

High-Speed Data Conversion for Digital Ultra-Wideband Radio Receivers

by

Puneet Prashant Newaskar

Submitted to the Department of Electrical Engineering
and Computer Science

in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2003

© Massachusetts Institute of Technology 2003. All rights reserved.

Author

Department of Electrical Engineering
and Computer Science
May 23, 2003

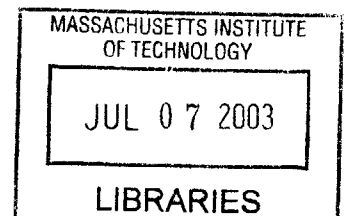
Certified by

Anantha P. Chandrakasan
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by

Arthur C. Smith
Chairman, Department Committee on Graduate Theses

BARKER



High-Speed Data Conversion for Digital Ultra-Wideband Radio Receivers

by

Puneet Prashant Newaskar

Submitted to the Department of Electrical Engineering
and Computer Science

on May 23, 2003, in partial fulfillment of the
requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

Abstract

Ultra wideband radio (UWB) is a new high-speed wireless technology that uses sub-nanosecond pulses to transmit information. Implementing an all-digital UWB receiver has numerous benefits ranging from low cost and ease of design to flexibility. Digitizing an RF signal near the antenna, however, introduces its own set of challenges and has traditionally been considered infeasible. A high-speed, high-resolution analog-digital converter (ADC) is difficult to design, and is extremely power-hungry. However, due to the unique characteristics of UWB signals and their noise environment, it can be shown that reliable detection is achievable with very few bits of resolution. In this thesis, the role of quantization noise in UWB systems is analyzed and the sufficiency of 4 bits of precision is demonstrated. Data conversion at several gigasamples/sec (GSPS) is challenging nevertheless, even at low resolutions. The feasibility of this problem has been investigated through the design and implementation of a 4-bit, 4 GSPS ADC for a prototype UWB system, using a 0.18 μm CMOS process. A time-interleaved architecture was chosen, with 4 FLASH channels each running at 1 GHz using offset clocks. The desired resolution was achieved through proper sizing of devices in the preamplifiers and comparators, and numerous circuit techniques were employed to keep dynamic offsets small. Pipelining was used extensively to support the high throughput. This design was laid out and fabricated. The resulting chip was shown to have 3.9 bits of relative accuracy and 3.3 bits of absolute accuracy at a sampling rate of 1.54 GSPS. Its total power consumption at this rate is 241 mW. Testing at the full designed speed of 4 GSPS was not possible due to problems with the on-chip test interface but the suspected cause has been identified. This work demonstrates that the problem of high-speed data conversion for digital UWB receivers is tractable in CMOS within a reasonable power budget.

Thesis Supervisor: Anantha P. Chandrakasan

Title: Professor of Electrical Engineering and Computer Science

Acknowledgments

I would like to thank Professor Anantha Chandrakasan, my research supervisor, for his guidance and encouragement throughout this project. Considering my only knowledge of analog-to-digital converters prior to this research was through a guest lecture in 6.775, the fact that he entrusted me with such an ambitious design has meant a lot. That faith, coupled with his unrelenting push to make things happen, kept this engine chugging along! In the process, I have learnt so much, not just about how to design fast circuits, but also how to effectively manage, inspire and lead.

I would also like to express my gratitude to Fred Lee and Raul Blazquez. The three of us have been like a band of brothers this past year, spending endless late nights in the lab all for the sake of a chunk of silicon! Raul was a co-author on a paper that we wrote on ADC precision requirements on which Chapter 2 of this thesis is based. Fred and I worked together on the test-chip described in Chapter 4 and many of his circuits are an integral part of the system (the clocks for my A/D for instance!) Technical collaboration aside, I believe that without their camaraderie and support, this project would not have been successful. I have now truly understood the importance of teamwork in engineering. I have also learnt a great deal from these guys. I have been struck by Fred's intuitive approach to thinking about analog circuits, his creativity and perseverance. Likewise, Raul's rigorous approach to problems and the elegance of his solutions have left quite an impression.

The rest of my labmates have been just as terrific. Dave would drop whatever he was doing if I had a question. His advice throughout the design and layout process was invaluable. Alice was very helpful with all my questions about Cadence. Ben and Rex would bring over food when Fred, Raul and I had been up two nights in a row working on our chip. All in all, this lab has been a wonderful place to work. Talking to Johnna about life in Austin, or to Frank about running marathons, playing softball indoors, it's all been part of a truly memorable experience.

And, of course, there's my family and friends. Mamma, Pappa, my kid sister Sonal, Satish mama, Shubha mami, Aji, my friends Sentheel, Zubin, Visvesh, Nehal,

Abe, Viji and my housemates Nigel, Vidya and Gaia. It's hard to put my debt of gratitude towards all of them into words. Whether it was listening to my concerns about whether this chip would work, or listening to my excitement about a new idea or some positive test results. Research has its ups and downs, and in both situations they have been great.

Last but not least, I would like to acknowledge the sources of funding that made my graduate studies possible, namely MIT's Presidential Fellowship and an assitanship from Hewlett-Packard Inc.

Contents

1	Introduction	13
1.1	Impulse Radio	13
1.2	Receiver Design	16
1.2.1	UWB Signal Structure	16
1.2.2	Receiver Architectures	17
1.3	ADC Challenges	20
2	ADC Precision Requirements	21
2.1	Traditional Approach	22
2.2	A New Framework	24
2.3	AWGN-Limited Case	25
2.3.1	No Quantization	26
2.3.2	Quantization Effects Included	26
2.4	Interference-Limited Case	28
2.5	Summary of Analysis	30
2.6	Simulations	30
2.6.1	AWGN-Limited Case	30
2.6.2	Interference-Limited Case	31
2.7	Conclusion	38
3	ADC Design	39
3.1	Specifications	39
3.2	Architecture	40

3.3	FLASH Design Issues	43
3.4	Analog Section	47
3.4.1	Track and Hold	48
3.4.2	PCC Block	52
3.4.3	Resistor Ladder	64
3.5	Digital Section	66
3.5.1	Thermometer to Binary Conversion	66
3.5.2	Retiming	71
3.6	External Interface	72
3.6.1	Clocks	73
3.6.2	Biasing	75
3.6.3	Input Interface	75
3.6.4	Test Interface	76
3.7	Top-Level Simulations	78
4	Chip Implementation	81
4.1	Analog Section	84
4.2	Digital Section	88
4.3	Complete ADC Layout	88
4.4	Top-Level Simulations	92
4.5	Chip-Level Integration	93
5	Testing and Analysis	95
5.1	ADC Test Methodology	95
5.2	Chip Measurements	99
5.3	Proposed Improvements	106
5.3.1	Ensuring Full Testability	106
5.3.2	Improving Dynamic Performance	107
6	Conclusion	109

List of Figures

1-1	Duality Between Sine Waves and Pulses	14
1-2	Demodulating a UWB Signal	18
1-3	UWB Receiver Architectures	19
2-1	Loss in SNR Due To Quantization	27
2-2	Probability of Error vs SIR (Theoretical)	29
2-3	Probability of Error vs SNR (Simulated)	31
2-4	Probability of Error vs SIR (Simulated)	32
2-5	Probability of Error vs SIR (Simulated)	33
2-6	2-bit Quantization Effects	34
2-7	Signal + Noise (post-correlation)	36
2-8	Simulation Results (OFDM interferer)	37
3-1	Time-interleaved FLASH ADC	42
3-2	Thermometer Code	42
3-3	FLASH Analog Section	48
3-4	Track and Hold	50
3-5	Preamplifier	53
3-6	Preamplifier Transfer Curves	53
3-7	Comparator Topologies	58
3-8	PCC Block	61
3-9	PCC Simulations	62
3-10	ADC Transfer Curve	64
3-11	Resistor Ladder and Connections	65

3-12	Decoding Logic	69
3-13	XOR Gate	70
3-14	Flip-Flop	71
3-15	Retiming Block	73
3-16	Phase-Locked Loop	74
3-17	Input Interface	76
3-18	Test Interface	77
3-19	Slowramp Test	78
4-1	Common Centroid Layout	83
4-2	Interdigitated Common-Centroid Transistor Pair	84
4-3	Common-Centroid Transistor Quad in Preamplifier	86
4-4	Regenerative Pair of First Comparator	87
4-5	Cross-Coupled NMOS Pair of Second Comparator	87
4-6	PCC Bank	87
4-7	Layout of Decoding Logic	89
4-8	Clock Distribution Options	90
4-9	Layout of Complete ADC	91
4-10	Die Photograph	94
5-1	Ideal 4-bit ADC	97
5-2	Non-Ideal 4-bit ADC	97
5-3	Channel 1 Performance	100
5-4	Channel 2 Performance	100
5-5	Channel 3 Performance	100
5-6	Channel 4 Performance	101
5-7	Dynamic Performance	103

List of Tables

3.1	Preamplifier Performance	56
3.2	PCC Performance	63
5.1	Accuracy of Channels	101
5.2	ADC Power Consumption	106

Chapter 1

Introduction

1.1 Impulse Radio

Ultra wideband radio (UWB) is an exciting new wireless technology that promises high data rates over short distances by employing bandwidths in excess of 1 GHz. UWB uses a train of sub-nanosecond pulses to carry information[18]. If the duration of the pulses is much smaller than the interval between successive pulses, they may be idealized as Dirac-Delta impulses. Since this approximation typically holds true, UWB is also referred to as *impulse radio*.

An impulse is narrow in time but wide in frequency. A sinusoid, in contrast, is narrow in frequency but wide in time. The two signals are thus mathematical duals of one another as illustrated in Figure 1-1. There is an interesting historical twist to this duality. Marconi's radio transmitted information using Morse code, comprising a sequence of *dots* and *dashes* that are essentially short and long pulses. The early origins of radio were thus wideband. However, this form of signalling was soon superseded by carrier-based narrowband systems. Today, almost all forms of wireless communication are narrowband in nature. 300 GHz of available radio spectrum is divided up into thousands of narrow bands as shown that are then allocated to different services and standards. In the United States, there is a government agency called the Federal Communications Commission (FCC) that is entrusted with managing spectrum. This is a herculean task considering the sheer number of radios that are

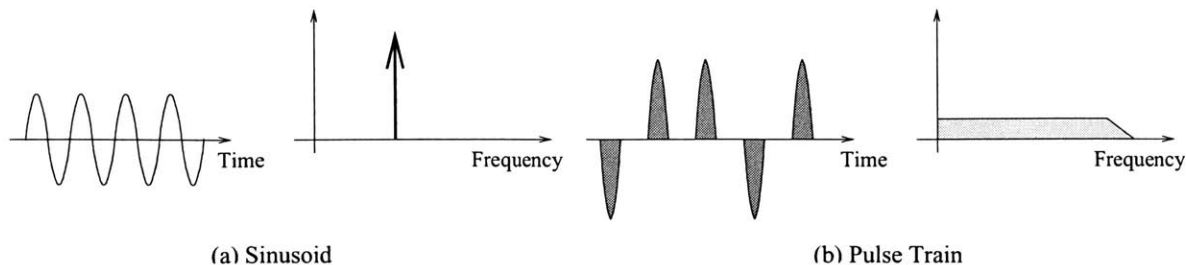


Figure 1-1: Duality Between Sine Waves and Pulses

concurrently operated. Applications include commercial ones like the cellular network, as well as government and military needs like weather satellites, radar and the geo-positioning system (GPS). In light of the current pervasiveness and dominance of narrowband radio, the *re-emergence* of ultrawideband is an intriguing phenomenon. Not surprisingly, the FCC's approval of this potentially groundbreaking technology came after years of rancorous debate and opposition from cellphone companies, their network operators, certain branches of government, the GPS community and other groups. Their primary concern was the interference problem posed by UWB to their services. In other words, is it possible for a UWB signal occupying several GHz of already-occupied bandwidth to co-exist with these narrowband services ?

