A Hardware Platform to Test Analog-to-Information Conversion and Non-Uniform Sampling

by

Miguel E. Perez

B.S., Massachusetts Institute of Technology (2012)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2012

© Massachusetts Institute of Technology 2012. All rights reserved.
A Hardware Platform to Test Analog-to-Information Conversion and Non-Uniform Sampling

by

Miguel E. Perez

Submitted to the Department of Electrical Engineering and Computer Science on September 4, 2012, in partial fulfillment of the requirements for the degree of Master of Engineering in Electrical Science and Engineering

Abstract

The Nyquist-Shannon sampling theorem tells us that in order to fully recover a band-limited signal previously converted to discrete data points, said signal must have been sampled at a frequency greater than twice its bandwidth. This theorem puts a burden on circuits like ADCs, in the sense that the higher the bandwidth of a signal, the faster the ADC must be by a factor of at least 2. This in turn translates into higher power consumption. The problem can be mitigated to a certain extent by the use of zero-crossing based ADCs which consume much less power than conventional op-amp based ones, while maintaining the same performance levels. However, the burden still remains, and with the increase in the use of biologically implantable devices, the need for the utmost power efficiency is essential. This is where the theory of compressed sensing seems to offer an alternate solution. Instead of solving the problem with the brute force approach of increasing power consumption to meet performance, compressed sensing promises to increase the effective figure of merit (FOM) by exploiting certain characteristics in the signal's structure. Compressed sensing tells us, that a signal that meets certain criteria, does not need to be sampled at twice its bandwidth in order to be fully recoverable. This means that an ADC no longer has to operate at the Nyquist rate to guarantee that the signal will not be distorted and as a result its power consumption can be reduced considerably. This allows for more robust and energy efficient data acquisition circuits. This means more efficient and longer lasting implantable monitoring devices along with the ability to perform on-site data processing.

Thesis Supervisor: Hae-Seung Lee
Title: Professor of Electrical Engineering
Acknowledgments

First and foremost, I want to give my sincerest thanks to Professor Hae-Seung Lee for having granted me the great opportunity to work alongside him and his group. I would also like to thank him for always being a source of guidance and having great patience. I must say, I will be leaving with a much greater appreciation for what it takes to know a subject as well as one possibly can. I would also like to thank Carolyn Collins for being always there to answer any administrative questions and for adding so much life to our group dinners. To everyone in lab who ever took time from their busy schedules to answer my queries or to just say hello, I'm grateful. I would especially like to thank SungWong, who has always been gracious enough to share his infinite wisdom with those who ask for it, and without whom, endless piles of cookies would have senselessly gone to waste. Finally, I would like to say thank you to my family, and especially to my mom for always being a source of strength and unconditional support. This thesis is dedicated to her.
Contents

1 Introduction .......................................................... 13
   1.1 Compressed Sensing ................................................. 13
   1.2 Analog-to-Information conversion ................................. 16
   1.3 Hardware architecture ............................................. 16

2 Compressed Sensing and AIC ...................................... 19
   2.1 Compressed Sensing ............................................... 19
      2.1.1 Sparsity and Incoherence .................................... 22
      2.1.2 Random Sensing Matrices ................................... 23
   2.2 Analog-to-Information Conversion ................................. 24
   2.3 Summary ........................................................... 27

3 Front-end amplifier .................................................. 29
   3.1 ECG signals ....................................................... 29
   3.2 Standard Instrumentation Amplifier ............................... 30
   3.3 Low noise front end amplifier ................................... 32
      3.3.1 Flicker Noise and Chopper Stabilization .................. 34
      3.3.2 Circuit Design .............................................. 43
   3.4 Low-Pass Filter .................................................... 49
   3.5 Front-End Noise Analysis ......................................... 55
      3.5.1 First Stage Noise Analysis .................................. 56
3.6 Summary ................................................................. 63

4 Two Step Triple Slope ADC ............................................. 65
  4.1 Integrating ADC ....................................................... 65
    4.1.1 Dual Slope ADC ............................................... 65
    4.1.2 Triple Slope ADC ............................................. 66
    4.1.3 Triple Slope Optimal Conversion Time .................... 69
  4.2 Two Step ADCs ....................................................... 70
    4.2.1 Two Step ADC Errors ....................................... 71
  4.3 Pipeline ADCs ........................................................ 73
    4.3.1 Standard Architecture ..................................... 73
    4.3.2 Accuracy Requirements ..................................... 74
    4.3.3 Digital Error Correction .................................... 74
  4.4 Two Step Triple Slope Architecture ............................. 77
    4.4.1 Circuit Diagram .............................................. 77
    4.4.2 Basic Operation .............................................. 80
    4.4.3 Zero Crossing Detector Design ............................. 81
    4.4.4 OpAmp Design .................................................. 83
    4.4.5 OpAmp Helper Circuits ..................................... 85
    4.4.6 OpAmp Open Loop Gain and Settling Time ................ 91
    4.4.7 Integrator Noise Analysis ................................ 95
    4.4.8 R2R Ladder Design .......................................... 101
    4.4.9 State Machine Design ....................................... 108
  4.5 Summary ............................................................. 111

A Noise Analysis of Selected Blocks .................................. 113
  A.1 OpAmp1 Noise Analysis .......................................... 113
  A.2 First Stage ZCD Noise Analysis ................................ 118
List of Figures

1-1 Overall block diagram illustrating multi-mode operation ............. 18

2-1 Matrix form for the equation \( y = \Phi \Psi x \) ......................... 20

2-2 Graphical representation of A-to-I encoder .......................... 25

2-3 Circuit level diagram of a single channel in the A-to-I encoder ..... 26

3-1 Standard instrumentation amplifier overall schematic ............... 31

3-2 Overall block diagram illustrating front end low noise amplifier .. 33

3-3 Power spectral density for current noise. Flicker noise is proportional to \( 1/f \) and thus thermal noise dominates at high frequencies greater than \( f_c \) ......................... 35

3-4 Closed-loop amplifier employing AZ technique a) \( \Phi_2 \) is the sample phase b) \( \Phi_1 \) is the input signal processing phase ......................... 37

3-5 SC amplifier employing the CDS technique on each phase \( \Phi_1 \) and \( \Phi_2 \) 38

3-6 Chopper switch connections ................................. 39

3-7 The different waveforms throughout the chopping operation, illustrated in the frequency domain ................................. 40

3-8 a) Effect of AZ on input referred white noise PSD. b) Effect of CHS on input referred white noise PSD ......................... 42

3-9 a) Adaptive element. b) The parasitic resistance from the chopper. 44
3-10 OpAmp1 circuit schematic. Vin+ and Vin- are the signal inputs, Vinv+/‐ are OpAmp1's inputs that connect in feedback. Vcmfb1 is the common-mode feedback node. 48

3-11 Arrows show flow of \( I_{net} = G_{m0} \Delta V_{in} \). 49

3-12 OpAmp2 circuit schematic. Vin+ and Vin- are the signal inputs. Vcmfb2 is the common-mode feedback node. 50

3-13 Classification of information processing systems across independent variables of magnitude and time. 51

3-14 a) SC equivalent of a resistor. b) SC equivalent of a RC integrator. 52

3-15 a) Block diagram of second order low-pass transfer function. b) Equivalent Gm-C values. c) Gm-C circuit diagram. 54

3-16 a) Equivalent noise power model of NMOS in strong inversion. b) Noise power model of NMOS in subthreshold. 57

3-17 a) Standard cascode amplifier with capacitive load. b) Cascode amplifier with noise generators included. 58

3-18 FFT of data points from amplifier simulation employing CHS technique. 62

4-1 a) Integrator block in Dual-Slope ADC. b) Output voltage waveform during conversion. 66

4-2 a) Integrator block in Triple-Slope ADC. b) Output voltage waveform during conversion. 67

4-3 Two step converter block diagram. 71

4-4 Residue voltage waveform. 71

4-5 (a) Real ADC transfer characteristic. (b) Residue with ideal DAC. 72

4-6 (a) Real DAC transfer characteristic. (b) Residue with ideal ADC. 72

4-7 Standard pipeline ADC. 74

4-8 Residue waveform with extra bit of precision in fine ADC. 76

4-9 1.5 bit residue waveform in pipeline converters. 76

4-10 Overall circuit diagram of analog blocks in ADC. 78
4-29 R2R ladder with binary sized switches in order to preserve binary distribution of $I_{ref}$ throughout the ladder. .......................... 107

4-30 Standard architecture of a clocked synchronous state machine. ... 108

4-31 State machine for control of first stage operation. ............... 109

4-32 State machine for control of second stage operation. ............ 110

A-1 First stage of OpAmp along with helper circuits. .................. 116

A-2 Input referred voltage noise power spectral density of first stage OpAmp. 117

A-3 Output waveform of first stage integrator illustrating noise requirements for first stage ZCD. ........................................ 119

A-4 ZCD pre-amplifier ................................................. 119

A-5 ZCD pre-amplifier transfer function ................................ 120
Chapter 1

Introduction

1.1 Compressed Sensing

Compressed sensing is the idea that Emmanuel Candes, David Donoho and Terence Tao developed, that allows to sample certain signals below the Nyquist sampling rate without incurring significant loss of information [1] [5]. It exploits the sparsity of the signal in a specific basis. In order to briefly illustrate the fundamental mathematical foundation of the theory of compressed sensing let us assume $f(t)$ to be a band-limited signal whose bandwidth is $BW$. The Nyquist-Shannon theorem tells us that in order to fully recover $f(t)$ from its samples, the sampling frequency $f_s$ must satisfy the following condition:

$$f_s \geq 2 \cdot BW$$  \hspace{1cm} (1.1)

This means that if $f(t)$ is of duration $T$ then $N$, being the number of samples must be:

$$N = f_s \cdot T$$  \hspace{1cm} (1.2)

According to the theorem, a lower bound to the number of samples is:
\[ N_{\min} = 2 \cdot BW \cdot T \]  \hspace{1cm} (1.3)

With \( N \) samples of \( f(t) \) we will have \( f[n] \) resulting in the following relation:

\[ f[n] = f(nT_0) \]  \hspace{1cm} (1.4)

where \( T_0 \) equals \( 1/f_s \).

Let's assume without loss of generality, that \( f[n] \) is sparse in the Fourier domain, with a sparsity level denominated by \( S \).

Let,

\[ x = \Psi f \]  \hspace{1cm} (1.5)

where \( f \) is a vector containing all \( N \) samples and \( \Psi \) is an \( N \times N \) matrix called the orthonormal base consisting in this case of complex exponentials of the following form:

\[ [\Psi]_{m,n} = \frac{1}{\sqrt{N}} e^{-j \frac{2\pi}{N} (m \cdot n)} \]  \hspace{1cm} (1.6)

\( x \) is a vector of length \( N \) containing \( f \)'s Fourier coefficients. Since \( f \) is sparse in the Fourier domain, this means that most coefficients of \( x \) are either zero or very close to zero. When this is the case, compressed sensing tells us that we can successfully reconstruct the signal by measuring only \( K \) samples, where \( K \ll N \). The measurements appear as:

\[ y = \Phi f \]  \hspace{1cm} (1.7)

where \( \Phi \) is a \( K \) by \( N \) sensing matrix. Another way of expressing \( y \) is \( y = \Phi \Psi x \), or \( y = \Phi' x \). More specifically, compressed sensing states that we can recover \( f[n] \) with about \( O\left(S \cdot \log\left(N\right)\right) \) samples [3]. This will be analyzed in greater detail in the
Another very important requirement is that the sensing matrix $\Phi$ be statistically incoherent with the orthonormal basis $\Psi$. This requirement has been proven to be met most successfully when $\Phi$ is created randomly. There are several successful candidates for the sensing matrix which will also be discussed later. In this project, however, binary measurements will be used as entries in $\Phi$. This means that $\Phi$ will be created by randomly choosing from a uniform Bernoulli distribution, where there are only two possible outcomes or values, namely, $\alpha = +1$ or $\alpha = -1$, and where the probability $P(\alpha = \pm 1) = \frac{1}{2}$. 
1.2 Analog-to-Information conversion

The majority of successful systems employing compressed sensing do so after the data has already been digitized. This scenario is beneficial in terms of power consumption, for example, in systems of wireless sensor nodes that must transmit data to a central station. Given that these systems are usually battery powered, energy efficiency is a major design constraint. The transmitter tends to be the most power hungry component in said systems. Furthermore the higher the data rate the higher the power consumption will be. This is where CS provides an advantage. However, since the amount of data transmitted is what must be decreased, CS computation can be carried in the digital domain. Therefore, in this type of architecture, the ADC must still sample at the Nyquist rate. Analog-to-Information conversion is the architecture implementation that allows us to compress the signal as it is being sampled [4]. In other words, we simply implement the sensing matrix computation in the analog domain before the signal is sampled by the ADC.

