPI ARCHITECTURE

The RMPI is a variant of the RD architecture that uses parallel channels. It encodes compressed samples by modulating the input signal with a PRBS (one per channel), integrating the output of the modulator, then sampling the output with a low-rate ADC. The mixing with the PRBS needs to be at the Nyquist rate $f_s = 4$GHz; our implementation mixes at precisely the Nyquist rate. The insight behind the RMPI is that it is much easier to accurately *mix* a signal at 4GHz than it is to accurately *sample* it at 4GHz.

Our RMPI consists of 8 parallel correlator channels that share a common input; see Fig. 1. Each channel is implemented as a modified direct down-conversion receiver with the oscillator replaced by a PRBS generator. PRBSs are used for correlation because they are incoherent with any structured basis [13]. The reconstruction process recovers fixed time windows $[0, T]$, where T is chosen as large as possible while still allowing quick computation; a typical value of T is $\approx 2\mu s$. If $x[n] = \int_{n\Delta t}^{(n+1)\Delta t} x(t)$ is a discretization of the input, where $\Delta t = 1/f_s$, then for a fixed time period $T = N\Delta t$ the system records $M \ll N$ digital measurements $y = \Phi x$. The measurement matrix $\Phi \in \mathbb{R}^{M\times N}$ is block-diagonal, and the nonzero entries are ± 1, distributed as signed Bernoulli variables.

Signal recovery consists of finding a solution $x \in \mathbb{R}^N$ to the under-determined linear system $y = \Phi x$. Any x solving the equation is called 'feasible', and in general there are infinitely many solutions. For many classes of matrices Φ, CS shows that if $y = \Phi x$ and x is sufficiently sparse, then x may be recovered by searching for the feasible solution which has minimum ℓ_1 norm; see further discussion in §4.

3. DESIGN STRATEGY AND IMPLEMENTATION

The mathematical analysis of the RD [6] appropriately assumed ideal blocks, measurements corrupted only by AWGN, and a finite-dimensional space of input signals. The relaxation of this model into a working prototype is by no means simple, since the design of each block involved tradeoffs that had heretofore not been explored. This section discusses those tradeoffs and the reasoning behind them.

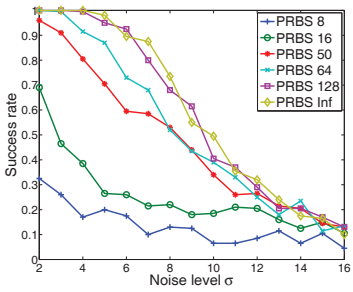


Fig. 2: Success rate of frequency estimation across noise, for various PRBS repetition lengths, via simulation. The noise level is unphysical, and for comparison purposes only. A PRBS of length 128 or higher is most robust to noise. A "success" is defined as estimation error of less than 5MHz (each data point is the outcome of 200 independent trials)

3.1. Input Amplifiers

Two cascaded amplifiers precede the random demodulator. The first is a single high gain LNA which reduces the system noise figure. The second is a set of transconductance amplifiers, one per channel, which reduce cross-talk among channels. The cascade of amplifiers introduces a transfer function which, ideally, would have linear phase/constant magnitude response in the system bandwidth. While potentially realizable, designing amplifiers with a near "ideal" response would require excessive power consumption. Since power trades directly with gain-bandwidth product, the 3dB cutoff of the cascade was placed at the edge of the system bandwidth.

3.2. Random modulation

The random modulation is implemented with a passive CMOS-switch based mixer and programmable shift register[2] that contains a PRBS. The primary consideration in choosing the actual bit sequence is to minimize gain variation, which means that two input signals of the same energy should generate measurements of roughly the same energy. A completely random sequence is ideal, but any realizable bit sequence must have a finite period after which it repeats, and shorter periods consume less power.

Lacking a quantitative theory, we explore this tradeoff via different types of numerical simulation. For example, Fig. 2 shows frequency estimation error at different noise levels. Each line series is a model with the PRBS repetition indicated in the legend, varying from 8 to ∞ (∞ means the PRBS does not repeat with $[0, T]$). The inputs are pure tones with uniformly random phase and uniformly random frequency in the interval $[0, 2]$GHz, so recovery is possible via digital matched filtering [12]. The implemented RMPI uses a sequence length of 128.