Consider a single rectangular pulse of 1 nanosecond duration. Its Fourier transform is a sinc, with a main lobe bandwidth of 1 GHz. The power of the pulse is spread over a wide swath of frequencies, and the resulting power spectral density in any 1 MHz slice of bandwidth within the main lobe is roughly 1/1000th of the peak power. An impulse radio system, however, uses a periodic train of such pulses with data modulating the polarity or position of each pulse. Any inherent periodicity in the data stream produces lines in the spectrum, each of which carries considerable power and causes interference. A commonly proposed solution to this problem is the use of pseudo-random noise (PN) sequences to encode the pulse train. This form of coding sufficiently randomizes the transmitted signal and removes the above-mentioned spectral lines. Aside from whitening the signal spectrum, coding provides yet another benefit. If a code of length N_c is used, then N_c pulses may be used to represent a

single bit. This redundancy, referred to as *pulse integration*[18] can be used to reduce the peak power of the pulses while maintaining a desired bit error rate (BER). Finally, by setting the pulse-to-pulse interval to be large relative to the pulse duration (ie. low duty cycle), average power can be reduced further. Therefore, by adjusting three knobs : pulse duration, duty cycle and PN code length, we can ensure that the power spectral density of a UWB signal in any 1MHz slice of bandwidth is sufficiently small. Claims have been made that a UWB signal can be designed to operate below the ambient noise floor and terms such as *imperceptible radio* [17] have been proposed to describe this property.

Having addressed the co-existence problem, let us now turn to the tangible benefits of UWB that make it a compelling wireless technology. Due to the interference problem mentioned above, a UWB system is power-constrained and operates at low signal-to-noise (SNR) ratios. However, it uses an extremely large signal bandwidth. A narrowband radio, on the other hand, uses much smaller bandwidth but operates at high SNR. From a capacity standpoint, a UWB system fares better and offers higher data rates. Shannon's classical equation[20] for capacity in a wireless link illustrates this capability :

$$Capacity = Bandwidth \cdot \log_2(1 + SNR) \quad (1.1)$$

Capacity is thus directly proportional to the bandwidth used, but is logarithmically related to the SNR. The theoretical upper bound on achievable data rate is thus much higher for a typical UWB radio than for a narrowband system for a given amount of transmit power. Conversely, UWB can support the same throughput as narrowband using much lower transmit power. Although actual data rates for both cases fall far short of Shannon's theoretical bounds, the conclusions above about UWB vis-a-vis narrowband hold true. Early prototypes of UWB radios have demonstrated this. For instance, XtremeSpectrum Inc. has developed a chipset[10] that supports 100 megabits/sec (Mbps) links over 10m, using 7GHz of spectrum and -3 dBm of transmit power. While the IEEE 802.11a standard for wireless LAN is designed to

provide 54 Mbps over a similar range, it uses 20dBm of transmit power, two orders of magnitude larger than XtremeSpectrum's UWB solution. Furthermore, through the use of more sophisticated modulation schemes, the throughput of UWB can be raised even higher and pushed closer to capacity. Thus, UWB is an attractive physical layer for high-bandwidth short-range wireless networking.

A UWB radio is inherently secure and has low probability of detection. This attribute stems from the low SNR of UWB signals in conjunction with the use of PN sequences for encoding them. Only a highly-sensitive receiver with knowledge of the code in question can detect and demodulate such a signal. Other key features of UWB are its multi-path immunity and precise locationing capability. These attributes are both a result of the large bandwidth used and make UWB a compelling choice for high resolution radar applications¹.

Thus, it is clear that UWB technology offers considerable benefits. Having established its potential, the next logical step is to address questions of implementation and feasibility. To this end, a more detailed description of UWB signal structure is presented, followed by an analysis of two competing receiver architectures.

1.2 Receiver Design

1.2.1 UWB Signal Structure

As described earlier, information in a UWB system is transmitted using a collection of narrow pulses (0.2 ns to 1.5 ns) with a very low duty cycle ($\sim 1\%$)[18]. Each user is assigned a different pseudo-noise (PN) sequence that is used to encode the pulses in either position (PPM)[13] or polarity (BPSK²)[12]. Channelization is thus based on the assigned code, as in the case of CDMA systems.

Suppose the bitstream is denoted by a sequence of binary symbols b_j (with values

¹The technology was, in fact, first developed for radar by the US military.

²The term BPSK (binary phase-shift keying) is somewhat of a misnomer in the context of a UWB signal since the notion of pulse phase is ambiguous. The intended meaning is an antipodal signalling scheme, but the term BPSK has been used for convenience.

+1 or -1) for $j = -\infty, \dots, \infty$. A single bit is represented using N_c pulses, where N_c refers to the length of the PN code c_i . For BPSK, the code modulates the polarity of a pulse within each frame. For PPM, it modulates the pulse positions (incrementing or decrementing them by multiples of T_c). Data modulation is achieved by setting the sign of the block of N_c pulses for BPSK. For PPM, we append an additional time-shift τ_{b_j} whose value depends on whether b_j is +1 or -1. Each frame has duration T_f ; the duration of each bit is thus given by $N_c T_f$. Letting A denote the amplitude of each pulse $p(t)$, the transmitted signal $s(t)$ can be written as follows for the two different modulation schemes :

$$s_{BPSK}(t) = A \sum_{j=-\infty}^{\infty} \sum_{i=0}^{N_c-1} b_j c_i p(t - jN_c T_f - iT_f) \quad (1.2)$$

$$s_{PPM}(t) = A \sum_{j=-\infty}^{\infty} \sum_{i=0}^{N_c-1} p(t - jN_c T_f - iT_f - c_i T_c - \tau_{b_j}) \quad (1.3)$$

As mentioned earlier, the code length and duty cycle are knobs that determine the system's throughput, bit error rate and the average power in the transmitted signal. This complex relationship can be reduced to a simpler form using a metric called processing gain (PG). This quantity refers to the boost in SNR as a UWB signal is processed by a correlating receiver.

$$PG, dB = 10 \log(N_c) + 10 \log\left(\frac{1}{duty_cycle}\right) \quad (1.4)$$

Both the code length N_c and the duty cycle appear in the above expression. Their relevance, and the importance of processing gain in the system will be made clear in subsequent discussions.

1.2.2 Receiver Architectures

Optimal detection of a noisy signal is based on matched filtering[20]. This entails correlating the received signal $r(t)$ against a template $l(t)$ that is an exact replica of the original transmitted signal and then feeding the correlator output to a slicer.

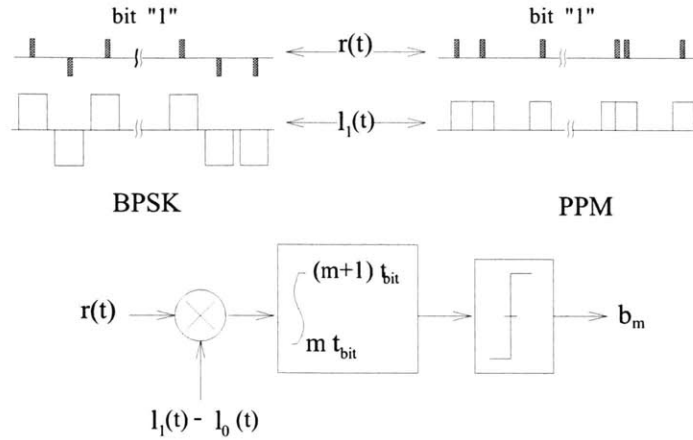
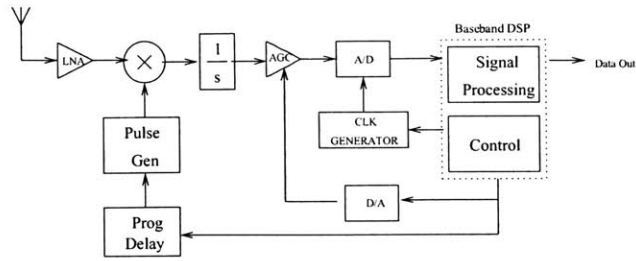


Figure 1-2: Demodulating a UWB Signal

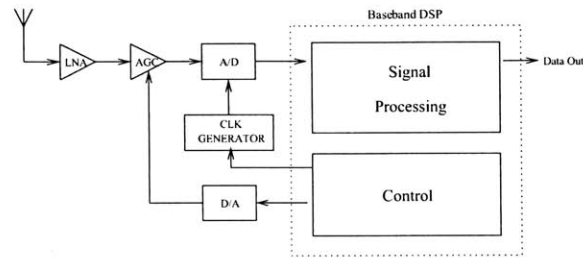
The latter compares the correlator output against a set of thresholds and yields the demodulated data. Since the various sources of noise that the signal is immersed in are *uncorrelated* with its underlying code, this process raises the effective signal-to-noise ratio and allows the desired signal to be successfully extracted.

However, generating exact replicas of sub-nanosecond pulses is a hard problem, and is highly susceptible to timing jitter. At the expense of some loss in performance, an approximate template signal may be used. For instance, the template signal could comprise a train of rectangular pulses that are, in general, wider than the actual received pulses but are coded with the same PN sequence. Fig. 1-2 illustrates the structure of the received and template signals, and depicts the operations necessary for demodulation. $l_0(t)$ and $l_1(t)$ represent the template signals corresponding to a transmitted 0 and 1 respectively. They are related to one another by either a sign-inversion or a time-shift, depending on whether BPSK or PPM was employed. Equations describing $l_1(t)$ and $l_0(t)$ can be obtained from (1.2) and (1.3) by simply replacing the original pulse $p(t)$ with a rectangular pulse $\text{rect}(t)$, and setting b_j to +1 and -1 respectively. By using such template signals, we are essentially performing a form of windowing. In other words, we are correlating the received signal against its underlying PN code over narrow windows.

In practice, the wideband signal may undergo considerable distortion in the chan-



RECEIVER with ANALOG CORRELATION



RECEIVER with DIGITAL CORRELATION

Figure 1-3: UWB Receiver Architectures

nel. The receiver design must then be modified to take these distortions into account by performing some equalization. If the channel is time-varying, this would further complicate the task. In light of this, let us consider the classical analog versus digital debate in the context of UWB receivers. Two possible architectures are illustrated in Fig. 1-3. One performs correlation in the analog domain and then digitizes the result, while the other samples the signal after sufficient amplification and performs correlation digitally. While the former approach relaxes constraints on the analog-to-digital converter (ADC) and backend signal processing, it represents an inflexible solution to the equalization problem. Estimating the channel and then controlling the delays and shapes of the template pulses generated in the analog domain is non-trivial. A UWB receiver based on digital correlation can more easily incorporate such functionality. Furthermore, it can provide additional flexibility by supporting multiple modulation schemes and bit rates, thus approaching the paradigm of fully-configurable software radios. The considerable benefits of technology scaling also lend weight to the digital solution. The caveat, however, is the design complexity of the ADC and the backend

signal processing given the sampling rates and throughput that must be supported. Not surprisingly, most of the early UWB prototypes that have been developed commercially employ analog correlation[10][5], presumably due to considerations like time to market. With continued technology scaling, however, it is likely that more digital solutions will emerge. The potential benefits and associated challenges make digital UWB radios a prime area for research.

1.3 ADC Challenges

In accordance with the Nyquist theorem, the ADC sampling rate for digitizing a UWB signal must be on the order of a few gigasamples/sec (GSPS). Even with the most modern process technologies, this constitutes a serious challenge. Most reported data converters operating at this speed employ interleaving[8], with each channel typically based on a FLASH converter. The latter is the architecture of choice for high-speed designs, but is not suitable for high-resolution applications[23]. An N-bit FLASH converter uses 2^N comparators so its power and area scale exponentially with resolution. Among recently reported designs (>1 GSPS) representing the state of the art[8],[19] in CMOS ADC's, none has a resolution exceeding 8 bits.

The minimum number of bits needed for reliable detection of a UWB signal is, therefore, a critical parameter. If excessively large, it can render an all-digital receiver infeasible. Fortunately, it can be demonstrated that reliable detection of a UWB signal can be performed with very few bits of resolution in the ADC. In fact, this work reveals that 4 bits is sufficient. A theoretical framework for explaining this is presented in the following chapter, together with simulations that reinforce the result.

Chapter 2

ADC Precision Requirements

The resolution required of the high-speed ADC in a digital UWB receiver is a critical parameter. If it is too large, it can render an architecture based on direct digitization infeasible. In this chapter, an analysis is presented to assess the impact of quantization noise on receiver performance, based on which a specification for ADC precision is derived[16].

Any form of distortion that is uncorrelated with the UWB signal's underlying PN code contributes to "noise" at the output of the correlator shown in Fig. 1-2. There are numerous sources of such noise : channel and circuit noise, quantization, timing jitter, in-band narrowband interferers and other UWB signals occupying the same swath of spectrum. The latter has not been considered since multi-user UWB interference is beyond the current scope of this work. Nevertheless, the validity of our result (ie. sufficiency of 4 bits) is not undermined by this omission. This is because UWB signals using different codes are largely uncorrelated and thus, the presence of multiple users only gradually raises the noise floor. Our result stems from the relative unimportance of quantization noise compared with the amount of channel noise that a UWB signal is immersed in. Raising of the ambient noise floor due to multiple UWB users does not change our analysis, therefore.