1.3 Hardware architecture

The front-end circuitry of the prototype chip presented in this thesis consists of a low noise programmable gain amplifier (PGA) followed by the signal conditioning block responsible for the implementation of the pseudo-random sequence that represents the sensing matrix Φ. The pseudo-random sequence is generated off chip. After the signal conditioning block comes the 16 bit pipelined triple slope ADC. For this work, we have decided to implement three different modes of operation that will allow us to compare the pros and cons of two different compressed sampling implementations, as well as standard Nyquist sampling. These are illustrated on Figure 1.1. Mode 1 will operate at the standard Nyquist sampling rate, which will sample the input signal at a frequency \( f_{s1} = 1/T_{NQ} \). Mode 2 will operate in uniform sampling mode at a rate lower than Nyquist through the implementation of an AIC architec-
ture that combines the pseudorandom sequence with a bank of M integrators. The effective sampling frequency for Mode 2 is given by $f_{s2} = \left(\frac{M}{N}\right) \frac{1}{T_{NQ}}$. In Mode 2, the bank of M integrators operate in periods of $T_{int} = NT_{NQ}$. Mode 3 will operate by sampling the signal randomly, in what is also referred to as non-uniform sampling. However, Mode 3 must still accommodate a possible highest sampling frequency of $f_{s1}$. 
Figure 1-1: Overall block diagram illustrating multi-mode operation
Chapter 2

Compressed Sensing and AIC

The theory of compressed sensing (CS) will be discussed in more detail along with one of its extensions, namely, analog-to-information conversion. The initial work as presented by in [5] treats signals as discrete vectors $x \in \mathbb{R}^N$. The general problem is centered around the recovery of vector $x$ of length $N$ from $M$ measurements $y = \Phi f$, where $M \ll N$.

2.1 Compressed Sensing

Suppose we have a vector $x \in \mathbb{R}^N$ and we measure it in the form of $y = \Phi'x$ (also depicted in matrix form in Figure 2.1), where each $y_k$ is

$$y_k = \langle x, \varphi_k \rangle, \quad k = 1, \ldots, M \quad (2.1)$$

In order to provide some context, we can assume that vector $x$ contains the Fourier coefficients of signal $f$, where $f$ represents a sampled sequence in the time domain. In this case $\Phi$ represents the sensing matrix and $\varphi_k$ is the $k_{th}$ column of $\Phi$. We are interested in solving for $x$, but since $M \ll N$, the system of equations is said to be underdetermined and thus there exist an infinite number of solutions. Compressed sensing states that if vector $x$ has a small $l_0$ pseudonorm ($\|x\|_{l_0} =$ number of nonzero
terms in $x$), meaning $f$ is sparse in the frequency domain, then the solution to the $l_1$ ($\|\tilde{x}\|_1 = \sum_{i=1}^{N} |\tilde{x}_i|$) minimization problem of

$$\min_{\tilde{x} \in \mathbb{R}^N} \|\tilde{x}\|_1 \quad \text{subject to} \quad \Phi \tilde{x} = y$$

will result in the right vector $x$ the majority of the time. In other words, there is a high probability that $\tilde{x} = x$.

In [3], a vector is said to be $S$-sparse if its support, or in this case the number of Fourier coefficients, is less than or equal to $S$. This also means that $l_0 \leq S$. To better illustrate the relation between the number of measurements $M$, the sparsity level $S$ in a given orthonormal basis $\Phi$, and the probability of successful reconstruction, the following Theorem is presented in [6]. It states that if the number of measurements $M$ obeys the following relation,

$$M \geq C \cdot S \cdot \log(N)$$

and if the constant $C$ is of the form $22(\delta + 1)$, then by minimizing $l_1$ the probability of exact reconstruction for $x$ exceeds $1 - O(N^{-\delta})$.

There are several implications that follow from the above statements. First, the orthonormal basis in which $f$ is sparse does not need to be known a priori, a favorable condition because it allows for a single hardware architecture to successfully encode
a broader class of signals, each sparse in different orthonormal bases. For example, many biosignals such as ECG and EEG are not sparse in the Fourier domain, but they are in the wavelet or Gabor domains [7]. In addition, it is not necessary to know the location, relative distribution or magnitude of the signal's basis coefficients.

It is worthwhile to reiterate the distinction between the sparsity and compressibility of a signal. In [3], a sparse signal is defined as one which is supported by a small set of coefficients in the $\Psi$ domain. A compressible signal is defined as one whose coefficients in the $\Psi$ domain are concentrated near a small set.
2.1.1 Sparsity and Incoherence

As mentioned earlier, the number of samples $M$ that need to be obtained in order to successfully reconstruct $f$ depends partly on the sparsity level of the vector $x$. However, the orthonormal character of the sensing matrix $\Phi$ also has an impact on the number of required measurements. Specifically, [8] introduced the concept of the uniform uncertainty principle which states that $\Phi$ obeys a restricted isometry hypothesis. For example, take $\Phi_T$, $T \subset 1,\ldots,N$ to be the $M$ by $|T|$ submatrix obtained by selecting the columns of $\Phi$ with indeces in $T$, then [9] defines the constant $\delta_S$ to be the smallest quantity for which the following statement is true:

\[
(1 - \delta_S)\|c\|_2^2 \leq \|\Phi_Tc\|_2^2 \leq (1 + \delta_S)\|c\|_2^2
\]  

(2.4)

for all subsets of $T$ and for $|T| \leq S$ with coefficients $(c_j)_{j \in T}$.

Essentially what the above relation represents is a measure of the orthogonality of $\Phi$. The smaller the $S$-restricted isometry constant $\delta_S$ the more the $|T|$ columns of $\Phi$ behave as an orthonormal system. It's further shown in [9] that if the columns of $\Phi$ behave orthogonal then an exact reconstruction is possible. The following theorem is also presented in the previous paper, and it provides a concrete bound on the isometry constant.

**Theorem 2.1.** [9]. Assume that $x$ is an $S$-sparse vector, then if $\delta_{2S} + \delta_{3S} < 1$, the solution to equation 2.2 will be exact. In other words $\hat{x} = x$. 

22
2.1.2 Random Sensing Matrices

It was mentioned that as long as the signal \( f \) was sparse in some orthonormal basis, that by employing the right sensing matrix \( \Phi \) the reconstruction of the sparse vector \( x \) would be exact. In order to take advantage of this characteristic and reduce hardware costs by being able to compressively sample signals that may be sparse in different domains, any subset of the columns of \( \Phi \) of size less than or equal to sparsity level \( S \) must behave as an orthogonal system. Random matrices are known to behave in such a way with great probability. Examples of these are given in [3] and a list of them is included next:

- **Gaussian measurements**
  
  In this case the entries in the \( M \) by \( N \) sensing matrix \( \Phi \) are made up of independently sampled points from a normal distribution with a mean of zero and a variance of \( 1/M \). Then as long as

  \[
  S \leq \frac{C \cdot M}{\log(N/M)} \tag{2.5}
  \]

  the sparsity measure \( S \) obeys the conditions of Theorem 2.1 with probability \( 1 - O(e^{-\gamma N}) \) for some \( \gamma > 0 \).

- **Binary measurements**
  
  Here the entries of the \( M \) by \( N \) sensing matrix \( \Phi \) are independently sampled from a symmetric Bernoulli distribution. In other words, entries in \( \Phi \) consists of \( \pm 1 \) where \( P(\Phi_{ij} = \pm 1) = 1/2 \). Then again, if sparsity level \( S \) obeys equation 2.5, results from Theorem 2.1 are attained with probability \( 1 - O(e^{-\gamma N}) \).

- **Fourier measurements**
  
  In this case, if entries in \( \Phi \) consists of \( M \) normalized randomly selected rows from a \( N \) by \( N \) Fourier matrix, it was shown in [8] that if

  \[
  S \leq \frac{C \cdot M}{\log(N)}
  \]

  then Theorem 2.1 holds accurate. This way of sens-
ing is also known as random sampling and is the implementation on which Mode 3 of our hardware platform is based.

- **Incoherent Measurements**

  For this type of measurements, the rows in the sensing matrix $\Phi$ are chosen uniformly at random from an orthonormal $N \times N$ matrix $U$. The columns of $\Phi$ are then normalized. $U$ can then represent the joint matrix $\Phi \Psi^*$, effectively performing a mapping from the $\Psi$ to the $\Phi$ domain. For this scenario, Theorem 2.1 holds accurate with great probability if

  $$S \leq C \cdot \frac{1}{\mu^2} \cdot \frac{M}{\log(N)^4} \quad (2.6)$$

  if $U = \Phi \Psi^*$, then

  $$\mu := \sqrt{N} \max_{i,j} |\langle \varphi_i, \psi_j \rangle| \quad (2.7)$$

  which is just the measure of coherence between the sensing matrix $\Phi$ and the orthogonal basis $\Psi$ [3]. The smaller the magnitude of $\mu$, the fewer the number of measurements needed.

### 2.2 Analog-to-Information Conversion

Even though up until now the signal of interest has been treated as a discrete vector $f$ of length $N$ for illustrative purposes, theoretically the same concepts apply to a continuous time signal $z$ that possesses the same sparsity characteristics in an orthogonal domain $\Psi$. The idea of analog-to-information conversion seeks to compressively sample said signal $z$ by implementing the sensing matrix $\Phi$ and the compression of the sensed samples before the implementation of the ADC [4]. The motivation behind such implementation is to lower the sampling speed requirements of the converter.
thereby reducing its power consumption and number of non-linearities for a given resolution.

The above processing in the analog domain is achieved by employing $M$ parallel paths where each represents a single row of the sensing matrix $\Phi$. In order to achieve this, a single path, or channel, must multiply the input signal $z$ by a random sequence of 1's and -1's and then integrate this multiplied signal for an amount of time $T_{\text{int}}$, where $T_{\text{int}} = N \cdot T_s$. $T_s$ represents the sampling period that would be required in order to sample $z$ at the Nyquist rate. Figure 2.2 illustrates the above architecture.

As seen in the above figure, the output of each channel only collects a data point once every $N \cdot T_s$ seconds. The interface between this block and the ADC can take one of two forms. One way is to use a sequential approach where the converter samples the $M$ data points being held by the integrating blocks over a period of $T_{\text{int}}$. The other is to use a bank of $M$ parallel ADCs each operating at a sampling frequency of $1/(NT_s)$ in an interleaving fashion. We have implemented the first approach in this work.

Figure 2-2: Graphical representation of A-to-I encoder.
The integrating blocks in Figure 2.2 consists of RC OpAmp integrators in order to maximize the linearity in our pre-processing block given that any non-linearities at this stage can limit the accuracy and success rate of signal reconstruction. Figure 2.3 shows a lower level circuit diagram for one channel in our pre-processing circuit.

Figure 2-3: Circuit level diagram of a single channel in the A-to-I encoder.
2.3 Summary

A brief description of the main concepts in the new sampling theory of compressed sensing have been presented in this chapter. In addition, the analog version of compressed sensing along with its architectural implementation has been discussed. It has been shown that compressed sensing is for the most part a probabilistic theory in the sense that it assigns probabilities to the success of signal reconstruction if said signal meets certain criteria. Although, in some cases, deterministic predictions can be done if more stringent prerequisites are met.
Chapter 3

Front-end amplifier

The design and operation of the analog-front-end (AFE) amplifier will be described in the following sections. The intended application is medical signal acquisition, in particular, ECG signals. The most important specifications for low power ECG front-end amplifiers are the input referred noise, common-mode rejection ratio (CMRR) and the power consumption.

3.1 ECG signals

The ECG signal consists of three main components [10]. These are:

- The differential ECG signal
- The differential electrode DC offset
- The common-mode signal (mostly at 60Hz)

The differential ECG signal contains the PQRS sequence and is thus the component of interest. It ranges between a few tens of uV to about 5mV in amplitude and between .05 Hz to 150 Hz in bandwidth. The differential electrode DC offset can vary between ±300 mV which requires the use of a high pass filter at the input to avoid
amplifier saturation. The 60 Hz common-mode signal is due to interference signals picked up by the human body from power lines and fluorescent lights among others. Current ECG front-end amplifiers require an input-referred noise of $\leq 20uVrms$ over a bandwidth of 150 Hz and a CMRR of at least 100dB. For battery powered devices, a power consumption of $\leq 20 uW$ is desired.

3.2 Standard Instrumentation Amplifier

Following is a brief review of a standard instrumentation amplifier (IA) architecture to serve as a reference point for the reader when analyzing the implementation used in this work. The classic instrumentation amplifier architecture is shown in figure 3.1. This architecture consists of two stages each of which can provide a fraction of the overall gain. If stage 1 has gain $G_1$ and stage 2 has gain $G_2$, then the overall gain is the product $G_1 \cdot G_2$. This design uses resistors to set the gain of the amplifier. Let’s denote $A_{dm1}$, $A_{dm2}$, $A_{cm1}$ and $A_{cm2}$ as the differential and common-mode gain of the first and second stages respectively. If we assume that $R_1 = R_2$, then it’s simple to derive that $A_{dm1} = (2 \cdot R_1)/(R_0) + 1$ and $A_{dm2} = ((R_5/R_3) + 1) \cdot (R_6)/(R_4 + R_6) + (R_5)/(R_3)$. It’s also easy to see that $A_{cm1} = 1$. For $A_{cm2}$, we have the following relation:

$$A_{cm2} = \left( \frac{R_5}{R_3} + 1 \right) \cdot \left( \frac{R_6}{R_4 + R_6} \right) - \frac{R_5}{R_3}$$ (3.1)

If we assume that $R_5 = R_6$ and that $R_3 = R_4$, then from equation 3.1 we can conclude that $A_{cm2} = 0$ ideally. However, resistor matching is never perfect and thus $A_{cm2}$ is never completely zero. The overall amplifier gains are $A_{dmov} = A_{dm1} \cdot A_{dm2}$ and $A_{cmon} = A_{cm1} \cdot A_{cm2}$ respectively. More explicitly, these can also be expressed as:

$$A_{dmov} = \left( \frac{2 \cdot R_1}{R_0} + 1 \right) \cdot \left( \frac{R_5}{R_3} + 1 \right) \cdot \left( \frac{R_6}{R_4 + R_6} \right) + \frac{R_5}{R_3}$$ (3.2)
As mentioned earlier, for ECG applications the CMRR, defined as \( Adm_{ov}/Acm_{ov} \), should be greater than 100dB. However, with the standard architecture shown in figure 3.1 it becomes very difficult to achieve higher than 90 dB of CMRR. This difficulty arises because of several factors, chief among them the mismatch between resistors R3, R4, R5, and R6, and the mismatch in the devices of amplifier A3. The former however presents the bigger limiting factor. From equation 3.3, we can see that in order for \( A_{cm_{ov}} \) to be zero, \((R5 + R3)\) must equal \((R6 + R4)\) and \(R5\) must equal \(R6\). However, taking resistor mismatch into account we can say \((R5 + R3) = \delta_1(R6 + R4)\) and \(R6 = \delta_2R5\). Then, after substituting these in equation 3.3, we can rewrite \( A_{cm_{ov}} \) as \( A_{cm_{ov}} = (R5/R3) \cdot (\delta_1\delta_2 - 1) \), which results in a non-zero \( A_{cm_{ov}} \) so long as \(\delta_1\delta_2 \neq 1\).

