Analog to Digital Converters for CMOS Imagers

by

Susan Dacy

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

Master of Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1998

Author .....
Department of Electrical Engineering and Computer Science
May 8, 1998

Certified by .....
Charles Sodini
Professor of Electrical Engineering
Thesis Supervisor

Certified by .....
Marc Loinaz
Member of Technical Staff, Bell Laboratories, Lucent Technologies
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Analog to Digital Converters for CMOS Imagers

by

Susan Dacy

Submitted to the Department of Electrical Engineering and Computer Science
on May 8, 1998, in partial fulfillment of the
requirements for the degree of
Master of Engineering

Abstract

A/D converters for single chip CMOS imagers have often been designed using the column-parallel approach, employing a slow A/D converter for each column of the sensor array. This thesis investigates a serial approach utilizing a single fast A/D converter to process all of the imager pixels. If power scales linearly with frequency in a given A/D architecture, power dissipation for the two approaches should be comparable. However, the serial approach should occupy less area since only the cost of one A/D converter is incurred. A figure of merit ($\frac{1}{power*area}$) is introduced to verify this theory by comparing previously reported A/D approaches after appropriate technology, speed, and supply scaling.

Camera system specifications require a single serial A/D converter to have 10b resolution at a 3MHz sampling rate for a CIF (352x288) imager array running at $30\frac{frames}{second}$. Area minimization, power minimization, and the ability to build the A/D in a standard CMOS process are extremely important for consumer product applications. A single slope A/D architecture with a subnanosecond time digitizer shows promise for optimizing figure of merit over pipelined and folding interpolating approaches. This work focuses on the design issues of the 3MHz single-slope based A/D converter. Architectures appropriate for extending this A/D converter to 12MHz for four times CIF image arrays (704x576) are discussed.

The 3MHz converter was designed, simulated, and laid out in a 0.35um CMOS technology. At 3.3V supply, 25°C and nominal process conditions, the converter dissipates 29 mW while occupying 0.3 mm^2. A 12MHz trislope extension of this converter is estimated to dissipate 37 mW in 0.4 mm^2.

Thesis Supervisor: Charles Sodini
Title: Professor of Electrical Engineering

Thesis Supervisor: Marc Loinaz
Title: Member of Technical Staff, Bell Laboratories, Lucent Technologies

Acknowledgments

This research was completed at Bell Labs Innovations for Lucent Technologies in Holmdel, NJ under the MIT VI-A program. Many people in the DSP & VLSI Research Department deserve thanks for their friendship, advice, and technical assistance. I would like to thank Patrik Larsson for helping with the PLL and reading this thesis, Chris Nichol for help with the low-power digital circuits, Venu Gopinathan, Turi Aytur, Dan McMahill, Steve Decker, and Harry Lee for attending my design reviews and brainstorming with me, and Bryan Ackland for reading drafts of my thesis proposal and thesis. Thanks also to Dave Inglis, Kim Dugo, and Jay O'Neill for their support.

I'd like to give a special thanks to my Thesis Supervisor, Marc Loinaz. Marc provided the inspiration for this research. He has taught me to be an independent researcher, to give clear presentations, and to justify my design decisions. Thanks are also due to Charlie Sodini for his support and encouragement during my internship at Bell Labs.

My gratitude goes to the DOD/NDSEG Fellowship Program and Lucent Technologies for funding this research.

Finally, I'd like to thank my family for their support during my education at MIT and Bell Labs.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

CMOS technology is used for most microprocessors and ASICs, and is backed by an enormous worldwide research and development effort. Device feature sizes are decreasing by about a factor of two every five years. The CMOS Camera Project at Lucent Technologies explores the use of this booming CMOS technology for imaging applications. Building an imager in CMOS technology allows processing circuits to be integrated on the same monolithic chip. This system integration will allow CMOS cameras to provide low-cost, low-power solutions for applications such as video conferencing, document imaging, and security/surveillance [1].

Power minimization, area minimization, and the ability to build the imager in a standard CMOS digital process are extremely important for consumer product applications. High localized power dissipation can degrade the performance of the imager sensor and reduce the benefits of on chip Analog to Digital (A/D) conversion [2]. Area minimization is essential for reducing fabrication costs. Building the imager in a standard, digital CMOS process is also important for cost reduction and for initiating widespread use of CMOS imagers. This thesis focuses on the design of an A/D converter in a standard, digital CMOS process for a 352x288 imager that minimizes power and area. Avenues for extending this design to a larger (704x576) imager array are discussed.

The thesis is organized as follows. In Chapter 2, the analog signal chain through the imager is described. The operation of CMOS pixels, readout circuits, pro-

grammable gain amplifier (PGA), analog to digital converter, control logic, and the overall single chip CMOS camera are explained. Performance specifications for the A/D converter are derived from the overall camera requirements.

Chapter 3 addresses the number of A/D converters needed in an imager to minimize power and area. Column-parallel, semi-column parallel, and serial approaches are compared. The serial approach is selected for its potential to minimize area. Architectures appropriate for the serial approach are described. The new single slope architecture is investigated for its potential to minimize the power-area figure of merit over pipelined and folding interpolating approaches.

Single slope approaches are explained in Chapter 4. The new single slope architecture with a subnanosecond time digitizer is compared with a traditional single slope design. The overall architecture and calibration scheme for this new single slope A/D are detailed.

Chapter 5 details the analog circuit design involved in implementing the new single slope architecture described in Chapter 4. The major analog blocks include the time digitizer, track and hold, ramp generator, and comparator. Design tradeoffs are justified in terms of their minimization of power and area. The digital calibration loop algorithm and implementation is also discussed. Layout of the chip is explained. Issues encountered include optimizing layout to minimize mismatch, area, and noise coupling.

Chapter 7 elucidates an extension of the 3MHz design to 12MHz. Architectures based on increasing LSB resolution and subranging are discussed. Power and area numbers for the 12MHz design are predicted from the achieved 3MHz power and area numbers.

Chapter 8 reports simulation results for the 3MHz A/D. The power-area figure of merit for the 12MHz extension is compared to the current imager A/D design, column parallel and semi-column parallel approaches, and previously reported pipeline and folding interpolating designs.

Chapter 2

Analog to Digital Converter Performance Requirements

The A/D performance specifications are derived from the overall camera requirements. These specifications aid the selection of an A/D architecture. Section 2.1 describes the operation of a CMOS photogate pixel. The analog signal chain is then traced from the pixel through the PGA and A/D to the control logic and calibration. The operation of the overall single chip CMOS imager is detailed. Section 2.2 derives the A/D specifications from the desired imager system performance.

2.1 Analog Signal Chain of CMOS Imager

Active pixel image sensors (APS's) are the origin of the analog signal in the CMOS imager. APS technology is a low power, low cost, easily integrable alternative to CCD technology. Active pixel sensors can easily be built in CMOS technology with analog processing circuits, and digital timing and control electronics. The main disadvantage of APS technology compared to CCDs is process dependent leakage current and low quantum efficiency [3].

A pixel schematic with readout circuits is shown in Figure 2-1. Charge is integrated under the photogate for a fixed period of time (integration time). During integration, the polysilicon *Photogate* is held at Vdd and photo-generated carriers are

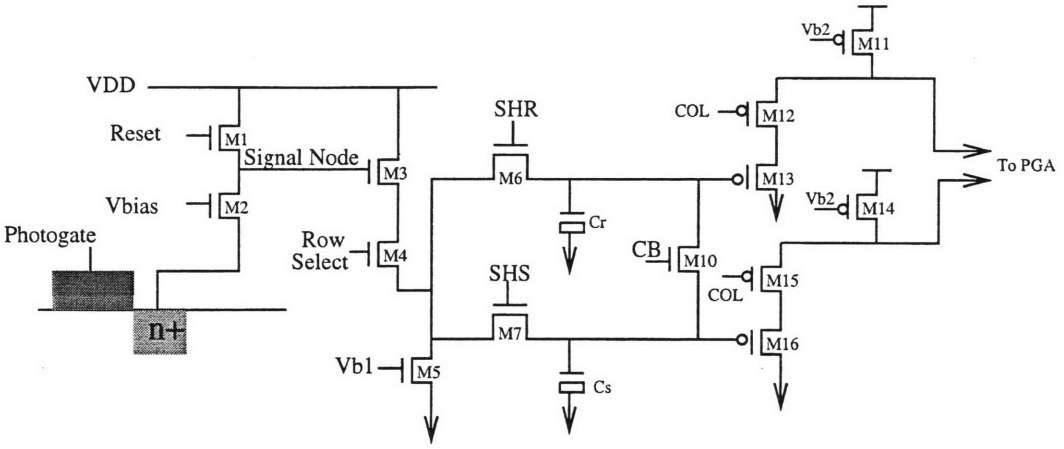


Figure 2-1: Active Pixel Sensor

collected beneath the gate oxide. *Vbias* is held at around 1.0V to isolate the collected charge from the *Signal Node* when the photogate voltage is high. During readout, the *Signal Node* is reset by pulsing *Reset* high and turning on M1. The reset level of the signal node is copied by the source follower M3-M5 onto the gate capacitance Cr by pulsing *SHR* to Vdd. *Photogate* is then driven to ground, turning on M2 and transfering the collected charge to the *Signal Node.* This charge displaces the voltage on the *Signal Node* by an amount that depends on the incident light intensity. The charge is then sampled on the gate capacitance Cs by pulsing *SHS*. Column Source followers M11-M16 drive the double-sampled signal into the fully differential, switched capacitor PGA. On one clock phase of the PGA, these signal and reset values are sampled. On the other clock phase, crowbar (*CB*) is pulsed and the offset difference between the column source followers is sampled and subtracted. This two-level correlated double sampling suppresses column offsets, pixel $\frac{kT}{C}$ noise associated with the reset operation, 1/f noise, and fixed pattern noise due to threshold voltage variations [4].

The pixel consisting of the photogate and M1-M4 of Figure 2-1 is repeated in a 352x288 array as in Figure 2-2. The two-level correlated double sampling circuit consisting of M5-M16 in Figure 2-1 is repeated at the end of each column in Figure 2-2. Each pixel in the array has a color filter that allows either red, green, or blue light to pass through. The signal and reset voltages of these pixels are fed as a fully

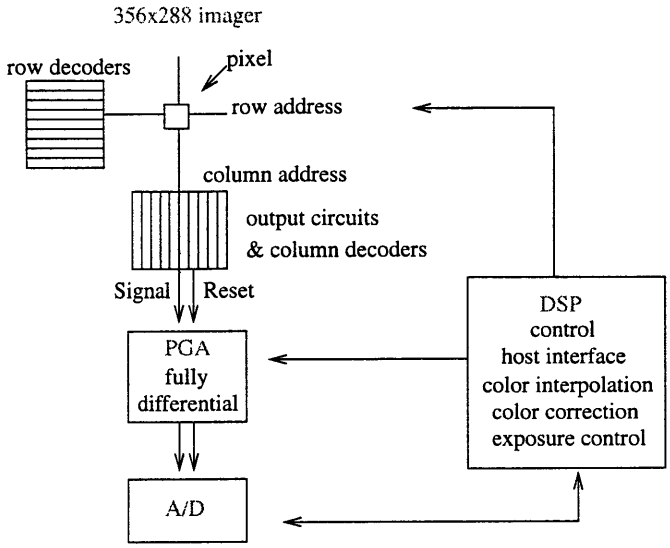


Figure 2-2: Single Chip Camera Architecture

differential signal into the Programmable Gain Amplifier (PGA).

The PGA provides variable gain to individual pixels. Programmable gain allows for correction of silicon's varying spectral response as shown in Figure 2-3. Specifically, the photogate sensor has a poor blue spectral response. When a blue pixel is read out, the PGA can amplify that pixel relative to the red and green pixels in order to achieve white balance. In low light conditions, the PGA can amplify all the pixels. This form of automatic gain control helps relax the resolution needed by the subsequent analog to digital converter.

The PGA design was investigated in Summer 1996. The PGA had a programmable gain from 1-16 and a 3_{db} bandwidth of 60MHz with a gain of 16 under a 2pF load. The fully differential output signal had a 2V range with a 3.3V supply. The PGA consisted of a fully differential, high gain-bandwidth op amp in a switched capacitor integrator configuration. Two pipelined PGA stages, each with a programmable gain from 1-4 were used to meet the specifications [5].

The fully differential output of the PGA is sent to the analog to digital converter

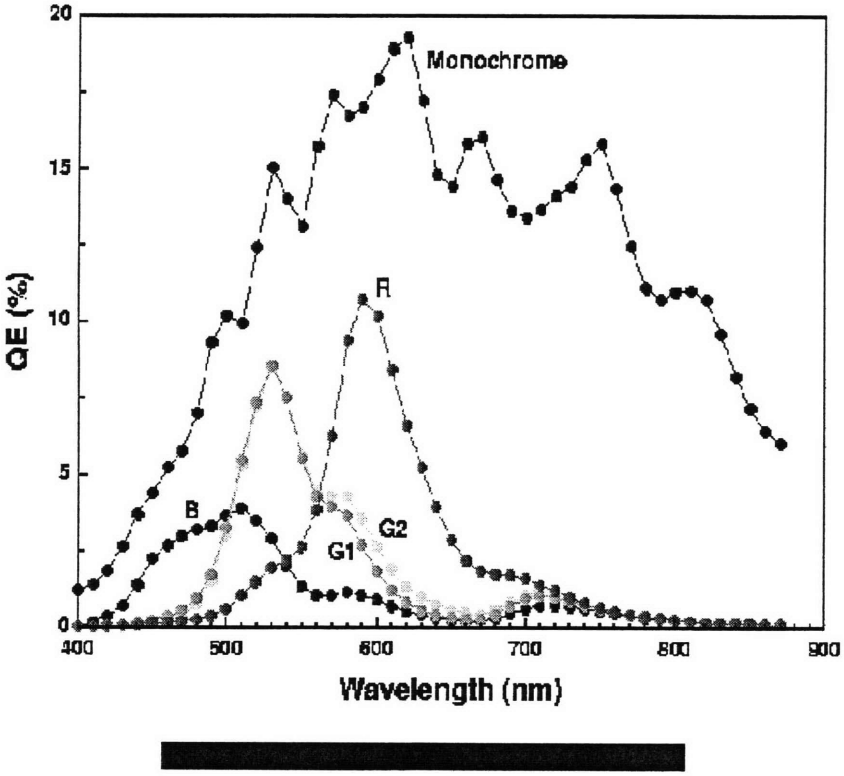


Figure 2-3: Spectral Response of Photogate Sensor

(A/D). The digital output of the A/D is then processed by the Digital Signal Processor (DSP). The DSP is responsible for color interpolation and color correction of this digital output. The R, G, and B values for a single pixel are interpolated from the surrounding pixels by interpolating from a 3x3 neighborhood surrounding each pixel. Color correction involves a linear 3x3 matrix transformation that minimizes mean colorimetric error [6].

2.2 Analog to Digital Converter Specifications

The A/D in Figure 2-2 converts the fully differential PGA output into digital codes that can be processed by the DSP. Performance metrics for the A/D are defined in Section 2.2.1. Specifications for the A/D performance are derived from the overall imager system requirements in Section 2.2.2.

2.2.1 Performance Metrics

Several parameters are used to measure the performance of an A/D. The parameters relevant to this A/D design are discussed below.

Conversion Rate

Conversion rate is the number of digital samples an A/D can convert in a given amount of time, measured in $\frac{MSamples}{second}$. For Nyquist rate A/Ds, the input signal frequency is limited to half of the conversion rate. In CMOS imager applications, the analog input voltage is held constant by the PGA. The conversion rate (f_c) in the imager application is governed by the size of the imager (r*s pixels), frame rate (f) and the number of A/D converters used (N):

$$f_c = \frac{r * s * f}{N} \tag{2.1}$$

The tradeoff between f_c and N will be discussed further in Chapter 3.

Differential Nonlinearity (DNL)

For an ideal A/D, the digital output codes as a function of the analog input voltage are shown in Figure 2-4.

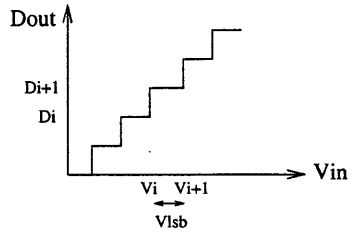


Figure 2-4: Ideal A/D transfer characteristic

In an ideal A/D, the analog input voltage change corresponding to two adjacent output codes is equal to the voltage V_{lsb}. V_{lsb} is the voltage corresponding to one Least Significant Bit (LSB). One LSB is $\frac{V_{FS}}{2^n}$ where V_{FS} is the full scale input voltage and n is the number of bits of the converter. A/D non-idealities can cause the spacing between adjacent digital output codes to be greater than or less than one LSB. Differential Nonlinearity (DNL) is a measure of this error and is defined as [7]

$$DNL[i] = \frac{V_{i+1} - V_i}{V_{LSB}} - 1 \tag{2.2}$$

where i is the index at which DNL is being measured. This measurement is defined in units of LSB's.

An extreme example is shown in Figure 2-5. The digital output code D_i never appears at the output. In this case, the DNL=-1 for the code D_i. From equation 2.2 a DNL of -1 is the worst case negative DNL and a code is missing. On the other end of the spectrum, a code can be very wide and have a positive DNL greater than 1. DNL is a measure of the error in the resolution of the A/D.

Integral Nonlinearity (INL)

Integral Nonlinearity (INL) measures the absolute accuracy of the A/D. INL in LSBs is the difference between the actual A/D transfer curve and the ideal A/D

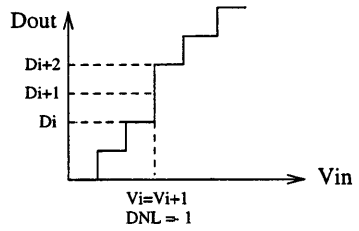


Figure 2-5: Poor DNL Example: Missing Output Code

transfer curve as shown in Figure 2-6. INL is defined after correction for gain and offset error. In the example of Figure 2-6, the DNL is small, but integrates to give a large INL in the middle of the transfer curve [7].

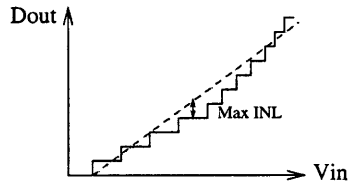


Figure 2-6: Transfer curve with large INL

Signal-to-Noise Ratio and Effective Number of Bits (ENOB)

Quantization error is defined as the difference between the original input and the digitized output converted back to an analog signal using an ideal D/A. Quantization Error for an ideal A/D transfer curve is shown in Figure 2-7. The quantization error decreases as the resolution of the converter increases. Quantization error can be modeled as an additive noise source appearing at the output [7].