In this work, the following three categories of noise are considered : Additive white gaussian noise (AWGN), narrowband interference and quantization noise. Note that several disparate sources of noise have been lumped into AWGN (thermal noise in

circuits, noise in the channel and the effects of timing jitter in transceivers). Before presenting the analysis, classical approaches to setting the ADC resolution in a narrowband radio receiver are discussed. A framework appropriate for a UWB system is then developed.

2.1 Traditional Approach

The minimum ADC resolution required for a narrowband receiver is typically obtained [25] by either one of two methods. The applicability of each depends on the point in the signal path where the A/D is placed.

(1) In a classical super-heterodyne receiver, there are multiple stages of gain and filtering. Both out-of-band and out-of-channel noise and interference are significantly attenuated before the signal reaches the A/D. At this point, the dominant source of noise is the quantization noise added by the data converter. The A/D is followed by the demodulator and the latter requires a certain minimum SNR for a desired bit error rate (BER). The number of bits is thus chosen so that the overall SNR is above this minimum value.

(2) In several modern architectures (like low-IF and the zero-IF homodyne receiver)[24], the A/D is placed closer to the antenna. Alongside the signal, therefore, there are powerful blocking interferers that get digitized. Typically, these blockers are out-of-band and are easily filtered out in the digital domain. However, the nonlinearity of the A/D could generate spurs that fall in the same band as the desired signal. The spurious-free dynamic range (SFDR) quantifies this non-linearity; it is defined as the ratio of the sinusoidal signal power to the peak-power of the largest spurious signal in the ADC output spectrum. For reliable detection of the desired signal, the power of the largest spur must be sufficiently below signal power. This, in turn, constrains the ADC resolution to be above a certain minimum.

Note that in both the cases outlined above, quantization noise is significant. The high SNR or SFDR needed for reliable detection imply a high A/D resolution.

Unlike most narrowband radios, a UWB system is power-limited instead of bandwidth-

limited. The FCC allows UWB signals to occupy several gigahertz of bandwidth while placing stringent restrictions on transmit power. The large bandwidth is utilized to enable reliable detection at low power-levels and low received SNR's. Signal bandwidth in a typical UWB system exceeds target data rates by 1 or 2 orders of magnitude and this translates to large inherent processing gain. The latter is, in fact, equal to the ratio of signal bandwidth to data bandwidth (ie. symbol rate) and this definition is consistent with the formulation in Eq. 1.4; a long PN code and a low duty-cycle for the pulses widen the spectrum from the symbol rate up to the full signal bandwidth. By then performing correlation of the received signal over narrow windows as described in the previous section, the effective SNR of the UWB signal can be boosted by an amount equal to the processing gain.

$$SNR_{post-corr} = SNR_{pre-corr} + PG \quad (2.1)$$

The received SNR and the SNR after the A/D can thus be considerably lower than in narrowband systems. In fact, with the transmit power levels mandated by the FCC, a UWB signal is immersed in noise and interference. To make this point clearer, some numbers are presented below.

According to the spectral mask stipulated in the FCC First Report and Order [4], the power spectral density of a UWB signal at a distance of 3m from the transmitter must be no more than -41.3dBm/MHz. For a 2 GHz signal, this corresponds to an average power of -8.3dBm. The following equation [26] is used to estimate the propagation loss of the channel :

$$Loss(dB) = 20 \log \frac{4\pi d}{\lambda} + 0.7(d - 4)(d > 4) \quad (2.2)$$

At a distance of 10m, this corresponds to a loss of around 57 dB (taking the center frequency of the UWB signal to be 1 GHz). This brings the received signal power down to -65 dBm. Interference from cellphones or wireless LAN's can easily exceed this number by 30dB or more.

Consider thermal noise in the channel and in the receiver. With a bandwidth of

2 GHz and a receiver noise figure of 10dB, this amounts to thermal noise power of -74dBm. There are, of course, a variety of other sources of white gaussian noise that would raise this number higher.

In conclusion, a typical UWB signal is immersed in noise and interference. Therefore, the assumption that quantization noise is dominant cannot be applied in this context. A different line of analysis must be pursued to arrive at a specification for the A/D.

2.2 A New Framework

Upon sampling, the received signal is given by the following equation :

$$v[n] = s[n] + w[n] \quad (2.3)$$

where $s[n]$ is the desired signal. For BPSK¹, it has the following expression :

$$s[n] = A \sum_{j=-\infty}^{\infty} \sum_{i=0}^{N_c-1} b_j c_i p[n - jN_c N_f - iN_f] \quad (2.4)$$

N_f is equal to $\lfloor \frac{T_f}{T_s} \rfloor$ where T_f and T_s denote the frame duration and sampling interval respectively. $p[n]$ is taken to be a rectangular pulse of width W :

$$p[n] = \begin{cases} 1 & \text{if } 0 \leq n < W \\ 0 & \text{otherwise} \end{cases} \quad (2.5)$$

In reality, such a pulse cannot be generated and represents an abstraction. More practical shapes include a gaussian, monocycle[18] or more generally, a wavelet. However, there is no loss of generality in assuming a rectangular pulse. As will be demonstrated, the basis of our analysis is that the various noise sources identified are uncor-

¹For PPM, a similar analysis can be carried out. The only difference from BPSK is the appearance of a factor that divides the signal-to-noise ratio or signal-to-interference ratio in the expression for overall probability of error. This factor is 2 if the signals representing bits 1 and 0 are orthogonal.

related with the signal. This is true irrespective of the pulse shape. Our claim that 4 bits are sufficient can thus be extrapolated to other pulse shapes.

In expression (2.3), $w[n]$ represents any form of distortion prior to the ADC. Following quantization by the ADC, an additional noise component $q[n]$ is added :

$$r[n] = s[n] + w[n] + q[n] \quad (2.6)$$

Correlation is then performed against a template signal $l[n]$ (same form as $s[n]$ but with wider rectangular pulses of width W_o). The signal at the output of the correlator $y[m]$ is given by :

$$y[m] = \sum_{n=mN_cN_f}^{(m+1)N_cN_f-1} r[n] \cdot l[n] = s'[m] + w'[m] + q'[m] \quad (2.7)$$

$s'[m]$, $w'[m]$ and $q'[m]$ represent the contributions of $s[n]$, $w[n]$ and $q[n]$ respectively to the final correlation value. $s'[m]$ can be viewed as the sum of samples from N_c pulses (each with amplitude A or $-A$, depending on the value of the information bit, and width W). Correlating against the right code ensures that these samples are added with the right signs. Thus,

$$s'[m] = \begin{cases} -AN_cW & \text{if } b_m = 0 \\ +AN_cW & \text{if } b_m = 1 \end{cases} \quad (2.8)$$

$w'[m]$ and $q'[m]$ are both random variables, whose statistical properties and effect on post-correlation SNR is discussed below. In doing so, the AWGN-limited and interference-limited cases are treated separately.

2.3 AWGN-Limited Case

The analysis begins by ignoring the quantization noise $q'[m]$ and considering only the effect of $w'[m]$. This is equivalent to using an infinite-resolution ADC.

2.3.1 No Quantization

During correlation, $N_c \cdot W_o$ uncorrelated samples of noise are added. Since each individual sample is a gaussian random variable with zero mean and variance σ^2 , their sum $w' [m]$ is also gaussian, and has zero mean with variance equal to $\sigma^2 N_c W_o$. So $y [m] = s' [m] + w' [m]$ is gaussian with mean AN_cW and variance $\sigma^2 N_c W_o$. Thus the probability of error can be expressed as :

$$P_e = Q \left(\frac{AN_cW}{\sqrt{\sigma^2 N_c W_o}} \right) = Q \left(\sqrt{SNR \cdot N_c N_f \frac{W}{W_o}} \right) \quad (2.9)$$

Measured at the input of the receiver, SNR is equal to $\frac{A^2 d}{\sigma^2}$, where d is the duty cycle (given by the fraction W/N_f). Note that having a window larger than the pulse width thus results in a loss factor $L = \frac{W_o}{W}$. Capturing the full processing gain, therefore, requires setting $W_o = W$.

2.3.2 Quantization Effects Included

The ADC in our model has a fixed input range, from -1 to 1 . So if the number of bits is b , the quantization step is given by $\Delta = \frac{1}{2^{b-1}}$. Due to the presence of gaussian noise at the input, the quantization noise can be reliably modeled as a uniform random variable of variance $\Delta^2/12$ [23]. The implicit assumption here is that the ratio of the AWGN standard deviation to the quantization step falls within a certain range. If this ratio is too large, the probability of ADC saturation/clipping is high and the quantization noise added has variance well above $\Delta^2/12$. If the ratio tends to zero, on the other hand, the quantization noise power tends to $\Delta^2/4$. In order to avoid either of these sub-optimal regimes, an automatic-gain-control (AGC) circuit is placed before the A/D converter. The AGC scales its noisy input signal by a factor α such that the A/D is fed an “optimal” input power of σ_o^2 . For resolutions of 2, 3 and 4 bits, optimal σ_o values are 0.2850, 0.2025 and 0.1425 respectively. Due to the AGC, we can safely assume henceforth that the quantization noise power added by the ADC for all input SNR’s is $\Delta^2/12$.

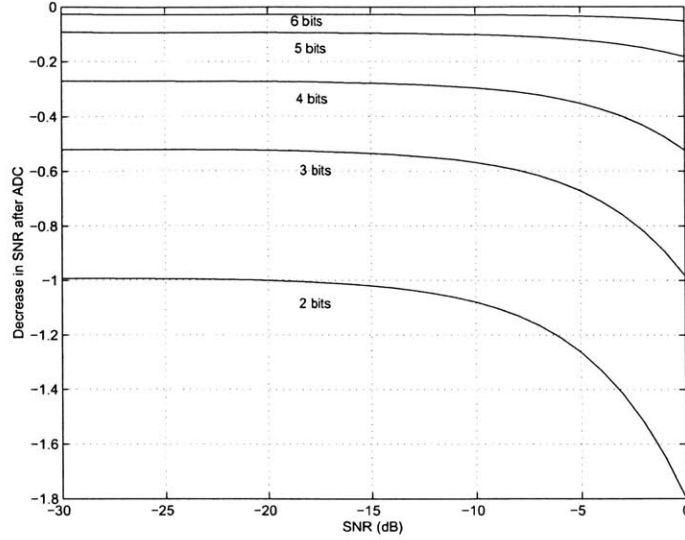


Figure 2-1: Loss in SNR Due To Quantization

Assuming the quantization noise is uncorrelated with the gaussian noise at the input, the SNR after A/D conversion can be formulated as follows :

$$SNR_{\text{after ADC}} = \frac{\frac{A^2 d}{\alpha^2}}{\frac{\sigma^2}{\alpha^2} + \frac{\Delta^2}{12}} \quad (2.10)$$

Correlation is then performed. One must account for the $N_c W_o$ samples of quantization noise that are added along with the signal and its input noise. These three components can be assumed to be uncorrelated. Since $N_c W_o$ is typically a large number, the central limit theorem implies that the distribution of the summed quantization noise samples approximates a gaussian with zero mean, and variance equal to $\sigma_{\text{quant}}^2 = \frac{\Delta^2 N_c W_o}{12}$.

The σ^2 in (2.9) is now replaced by $\sigma^2 + \sigma_{\text{quant}}^2$. Combining the resulting expression with (2.10) yields the revised approximation :

$$P_e = Q \left(\sqrt{SNR_{\text{after ADC}} N_c N_f \frac{W}{W_o}} \right) \quad (2.11)$$

For an amplitude A of 1 and a duty cycle of 2%, Fig 2-1 shows the decrease in SNR

after the ADC for various input SNR's and bit resolutions. This loss is marginal (less than 2dB^2) over the range of SNR's within which a UWB signal is expected to operate. Furthermore, increasing the number of bits provides diminishing improvement.

2.4 Interference-Limited Case

Again, let us start with no ADC and thus, zero quantization noise. First the impact that the correlation process has on a narrowband signal $w(t)$ at the input must be modeled.

$w(t) = B \cos(\Omega t + \phi_o)$. If this signal is sampled with sampling period T_s , it takes the following form :

$$w[n] = w(nT_s) = B \cos(\Omega T_s n + \phi_o) = B \cos(\omega_o n + \phi_o) \quad (2.12)$$

These interference samples are then processed by the correlator This step can be viewed as running $w[n]$ through a filter $F(z)$ that yields an output of the form:

$$I[n] = BF_o \cos(\omega_o n + \phi_o + \theta_o) \quad (2.13)$$

After downsampling by a factor $N_c N_f$, this sinusoid may be aliased to a different frequency ω'_o , but its amplitude does not change. This leaves the random variable:

$$y[m] = s'[m] + w'[m] = WN_c A + BF_o \cos(\Phi) \quad (2.14)$$

If the initial phase ϕ_o is a uniform random variable between 0 and 2π , then Φ is also a random variable between 0 and 2π , regardless of the distribution of the interferer's

²Without the AGC, the loss would be somewhat larger since the quantization noise power added by the ADC would be more than $\Delta^2/12$ as explained earlier.

