Compressed Sensing ICs for Bio-Sensor Applications

Kevin Park

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2014-102
http://www.eecs.berkeley.edu/Pubs/TechRpts/2014/EECS-2014-102.html

May 16, 2014

University of California, Berkeley College of Engineering

MASTER OF ENGINEERING - SPRING 2014

Electrical Engineering and Computer Science

Integrated Circuits

Compressed Sensing ICs for Bio-Sensor Applications

Kevin Park

This **Masters Project Paper** fulfills the Master of Engineering degree requirement.

Approved by:

1. Capstone Project Advisor:

Signature: _____ Date _____

Print Name/Department: David J. Allstot / EECS

2. Faculty Committee Member #2:

Signature: _____ Date _____

Print Name/Department: Jan M. Rabaey / EECS

Compressed Sensing ICs for Bio-Sensor Applications

by

Kevin Park

A project report submitted in partial satisfaction of the
requirements for the degree of
Master of Engineering

in

Electrical Engineering and Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David Allstot
Professor Jan Rabaey

Spring 2014

Compressed Sensing ICs for Bio-Sensor Applications

by

Kevin Park

Abstract

Compressed Sensing ICs for Bio-Sensor Applications

by

Kevin Park

Master of Engineering in Electrical Engineering and Computer Science

University of California, Berkeley

Professor David Allstot

In recent years, interest towards wearable sensors and medical monitoring has exploded. Despite efforts to improve health care with medical technology, many people still die today from the lack of proper health management. Bio-sensors can provide long-term, continuous monitoring of key bio-signals and enable proactive personal health management [1]. Since the bio-sensors are worn in some manner, ultra-low power consumption is critical. Conventional sensors use Nyquist sampling to capture key information in bio-signals; hence, the operation is very costly. Compressed sensing exploits data redundancy arising from a low rate of significant events in a sparse signal. The data is compressed at the sensor node and therefore, the power required to transmit the compressed information to a receiver (e.g., smartphone) is minimized. The compression factor of a sparse bio-signal can be much larger than one, thus providing significant power efficiency in wireless sensors.

[2] and [3] assume that the bio-signals have constant rate of information with minimal

noise. This report presents a compressed sensing front-end for bio-signal applications that dynamically adjusts the compression factor, and increases the accuracy of reconstruction by removing noise with spike detection. This report includes the design and simulation of a low power differential ADC and the analysis of the proposed compressed sensing DSP with a 32 nm CMOS process. The DSP is estimated to consume 0.525 μW at $V_{DD} = 0.4\,\text{V}$. The ADC uses the SAR architecture with a hybrid capacitive array with floating voltage shield, a Strongarm comparator, and digital SAR logic. The ADC consumes 9.8 nW at $V_{DD} = 1.05\,\text{V}$ and has 9.7-bits of resolution running at 20 kHz. Significant power savings are possible with compressed sensing and its influence will be key in driving the revolution of wearable sensors and medical monitoring.

Contents

List of Figures

List of Tables

1. Introduction

In recent years, interest towards sensors for scientific research, surveillance, medical applications, and imaging has exploded. In particular, wearable sensors and medical monitoring are two promising research areas that have been getting a great deal of attention. The nature of their applications imposes a highly stringent power requirement on the technology. Since sensors will be worn in some manner, ultra-low power consumption is critical. With compressed sensing, bio-sensors can save a significant amount of power, and therefore have increased lifetimes.

1.1 Bio-sensors

According to the World Health Organization, almost 20 million people will die from various heart diseases in 2015 [4]. With wearable sensors and proper medical monitoring systems, many of these deaths can be avoided [4]. Wearable/Implantable network of bio-sensors can provide long-term, comprehensive monitoring and evaluation of one's body at all times, thereby increasing the awareness and facilitating distribution of critical health conditions appropriately.

The global bio-sensor market is predicted be a USD $18.9 billion market by 2018 [5]. With the recent advances in technology, wireless and wearable sensors are rapidly being integrated into our lives. However, many challenges have prevented universal usage of bio-sensors. One of the major challenges is power consumption, as bio-sensors often have a very small form factor. A wireless bio-sensor system consists of bio-sensor nodes that transmit information to

an aggregator (e.g., smartphone) that receives and processes data. Compressed sensing is a new signal processing paradigm that is applicable for applications where sparsity is evident. Most bio-signals are sparse in many domains with low bandwidth [2]. A compressed sensing based architecture exploits data redundancy arising from a low rate of significant events in a sparse signal to achieve ultra-low power operation, while capturing important information with high probability.

1.2 Existing Work

The Shannon-Nyquist theorem suggests that to recover the original signal accurately, sensors have to acquire and transmit a lot of data. This is a problem for bio-sensors as their power requirements are very stringent. The ultimate goal is to reduce the data-rate while preserving important information in the signal. A great deal of research has been done in this area and many solutions use some form of data compression or filtering.

Since key bio-signals have a low rate of information, data transmission can be done more efficiently. Karkare developed an energy efficient neural sensing system that only sends data when significant events are detected using signal conditioning [6]. It is a low-power DSP chip that performs spike detection and spike sorting on neural recordings, which is implemented in a 90-nm CMOS process. It dissipates 2μW of power per channel and has a data-rate reduction of 91%.

A few compressed sensing chips have been presented for bio-signal applications. A digital implementation of a compressed sensing chip has been developed by Chen with a Nyquist-rate ADC and 50 Multiply & Accumulators (MACs), which consumes 1.9 uW at 0.6 V and 20 kS/s [3]. The chip uses MACs to multiply the input signal with a random Bernoulli matrix to

apply the compressed sensing operation. Gangopadhyay presented an analog implementation of a universal compressed sensing front-end in 0.13 μm CMOS with 384-bit Fibonacci/Galois hybrid linear feedback shift register and 64 digitally-selectable CS channels with a 6-bit C-2C MDAC/integrator and a 10-bit C-2C SAR ADC in each [2]. With a running frequency of 2 kHz, the total power dissipation is 28 nW and 1.8 μW for one and 64 channels, respectively.

1.3 Capstone Project

This report presents a compressed sensing front-end for bio-signal applications. This front-end dynamically adjusts the compression factor, and increases the accuracy of reconstruction by removing noise with spike detection.

1.3.1 Contributions

The M.Eng Compressed Sensing group is composed of Brian Wang, Daniel Oh, Xinping Zhang, and myself. My particular contributions include the design and analysis of:

1. A low power ADC for bio-signal applications
2. A compressed sensing DSP for bio-signal applications

2. Literature Review

2.1 Compressed Sensing

Traditional sensors operate based on the highest frequency of a continuous signal. A compressed sensing based architecture exploits sparse nature of bio-signals, where the sampled bio-signals are compressed before transmitting. The RF transmitter can operate at sub-Nyquist-rates, allowing the sensor to achieve low power consumption. Then, the reconstruction of compressed samples, which is nearly impossible to do on chip without a processor or an ASIC, can be done at the receiver. The ADC can also operate at sub-Nyquist-rates in an analog implementation, because the signal can be compressed before digitizing. However, in a digital implementation, the ADC needs to operate at Nyquist-rates if the signal is sparse in the time domain. If the signal is sparse in the frequency domain, the ADC can operate at sub-Nyquist-rates in both implementations.

2.1.1 Compressed Sensing Construction

Compressed sensing is defined by the following matrix operation, $[Y] = [\Phi][X]$. $[X]$ is the input signal of size $N \times 1$ and is multiplied by a random matrix $[\Phi]$ to produce a

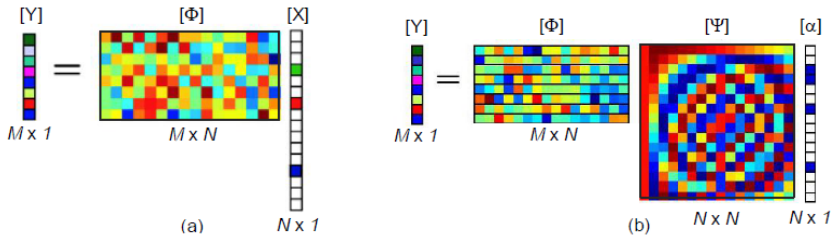


Figure 2.1: (a) Compressed Sensing Construction (b) Reconstruction

compressed vector [Y] of size $M \times 1$, where $M < N$ as shown in Fig. 2.1(a) [2][7][8][9]. Therefore, the compression factor is N/M. With more sparsity, the compressed sensing operation can be more aggressive towards the compression factor to save more power. To reconstruct the original signal $[\hat{X}]$, there needs to be at least M number of measurements, where $M \geq p \log(N/K) > 2K$, where K is the number of nonzero entries in N input samples, and p is a constant [2][10]. Sparse signals have a very small K/N ratio and M random measurements of a sparse signal can capture all the essential information accurately with high probability [11]. $[\Phi]$ is a matrix of random numbers (e.g., Bernoulli, Gaussian, Uniform, etc.), which can be generated using linear feedback shift registers. Ideally, white noise is multiplied with $[\Phi]$ and integrated so that the noise is averaged out and the contribution is very small [2].