Assuming the quantization error is uniformly distributed and independent of the analog input, the quantization noise power can be expressed as the mean square of the quantization error in Figure 2-7 [7]

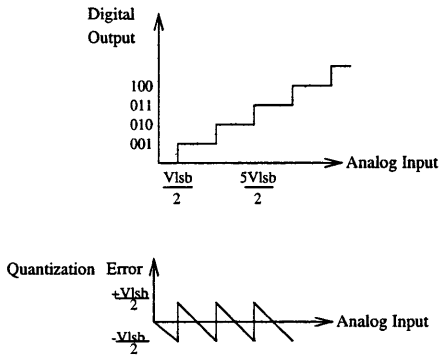


Figure 2-7: Quantization Error

$$V_{noise,rms} = \sqrt{\frac{\int_{-\frac{1}{2}V_{lsb}}^{\frac{1}{2}V_{lsb}} V^2 dV}{V_{lsb}}} = \frac{1}{\sqrt{12}} V_{lsb} \tag{2.3}$$

A sinusoidal analog input with a peak-to-peak voltage of $VFS = 2^n * V_{lsb}$ (where n is the number of bits) has an RMS value:

$$V_{FSRMS} = \frac{1}{2\sqrt{2}} 2^n V_{lsb} \tag{2.4}$$

Thus an ideal n-bit A/D with a sine wave input has an SNR:

$$SNR_{FS}[dB] = 20 \log_{10} \frac{V_{FSRMS}}{V_{noise,rms}} = 6n + 1.76 \tag{2.5}$$

For example, an ideal 10-bit A/D has a peak SNR of 61.76 dB. While SNR is a measure of the noise level of an A/D relative to its peak input signal, Signal to Noise plus Distortion Ratio (SNDR) is a measure of the noise plus harmonic distortion relative to the input. This distortion is a result of nonlinearity (INL) of the A/D. Effective Number of Bits (ENOB) is computed by measuring the SNDR and using equation 2.5:

$$ENOB = \frac{SNDR[dB] - 1.76}{6} \tag{2.6}$$

ENOB, INL, and DNL are the primary measures of the linearity and noise of the

A/D. ENOB can be measured as a function of the sample rate and bandwidth of the incoming signal, whereas INL & DNL are "DC" parameters.

2.2.2 Analog to Digital Converter Requirements

The A/D specifications are listed in order of importance in Table 2.1. These specifications are justified in terms of the overall camera system requirements.

Table 2.1: A/D Specifications

Resolution	10 bits
Conversion Rate	$356 * 288 * 30 \frac{frames}{second} = 3 \frac{Msamples}{second}$, $12 \frac{Msamples}{second}$
DNL	$\leq 0.5LSB$
INL	Not critical - $\leq$ 5LSB
Area	$\leq 1mm^2$
Power	$\leq$ 20 mW
ENOB	$\geq$ 8b
Supply	3.3 Volts
Technology	0.35um, no linear capacitors or resistors
Calibration	flexible- several lines of blanking intervals
Input	fully differential
Temperature	0°-70°C
Supply variation	3.3V $\pm$ 10 %
$v_{incommonmode}$	1.5V
$v_{indifferential}$	$\pm$ 1V

The resolution needed by the A/D is reduced by the gain control function of the PGA. In medical and scientific imaging applications, 12b resolution is often needed with a CCD imager. In consumer electronics, 8b or 10b resolution is commonly in use. Figure 2-8 shows the output signal level relative to the photon shot noise and dark current shot noise for a CMOS photogate imager [3]. This shows that the imager signal to noise ratio is only about 50dB, or 8 bits. Thus the A/D should achieve at least 8 effective bits (ENOB) so that camera performance is limited by the imager and not the A/D. An A/D resolution of 10 bits is needed for digital post processing (about 2 bits of resolution are thrown away in digital post processing). Additionally, as the signal to noise ratio of the photogate pixel improves, higher resolution converters will

be needed. For CCD applications where the imager signal to noise ratio is higher, a higher resolution converter will be needed.

For the 356x288 imager running at a CIF standard frame rate of 30 $\frac{frames}{second}$, a conversion rate of 3 $\frac{Msamples}{second}$ is needed in a single A/D from Equation 2.1. Also of interest is extending this design to a four times CIF image array (704x576), requiring a conversion rate of 12 $\frac{Msamples}{second}$.

The A/D does not require good integral nonlinearity (INL), but does require good differential nonlinearity (DNL) - less than half an LSB. This is because the human eye is logarithmically sensitive to light intensity. The eye is sensitive to the difference in pixel intensities, not the absolute linearity of the intensity difference [8].

Area and power minimization are the main challenges in this A/D design. Area consumption will be minimized, with a target of $1mm^2$ in 0.35um CMOS. The area of the A/D is part of the cost of fabrication. Cost reduction is essential for the consumer products market. The current single chip imager in [6] has a total die size of $100mm^2$. This includes the imager, readout circuits, PGA, A/D, and digital color processing. The target $1mm^2$ is about 1% of this total chip area, and a factor of three reduction in area over the $3.6mm^2$ A/D reported in [6]. Power dissipation will also be minimized, and should not exceed 20 mW. The power dissipation for the total single chip imager reported in [6] is 188mW. The target of 20mW is 20% of this total power and a factor of two reduction in A/D power over the converter in [6]. The analog circuits on the imager chip will operate with high quiescent power dissipation compared to the dynamic power dissipation of the digital circuits. Thus, as technology and supply are scaled, an increasing fraction of the total power of the imager chip will be dissipated by the analog circuits. Power minimization techniques are essential for the design of a scalable imager A/D [9].

The technology available for fabrication of this A/D is 0.35um CMOS with a 3.3 Volt supply. Design challenges in this technology include using a standard digital process. This means that there is no high-resistance poly available for making large resistors in a reasonable area. The sheet resistance of the polysilicon in this digital process is about 30 $\frac{\Omega}{square}$, indicating that large resistors will consume a lot of area.

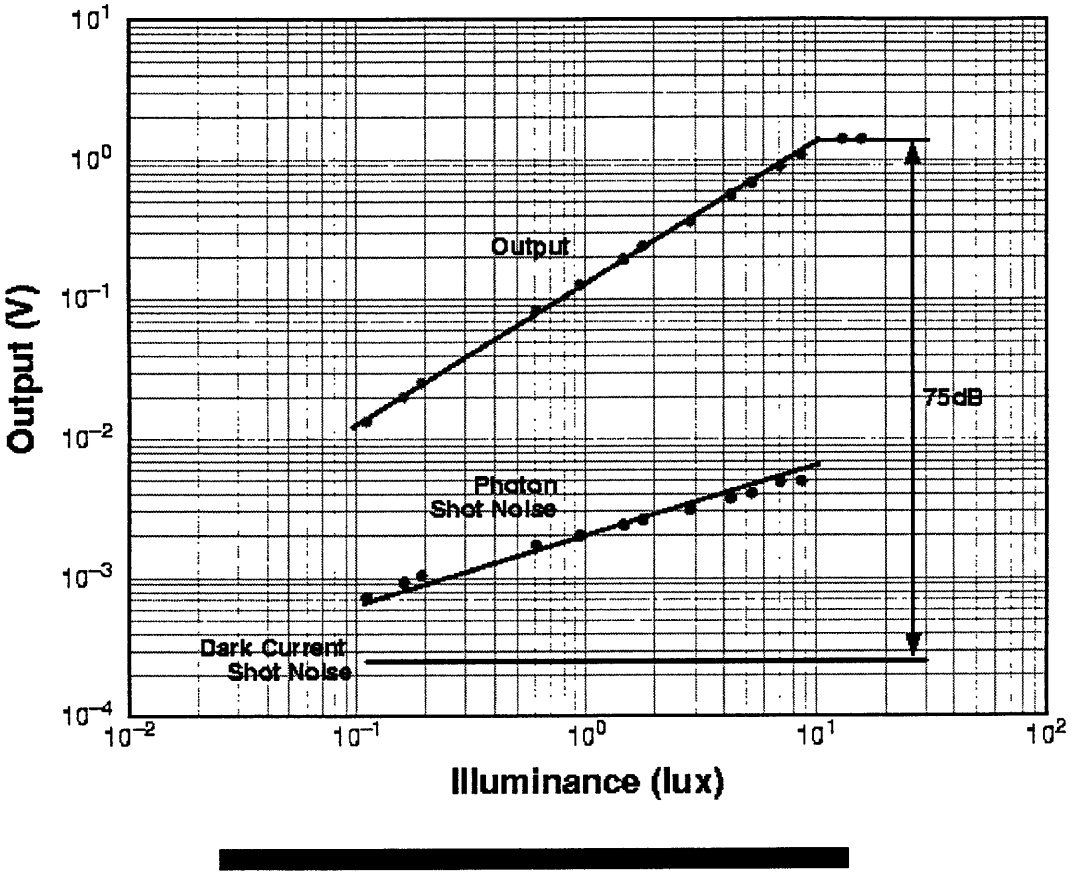


Figure 2-8: Noise of Imager Pixels

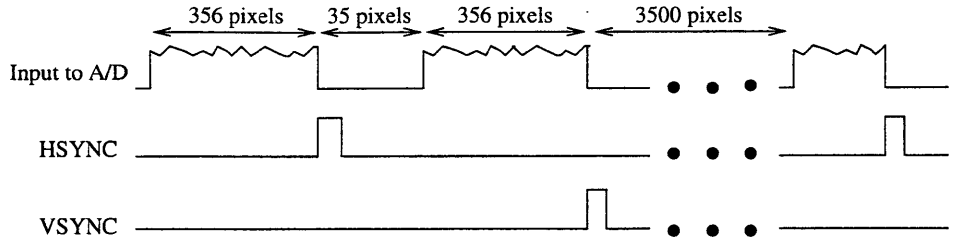


Figure 2-9: Blank Pixels Between Lines and Frames

There is also no thin dielectric available for making large capacitors in a small area. Standard Poly/M1/M2 have a capacitance of around $6 * 10^{-2} \frac{fF}{um^2}$, so large capacitors will also consume a large amount of area. Ideally, the A/D will be an all-MOS design utilizing only small resistors and small capacitors. However, as capacitors and resistors become smaller, they are more sensitive to mismatches.

The input signal for the A/D is shown in Figure 2-9 (assuming one A/D is used). There are 35 "blank" pixel times between each line of 356 pixels. Blanking intervals exist for line and frame synchronization in video systems. For a 3MHz clock, this allows about 12usec for calibration. Between frames, there are 10 blank lines (3500 pixels), allowing 1msec. This dead time gives flexibility in the calibration scheme that may allow extra design opportunities for power and area minimization.

Since the A/D will be embedded on the same chip with the imager, the A/D should not produce too much substrate or supply bounce. This bounce is especially critical during pixel readout. By minimizing full scale voltage switching at high frequencies, this bounce should be reduced. It is even more important that the A/D converter be insensitive to switching noise caused by digital control and processing circuits on the chip.

Circuits that are not part of the pixel array, including the A/D, will be covered by a light shield. This avoids high leakage currents and degradation of signals caused by photo-generated carriers.

Chapter 3

Number of Converters

The A/D architecture chosen for imaging applications is highly dependent on the number of converters, since the number of converters determines the speed of A/D operation. The optimal A/D solution for consumer products applications minimizes the total power and area of all the converters needed to process an image.

A power/area figure of merit is defined as $\frac{1}{power*area}$. Power and area are weighted equally in this formula, so designs are equivalent even if they dissipate twice the power as long as they occupy half the area. This formula assumes that all designs are 10b and that they have been scaled for 0.35um technology, 3.3V, and the appropriate conversion rate. The optimal A/D solution has the largest figure of merit. The extension of this design to a four times CIF imager array size (704x576) is ultimately important. Thus, the comparisons in this chapter are done for a 704x576 imager operating at 30 $\frac{frames}{second}$, giving a conversion rate of 12MHz.

Section 3.1 describes the column parallel approach and gives some previously reported data. Section 3.2 details the semi-column parallel approach. Section 3.3 describes the serial approach. Previously reported power and area numbers and figure of merits are compared for all approaches. These numbers indicate that the serial approach should result in maximum power/area figure of merit. Architectures appropriate for the serial approach are described. The new single slope architecture is chosen for its potential to minimize power and area relative to folding interpolating and pipeline approaches.

3.1 Column Parallel Approach

The column parallel approach involves building an A/D at the bottom of each column. This A/D processes all the pixels in that column. The number of A/D's needed is now equal to the number of columns - 704 in this case. The speed of the converter is smaller than would be necessary with a single A/D, down by a factor of 704 to 17 $\frac{ksamples}{second}$. However, a PGA will be needed on every column, thereby decreasing the bandwidth but increasing the overall area costs of the PGA. If no PGA is used, the resolution of the converter will need to be increased. Although some reported designs have used 10 bits without a PGA, the use of color filters degrade the sensor signal by a factor of around 10. Thus, about 12 bits of resolution will be needed for operation down to 1 lux [10]. Each A/D in the column-parallel approach needs to achieve 12 bit resolution, less than $\frac{20mW*2}{704} = 56.8uW$ of power dissipation and minimal area consumption at $17\frac{ksamples}{second}$. The factor of two in the power spec is because no PGA is needed. The 20mW of power budgeted for the PGA can therefore be added to the 20mW already budgeted for the A/D.

A column parallel approach utilizing a successive approximation converter was taken in [11]. A slow 10b successive approximation converter was built in the pixel pitch width at the end of each column. Each A/D occupied 0.094 mm^2 in 1.2um technology. Scaling for 0.35um technology and multiplying by 704 gives a total area of 5.6 mm^2. Scaling the quoted single A/D power consumption of 8.6uW from 5V to 3.3V, 333Hz to 17kHz, 1.2um to 0.35um, and multiplying by 704 gave the total power number of 59.6mW. Scaling of power with conversion rate, technology, and supply was done linearly.

A column parallel approach was also taken in [12]. Slow 8-b standard single slope converters were built at the end of each column of the imager. Each A/D occupied 0.1 mm^2 in 2um technology. Note that this area does not include the counter, ramp generator or any calibration circuitry, which were built off chip. Scaling for 0.35um and multiplying by 704 columns gave a total best case area of 2.2 mm^2. The 125uW of power of each single slope was dominated by the comparator (again, the power

for the ramp generator and counter are not given). Scaling from 5V to 3.3V, 1.2 $\frac{ksamples}{second}$ to 17 $\frac{ksamples}{second}$, 2um to 0.35um, and multiplying by 704 gave a total power consumption of 144mW. Scaling of power for supply, conversion rate, and technology were done linearly.

Both of these column parallel approaches were used for a small array size at low conversion frequencies. When appropriately scaled for a large array size at high conversion frequencies, the total power and area consumption are large as summarized in Table 3.1. These two converter power/area figure of merits were the highest of those found in the literature. These A/Ds were probably not aggressively optimized for power and area, since their overall power and area were small in absolute terms for small array sizes operating at low conversion frequencies.

Designing column parallel A/Ds has the additional layout challenge of fitting the A/D in a pixel pitch width, which in this application is a factor of 2 smaller than in [11] or [12]. Column parallel A/Ds also have the potential problem of mismatches between A/Ds in different columns causing fixed pattern noise. This issue was not addressed in [11] or [12].

3.2 Semi-Column Parallel Approach

Another option is a semi-column parallel approach in which one A/D is used for every x columns [2]. This decreases the conversion rate over the serial approach by a factor of $\frac{704}{x}$. However, it increases the area by a factor of $\frac{704}{x}$.

A semi-column parallel approach was used in [13]. A 10-b cyclic A/D converter was designed for every 2 columns of the imager. Each A/D occupied $0.07mm^2$ and dissipated 100uW of power. Scaling for 0.35um technology, the total area of $\frac{704}{2}$ such converters is 4.7 mm^2. The power dissipation is 22.6mW after scaling from 15.3 $\frac{kSamples}{second}$ to 34 $\frac{kSamples}{second}$, 5V to 3.3V, and 0.8um to 0.35um. Scaling for conversion rate, power supply, and technology was done linearly. The power/area figure of merit is better than the column parallel approach, but still smaller than the desired figure of merit as shown in Table 3.1. The semi-column parallel approach still has the pixel

Table 3.1: Scaled Comparison of Previously Reported A/D Approaches

Approach	Total Power	Total Area	Figure of Merit ($\frac{k}{power*area}$)
Column Parallel Single Slope[12] (Note: 8b, some circuits not included)	144mW	2.2 mm^2	3
Column Parallel Successive Approximation [11]	59.6mW	5.6mm^2	3
Semi-Column Parallel Cyclic [13]	22.6mW	4.7 mm^2	9
Serial Pipeline [9] (Note: 12b, 11 ENOB, double poly)	30.5 mW	1.4mm^2	24
Serial Folding Interpolating [15] (Note: 8.7 ENOB at 50$\frac{MSample}{sec}$)	26.6mW	0.59mm^2	64
Desired Converter (Standard process, 10 ENOB)	20mW	1mm^2	50

pitch matching and column fixed pattern noise issues. Measurements of column fixed pattern noise due to different A/D converters were not reported in [13].

3.3 Serial Approach and Architectures

The serial approach to A/D conversion for CMOS imagers involves using a single A/D and PGA to process the whole image. Since the same A/D is used to process every pixel, the issue of fixed pattern noise caused by A/D mismatch is eliminated. Although the serial approach only has the hardware costs associated with one A/D, this A/D has to operate at a high conversion rate. Intuitively, if power scales linearly with frequency, the column parallel, semi-column parallel and serial approaches should have comparable power dissipation. However, the serial approach should minimize area since only one A/D is needed. Table 3.1 summarizes power, area, and power/area figure of merit for previously reported column parallel, semi column parallel, and serial designs. This figure of merit indicates that the serial approach is optimal for a low power, small real estate A/D solution.

While the serial approach shows promise for minimizing power and area, the conversion rate is the same as the pixel data rate, and is therefore proportional to

the total number of pixels. This conversion rate of 12 $\frac{MSamples}{second}$ limits the choice of architecture. A conversion rate of $12\frac{Msamples}{second}$, power dissipation $\leq$ 20mW and an area around $1mm^2$ in 0.35um standard CMOS are needed if a serial approach is used. Architectures capable of achieving the necessary conversion rate are outlined in the following sections. Previously reported data for these architectures has been scaled and is compared in Table 3.1.

3.3.1 Pipeline

The pipelined converter architecture shown in Figure 3-1 is based on an Analog to Digital Subconverter (ADSC) to perform a 1 bit comparison of the output voltage and the reference voltage. The Digital to Analog Subconverter (DASC) is basically a switch that subtracts the reference voltage from the input if the output of the comparator was a 1. There are B pipeline stages, each of which finds the value of a bit and passes on the residue voltage. The conversion speed of the A/D is therefore equal to the conversion speed of a single stage, allowing high throughput. The resulting latency can be tolerated as it only gives an initial delay in reading out the image [9].