In addition to the above, 1/f noise becomes dominant at low frequencies of interest in our application, thus requiring prohibitively large MOS transistors at the input of the IA. Therefore a few innovations to the architecture are necessary.

![Figure 3-1: Standard instrumentation amplifier overall schematic](image-url)
3.3 Low noise front end amplifier

In order to substantially increase the CMRR and still be able to preserve a high input impedance a differential Opamp with two pairs of differential input stages (OPA1) is implemented [17]. The overall block diagram is shown in Figure 3.2. Chopper stabilization has also been included to mitigate the $1/f$ noise at the input stage. Capacitors have been used in place of resistors in order to both reduce thermal noise in the circuit and to achieve better matching and therefore higher CMRR. However, since capacitors do not provide a DC path, active elements have been used to maintain a stable DC operating point. The gain is distributed between both stages with the first stage burning the most power in order to reduce input referred noise. Low pass filters are also included at the output of the circuit to filter out higher frequency noise.
Figure 3-2: Overall block diagram illustrating front end low noise amplifier
3.3.1 Flicker Noise and Chopper Stabilization

Given that the band of interest for ECG signals lies between .05Hz and 150Hz, $1/f$ noise becomes an important factor in limiting the SNR of the front-end amplifier. $1/f$ noise is ubiquitous and can be found anywhere from earthquake vibrations to nerve membranes [11]. Although there is still no universal theory for the occurrence of $1/f$, several models have been proposed based on empirical measurements. Specifically in MOSFETS, $1/f$ noise is attributed to the random trapping of channel electrons in the gate oxide. The resulting effect can be modeled as variance in the transistor current $I_{ds}$ with a power spectral density described by:

$$
\Delta I_{ds}^2 = \frac{KI_{ds}^2}{f_n^1} \Delta f, \quad \text{in subthreshold} \quad n_1 \approx 1
$$

$$
= \frac{KI_{ds}}{f_n^2} \Delta f, \quad \text{in above threshold} \quad n_2 \approx 1
$$

Figure 3.3 illustrates the above relation. The frequency at which the power spectral density of flicker noise and thermal noise are equal is denoted as the corner frequency $f_c$. The transistor output current noise PSD can also be modeled as an equivalent input referred voltage noise with power spectral density $S_v^2(f)$ given by:

$$
S_v^2(f) = \frac{K}{C_o^2WL^n}
$$

where $K$ is a fabrication process dependent parameter, $C_o$ is the oxide capacitance per unit area, $WL$ is the gate area, and $n \approx 1$ [12]. According to the above expressions, the flicker noise input-referred voltage power spectral density $S_v^2(f)$ can be reduced by increasing the transistor’s gate area $WL$. More intuitively, the larger the area the larger the fixed charge $Q_0$ in the oxide and the greater number of traps. This has an averaging effect on the fluctuation of electron trappings and hence lowers flicker noise.

Empirically, PMOS transistors tend to suffer from less flicker noise than NMOS
Figure 3-3: Power spectral density for current noise. Flicker noise is proportional to $1/f$ and thus thermal noise dominates at high frequencies greater than $f_c$. 

$S_n(f) = P(f)$
transistors because electrons in a PMOS need more energy to enter the oxide region [11]. For this reason, the input transistors in the input differential amplifier of this design are PMOS.

Several techniques exist that mitigate amplifier input-referred errors, such as DC offset and low frequency noise. For a thorough review and implementation of said techniques, the reader is referred to [13]. We will present a brief overview of the aforementioned techniques in order to provide some context. The two most widely used techniques to mitigate the errors mentioned above are the autozero (AZ) technique and chopper stabilization (CHS). Correlated double sampling (CDS) involves extra steps but falls under the category of AZ. The main difference between AZ and CHS lies in the fact that AZ is a sampling technique and CHS is a modulating technique. Figure 3.4 shows a closed-loop amplifier offset cancellation arrangement where the input referred offset and 1/f noise at that instant are sampled on capacitor C during phase Φ2, and then during phase Φ1 the input signal is processed. It should be noted that the amplifier could also have a feedback loop to set the closed loop gain for V_m and the AZ operation would still be the same.

While AZ samples the error signals during one phase, CDS does so during every phase. In other words, it samples the input referred DC offset and 1/f noise while simultaneously performing signal amplification. This is illustrated on Figure 3.5, where one half of the circuit composed of capacitors C1 and C2 samples during phase Φ2 while the input signal is amplified through capacitors C1' and C2'. During phase Φ1 the roles are reversed. However, while both techniques are successful in correcting for DC offsets and reducing 1/f noise, their sampling nature causes the amplifier's input referred thermal noise to be aliased down to the baseband.

The amount of noise that is folded into the baseband is proportional to the ratio of noise bandwidth to sampling frequency. If the aforementioned ratio is considerably greater than unity \( \frac{N_{BW}}{f_s} \geq 5 \), then the folded component dominates the input referred noise of the architecture [13]. For this reason, amplifiers employing the AZ
technique should be designed to minimize the input referred white noise component.

The other technique to minimize the effect of 1/f noise is the CHS technique and is the one we employ in our design. This technique relies on the up-modulation of the input signal before it is fed to the amplifier while the 1/f noise and DC offsets are left unaltered. The up-modulation of the input signal is achieved by multiplying it with a square wave. Since we are employing a differential architecture, this effective up-modulation is implemented by using the switching configuration shown in Figure 3.6.

An overall representation of the chopping operation is illustrated by Figure 3.7.
On this figure, \( m_1(t) \) represents the chopping sequence applied at the input of the amplifier. In this case \( m_1(t) \) can be represented by a square wave with chopping frequency \( f_{ch} \). The input signal \( v_{in} \) is subsequently multiplied by \( m_1(t) \) resulting in the signal \( m_1(t) \cdot v_{in}(t) \). Since multiplication in the time domain is equivalent to convolution in the frequency domain, the resulting signal's frequency Fourier transform can be expressed as follows,
The input signal is effectively copied at every odd multiple of the chopping frequency while its magnitude is scaled accordingly. The signal $V_{Mod}(t) = m_1(t) \cdot v_{in}$ is then amplified along with the 1/f noise and input referred offset. At the output of the amplifier there is another chopper modulator to process the output. This operation effectively up-modulates the 1/f noise and DC offset. The up-modulation of $V_{Mod}(t)$ produces a copy of $v_{in}$ in the baseband region as shown by the frequency domain.
Figure 3-7: The different waveforms throughout the chopping operation, illustrated in the frequency domain.
graph of $V_{out}$ at the top of Figure 3.7. If $V_{out}$ is then fed through a low pass filter with a cutoff frequency smaller than the chopper frequency but larger than the signal bandwidth, then $v_{in}$ can be successfully recovered while the 1/f noise and input referred DC offsets are filtered out.

It must be noted, that if the amplifier had a gain of $A_0$ and infinite bandwidth the recovered output signal $V_{out}$ would equal $A_0 \cdot v_{in}(t)$. The effect of finite bandwidth is illustrated by the following example. If the amplifier has a gain of $A_0$ and a bandwidth of $2f_{ch}$, and if we model it as an ideal low pass filter with zero gain beyond $f_{ch}$, then we would obtain a $V_{out}(t)$ equal to \((\frac{1}{\pi})A_0 \cdot v_{in}(t)\). The effective gain would decrease by a factor of 0.8 [13]. However, it should also be noted that amplifiers in real systems do not present an ideal low pass filter response, and instead follow a gradual roll-off, such as 20 dB/dec in the case of single pole systems.

Whereas the AZ and CDS techniques suffer from a large baseband accumulation of foldover broadband noise due to their sampling nature, the CHS technique does not have this drawback. The reason is that the noise replicas are located at odd multiples of the chopping frequency and these scale proportional to $1/n^2$ where $n$ is the harmonic number. Therefore, higher frequency noise replicas contribute less to baseband foldover noise. More precisely, if we denote $S_N(f)$ as the input referred noise power spectral density (PSD) for the amplifier, then after chopping the output noise PSD ($S_{N_{out}}(f)$) will be:

$$S_{N_{out}}(f) = \left(\frac{2}{\pi}\right)^2 \sum_{n=-\infty, n=\text{odd}}^{\infty} \frac{1}{n^2} S_N(f - f_{ch}). \quad (3.7)$$

Another way to illustrate the above result is shown on Figure 3.8. Here we assume that the input referred noise $S_N(f)$ for our amplifier has an effective noise bandwidth of $NBW$, and that $NBW = 2f_s = 2f_{ch}$, where $f_s$ is the sampling frequency used in autozeroing. Then, on Figure 3.8a. is shown what happens to $S_N(f)$ when the AZ technique is applied, while the bottom graph shows what happens when the
CHS technique is used. It is easy to see how rapidly the foldover broadband noise accumulates the higher the ratio of $NBW$ to $f_s$ and how this effect is much less pronounced when implementing the CHS.

![Normalized PSD](image)

Figure 3-8: a) Effect of AZ on input referred white noise PSD. b) Effect of CHS on input referred white noise PSD.
3.3.2 Circuit Design

Given that an ECG signal can be as low as just a few $\mu$V in amplitude, front-end ECG amplifiers usually require gains of 40dB or more. Therefore, a multistage amplifier such as the one shown on Figure 3.2 is implemented. Chopper 1 is preceded by a high pass filter that eliminates the differential DC offset from the signal electrodes. The high pass corner must be placed low enough so that information contained in the low frequency band passes unaltered. Since the desired signal resides between 0.05Hz and 150Hz, the high pass corner should satisfy the following relation:

$$f_{hp} = \frac{1}{2\pi C_0 R_{ad}} < 0.05 \text{ Hz}$$  \hspace{1cm} (3.8)

where $C_0$ will be an off-chip capacitor due to its large size and $R_{ad}$ represents the parallel combination of the resistance from the adaptive elements, shown in Figure 3.6b, and the parasitic resistance formed between the chopping switches and the parasitic capacitance at the input nodes of OpAmp1 [14]. The adaptive element mentioned above, also referred to as a pseudoresistor, consists of the structure shown in Figure 3.9a. The reason it is called adaptive is because it presents a large resistance to small signals and a large conductance to a large voltage applied across its terminals. For small enough signals, this particular structure appears as two reversed MOS diodes in series, thus presenting a large impedance. For a more detailed description of the adaptive element architecture, the reader is referred to [15].

The parasitic resistance formed by the chopper modulator is equivalent to a switched capacitor resistor. Figure 3.9b illustrates this parasitic resistor with an effective value of $R_p = \frac{2T_{dk}}{C_p}$. From the expression for $R_p$ we can see that $T_{dk}$ must be sufficiently high or $C_p$ sufficiently low so that $R_p$ does not limit the effective input impedance of the front-end amplifier. From simulations of transistor parasitic capacitance $C_{gs}$, we obtain a $C_p$ of about 18 fF for the present size of the devices. This translates into a parasitic resistance $R_p$ of 430 Mohms, which is large enough.
The first stage consists of a fully differential non-inverting amplifier where the closed loop gain is given by the following equation:

\[ H_1(s) = 1 + \frac{sR_{adel}C_2 + 1}{sR_{adel}C_1 + 1} \]  \hspace{1cm} (3.9)

where \( R_{adel} \) denotes the value of the pseudoresistors. One important point to note about the first stage is that it has a closed loop DC gain of 2 which would threaten to...
saturate the front-end amplifier if large enough DC signals reached its inputs. However, since we employ a high-pass filter at the input, we need not worry about this. In addition, as will be explained later, in contrast to standard instrumentation amplifier topologies, the first stage provides common-mode signal attenuation. The first stage provides a pass-band differential gain of \( \frac{C_2}{C_1} \), or 10X in this particular application. Values for \( C_1 \) and \( C_2 \) in this work are 100 \( fF \) and 1 \( pF \). Given that at this point the signal of interest has been up modulated by the chopper, the high pass corner of the closed-loop transfer function can be increased. The gain bandwidth product of the first stage Opamp must still be high enough to accommodate the up modulated signal.