[2]The use of a programmable shift-register is strictly for testing purposes and could be replaced with an appropriate combination of much lower power consumption LFSRs.

Although not evident from the figure, the necessary PRBS length is actually tightly coupled with the gain-bandwidth product of the integrator (Fig. 2 uses fixed integrator parameters); if the integrator gain-bandwidth product is small, the PRBS length must be longer. For intuition, consider a pure tone input. The PRBS is periodic, so its spectrum has distinct harmonics, and these harmonics convolve with the pure tone to create shifted versions of the harmonics. The integration block attenuates high frequencies, so the shifted harmonic nearest DC (equivalently, the nearest PRBS harmonic to the input tone) will dominate.

If the PRBS sequence is short and hence the harmonics are widely spread, this effect is pronounced, and some input frequencies have very large output (hence high SNR), while others have small output. The effect is reduced by either decreasing the distance between harmonics (a long PRBS sequence), or increasing the integrator bandwidth so that more than just the dominant harmonic passes through.

3.3. Integration and low-rate sampling

The memory provided by the integrator is crucial for sampling short pulses. A realizable filter has a dominant pole near DC as well as several non-dominant poles. In our system, the dominant pole is placed at 300kHz for an integration window of 25ns. A shorter time constant reduces the memory effect and decreases the ability of the system to detect short pulses. The second pole of the system is at 300MHz, and as suggested in the previous section, must be great enough to allow several PRBS harmonics to fit inside the bandwidth.

3.4. Considerations in parallelization

For a fixed overall sampling rate, it is possible to decrease the ADC sampling rate of each channel by increasing the number of channels. From a purely theoretical standpoint, more channels is better, since the matrix representation of the RMPI Φ better approximates a signed Bernoulli matrix as opposed to a block-diagonal signed Bernoulli matrix. Intuitively, there is more randomization involved, and less chance that a signal passes through the sieve. Fig. 3 shows numerical experiments supporting the superiority of multi-channel designs.

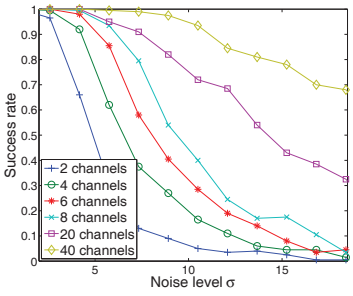


Fig. 3: Same setup as Fig. 2. The overall data output rate is fixed, and the number of channels is varied (as the number of channels increases, the sampling rate of each channel decreases). Designs with more channels are more robust to noise.

However, there are numerous reasons to use fewer channels. There are spatial constraints imposed by the IC and difficulties involved with synchronization. Additional channels require more power, and make calibration more complicated. Since the sampling rate is proportionally lower, the integration time is longer, and the dominant pole of the integrator must be made smaller in order to increase the time-constant. Given these considerations (see [12]), and other pragmatic testing details, the final design uses 8 channels.

Even with 8 channels, timing differences can be significant, and differences in time and phase delays hurt system performance (although this can be compensated for by calibration). In order to minimize the timing differences, minimize jitter at the mixer, and minimize power consumption, the system clock is distributed in a binary symmetric tree topology and converted to CMOS levels at the place closest to the PRBS generators. This minimizes duty-cycle distortion and clock jitter.

3.5. Accounting for non-idealities

As in a conventional receiver, the system is affected by thermal noise and jitter, and a limited range of linear response, and our techniques to deal with these effects are standard. In addition, there is preliminary work that non-linear response may be partially accounted for; see [12].

One of the benefits of the RMPI is that it is possible to calibrate the system and learn the transfer function. This is done by sending in a basis of known inputs and measuring the outputs. On simulation results, this shows enormous promise, and hardware calibration results are ongoing.