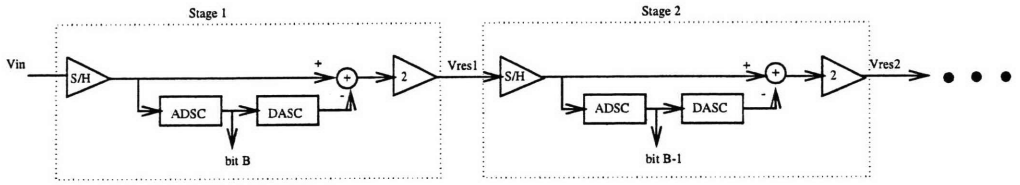


Figure 3-1: Typical Pipelined Converter

A 12b dynamic range, 11 effective number of bits (ENOB), 1 bit per stage pipeline converter was designed in [9]. This converter dissipates 33mW of power in $17mm^2$. Scaling from 1.2um to 0.35um, 2.5V to 3.3V, and 5MHz to 12MHz, gives 30.5mW of power dissipation in 1.4 mm^2. Scaling of power with supply voltage, conversion rate, and technology was done linearly. Scaling of area with technology was done as the square of the feature size ratio. This design utilized a process with double poly, allowing small, well-matched pipeline capacitors. In a standard, digital CMOS process without double poly, capacitors will be about a factor of 7 larger [14]. More power

would also be dissipated due to the bottom plate capacitance of metal capacitors. Thus the power and area numbers reported in [9] will both be increased in a CMOS process without double poly capacitors.

3.3.2 Folding Interpolating

A typical folding architecture is shown in Figure 3-2. The first m bits are resolved by a course m-bit A/D. The folding circuit produces a residue voltage from which the least significant k-bits are determined. A folding architecture has a conversion rate comparable to a flash A/D, while the folding circuit reduces the hardware costs of the flash A/D [9]. Interpolation techniques can be applied to the k-bit A/D to increase the resolution of the converter [7].

A 10b dynamic range, 8.7ENOB folding interpolating converter was designed in [15]. This A/D dissipated 240mW in $1.2mm^2$. Scaling from 0.5um to 0.35um, 50MHz to 12MHz, and 5V to 3.3V gave 26.6mW in 0.59 mm^2. Scaling of power for technology, conversion rate and supply was done linearly. These power and area numbers, when scaled, were the smallest of those found in the literature for this type of architecture.

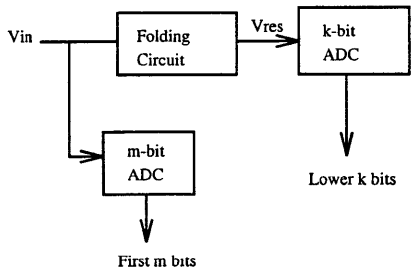


Figure 3-2: Folding Architecture

3.3.3 Single Slope

The serial pipeline and folding interpolating figure of merits are close to the desired figure of merit. However, the new single slope architecture described in Chapter 4 shows promise for achieving comparable power dissipation, smaller area, and a high ENOB in a standard CMOS process.

Chapter 4

Single Slope Architecture

A single slope converter converts an input voltage into a time interval, the duration of which is proportional to the value of the input voltage. A time digitizer converts this interval into a digital output. The speed and accuracy of the time digitizer typically limits single slope performance.

Section 4.1 describes a traditional single slope architecture and outlines problems at high conversion rates. Section 4.2 describes the new single slope architecture utilizing a subnanosecond time digitizer. This time digitizer increases the achievable conversion rate over the traditional single slope design. Section 4.3 outlines an endpoint calibration scheme for the new single slope converter.

4.1 Traditional Single Slope

A traditional single slope converter is shown in Figure 4-1. V_{in} is one input to a comparator. A ramp generator starts ramping from Vdd at time Tstart. When the ramp voltage reaches the input voltage at time Tstop, the comparator is triggered, generating Vstop. The duration of Tpulse=Tstart-Tstop is proportional to the input voltage. Tstart and Tstop are triggering signals for a time digitizer. A counter acts as the time digitizer by counting the number of periods of a fixed frequency clock in Tpulse. The digital output of the counter is the output of the A/D.

The speed of the time digitizing counter limits how quickly a given number of bits

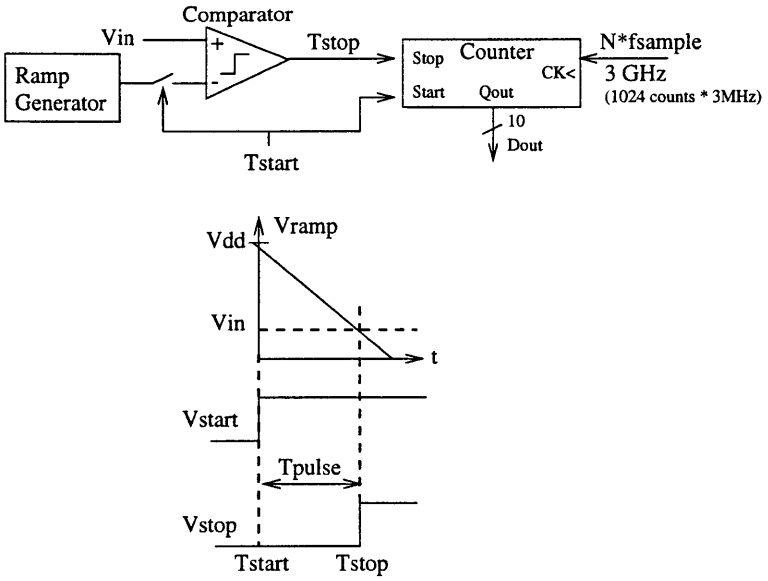


Figure 4-1: A 10b, 3MHz Traditional Single Slope Architecture

can be resolved. For 10b to be resolved at a 3MHz conversion rate, a 3GHz counter is needed. A 3GHz clock is not available on the single chip imager. Additionally, such high frequency clocking creates substrate bounce. This counter clock also must be low jitter to achieve the desired effective number of bits. Traditional single slope converters have been used in low conversion rate applications, such as a column parallel A/D approach. An alternative method of time digitization is described in the next section which eliminates the need for a high frequency clock, thereby increasing the conversion rate achievable in a single slope converter.

From Figure 4-1, a single slope converter has minimal power and area requirements. There are no large capacitors, reference ladders, or resistors that require good accuracy and matching. There are also not a large number of power consuming op amps or comparators. The critical components are a low jitter time digitizer, ramp generator, and comparator. Thus, when initially comparing the single slope converter to pipeline and folding interpolating architectures, it showed promise for maximizing the power-area figure of merit.

4.2 New Single Slope with Sub-nanosecond Time Digitizer

The new single slope converter with a sub-nanosecond time digitizer is shown in Figure 4-2. The single slope part of the A/D is as described in Section 4.1 with a ramp generator and comparator. The time digitizer uses a gate delay to set the resolution of the converter (t_{lsb})[16]. This gate delay is controlled by phase locking the ring oscillator to a lower frequency system clock, often at the conversion rate of the A/D. If there are N stages in the ring oscillator, there are 2N possible states of the ring oscillator. Thus, the ring oscillator provides the least significant $\ln_2 2N = 1 + \ln_2 N$ bits and the coarse counter provides the remaining bits. This architecture makes the speed and resolution performance of the A/D independent of the availability of a high frequency, low noise, on-chip clock. The maximum frequency on chip is now $\frac{f_{conv}*2^m}{2*N}$ where N is the number of stages in the ring oscillator, and m is the number of bits of resolution. Additionally, using a gate delay as the time measurement unit allows the resolution of the time digitizer and thus the overall performance of the A/D to improve with technology scaling.

Another interesting feature of the single slope converter is the tradeoff between conversion rate and number of bits for a given t_{lsb}. This tradeoff is depicted in Figure 4-3. 10b can be resolved at 3MHz using t_{lsb}=325psec. Or, 9b can be resolved at 6MHz using the same t_{lsb}. The Phase Locked Loop (PLL) and ring oscillator frequency stay the same. The A/D converter could potentially be programmable along this resolution/speed curve at minimal additional hardware cost. At high resolutions (low conversion frequencies), the accumulation of jitter will reduce the effective number of bits. However, there is still a linear tradeoff between dynamic range and speed.

At 12MHz, t_{lsb}=81psec, which is too fast for 0.35um technology given variations in temperature, process and power supply. Sub-gate delay resolution has been obtained in [16] and [17]. These extensions will be elaborated on in Chapter 9.

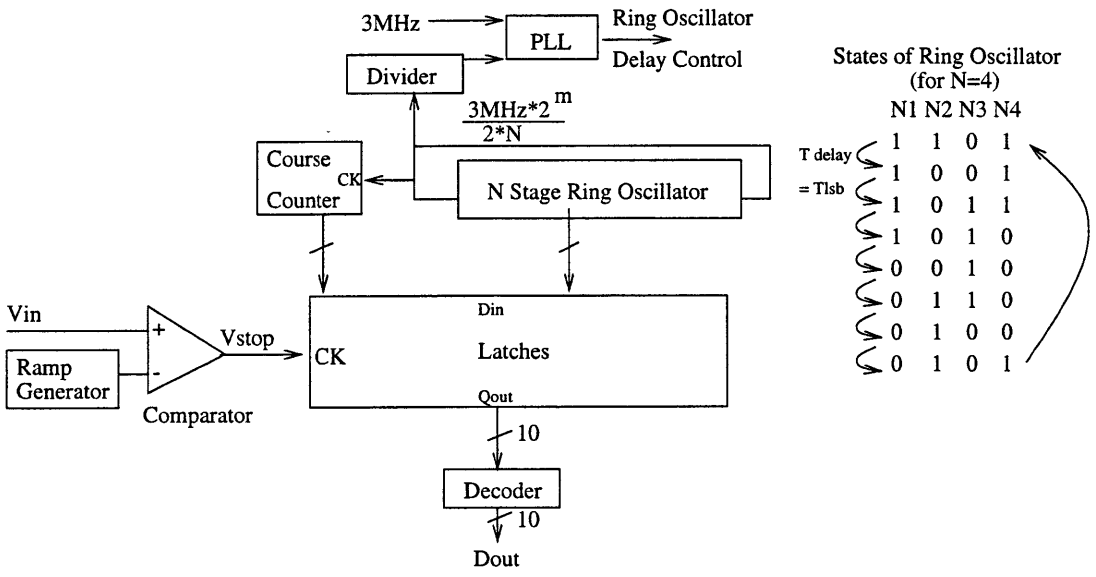


Figure 4-2: New Single Slope Architecture with Subnanosecond Time Digitizer

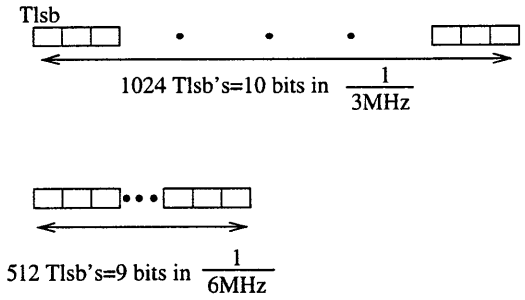


Figure 4-3: Linear Tradeoff Between Resolution and Speed

4.3 Calibration

Calibration of the A/D can be performed at the end of each line of the image during the 35 pixel period blanking interval. The proposed calibration technique consists of measuring and storing the digital value of the comparator delay. This delay is subtracted from the digital output of the converter. Then, the full scale input voltage is applied. The slope of the ramp is adjusted to achieve D_{max} when the input is full scale. The effect of this calibration is shown in Figure 4-4. This calibration scheme only aligns the endpoints of the A/D transfer curve. Since no calibration is performed in the middle of the curve, the A/D must be designed for good linearity. This translates to the need for a linear ramp generator.

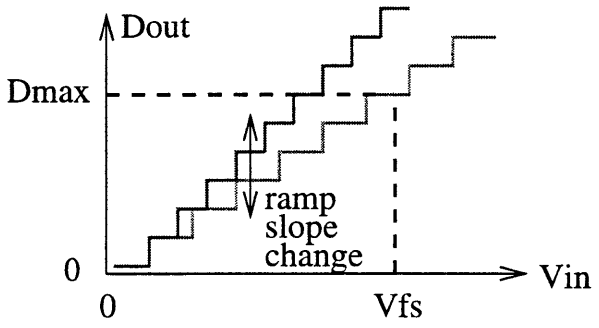


Figure 4-4: Single Slope Endpoint Calibration

Chapter 5

Design of a 10b, 3 MS/s Single Slope A/D

5.1 Introduction

The new single slope architecture described in Chapter 4 is detailed in Figure 5-1. The single slope compares the input voltage to a ramp to generate V_{strobe}. The time digitizer uses V_{strobe} to latch the state of a ring oscillator and coarse counter. The final output is decoded from the latch bank. Endpoint calibration is performed during the blanking intervals at the end of each line and at the end of each frame by adjusting the slope of the ramp.

This chapter details the design of the analog blocks in Figure 5-1. Design choices are motivated by the minimization of power and area. The analog blocks include the time digitizer, track and hold, ramp generator, and comparator. Bias point generation and critical timing issues are also discussed.

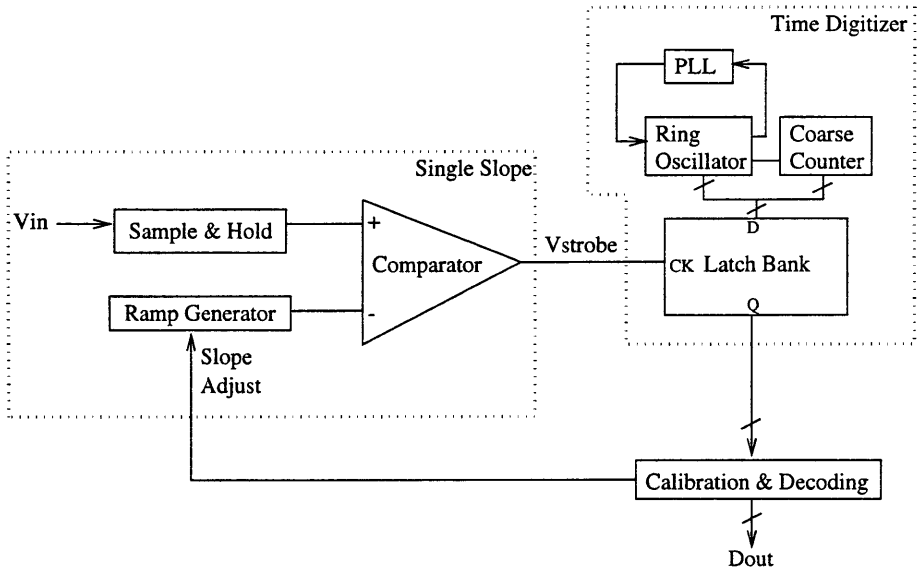


Figure 5-1: Single ended representation of Single Slope Architecture

5.2 Time Digitizer

5.2.1 Overall Architecture

A high resolution time digitizer is a critical part of increasing the conversion rate of time based A/D converters. In the architecture shown in Figure 5-2, the state of a phase locked ring oscillator gives the least significant $1 + \ln_2 N$ bits (where N is the number of ring oscillator stages), while a coarse counter provides the most significant bits. The resolution of the time digitizer is given by $t_{lsb} = \frac{1}{f_c * D * N}$ where N is the number of stages in the ring oscillator, D is the divider ratio, and f_c is the conversion rate of the A/D. t_{lsb} is the delay of a single ring oscillator stage, and the minimum value of t_{lsb} will scale with smaller feature sizes. Thus the achievable performance of this A/D architecture (in terms of resolution and conversion rate) follows technology scaling curves. The divider ratio, D, can be changed depending on the value of N and f_c. Given f_c, N is chosen based on power, area, and noise considerations.

Area stays approximately fixed over a small range of increasing N. This is because the small area of the ring oscillator increases, but the area of the coarse counter decreases. Power, however, increases with increasing N. This is because the ring os-

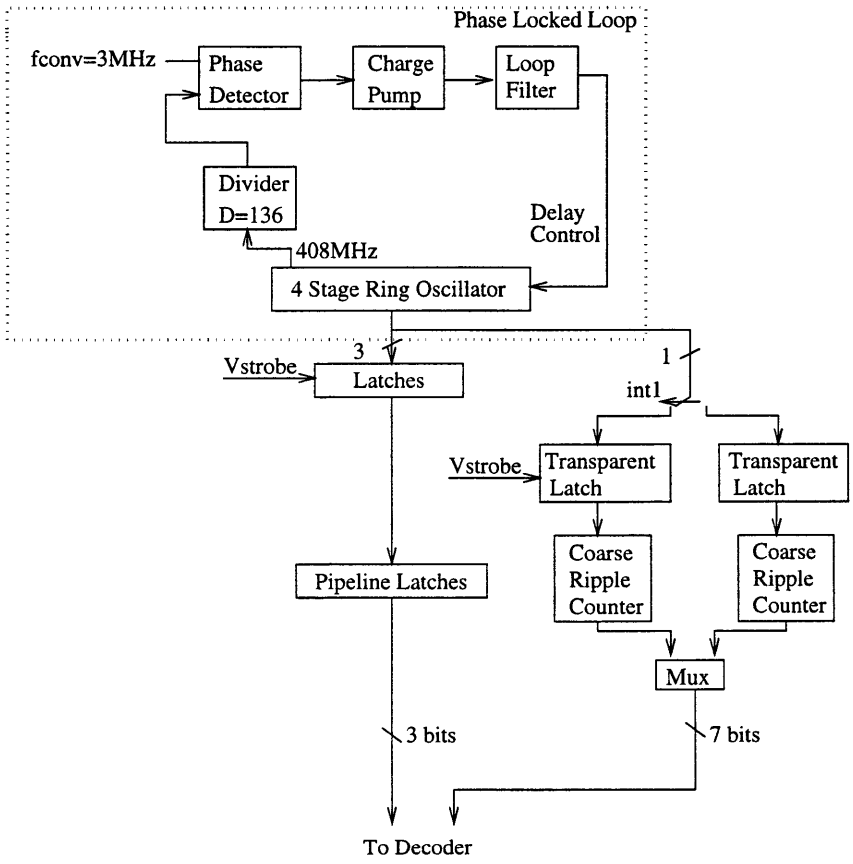


Figure 5-2: Single ended representation of Time Digitizer

cillator is implemented using fully-differential current-steering circuits to minimize noise caused by and coupled into the ring oscillator. Each ring oscillator stage dissipates static power, as do the extra fully-differential buffers and fully-differential to single ended converters. The fewer the number of ring oscillator stages, the faster the frequency of the oscillator and the more dynamic power is dissipated in the ring oscillator and coarse counter. There is a tradeoff between static power dissipated in the ring oscillator and dynamic power in the single ended coarse counter. A four stage ring oscillator was chosen to make the static power dissipated in the ring oscillator approximately equal to the switching power dissipated in the coarse counter. The four stage ring oscillator was made by connecting three inverters and a buffer in a chain, as shown in Figure 5-3.