2.1.2 Compressed Sensing Reconstruction

Any signal can be written as a weighted linear combination of some basis. Most bio-signals are sparse in many domains. Therefore, with proper basis, a sparse signal can be written as $[X] = [\Psi][\alpha]$, where $[\alpha]$ is a sparse representation of $[X]$ in the basis $[\Psi]$ [2]. For compressed sensing, [X] must be sparse in $[\Psi]$, and $[\Psi]$ and $[\Phi]$ must be incoherent [2]. For a signal that is sparse in the time domain, $[\Psi]$ is just an $N \times N$ identity matrix [2]. This means that the rows of $[\Phi]$ must be uncorrelated for accurate reconstruction. A compressed signal *cannot* be reconstructed from $[\hat{X}] = [\Phi]^{-1}[Y]$, because $[\Phi]$ is non-invertible since $M < N$. Alternatively, l_1 norm can be used to determine the sparsity of $[\alpha]$ and recover the original signal by solving $\min ||\alpha||_{l_1}$ subject to $[Y] = [\Phi][\Psi][\alpha]$, then the recovered signal is just $[\hat{X}] = [\Psi][\alpha]$ [7][8]. CS reconstruction can be done on the receiver that has a processor or an ASIC (e.g smartphone).

A simple thresholding (zeroing insignificant values) can make a noisy signal more sparse and increase the accuracy of the reconstruction [11].

2.1.3 Compressed Sensing Frond End (CS-FE)

The goal of the CS-FE is to compress the input signal by multiplying it with $[\Phi]$. This operation simply consists of column-by-column matrix multiplication of $[\Phi]$ with a sampled bio-signal $[X]$. For example, the first sample X_1 is multiplied by the first column of $[\Phi]$. The second sample X_2 is multiplied by the second column of $[\Phi]$ and so on. After taking N samples, the products are summed to produce $[Y]$ as shown in Equation (2.1.1).

$$[Y] = \begin{bmatrix} Y_1 = \Phi_{11}X_1 + \Phi_{12}X_2 + \cdots + \Phi_{1\mathrm{N}}X_{\mathrm{N}} \\ \vdots \\ Y_{\mathrm{M}} = \Phi_{\mathrm{M1}}X_1 + \Phi_{\mathrm{M2}}X_2 + \cdots + \Phi_{\mathrm{MN}}X_{\mathrm{N}} \end{bmatrix} \tag{2.1.1}$$

In an analog-FE, an MDAC and an integrator can be used to perform the matrix multiplication. There needs to be M MDACs and integrators to calculate $Y_1 \cdots Y_M$. In [2], M is digitally selectable from 1 - 64 based on the desired compression factor. In [2], the CS-AFE also includes 6-bit hybrid linear feedback shift registers (HLFSR) that generates a random sequence of coefficients to form $[\Phi]$ on chip. It generates the coefficients using Fibonacci-Galois hybrid linear feedback shift register, where Fibonacci and Galois LFSRs are interleaved to provide additional randomization. It is suggested that the 1-bit Bernoulli $[\Phi]$ is sufficient in most practical applications of compressed sensing [2]. With a digital-FE, the multipliers aren't needed if the coefficients of $[\Phi]$ are 1-bit, because the input can be either accumulated or discarded. A digital-FE would mostly consist of adders, multiplexers, and large banks of registers.

2.1.4 Spike Detection

Noise in bio-signals often reduces sparsity. Signal thresholding can effectively remove noise by allowing only the spikes to pass through. One way to do this is with spike detection with nonlinear energy operator (NEO). An input data undergoes signal conditioning, where the data is sent out only when the input exceeds a certain threshold, thereby reducing the data-rate significantly if the signal is sparse. The input signal is subjected to the following NEO operation, $\psi[x(n)] = x^2(n) - x(n+1) \cdot x(n-1)$, and $\psi[x(n)]$ is conditioned using a threshold, which can be calculated by $Thr = C\frac{1}{N}\sum_{n=1}^{N}\psi[x(n)]$, where C and N are empirically chosen constants [12]. In addition, the sparsity of bio-signals can be estimated in real-time with spike detection, which may be helpful in determining the appropriate compression factor for compressed sensing.

2.2 Specifications

The specifications for the compressed sensing DSP chip are given in Table 2.1. The specifications are derived based on neural signals, but it is designed to be a universal compressed

Table 2.1: Proposed Specifications for Compressed Sensing DSP

Parameter	**Description**	**Value**
Input Rate	Incoming rate of 10 bit data	20 kHz or 200 kb/s
Error	Reconstruction error	$< 1\%$
P_d	Power dissipation	**Minimize**

Table 2.2: Proposed Specifications for ADC

Parameter	**Description**	**Value**
ENOB	Effective Number of Bits	10 bits
f_s	Sampling Frequency	20 kHz
HD3/SFDR	Harmonic Distortion / Spurious Free Dynamic Range	$\sim$62 dB
P_d	Power Dissipation	**Minimize**

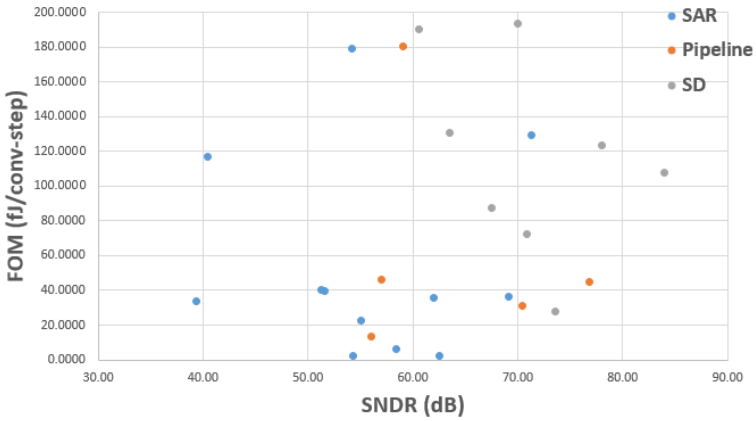


Figure 2.2: ADC Figure of Merit (2011 - 2013)

sensing DSP for any sparse bio-signals. The specifications for the ADC is given in Table 2.2.

2.3 ADC Architecture

Some key things to note in the ADC specifications are that the signal bandwidth is relatively low, the power must be extremely low (power is going to be a constraining factor), and 10 bits of ENOB corresponds to 11-12 physical bits, which may be difficult to achieve for some architectures.

2.3.1 Choice of ADC Architecture

Fig. 2.2 compares the ADCs presented in ISSCC in the last couple of years (2011 - 2013) [13]. Although high SNDR converters are dominated by sigma delta ADCs, they are mostly clustered above 50 fJ/conv-step. A flash ADC is fast, but uses a lot of power as it requires $2^B - 1$ comparators. It is also difficult to achieve high accuracy with the flash ADC. The main advantage of a pipeline ADC is speed, which is not necessary for this application. The figure of merit in Fig. 2.2 suggests that pipeline ADCs can be fairly energy efficient, but not as efficient as SAR ADCs. There are many SAR ADCs that perform below 50 fJ/conv-step, achieving up to 2.2 fJ/conv-step [14]. To put this into perspective, a 10-bit ADC with

$f_s = 20\,\text{kHz}$ that has 2.2 fJ/conv-step consumes about 45 nW ($P = \text{FOM} \cdot 2^{\text{ENOB}} \cdot f_s$). With low speed, low power, and high accuracy in mind, the SAR ADC architecture was chosen to be the most promising candidate for bio-sensor applications.

2.3.2 SAR ADC

A SAR ADC uses a binary search algorithm over the DAC output for analog-to-digital conversion. High resolution is possible with this architecture, but this requires the comparator and the capacitive DAC to be highly accurate. A disadvantage of this architecture is that the ADC requires N number of clock cycles to convert N bits, because it needs to evaluate each bit individually. The SAR ADC would need to run at $12f_s = 240\,\text{kHz}$ to produce an output at f_s.

Without any calibration techniques, ENOB tends to be 1$\sim$2 bits less than the physical bits. To achieve a 10-bit ENOB, a 12-bit SAR ADC can be implemented to budget for parasitics, mismatches and non-linearities. In a SAR ADC, most of the power is consumed in the capacitive DAC and the digital logic. One conventional way to design the DAC is to use a binary-weighted DAC, which is inherently monotonic and more immune to parasitic capacitors. However, the binary-weighted DAC requires $2^n - 1$ unit elements. Therefore, the input capacitance of a binary DAC is very high, $C_{in} = 2^n C_u$. Instead, a C2C DAC can be used to reduce the number of capacitors. The input capacitance of a C2C DAC is $C_{in} = 2C_u$ and the total capacitance is approximately $3nC_u$ [15].

3. Methodology

This section describes the architecture of the proposed compressed sensing DSP as well as the design of a low-power ADC that interfaces with the DSP. The ADC and the DSP were designed with the 32-nm PDK from Synopsys.

3.1 Compressed Sensing DSP

Noise can often reduce the sparsity of a signal and impair the accuracy of reconstruction. Therefore, in this architecture, compressed sensing is done after thresholding. Spike detection with NEO is used in this architecture. For the first 1024 samples, the spike detection block calculates the threshold. When an input sample exceeds this threshold (when a spike is detected), subsequent 40 samples are stored in the memory. 50 output samples are prepared for every neural spike and the output is aligned so that the 22^{th} sample has the maximum derivative. A preamble buffer of size 22 is used, in case the maximum derivative occurs within the first 22 samples from the detection. The details of this spike detection architecture can be found in [6].