The second stage, as depicted in Figure 3.2, consists of a fully differential inverting amplifier where the closed loop transfer function is given by:

\[
H_2(s) = \left[ \frac{C_3}{C_{4/5/6}} \right]
\]

For this stage the pass-band gain becomes \( \frac{C_3}{C_{4/5/6}} \) and depending on which capacitor is switched on, we obtain a programmable gain of 6X, 8X, or 10X. The output of the two stages is subsequently down-modulated by the output choppers and later low pass filtered to eliminate the up-modulated 1/f and thermal noise. The overall front-end amplifier transfer function is then given by \( H_{\text{amp}}(s) = H_1(s)H_2(s) \) or more precisely:

\[
H_3(s) = \left[ 1 + \left( \frac{sR_{\text{adi}}C_2 + 1}{sR_{\text{adi}}C_1 + 1} \right) \right] \left[ \frac{C_3}{C_{4/5/6}} \right]
\]

Thus, the pass band gain results in \( \left[ 1 + \left( \frac{C_2}{C_1} \right) \right] \left[ \frac{C_3}{C_{4/5/6}} \right] \).

In an effort to further analyze the front end architecture, we will examine the transistor level schematic of the first and second stage OpAmps. OpAmp1 consists of a fully differential folded architecture with two pairs of inputs. The schematic for OpAmp1 is shown in Figure 3.10. The purpose behind the use of two pairs of inputs is to maintain a high input impedance and achieve a higher CMRR. Let's analyze the
response of the circuit to common mode signals.

Let \( A_{cm} \) be the common-mode gain of the amplifier. Let’s assume that \( V_{in+} \) and \( V_{in-} \) both see the same change in input signal \( \Delta V_{in} \). Furthermore, let \( G_{m0} \) be the transconductance of each differential pair, which we will assume for now to be perfectly matched. Thus, \( \Delta V_{in} \) will cause a net current \( I_{net} = G_{m0} \Delta V_{in} \) to flow through one branch of each differential pair and \( -I_{net} \) through their mirror branch. This is illustrated on Figure 3.10. Here the current sources formed by \( N1, N2, N3 \) and \( N4 \) along with any load are substituted by impedance blocks \( Z \). In order for \( V_{out+} - V_{out-} \) to change there must be some net current flowing through impedance \( Z \). However, if both differential pairs are matched perfectly then the net current will flow full circle around the loop formed by transistors \( P2 \) through \( P10 \), and thus \( V_{out+} - V_{out-} \) will remain unchanged. Therefore, \( A_{cm} \) should ideally be zero. However, perfect matching is not possible and some net current, albeit considerably small, will still flow through impedance \( Z \). How low we can make \( A_{cm} \) for a specific range of common-mode signal amplitudes depends on the matching of the differential branches. In addition, another factor that can limit this architecture’s CMRR is the limited linear region of the source-coupled pairs. Given that a common-mode signal effectively moves the biasing point of the differential pairs, \( \Delta V_{in} \) must remain within said linear region of operation to maintain good CMRR and linear operation of the closed-loop differential amplifier. Once \( \Delta V_{in} \) exceeds this linear region, a transistor in each differential pair will conduct all of the tail current thus leading to heavy distortion.

The linear region of differential pairs operating in subthreshold and not employing any further linearization techniques such as source degeneration, is limited by fundamental factors such as a process dependent parameter and the thermal voltage. \( V_L \) is given by:

\[ V_L \]
where $k_s$ is usually around 0.7 and $\phi_t$ is the thermal voltage also given by the above expression. Here $k$, $T$, and $q$ represent Boltzmann's constant, the temperature and the electron charge, respectively. $\phi_t$ is around 26 mV at room temperature. Thus, $V_L$ for differential pairs in subthreshold is about 75 mV.

A small signal analysis of the OpAmp1 topology yields the following relation for the differential gain $A_{d1}$:

$$A_{d1} = g_{m1}r_{out1}$$

$$g_{m1} = \frac{I_{bias_1}}{k\Phi_T}$$

$$r_{out1} = r_{cas1}||r_{cas2}||r_{cas3}$$

where $g_{m1}$ represents the transconductance of the source-coupled pair and $r_{out1}$ is the output impedance looking into $V_{out+]$. In the expression for $g_{m1}$, $\Phi_T$ represents the thermal voltage and $k$ is a technology process dependent constant. Lastly, $r_{out1}$ is the parallel combination of output impedances from the cascoded transistors N4 and N3; P3 and P5; and P8 and P10.

The second stage OpAmp2 consists of a standard folded cascode amplifier. The main reason behind the use of this particular architecture is its ability to simultaneously achieve a large output swing and a high gain. Figure 3.12 illustrates the schematic of OpAmp2. The small signal gain of this topology is given by the following relation:
\[ A_{ol2} = g_m2 r_{out2} \]
\[ g_m2 = \frac{I_{bias2}}{k_h} \]  
(3.15)
\[ r_{out2} = r_{cas} || r_{casn} \]

Figure 3-10: OpAmp1 circuit schematic. Vin+ and Vin- are the signal inputs, Vinv+/− are OpAmp1’s inputs that connect in feedback. Vcmfb1 is the common-mode feedback node.
3.4 Low-Pass Filter

We need to provide a low pass filter at the end of our two stage amplifier to ensure that aliasing does not occur later during data conversion and to filter out the up-modulated 1/f noise and input referred DC errors. There are two main types of implementation when it comes to analog filters [16]. One is based on switched-capacitor (SC) circuits and the other is continuous-time (CT) circuits. These two types of architectures represent the two categories among which all analog electronic systems can be divided, when the independent variables are magnitude and time. Figure 3.13 shows the above classification with examples.

The main advantage in SC filters is their high precision. The reason behind such precision is the fact that resistors, which can have tolerances as high as twenty percent are implemented as switching capacitors with a value of \( R \propto 1/(f_s C) \), where \( f_s \) denotes the frequency at which capacitor \( C \) is switched. Figure 3.14.a shows the above implementation of a resistor. This technique allows for a very accurate implementation of filter poles and zeros whose precision depends on the relative tolerance.
of capacitors which can be as high as 0.1 percent. The other variable which is the switching frequency can also be controlled tightly by using a high precision clock. For example, the time constant of a standard switched capacitor integrator as shown in Figure 3.14.b is $\tau = C_1/(f_s C_0)$. However, the main drawback in switched-capacitor systems is the fact that the switching frequency must be many times higher than the bandwidth of the signal of interest in order to reduce aliasing and distortion. The Opamp used in SC circuits must be able to settle to a certain degree of accuracy within half the switching period in some instances, and this then requires active components to be designed with gain-bandwidth (GBW) products much larger than the bandwidth of the signals to be processed.
On the other hand, since CT filters are not clocked, the GBW specifications can be relaxed and this usually results in a much lower power consumption than their SC counterparts for a given signal bandwidth. However, non-linearities in active transconductors and the high tolerances in passives such as resistors limit the accuracy of integrated CT filter coefficients to about 20 to 30 percent.

Different kinds of CT filters include passive RLC filters, op-amp RC, Gm-C and MOSFET-C filters. Passive RLC are the most power efficient but the large size of the components at times restricts their use on integrated platforms. Op-amp RC are

<table>
<thead>
<tr>
<th>Discrete</th>
<th>Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switched Capacitor Circuits</td>
<td>Continuous Time Analog Circuits Opamps, OTAs.</td>
</tr>
<tr>
<td>Clocked Digital Circuits</td>
<td>Asynchronous Digital Circuits</td>
</tr>
<tr>
<td>DSPs, CPUs</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3-13: Classification of information processing systems across independent variables of magnitude and time.
highly linear, but since they are configured in a negative-feedback loop, the op-amp’s gain-bandwidth product must be quite large, thereby increasing power consumption. MOSFET-C filters, on the other hand, reduce their size by employing MOSFETs in the triode region, however, this reduces their margin of linear operation.

In this work, we employ Gm-C filters due to their ease of implementation and low power consumption given their operation in an open-loop configuration. However, given the limited linear range in transconductors, care must be taken in their design to reduce signal distortion.

A standard second order low-pass transfer function with two conjugate poles is
given by the following relation:

\[
H(s) = \frac{1}{\tau^2 s^2 + \frac{\tau}{Q} s + 1}
\]  
(3.16)

\[
H(s) = \frac{1}{(\tau s - p_1) \cdot (\tau s - p_2)}
\]  
(3.17)

\[
p_{1,2} = -\frac{1}{2Q} \pm j \sqrt{1 - \left(\frac{1}{2Q}\right)^2}
\]  
(3.18)

\[
|H(j\omega)| = \frac{1}{|j\tau\omega - p_1| \cdot |j\tau\omega - p_2|}
\]  
(3.19)

where \(\tau = 1/\omega_n\) is the inverse of the natural frequency and \(Q\) is the quality factor of the network.

Figure 3.15 shows the block diagram along with the circuit architecture of the low-pass filter implemented in this work. For this particular architecture \(\tau_1 = C_1/G_{m1}\), \(\tau_2 = C_2/G_{m2}\) and \(Q = \sqrt{\frac{\tau_2}{\tau_1}} = \sqrt{\frac{C_2G_{m1}}{C_1G_{m2}}}\). The filter has been implemented using standard 5 transistor transconductance amplifiers as the \(G_{m1/2}\) blocks.
Figure 3-15: a) Block diagram of second order low-pass transfer function. b) Equivalent $Gm$-$C$ values. c) $Gm$-$C$ circuit diagram.
3.5 Front-End Noise Analysis

Fundamental noise in circuits arises from the discrete and stochastic nature of electrical currents. Electrical currents are made up of electrons which carry discrete packets of charge approximately equal to 1.602e-19 coulombs. Following is a short description of other sources of noise regularly encountered in electronic systems. For a more detailed analysis the reader is referred to any of the following sources [11], [16], [17].

- **Shot noise**
  Shot noise arises in the flow of electrons by way of diffusion, or as minority carriers jump across a potential barrier in a PN junction. This is due to the fact that for electrons or holes to flow across a junction they must have enough energy to overcome the barrier potential. The flow across a junction where electrons become minority carriers can be modeled as a Poisson event. Shot noise has a frequency independent PSD of $2qI_D$.

- **Thermal noise**
  Thermal noise is another fundamental type of noise that results from the random thermal motion of electrons. As such, its PSD is directly proportional to temperature and also frequency independent. The thermal noise in a resistor is given by $4KTR$ if modeled as a variance in voltage.

- **Burst noise**
  This particular source of noise, also known as popcorn noise, is not yet fully understood. However, it is theorized that it might be a result of heavy-metal ion contamination, and is present in some integrated circuits. It’s PSD has been shown to be of the form:
Avalanche noise arises from the avalanche breakdown of a pn junction or the operation of Zener diodes. In this type of noise, the collision from highly energetic holes and electrons with silicon atoms causes generation of hole-electron pairs that give rise to random noise spikes.

3.5.1 First Stage Noise Analysis

We will now estimate the input referred noise for the first stage. We begin by identifying the operation regime of the transistors in the OpAmp1, given that this affects their noise power spectral density (PSD) levels. We must also clarify that the following analysis assumes that the transistors are always in saturation no matter the operating regime. Two distinct operational regimes in MOS transistors are subthreshold and strong inversion. Subthreshold operation happens when $V_{gs} < V_{th}$, where $V_{th}$ represents the threshold voltage of the MOS transistor. Even though in this analysis we will only consider NMOS transistors, the same reasoning applies to PMOS devices. In the subthreshold regime, the current through the NMOS channel is dominated by the process of diffusion, much like in a NPN transistor. Therefore, the respective current PSD from a NMOS transistor in subthreshold can be modeled as shot noise. On the other hand, in the strong inversion regime, $V_{gs} > V_{th}$, and the flow of electrons through the channel is dominated by drift currents. As a result, the respective current PSD of NMOS transistors in strong inversion can be modeled as thermal noise much like that in a resistor. The noisy NMOS device in either operational regime can then be modeled as shown in Figure 3.16.

As shown in the above figure, noise in transistors can be modeled as independent
Figure 3-16: a) Equivalent noise power model of NMOS in strong inversion. b) Noise power model of NMOS in subthreshold.

Current sources in parallel with the devices and as such, these can also be input referred to the gate of the transistors by merely dividing their PSDs by the square of the device’s transconductance.

The total PSD is calculated by shorting the circuit’s inputs and summing the PSDs of all noise generators in the circuit scaled by the square of the transfer function from their location to the output. The total output noise PSD is given as:

\[ S_{total}(f) = \sum_{i=1}^{N} k_i(f)S_i(f) \]  

(3.21)

where \( N \) is the total number of noise generators in the circuit and \( k_i(f) \) is the squared transfer function for the \( i \)th generator. If \( S_i(f) \) has units of \( [A^2/Hz] \) then \( k_i(f) \) would
have units of \([\text{ohm}^2]\). Once we have \(S_{\text{total}}(f)\) we can calculate the total output noise power by integrating over all frequencies. This operation can be expressed as:

\[
\overline{v_n^2} = \int_0^\infty S_{\text{total}}(f) \, df
\]  

(3.22)

where \(\overline{v_n^2}\) represents the total output voltage noise power in \([V^2]\).

Let us first analyze a simple cascode amplifier which will make the analysis of OpAmp1 much simpler. Figure 3.17.a shows a simple cascode amplifier topology with a capacitive load and b) shows the set-up for noise analysis. Since noise signals are small, the circuit can be analyzed with a small-signal model where all DC voltage sources are shorted and DC current sources are opened. In Figure 3.17.b, the output is also shorted to estimate the total output noise current PSD of the amplifier.

Figure 3-17: a) Standard cascode amplifier with capacitive load. b) Cascode amplifier with noise generators included.