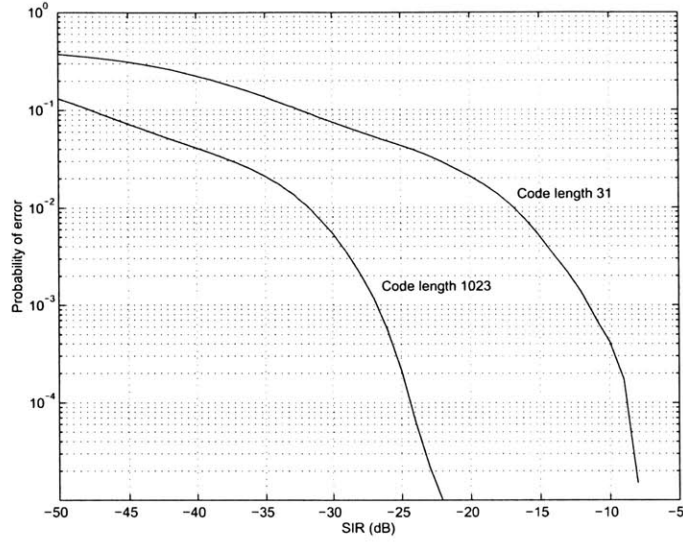


Figure 2-2: Probability of Error vs SIR (Theoretical)

frequency. The probability of error can be shown to be:

$$P_e = \begin{cases} 0 & \text{if } WN_cA > BF_o \\ \frac{1}{2} - \frac{1}{\pi} \sin^{-1} \frac{N_c}{F_o} \sqrt{\frac{SIR \cdot N_f W}{2}} & \text{if } WN_cA < BF_o \end{cases} \quad (2.15)$$

where SIR (signal-to-interference ratio) is equal to $\frac{2dA^2}{B^2}$. Thus, the behaviour depends on the length of the code. For comparison purposes, simulations were run using the above equation for Gold codes of length 31 and length 1023. Their results are observed in Fig. 2-2. The longer code yields a probability of error of 10^{-3} for an SIR that is 15dB lower. This difference is equal to the difference in processing gains.

Developing closed-form expressions for the cumulative effects of interference and quantization noise on the error probability is cumbersome. Instead, simulations are used to demonstrate the effect of increasing ADC resolution for the interference-limited case.

2.5 Summary of Analysis

Bit error rate (BER) in a UWB system is tied to the post-correlation SNR, which, in turn, is some function of signal power, AWGN noise, interference and quantization. High processing gain inherent in a UWB signal allows us to operate reliably (BER of 10^{-4} say) with a pre-correlation SNR that is considerably lower than for a typical narrowband system. This implies that the power levels of AWGN and interference at the input, corresponding to this target BER, are likely to be high. Our analysis confirms this. The models developed demonstrate the performance achievable in the absence of any quantization noise. It can be expected that introducing the latter should not cause a significant departure from the performance curves obtained without it. Simulations have been used to validate this hypothesis and to determine the minimum number of bits needed to get close enough to these “ideal” performance curves.

2.6 Simulations

2.6.1 AWGN-Limited Case

Again, the AWGN-limited and interference limited cases are treated separately. The representation of the UWB signal is the same in each, however. Rectangular pulses of width 2 samples are used, with a pulse-to-pulse interval of 100 samples (ie. duty cycle of 2 %). The PN code used is a Gold code of length 31 bits. Correlation is performed using a window of 10 samples per pulse.

Noise samples are uncorrelated. The ADC is preceded by an AGC which sets the power fed to the converter based on the “policy” described in the previous section. For each value of SNR simulated, 10^6 independent trials were carried out, based on which a probability-of-error P_e was assigned. Montecarlo simulations were carried out that provided a standard deviation under under 10 % for a P_e of 10^{-4} , and less than 1 % for a P_e of 10^{-3} (or higher). Our results are shown in Fig. 2-3 for ADC resolutions 2,3,4 and ∞ (no ADC).

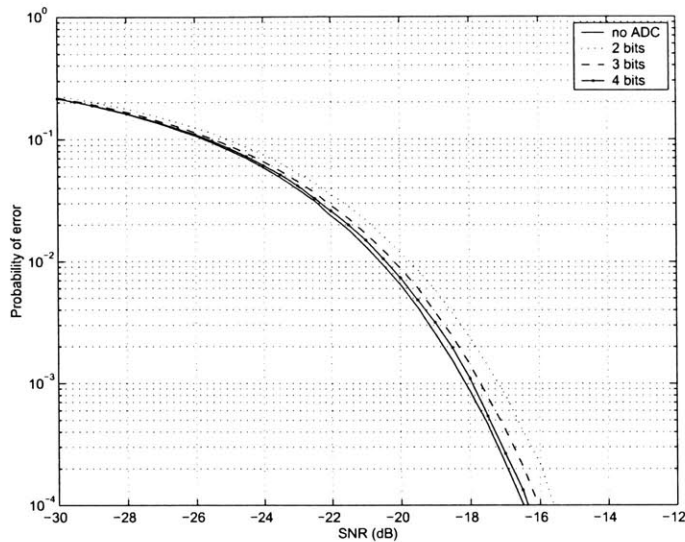


Figure 2-3: Probability of Error vs SNR (Simulated)

The simulations closely match the results of our analysis. The shape of the curves agrees with the predictions of (2.9). For a P_e of 10^{-3} , the gap in SNR terms from the infinite resolution case decreases from 0.9 dB to 0.4 dB in going from 2 to 3 bits. Moving up to 4 bits lowers this gap to a mere 0.2 dB . There is little to be gained by increasing the resolution further.

2.6.2 Interference-Limited Case

The interference to be modeled is narrowband in the real sense of the word, and is thus described by a pure sinusoid, rather than as a modulated carrier with a finite data bandwidth. Thus, there are no abrupt changes of phase over the duration of one bit (ie. $N_c N_f$ samples). Its frequency is a uniform random variable in the range from 0 to half the sampling rate. Its initial phase is an independent uniform random variable from 0 to 2π . Our Montecarlo simulation provided a standard deviation of 10 % for a P_e of 10^{-4} , and less than 1 % for a P_e of 10^{-3} (or higher). The simulation results are shown in Fig. 2-5 for bit resolutions of 2,3,4 and ∞ .

It should be noted that for these simulations, the same automatic-gain control

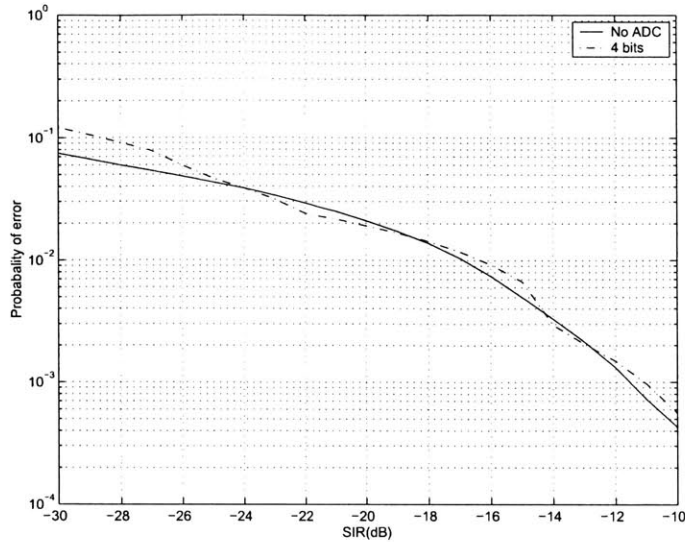


Figure 2-4: Probability of Error vs SIR (Simulated)

policy was used as in the noise-limited case. This policy was devised to minimize the quantization noise added by the ADC assuming the AGC input has a gaussian distribution. Therefore, it is presumably sub-optimal for the interference-limited case. However, its performance is still satisfactory as our simulations reveal. In principle, a system could be designed to detect which regime the receiver is currently in and adapt the AGC policy accordingly. However, there is significant computational overhead involved. For this work, we assumed straightforward and statically configured automatic gain control.

The simulation results for the interference-limited case are much more interesting, and deserve an elaborate examination. The discussion is organized around the following points :

(1) The 4-bit curve lies close to the curve for infinite-resolution (labelled “no ADC”); for a P_e of 10^{-3} , going to 4 bits results in a performance loss of only 0.5dB. Simulated curves for 5 and 6 bits have been omitted from Fig. 2-4 because they are indistinguishable from the ∞ resolution curve. Therefore, in a UWB system with such large processing gain there is nothing to be gained in going to resolutions higher than 4 bits. This is because quantization noise power is relatively insignificant compared with the

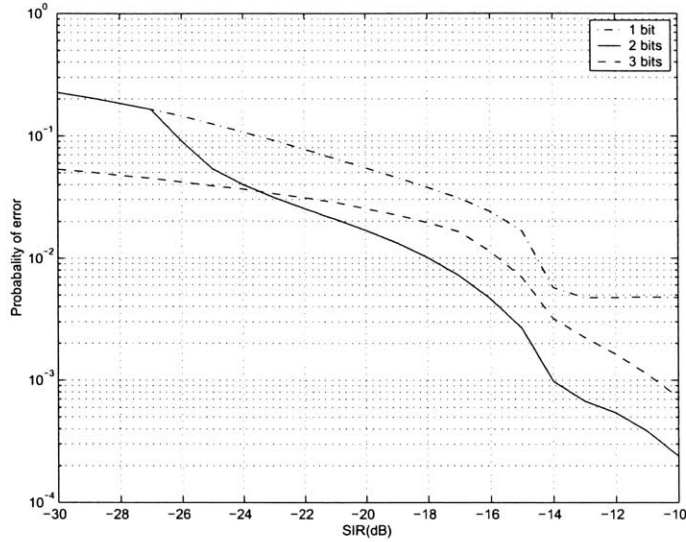


Figure 2-5: Probability of Error vs SIR (Simulated)

range of interference powers swept. This range of SIR is reasonable considering the fact that received UWB signal strength at a distance of 10m is around -65dBm as esimated earlier.

In fact, SIR values even less than -30dB are certainly possible depending on the type of indoor environment. In such a situation, more processing gain (ie. longer codes and/or lower duty-cycle) must be added to the system to help it cope with the stronger interference. This would shift the curves in Fig. 2-4 to the left, and allow the same target P_e to be achieved at lower SIR. This would, however, come at the cost of a lower data rate. An alternative solution would be to use high-Q notch filters to provide specific interference rejection.

While additional processing gain or the use of notch filters would help combat stronger interference, additional bits of ADC resolution would not. At negative SIR values, quantization noise is a small component of total noise.

(2) For 2 and 3 bits, kinks are seen in the curve stemming from the severe nonlinearity of the ADC characteristic at low bit resolutions. For SIR's between -23dB and 0dB, the probability of error is actually lower for 2 bits than infinite bits (no ADC). There is an explanation for this phenomenon. Consider the case illustrated in Fig. 2-6.