The compressed sensing block generates M by N random numbers by interleaving an M-

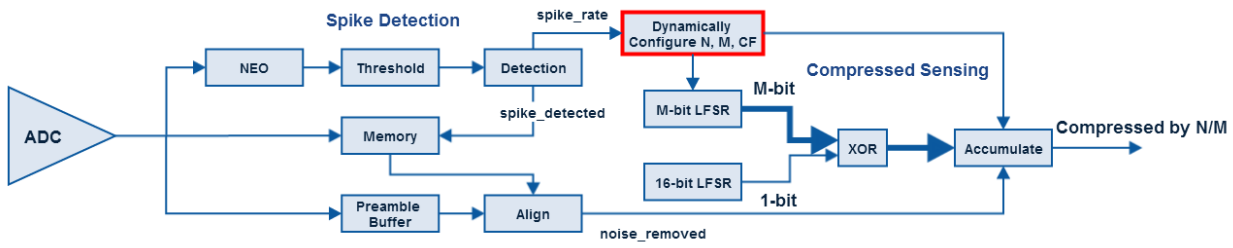


Figure 3.1: Proposed Compressed Sensing DSP

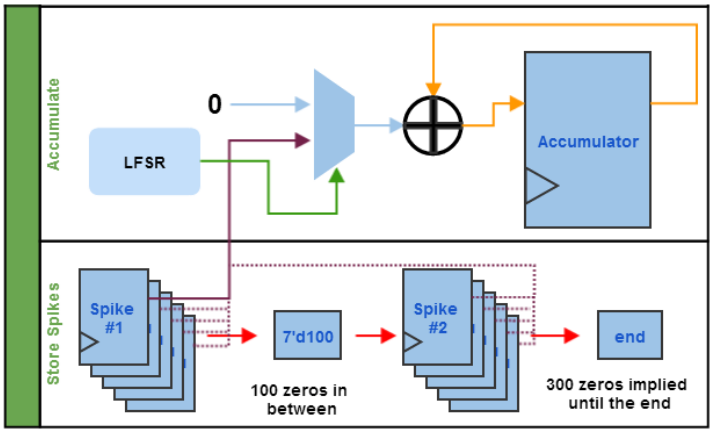


Figure 3.2: Block Diagram of a Single Accumulator

bit LFSR with another independent 16-bit LFSR as shown in Fig. A.1 [3]. Each entry in $[\Phi]$ is either 1 or 0, so multipliers are not needed since the input can be either accumulated (in case of 1) or discarded (in case of 0). Since power is directly related to the choice of the compression factor, CF, and hence the size of N and M, they should be sized optimally based on the sparsity of the input signal. Fig. 3.1 shows the proposed block diagram of the compressed sensing DSP, where N, M, and CF can be dynamically adjusted based on the sparsity calculated by the spike detection block. Dynamically adjusting the CF not only improves robustness of the system, it also gives the ability to power down resources that are not used, thus saving power.

Since the CF is dynamically adjusted, the receiver needs to know the CF that is associated with each N input. This information can be tagged on with each output. Furthermore, a memory of size N is required, because the spike detection block has to calculate the sparsity for every N input before configuring the CF. This memory buffer costs an overhead of $N \cdot 10\text{-bits} = 10N$ registers. This is very costly because most of the data stored in these registers will be zeros. Instead, only the spikes and the number of zeros in between spikes are

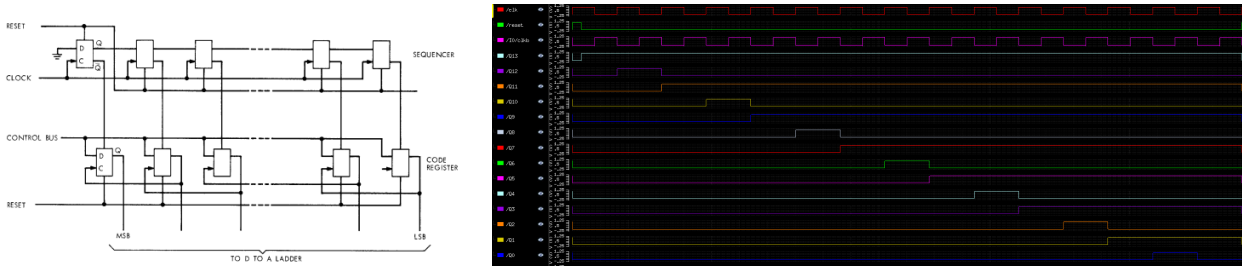
stored, which only requires about $10N \cdot (1-s)$ registers, where s is the sparsity. Fig. 3.2 shows a simplified diagram of a single accumulator. There would be M copies of this accumulator. The figure shows that there are 2 spikes detected in N and there are 100 zeros in between the two spikes. If the spikes are 50 samples each and $N = 500$, the rest of the samples after the second spike are implied to be zeros (300 zeros). In this case, only $50 \cdot 10\text{-bits} + 50 \cdot 10\text{-bits} + 7$ bits $= 1007$ registers are used to store the entire N data as opposed to 5000 registers. Low power techniques such as parallelism, pipeline, multi-vdd, multi-Vt, and clock gating, and power gating can be used to further minimize power.

3.2 SAR ADC

The ADC is fully differential and is operated at $V_{DD} = 1.05$ V and at a clock frequency of $12f_s = 240\,\text{kHz}$. Since the compressed sensing DSP can run at a lower V_{DD}, level shifters or inverter buffers can be used to interface with the DSP.

3.2.1 SAR Logic

Transmission gate flip flops with set and reset were used to design the SAR logic. NOR2 gates were used to implement the asynchronous set and reset functionality of the flip flop. The optimal PMOS/NMOS sizing ratio is about 1.6:1 for the 32 nm technology [16]. In NOR2, the width of the two stacked PMOS transistors are sized up by 1.8x since stacked transistors are less saturated [16]. Therefore, the size of the PMOS transistors in NOR2 is $\text{W}_{\text{NMOS}} \times 1.6 \times 1.8$. Since the NMOS and the PMOS transistors are not competing in transmission gates, they are sized the same. High V_T devices are used to minimize leakage. A simulation shows that the transmission gate flip flop consumes 264 pW with $\text{W}_{\text{NMOS}} = 100\,\text{nm}$ and L = 70 nm at a switching frequency of 240 kHz. Increasing the length helps decrease leakage, but further



(a) SAR Logic Block Diagram (b) SAR Logic Functional Simulation

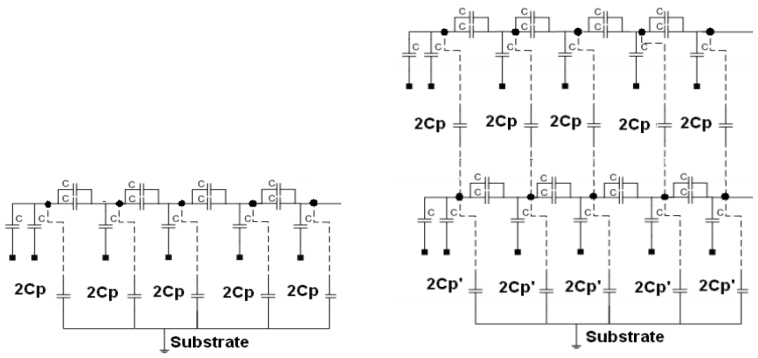
Figure 3.3: SAR Logic

increase of length beyond L = 70 nm increases capacitance with very little gain in power. Even at 240 kHz, W_{NMOS} = 100 nm with L = 70 nm provides sufficient speed for the flip flops.

The employed 12-bit SAR logic, which is shown in Fig. 3.3, is based on the standard sequencer and code register design presented in [17]. It was shown that this SAR logic type uses the least amount of power from the 3 logic types that were studied in [18]. The top flip-flop array shifts a '1' every clock cycle. This is used to set the bottom flip-flop array to store the output of the comparator for each bit. For example, the most significant bit is set first. This drives $1000\cdots$ to the DAC and the comparator compares this value with the input signal and produces an output. At the next clock edge, the result of the comparator is stored and the next significant bit is set and so on. Fig. 3.3 shows the simulation for the case where the output of the SAR logic is $10101010\cdots$.

3.2.2 Hybrid DAC

A C2C array has significantly lower power consumption than a binary weighted array. However, there are many floating nodes in the C2C array that couple to ground due to parasitic capacitors. This introduces error as some charge is lost to the parasitic capacitors. One way to deal with this error is to insert a secondary capacitor array between the floating



(a) Conventional C2C DAC (b) C2C DAC with Shielding

Figure 3.4: C2C DAC

nodes and ground as shown in Fig. 3.4 [19]. The secondary capacitor array shields the floating nodes from ground and reduces the voltage that appears across the parasitic capacitors so that the linearity loss associated with the parasitic caps is mitigated.

Since this linearity loss is minimal in a binary weighted array, a hybrid DAC can be used, where B_1-bit are used with a C2C DAC and B_2-bit are used with a binary DAC. The power and accuracy can be traded off with the choice of B_1 and B_2. Therefore, there exists some optimal solution that achieves the required accuracy for minimum power. A 10-2 DAC ($B_1 = 10$, $B_2 = 2$) uses the same amount of capacitance as a 12-bit C2C DAC. With about $C_{\text{parasitic}}/C_{\text{u}} = 0.1$, experiments show that 9-3 hybrid DAC with FVS gives the lowest power while achieving the required accuracy. The linearity depends heavily on the capacitor matching in the C2C array. The Metal-insulator-Metal process gives the best capacitor matching with 1-2 fF/um^2.