4. RECOVERY ALGORITHMS

CS theory suggests basis pursuit to recover the signal, but there are many variations. We refer the reader to [21] for a review, and to [12] for techniques used specifically for RMPI samples. Table 1 shows a representative sample of state-of-the-art solvers, and their performance on a sample problem generated from the RMPI simulations. The table plots relative reconstruction error $\|x - \hat{x}\|_2/\|x\|_2$ where $\hat{x}$ is the estimate produced by the solver. The input x was a radar pulse, and measurements were recorded from the Simulink model using realistic values for the non-ideal blocks and noise sources. Reconstruction requires knowledge of some sparsifying dictionary

	Reconstruction error		
Solver	8× Gabor	32× Gabor	DCT
OMP [14]	1.3e-1	1.6e-1	9.9e-2
OMP (SPAMS [15])	1.0e-1	6.2e-2	8.3e-2
CoSaMP [16]	1.5e+1	DNC	1.5e-1
CoSaMP (modified)	1.3e-1	9.2e-2	2.0e-1
ℓ_1 synthesis [17]	2.9e-1	3.4e-1	2.4e-1
" " with reweighting	1.3e-1	2.1e-1	2.0e-1
ℓ_1 analysis [17]	1.1e-1	1.4e-1	2.8e-1
" " with reweighting	6.6e-2	5.4e-2	2.1e-1
LARS (SPAMS [15])	1.3e-1	9.5e-2	1.7e-1
ALPS [18]	1.4e-1	1.1e-1	2.3e-1
SL0 [19]	2.6e-1	4.2e-1	1.2e-1
AMP [20]	DNC	DNC	DNC

Table 1: State-of-the-art solvers on a realistic sparse recovery problem. The problem uses realistic measurements, and the signal is compressible but not exactly sparse. When the algorithm diverged or failed to converge, we report DNC.

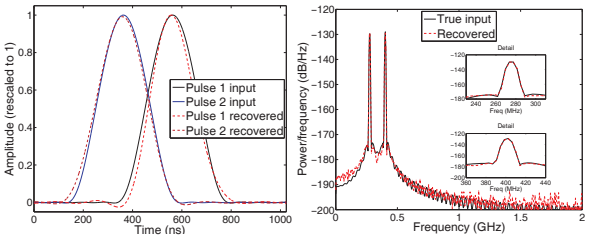


Fig. 4: Pulse-on-pulse recovery. Two superimposed pulses, one of 275MHz and one of 401MHz, are recovered from hardware data. The carrier frequency of both pulses is estimated to within .234MHz.

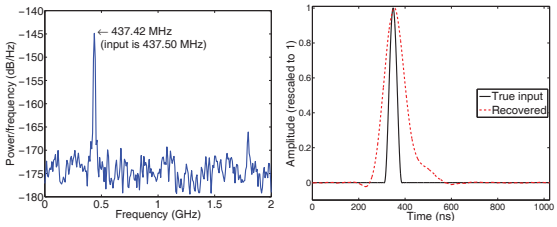
Ψ, and the three columns of numbers repesent three choices of Ψ: an $8\times$ oversampled multi-scale Gabor time-frequency dictionary, a $32\times$ oversampled Gabor dictionary, and a discrete cosine basis. The discretized version x of the input was of size $N = 2048$, and all solvers took on the order of 1 minute or less.

The results suggest that many solvers do quite well and achieve less than 10% relative error. It also appears that the reweighted techniques used in [12] are among the best. Since this was a single trial, and algorithm parameters need to be tailored for any application, one should not view the results as a comparison. Rather, these give an idea that there are many alternatives to basis pursuit (basis pursuit is equivalent to ℓ_1 synthesis in the table) that display competitive performance.

5. EMPIRICAL RESULTS AND DISCUSSION

Reconstructions (from hardware measurements) of radar-pulses with carrier frequencies from 100MHz-2GHz have been obtained [1]. Some typical reconstructions of interesting scenarios are shown in Figs. 4 and 5. The test inputs were generated by an arbitrary waveform generator and the reconstructions were performed using a variant of basis pursuit with reweighting (see [12]). Figure 4 shows the reconstruction of two 450ns pulses that overlap in time (pulse-on-pulse) with center-to-center time separation of 200ns and carrier frequencies 275MHz and 401MHz. This is a challenging reconstruction for even Nyquist-rate systems. Fig. 5 shows limits of the system; Fig. 5(a) is a reconstruction of a small tone 54 dB below the full-scale amplitude, thus the system dynamic range is $\geq$ 54dB. Fig. 5(b) shows the reconstruction of a 75ns pulse which for our integration window time of 25ns represents only 24 samples worth of data. It is possible to recover a 50ns pulse with only 1.4MHz frequency estimation error, although the pulse window is not recovered well.