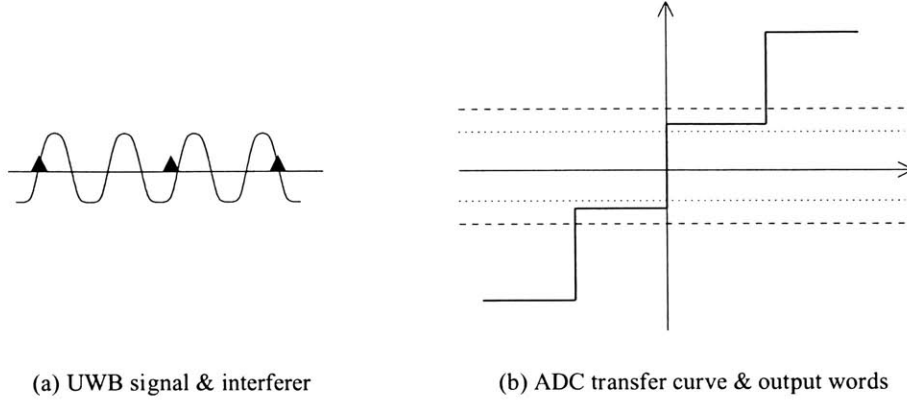


Figure 2-6: 2-bit Quantization Effects

The SIR here is around -20 dB and a 2-bit ADC is used. During the portion of the window where the UWB signal is present, the total received signal has large enough magnitude to enable the MSB of the ADC. During the rest of the window, there is only interference and no signal. The ADC output is now either a plus or a minus LSB (least-significant bit) and exhibits less variation than it would for higher bit resolutions. The signal-to-noise ratio in the output codeword is thus considerably higher than -20dB. Due to strong *negative* correlation between the interferer and the quantization noise, the ADC has raised the effective SNR as the signal goes through it. The typical assumption made for ADC resolutions above 4 bits is that the input and quantization noise are uncorrelated, but that cannot be applied to resolutions lower than that. As a result, an additional term must be introduced into the denominator of the expression for SNR after the ADC :

$$SNR_{\text{after ADC}} = \frac{\frac{A^2 d}{\alpha^2}}{\frac{\sigma^2}{\alpha^2} + \frac{\Delta^2}{12} + 2\rho_{\text{corr}} \frac{\sigma}{\alpha} \frac{\Delta}{\sqrt{12}}} \quad (2.16)$$

For an SIR of -20dB, the correlation coefficient ρ_{corr} is -0.27 for a 2-bit ADC and +0.19 for a 3-bit ADC. This explains why the effective SNR rises in the former case, but falls in the latter as the signal gets quantized. Sweeping the range of SIR's, obtaining the resulting correlation coefficients and plugging them back into Eq. 2.16 above yields new SNR values that appear to be consistent with the kinks in

the interference-limited curves for 2 and 3 bits. Closer agreement would, of course, require further modifications to the simple model contained within Eq. 2.16. This is because other implicit assumptions are also inapplicable at very low bit resolutions, such as the uniformity of each quantization error sample and the gaussianity of a sum of several such samples.

(3) Consider the question of whether 1 bit of resolution would suffice. A 1-bit ADC is just a comparator and would be an extremely small, cheap and low-powered solution to the problem of high-speed data conversion. In both the noise-limited and interference-limited regimes, there is considerable degradation in performance in going from 4 bits to 1 bit. In the noise-limited case, the performance penalty grows as the SNR increases. For a P_e of 10^{-3} , using 1 bit implies a 5dB loss. Another phenomenon that occurs with a 1-bit ADC in the interference-limited case is the “clipping” of P_e for UWB signal amplitudes larger than the interferer amplitude. With a 1-bit quantizer, this means that the signal is always seen whenever it is present. But it also means that during parts of the window where there is no signal, the interferer is also always seen with the same 1-bit magnitude. Thus, the signal-to-noise ratio after the ADC stays constant even if the input SIR improves above this critical value. As for the latter, its value depends on the duty cycle d . Consider a UWB signal and interferer, each with amplitude A . The average powers are given by A^2d and $\frac{A^2}{2}$ respectively. For a duty-cycle of 2%, this corresponds to an SIR of -14dB. As seen in the 1-bit curve in Fig. 2-5, this is the critical SIR above which P_e gets clipped at 5×10^{-3} .

A raw bit error rate of 5×10^{-3} is higher than what is needed for most applications. Let us consider ways of improving this “clipped” P_e . As mentioned above, the SIR at which clipping occurs is determined by the duty cycle, and decreasing the latter only pushes this threshold to a lower SIR and thus a higher P_e . The other component of processing gain is the code length, and increasing this does help. The curves shown in Fig. 2-5 move leftward and so the P_e corresponding to -14dB is lower. The trade-off is a lower data rate. There is also the 2-3dB performance penalty referred to earlier in using 1 bit instead of 4 bits, and this translates to lower operating range for the same transmit power.

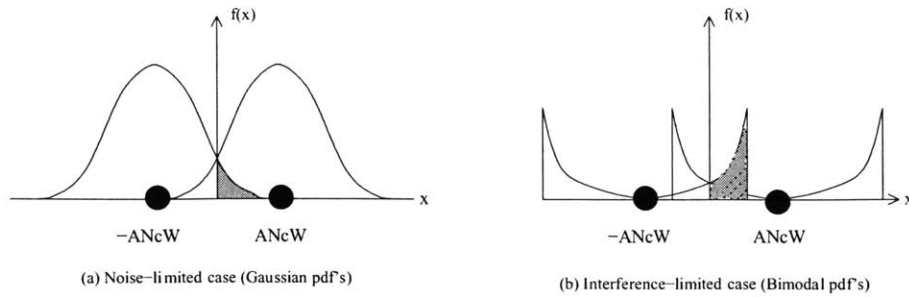


Figure 2-7: Signal + Noise (post-correlation)

Therefore, it can be concluded that a 1-bit ADC may be used in the receiver if raw performance (data rate and range) is not critical to the application. The associated power saving would be huge and would presumably enable a whole class of low-power digital UWB radios.

(4) An exciting possibility raised by the above observations is that of dynamic A/D resolution scaling. It has been noted that for a certain range of SIR values, 2 bits *improves* performance over 3 and 4 bits. If the backend DSP can be designed to detect such a condition, the ADC resolution can be scaled down to squeeze out a few extra dB of performance. The simplest way of scaling down resolution is simply to throw away LSB's, but a clever scheme that shuts down certain comparator banks can be devised to reduce power consumption. If, on the other hand, performance requirements fall and a lower data rate is acceptable, a single comparator can be used for 1-bit A/D conversion. As stated earlier, there is an exponential relationship between power consumption and resolution for a FLASH A/D, so the potential power saving would be huge.

Dynamic ADC scaling would allow for operation at the right power/performance point, making it a key feature of a reconfigurable UWB software radio.

(5) The difference in shape between the curves in Fig. 2-5 and Fig. 2-3 is due to the AWGN and interference amplitudes having different probability distributions. The amplitude of a single sinusoidal interferer has a bimodal pdf, since the sinusoid spends considerably more time near its peaks than its zero crossings. AWGN, as the name

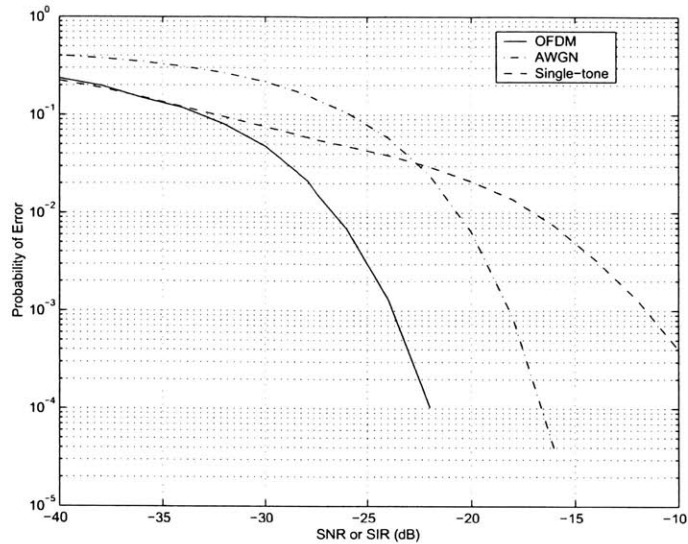


Figure 2-8: Simulation Results (OFDM interferer)

implies, has a gaussian distribution. Fig. 2-7 illustrates the two pdf's added to the post-correlation signal; the amplitude of the latter is AN_eW or $-AN_CW$ depending on the sign of the underlying bit. The shaded area represents the probability of error for that SNR or SIR.

If the total interference power were instead distributed across multiple tones, the probability distribution of the resulting signal would be a convolution of the individual bimodal pdf's. For a large number of tones (as in the case of OFDM), this resulting distribution would be Gaussian for the real and imaginary parts of the signal, and Rayleigh for the signal magnitude. In this case, one can expect the SIR-versus- P_e curves to look more like the noise-limited ones. Simulations using a 64-carrier OFDM signal support this hypothesis. Results are shown in Fig. 2-8. Curves for AWGN and single-tone interference have been superimposed for comparison.

In a real home or office environment, there may be a situation with only one or few strong interferers. The interference-limited regime would then apply and the curves shown in Fig. 2-5 would dictate performance. If not, the noise-limited curves of Fig. 2-3 would have greater validity.

2.7 Conclusion

In this chapter, the problem of determining the minimum ADC resolution for reliable detection of a UWB signal was addressed. As explained, the unique nature of these signals requires a different approach for arriving at such a specification than traditional methods. Based on analysis and simulations, it has been demonstrated that 4 bits of resolution are sufficient. This result eases one of the main obstacles to implementing an all-digital, software-defined UWB radio, namely high-speed A/D conversion within a reasonable power budget.

Chapter 3

ADC Design

3.1 Specifications

The key result of the preceding chapter is that a 4-bit ADC is sufficient for reliable UWB detection. This suggests that implementing a data converter running at several GSPS ought to be feasible. The primary goal of the design effort that followed was to validate this hypothesis through an integrated prototype. The ADC thus designed was to be part of a complete UWB system being developed by this research group.

Having fixed the converter's resolution at 4 bits, the next step was determining its target sampling rate. The driving application was a UWB system employing 1 nanosecond pulses. The question is what sampling rate is required to process such a signal ?

The simulations carried out in the previous chapter assumed 1 sample per pulse. Reliable detection was found to be possible despite large amounts of noise and interference due to the processing gain associated with integrating N samples over narrow windows spaced far apart. Following this argument, taking 2, 4 or more samples per pulse and integrating them should yield even more processing gain and thus better noise immunity¹.

Oversampling provides yet another benefit, namely the ability to do *fine track-*

¹Strictly speaking, there is a fixed amount of processing gain inherent in the system. Taking more samples per pulse gets one closer to the total amount available.

ing[1] within a digital feedback loop. When a receiver starts up, the position of the pulses are not known a priori and coarse acquisition is needed to lock on to the right pulse positions. Subsequently, this lock must be *maintained* since the transmitter and receiver clocks may drift with respect to one another. This process is called fine tracking. The objective is to detect small offsets between a pulse's actual and expected positions and correct for it. With just 1x sampling, there is no way of knowing which direction the pulse moved since a slight time advance or delay both cause equal drops in the correlation value. With 2 samples per pulse, it is possible to perform correlation over two sub-windows, one that is *early* relative to the center of the pulse and the other being *late*. By taking the dc value of the difference between the two correlation values over a certain duration and inspecting its sign and magnitude, one can then ascertain both the direction of the offset and its extent. Correcting for this may be done by feeding back to a phase-locked loop (PLL) or delay-locked loop (DLL) generating the sampling clock for the ADC and adjusting its sampling edges accordingly. However, this is rather complex and there are issues of stability involved. A simpler solution is to oversample by an even larger factor. With 4 or more samples per pulse, there is finer granularity available within the digital domain. Offsets can be corrected simply by sliding the correlation window across the set of samples already gathered. It has been shown that 4 samples per pulse provides sufficient granularity for full-digital tracking given typical amounts of clock drift[1]. Oversampling rates higher than this yield diminishing returns especially when the power cost associated with such high-speed A/D conversion is considered.

Based on the above considerations, a 4 GSPS, 4 bit converter is needed for processing 1 nanosecond UWB pulses. Therefore, this was the target pursued in this design effort.

3.2 Architecture

Implementing such a fast data converter, even at low resolutions, poses a serious challenge. A useful metric for the achievable speed of a process technology is its

FO4 inverter delay, and this roughly 90 ps for 0.18 μm CMOS. Extrapolating this metric to comparator design, it is plausible that a track-and-latch stage using devices near minimum size can be designed to run at 4 GHz. However, offsets are inversely proportional to \sqrt{WL} and the use of minimum-sized devices would make even 4-bit performance impossible without some form of offset calibration and compensation.

A more tractable architecture for such a high-speed ADC, and one that is widely employed, is time-interleaving. M distinct ADC channels are designed to run at a speed of $4/M$ GHz, using clocks that are offset in phase by $360/M$ degrees. The overall sampling rate is thus 4 GSPS as desired. The block diagram in Fig. 3-1 illustrates such a system in which M is equal to 4. A phase-locked loop (PLL) generates 4 phases of a 1 GHz clock, each of which is used for a separate FLASH ADC channel. Outputs from the 4 channels are then re-synchronized to one arbitrarily-chosen phase.