58
We proceed by examining the noise contribution of every noise generator to the output. We start with $i_1^2$ and thus we ignore $i_2^2$ initially by opening said current source. Since looking into the source of N2 we see an effective impedance of $1/gm_2$ then $i_1^2$ will effectively appear at the output node and we can say that transistor N1 contributes all its noise power. On the other hand when performing the same analysis on N2, we see that transistor N2 effectively shunts its own noise, given that to $i_2^2$ there are only two paths to take. One path is up the source node of N2 with an effective impedance of $1/gm_2$ and the other path is down the drain of N1 with an impedance of $r_{o1}$. Clearly since $r_{o1} \gg 1/gm_2$, we can effectively ignore the noise from N2 at the output of the amplifier. Another way of looking at it is by referring $i_2^2$ to the gate of N2 as a voltage source and realizing that this then becomes a source degenerated common source amplifier whose effective transconductance is $1/r_{o1}$.

To finally calculate the total output noise power we refer the total output noise current PSD to the input of our cascode amplifier by dividing $i_{out}^2(f)$ by $g_{m1}^2$, multiplying this result by the square of the amplifier’s transfer function and integrating over all frequencies as shown in the following equations.

\[
\overline{i_{out}^2}(f) = \overline{i_1^2}(f) \tag{3.23}
\]
\[
\overline{v_{in}^2}(f) = \overline{i_{out}^2}(f) \overline{g_{m1}^2} \tag{3.24}
\]
\[
\overline{v_{out}^2}(f) = \overline{v_{in}^2}(f) \left[ \frac{(g_{m1}r_{out})^2}{1 + (2\pi f)^2 r_{out}^2 C_L^2} \right] \tag{3.25}
\]
\[
\overline{v_{out}^2} = \int_0^\infty \overline{v_{out}^2}(f) \, df = \int_0^\infty \overline{v_{in}^2}(f) \left[ \frac{(g_{m1}r_{out})^2}{1 + (2\pi f)^2 r_{out}^2 C_L^2} \right] \, df \tag{3.26}
\]

The main conclusion here is that when dealing with a set of cascoded transistors in the output stage of amplifiers, the cascoded transistor contributes an insignificant amount of noise as compared to the bottom transistor and thus we can ignore it. This is another reason why the cascode is such a popular choice given that it provides
a large boost to the gain while keeping the noise contribution limited to that of one device. Now we can proceed to identify those transistors in OpAmp1 that will contribute noise to the output of the amplifier. Looking back at Figure 3.10, we see that each of the source-coupled transistors contributes its noise to the output, as do N2 and N4. The rest of the transistors are effectively cascoded devices and thus we can ignore their noise footprint.

All source-coupled transistors in our design are operated in the subthreshold region. The main reason behind this is that in subthreshold operation we obtain the biggest transconductance for a given bias current. On the other hand, transistors N2 and N4 operate in strong inversion. Since N2 and N4 are not on the signal path, in order to reduce their noise contribution we must minimize their transconductance by increasing their overdrive voltage $V_{ov}$ for a given bias current. To obtain the total output noise PSD we sum the individual contributions given by:

$$i_{OpAmp1}^2(f) = \sum_{k=1}^{6} i_{k}^2(f)$$

$$i_{OpAmp1}^2(f) = 4 \cdot 2qI_{DSat} + 2 \cdot 4KT \left( \frac{2}{3} \right) g_{mN2/4}$$

$$v_{in}^2(f) = \frac{i_{OpAmp1}^2(f)}{g_{mP3/7}^2} = \frac{4 \cdot 2qI_{DSat}}{g_{mP3/7}^2} + 2 \cdot 4KT \left( \frac{2}{3} \right) \frac{g_{mN2/4}}{g_{mP3/7}^2}$$

Given that P3 and P5 on Figure 3.11 are in the subthreshold region, we can approximate their transconductance as $g_{mP3/7} = kI_{DSat}/\phi_{th}$, where $\phi_{th}$ is the thermal voltage and $I_{DSat} = 400 \text{ nA}$. Next we also have $g_{mN2/4} = 2 \cdot 2I_{DSat}/V_{ov}$ where $V_{ov}$ is the overdrive voltage applied to transistors N2 and N4. After substituting for these values in the equation above and obtaining a value for $v_{in}^2(f)$ we must take into account the effect of chopping at the output of the amplifier. From equation 3.7, we know that the white spectral density $S(f)$ contributes foldover wideband noise into the baseband after being chopped. However, each of these replicas has been scaled.
by \((2/\pi)^2 1/n^2\), where \(n\) represents the odd harmonic number. Therefore, it is safe to ignore the spectral density from replicas above the first harmonic. Then, with an effective bandwidth of 11 KHz set by the low-pass filter, an approximation of the input referred noise is given by:

\[
\begin{align*}
\overline{v_{\text{total}}^2}(f) &= \left(\frac{2}{\pi}\right)^2 \sum_{n=1, n \text{ odd}}^{1} \frac{1}{n^2} v_{\text{in}}^2\left(f - \frac{n}{T_{ch}}\right) \\
v_{\text{INrms}} &= \sqrt{\overline{v_{\text{total}}^2}(f) \cdot 11000 \cdot \left(\frac{\pi}{2}\right)}
\end{align*}
\]

(3.30) (3.31)

The above equation returns a \(v_{\text{INrms}}\) of 9 uVrms, which is in good agreement with the input referred noise of 9.3 uVrms we obtain from integrating the noise PSD on Figure 3.18. This result demonstrates that when using the CHS technique, the input-referred thermal noise of the first stage is what contributes to the bulk of the resulting noise power at the output of the low-pass filter.
Figure 3-18: FFT of data points from amplifier simulation employing CHS technique.
3.6 Summary

In this chapter, we have presented the front-end amplifier architecture and have discussed the benefits of using the chopping technique to mitigate 1/f noise. In addition, thermal noise analysis of the first stage amplifier was carried out to point out which are the main contributors to our input referred noise. Overall, the front-end consumes 3.12 uW of power, divided in the following manner: The first stage OpAmp consumes 1.92 uW, followed by the second stage OpAmp with 480 nW. Finally, the low-pass filter and common-mode feedback circuitry consume a combined 720 nW. The amplifier presents an input referred noise of 9.3 uVrms when operating with an effective bandwidth of 11 KHz and 1.98 uVrms with a bandwidth of 500 Hz while providing a gain of 40 dB in both cases. Other gain settings are 35.5 and 38 dB. A CMRR of 110 dB is achieved when operating at 60 Hz.
Chapter 4

Two Step Triple Slope ADC

In this chapter the architecture of a 16 bit two step triple slope ADC is discussed. The ADC is primarily intended for use in a front-end ECG acquisition chip with pacemaker signal detection capabilities. For ECG applications, the ADC operational space is usually in the 8 bit and 100 Hz to 500 Hz bandwidth space. However, pacemaker signals can go up to 5 KHz which increases the bandwidth requirements. Furthermore, this particular architecture can also be exploited in wider operational spaces which are further explored.

4.1 Integrating ADC

4.1.1 Dual Slope ADC

Integrating ADCs are usually used in high precision and low bandwidth operations given that they tend to be slow [18]. They can be implemented in a variety of architectures the simplest of which is the dual-slope configuration as shown in Figure 4.1. The dual-slope ADC consists of an integrator as shown in Figure 4.1 a. Figure 4.1 b shows the output voltage waveform during the conversion operation. It operates in the following manner: At time $t_0$ switch $S_1$ closes and an unknown input $V_{in}$ is applied to the integrator for a pre-determined amount of time $T_0$. After $T_0$, switch $S_1$
open and switch $S_2$ closes applying $V_{ref}$, which is of opposite polarity with respect to $V_{in}$, until the output crosses through zero. The time it takes for $V_{out}$ to cross zero is proportional to the magnitude of $V_{in}$ and can be digitized by counting clock cycles. As can be seen, the conversion time grows proportionately to $2^N$ where $N$ is number of bits.

### 4.1.2 Triple Slope ADC

In order to speed up the conversion time, other architectures have been proposed such as the triple slope configuration [19]. The main block and output waveform for this model are shown in Figure 4.2.
Here, let's assume resistor $R_1 = K \cdot R_2$, where $K > 1$ in order to speed up the downslope ramp. To analyze the triple slope operation it is easier to map input voltage $V_{in}$ to input current $I_{in}$ and analyze the conversion in terms of the following variables.

Figure 4-2: a) Integrator block in Triple-Slope ADC. b) Output voltage waveform during conversion.
\[ I_{in} = \frac{V_{in}}{R_1} \]  
\[ I_{ref} = \frac{V_{ref}}{R_2} = \frac{K \cdot V_{ref}}{R_1} \]  
\[ I_{refm} = \frac{K \cdot V_{ref}}{R_1 \cdot 2^m} = \frac{I_{ref}}{2^m} \]

For analysis purposes we can think of \( V_{ref} \) as the maximum input voltage \( V_{in_{max}} \). Up to \( t = t_1 \) the triple slope operation is the same as in dual-slope. At \( t = t_1 \), \(-V_{ref}\) is applied, however since \( R_2 \) is smaller by a factor of \( K \) the slope of the downward ramp will also be larger by the same factor. This means that \( V_{out} \) will cross zero much faster than in the dual slope version and hence we will have \( K \) times less counts between \( t_1 \) and \( t_2 \) throughout the entire input range. Thus, we have basically traded precision during the downward slope for speed. In other words, between \( t_1 \) and \( t_2 \) we can only resolve the \( N - \log_2(K) \) MSBs. In order to resolve the remaining \( \log_2(K) \) bits we need to continue integrating past the zero crossing until the next clock edge. Therefore, \( V_{out} \) at time \( t_2 \) contains the residual information needed to resolve the \( \log_2(K) \) LSBs. At time \( t = t_2 \), switch \( S_2 \) opens and switch \( S_3 \) closes and we integrate with \( V_{ref}/2^m \) until \( V_{out} \) crosses zero again. The maximum number of counts in this third ramp will be \( K \) counts. As can be noticed, we have had to reduce the ramp slope by a factor of \( 2^m \) where

\[ m = \log_2(K) \]

in order to obtain the overall \( N \) bit precision. This division effectively reduces the exponential factor by half for the conversion time, going from \( T_{total} \propto 2^N \) to \( T_{total} \propto 2^{N/2} \).
4.1.3 Triple Slope Optimal Conversion Time

The conversion time analysis can be broken into two parts. The first part concerns the first upward ramp resulting from the integration of $V_{in}$ for a preset time of $T_0$. The second part concerns the rest of the conversion. First let us set the resolution of the triple slope converter to $N$ bits, with $N$ being even. Next let us analyze the second part of the operation defined within the time interval from $t = t_1$ to $t = t_3$, as depicted in Figure 4.2b. In the dual slope ADC, the equivalent maximum time for this portion of the operation, which we define as $T_{2nd} = t_3 - t_1$ equals $2^N$. In the triple slope ADC, we devide the $N$ bits into $M + P$ bits, where $M$ bits are resolved by the downward ramp and the remaining $P$ bits are resolved by the final upward ramp. This new division provides the following expression for $T_{2nd}$:

$$T_{2nd} = 2^M + 2^P$$

In order to minimize $T_{2nd}$ we must make $M = P$, and thus $M = N/2$. This makes $T_{2nd_{\text{min}}} = 2 \cdot 2^{N/2} = 2^{(N/2)+1}$. As for the $t = 0 \to t_1$ portion of the conversion, in order to reduce $T_0$ by a certain factor, $R_1$ must also be reduced by the same factor. However, since we wish to use the same value of $R_1$ to produce the reference current $I_{\text{ref}}$ that resolves the $M$ MSBs, we have decided to make $T_0 = 2^{N/2}$. This operation results in $T_{\text{total}} = 3 \cdot 2^{N/2}$, where $T_{\text{total}}$ represents the maximum value for $t = t_3$ as depicted on Figure 4.2b. This maximum value of $T_{\text{total}}$ represents an effective speed-up of a factor of $(2/3) \cdot (2^{N/2})$ for the maximum conversion time as compared to the dual slope ADC architecture.
4.2 Two Step ADCs

Two step ADCs offer power consumption and area savings in comparison to flash converters. Flash converters perform an $N$ bit conversion in one clock cycle by employing $2^N - 1$ comparators that compare the input voltage to every one of the $2^N - 1$ quantization levels. The two step algorithm performs the conversion in two or three clock cycles by distributing it across two stages. One stage resolves the $M$ MSBs and the other stage the $P$ LSBs. Figure 4.3 shows the block diagram for a two step $N$ bit converter. Both coarse and fine ADCs are usually implemented as flash converters. This means that a total of $2^M + 2^P$ comparators are used instead of $2^{M+P}$ before. The timing diagram is also shown in Figure 4.3. Here it is assumed that each of the listed operations can be completed in half a clock cycle. Once the input signal has been sampled, the coarse ADC resolves the $M$ MSBs. This $M$ bit word is fed to the DAC which then produces the respective analog value. The difference between $V_{in}$ and $V_{dac}$ constitutes the residue $V_{res}$ which is computed in the following phase of the clock. This residue may be later amplified by a factor of $K$ and finally resolved to $P$ LSBs by the fine ADC. The amplification by $K$ helps ease the requirements for the $P$-bit converter. If $K$ is made equal to $2^M$, then the residue signal is restored to full range and thus the same reference signals can be used in the $P$-bit converter. Figure 4.4 shows the residue generator transfer characteristic for a 2 bit stage.
4.2.1 Two Step ADC Errors

Deviations in the coarse ADC and DAC transfer characteristics can cause a series of conversion errors. In particular, DNL errors in the coarse ADC cause the residue generator to fall short from or go past the maximum residue range. This is better illustrated in Figure 4.5. For this case we assume that the DAC is ideal. The first ADC error causes the residue to go out of range and thus the fine ADC continues to output a code of 11..1. However, at the break point the residue comes back to the correct value. Therefore, an error in the coarse ADC only causes local conversion errors due to the fine ADC’s inability to sense outside of its input range. Errors in the DAC however remain present throughout an entire MSB interval, as shown in Figure 4.6. They effectively introduce an offset in the LSBs. Given that the DAC's
$DNL(i)$ affects the LSBs within the residue interval pertaining to $MSB(i)$, it is often required to have greater accuracy than the coarse ADC.