The four stage ring oscillator has 8 possible states of nominally equal duration as shown in Figure 4-2. The least significant 3 bits are resolved. The divider value D for the phase locked loop is 136 given a system clock at the conversion rate of 3MHz. The frequency of oscillation of the ring oscillator is 408 MHz. This gives a t_{lsb} of 306psec and an overcount of $\frac{1}{3MHz} - 1024 * t_{lsb}$ or 19.6nsec. The overcount allows for the delay through the comparator and time needed to reset the ramp and track and hold.

Two ripple counters are interleaved to form the coarse counter as in Figure 5-2. This gives the transparent latch time to resolve the input from the ring oscillator and the coarse counter time to ripple that decision through. Since the two counters are interleaved, this resolution and ripple time does not deduct from the conversion time. The fine outputs are pipelined to match the latency in the coarse ripple counter and to increase the time for resolving the LSBs.

Section 5.2.2 describes the circuit design and issues involved in the ring oscillator and phase locked loop. Prominent issues include mismatches which cause DNL, jitter analysis and minimization, power reduction, and designing the loop to lock under all process, temperature and supply variations. Section 5.2.3 describes the fully-differential latches used to latch the state of the ring oscillator. Section 5.2.4 details the interleaved coarse counter implementation.

5.2.2 Ring Oscillator and Phase Locked Loop

The overall block diagram for the phase locked ring oscillator is shown in Figure 5-3. One output of the oscillator is buffered and converted from a fully-differential to a single ended signal. The divider was implemented in a single ended fashion. The output of the divider is fed to a phase detector with an off chip reference clock running at f_c=3MHz. The phase detector then causes the charge pump to pump up or down, depending on the phase error between the output of the divider and the off chip reference clock. The charge pump circuit pumps current into a loop filter, which changes the control voltage that determines the ring oscillator frequency. A higher frequency clock could also be used as the reference clock, if available on chip.

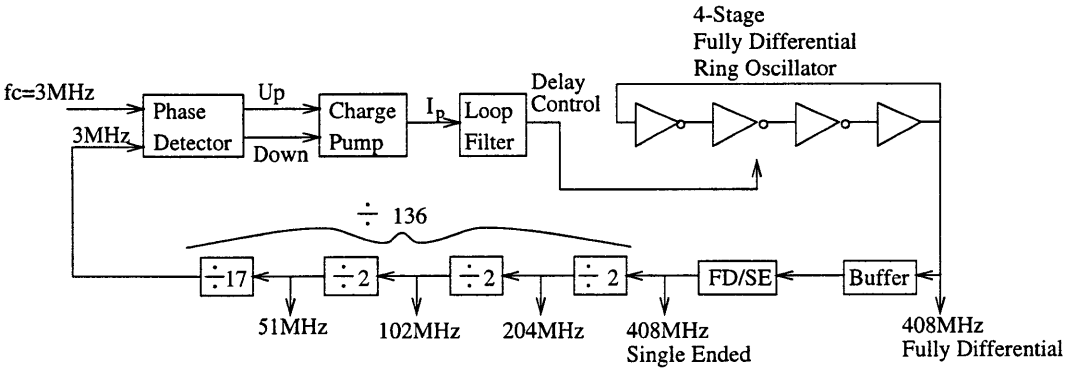
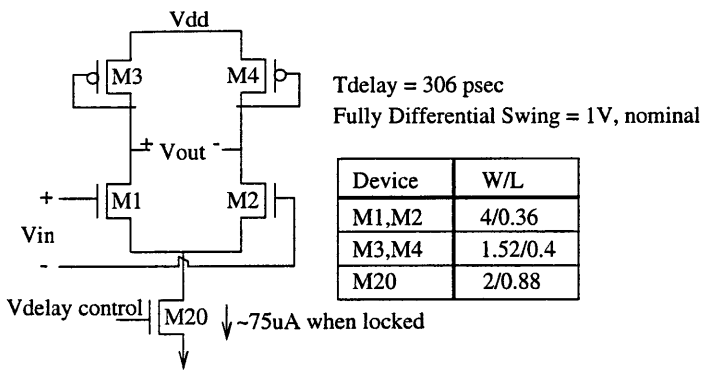


Figure 5-3: Phase Locked Loop Block Diagram

Ring Oscillator

A schematic of one fully-differential ring oscillator stage for the time digitizer is shown in Figure 5-4. The ring oscillator frequency is controlled by $V_{delaycontrol}$ changing the current through M20. M1 and M2 form a current-steering NMOS source-coupled pair with diode connected PMOS load devices. The output is shown such that this stage is an inverter. The output terminals are simply flipped for a non-inverting buffer connection, as in the last stage of the ring oscillator shown in Figure 5-3.

The output swing at the drains of M3 and M4 changes over temperature, process and supply by about 0.5Volts. This swing variation imposed some extra design challenges to subsequent circuits. Another solution would have been to actively bias M3



Device	W/L
M1,M2	4/0.36
M3,M4	1.52/0.4
M20	2/0.88

Figure 5-4: Ring Oscillator Stage

and M4 in triode, using a replica bias circuit for the gate voltages of M3 and M4 to keep the output swing of the ring oscillator constant [18].

A potential problem in this design is mismatch between the input-referred offset voltage of different oscillator stages causing a fixed difference in delay. This mismatch translates directly into variations in time measurement units, and results in DNL. Since M1 and M2 are minimum size, it can be expected that they have a large threshold mismatch. Using a worst case estimate of a 10mV mismatch and multiplying by the inverse of the simulated output edge rate of the stage near switching gives a delay mismatch of 23psec. This translates to 0.08LSB DNL ($\frac{23psec}{306psec}$). The channel lengths of M1 and M2 could be increased to minimize this mismatch induced DNL, but then more power (current through M20) would be needed to achieve the same delay. Since 0.08LSB DNL is sufficiently smaller than the 0.5LSB DNL specification, the channels of M1 and M2 were designed at minimum length to minimize power.

Mismatches in the M20 $\frac{W}{L}$'s of different stages can cause different currents between stages. This current difference translates into a time delay difference, and also results in DNL. However, M20 was designed with twice minimum channel length and width to reduce mismatches. Careful, symmetric layout was also done on the ring oscillator stages to reduce mismatches.

Jitter is an uncertainty in the delay of the ring oscillator stage. This uncertainty is caused by supply noise and thermal noise in the MOSFETS. Jitter is a time varying error that contributes to the noise floor of the A/D and reduces the converters effective

number of bits (ENOB). Assuming that jitter should be less than quantization noise (0.3LSB RMS [7]) gave a jitter specification of less than 0.15LSB RMS. Multiplying this by $T_{LSB} = 306psec$ and dividing by $\sqrt{1024}$ gave a $1.5\frac{psec}{stage}$ RMS jitter spec. The jitter was divided by $\sqrt{1024}$ because in the worst case, jitter is accumulated in quadrature over 1024 delays for a 10 bit measurement. This assumes that jitter in each stage is uncorrelated and that noise processes are rapidly varying.

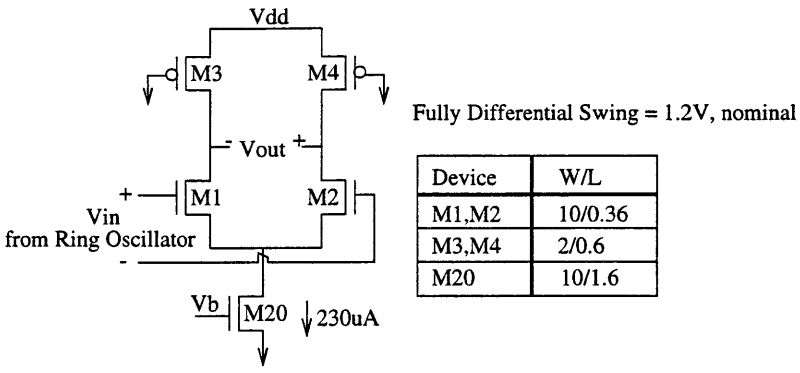
Jitter due to thermal noise in the ring oscillator stage of Figure 5-4 was simulated. A small signal noise analysis was performed to obtain a value for the RMS output voltage noise at the buffer stage switching point. This voltage value was multiplied by the inverse of the edge rate of the oscillator to approximate a delay uncertainty. This simulated uncertainty was 0.7 psec per stage, a factor of two lower than the specification.

As the resolution of the converter is increased, jitter will start to degrade the effective number of bits. At a resolution of 14 bits, noise caused by jitter becomes on the order of quantization noise. Although this diminishes the advantages of the linear tradeoff between conversion rate and resolution (ENOB) as described in Chapter 4, there is still a linear tradeoff between conversion rate and dynamic range.

Buffer

The fully-differential circuit shown in Figure 5-5 buffers the output of each ring oscillator stage. The buffers provide the ring oscillator with protection from kickback due to the operation of subsequent circuits, such as the latches or fully-differential to single ended converters.

Transistors M3 and M4 are biased in the triode region by grounding their gates. Again, the output swing of the buffers varies over process, temperature and supply. The buffers were designed to have slightly larger swing than the ring oscillator to ensure that there is enough signal to trigger the differential to single ended converters over these variations. Matching between buffer stages is important so that delays are the same and DNL is minimized. Assuming, as above, a worst case 10mV mismatch in threshold voltages, the DNL error due to buffer mismatch should be less than



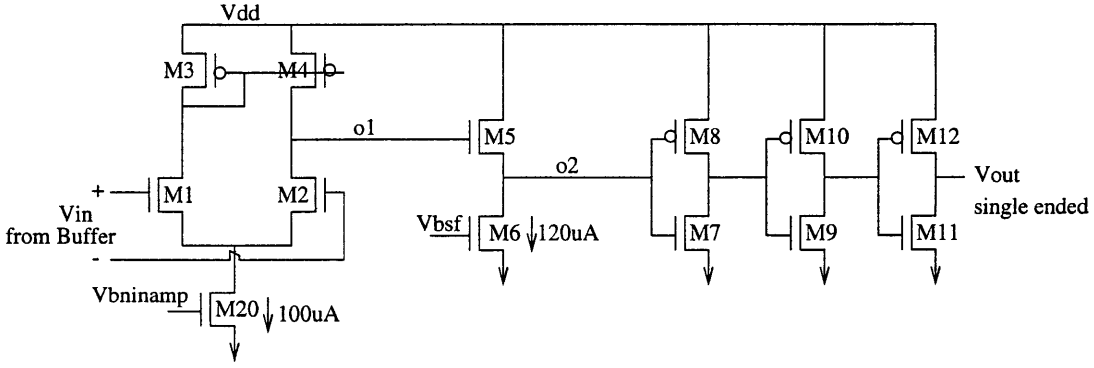
Device	W/L
M1,M2	10/0.36
M3,M4	2/0.6
M20	10/1.6

Figure 5-5: Buffer

0.1LSB. To minimize bias current mismatch, the current source transistor, M20, was designed to have a long channel length. Additionally, layout of the buffers was as symmetric as possible.

Differential to Single Ended Converter

The differential to single ended converter in Figure 5-6 converts the output of the ring oscillator buffer to a single ended signal for the phase locked loop (PLL) divider.



Device	W/L	Device	W/L	Device	W/L
M1,M2	1/0.36	M5	10/0.36	M7,9,11	1/0.36
M3,M4	2/0.4	M6	2/0.72	M8,10,12	2/0.4
M20	4/0.7				

Figure 5-6: fully-differential to single ended converter

In the circuit of Figure 5-6, the input source-coupled pair with active load (M1-M4) acts as an open loop op amp that converts the fully-differential signal to a

higher swing single-ended signal at node o1. Source follower M5-M6 level shifts o1 to be symmetric around the switching voltage of inverter M7-M8. Since the swing of the fully-differential input changes over temperature, process and supply variations, the swing of the single ended outputs o1 and o2 varies too. Care was taken to ensure that the swing of o2 included the switching point of the subsequent inverter chain over process, temperature, and supply variations.

Divider

The divide-by-136 in Figure 5-3 divides the ring oscillator 408MHz oscillation down to 3MHz for comparison to the system clock. The divide by 136 was implemented as shown in Figure 5-3 with three divide by twos and a divide by 17. There was a tradeoff between dynamic power in a single ended, full swing CMOS divider given by

$$P_{Single-Ended} \approx C * V_{dd}^2 * f \tag{5.1}$$

and power in a fully-differential, low swing divider given by

$$P_{Fully-Differential} \approx StaticPower(constant) + C * V_{dd} * V_{sw} * f \tag{5.2}$$

Where V_{sw} is the fully-differential output swing. At low frequencies, a full-swing, single ended divider will consume less power than a low-swing, fully-differential divider. At higher frequencies, the dynamic power of a single-ended implementation may approach the static power of a low swing, fully-differential divider. Thus, operating frequency determines if a fully-differential or single ended implementation saves power. A design tradeoff can be made where the high frequency stages are fully-differential and the low frequency stages are single ended. A low-swing, fully-differential implementation also has the advantage of minimizing the effect of switching noise coupled into the divider. In this design, the divider frequencies were low enough so that a full-swing single ended divider saves power over a low-swing, fully-differential divider. If the resolution of the converter were increased, or the number

of ring oscillator stages decreased, the frequency may become high enough that a low-swing, fully-differential divider makes sense.

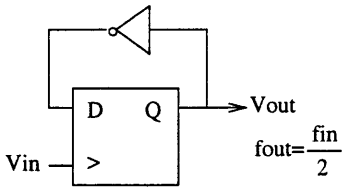


Figure 5-7: Single ended divide by 2

The single ended divide by two is shown in Figure 5-7. A standard cell T flip flop is used. On every period of Vin, Vout transitions from high to low or vice versa. This gives a division of two between the frequency of the input and the frequency of the output.

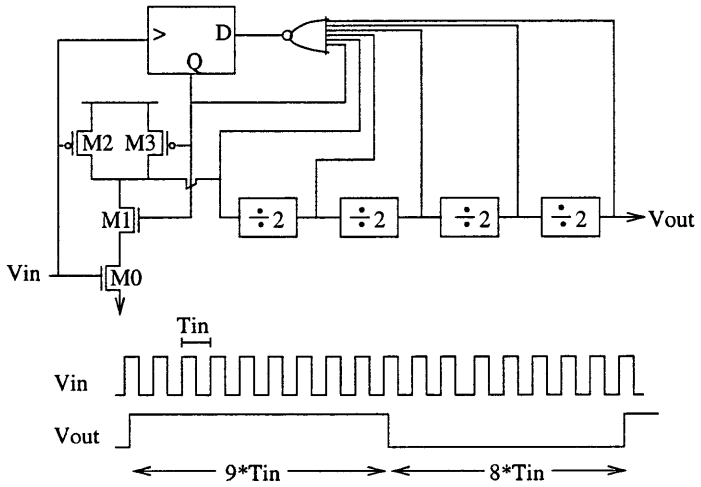


Figure 5-8: Divide by 17

The single ended divide by 17 of Figure 5-8 is based on [19]. Four divide by 2 circuits, implemented as shown in Figure 5-7, divide the input by 16 when Q is high (M1 is on, M3 is off, and M2 and M0 function as an inverter). However, on every cycle, when the "count" of the divider gets to a certain value, Q goes low for one cycle, causing M1 to turn off and M3 to turn on. This keeps the input to the divide by 2 chain high until the flip flop is clocked again. Thus, one period of the input

pulse is skipped. The duty cycle ($d = \frac{9}{17}$) of the output shown in Figure 5-8 does not affect PLL operation since the subsequent phase detector is only sensitive to the falling edges of its input.

Phase Detector

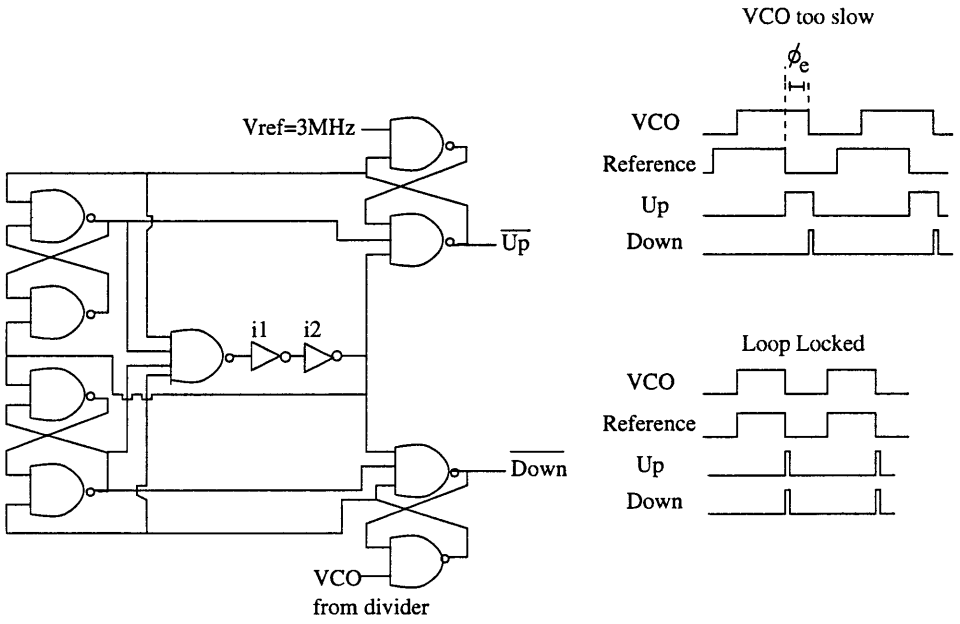


Figure 5-9: Phase Detector

The phase detector shown in Figure 5-9 compares the negative edges of V_{ref} and VCO. Up and Down output signals are activated for a duration proportional to the phase error between V_{ref} and VCO. These Up and Down pulses indicate if the charge pump should pump the control voltage up or down.

Two examples are shown in Figure 5-9. In the top example, the VCO frequency is lower than the reference frequency and the phase detector indicates that the charge pump should pump up. In the bottom example, the VCO and reference frequency are locked and the phase detector outputs short up and down pulses.