It is important to mention that a common problem with time-interleaved architectures is mismatch across the channels. If one of these channels has especially large gain and offset errors, for instance, a tone appears in the output spectrum at $1/4\text{th}$ the effective sampling rate. This corresponds to the fact that 1 in every 4 outputs generated by the ADC is from the faulty channel. Overall resolution of the data converter is thus limited by the worst-case channel. In our application, however, 4 successive samples correspond to a single pulse. They are thus added together and the sum multiplied by 1 or -1 in the correlator. This operation tends to average out the errors across the channels. Overall resolution of the ADC is now determined by the average case. Consequently, calibration across channels that is required in most time-interleaved ADC's is not needed in this case.

Each ADC channel in this system is implemented using FLASH, the architecture most widely employed for high-speed designs. FLASH converters are comprised of a bank of $2^N - 1$ comparators[23] that sample the input at the same instant and concurrently process it to yield an N -bit output. As their name implies, these circuits compare the input with a reference voltage, yielding a 1 or 0 depending on the sign of the difference. Comparators in the bank are fed $2^N - 1$ distinct reference voltages. Their outputs constitute a thermometer code as depicted in Fig. 3-2 in which the

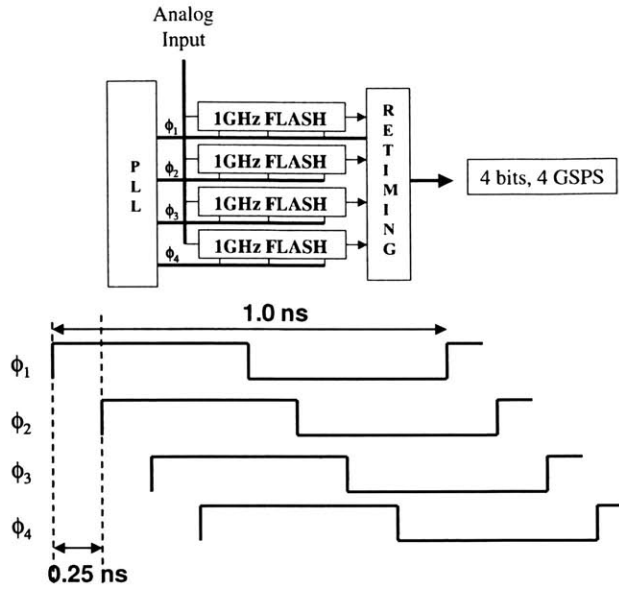


Figure 3-1: Time-interleaved FLASH ADC



Figure 3-2: Thermometer Code

position of the 0-1 transition indicates which pair of reference voltages the input lies between. A simple decoder can then generate the desired N-bit output. FLASH conversion is markedly different from other architectures like successive approximation or pipelining that break up the N-bit conversion into several smaller m-bit operations that are carried out sequentially. While the latter are useful for high-precision applications, they are not as well-suited for high-speed since they require high-gain operational amplifiers.

The choice of interleaving factor is dictated by several considerations. These include the maximum achievable speed of each channel, the challenges involved in multi-phase clock distribution and the cost of the total solution. An interleaving

factor of 8, for instance, would reduce the speed requirement on each channel to 500 MHz. However, 8 FLASH channels occupy twice as much area as 4 such channels. Furthermore, the distribution of 8 phases of a 1 GHz clock across the chip is a non-trivial problem since adjacent phases are a mere 125 ps apart. De-skew mechanisms must be incorporated into the design and these add considerable power and routing overhead. With 4 clock phases, the latter may be avoided but considerable time and attention must still be devoted to balancing path lengths and reducing coupling between phases.

3.3 FLASH Design Issues

While a FLASH architecture is an attractive choice for each of the 1 GSPS channels, it has numerous design issues that must be resolved. Common problems are outlined below, together with some techniques for alleviating them[23]. Most of these were incorporated into the design.

Problem : *Kickback Noise*

When clocked comparators are switched from track to latch mode, large amounts of charge may be injected from the devices in question. This charge couples through gate-drain capacitances and other parasitics to the input node of the comparator. The resulting noise is called kickback and is especially problematic in FLASH designs where a bank of comparators share the same input. Also, the different impedances at the input and reference nodes causes different errors on each of them and thus creates frequency-dependent offsets, also called *dynamic* offsets.

Solution :

By preceding each comparator with a preamplifier, the shared input and reference nodes are isolated from the switching comparator nodes, thus greatly reducing the impact of kickback noise. For further protection, a fully *differential* preamplifier that takes both a differential input and a differential reference may be used. Whatever kickback noise is injected onto the pair of input nodes or onto the pair of reference

nodes is thus common-mode. For further protection of the reference nodes, large capacitors can be placed there to reduce the associated voltage errors and keep the references stable.

Problem : *Overdrive Recovery*

If a large signal is applied to the comparator inputs in one clock period, and a small signal of the opposite polarity is then applied on the next one, there is a delay involved in recovering from the large output swing induced by the former.

Solution :

Addressing this problem requires keeping the time-constants of internal nodes of the preamplifier and comparator small, and including a reset switch between the comparator's output nodes that is on during the track phase. Both of these fixes reduce gain. However, this is not a problem since a huge amount of amplification is subsequently achievable through positive feedback in the latch phase of the comparator. This will be explained in more detail later in this chapter.

Problem : *Signal and Clock Delay*

An N-bit FLASH converter has 2^{N-1} comparators. Delays in the arrival of the input signal or clock across all these comparators can cause dynamic offsets. For an input signal with a peak amplitude of 1V and a frequency of 1GHz, its maximum rate of change is near the zero crossings and is equal to 6.28V/ns. For an 8-bit converter, it takes just 1.25ps for the input signal to change by 1 least significant bit (LSB). Skews on this order can easily occur due to slightly differing path lengths or loads. This uncertainty in defining the sampling instant of the comparator, referred to as *aperture* uncertainty, can considerably degrade the performance of the ADC at high input and clock frequencies. With a 4-bit ADC, the skew constraint is more lax. A single LSB is 16 times larger than in the 8-bit case. Accordingly, the time for a signal to go through 1 LSB is 16 times larger and is thus equal to 20ps. Keeping skews under this value is challenging, nevertheless, if the comparator bank spans a large area.

Solution :

The problem can be significantly attenuated through the use of a track-and-hold (T/H) circuit prior to each 1GHz FLASH channel. All comparators thus see and sample the same held input voltage. With only 4 bits of resolution, a T/H is relatively simple to design as will be discussed.

Problem : *Input Capacitive Loading*

The large number of comparators connected to the input node of a FLASH ADC results in a large input capacitance.

Solution :

A strong buffer capable of driving such a large load may be placed between the track-and-hold and the input node of the comparators.

Problem : *Skew Between Clock Phases*

Skew between the 4 clock phases is also of great concern in a time-interleaved ADC. The delay between adjacent phases is just 250ps. Moderate amounts of skew between them would have a large negative impact on the effective number of bits (ENOB) achievable. If larger than this, it could even prevent proper functioning of the chip (in particular, the re-timing block that attempts to align all 4 channel outputs to the same clock edge).

Solution :

Keeping skew within manageable limits requires careful layout, with close attention paid to balancing path lengths and loads for the 4 phases. In the re-timing circuit mentioned above, hold time constraints in the registers are the biggest concern and they can be alleviated by routing the signal and clock in opposite directions. However, this sacrifices some speed performance for the sake of robustness. This wiring scheme and others related to skew management will be discussed later in this chapter.

Problem : *Substrate and Power-Supply Noise*

The presence of a large number of high-frequency digital circuits in the vicinity of

sensitive analog circuitry makes the design highly susceptible to noise coupling. The large transient currents drawn by the decoding logic and clock drivers create ripple on the power and ground lines of nearby analog circuits, and dynamic offsets result if this noise couples in different ways to the two inputs or outputs of a comparator.

Solution :

- (1) Reducing the *amount* of coupling
- (2) Reducing *asymmetry* in the coupling to differential circuits

The first objective can be met by keeping analog and digital sections physically separated on the die and inserting thick guard rings between them. These consist of a ring of substrate and nwell ties to ground and Vdd respectively. The structure is equivalent to a reverse-biased diode and its huge resistance minimizes the amount of coupling through the substrate. Generous use of substrate ties within the digital circuits themselves also helps by ensuring low resistance paths to their local grounds. The amount of ripple on digital power supplies can be reduced using large bypass capacitors. As a further line of defense, separate power supplies can be used for analog and digital sections.

Some power supply and ground bounce is inevitable, however. Coupling occurs through parasitic capacitances in both the devices and interconnect, and thus gets worse with frequency. Maintaining good dynamic performance of the ADC requires that this noise couple in the same way to the two legs of any differential circuits. Device matching and balancing of path lengths is key, therefore, and this is more important in the preamplifier than the two comparators as will be explained in the following implementation chapter.

Problem : *Bubble Error Susceptibility*

The outputs of the 15 comparators in each ADC channel form a thermometer code as previously described, in which there is a single 0-1 transition indicating the level of the input in relation to a set of reference voltages. Under certain conditions, however, it is possible for a solitary 1 to occur within a string of 0's or vice-versa. This could happen due to comparator metastability, noise, cross-talk or large aper-

ture uncertainty. Referred to as *bubble* errors, these usually occur near the transition point of the thermometer code. The amount of aperture uncertainty can be significantly reduced by the use of a T/H as explained earlier. However, other causes of bubble errors remain and ways of guarding against them are needed. One possible fix is to use 3-input AND's instead of 2-input AND's in the decoder to detect the transition point. With this modification, there must be two 1's immediately above a 0 in the thermometer code to ascertain whether that is a transition point. However, this does not guard against a stray 0 being two places away. A better solution is to use Gray encoding as an intermediate between the thermometer code input and binary output[23]. Such a decoder is quite resilient to one or two bubble errors, the underlying reason being that adjacent numbers in a Gray code vary only by one bit.

Being reasonably assured that problems associated with the FLASH architecture can be alleviated as described, an ADC incorporating these features was designed. It should be mentioned that the main objective of this design effort was meeting the resolution required at the desired sampling rate. Less emphasis was placed on low power consumption for this first prototype. The individual circuit blocks of the ADC will now be presented.

3.4 Analog Section

Fig. 3-3 is a block diagram of the analog section of each FLASH channel. Before delving into the design of each constituent circuit, the signalling scheme and signal levels will be discussed.

Differential signalling was chosen since it improves immunity to kickback and power supply noise as described earlier. As for input swing, it must match the ADC's full-scale voltage which was nominally set to 1V peak-to-peak². The differential input ($v_{inp} - v_{inm}$) thus ranges from -500mV to +500mV. Accordingly, v_{inp} and v_{inm} swing from -250mV to +250mV about a certain common-mode and with opposite

²It is adjustable for testing purposes, however.

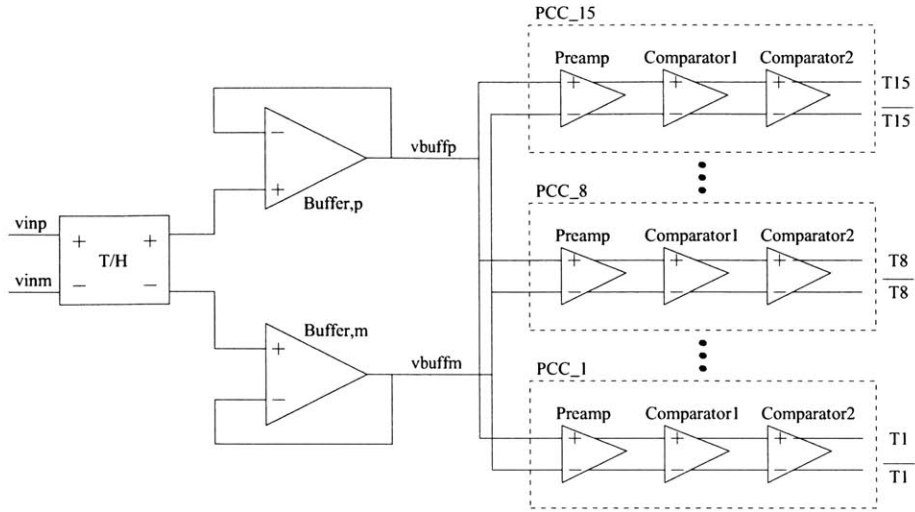


Figure 3-3: FLASH Analog Section

polarities. The use of a fairly large full-scale makes it easier to achieve the target resolution. A 4-bit ADC has 16 levels and a least-significant-bit of resolution is given by $A/16$ where A is the full-scale range. Offsets must be less than 0.5 LSB for the ADC to have 4 effective bits of precision. Raising A , therefore, means that larger offsets can be tolerated.