As a way of dealing with the coarse ADC's errors, the fine ADC is often made to have an extra bit of precision. The purpose of this extra bit is to account for the out of range residue that results from the coarse ADC's INL. This technique is further explained in section 4.3.3.

Figure 4-5: (a) Real ADC transfer characteristic. (b) Residue with ideal DAC.

Figure 4-6: (a) Real DAC transfer characteristic. (b) Residue with ideal ADC.
4.3 Pipeline ADCs

In this section we will provide a brief overview of the standard pipeline configuration and its operation. Pipeline converters work most optimally in a space of low-to-medium resolution and medium-to-high sampling speeds [18]. They can implement sequential algorithms at greater speeds by expanding in space what was sequentially done throughout time.

4.3.1 Standard Architecture

A standard pipeline block diagram is shown in Figure 4.7. Figure 4.7.b illustrates the block diagram configuration of each pipelined stage. As it shows, each stage contains a sample and hold followed by a subADC. The subADC output bits represent the digital output of that individual stage, which is read by the DAC to produce an output $V_{dacj}$. The difference between $V_{in}$ and $V_{dacj}$ is amplified by a factor of $K_j$ and output to the next stage. From Figure 4.7.a, if each stage $subADC_j$ produces $N_j$ bits then the total resolution of the pipeline converter is $N_{total} = \sum_{j=1}^{K} N_j$ bits. The digital logic block groups and then outputs the $N_{total}$ bit word, at first with a latency of $(K+1)$ clock cycles and then every $T_{clk}$ after that. This initial delay must be taken into account when using the converter in a feedback loop.
4.3.2 Accuracy Requirements

As previously shown, the S&H, subADC, DAC and the residue generator can all contribute to a decrease in absolute accuracy. However, the number of bits to be resolved decreases as we go down the pipeline which in turn relaxes the accuracy requirements for the subsequent stages. This makes the first stage the most critical one, especially the S&H block whenever it does not provide any amplification.

4.3.3 Digital Error Correction

The subADC errors mentioned previously can be corrected with digital techniques [20]. The ability of digital techniques to correct for ADC errors relies on the fact that as long as the difference between the sDAC (subDAC) and $V_{in}$ is within the input range of the subsequent stages, the information needed to resolve the rest of the LSBs is still intact. The problem arises due to the fine ADC's inability to
resolve an input value beyond its input range. Two popular ways of implementing
digital correction consists of adding extra bits to the fine ADC in order to increase its
input range and the 1.5 bit implementation. Both rely on added bits for redundancy.
Figure 4.8 shows the residue voltage waveform for the case where an extra bit is added
to the fine ADC in a 4 bit two step architecture. In this example the coarse ADC has
2 bits while the fine ADC has 3. Once the residue goes beyond the range of a 2 bit
converter the fine ADC goes from having an output code of 011 to 100. The MSB in
the fine ADC indicates whether or not 01 should be added to the coarse ADC output.

The residue waveform shown in Figure 4.9 illustrates the concept of the 1.5 bit
implementation in pipelined converters. In this particular case the coarse ADC only
has one bit of resolution and in order to keep the residue from going out of range a
third region is added around the original break point. A 1 bit ADC divides the residue
waveform into two regions whereas a 2 bit ADC divides the residue into four regions.
Since this error correction method splits the residue waveform into three regions, it
is equivalent to a resolution of 1.5 bits. The middle region is $V_{th_H} - V_{th_L}$ wide where
for analysis purposes we have $V_{th_L} < 0$. If $V_in < V_{th_L}$ then we have a certain 0 digital
output for that particular stage and $V_{ref-}$ is subtracted from the input. If $V_in > V_{th_H},$
then we have a certain 1 digital output and $V_{ref+}$ is subtracted from $V_in$. In the case
where $V_{th_L} \leq V_in \leq V_{th_H}$ then $V_in$ is considered to be in the uncertain and therefore
is left as is. In this case the decision is postponed until a later stage. This means
that when $V_in$ is in the uncertain region, it is amplified by two, without the addition
or subtraction of $V_{ref}$, before being passed on to the next stage. The digital output
of a 1.5 bit stage consists of 10 for a certain 1, 00 for a certain 0 and a 01 for the
uncertain region, 11 is not used because the resolution is of only 1.5 bits.
Figure 4-8: Residue waveform with extra bit of precision in fine ADC

Figure 4-9: 1.5 bit residue waveform in pipeline converters.
4.4 Two Step Triple Slope Architecture

We have combined aspects of the two step, pipeline and triple slope architectures into a two stage 16 bit converter. This takes advantage of the integrating ADC's high linearity, the speed-up provided by the triple slope and pipeline configurations, and the area and power savings provided by the two step architecture.

4.4.1 Circuit Diagram

The overall circuit diagram for the analog blocks is illustrated in Figure 4.10. The source voltages for $V_{ref}$ and $V_{ref}/16$ are produced by an R2R ladder in a setup that will be illustrated in greater detail later. In the case of 16 bit operation, the 8 MSBs are resolved by the first stage and the 8 LSBs by the second stage. $OpAmp_1$ and $OpAmp_2$ consist of folded cascodes each with switched capacitor common-mode feedback circuits. $OpAmp_1$ requires the more stringent specifications with respect to noise and gain given that it is resolving the MSBs. Therefore, the bulk of the analog power consumption is consumed by the first stage. Next let us analyze the overall operation of the converter. Given that the architecture is fully differential we can analyze the single-ended version of the circuit without loss of generality, to better illustrate its principles of operation. Figure 4.11 shows the single-ended version of the pipeline triple slope configuration.
Gain = 1 + Cbig/Csmall = 2^k
M = 2^8/Gain
Vf+ = 900 mV
Vf- = 300 mV
Vmid = 600 mV

First Stage in Pipeline

Second Stage in Pipeline

Figure 4-10: Overall circuit diagram of analog blocks in ADC
Gain = 1 + Cbig/Csmall = 2^k
M = 2^B/Gain
+Vref = 900 mV
-Vref = 300 mV
Vmid = 600 mV

Figure 4-11: Single-ended version of analog blocks in ADC
4.4.2 Basic Operation

At the start of conversion, the entire circuit is reset in order to set the initial conditions. Since the output of OpAmp1 and OpAmp2 cannot swing rail-to-rail due to the presence of output transistors that must remain biased in saturation, \( V_{out1} \) is preset to \( V_{ref+} \). Having a power supply voltage of 1.2 V, we have chosen \( V_{ref+} = 900 \text{ mV} \) and \( V_{ref-} = 300 \text{ mV} \) so that \( V_{out1} \) can have a maximum swing of 600 mV. This particular setup provides us with a maximum differential swing of 1.2 V. This DC bias is also presented across level-shifting capacitors \( C_{comp} \) so that the comparator can detect the crossing at the common-mode voltage level of 600 mV.

Switch \( S_a \) closes and an input current proportional to \( V_n \) charges the integrating capacitor \( C_{big} + C_{small} \) for a predetermined amount of time. Switches \( S_b \) and \( S_c \) will each close and open in a sequence typical of the triple slope integrator architecture explained in section 4.1.2, thus resolving the 8 MSBs. At the end of the triple slope conversion we are left with a residual output voltage that we denote as \( V_{res} \). If we were to continue with the 16 bit conversion using the quadruple slope, we would eventually need a reference equal to \( V_{ref}/2^{12} \) and in addition this reference would have to be precise to 16 bits. This requirement certainly complicates the design of the reference circuitry and therefore we have taken the following approach, which is to amplify \( V_{res} \) by a factor of 16. This alleviates the precision constraint on the reference circuitry but it does pose a bigger challenge to the OpAmp design in terms of necessary open loop gain. This requirement will be reviewed in greater detail in a subsequent section.

Once \( V_{res} \) has been amplified and sampled by \( C_{samp} \), \( C_{samp} \) is connected across OpAmp2 and the subsequent triple slope process ensues.
4.4.3 Zero Crossing Detector Design

Since we want to know the time when the integrator output crosses a given level, we must employ a zero crossing detector (ZCD) in place of a standard clocked comparator that compares two voltage levels every clock cycle. The schematic for the ZCD is shown in Figure 4.12. During one cycle of operation, we have four distinct zero crossings between the first and second stages. The ZCD consists of a single-ended differential preamplifier followed by two inverters each optimally sized to detect a specific crossing in each stage. During the first ramp in stage one, the discharging current is $I_{ref}$ and during the second $I_{ref}/16$ is used, thus decreasing the ramp’s rate of change by a factor of 16.

The signal $CompOut$ tells the state machine to stop the corresponding counting sequence. Since $Q_1$ triggers at the first crossing and $Q_2$ triggers at the second, $CompOut$ must depend on these two signals. The logic controlling switches $SwQ1$ and $SwQ2$ implement this functionality. When both $Q_1$ and $Q_2$ are down, this means that $(V_{in+} - V_{in-})$ is positive and thus we want $CompOut$ to follow $Q_1$ in the next transition. If $Q_1$ and $Q_2$ are both up, then $(V_{in+} - V_{in-})$ is negative and we want $CompOut$ to follow $Q_2$ instead. The problem arises when $Q_1$ is low and $Q_2$ is high, or vice versa. However, we notice that between $Q_1$ and $Q_2$ a hysteresis curve is formed that is distinguishable between each transition. One dynamic logic gate, with output signal $In0$ keeps switch $SwQ1$ closed so that $CompOut$ follows $Q_1$ until both $Q_1$ and $Q_2$ have gone from low to high, and another, with output signal $SwQ2$ keeps the corresponding switch closed so that $CompOut$ follows $Q_2$ until both $Q_1$ and $Q_2$ have gone from high to low.
Figure 4-12: Schematic for dual phase zero crossing detector
4.4.4 OpAmp Design

The circuit schematic for the OpAmp design used is shown in Figure 4.13. Given the need to provide a precise closed-loop gain of 16 for residue amplification and present low input referred noise, the first stage OpAmp presents the toughest requirements. Furthermore, given the fact that a large output swing is needed, it is best to achieve the desired open loop gain by using a two stage OpAmp design.

Integrating ADCs are generally used in applications where the sampling speeds are relatively low. When used at high speeds, finite bandwidth and open loop gain in the OpAmp introduce errors. As shown in Figure 4.2b, the output curve of the integrator consists of a ramp with different rates of change depending on the state of operation. As the rate of change for a ramp increases, the non-idealities around the switching points of the curve become more apparent and can introduce conversion errors. Figure 4.2b shows an ideal integrator output curve, a real waveform is depicted in Figure 4.14. Here, it can be seen how at the switching instant the waveform appears to shoot up and then follow a first order transient profile as it settles back to a ramp. The small overshoot at the switching instant is dependent on the OpAmp’s first stage’s transconductance \( Gm1 \) as will be shown next. Any non-linearity in \( Gm1 \) will propagate through the rest of the conversion.
Figure 4-13: OpAmp schematic
Figure 4.14: Ideal versus non-ideal behavior of integrator output waveform at switching times.

4.4.5 OpAmp Helper Circuits

As mentioned in the preceding section, any non-linearities present in the transconductance of the OpAmp's first stage will propagate from the switching instant onward along the conversion cycle. The switching perturbation's dependence on the OpAmp's first stage's $Gm1$ can be better illustrated in Figure 4.15. Here, the overall circuit is depicted on the left hand side of the figure. The right-hand side shows the equivalent of the inside of a two stage OpAmp. Right before switching from one ramp to another,
there is a constant current \( I_{in} = \frac{Vin}{R} \) flowing through \( C \), thus giving an output voltage rate of change of \( \frac{I_{in}}{C} \). Since the OpAmp’s transconductance is finite, the first stage must provide the necessary current to charge the compensating capacitor \( Cc \) at the same rate of \( \frac{I_{in}}{C} \), with the rest if \( I_{in} \) being sunk by the OpAmp’s output stage. The input voltage \( V_x \) to the OpAmp that is needed to charge \( Cc \) is dependent on the first stage’s transconductance and is given by the following equation:

\[
V_x = \frac{-V_{in} \cdot Cc}{gm \cdot R \cdot C}
\]  

(4.6)

As the ramp changes direction, a new \( I_{in} \) starts flowing through \( C \) which results in a corresponding change in \( V_x \). It is this change that propagates to the output of the OpAmp as the output ramp switches direction. Looking back at equation 4.6, \( C \) is determined by thermal noise requirements while \( R \) is limited by the sampling frequency and resolution, so we can regard these as hard constants. Then the only way to minimize \( V_x \) would be to either decrease \( Cc \) or increase \( Gm \). However, these two variables are involved in the OpAmp’s stability, and changing them in the desired direction would actually be moving the amplifier towards instability. Obviously, increasing \( Gm \) also increases the amplifier’s power consumption given that transistors P1 and P2 operate in the subthreshold regime.