Future efforts will be focused on hardware calibration and improved signal recovery models.



(a) (b)

Fig. 5: (a) Recovery of a tone with amplitude $400\mu V_{pp}$, which is -53.98dB below the full-scale amplitude. The frequency estimation is accurate to within 0.08MHz. 512ns worth of samples were used. (b) Recovery of a 75ns pulse. Frequency estimation accurate to 1.02MHz.

6. ACKNOWLEDGEMENTS

This work was funded under DARPA grant FA8650-08-C-7853; in particular, the authors are indebted to Dr. Denis Healy for his foresight and encouragement. Justin Romberg, Michael Wakin, and Michael Grant also contributed in many ways.

7. REFERENCES

[1] J. Yoo, S. Becker, M. Loh, M. Monge, E. Candès, and A. Emami-Neyestanak, "A 100MHz-2GHz 12.5x sub-Nyquist rate compressed-sensing signal acquisition system," ISSCC, *submitted*, 2012.

[2] B. Murmann, "Trends in low-power, digitally assisted A/D conversion," *IEICE Trans. Electron.*, vol. E93-C, no. 6, pp. 718–727, 2010.

[3] H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Area Comm.*, vol. 17, no. 4, pp. 539–550, 1999.

[4] W. Skones, B. Oyama, S. Stearns, J. Romberg, and E. Candès, "Analog to information (A-to-I), technical and management proposal," August 2005, (in response to DARPA BAA 05-35).

[5] S. Kirolos, J. N. Laska, M. Wakin, M. F. Duarte, D. Baron, T. Ragheb, Y. Massoud, and R. G. Baraniuk, "Analog-to-information conversion via random demodulation," in *Proc. IEEE DCAS*, Sept. 2006.

[6] J. A. Tropp, J. N. Laska, M. F. Duarte, J. Romberg, and R. G. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Inform. Theory*, vol. 56, no. 1, pp. 520–544, 2010.

[7] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[8] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[9] M. Mishali, Y. C. Eldar, O. Dounaevsky, and E. Shoshan, "Xampling: Analog to digital at sub-Nyquist rates," *IET Cir. Dev. and Systems*, vol. 5, no. 1, pp. 8–20, Jan. 2011.

[10] J. Romberg, "Compressive sensing by random convolution," *SIAM J. Imag. Sci.*, vol. 2, no. 4, pp. 1098–1128, 2009.

[11] J. P. Slavinsky, J. N. Laska, M. A. Davenport, and R. G. Baraniuk, "The compressive multiplexer for multi-channel compressive sensing," in *IEEE Intl. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Prague, may 2011, pp. 3980–3983.

[12] S. Becker, *Practical Compressed Sensing: modern data acquisition and signal processing*, Ph.D. thesis, California Institute of Technology, 2011.

[13] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[14] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Tran. Info. Theory*, vol. 53, no. 12, 2007.

[15] J. Mairal, *SPAMS: SPArse Modeling Software*, WILLOW, INRIA, 2.1 edition, June 2011.

[16] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal*, vol. 26, pp. 301–321, 2009.

[17] S. Becker, E. J. Candès, and M. Grant, "Templates for convex cone problems with applications to sparse signal recovery," vol. 3, no. 3, 2011.

[18] V. Cevher, "An ALPS view of sparse recovery," in *IEEE Intl. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Prague, may 2011, pp. 5808–5811.

[19] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed l0 norm," *IEEE Trans. Sig. Proc.*, vol. 57, no. 1, pp. 289–301, jan 2009.

[20] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18914–18919, 2009.

[21] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, 2010.