The phase detector must output these short up and down pulses when locked to avoid a "dead zone" in the phase detector transfer characteristic, as shown in the solid line of Figure 5-10. The phase detector transfer curve plots the phase error, ϕ_e between the VCO and V_{ref} verses Q, the total charge output of the charge pump. Q

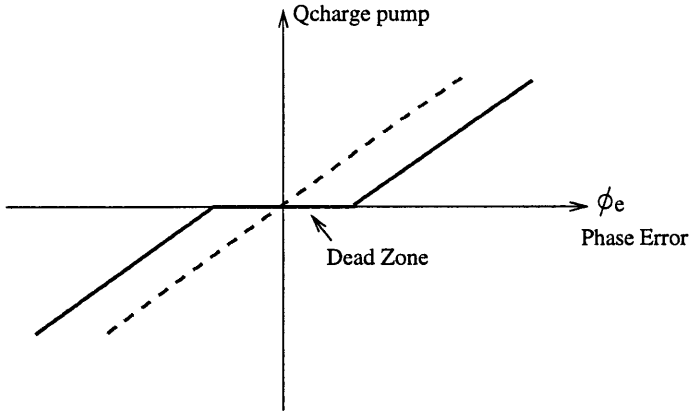


Figure 5-10: Dead zone in phase transfer characteristic

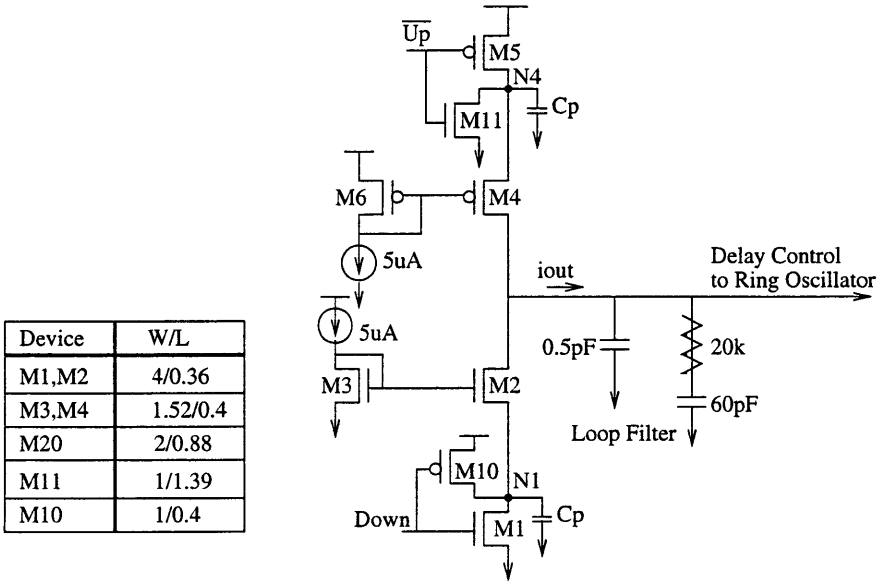
is proportional to how long the Up or Down signals are pulsed. A dead zone occurs when there is a range of $\emptyset_e$ over which there are no Up or Down pulses. This dead zone may be very small, but the phase locked loop can be "locked" anywhere in this zone, creating the potential for jitter in the ring oscillator. With the Up and Down pulses present even when in lock, the phase detector transfer curve becomes linear as indicated by the dotted line of Figure 5-10. For the phase detector of Figure 5-9, the dead zone is eliminated by inverters i1 and i2. These inverters add delay in the resetting of the up and down pulses, giving a short down and up pulse every time the phase detector is reset.

Charge Pump

The charge pump shown in Figure 5-11 is based on [20]. When the Up signal from the phase detector is pulsed, M5 turns on. This allows 5uA of current to be mirrored through M4. If Down is pulsed, M1 turns on allowing 5uA to be mirrored through M2. This current goes into the loop filter and changes the delay control voltage. If both Up and Down are pulsed at the same time, the 5uA should ideally cancel and there should not be a change in the delay control voltage.

Transistors M10 and M11 help reduce output current spikes when Up or Down are turned off. Transistor M10 turns on whenever Down goes low, pulling node N1 up to V_{dd}. M10 provides the current to pull up the parasitic capacitance C_p at node

N1. If M10 were not there to provide this current, the current would be drawn through M2, giving a current "spike" when Down turns off. Transistor M11 performs a complementary function when the Up pulse goes low.



Device	W/L
M1,M2	4/0.36
M3,M4	1.52/0.4
M20	2/0.88
M11	1/1.39
M10	1/0.4

Figure 5-11: Charge Pump

Loop Filter

The loop filter implementation consists of a resistor and two capacitors as shown in Figure 5-12. If only a single capacitor C were used, the loop would be unstable (see loop block diagram in next section). The resistor adds a zero to stabilize the feedback loop. Capacitor C3 provides additional high frequency filtering to limit the noise bandwidth. Values of these components were chosen to ensure good stability over temperature, process, and supply variations, as discussed in the next subsection.

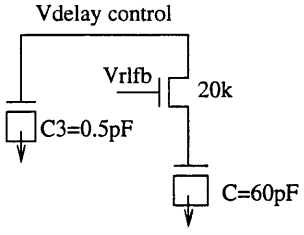


Figure 5-12: Loop Filter

The resistor was implemented with a MOSFET in the triode region. V_{rlfb} was

taken off chip so that the resistor value could be changed, but the nominal value of V_{rlfb} was V_{dd}. With V_{rlfb} constant, the value of the resistance changed by about 60% over variations in process and temperature. The capacitors C and C3 were implemented with MOSFETS biased in the inversion region. Their capacitances varied by about 8% over process and temperature variations.

Loop Analysis and Stability

The PLL parameter block diagram is shown in Figure 5-13. Since the PLL is designed to act as a bias generator for setting ring oscillator delays, the delay control voltage should be insensitive to high frequency jitter in the system clock or VCO. Thus, the PLL was designed to have a low loop bandwidth.

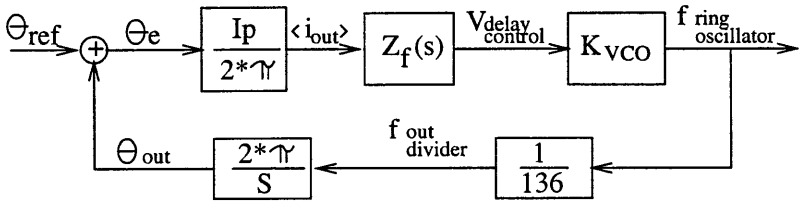


Figure 5-13: Phase Locked Loop Parameter Block Diagram

Since the bandwidth of the PLL was low compared to the signal frequency, the time varying nature of the charge pump can be averaged. The phase detector signals the charge pump to pump $I_p * sign(\theta_e)$ for a fraction $\frac{\theta_e}{2*\pi}$ of the time. This gives the average output current of the charge pump

$$< i_{out} >= \frac{I_p \theta_e}{2\pi} \tag{5.3}$$

This current is converted to the delay control voltage through the loop filter described above. The transfer function of this current to voltage transfer function, $Z_f(s)$, is given by

$$Z_f(s) = \frac{RCs + 1}{s(RCC_3 s + C_3 + C)} \tag{5.4}$$

The delay control voltage is then multiplied by K_{vco} to give the frequency of oscillation of the ring oscillator. This signal is divided by 136 and the phase error is compared with the reference.

The total loop transmission is

$$LT = \frac{I_p K_{VCO}(RCs+1)}{136s^2(RCC_3s + C_3 + C)} \tag{5.5}$$

$Z_f(s)$ is functioning as lead compensation. To maintain stability, the crossover frequency must occur after the zero in $Z_f(s)$ at $w = \frac{1}{RC}$, but before the pole in $Z_f(s)$ at $w = \frac{C_3+C}{RCC_3}$. The nominal loop gain bandwidth was designed to be 180kHz with a phase margin of 55°. Table 5.1 shows the loop unity gain bandwidth and phase margin over process, temperature and supply variations. This includes variations in K_{VCO}, R, C, and C_3.

Table 5.1: PLL Loop Crossover and Stability

Parameter	Low	Nominal	High
K_{VCO}	$930\frac{MHz}{V}$	$1476\frac{MHz}{V}$	$2212\frac{MHz}{V}$
R	33k	20k	13k
C	57.7pF	62.8pF	67.8pF
C_3	0.48pF	0.52pF	0.56pF
Unity-gain Bandwidth	180kHz	183kHz	200kHz
Phase Margin,ϕ_m	65°	55°	50°

5.2.3 Latches

A block diagram of the latches and coarse counter following the ring oscillator buffers is shown in Figure 5-14. The buffer outputs are latched by fully-differential latches. The latch outputs are converted to single ended and either pipelined through or counted.

The schematic for the master slave latch is shown in Figure 5-15. The latch inputs come from the output of the buffers connected to the ring oscillator. The clock input to the latch comes from the fully-differential output of the comparator,V_{strobe}. If the

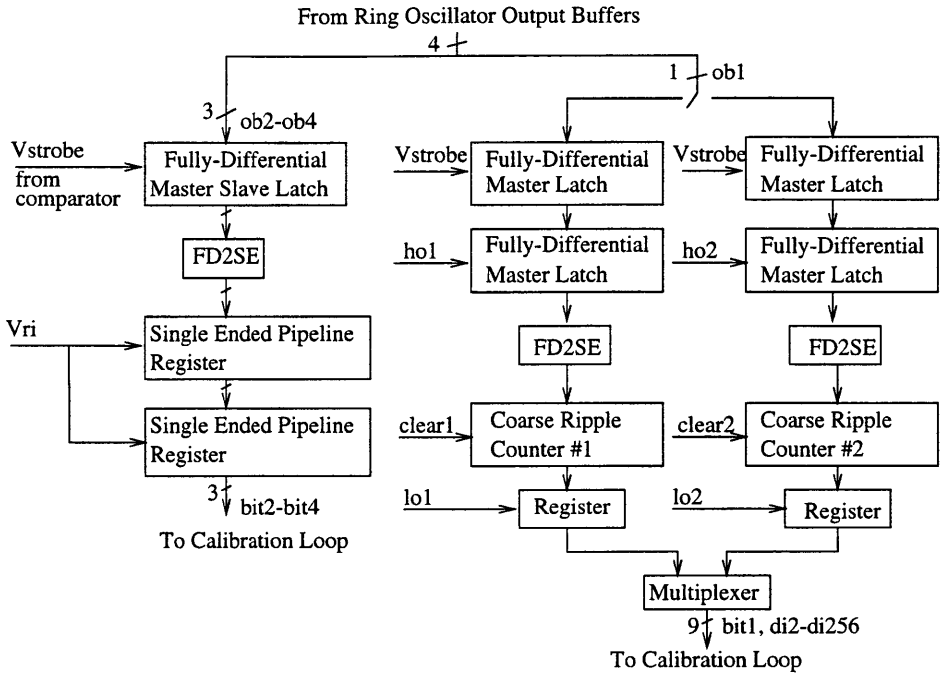


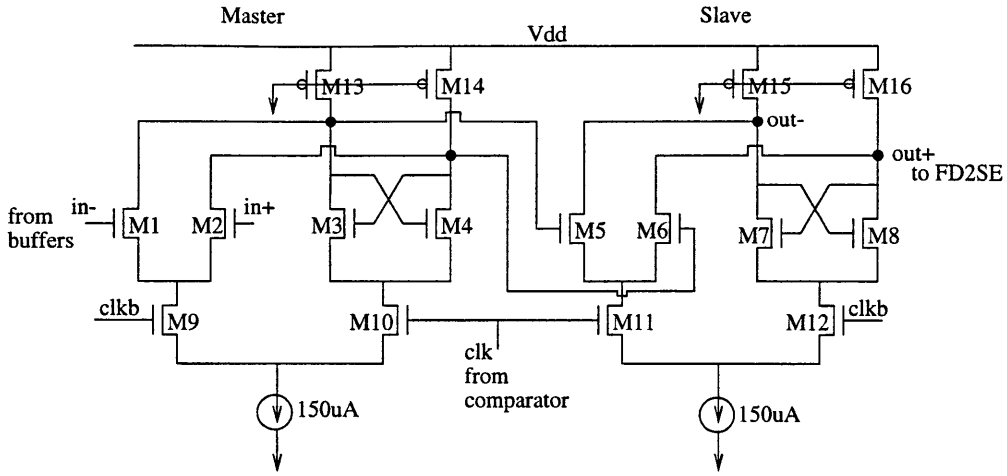
Figure 5-14: Block Diagram from Output of Buffers

comparator output is low, the master stage tracks the input (current through M9) while the slave is holding the previous output (current through M12). When latched, the master regenerates and the slave tracks the master output. Load transistors M13-M16 were biased in the triode region with their gates connected to ground. Simulations were performed to verify that the output swing was large enough to activate subsequent circuits over process, temperature, and supply variations.

Mismatches in the offset voltages of the latches can also create a DNL error. Assuming a 10mV worst case mismatch and dividing by the edge rate gave a DNL error of $\frac{23psec}{306psec} = 0.1LSB$. Layout of the latches was done symmetrically to minimize this mismatch.

The fully-differential to single ended converters are the same as those of Figure 5-6. Two single ended pipeline registers are used to match the latency through the coarse ripple counter. The coarse ripple counter latency is due to allowing the second latch time to resolve the metastable state and the registers that latch the final count.

In the coarse counter path, the output of the buffer is switched between two sets



Device	W/L
M1,M2,M5,M6	10/0.36
M9-M12	10/0.36
M3,M4,M7,M8	40/0.36
M13-M16	2/0.6

Figure 5-15: Fully-Differential, Master Slave Latch

of latches and coarse ripple counters. The first latch is a master latch as shown in Figure 5-15. This master latch is clocked with V_{strobe} from the comparator. Instead of a slave latch, another master latch is used so that the latches are transparent to the coarse counter when an input signal is being converted. The second master latches are clocked by ho1 and ho2. Ho1 and ho2 are latched versions of V_{strobe}. They are delayed by a few gate delays from V_{strobe} and are held until clear1 or clear2 are pulsed. lo1 and lo2 are pulsed half way through the clock period, giving half a conversion cycle for the second master latch to resolve and the count to ripple through.

When the master latch in Figure 5-15 enters positive feedback (regeneration), the output voltage increases approximately by the function [21]

$$V_o = AV_{in}e^{t/\tau} \tag{5.6}$$

Where $\tau = \frac{C}{g_m}$ is the time constant of the latch, t is the time spent in regeneration, and A is the gain of the latch in the transparent mode. Solving for V_{in} when V_o is on the threshold of detection gives a window of V_{in} over which the input is undetectable. The ratio of this window of V_{in} to the full scale input voltage is the percent chance

of a metastable state. For the master latches of Figure 5-15 given half a clock cycle to resolve (150nsec), the calculated percent chance of metastability, $1 * 10^{-13}$, much less than one pixel per frame ($1 * 10^{-5}$).

5.2.4 Coarse Counter

Two coarse counters are used in parallel as shown in Figure 5-16. The counters are interleaved in the A/D conversion. This is so that the latch has time to resolve the bit the counter is counting and the appropriate count has time to ripple through. If parallel counters were not used, the time for latch resolution and the counter delay would be subtracted from the time available to do the time digitization, necessitating an increase in the overcount (faster ring oscillator). The nominal delay through the ripple counter was 4.5nsec, giving the latch about 145nsec to resolve. Using the parallel counters resulted in a 2% increase in area and a 6% increase in power due to the extra latch and fully-differential to single ended converter.

The coarse ripple counter design is detailed in Figure 5-16. It consists of a string of eight divide by two cells with an asynchronous clear. The divide by two cells were implemented as in Figure 5-7. Since the three least significant bits were resolved in the ring oscillator, theoretically the coarse counter should only need seven stages. However, an extra stage was needed for the time digitizer overcount. The overcount allows for the comparator delay and resetting of the ramp and track & hold. When the delay through the comparator is subtracted from the digital output, this MSB (the 11th bit) should subtract to be zero. The outputs of the two coarse counters are multiplexed and sent to the calibration loop.

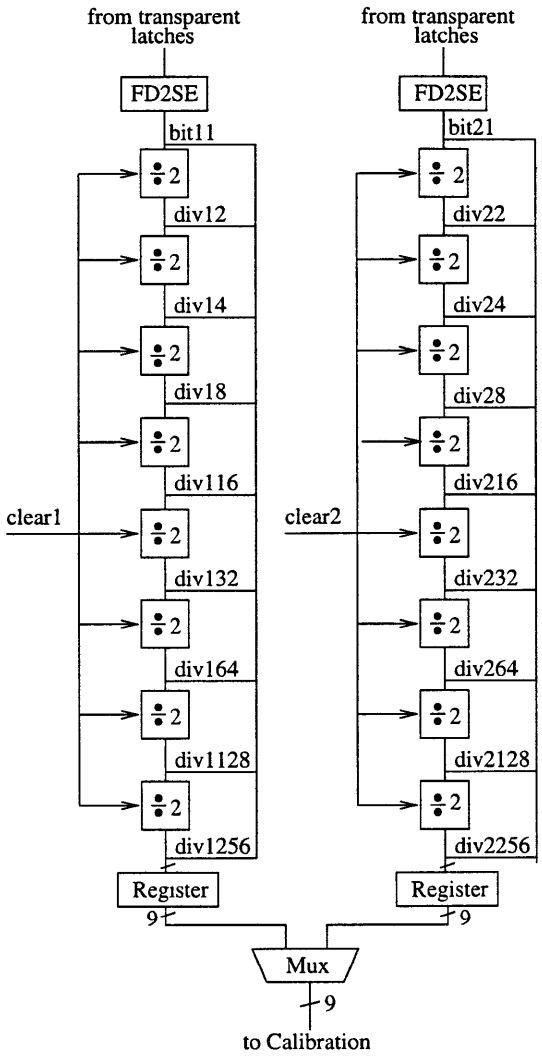


Figure 5-16: Coarse Counter

5.3 Track & Hold

The fully-differential input voltage from the PGA has to be held over the entire conversion period so that it can be compared with the ramp. However, any time spent sampling and settling takes time away from the time digitization, thereby increasing the overcount in the time digitizer and requiring the ring oscillator to oscillate faster. So, an interleaved track and hold architecture shown in Figure 5-17 is used. Note that in this figure, only the track and hold for V_{in+} is shown. There is another track and hold for V_{in-}.