We have thus devised a way to minimize the change in \( V_x \) without having to change either \( Gm \) or \( Cc \). The circuits used to implement the minimization of \( V_x \) are shown in Figure 4.13 inside the green dotted lines. They essentially provide a second set of inputs through which \( I_{in} \) can also flow. Focusing on node \( Vin2_{+} \), we can see that transistors N3 and N2 form a cascode. Therefore, node \( Vin2_{+} \) is a low impedance node since it is looking into N2’s source. Furthermore, the feedback loop formed by transistors N3, P4, and N5 further decreases the impedance looking into \( Vin2_{+} \) by the loop gain. The fact that \( Vin2_{+} \) is such a low impedance node makes it also a good virtual ground. When input current \( I_{in} \) flows through N3, it is mirrored by transistor N9 of the input differential pair branch. N9 thus sinks the current needed
Figure 4-15: Overall circuit depicted on the left. Equivalent of inside of OpAmp on the right.

to charge or discharge the compensating capacitor between the two stages. Since the compensating capacitor and the integrating capacitor might not be the same, which is the case in our design, N3 and N9 of the current mirror must be sized by the same scale factor $C_{int}/C_{comp}$, or in other words:

$$\frac{W_{N3}}{W_{N9}} = \frac{C_{int}}{C_{samp}}$$

(4.7)

where $W_{N3}$ and $W_{N9}$ are the widths of transistors N3 and N9 respectively. Since any differential current through the compensating capacitors $C_{comp}$ is provided or sunk by transistors N8 and N9, the input nodes $Vin_{+}$ and $Vin_{-}$ of the differential pair do not have to change in order to supply said current thus effectively minimizing any non-linear overshoots at the output of the OpAmp on each switching instant. This technique increases the OpAmp's effective transconductance as seen from its inputs $Vin_{+}$ and $Vin_{-}$.
The effect of the above technique was simulated and results are shown in Figures 4.16 and 4.17, where the nodes for which the waveforms are depicted are referenced to the OpAmp schematic on Figure 4.13. Figure 4.16 shows the OpAmp operating without the helper circuits. More specifically, Figure 4.16a shows the differential signal at the input of the OpAmp over the switching sequence for integrating currents $I_{in}$, $I_{ref}$ and $I_{ref}/16$. These waveforms can be compared with those of Figure 4.17d and 4.17f, where it is shown that the auxiliary circuits contribute to a reduction of the differential input voltage of the OpAmp by a factor of 9X. Essentially, both figures depict the OpAmp response to the same input signal value, but without the helper circuits the OpAmp inputs must change by 72 mV at one switching point, while in the presence of the auxiliary circuits, said change is reduced to 8.3 mV.
Figure 4-16: Without auxiliary circuits. a) Differential signal $V_{in+} - V_{in-}$. b) Signals $V_{in+}$ and $V_{in-}$. c) Differential signal $V_{out+} - V_{out-}$. d) Signals $V_{out+}$ and $V_{out-}$. 
Figure 4-17: With auxiliary circuits. a) Differential signal $V_{in2+} - V_{in2-}$. b) Signals $V_{in2+}$ and $V_{in2-}$. c) Signals $V_{out+}$ and $V_{out-}$. d) Signals $V_{in+}$ and $V_{in-}$. e) Differential signal $V_{out+} - V_{out-}$. f) Differential signal $V_{in+} - V_{in-}$. 
4.4.6 OpAmp Open Loop Gain and Settling Time

During residue amplification, the first stage operates as a multiplying DAC with a closed-loop gain of 16. At this point, the 8 MSBs have been resolved by the first stage, and the amplified residue is sampled so that the second stage can resolve the 8 remaining bits. However, if OpAmp1 does not have enough open-loop gain, or enough time is not given for it to settle to within a certain degree of accuracy, errors are introduced. Figure 4.18 shows the residue amplification waveform and Figure 4.19 shows OpAmp1's open loop transfer function.

As it is well known, the actual closed-loop gain of a system that uses negative-feedback is given by the following equation:

\[ G_c(s) = \frac{A(s)}{1 + FA(s)} \]  

(4.8)

where \( F \) is the feedback factor and \( A(s) \) is the open-loop amplifier gain. For a typical open-loop, single pole transfer function \( A(s) \), such as the one in Figure 4.19, with a time constant \( \tau_{ol} \), the step response for \( G_c(s) \) is given by the following expression:

\[ v_o(t) = \frac{1}{F} \left[ 1 - \exp \left( -\frac{A_0 F t}{\tau_{ol}} \right) \right] \]  

(4.9)

where \( A_0 \) is the DC open-loop gain of the amplifier, or \(|A(0)|\). As we can see, the effective time constant for the closed-loop network has been reduced by a factor of \( A_0 F \). Moreover, we note that if we keep \( \tau_0 \) and \( F \) constant, there is a fundamental trade-off between settling accuracy and how soon we sample the output. Thus, we must calculate how long we have to wait in order to assure a specific degree of accuracy.

First, we must estimate how much error is permissible. Let us assume that we have a gain error of \( \delta \) so that the effective closed-loop gain becomes \( 1/F + \delta \). Let us further define \( v_{res} \) as the maximum magnitude of the residue, which in this case is equal to \( v_{lsb} \cdot 2^8 \). Given the above assumptions, the magnitude of the amplified residue becomes \( v_{res} \cdot (1/F + \delta) \), which results in an error of \( \delta v_{res} \).
In order to illustrate how this error affects the linearity, assume that $\delta = 1$, then we have an error of $v_{lab} \cdot 2^8$, in other words, an 8 bit error. Therefore, in order to maintain 16 bit linearity, we can only allow an error of $v_{lab}/2$, which translates to an effective $\delta$ of $1/512$. In conclusion, the closed-loop network must settle to within 0.195%, or $\exp\left(-\frac{A_0E\tau}{\tau_{ol}}\right) = 0.00195$. If we substitute $A_0 = 10^{5.2}$, $\tau_0 = 1/(2\pi \cdot 16)$ and $1/F' = 16$, we obtain a settling time of $t = 6.26\mu s$, which translates to $12T_{clk}$ for $T_{clk} = 500\mu s$. 
Figure 4-18: MDAC configuration during residue amplification.
Figure 4-19: OpAmp1 open loop transfer function and phase.
4.4.7 Integrator Noise Analysis

Given the high target resolution of 16 bits, the thermal and shot noise will be the theoretical limiting factors on the effective number of bits (ENOB) that can be achieved. Therefore, it is necessary to perform a detailed analysis to determine which are the main contributors and if any design techniques can be employed to reduce said contributions.

We are interested in the total noise power measured at $V_{out1}$, at the end of the input signal integration phase. Figure 4.20 shows the first phase of integration. At $T_0$, in order to preserve a 16 bit SNR of 96 dB, the following relation must be satisfied:

$$\sigma_{out1}^2 < \left(\frac{V_{LSB}}{(2\sqrt{2})}\right)^2$$

(4.10)

for $V_{LSB} = V_{FS}/2^{16}$, where $V_{FS}$ is the full scale voltage, which in this case is 1.2V. $\sigma_{out1}^2$ is the variance around the ideal value of $V_{out1}(T_0)$.

![Figure 4-20: Output waveform during integration of input signal with added noise.](image)

95
The chief noise sources that can limit the effective resolution of the converter are the OpAmp, the R2R ladder and the ZCD. Let us analyze each as it pertains to $\sigma_{out}^2$. We can treat the OpAmp, R2R ladder and the ZCD as individual elements each with their respective input referred noise PSDs. Figure 4.21 illustrates this idea. For the moment, let us concentrate on noise generators $V_{R2R}(f)$ and $V_{OpAmp}(f)$ since these two are the ones that contribute to noise across $C_s$ at $T_0$.

![Figure 4-21: First stage with respective noise generators](image)

The transfer functions from $V_{R2Rrms}$ and $V_{OpAmprms}$ to $V_{out1}$ are:

\[
H_{r2r}(s) = \frac{1}{sRC_1} \tag{4.11}
\]

\[
H_{OpAmp}(s) = 1 + \frac{1}{sRC_1} \tag{4.12}
\]

Given that the time constant for $H_{r2r}$ is $\infty$, $V_{R2R}^2(f)$ cannot be treated as a wide sense stationary (WSS) process because the circuit never reaches steady-state. Therefore, in order to correctly estimate $\sigma_{out}^2$ we must perform a non-stationary noise analysis. On the other hand, we can see that $H_{OpAmp}(s)$ in addition to a non-stationary component also presents a wide-sense component with unity gain. For a more detailed analysis of non-stationary processes the reader is referred to the following sources [22], [23].

96
We will begin by analyzing $\overline{V_{R2R}^2}(f)$ as if it were driving an ideal integrator as shown in Figure 4.22. We want to know what $\overline{V_{O2r}^2}$ (variance of $V_{out1}$ due to $\overline{V_{R2R}^2}(f)$) will be at time $T_0$. From [24], we can use the following equation to calculate $\overline{V_{O2r}^2}$:

$$\overline{V_{O2r}^2}(t_i) = \frac{2kT G_n}{C^2} t_i u(t) \quad \text{for} \quad t_i \ll T_0/2$$

$$= (1/2) \frac{S_i(0)}{C^2} t_i u(t)$$

(4.13)

(4.14)

where $G_n$ represents the effective white noise transconductance (i.e. $1/R$ for the thermal noise current in a resistor $R$), and $t_i$ is the integration time, which in our case is equal to $16T_{dk}$. If we now use the noise current PSDs equivalents of $\overline{V_{R2R}^2}(f)$ and $\overline{V_{OpAmp}^2}(f)$, we are then able to estimate their noise contributions at the output of the integrator. The expressions for the above quantities are given as follows:

$$\overline{i_{2r}^2}(f) = \overline{V_{R2R}^2}(f)/R^2$$

(4.15)

$$\overline{i_{OpAmp}^2}(f) = \overline{V_{OpAmp}^2}(f)/R^2$$

(4.16)

Finally, substituting $\overline{i_{2r}^2}(0)$ and $\overline{i_{2r}^2}(0)$ for $S_i(0)$ in equation 4.14, we obtain the non-stationary component of the variance at $V_{out1}$.

$$\overline{v^2}_{Non-Stat} = (1/2) \frac{\left[ \overline{i_{2r}^2}(0) + \overline{i_{OpAmp}^2}(0) \right]}{C^2} t_i$$

(4.17)
Next, we need to calculate the WSS component of $V_{OpAmp}^2(f)$. First, we must clarify that this noise component must be assessed under the circuit configuration shown in Figure 4.23 which shows the residue amplification phase. This is because during residue amplification, we must wait until the circuit reaches steady state so that the correct value is sampled across $C_s$, and thus $V_{OpAmp}^2(f)$ also gets amplified along with the residue and ultimately sampled across $C_s$. In order to calculate the output noise contribution we must integrate the following expression over all frequencies:

$$
\bar{v}_{wss}^2 = \int_0^\infty V_{OpAmp}^2(f) \left| \frac{L(f)}{1 + L(f)} \right|^2 df
$$

(4.18)

where $|L(f)|$ is the magnitude of the loop gain for the feedback configuration around OpAmp1. Since the open loop gain of OpAmp1 can be approximated as a first order transfer function, the above integral can also be approximated as:

$$
\int_0^\infty V_{OpAmp}^2(f) \left| \frac{L(f)}{1 + L(f)} \right|^2 df = \frac{V_{OpAmp}^2(0)}{1 + L(f)} \cdot NBW_{16} \cdot \pi/2
$$

(4.19)

where $NBW_{16}$ represents the bandwidth of the loop gain $L(j2\pi f)$ of the circuit in Figure 4.23.

Finally, since each noise generator in this example can be considered as an independent stochastic process, we can sum their variances to obtain the final noise power contribution and relate this result to the initial constraint for 16 bit SNR as:

![Diagram][1]

Figure 4-22: With noise generator driving integrator.
\[
\overline{v_\text{total}^2} = \overline{v_\text{Non-Stat}^2} + \overline{v_\text{wss}^2}
\]

(4.20)

\[
\overline{v_\text{total}^2} < \left( \frac{V_{\text{LSB}}}{2\sqrt{2}} \right)^2
\]

(4.21)

Figure 4-23: First stage during residue amplification phase with noise generator included.

where \(V_{\text{LSB}}\) in equation 4.21 represents the magnitude of the LSB referred to the output \(V_{\text{out1}}\), or equivalently \(1.2/(2^{16})\) \(V\). Given our result from Appendix 1 and the thermal spectral density from the R2R ladder, we can estimate \(\overline{v_\text{Non-Stat}^2}\) and \(\overline{v_\text{wss}^2}\). To calculate \(\overline{v_\text{wss}^2}\) we also need the effective \(N BW_{16-n/2}\) that we extrapolate from Figure 4.19 to be about 92 \(KHz\). Lastly, we sum the output noise powers in order to obtain \(\overline{v_\text{total}^2} = \overline{v_\text{Non-Stat}^2} + \overline{v_\text{wss}^2} = 7.87 \times 10^{-2} \, V^2\). This final value results in an output noise rms quantity of \(v_{\text{rms}} = 28.1 \, \mu V\). In order to assess the theoretical SNR obtained from this result, we must compare \(v_{\text{rms}}\) with \(v_{\text{LSBRms}}\), where \(v_{\text{LSBRms}} = \left( \frac{V_{\text{LSB}}}{2\sqrt{2}} \right)^2 = 6.47 \, \mu V_{\text{rms}}\). According to our calculations and the current values used, we obtain a theoretical \(ENOB\) of 13.9 bits.