3.4.1 Track and Hold

A FLASH converter does not, strictly speaking, need a track and hold (T/H) circuit. This is because all $2^N - 1$ comparators sample the input signal at the same instant and concurrently process it. Thus, there is no compelling need to hold the input signal for a certain amount of time. However, it has been widely reported that the inclusion of a front-end T/H improves the dynamic performance of a FLASH ADC[3]. As explained in the previous section, the problem of aperture uncertainty and the resulting bubble errors can be mitigated by using a front-end T/H. The circuit must support a sampling bandwidth of over 1 GHz in order to process 1 nanosecond pulses. Also, it must provide differential outputs to drive the preamplifier and support the latter's required common-mode and swing. For our 4-channel ADC, 4 accompanying

T/H circuits are needed. Making this block small and low-powered is thus highly desirable.

Fig. 3-4(a) illustrates a circuit that satisfies the above design criteria. Comprising a pass-gate and sampling capacitor in each branch, it is exceedingly simple. When CLK is high, the switch is on and the output tracks the input. When CLK goes low, the switch turns off. The charge stored on the capacitor remains there and the input is held during this half-cycle. A pass-gate is used instead of just an NMOS device to allow the output to go higher than a threshold voltage below Vdd³. This is an important requirement considering the preamplifier input must be able to swing up to 250mV above the 1.1V common-mode.

This circuit has some important drawbacks, however. When the switches turn off in response to CLK going low, charge in their channels is injected on to the sampling capacitor. The magnitude of the charge is given by $WLC_{ox}(V_{gs} - V_t)$ [11] for each on device and it causes a voltage error as described to first order by Eq. 3.1. The form of the expression depends on the regime of operation of the switches (only NMOS on, both devices on or only PMOS on).

$$V_{error} = \begin{cases} \frac{q_{ch,n}}{2C_{sample}} = \frac{WLC_{ox}(V_{dd}-V_{in}-V_{tn})}{2C_{sample}} & \text{if } V_{in} < |V_{tp}| \\ \frac{q_{ch,n}-q_{ch,p}}{2C_{sample}} = \frac{WLC_{ox}(V_{dd}-2V_{in}-V_{tn}+|V_{tp}|)}{2C_{sample}} & \text{if } |V_{tp}| < V_{in} < V_{dd} - V_{tn} \\ \frac{-q_{ch,p}}{2C_{sample}} = \frac{WLC_{ox}(-V_{in}+|V_{tp}|)}{2C_{sample}} & \text{if } V_{in} > V_{dd} - V_{tn} \end{cases} \quad (3.1)$$

The first observation that can be made from the above equations is that the voltage error has two components, one that is independent of the input and one that is not, regardless of the regime of operation of the switches. By using a differential T/H, one might expect the input-independent component of the charge injection error to disappear provided the switches in both branches are in the same regime of operation. Then we are left with only the input-dependent portion which, if less than 0.5 LSB

³The source nodes of the NMOS and PMOS are not tied to their respective body terminals. Thus V_{tn} and V_{tp} are higher than their nominal values V_{tno} and V_{tpo} due to body effect.

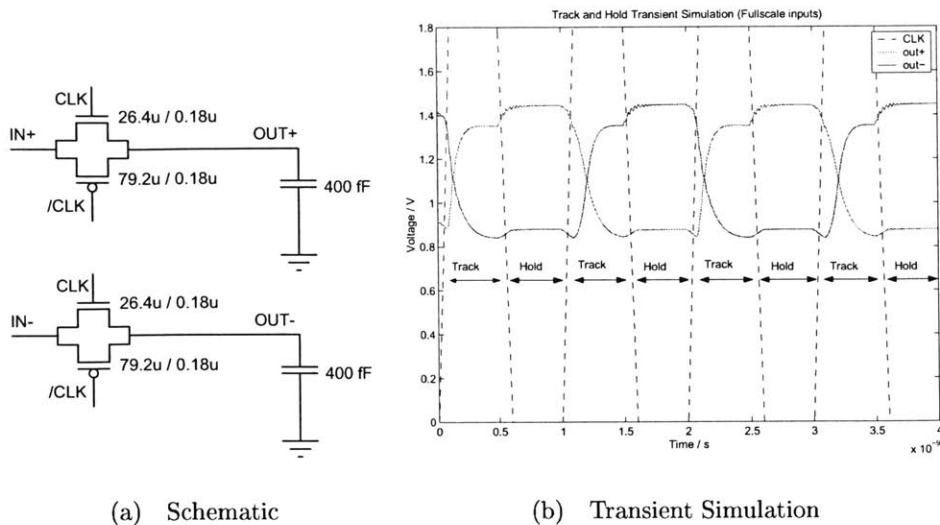


Figure 3-4: Track and Hold

(32mV) poses no problems.

In practice, large input swings about a 1.1V common-mode imply that switches in the two branches of the T/H may indeed be in two different regimes of operation. Furthermore, the equations above assumed constant threshold voltages for simplicity but this is a flawed assumption; source nodes of the switches float and thus body effect cannot be ignored. Even the component of charge injection error that was previously assumed constant, therefore, has some input dependence. This implies that perfect cancellation cannot be achieved through differential operation. Also, it underscores the fact that accurately predicting charge injection error in this circuit based on hand analysis would be rather cumbersome.

Instead, simulations were carried out to ascertain the size of charge injection errors across the entire input range. Fig. 3-4(b) shows a sample plot in which inputs to the T/H are toggled from negative fullscale to positive fullscale from one clock cycle to the next. This situation represents the worst case for charge injection error since the positive and negative input are furthest apart and thus the input dependence referred to above has the greatest impact. As seen in the plot, the step associated with charge injection at the falling clock edge has different magnitudes for out+ and out-. The

total voltage error is 62mV which is just under 1 LSB. As mentioned earlier, this number corresponds to the worst case. Inputs that are less than than full-scale cause smaller errors. A maximum differential or integral non-linearity of 1 LSB is consistent with 4-bit performance as explained in Chapter 5.

Sizing of devices and the sampling capacitors in this circuit was guided by two considerations, speed and precision. Its 3dB bandwidth is inversely related to the product of the sampling capacitance and the total resistance from the preceding drive stage to the output node. The latter is given by the sum of the pass-gate on-resistance and the source resistance of the drive stage. A sampling capacitor that is too large limits the speed achievable. On the other hand, choosing too small a value affects the accuracy of the output during both track and hold modes since parasitic capacitances at the sampling node become important. These include junction capacitances that have non-linear characteristics and introduce input-dependent errors. Accordingly, the sampling capacitor was made ten times larger than parasitics at that node. The size of these parasitics is determined largely by the size of the buffer stage that follows the T/H. The switches also contribute some parasitic capacitance, however. For this reason, they have not been made excessively large. In any case, sizing them up beyond a point has marginal benefit in terms of speed since the total resistance is then dominated by the source resistance of the drive stage. Final values chosen for the switches and sampling capacitor are indicated in Fig. 3-4. The buffers following the T/H are simple differential amplifiers in unity-gain configuration⁴. They absolve the T/H from the responsibility of driving the huge capacitive load of 15 preamplifiers. Without the buffers, this load would have have severely limited the achievable speed of the T/H.

The fact that only 4 bits of resolution are demanded from the ADC considerably simplified the design of this block. In higher-precision data converters, the T/H is indeed one of the performance limiters and elaborate techniques have been devised

⁴The design and layout of this circuit was carried out by Fred Lee, a member of this research group, for use elsewhere in the receiver. They were found to be convenient for use with the T/H also.

for cancelling the effects of charge injection and clock feedthrough. Employing such techniques requires extra area and power, however, and is not necessary given the low precision requirements of the driving application.

The 3dB bandwidth of the input buffer and T/H combination was determined to be 1.48 GHz, which comfortably exceeds specifications. The T/H circuit consumes no static power, while its dynamic power dissipation is under a milliwatt.

3.4.2 PCC Block

The PCC block (preamplifier, comparator1, comparator2) is the analog signal processing core of the ADC and each channel contains 15 of these. It comprises a preamplifier, a track-and-latch comparator and a regenerative latch as the second comparator. Its input is the differential signal coming out of the T/H, and its output is high or low depending on whether the input exceeds the reference. The design of each constituent circuit will now be discussed.

Preamplifier

The primary purpose of the preamplifier is to isolate the switching nodes of the comparators from the input and reference nodes, while providing a small amount of amplification. A fully-differential design is desirable for good rejection of both kickback and power supply noise. Such a circuit has a total of four inputs, two coming from the differential T/H and two for the differential reference that is tapped off the resistor ladder. The signal to be amplified is thus $((V_{refp} - V_{refm}) - (V_{inp} - V_{inm}))$.

The designed circuit is illustrated in Fig. 3-5 while its dc transfer characteristic is shown in Fig. 3-6 for three different differential reference values or trip points. In each FLASH ADC channel, there are 15 such preamplifiers with 15 distinct trip points. The main design decisions for this circuit are outlined below.

Sizing of Input Transistors

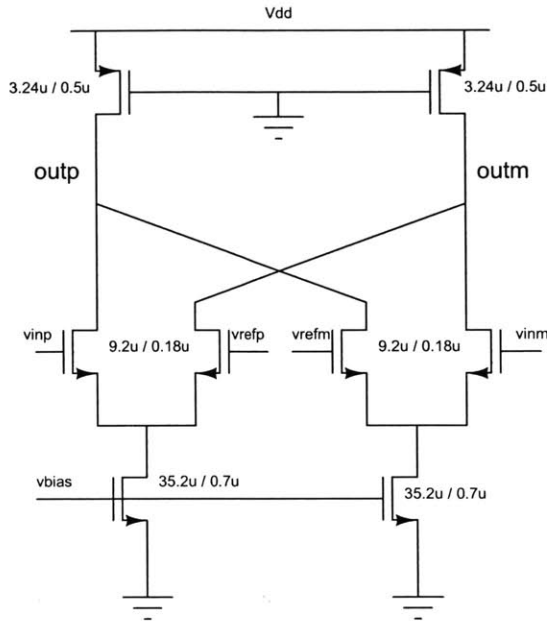


Figure 3-5: Preamplifier

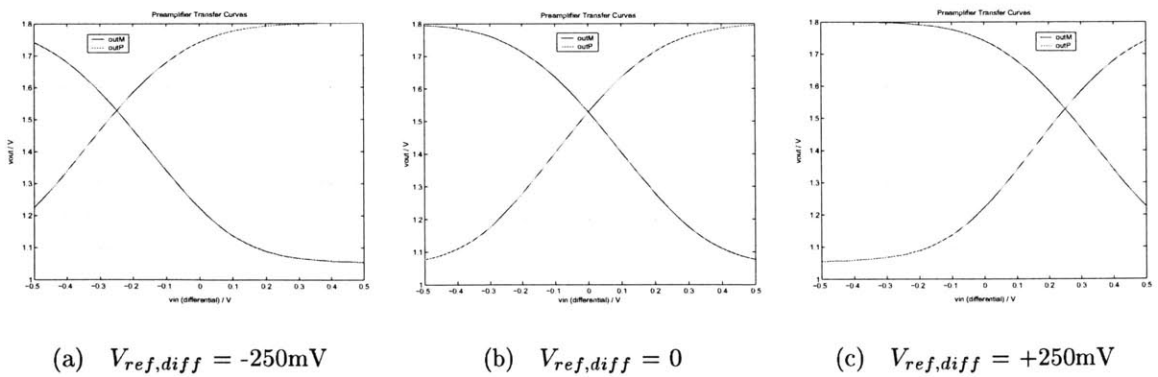


Figure 3-6: Preamplifier Transfer Curves

Minimum-length devices were used for maximum speed. Their width was chosen based on the following considerations. If the load capacitance of the amplifier is dominated by external parasitics, a larger device width yields a larger g_m and thus, a larger gain-bandwidth product. Beyond a point, however, the dependence flattens out since the device's own parasitics dominate the load capacitance. Increasing the device width further needlessly raises the power consumption. Another benefit of using large input devices is improved input-referred offset. Larger devices have smaller fractional width variation. Also, they yield a larger g_m which divides the current mismatch contributed by both the input and load devices.

Choosing the device width thus entails a trade-off between power consumption on one hand with bandwidth and input-referred offset on the other. A value of 9.2μ was found to be a suitable compromise. $200\mu\text{A}$ of bias current were assigned to the preamp, with $50\mu\text{A}$ flowing through each input device when the circuit is in the balanced state.