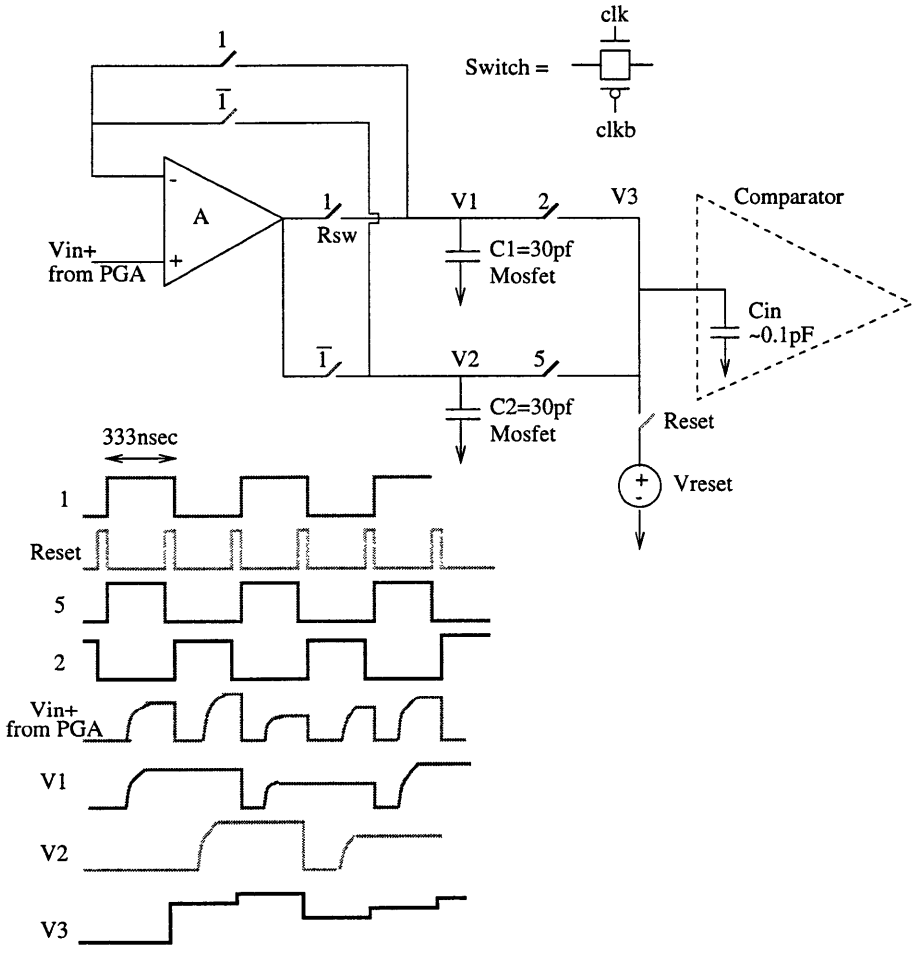


Figure 5-17: Interleaved Track & Hold

In typical interleaved A/D architectures, clock skew in the sampling instant has to be very small. In this application, the analog input voltage is held constant by the PGA and skew in the sampling instant is not an issue.

On Clock 1, C1 tracks the input and C2 connects the previous input value to the input of the comparator. In this manner, none of the time available for conversion is wasted settling a sample and hold. The resulting latency in the conversion time is not critical in imager applications.

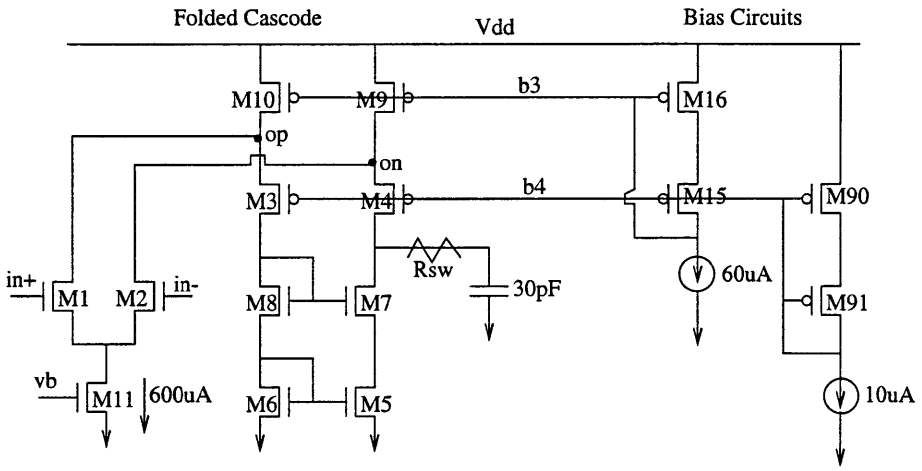
The charge injection and settling through the two paths in Figure 5-17 must be matched to greater than 10b accuracy to avoid "banging" at the endpoints of each A/D code. "Banging" refers to the situation where a constant input voltage near a code transistion is input to the track and hold. Two different output codes are given depending on which path the input went through. Basically, this translates into the need for charge injection matching between the two paths. The smaller the switches in the track and hold, the less the overall charge injection, but the larger the mismatches. The larger the switches, the more charge injection, but the smaller the mismatches in charge injection. 10b path matching was achieved by using a large 30pF sampling capacitor and small switches.

The charge sharing between C_{in} in Figure 5-17 and the sampling capacitors was turned into a constant offset voltage by resetting the input to V_{reset} before every conversion. Constant offsets (pedestal error) essentially change the input voltage common mode. This is only a problem if the common mode mismatch between the ramp and the input becomes larger than 83mV (see comparator section for more discussion). Since C1 was so large, offset error due to both clock feedthrough and charge sharing resulted in a worst case common mode variation of 5mV.

The amplifier feedback configuration in Figure 5-17 includes the sampling switch in the feedback path to decrease the $R_{sw} * C1$ settling time constant. Assuming that $R_{sw} * C1$ is the dominant pole compensating the opamp, the effective switch resistance is decreased by a factor of the loop gain. In this case, $R_{sw} = 500\Omega$ was decreased to less than 1 Ohm.

To achieve 10b (0.1%) settling in 300nsec, an amplifier with a unity gain bandwidth of at least 4MHz was needed. The folded cascode amplifier shown in Figure 5-18 was designed to have a 26MHz unity gain bandwidth. The 30pF sample capacitor provided dominant pole compensation. Nondominant poles at nodes op and on were

well above crossover, giving the amplifier good unity gain stability. Bias circuits M15, M16, M90, and M91 biased the folded cascode to maximize output swing.



Device	W/L	Device	W/L
M1,M2	600/0.36	M16	11/0.72
M11	67.2/0.72	M15	2.9/0.4
M10,M9	110/0.72	M90,M91	5/5
M3,M4	29/0.4	M5-M8	26/0.52

Figure 5-18: Folded Cascode Amplifier

5.4 Ramp Generator

A linear ramp is needed to compare to the output of the track and hold. When the ramp crosses the input level, the comparator trips and the ring oscillator state is latched. Ramp linearity is directly related to the INL and DNL of the A/D transfer curve.

The fully-differential ramp consists of a constant current source charging and discharging a capacitor, as shown in Figure 5-19. At the beginning of each conversion, in1n was reset to 1V and in2n was reset to 2V. When the conversion period started, the reset switch was opened and in1n ramped up to 1V while in2n ramped down to 1V. The charging and discharging currents were calibrated during the blanking intervals.

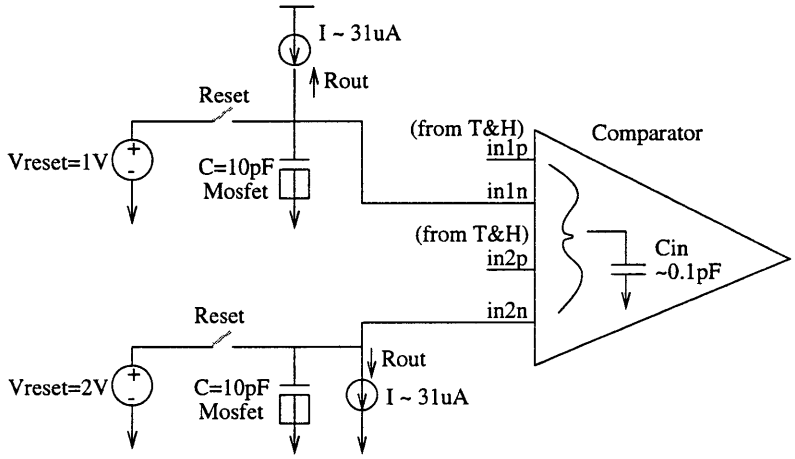


Figure 5-19: Ramp Generator Model

Assuming the converter is calibrated, the ramp rate is fixed at $\frac{dv}{dt} = \frac{1V}{333nsec} = \frac{I}{C}$. From a power and area point of view, C and I should be minimized. However, C needs to be large enough so that the capacitance at the ramp node is not dominated by nonlinear parasitic junction capacitances or by the input capacitance of the comparator. Assuming that the nonlinear parasitic capacitance C_p at nodes in1n and in2n changes by no more than 10% over the ramp voltage range (1-2V), C is made one hundred times larger than C_p. This gives a worst case capacitance change of 0.1% over the ramp voltage range. Additionally, C should be large to minimize the effects

of thermal noise. Once C was chosen, I was calculated by multiplying by the ramp rate.

The minimum current source output resistance was determined using the model of Figure 5-20. This minimum value ensures that most of the current goes to charging the ramp capacitor instead of being dissipated through R_o. Solving for in2n as a function of time,

$$in2n = (2 + IR_o)e^{\frac{-t}{R_oC}} - IR_o \tag{5.7}$$

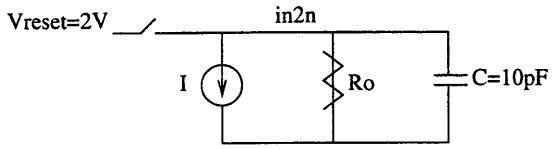


Figure 5-20: Ramp Model

This assumes that at $t = 0$, in2n was reset to 2V. When the converter is calibrated, I will be set so that $in2n(t = 333n) = 1V$. This gives

$$I = \frac{1 - 2e^{\frac{-333n}{R_oC}}}{R_o(e^{\frac{-333n}{R_oC}} - 1)} \tag{5.8}$$

Plotting in2n as a function of R_o gave the curves in Figure 5-21. The DNL caused by the ramp nonlinearity is small, but the INL is large in the middle of the curve. A similar analysis was conducted for in1n and the INL was computed. INL curves for various values of R_o are shown in Figure 5-22. These INL curves are plotted as a function of time from the ramp voltages in Figure 5-21. Assuming the output resistance R_o of the ramp up and ramp down currents is the same, R_o had to be at least $850k\Omega$ to achieve less than 5LSB INL.

An even tighter constraint on the output resistance of the current source is the sensitivity of the ramp to supply voltage variation. Assuming a worst case supply variation of 100mV, the current should not change by more than 3nA to achieve an output error of less than 0.1LSB. This gave a minimum current source output

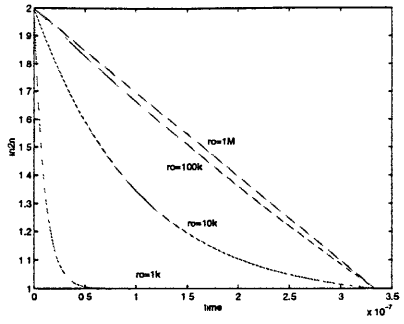


Figure 5-21: Ramp Voltage versus Time

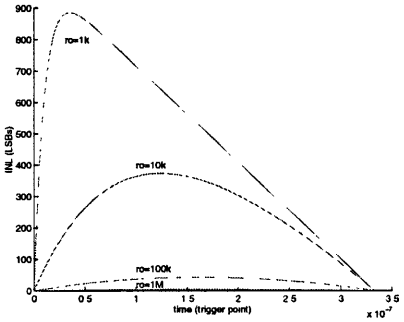


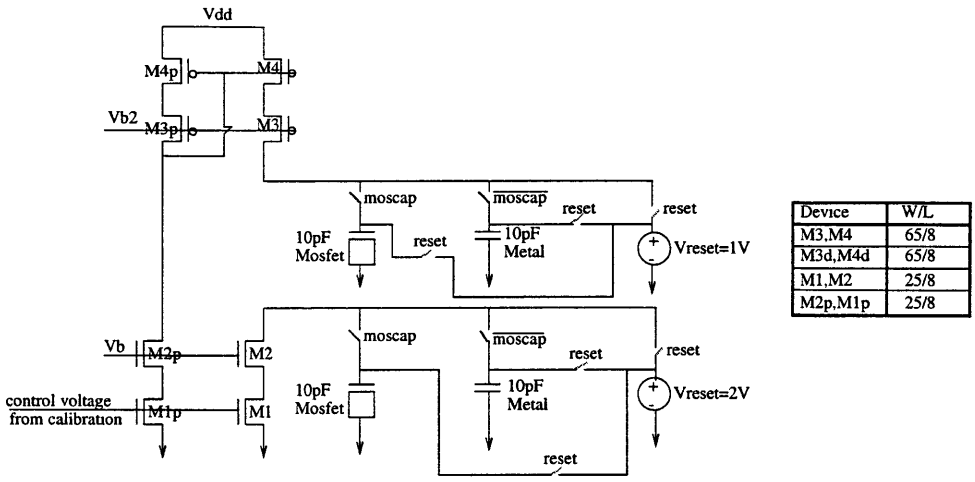
Figure 5-22: INL versus Time(Trigger Point or Vin)

resistance spec of $33M\Omega$, much larger than the INL constraint.

The ramp implementation is shown in Figure 5-23. The control voltage from the calibration loop sets the current through a cascoded current source. That current is mirrored through M1p-M4p to provide charging and discharging currents. Vb and Vb2 are set to maximize the output swing of the cascode. The output resistance for both the charging and discharging current sources is nominally $50M\Omega$.

A MOSFET capacitor was chosen to implement the ramp capacitors. MOSFET capactiors are typically a factor of 20 smaller than metal capacitors [14]. However, they are known for their poor voltage coefficient, on the order of $28\frac{kppm}{V}$ [14]. Since the ramp is implemented fully-differentially, even order harmonics of the capacitor nonlinearity cancel. On the chip, 10pF metal ramp capacitors are also built from a stack of poly, M1, M2 and M3. The MOSFET and metal capacitors are in series with low resistance switches so they can be switched for testing. As seen in the layout section, the metal capacitors are a factor of 26 larger than the MOSFET capacitors.

The ramp reset switches are large to minimize the RC settling of the reset.



Device	W/L
M3,M4	65/8
M3d,M4d	65/8
M1,M2	25/8
M2p,M1p	25/8

Figure 5-23: Ramp Implementation

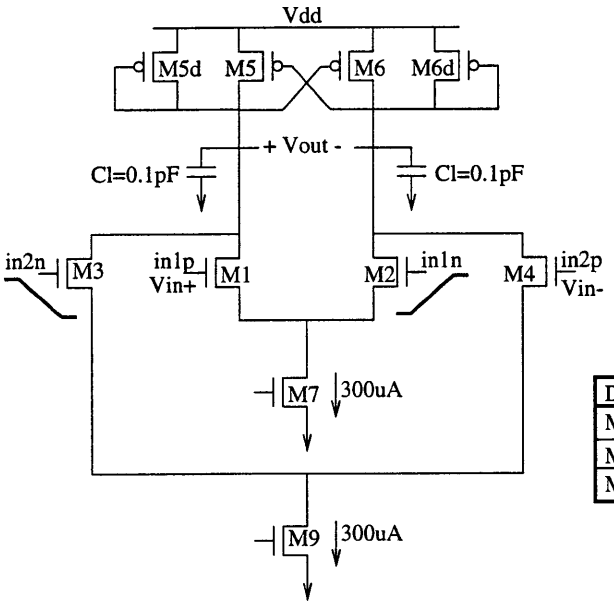
Switches are built from V_{reset} to both the comparator input and to nodes with the capacitors. As a result, the switches that choose between MOSFET or metal capactiors are not critical in the settling time. The switches were designed so that 0.1% settling of the 10pF MOSFETs is achieved in the reset pulse time of 5nsec. Simulations indicate that there should be less than 0.3LSB INL and 0.01LSB DNL due to both the finite ramp output resistance and capacitor nonlinearity.

The amount of change in the ramp due to thermal noise in the cascode current source was simulated. The ramp voltage is reset at the beginning of each conversion, so changes in ramping current which cause a change in the slope of the ramp are critical. Additionally, since the ramp is calibrated at the end of each line ($\frac{3MHz}{356} = 8kHz$), only noise above 8kHz is of interest. This assumes noise below 8kHz is small enough so that there are enough calibration steps at the end of each line to completely calibrate for it. Simulations indicate that over a line, the nominal current noise is 1.3pA. This number was obtained by multiplying the input referred voltage noise by g_m of M1p. This means that noise will cause a shift of $4 * 10^{-5}$ LSB in the A/D transfer characteristic over a line time.

5.5 Comparator

The fully differential comparator design is shown in Figure 5-24. The initial comparator design had one source-coupled pair connected to the output of the track and hold. The other source-coupled pair was connected to the ramp generator. The comparator will flip when the difference between the input voltage and the fully-differential ramp are the same,

$$V_{out} \propto (V_{ramp+} - V_{ramp-}) - (V_{in+} - V_{in-}) \tag{5.9}$$



Device	W/L
M7, M9	63/0.88
M1-M4	216/0.36
M5,M6,M5d,M6d	12.48/0.52

Figure 5-24: Comparator

However, connecting the comparator inputs in this manner means that the linear range of the source coupled pairs has to equal the entire differential input voltage range of 2V. If designed for such a large linear input voltage range, the comparator will switch slowly. So, the comparator inputs were implemented as in Figure 5-24, with V_{in+} being compared to the ramp sloping up and V_{in-} with the ramp sloping down. This is essentially a rearrangement of Equation 5.9,

$$V_{out} \propto (V_{ramp+} - V_{in+}) - (V_{ramp-} - V_{in-}) \tag{5.10}$$

This increases the linear input range of the comparator to allow for faster current switching (lower comparator delay) as the ramp voltage passes the sample and hold voltage. For example, if the mismatches are such that M1 and M2 start switching before M3 and M4, M3 and M4 must start switching before M1 and M2 completely finish switching. If M3 and M4 do not start switching in time, the comparator may sit at a metastable state and the slope of the output at the switching threshold will be low, leading to comparator delay jitter. The linear input range of the source coupled pairs is around 167mV. This gives a maximum mismatch between the track and hold output common mode and the ramp common mode of 84mV. The change in common mode between the input and output of the track and hold is 5mV. This is due to the opamp offset and switch charge injection. The latter is minimized by the large track and hold capacitor C1 in Figure 5-17. The ramp common mode is determined by the starting point voltages. These voltages come from off chip.

The linear range of the comparator also has to allow for mismatches in the ramp current mirror. If the common modes of the ramp and input are matched, the maximum current difference such that the source coupled pairs overlap their linear regions is 6uA, or 20% of the total current. Cascode current mirror matching was simulated to be within 6%.

The delay of the comparator needs to be within the overcount of the time digitizer. The delay is designed to be less than 10nsec over temperature, process, and supply variations.

An even more challenging design goal for the comparator is minimizing output jitter. The design goal is an output jitter less than 0.1LSB (below quantization noise). If jitter is noise divided by edge rate, a large edge rate is desirable. Near the switching point, the cross coupled load M5,M6 cancels the $\frac{1}{g_m}$ of the diode connected transistors M5d and M6d. This gives a load output resistance of $\frac{R_o}{2}$. The edge rate near switching is determined from Figure 5-25 where the current ΔI_d is given by [22]

$$\Delta I_d = \frac{u_n C_{ox} W V_{id}}{2L} \sqrt{\frac{2I_{ss}}{u_n C_{ox} \frac{W}{2L}} - V_{id}^2} \tag{5.11}$$

where V_{id} is the fully differential input voltage. Assuming $V_{id} = Kt$ is a small ramp near switching and simplifying gives

$$\Delta I_d = \frac{V_{id} g_m}{\sqrt{2}} = \frac{K g_m t}{\sqrt{2}} \tag{5.12}$$

Where K in equation 5.12 is the ramp rate of the input. Solving the circuit of Figure 5-25 for V_{out} with the ramp input gives

$$V_{out} = \frac{R_o K g_m}{\sqrt{2}} (t - T + T e^{-\frac{t}{T}}) \tag{5.13}$$

where $T = R_o C_l$ is the time constant of the circuit of Figure 5-25.