The thermal noise in the ADC is limited by the LSB of the required ENOB. For a 10-bit ENOB, we have $\sqrt{\frac{kT}{C_{in}}} < \frac{1}{2}\frac{V_{DD}}{2^{10}}$. This sets the minimum C_{in} at 15.2 fF and hence, C_u at 7.6 fF, assuming $C_{\text{in}} = 2C_u$. $C_u = 10\,\text{fF}$ is chosen as the unit capacitance in the hybrid DAC. The sample and hold circuit inherently comes from the switches in the differential hybrid

DAC. The input signals gets cut off after reset turns off.

3.2.3 Comparator

For a single stage amplifier, the gain bandwidth is directly related to ω_T of the device. To break this limit, N stage amplifier can be used, where the gain increases as $(A_{stage})^N$, but the speed decreases with each additional stage. From the gain and bandwidth optimization standpoint, we want $N = ln(A_{total})$ and $A_{stage} = e$, then the bandwidth becomes $\frac{\omega_T}{e \cdot ln(A_{total})}$. A regenerative latch is almost like an N stage amplifier, which can be formed using cross-coupled inverters. With time, the signal gains A_{stage} each time it circulates the positive feedback loop of the cross-coupled inverters. The time constant for this circuit is $\tau = g_m/C_{gs}$, then the time required to regenerate is $t_{regen} = ln(A_{total}) \cdot \frac{Cgs}{g_m} = ln(A_{total}) \cdot {\omega_T}^{-1}$. The regeneration time is set by ω_T of the device and to minimize t_{regen}, the devices should be made as small as possible, though the settling time is not a stringent requirement at 20 kHz. The comparator needs to amplify a minimum difference of 1/2 LSB input voltage up to V_{DD}, i.e $A_{total} = \frac{1.05}{0.5*1.05/2^{10}} = 66.2$ dB. An NMOS transistor at around $V^* = 0.2$ V gives $\omega_T = 2\pi \cdot 150$ GHz, which requires $t_{regen} = 1.23$ ns. Therefore, the width of the transistors should be of minimal size (W = 100 μm, L = 70 μm) to minimize power as much as possible.

A current-mode latch (CML) comparator uses a diff pair (preamp) to inject an input signal to a regenerative latch. The power consumption is set by the bias current I_b. The output of the CML comparator does not swing rail-to-rail. Instead, the output swings from $V_{DD} - I_b R$ to V_{DD}. The digital circuits that interface with this comparator will be slower and consume more leakage power, because it may not fully turn on/off the devices. The rail-to-rail output swing can be obtained with a Strongarm comparator, which is shown in Fig. 3.5.

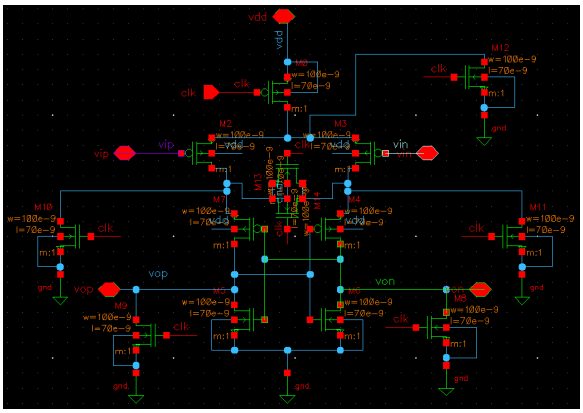


Figure 3.5: Strongarm Comparator Schematic

The operation of this circuit is not much different from the CML comparator; the difference is that the output is precharged to V_{DD} or GND with a clock. Once the output is precharged, the input signal is amplified by a preamp, and the output starts to follow the input. Once the output has moved by V_T from the precharged voltage, the cross-coupled inverters kick in and the differential output regenerates exponentially to VDD or gnd. This circuit functions like a digital circuit and the power consumption is just going to be $P = \frac{1}{2}CV^2f$. Comparators tend to remember the last decision it made. Hysteresis is mitigated by precharging/shorting all node capacitors during the precharging phase. High V_T devices are used for low power and the input is driven to PMOS devices to reduce flicker noise. The comparator is fully differential and the systematic offset is zero, since everything is balanced. Large transistors can be used to further minimize mismatches and flicker noise.

4. Results

4.1 Compressed Sensing DSP

4.1.1 Bio-signals with Noise

If a sampled signal is compressed by mere downsampling, it's very possible to miss a part or an entire spike as shown in Fig 4.1. The spikes, which occupy the high bins in FFT, are either aliased or filtered with LPF, if the signal is downsampled below the Nyquist frequency. Compressed sensing allows the signal to be effectively 'downsampled' below the Nyquist frequency. Compressed sensing relies on the signal's sparsity for accurate reconstruction. In reality, there is always noise associated with bio-sensors, which can significantly reduce the

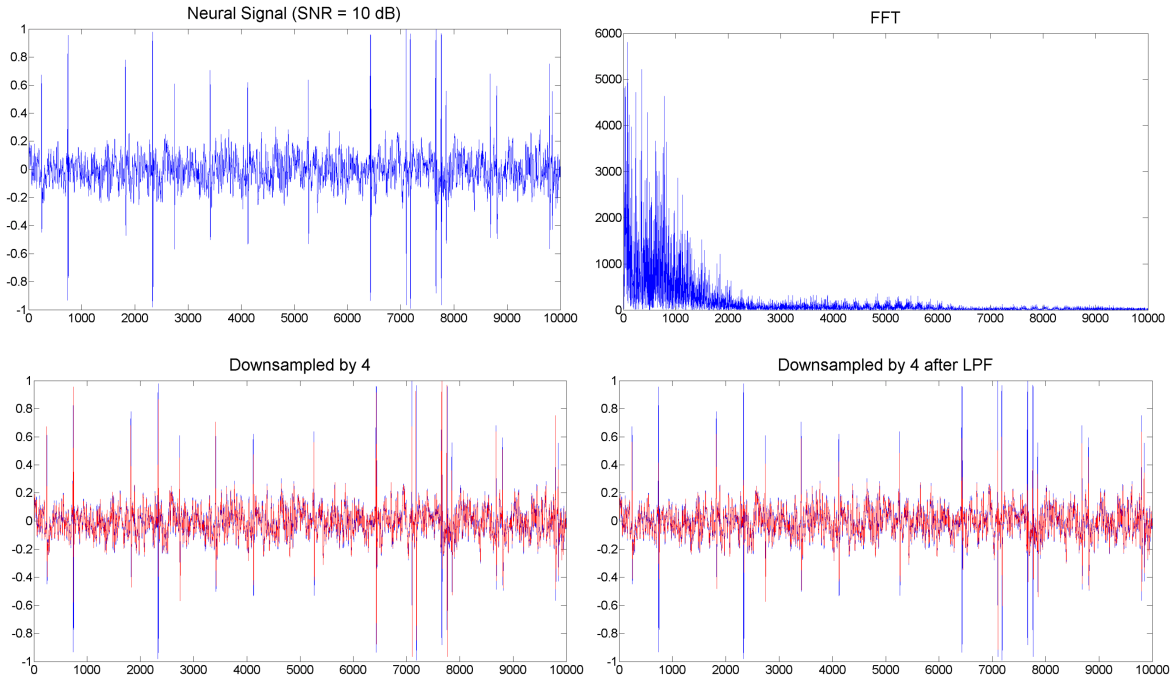


Figure 4.1: Compression by Downsampling (N = 10000, M = 2500)

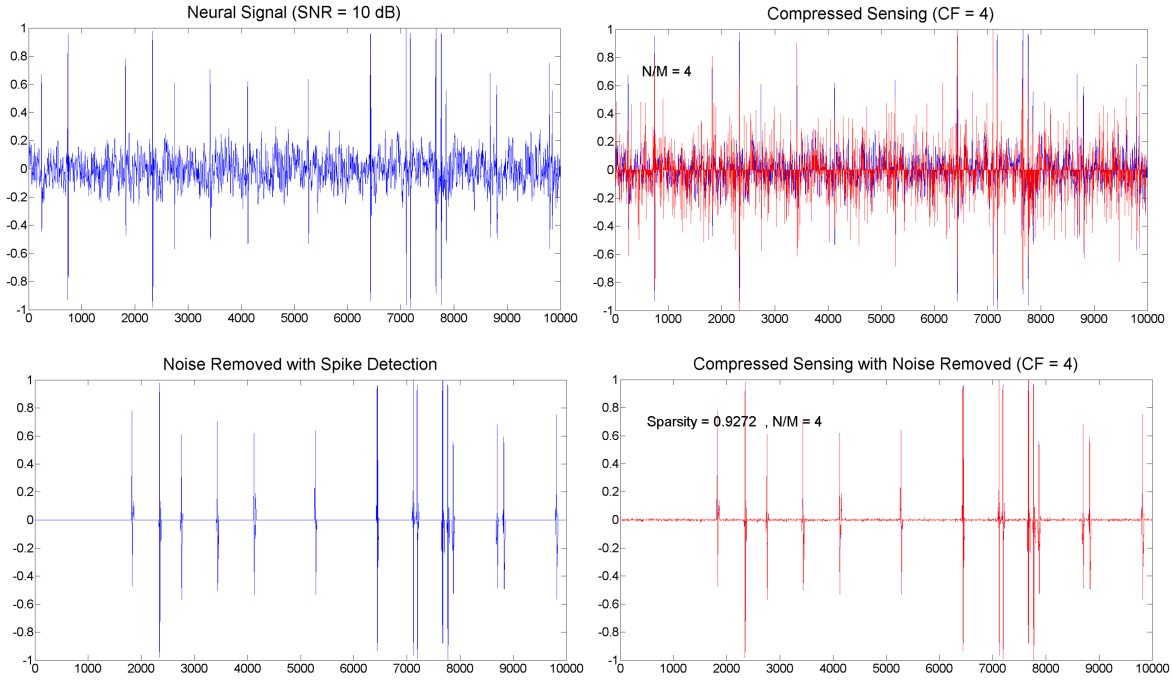


Figure 4.2: Compressed Sensing With and Without Noise (N = 10000, M = 2500)

sparsity of a signal. Fig. 4.2 shows a neural signal with SNR = 10 dB, which was compressed by 4 times (CF = 4) using compressed sensing (N = 10000, M = 2500), and reconstructed using L1 with equality constraints. The reconstructed signal does *not* accurately recover the spikes, which store important information. However, when noise is filtered using spike detection, Fig. 4.2 shows that the signal can be recovered pretty accurately with the same compression factor. Therefore, the compressed sensing DSP will be more accurate and robust by using some form of spike detection in front. Then, the compressed sensing reconstruction can tolerate noise as long as the spikes can be detected.