The reason why the noise component from the ZCD can be ignored in the first stage is that it only needs be accurate to 8 bits. This means that so long as the input referred noise of the ZCD obeys the inequality 4.22, it will not degrade the
converter's linearity. For a more detailed treatment of the ZCD noise analysis, the reader is referred to Appendix A.2.

\[ V_{ZCD_{\text{rms}}} \ll \frac{1.2}{(2\sqrt{2}) \cdot 2^8} = 1.65 \text{ mV} \]  

(4.22)
4.4.8 R2R Ladder Design

An R2R ladder is used to provide the reference current $I_{ref}$ and its scaled versions $I_{ref}/16$ and $I_{ref}/256$. The R2R ladder can be divided into two operational modes [18], namely, voltage and current as illustrated in Figure 4.24. This architecture can be successfully used without calibration in low to medium resolution applications, or up to 12 bits [18]. The operation of this design can be more easily understood if we realize that the impedance looking to the left of every branch is equal to 2R. This fact results in a binary distribution of a current $I_{ref}$ through each branch of the ladder in Figure 4.24b. As for the voltage mode, it's easy to see that if the MSB branch is the only one connected to $V_{ref}$ then $V_{out}$ will equal $V_{ref}/2$. Each subsequent branch in the ladder will contribute to $V_{out}$ with a weight of $1/2^i$, with $i = 1$ corresponding to the MSB branch.

The operation in our design is the same but the switching setup is slightly different in order to accommodate for the desired operation. In order to make use of the OpAmp helper circuits, we need to reproduce the same input charging or discharging current across the integrating capacitor through the alternate set of inputs in the OpAmp. Therefore, we need to create two versions of the R2R ladder. Shown in Figure 4.25 is the present R2R ladder setup used to achieve the desired functionality along with OpAmp1 at the bottom to show where the respective nodes connect. During the initial charge up segment, the input current along with an offset current flow across the integrator. During this time only, $\theta_{1a}$ is connected to Out1+, $\theta_{1b}$ is connected to Out1-, $\theta_{2a}$ to Out2+, $\theta_{1b}$ to Out2-, and both S1 and S2 are closed. The rest of the R2R branches are connected to $V_{mid} = 0.6V$. Figure 4.27a illustrates the above set-up. At the end of this segment, we start integrating with a ramp in the opposite direction to resolve the 4MSBs. This is accomplished by connecting $\theta_{1a}$ and $\theta_{1c}$ to Out1-, $\theta_{1b}$ and $\theta_{1d}$ to Out1+, $\theta_{2a}$ and $\theta_{2c}$ to Out2-, $\theta_{2b}$ and $\theta_{2d}$ to Out2+, and opening both S1 and S2 with the rest of the R2R ladder branches connected to $V_{mid}$. The same set-up is illustrated on Figure 4.27b. Then during the final stage of operation
in the first stage, we must resolve the next 4 bits. This is accomplished by connecting \( \theta_{1e} \) to Out1+, \( \theta_{2e} \) to Out2+, \( \theta_{1f} \) to Out1- and \( \theta_{2f} \) to Out2- while the rest of the branches are left connected to \( V_{mid} \) as shown in Figure 4.27c.

The switching operation in the second stage is very similar but we will briefly go over it to assure that it is fully clear. The second stage OpAmp receives the amplified residue voltage from the first stage across the sampling capacitor. The next step involves applying another scaled reference current in order to resolve the next 4 bits. The R2R ladders shown in Figure 4.26 provide the scaled reference currents for this particular stage. We start by connecting \( \theta_{ic} \) to Out1+, \( \theta_{ia} \) to Out1-, \( \theta_{2c} \) to Out2+ and \( \theta_{2a} \) to Out2- while keeping the rest of the branches connected to \( V_{mid} \) as shown in Figure 4.28a. After this segment is over, we have the 4LSBs left to resolve and we accomplish this by connecting \( \theta_{ib} \) to Out1+, \( \theta_{id} \) to Out1-, \( \theta_{ib} \) to Out2+ and \( \theta_{2d} \) to Out2- as illustrated by Figure 4.28b.

There is another issue that can significantly limit the precision of the R2R ladder, and that is the effective resistance of the switches connecting the branches of the ladder to its output. Each branch must present an impedance of 2\( R \) so that the current is precisely distributed in a binary fashion throughout the ladder. However, if a switch impedance of \( R_{sw} \) is added to each branch, then we effectively have \( 2R + R_{sw} \) and the binary distribution is broken. This effect is easily demonstrated by calculating the parallel impedance of the two leftmost branches on Figure 4.24a, which results in \( R + R_{sw}/2 \). Then we add \( R \) to obtain \( 2R + R_{sw}/2 \) which we can see is not equal to the effective impedance of the next branch, \( 2R + R_{sw} \). However, if we can half the switch impedance in the next branch, then both branches would have \( 2R + R_{sw}/2 \) as their effective impedance and the binary distribution of current would be preserved. This technique can be extended to the rest of the R2R ladder by doubling the width of every subsequent transistor or switch. The technique is illustrated in Figure 4.29, where \( W \) represents the unit width of the smallest switch.
Figure 4-24: a) Voltage mode R2R ladder. b) Current mode R2R ladder.
Figure 4-25: R2R ladder for current reference in first stage.
Figure 4-26: R2R ladder for current reference in second stage.
Figure 4-27: a) Vin integration set-up b) Discharge with Iref c) Charge-up with $I_{ref}/16$ d) OpAmp1 and integrating capacitor.
Figure 4-28: a) Initial discharge after residue sampling b) Final integration with $I_{ref}/256$ c) OpAmp2 and integrating capacitor.

Figure 4-29: R2R ladder with binary sized switches in order to preserve binary distribution of $I_{ref}$ throughout the ladder.
4.4.9 State Machine Design

A state machine was implemented to control the switching operations of each stage. We will provide a brief note on clocked synchronous state machine architecture. For a more detailed analysis of state machine design the reader is referred to [21]. Clocked refers to the fact that all storage elements possess a clocked input, while synchronous means that all clocked elements use the same clock signal. A block diagram of a standard clocked synchronous state machine is given in Figure 4.30. The basic blocks consist of a next-state logic, a state memory containing storage elements such as flip-flops, and the output logic. The output of the next-state logic is referred to as excitation variables, and these only depend on the inputs and the current state of the machine. The state-memory consisting of flip-flops is clocked synchronously and holds the current state of the machine in its state variables. Lastly, the output logic is composed of logic gates that implement a given logic function that maps [inputs + current state] to outputs.

Figure 4-30: Standard architecture of a clocked synchronous state machine.
Figure 4-31: State machine for control of first stage operation.
Figure 4-32: State machine for control of second stage operation.
4.5 Summary

A short overview of the ADC configurations that are used in our design was given in this chapter. The operation of the ADC in this work was explained in detail. In addition, a new current cancellation technique was introduced in order to eliminate the switching non-linearities inherent in our integrator. Noise analysis of the main blocks in the converter was carried out to assess the major noise contributors and predict thermal and shot noise limited performance.
Appendix A

Noise Analysis of Selected Blocks

In the sections below, we will analyze in more detail the noise profile of the OpAmp operating as the integrator in the first stage of the ADC. The analysis of this block is important because noise from this stage is one of the main limiting factors to achieving a high SNR. Noise analysis of the ZCD is also performed in order to assess its effect on the converter's dynamic range.

A.1 OpAmp1 Noise Analysis

Let us focus on the first stage and helper circuits. Figure A-1 illustrates the parts of the amplifier involved in the analysis for reference purposes. We will start the analysis by identifying the transistors whose noise generators contribute to the total output noise power and identify their respective regions of operation. We know from the analysis in Chapter 3 that cascoded transistors in the first stage of Figure A-1 can be ignored. Therefore, in the cascoded OpAmp, only transistors P1, P2, N8 and N9 contribute to $i^2(t)$.

Next, we analyze the devices in the helper circuits. We only need to examine one of the branches since the other will contribute the same amount of noise. Noise from the helper circuitry appears at the output of the first stage through the gates of N8.
and N9. Since, N3 forms a current mirror with N9, any noise produced by N3 will directly appear at the output. N2 is a cascoded transistor and will therefore shunt its own noise and thus can be ignored. We have a low impedance node looking into the drain of N2 from the source of P4 and thus noise from P4 will also contribute to total output noise power. Transistors N5 and N13 effectively form a source follower, which means that noise from N5 and N13 can be input referred to the gate of N5, and this noise in the form of $\overline{v^2} = \overline{v_{N5/N13}}^2/9g^2_{m5}$ will also contribute to the total output noise current.

Transistors P1 and P2 are both operated in the subthreshold regime in order to maximize the $g_m/I_{DSat}$ ratio and thus minimize input referred noise. Therefore, they each contribute noise in the form of $2qI_{DSat}$, where q is the electron charge and $I_{DSat}$ represents the biasing current. N3 also operates in subthreshold given that it needs to have a large $I_{DSat}$ and low $V_{DS}$ for correct operation of the helper circuits. The rest of the noise contributing transistors identified above operate in strong inversion. Following is a list of expressions for the noise contributions from each of the noise generating devices as they appear at the output.

\[ \overline{i_{P1}^2}(f) = 2qI_{DSat}P1 \]
\[ \overline{i_{N3}^2}(f) = 2qI_{DSat}N3 \]
\[ \overline{i_{P4}^2}(f) = 4kT(2/3)g_{m4} \]
\[ \overline{i_{N5}^2}(f) = \frac{4kT(2/3)}{g_{m5}} \cdot g^2_{m9} \]  \hspace{1cm} (A.1)
\[ \overline{i_{N13}^2}(f) = \frac{4kT(2/3)9g_{m13}}{g^2_{m5}} \cdot g^2_{m9} \]
\[ \overline{i_{N9}^2}(f) = 4kT(2/3)g_{m9} \]
Following these set of expressions, we can estimate the total output noise PSD as:

$$\overline{i_{\text{total}}^2}(f) = 2 \left( \overline{i_{P1}^2}(f) + \overline{i_{N3}^2}(f) + \overline{i_{P4}^2}(f) + \overline{i_{N5}^2}(f) + \overline{i_{N13}^2}(f) + \overline{i_{N9}^2}(f) \right)$$  \hspace{1cm} (A.2)$$

where the factor of 2 arises from the fact that we must account for the other half of the amplifier and helper circuits. To calculate the input referred noise PSD we divide $\overline{i_{\text{total}}^2}(f)$ by $g_{m0}^2$ which is the transconductance of the source-coupled pair P1 and P2 giving us:

$$\overline{v_{\text{in}}^2}(f) = \frac{2}{g_{m0}^2} \left( \overline{i_{P1}^2}(f) + \overline{i_{N3}^2}(f) + \overline{i_{P4}^2}(f) + \overline{i_{N5}^2}(f) + \overline{i_{N13}^2}(f) + \overline{i_{N9}^2}(f) \right)$$  \hspace{1cm} (A.3)$$

After substituting for the respective transconductances in our equations, we estimate a value for $\overline{i_{\text{total}}^2}(f)$ of $2.88 \times 10^{-24} \, A^2/Hz$, which when input-referred results in $\overline{v_{\text{in}}^2}(f) = 2.18 \times 10^{-15} \, V^2/Hz$. We must now verify that this result is in agreement with transient noise simulations of the OpAmp. Figure A.2 shows the transient simulation results for the OpAmp’s input referred voltage noise power spectral density. The white noise PSD is approximately $3.5e-15 \, V^2/Hz$, which is indeed higher than our calculations predict but close enough to support our model of the main noise contributors.
Figure A-1: First stage of OpAmp along with helper circuits.
Figure A-2: Input referred voltage noise power spectral density of first stage OpAmp.
A.2 First Stage ZCD Noise Analysis

Before we calculate the input referred noise PSD for the ZCD, let us establish its accuracy requirements. Figure A.3 shows the output of the first stage integrator as the 8 MSBs are being resolved. As shown, the markings on the x-axis represent time periods equal to $T_{dk}$, each increment representing a unit increase in the count. Since the largest possible count can only be of 16, it is not possible to detect changes in $V_{out}$ that lie within a distance of $\Delta/2$, where $\Delta = V_{FS}/2^8$. This means that the first stage ZCD need only be accurate to 8 bits during first stage operation. More specifically, we say that:

$$I_{ref} T_{dk} V_{zcd,rm}, < I_{ref} 6c$$  \hspace{1cm} (A.4)

For a full scale of 1.2V, the above limit becomes 1.65mV. The pre-amplifier for the ZCD consists of a standard single-ended transconductor shown again on Figure A.4 for reference purposes. Source-coupled transistors operate in subthreshold and active currents in strong inversion. The input referred noise voltage PSD is given by the following expression:

$$\overline{v_{in}^2(f)} = \frac{2}{2 g_{msub}} \left[ 2 \cdot 2qID_{Sat} + 2 \cdot 4kT \left( \frac{2}{3} \right) g_{msi} \right]$$  \hspace{1cm} (A.5)

where $g_{msub}$ and $g_{msi}$ represent the transconductances for the devices operating in subthreshold and strong inversion respectively. Given that the transconductor has a bandwidth of 400 KHz as shown in Figure A.5, we estimate the input-referred noise as:

$$v_{zcd,rm} = \sqrt{\overline{v_{in}^2(f)} \cdot NBW \cdot \pi/2}$$  \hspace{1cm} (A.6)

where $NBW = 400 \text{ KHz}$. Substituting for NBW and $\overline{v_{in}^2(f)}$ in the above equation returns $v_{zcd,rm} = 32 \text{ uV}$. This figure satisfies the limit from equation A.4. It also
satisfies the same constraint but for the second stage, which is just 16 times smaller than the one given in equation A.4.

Figure A-3: Output waveform of first stage integrator illustrating noise requirements for first stage ZCD.

Figure A-4: ZCD pre-amplifier
Figure A-5: ZCD pre-amplifier transfer function.
Bibliography