Assuming the input has been ramping for a long time and initial transients have died out,

$$\frac{dV_{out}}{dt} = \frac{K g_m R_o}{\sqrt{2}} \tag{5.14}$$

The output voltage noise is given by

$$\overline{V_{out}} \approx \sqrt{4kT \frac{2}{3} \frac{g_{m1} R_o}{C_l}} \tag{5.15}$$

Dividing the output voltage noise of Equation 5.15 by the edge rate in Equation 5.14 gives

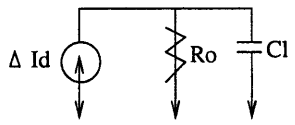


Figure 5-25: Small Signal Model for determining output edge rate

$$t_{jitter} \approx \frac{\overline{V_{out}}}{\frac{dV_{out}}{dt}} \approx \frac{\sqrt{4kT\frac{2}{3}\frac{g_{m1}R_o}{C_l}}}{\frac{Kg_mR_o}{\sqrt{2}}} \propto \sqrt{\frac{1}{Kg_mR_oC_l}} \tag{5.16}$$

This indicates that a low jitter comparator should be designed with high gain (g_mR_o) and large output capacitance C_l. This translates to a low noise, low bandwidth comparator design. Considering $t_{jitter} * t_{delay}$ as a figure of merit where $t_{delay} \propto R_oC_l$,

$$t_{jitter} * t_{delay} \propto \sqrt{\frac{R_oC_l}{Kg_m}} \tag{5.17}$$

This indicates a low jitter, low delay comparator, should have a large bandwidth ($\propto \frac{1}{R_oC_l}$) and g_m. The comparator jitter was designed to be less than $0.1t_{lsb} = 30psec$ with a nominal delay of 6nsec.

When simulating the operation of the comparator with the ramp and track & hold, kickback from the comparator switching causes changes in the ramp and track & hold voltages. Note that kickback occurs as the comparator flips and does not change the inputs before they determine the time of switching (assuming that the common modes of the ramp and track & hold are matched as above). As long as the track & hold and ramp reset operations settle so that kickback does not effect the value of the inputs on the next cycle, comparator kickback will not be a problem.

5.6 Calibration

When performing an A/D conversion, the calibration loop for the A/D digitally subtracts the delay through the comparator from the count of the ring oscillator and coarse counter. When calibrating at the end of each line and the end of each frame, the calibration loop subtracts the digital output when the input is full scale from the comparator delay. This number is compared to 1024, and the control voltage for the ramp is increased or decreased accordingly.

The calibration loop is shown in Figure 5-26. The inputs (di2-di256, bit1-bit4) to the calibration loop come from Figure 5-14. When not calibrating (clb=0), the decoded outputs of the ring oscillator (LSBs) are combined with the output of the coarse counter (MSBs), and the comparator delay D_d is subtracted. D_d is stored during the calibration cycle at the end of each line. The MSB is dropped from the subtractor output (it should be zero if the A/D is calibrated) and the 10b digital output is sent off chip.

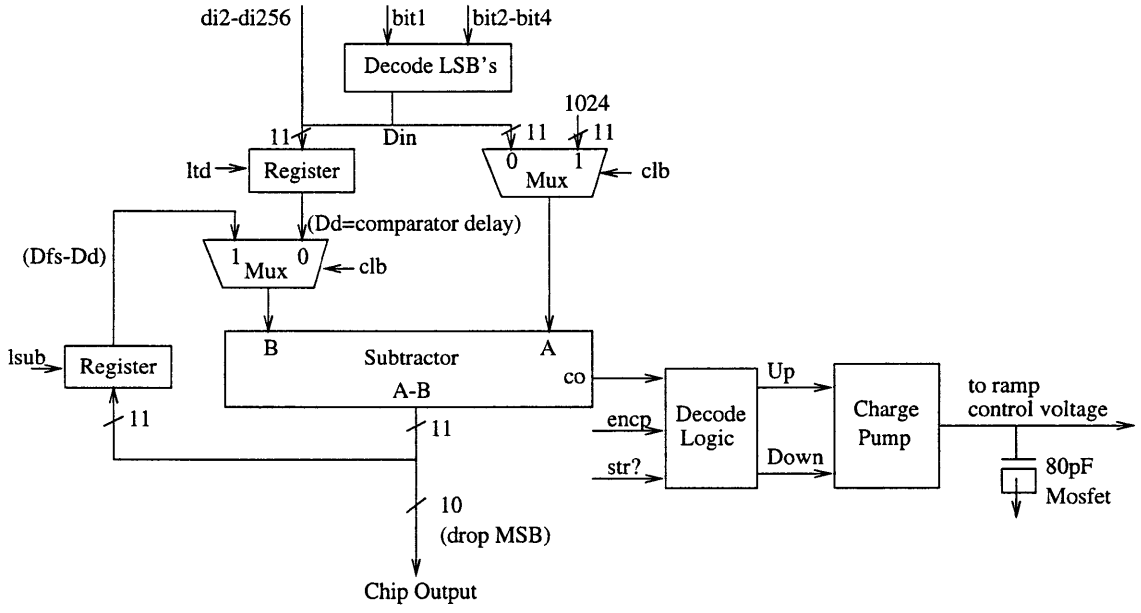


Figure 5-26: A/D Calibration loop

When calibrating, a series of steps are taken. First, the negative full scale input

voltage is switched to the track & hold. l_{td} is pulsed and the digital output (D_d) representing the comparator delay and any initial ramp overshoot is latched. Then, the positive full scale voltage is switched to the input. l_{sub} is pulsed and the difference between this digital output (D_{fs}) and D_d is latched. *Clb* is pulsed and $D_{fs} - D_d$ is subtracted from the desired output of 1024. The carry out of the subtractor (*co*) indicates if the actual output was less than or greater than 1024.

co is combined with *str?* and *encp* to determine if the charge pump should pump up, down, or not at all. *str?* indicates if the comparator tripped when the positive full scale voltage was input. If the comparator did not trip, the ramp is slow and the charge pump pumps up. *encp* is a pulse that activates the charge pump. The length of the *encp* pulse determines how long the charge pump is enabled, and thus the incremental change in voltage at the ramp control voltage node (see also Figure 5-23).

Once the ramp control voltage is updated, calibration starts again. This loop is performed until the end of the calibration time. Between lines, about 16 calibration cycles can be performed. Between frames, 1750 calibration cycles can be performed. Assuming at startup that the ramp control voltage is far away from its locked value, there is a startup time of a few frames. This start up time is not critical in consumer product applications.

The nominal increment of the ramp control voltage is designed to change the A/D transfer curve by 0.1LSB. This step size was chosen to minimize mismatches in the A/D transfer characteristic between lines (calibration is performed at the end of each line). This gives a calibration range between lines of 16*0.1LSB=1.6LSB. The "droop" in ramp control voltage due to leakage during a line was designed to be less than 1LSB, so this droop should be completely calibrated out between lines.

All of the digital calibration circuitry of Figure 5-26 are implemented in static CMOS circuits. Switching frequencies are on the order of 3MHz, so power dissipation from the calibration circuitry is small.

5.7 Layout

The single slope based A/D was laid out in the Lucent Technologies 0.35um standard CMOS technology. This process had three levels of metal and one layer of poly. The layout was completed in Magic and verified by extraction and simulation in ADVICE, which is a SPICE-like simulator.

Layout of the sensitive analog blocks was done symmetrically from all angles. This minimizes device mismatch due to varying implant angles. Any metal routed over sensitive analog circuits can also cause processing mismatches. Thus, routing was done around the sensitive analog blocks.

In the ring oscillator, unequal delay due to device or routing mismatch couples directly to DNL. The connection of ring oscillators shown in Figure 5-28 minimizes delay variation due to routing mismatch.

Sensitive analog circuits were built near each other. For instance, the ring oscillator, buffers, fully-differential latches, and fully-differential to single ended converters were built in a row to minimize routing delay and mismatches due to routing. Digital circuits were built together above these sensitive analog circuits. There is no separate analog power supply, so power supply noise due to switching had to be taken into account in the analog circuit design.

The layout of digital circuits was done to minimize area, not mismatches. Routing was often placed over transistors and devices were not always laid out symmetrically.

The breakdown of area consumption from most area to least area is shown in Table 5.7. The 80pF capacitor at the ramp control voltage node and the standard cell digital circuits in the calibration loop dominate the area consumption. This is followed closely by long channel length ramp current sources and 10pF charge/discharge capacitors. The track & hold area is dominated by the four 30pF MOSFET hold capacitors. The bias circuit area is dominated by decoupling capacitors. Area consumption of the other analog and digital circuits is very small.

The area of the chip ($0.3mm^2$) is small, primarily due to the use of MOSFET capacitors. Since this design does not depend on capacitor matching or linearity,

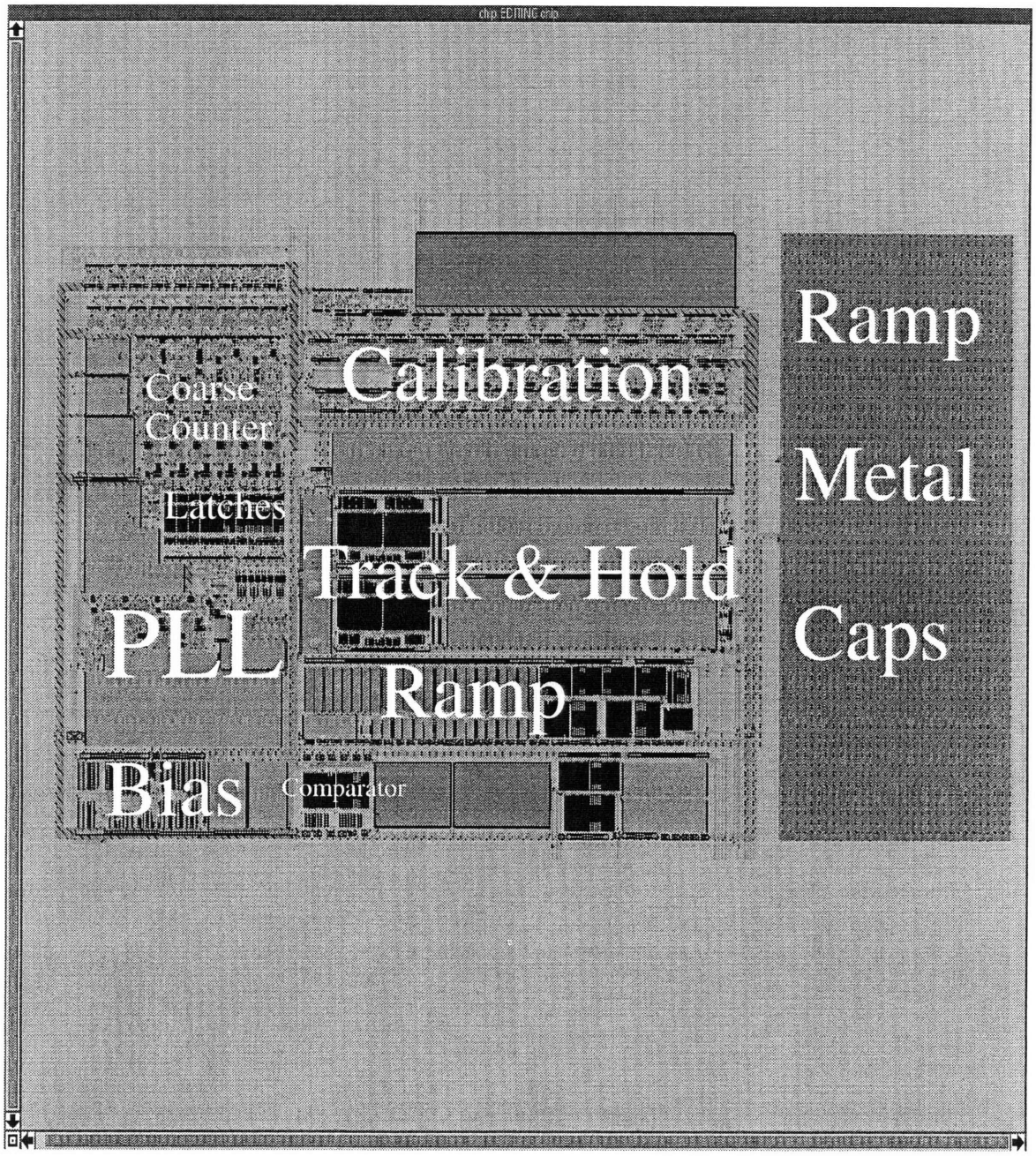


Figure 5-27: A/D Layout

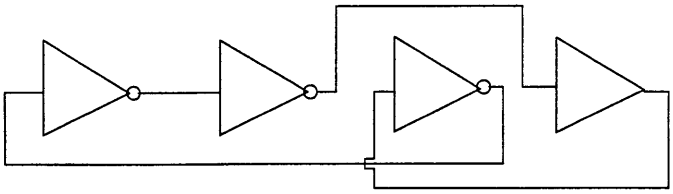


Figure 5-28: Ring Oscillator Routing

Table 5.2: Chip Area Breakdown

Calibration	$0.08mm^2$	27%
Ramp	$0.06mm^2$	20%
Track & Hold	$0.055mm^2$	18%
Bias	$0.053mm^2$	18%
Ring Oscillator & PLL	$0.025mm^2$	8%
Divider & Coarse Counter	$0.012mm^2$	4%
FD Latches	$0.008mm^2$	3%
FD2SE	$0.006mm^2$	2%
Comparator	$0.005mm^2$	2%
Buffers	$0.0025mm^2$	1%
Total Area	$0.3mm^2$	100%

MOSFET capacitors are used to save area. In the case of the ramp discharge capacitors where linearity is important, metal capacitors are built on the right hand side of the layout. Built as a poly/M1/M2/M3 sandwich, these 10pF capacitors are a factor of 26 larger than the MOSFET capacitors. The metal capacitors can be switched with the MOSFET capacitors as shown in Figure 5-23. They are built so that the effect of the nonlinearity of the MOSFET capacitors on the A/D transfer characteristic can be tested.

Chapter 6

Extension of Design to 12MHz

The 10b, 3MHz A/D converter designed in this thesis can be extended in several ways to a 10b, 12MHz converter. A 10b, 12MHz converter is necesssary for increasing the imager array size to four times CIF (704x576). Two approaches that can be used to increase the conversion rate are described in this chapter. The first approach, increasing timing (LSB) resolution, can be achieved by running the ring oscillator faster or by interpolating between ring oscillator gate delays. The second approach involves a coarse subranging of the input voltage before the new single slope converter. Power and area numbers for these architectures are estimated from the 3MHz design and are compared with previously reported 10b, 12MHz power/area figures of merit in Chapter 8.

6.1 Increasing LSB Resolution

T_{LSB} of the time digitizer can be reduced by a factor of four, from 306psec to 76psec, to increase the conversion rate of the A/D from 3MHz to 12MHz. This can be accomplished by running the ring oscillator faster or by interpolating between ring oscillator edges.

6.1.1 Faster Ring Oscillator

Increasing the speed of the ring oscillator is a straightforward way to increase the conversion rate of the A/D. However, at high ring oscillator frequencies, substrate bounced produced by the ring oscillator and subsequent dividers will increase. The power dissipated in the ring oscillator, divider, and latches will also increase. In this case, more of the divider circuit design might be fully-differential to reduce dynamic power and substrate bounce. This approach is ultimately limited by the minimum stage delay achievable in a given technology. Simulations show that t_{LSB}=76psec is not achievable over temperature, process, and supply variations. However, as technology is scaled, t_{LSBmin} can be reduced and the conversion rate of the A/D increased.

6.1.2 Interpolation Between Edges

Time digitizer performance can also be increased by interpolating between ring oscillator edges to decrease the effective t_{LSB}. Knotts [16] used the weighted summer shown in Figure 6-1. This summer generates an interpolated edge between two given edges (tune=0). An effective t_{LSB} of 15.6psec is achieved in bipolar technology by subdividing ring oscillator edges into four equal bins. A tune voltage calibration was needed to set the location of the interpolated edge.

Using this approach to expand the A/D to 12MHz would require 2 levels of interpolation as shown in Figure 6-2. The first phase interpolator generates an edge between adjacent ring oscillator edges. The edges themselves are also buffered (tune=1 or tune=-1) to preserve a constant delay through all paths. For example in Figure 6-2, F is the edge interpolated between edges A & B and E and G are buffered versions of A and B. The outputs of the first phase interpolator are interpolated and buffered again by the second phase interpolator to produce an effective t_{LSB} one-fourth of the time between the original ring oscillator edges. Note that the maximum frequency on the chip is still the original ring oscillator frequency.