4.1.2 Ill-Behaved Bio-signals

All published compressed sensing chips assume that the input signals are 'well-behaved' and use fixed N, M, and CF during the operation, though they can be configured at the beginning. However, most bio-signals are changing all the time along with their sparsity. In

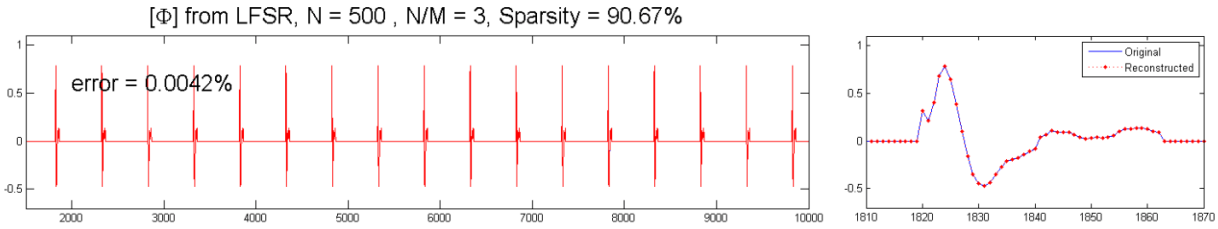


Figure 4.3: Compressed Sensing with a Well-Behaved Bio-signal

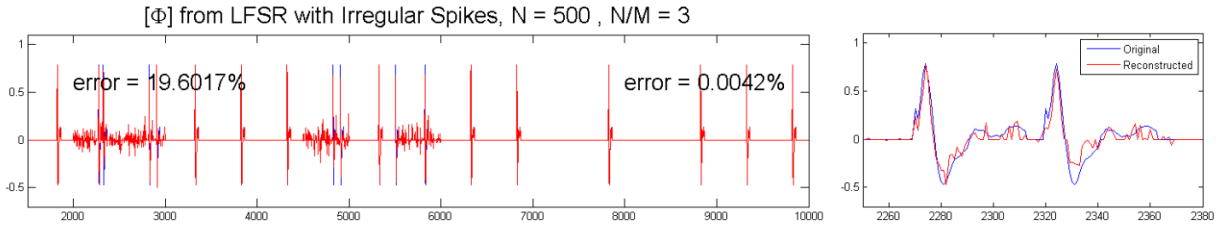


Figure 4.4: Compressed Sensing with an Ill-Behaved Bio-signal

these chips, irregular spikes can drastically increase or decrease the error of reconstruction. For example, suppose that neural spikes occur exactly every 500 samples and about 50 samples make up a spike. If N is chosen such that it contains exactly one spike ($N = 500$), it results in a sparsity of $1 - 50/500 = 90\%$. Experiments show that the error with CF of 2, 3, and 4 are 0.00078%, 0.0042%, and 22.3% respectably. The change in error from CF from 3 to 4 is quite substantial. To minimize power consumption, the highest compression factor should be chosen such that it gives a corresponding error that is tolerable for a given application. Therefore, the CF of 3 is chosen for further analysis and the reconstructed signal is shown in Fig. 4.3. The error is calculated by $norm(\hat{x} - x)/norm(x)$, where $\hat{x}$ is the reconstructed signal of x.

Suppose now that two spikes happened to occur in one period as shown in Fig. 4.4. The reconstruction with CF of 3 will now be done with a much greater error because the signal's sparsity has reduced to about 80%. To maintain the same error as before, the CF has to be decreased to 2 and use more power. If the spike rate returns to one per N, the CF can be

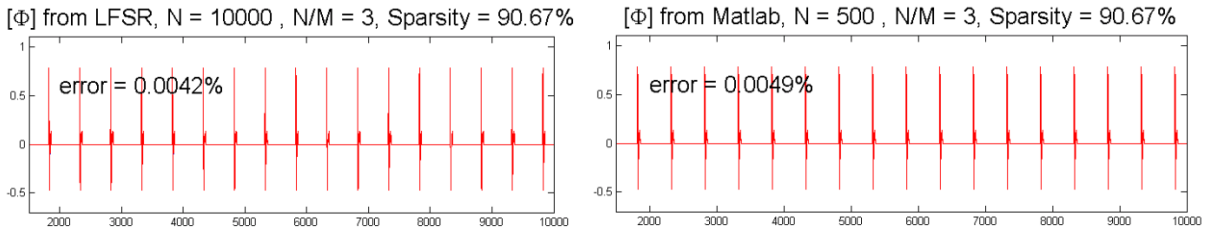


Figure 4.5: Incoherence of [Φ] from LFSR vs. Matlab (*randi* function)

returned to 3. However, if no spike was detected in N, the whole compressed sensing DSP can be powered down. If the CF can be dynamically adjusted depending on the sparsity of the input signal, the compressed sensing DSP can be robust and consume less power through powering off unused resources.

4.1.3 Random Matrix ([Φ]) with LFSR

Fig. 4.5 shows that the rows of [Φ] generated using the LFSRs exhibit a similar level of incoherence of rows of [Φ] generated using Matlab's random number generating function, *randi*. The actual incoherence between [Φ] and [Ψ] are $\mu(\Phi_{\text{LFSR}}, \Psi) = 1.0239$ and $\mu(\Phi_{\text{Matlab}}, \Psi) = 1.0330$ with $N = 10000$, which are comparable to the measurements in [2]. The coherence is calculated by $\mu(\Phi, \Psi) = \sqrt{N} \max_{1 \leq k,j \leq N} |\langle \phi_k, \psi_j \rangle|$.

4.1.4 Power

The power consumption of the proposed compressed sensing DSP is estimated. The power consumption is dominated by leakage in registers at $f = 20$ kHz. The size of each accumulator needs to be 10-bit + $log_2(\frac{N}{2})$, because about half of the inputs are discarded due to the random nature of [Φ]. Since there are M accumulators and the sparsity, s, is known, the total number of registers used by the accumulator is $M(10 + log_2(\frac{(1-s)N}{2}))$. The LFSRs use M + 16 registers and the buffer requires $10N(1-s)$ registers. Therefore, the total number of registers used is $M(10 + log_2(\frac{(1-s)N}{2})) + M + 16 + 10N(1-s)$.

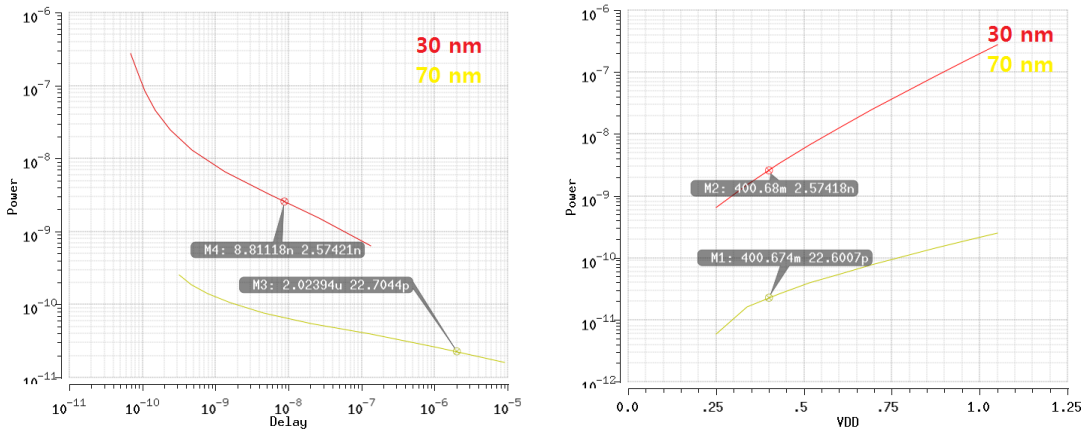


Figure 4.6: Power-Delay Curve of DFFARX1_HVT at L = 30 nm and L = 70 nm

At 20 kHz, there is going to be a lot of slack that could be traded for power. V_{DD} = 0.4 V was chosen to give enough margins to be safe against variability. A high-V_T register (DFFARX1_HVT) from the library consumes a lot of leakage power because of the small length devices. Instead, the transistor length in DFFARX1_HVT was increased to 70 nm. Further increases of length resulted in diminishing returns. The power-delay curves for DFFARX1_HVT with L = 30 nm and L = 70 nm are shown in Fig. 4.6. The power consumption is 22.6 pW and the t_{c-q} is 2 μs with L = 70 nm at $V_{DD} = 0.4\,\mathrm{V}$, which is only 4% of the clock period. The critical path exists somewhere in the adders, but there is still big enough slack left to meet timing.

First, the total number of registers is multiplied by the power consumption of a single register. Then, a generous margin of 40% is added to compensate for uncertainties, and also to account for adders, combinational logic, and other registers. The spike detection block is estimated to consume about 0.5 μW at $V_{DD} = 0.4\,\mathrm{V}$ with all the power reduction techniques applied. The total power of the proposed DSP is compared against [2] and [3] in Fig. 4.7.

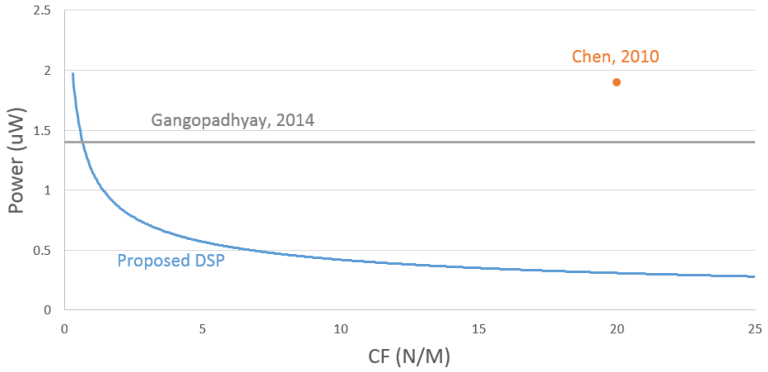


Figure 4.7: CS FE Power Consumption (N = 1000)

4.1.5 Summary

[2] is an analog implementation of CS FE and uses 1.8 μW with $M = 64$ at 2 kHz. This power consumption depends only on M and the clock frequency, and not N. In other words, the power consumption does not depend on CF. Table 4.1 compares the proposed DSP with [2] and [3]. The DSP is estimated to use about 0.5 μW (spike detection) + 0.025 μW (CS) = 0.525 μW at $V_{DD} = 0.4$ V. At such high CF, the compressed sensing block consumes negligible power compared to the spike detection block, which is its key strength.

Table 4.1: Compressed Sensing DSP Comparison with Other Works

Author	**Tech.**	**V_{DD}**	**f**	**CF**	**M**	**Power**
This DSP	32 nm	0.4 V	20 kHz	20	50	0.525 μW
Gangopadhyay [2]	130 nm	0.9 - 1.2 V	2 kHz	1 - 20	64	1.8 μW
Chen [3]	90 nm	0.6 V	20 kHz	20	50	1.9 μW

4.2 SAR ADC

4.2.1 INL/DNL and Noise

The INL/DNL simulation shows that there are a lot of missing codes with bits 1 and 0, which means that the ADC is accurate to about 10 bits. In addition, there appears to be some missing codes in the 7^{th} bit whenever it switches. The nonlinearity essentially all comes from

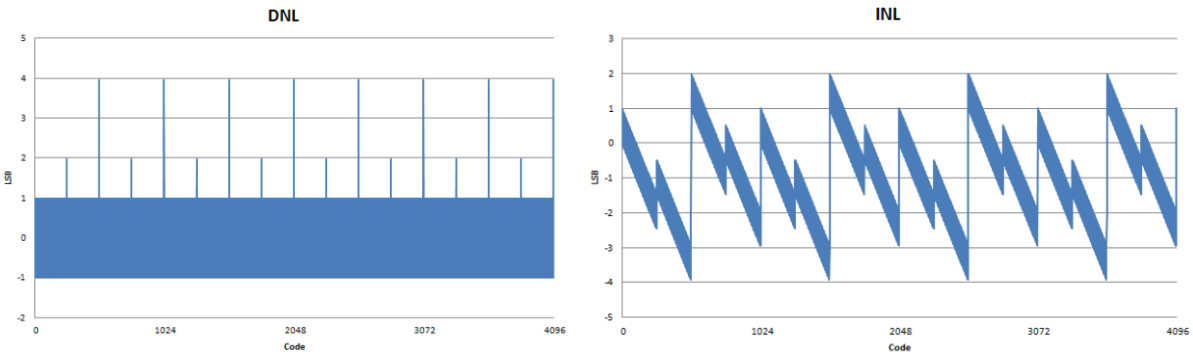


Figure 4.8: DNL and INL

the parasitic capacitors in the hybrid DAC. The simulations are based on $C_{\text{parasitic}}/C_u = 0.1$. The missing codes are eliminated with $C_{\text{parasitic}}/C_u = 0.01$. At 20 kHz, the input referred noise is 1.22 $\mu\text{V}/\sqrt{\text{Hz}}$ or 172.5 μV, which is less than 1/2 LSB.

4.2.2 ENOB

ENOB can be calculated from SNDR by (SNDR − 1.76)/6.02. SFDR is approximately SFDR $\approx 20\log_{10}(2^B/\text{INL}) = 20\log_{10}(2^{12}/4) = 60.2$ dB, where INL = 4. Assuming SFDR $\approx$ SND $\ll$ SNR, the ENOB is $(60.2 - 1.76)/6.02 = 9.7$-bits. Fig. 4.9 shows the transient simulation of the ADC giving an output of $10101010\cdots$, where the output is accurate up to the 10^{th} bit.

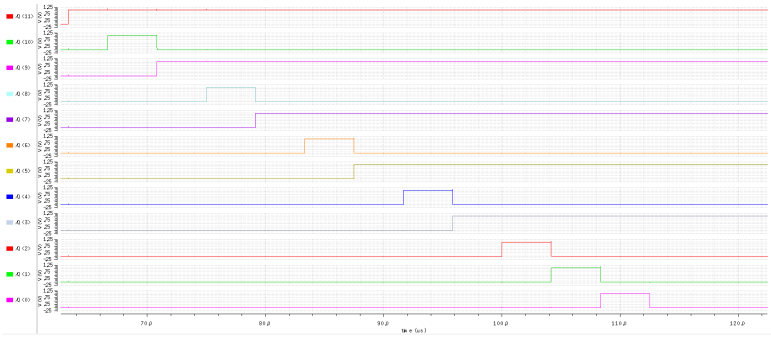


Figure 4.9: ADC Transient Simulation

4.2.3 Power

The ADC is clocked at $12f_s$ internally. There are 26 registers in the SAR digital logic. Since each registers consume 264 pW at 240 kHz, the total power consumed by the SAR digital logic would be about $26 \cdot 264\,\text{pW} = 6.86\,\text{nW}$. A 9-3 hybrid capacitive array with FVS has a total capacitance of $78C_u$. Since it is differential, it has a total capacitance of $156C_u$. However, some capacitors in the C2C array never switch. A more accurate capacitance to use in the power calculation is $96C_u$. Assuming a switching activity of 1/12, the total switching power of the hybrid DAC is approximately $\frac{1}{2} \cdot \frac{1}{12} \cdot 10\,\text{fF} \cdot 0.525^2 \cdot 12 \cdot 20\,\text{kHz} \cdot 96 = 2.65\,\text{nW}$. Here, $V_{DD} = 0.525\,\text{V}$ is used, because the capacitors are switching between $V_{DD}/2$ and GND in differential circuits. The total power consumed by the digital logic and the hybrid DAC together should be about 9.51 pW. The simulation shows that the actual power consumption is 8.76 nW.

The comparator has to switch 5 capacitive nodes of about 0.2 fF, so the dynamic power consumption alone would be around $\frac{1}{2}CV^2fN = \frac{1}{2} \cdot 0.2\,\text{fF} \cdot 1.05^2 \cdot 12 \cdot 20\,\text{kHz} \cdot 5 = 132.3\,\text{pW}$. The comparator has to drive 12 registers in the digital logic. Two FO4 inverters are placed at the output of the comparator to help drive the digital logic. The inverter buffers have to burn about $\frac{1}{2}CV^2fN = \frac{1}{2} \cdot (0.5\,\text{fF} + 0.1\,\text{fF} \cdot 12) \cdot 1.05^2 \cdot 12 \cdot 20\,\text{kHz} = 225\,\text{pW}$ of dynamic power to drive the 12 registers. The simulation shows that the comparator uses 324.5 pW, while the inverter buffers use 711 pW. Since the power consumed by the comparator is very low compared to the digital logic and the DAC, more power can be used to reduce random offsets and noise in the comparator, which may include using bigger sized transistors, chopper stabilization (1/f noise), and offset cancellation, if necessary. The total power consumption

of the differential SAR ADC at $f_s = 20$ kHz and $V_{DD} = 1.05\,\text{V}$ is summarized in Table 4.2.

Table 4.2: Total Power Consumption of the Differential SAR ADC

SAR Logic + Hybrid DAC	**Comparator**	**Inverter buffers**	**Total**
8.76 nW	324.5 pW	711 pW	9.8 nW

4.2.4 Summary

The overall performance of the SAR ADC is summarized in Table 4.3. The full schematic of the ADC is included in Fig. B.1. Figure of Merit $\left(\text{FOM} = \frac{\text{P}}{2^{\text{ENOB}} \times f_s}\right)$ is a good measure

Table 4.3: ADC Summary of Performance

ENOB	**f_s**	**SFDR**	**INL/DNL**	**Power**	**FOM $\left(\frac{\text{fJ}}{\text{conv-step}}\right)$**
9.7-bits	20 kHz	60.2 dB	4/4	9.8 nW	0.59

of efficiency that captures both power and performance. With an ENOB of 9.7 bits, the ADC's FOM is 0.59 fJ/conv-step. The FOM of the ADC is very low, even lower than some of the most sophisticated ADCs. In reality, the ADC may lose some accuracy due to offsets, mismatches, and variations, especially without any calibration techniques, thereby bringing up the FOM to a more reasonable value. Table 4.4 compares the results against the state-of-the-art SAR ADCs presented in ISSCC in 2013.