The extra hardware cost includes 24 weighted summers, 12 additional fully differential latches, and 12 additional fully differential to single ended converters. The

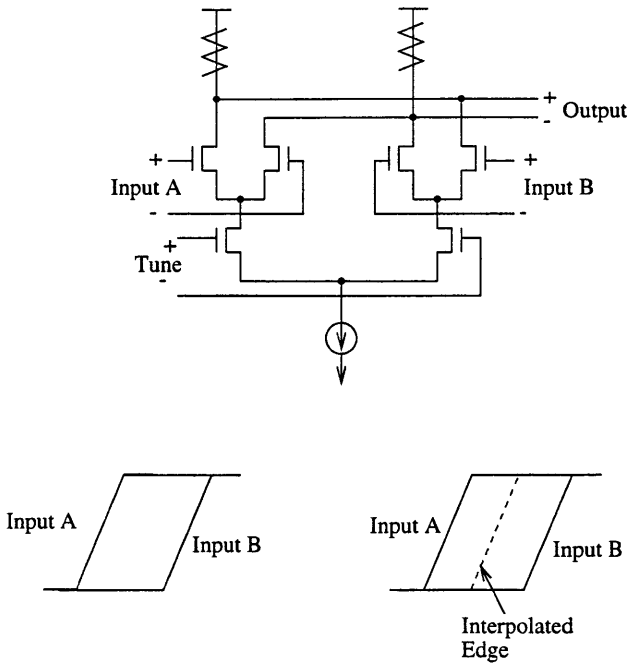


Figure 6-1: Weighted Summer used as an Interpolator

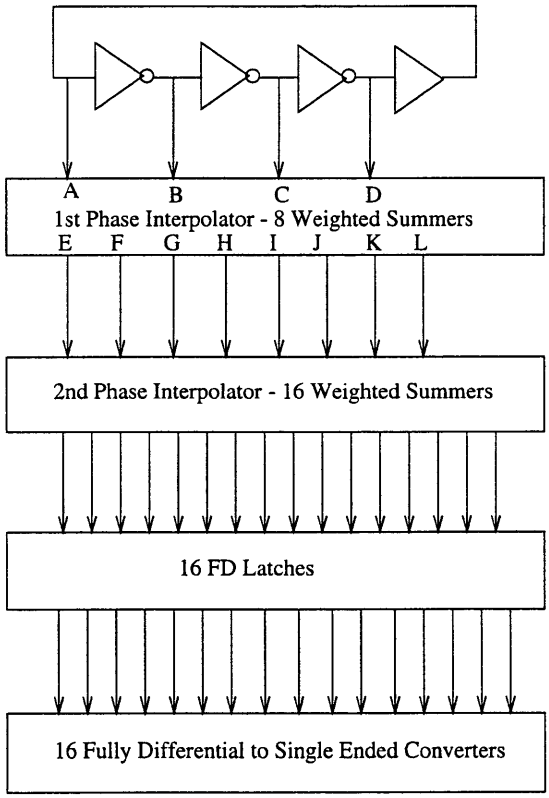


Figure 6-2: Architecture using Interpolators

power dissipated in the weighted summers was estimated to be equal to the power dissipated in the ring oscillator stages. The additional power dissipated in the fully-differential latches and fully-differential to single ended converter was estimated to be the same as the power dissipated in the circuits already designed. This gave a total estimated power dissipation of 53mW. Some additional power may also need to be dissipated in the track & hold to ensure that it acheives 12b settling. This power should be small copmared to the extra power of the latches and fully-differential to single ended converters. The additional area cost is estimated around 50%, giving a total area of $0.45mm^2$.

6.2 Subranging

Another strategy for increasing the speed of the converter is to divide the input into subranges. In other words, a coarse A/D conversion is performed to determine which subrange the input is in. The single slope conversion is then performed to resolve the fine bits. If the input is divided into 4 subranges, the conversion rate of the A/D can be increased from 3MHz to 12MHz. One potential problem with subranging techniques is calibrating the alignment of the endpoints of each subrange.

6.2.1 Tri-Slope Converter

A tri-slope converter combines a coarse and fine single slope conversion as shown in Figure 6-3. On clock 1, Vramp is initialized to Vss. Then I_{coarse} is switched in parallel with the ramp capacitor. The ramp capacitor is charged for a fixed number of Tlsbs. In this case, 4 coarse subranges are needed and C_{ramp} is charged for 1-4 Tlsbs, depending on when V_{ramp} crosses V_{in}. When V_{ramp} crosses V_{in}, clock 3 is activated (clock 2 deactivated) on the next clock pulse of the coarse counter to slowly discharge C_{ramp} until V_{ramp} is less than V_{in}. This interval is measured in Tlsbs as the fine count. The coarse count is the number of fixed Tlsbs before V_{ramp} crosses V_{in} the first time. The coarse and fine counts are decoded to produce a final 10b digital output.

Theoretically, the input can be divided into a large number of subranges. In

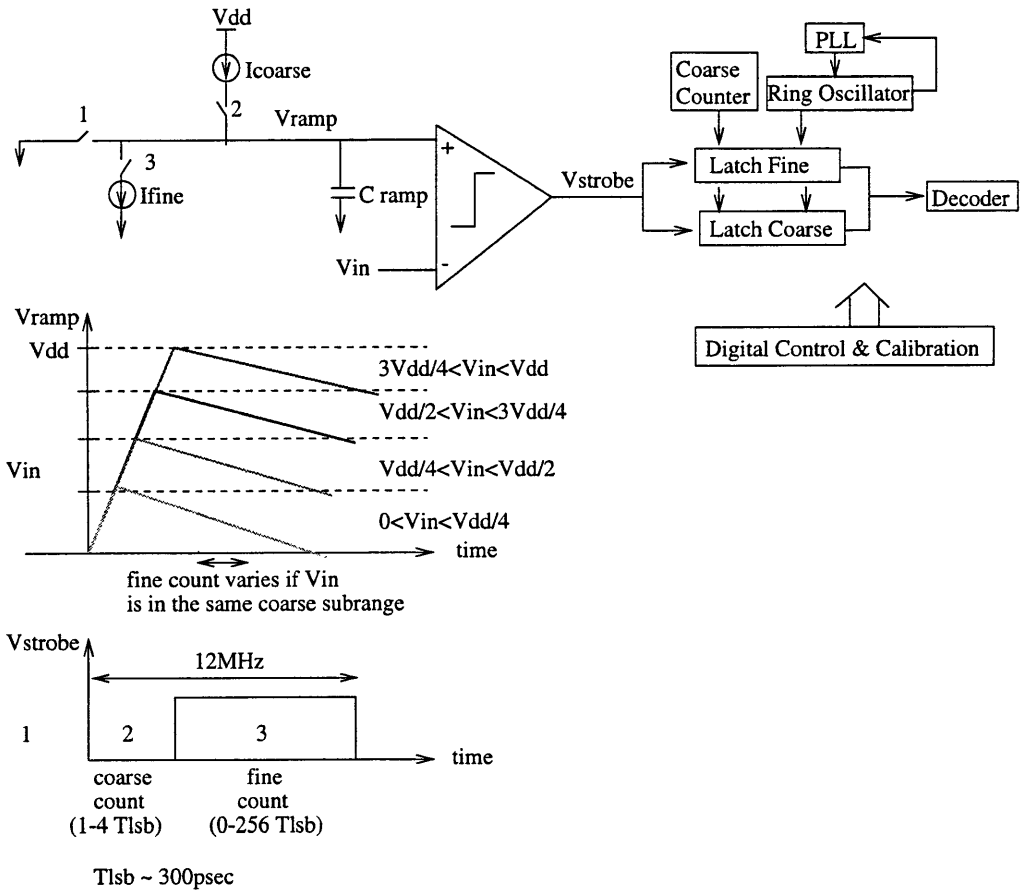


Figure 6-3: Trislope Converter

this case, only 4 are needed to increase the conversion rate to 12MHz. This gives the fine conversion a resolution of 256 or 8 bits, and the ring oscillator frequency has to increase only slightly to account for the delay in making the initial coarse conversion. The four coarse time intervals can be ring oscillator $t_{LSB's}$ or the period of an external clock. If they are ring oscillator $t_{LSB's}$, an extra set of fully differential latches is needed to latch the state of the ring oscillator during the coarse counting interval. Additional decoding circuitry is also needed, but the area and power cost of these digital circuits should be negligible.

One design challenge is matching the ratio between the coarse and fine currents to 10b accuracy. If they are not matched, subrange endpoints will not be aligned. Another concern with this design is charge injection on the ramping node when the coarse and fine currents are switched. Additionally, startup transients associated with these charging and discharging currents may cause nonlinearities in the A/D transfer curve.

The additional power consumption of this design is dominated by the fully differential latches. The estimated power consumption for this design is 37mW. The additional area of the extra current source and latch banks gives an estimated 20% increase in area, giving a total A/D area of $0.37mm^2$.

6.2.2 Integrated 2b Flash

A similar approach is shown in Figure 6-4. Three comparators and latch banks are used to perform a coarse 2b flash conversion. The most significant bits are determined by analyzing which comparators flipped. The least significant bits are determined by the count that was latched when the comparators flipped. For instance, if the input is between Vr2 and Vr3, C1 and C2 will flip immediately. C3 will flip at some count, and C4 will not flip at all. I_{fine} is calibrated so that there are 256 states in each range.

The problem with this approach is calibrating so that there are no DNL spikes at the endpoints. Mismatches in comparator offset voltages appear as an effective reference offset in Vr1-Vr4. Thus some subranges may contain more than 256 counts.

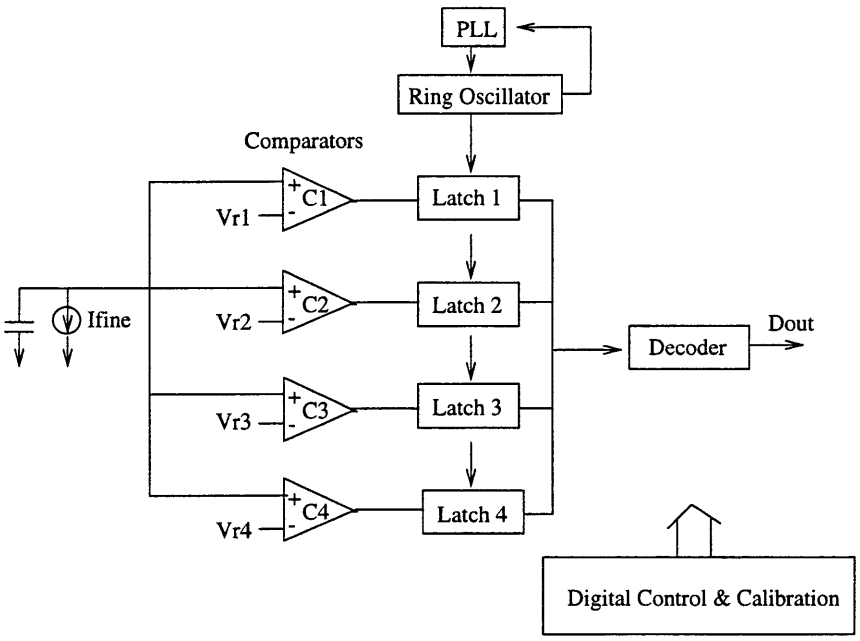


Figure 6-4: Coarse 2b Flash

The endpoints of these subranges must be properly aligned to avoid DNL errors at the endpoints.

This approach requires 3 extra comparators and 3 sets of extra latch banks. Estimating the additional power dissipation based on the comparator and latches already designed gave a power dissipation of 50mW. The additional area consumption is estimated to be an additional 13%, giving a total area of $0.34mm^2$.

Chapter 7

Simulation Results and Conclusions

Power results for the A/D are shown in Table 7.1, The track & hold power dissipation is large so that the op amps can settle the 30pF load capacitor in 333nsec. The load capacitor has to be large to achieve 10b matching between the two track & hold paths. Looking back on this design decision, power may have been saved if a single sample & hold was used. The settling time of the sample & hold would have reduced the time available for conversion, resulting in a higher ring oscillator frequency. This additional power dissipated in the ring oscillator may have resulted in less total power than the interleaved track & hold.

Another area for power improvement is in the fully-differential latches. The latches need to have a large enough bandwidth to trigger the coarse counter when in transparent mode. These latches were also used as master slave latches for the other ring oscillator outputs. The latch design was kept the same for both paths in Figure 5-2 to reduce mismatches in resolving the ring oscillator state. Perhaps the latches that did not have to be transparent could have been re-designed to save power. Another idea is to make the ring oscillator and latches all single ended. However, this would increase the sensitivity of these circuits to substrate bounce and noise in the power supply.

An improvement in the area consumption might be to make the ramp current less

Table 7.1: Power Consumption

Track & Hold	8mW	28%
FD Latches	5mW	18%
FD2SE	3.6mW	13%
Bias	3mW	11%
Ring Oscillator & PLL	2.5mW	9%
Buffers	2.6mW	9%
Comparator	2mW	7%
Divider & Coarse Counter	1.5mW	5%
Ramp	0.2mW	0.7%
Calibration	0.1mW	0.4%
Total Power	28.5mW	100%

sensitive to changes in the ramp control voltage. This would allow for a smaller capacitor at the ramp control voltage node, thereby decreasing the area of the calibration capacitor, or increasing the size of the calibration (encp) pulse.

Sources of A/D Nonlinearity are summarized in Table 7. These numbers are explained further in Chapter 7 and are based on simulations and calculations of worst case mismatches. These errors either translate to DNL, INL, or a degradation in SNDR (ENOB) as indicated in Table 7.

Table 7.2: Summary of Sources of A/D Nonlinearity

Parameter	Error
Ring Oscillator Mismatches	0.1LSB DNL
Buffer Mismatches	0.1LSB DNL
Latch Mismatches	0.1LSB DNL
Ramp Nonlinearity	0.01LSB DNL
Ramp Nonlinearity	0.3LSB INL
Ramp Noise	$4 * 10^{-5}$LSB RMS (ENOB)
Ring Oscillator Jitter	0.1LSB RMS (ENOB)
Comparator Jitter	0.1LSB RMS (ENOB)

Future design challenges include extending this 3MHz design to a 12MHz trislope converter, and decreasing the power dissipated in the track & hold and fully-differential latches. Other interesting challenges include measuring the ramp non-

linearity due to the MOSFET capacitors, the noise due to jitter, and the mismatch (banging at endpoints on the A/D transfer curve) due to mismatches through the interleaved track & hold.

The estimated power and area numbers for the 12MHz design are summarized in Table 7. Previously reported A/D power and area numbers are also shown. These numbers were scaled for conversion rate, supply, and technology as described in Chapter 3.

Table 7.3: Comparison to Literature

Approach	Total Power	Total Area	Figure of Merit $(\frac{1000mW*mm^2}{power*area})$
Desired Converter (Standard process, 10 ENOB)	20 mW	1 mm^2	50
Serial Pipeline [9] (Note: 12b, 11 ENOB, double poly)	30.5 mW	1.4mm^2	24
Folding Interpolating [15] (Note: 8.7 ENOB)	26.6 mW	0.59 mm^2	64
Single Slope - Edge Interpolation	53 mW	0.45 mm^2	42
Single Slope - 2b Flash	50 mW	0.34 mm^2	59
Single Slope - Trislope	37 mW	0.37 mm^2	73

The trislope implementation shows promise for a larger power/area figure of merit than previously reported architectures. If the power dissipated in the track & hold and fully-differential latches is decreased, the trislope converter has the potential to be a low power, low area solution.

The serial approach to A/D converters for CMOS imagers was taken. This approach involves using one fast A/D to process all of the imager pixels. The serial approach optimizes $\frac{1}{power*area}$ figure of merit over the column parallel and semi column parallel approaches. The single slope architecture with subnanosecond time digitizer shows promise as a low power, low area alternative to pipelined and folding interpolating approaches. A 10b, 29mW, 3MHz single slope architecture with subnanosecond time digitizer was designed and built in 0.3mm^2 in a standard 0.35um, 3.3V CMOS technology. This is a 28% reduction in power and 90% reduction in

area over the current single chip imager A/D design [6]. This architecture also shows promise as a low power, low area solution for a four times CIF imager array (704x576) when extended to a 12MHz trislope converter. With additional effort in power minimization, this design has the potential to have a larger figure of merit than previously reported designs in the literature. With the addition of background calibration, this A/D is a low power, low area solution for applications other than imaging.

Bibliography

[1] B. Ackland and A. Dickinson, "Camera on a Chip," *IEEE International Solid State Circuits Conference*, 1996, pp. 22-25.

[2] B. Pain, E. Fossum, "Approaches and analysis for on-focal-plane analog-to-digital conversion," *Proceedings of the SPIE, Aerospace Sensing - Infrared Read out Electronics II*, Vol. 2226, 1994, pp.1-11.

[3] A. Blanksby, M. Loinaz, D. Inglis and B. Ackland, "Noise Performance of a Color CMOS Photogate Image Sensor," *International Electron Devices Meeting Technical Digest*, December 1997, pp. 205-208.

[4] E. R. Fossum, "Ultra Low Power Imaging Systems Using CMOS Image Sensor Technology." *Proceedings of the SPIE, Advanced Microdevices and Space Science Sensors*, Vol. 2267, 1994.

[5] S. Dacy, "Programmable Gain Amplifier." Report written summarizing work done at Lucent Technologies in Summer 1996. Submitted as an MIT AUP.

[6] M. Loinaz, K.J. Singh, A. Blanksby, D. Inglis, K. Azadet and B. Acklandet, "A 200mW 3.3V CMOS Color Camera IC Producing 352x288 24-b Video at 30 frames/s," *International Solid State Circuits Conference Digest of Technical Papers,* February 1998, pp. 168-169.

[7] B. Razavi, *Principles of Data Conversion System Design.* New York: IEEE Press, 1995.

[8] A. Netravali, and B. Haskell, *Digital Pictures,* New York and London: Plenum Press, 1988.

[9] P. C. Yu, *Low-Power Design Techniques for Pipelined Analog-to-Digital Conve rters.* PhD Thesis, Department of Electrical Engineering and Computer Science, Massachusetts In stitute of Technology, May 1996.

[10] M. Loinaz, personal communication.

[11] B. Olson, T. Shaw, B. Pain, R. Paniccaci, B. Mansoorian, R. Nixon and E. Fossum, "A Single Chip CMOS APS Digital Camera," *Proceedings of 1997 IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, Burges, Belgium, June 1997, pp. R24-1-R24-4.

[12] S. Mendis, *CMOS Active Image Sensors with On-Chip Analog to Digital Conversion.* PhD thesis, Columbia University, 1995.

[13] S. Decker, *A Wide Dynamic Range CMOS Imager with Parallel On-Chip Analog to Digital Conversion,* PhD Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, February 1998.

[14] H. Yoshizawa, *High-Linearity Switched-Capacitor Circuits in Digital CMOS Technologies.* PhD Thesis, Department of Electrical and Computer Engineering, Oregon State University, May 1997.

[15] K. Bult, A. Buchwald, and J. Laskowski, "A 170mW 10b 50MSample/s CMOS ADC in $1mm^2$," *IEEE International Solid-State Circuits Conference*, pp. 136-137, 1997.

[16] T. A. Knotts, D. Chu, and J. Sommer, "A 500MHz Time Digitizer IC with 15.62ps Resolution," *IEEE International Solid-State Circuits Conference*, 1994, pp.58-59.

[17] J. Maneatis and M. Horowitz, "Precise Delay Generation Using Coupled Oscillators," *IEEE Journal of Solid-State Circuits*, pp. 1273-1282, December 1993.

[18] M. Loinaz, *Mixed-Signal VLSI Circuits for Particle Detector Instrumentation in High-Energy Physics Experiments.* PhD Thesis, Department of Electrical Engineering, Stanford University, July 1995.

[19] P. Larsson, "High-Speed Architecture for a Programmable Frequency Divider and a Dual-Modulus Prescaler," *IEEE Journal of Solid-State Circuits*, pp. 744-748, May 1996.

[20] P. Larsson, J. Lee, "A 400mW 50-380MHz CMOS Programmable Clock Recovery Circuit," *8th annual IEEE Int'l ASIC Conf. and Exhib.* pp. 271-274, 1995.

[21] C. W. Mangelsdorf, "A 400-MHz Input Flash Converter with Error Correction," *IEEE Journal of Solid-State Circuits*, pp.184-191, February 1990.

[22] P. Gray and R. Meyer, *Analysis and Design of Analog Integrated Circuits*, New York:John Wiley & Sons, Inc., 1977.