Table 4.4: ADC Comparison with Other Recent Works

Author	**Type**	**Tech.**	**ENOB**	**f_s (kHz)**	**P (nW)**	**FOM $\left(\frac{\text{fJ}}{\text{conv-step}}\right)$**
This ADC	SAR	32 nm	9.7	20	9.8	0.59
Harpe [14]	SAR	65 nm	10.1	40	97	2.2
Liou [20]	SAR	90 nm	8.7	500	500	2.4

5. Conclusion

5.1 Compressed Sensing DSP

In [2][3], the compressed sensing chips use bio-signals that have constant rate of information with minimal noise. However, noise can easily cause these systems to fail. Furthermore, since CF is fixed, any variations in bio-signals can also cause these systems to fail. If these systems were to use a small CF to compensate for variations, it would be wasting power. With spike detection, noise is suppressed and only the spikes are subjected to compressed sensing. It also serves to dynamically configure N, M, and CF, so that it is robust with any type of bio-signals with significant power savings.

Since the power is dominated by leakage, it would be worthwhile to look into re-using some parts of logic to even out dynamic and leakage power. At high CF, the spike detection block dominates the power consumption. Therefore, more studies could be done on the different ways of spike detection and determine which one is the most suitable for compressed sensing.

5.2 ADC

The power consumption of the ADC is much smaller than that of the compressed sensing DSP. The ADC achieves 9.7-bit accuracy, which is on par with the target, and uses 9.8 nW at 20 kHz. The majority of the power comes from the digital logic and switching capacitors in the hybrid DAC. The kT/C noise limits the unit capacitance used in the hybrid DAC so there is not much to be gained here. In the digital logic, the leakage in the registers consume a lot

of power. Using a lower V_{DD} would significantly reduce the power consumption. The switches used in the differential DAC should be kept at a relatively high V_{DD} to avoid linearity loss. The comparator may have issues with offsets, mismatches, and variations that arise from the minimum sized transistors. These issues can be solved with calibration techniques, offset cancellation techniques such as double correlated sampling and chopper stabilization, and careful sizing of the transistors. Since the power consumption of the ADC is negligible compared to the DSP, the voltage can be kept at $V_{DD} = 1.05\,\mathrm{V}$.

Bibliography

[1] A. Pantelopoulos and N. Bourbakis, "A Survey on Wearable Biosensor Systems for Health Monitoring," *30th Annual International IEEE EMBS Conference*, pp. 4887–4890, 2008.

[2] Gangopadhyay, E. G. Allstot, A. M. R. Dixon, K. Natarajan, S. Gupta, and D. J. Allstot, "Compressed Sensing Analog Front-End for Bio-Sensor Applications," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 2, pp. 426–438, 2014.

[3] F. Chen, A. Chandrakasan, and V. Stojanovic, "A Signal-Agonostic Compressed Sensing Acquisition System for Wireless and Implantable Sensors," *IEEE Custom Integrated Circuits Conference*, pp. 1–4, 2010.

[4] World Health Organization. (2013), [Online]. Available: `http://www.who.int/mediacentre/factsheets/fs317/en/index.html`.

[5] Transparency Market Research. (2013), [Online]. Available: `http://www.transparency-marketresearch.com/biosensors-market.html`.

[6] V. Karkare, S. Gibson, and D. Markovic, "A 130- W, 64-Channel Neural Spike-Sorting DSP Chip," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 5, pp. 1214–1222, 2011.

[7] E. Candes, J. Romberg, and T. Tao, "Stable Signal Recovery from Incomplete and Inaccurate Measurements," *Communications on Pure and Applied Mathematics*, vol. 59, pp. 1207–1223, 2006.

[8] D. Donoho, "Compressed Sensing," *IEEE Trans. on Information Theory*, vol. 52, pp. 1289–1306, 2006.

[9] D. Donoho, "For Most Large Underdetermined Systems of Linear Equations the Minimal L1-Norm Solution is Also the Sparsest Solution," *Communications on Pure and Applied Mathematics*, vol. 59, pp. 797–829, 2006.

[10] E. Candes and M. Wakin, "An Introduction to Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, pp. 21–30, 2008.

[11] A. Dixon, E. Allstot, D. Gangopadhyay, and D. Allstot, "Compressed Sensing System Considerations for ECG and EMG Wireless Biosensors," *EEE Trans. on Biomedical Circuits and Systems*, vol. 6, pp. 156–166, 2012.

[12] S. Mukhopadhyay and G. Ray, "A New Interpretation of Nonlinear Energy Operator and its Efficacy in Spike Detection," *IEEE Trans. Biomedical Eng.*, vol. 45, no. 2, pp. 180–187, 1998.

[13] B. Murmann. (2013), [Online]. Available: `http://www.stanford.edu/~murmann/adcsurvey.html`.

[14] P. Harpe, E. Cantatore, and A. Roermund, "A 2.2/2.7fJ/conversion-step 10/12b 40kS/s SAR ADC with Data-Driven Noise Reduction," *IEEE International Solid-State Circuits Conference*, 2013.

[15] E. Alpman, H. Lakdawala, L. Carley, and K. Soumyanath, "A 1.1V 50mW 2.5GS/s 7b time-interleaved C-2C SAR ADC in 45nm LP digital CMOS," *Dig. IEEE Intl. Solid-State Circuits Conf*, pp. 76–77, 2009.

[16] J. Rabaey, *EECS 241B Lecture Slides, 2014*.

[17] T. Anderson, "Optimum Control Logic for Successive Approximation Analog-to-Digital Converters," *Computer Design*, vol. 11, no. 7, pp. 81–86, 1972.

[18] R. Hedayati, "A Study of Successive Approximation Registers and Implementation of an Ultra-Low Power 10-bit SAR ADC in 65 nm CMOS Technology," *Master thesis, Linkoping Institute of Technology*, 2011.

[19] H. Balasubramaniam, W. Galjan, W. Krautschneider, and H. Neubauer, "12-bit Hybrid C2C DAC based SAR ADC with Floating Voltage Shield," *3rd International Conference on Signals, Circuits and Systems*, pp. 1–5, 2009.

[20] C. Liou and C. Hsieh, "A 2.4-to-5.2fJ/conversion-step 10b 0.5-to-4MS/s SAR ADC with Charge-Average Switching DAC in 90nm CMOS," *IEEE International Solid-State Circuits Conference*, 2013.

A. Compressed Sensing DSP

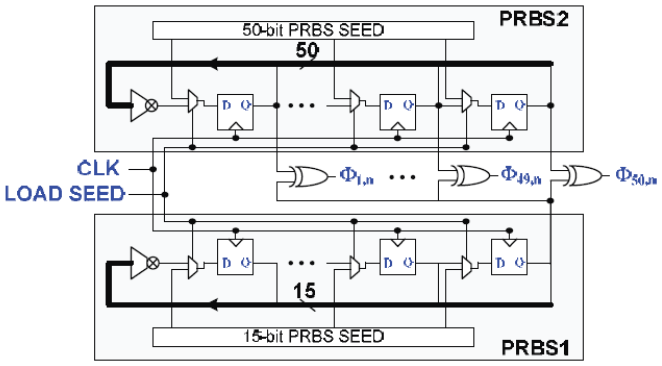


Figure A.1: [Φ] Generation using LFSRs [3]

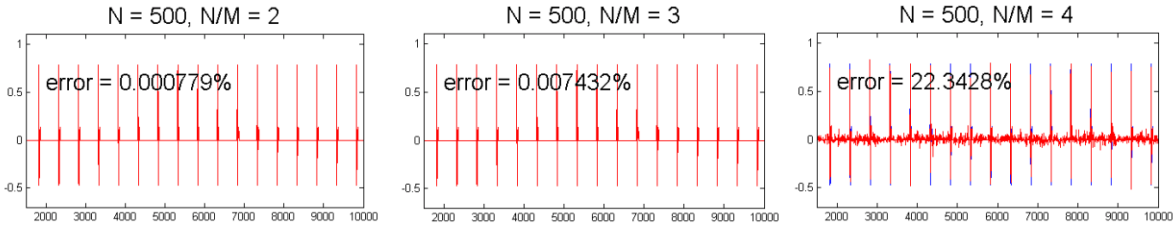


Figure A.2: Compression Factor

B. ADC

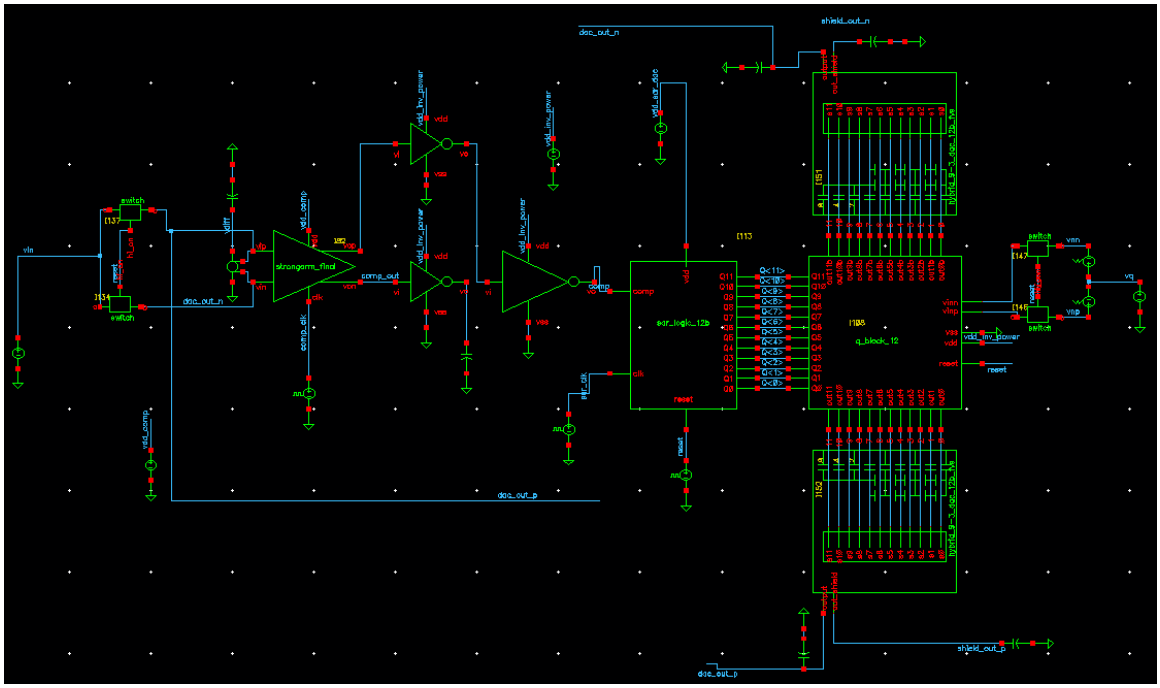


Figure B.1: Schematic of the Differential SAR ADC

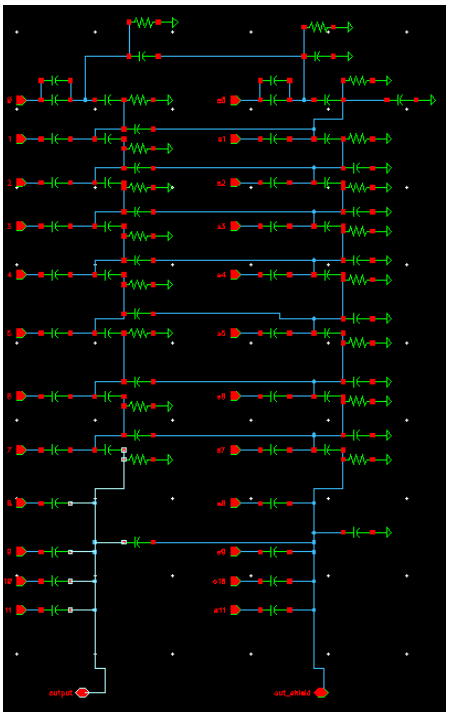


Figure B.2: 9-3 Hybrid DAC (Resistors to ground are very large)

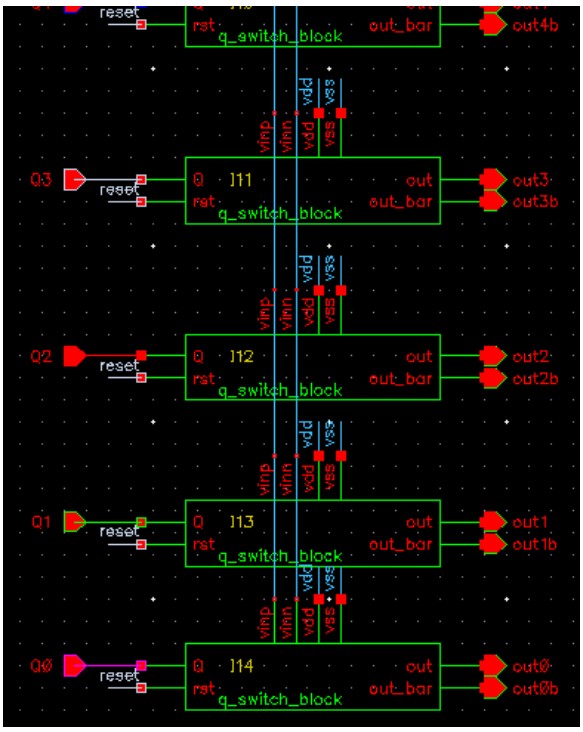


Figure B.3: Switch Block used in the Differential DAC

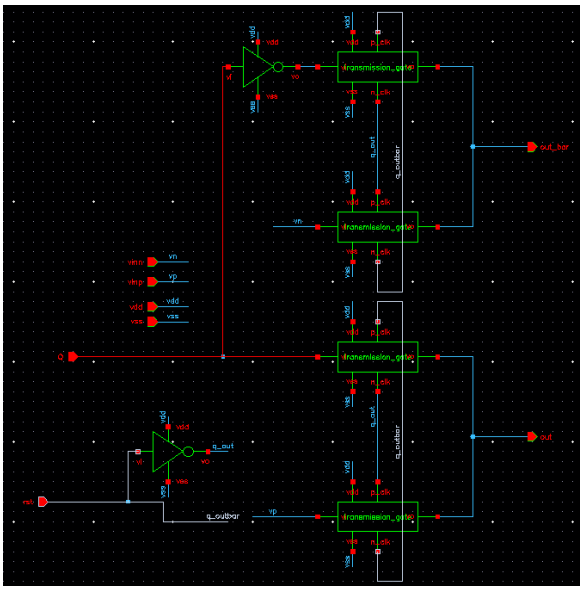


Figure B.4: Switches used in the Switch Block of the Differential DAC

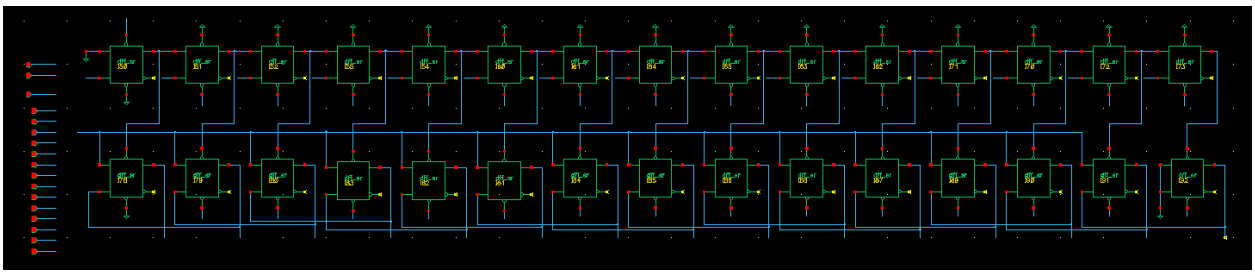
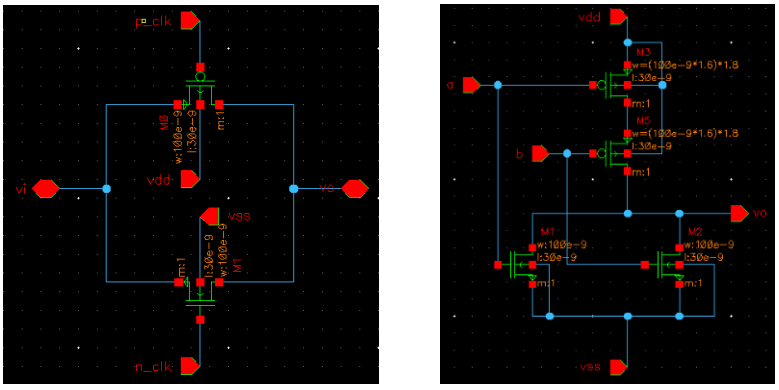
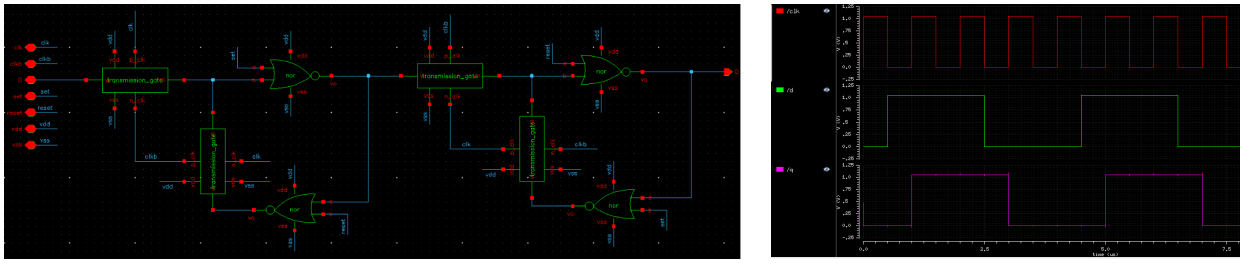


Figure B.5: SAR Logic Schematic



(a) Transmission Gate (b) NOR Gate

Figure B.6: Gate Sizing



(a) Transmission Gate Flip Flop (b) Flip Flop Functional Simulation

Figure B.7: Flip Flop for SAR Logic