Type of Loads

The choice of loads was decided based on bandwidth considerations as well as output common-mode and swing requirements. A separately biased PMOS is typically a more efficient current provider in terms of its current to capacitance ratio than a self-biased diode-connected device⁵. This is because the gate bias voltage of the former is not limited by headroom considerations; it can be set to ground for the maximum overdrive attainable, thus reducing the size of the load devices and the capacitive loading they contribute. Also, only C_{db} and C_{gd} contribute to this loading, and not C_{gs} that also appears in the case of diode-connected devices. Yet another consideration favoring a simple PMOS current source was the higher output common-mode and swing attainable. Keeping V_{gs} and V_{ds} separate is key for both these metrics.

⁵A plain resistor is even better in this respect, and the fastest amplifiers are typically resistively loaded designs. However, a resistor of the requisite value would occupy considerably more area than a triode-region MOSFET which is why this option has been dismissed.

The use of this type of load, however, typically requires an accompanying common-mode feedback network (CMFB) that sets the CM level of the output nodes. In the absence of such a CMFB circuit, the output common-mode may be poorly defined and is susceptible to large changes induced by process, supply and temperature variations. CMFB circuits add another layer of complexity to fully-differential amplifiers and also raise the total power consumption. The need for such a circuit could outweigh the above-mentioned benefits of the simple PMOS current source loads. It turns out, however, that we can use them without any CMFB. This is because the possible output common-mode variations in the preamplifier are not as large as one might expect. A small bias current flows through the circuit, and the output nodes are low impedance because the preamp was designed for a small dc gain. Simulations were done to verify robustness of the preamplifier and comparator combination against common-mode changes induced by device mismatch and variations in power supply and temperature.

Sizing of Loads

Having explained the choice of load type for this amplifier, its sizing must be explained. Firstly, it should be noted that a small amplifier gain is preferable for good overdrive recovery. As explained earlier, this means that the circuit should be able to handle a small input in the cycle *immediately after* a large one of the opposite sign has been applied. The output nodes of the circuit must charge or discharge to the balanced condition before responding to the small input. A large gain in the preamplifier requires large load resistance which raises the dominant RC time constant and results in poor overdrive recovery. Therefore, we assign a small gain of 2 to this stage with the knowledge that there is plenty of gain available in the subsequent latching comparators. Since the input pair has a high g_m , a relatively small output resistance is all that is desired of the output load pair to achieve a gain of 2. Therefore, these devices can be operated in the triode region, a condition ensured by giving them a large gate overdrive. The exact size of this overdrive and of the devices is determined by two design constraints. One is the correct on-resistance for the desired gain, and the other is the bias current that must be supported. A small

W/L ratio of 6.5 in conjunction with an overdrive of $(V_{dd} - V_{tp})$ satisfies both these constraints. This solution for overdrive is especially convenient since it means the gates of the PMOS loads can simply be tied to ground.

PMOS loads this small, however, are more susceptible to mismatch and a potential concern is their contribution to input-referred offset. That is partially alleviated by the fact that this contribution gets divided by the g_m of the input pair in being referred to the input. Nevertheless, some improvement is still desirable and is indeed possible without affecting the convenient biasing arrangement for the load pair. The solution is to increase both L and W, while keeping their ratio the same. This raises the parasitic capacitance of the load pair but they are so small as it is that any associated bandwidth degradation is negligible. The final L and W values chosen after performing this scaling, therefore, are $0.5\mu\text{m}$ and $3.24\mu\text{m}$ respectively.

Table 3.1 summarizes the preamplifier's performance as measured by its gain, 3-dB bandwidth, static power consumption and input-referred offset. 3σ value is quoted for the latter and corresponds to a 99% confidence level.

Table 3.1: Preamplifier Performance

Metric	Value
Gain	2.4
BW	1.42 GHz
Offset (3σ)	13.5 mV
Power	0.36 mW

Two other metrics of amplifier performance that are typically quoted are common-mode rejection ratio (CMRR) and power supply rejection ratio (PSRR). For a perfectly matched differential amplifier, variations in input common-mode level and power supply affect both output nodes in exactly the same fashion. As long as the resulting variations in output common-mode are not large enough to create problems in driving the next stage⁶, there is no discernible effect on performance. However,

⁶The effect of power supply variations on output common-mode has already been addressed and deemed to be within tolerable limits. The effect of input common-mode level has been reduced by elongating the tail current source, thereby increasing its output resistance.

if there is some device mismatch between the two legs of the amplifier, CMRR and PSRR become important. Mismatch causes the impedance at the two output nodes to differ. Consequently, power supply and input common-mode variations contribute additional offsets in the preamp and they typically grow with frequency. This susceptibility underscores the importance of the careful layout that was carried out for this block, and the rest of the analog circuits.

Comparators

While the preamplifier provides some gain, its output is still much smaller than the voltage levels needed to drive digital circuits. The bulk of the gain comes from the comparators that follow. Both comparators are latch-based designs, the reason being that positive feedback is a more efficient way of providing large gain than simply cascading several amplifier stages. Consider a pair of cross-coupled inverters. If the initial condition of this circuit had one node A at a higher voltage than the other node B, this relationship is reinforced by the large gain of two back-to-back inverters and the two nodes are driven to Vdd and Gnd respectively. The regenerative action induced by this positive feedback may be modelled by the following exponential relationship :

$$V_A(t) - V_B(t) = V_{diff|t=0} e^{t/\tau} = V_{diff|t=0} e^{t g_m / C_{tot}} \quad (3.2)$$

τ is the time constant of the rising exponential and is equal to the ratio of load capacitance to inverter transconductance. g_m is highest when both nodes are near the inverters' switching thresholds and degrades as the devices go from saturation to triode region. Nevertheless, much of the rail-to-rail swing occurs with high g_m . C_{tot} includes the two gate capacitances of the driven NMOS and PMOS, drain capacitances of the driving NMOS and PMOS and any other parasitic or wiring capacitance at the node in question. The total amount of gain provided is related to g_m , C_{tot} and the latching time t .

Latching comparators are clocked systems. On one phase of the clock, an input tracking or re-setting mechanism is activated that establishes an initial condition for

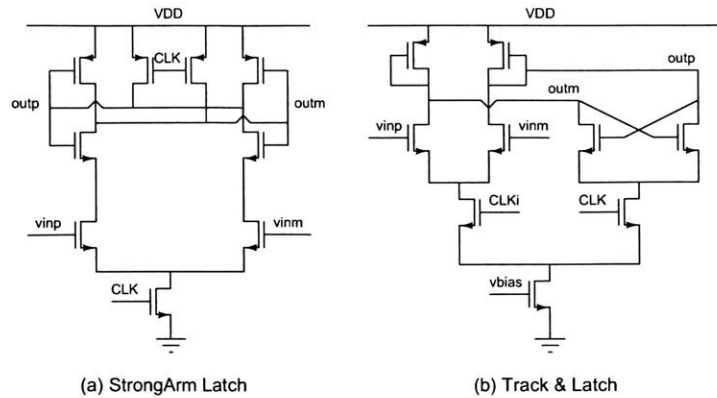


Figure 3-7: Comparator Topologies

the differential output nodes and sets $(V_A - V_B)$ prior to latching. This can be done by setting both nodes to the same rail or perhaps to $V_{dd}/2$. On the opposite phase of the clock, a positive feedback circuit similar to the one described above is enabled that regenerates even a small differential input to a full-swing rail-to-rail differential output.

Consider the circuit depicted in Fig. 3-7(a). It is a regenerative latch that was designed for the StrongArm[2] microprocessor. When CLK is low, the latch outputs are precharged high. When CLK goes high, the input pair is activated. A small differential input causes different discharge rates for the two output nodes. Positive feedback then kicks in and rails the outputs. Although designed for a digital application, it has found widespread use as a comparator. It is self-biased, has no static power dissipation and provides rail to rail outputs that are necessary for driving the decoding logic that follows. However, this last attribute has an associated cost. Charging and discharging the load capacitance from one rail to another is time consuming and draws large transient currents from the supply. This slows down the circuit and increases dynamic power consumption. On account of this speed limitation, the StrongArm latch has poor *metastability resolution*. This refers to its ability to process very small signals. It was found that for inputs smaller than 0.5 LSB (32mV), the latch yields

either a metastable output or a wrong decision⁷. In order to enhance metastability resolution, another comparator stage is needed before it. By cascading two comparators and providing more regeneration time, a higher gain is available as indicated in Eq. 3.2.

The classical track-and-latch stage[11] shown in Fig. 3-7(b) can be designed to run faster. It features two differential pairs, one for tracking the input when CLK is low and the other to provide positive feedback when CLK is high. The latter's cross-coupled drive pair is essentially equivalent to two cross-coupled inverters, but with reduced output swing and static power dissipation. It is thus akin to a ratioed-logic design style. The speed advantage comes primarily from the reduced swing. The product of bias current and load resistance determines the swing, and by restricting the latter we reduce the amount of time needed for charging and discharging. Also, the common-mode level of the outputs prior to latching can be set to $V_{dd}/2$ by proper ratioing of NMOS and PMOS devices. This initial condition allows faster settling towards the comparator's full swing than precharging both output nodes high. A related benefit comes from the comparator's tracking capability prior to latching. When CLK is low, the inputs are amplified, albeit with a small gain to facilitate overdrive recovery. At the start of the latching phase, therefore, the outputs are proportional to the value of the differential input at the rising clock edge. This speeds up latch operation. Although the source node of the cross-coupled drive pair must first discharge before those devices can turn on and positive feedback can kick in, their drain nodes already hold the sampled input. In the StrongArm latch, on the other hand, there is no tracking phase that precedes latching. The response to the inputs sampled at the rising clock edge is slower because of the indirect coupling between the inputs and the differential output nodes. Not only must the source node of the input pair discharge sufficiently, but so must their drain nodes, before positive feedback can kick in.

The two comparators described are used in conjunction, with the track-and-latch stage preceding the StrongArm latch. The fact that the final output appears one

⁷The latter could be induced by noise coupling or memory of the previous decision (hysteresis).

clock cycle later with two comparator stages is hardly a problem due to lax latency constraints imposed by the digital signal processor in the backend of the UWB receiver. Total power consumption is higher with the two-comparator solution but only by 30%. Although the track-and-latch stage burns static power, its dynamic power dissipation is small due to the reduced output swing.

Details of the final design such as device sizing and biasing will now be presented. Noting that the input pair of the first comparator has a dominant impact on the total input-referred offset, this pair of devices was designed to be large. A relatively small bias current was allocated to keep the required gate overdrive for the input pair small. This reduces susceptibility to variations in the input common-mode level that could occur due to mismatches in the preamplifier. Having established the bias current of the first comparator, its PMOS loads were sized so as to provide output swing from 0.8V to V_{dd}. A reset switch was added between the differential outputs to reduce the gain during the track phase to just over one. A low gain allows the circuit to recover from the large output swing of the previous latch phase and readies it for the next comparison. The regenerative pair was sized for speed, based on the expression for latch-mode time constant contained within Eq. 3.2. When the capacitance at the output nodes is dominated by the external load and wiring parasitics, sizing up the regenerative pair speeds up the positive feedback by increasing g_m . Beyond a certain size, however, the device pair's own parasitics begin to dominate and diminishing returns set in. Speed considerations such as this were the dominant constraint for sizing the devices in the second comparator.

PCC Performance

The complete PCC circuit is illustrated in Fig. 3.4.2. The two comparators are clocked using opposite phases to provide two successive stages of positive feedback. In other words, the second comparator enters latch mode just as the first comparator has finished latching. Additional circuitry must be added to the second comparator in order to facilitate a simple interface with the digital decoding logic that follows. The comparator's outputs are reset every half-cycle by being precharged high. Ideally, we

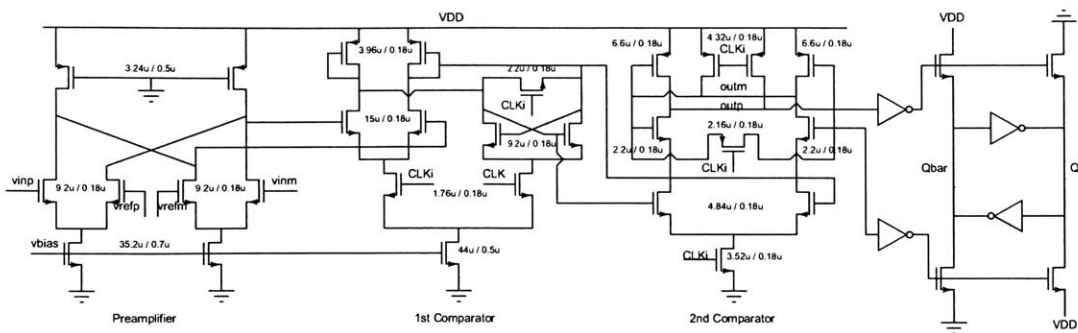
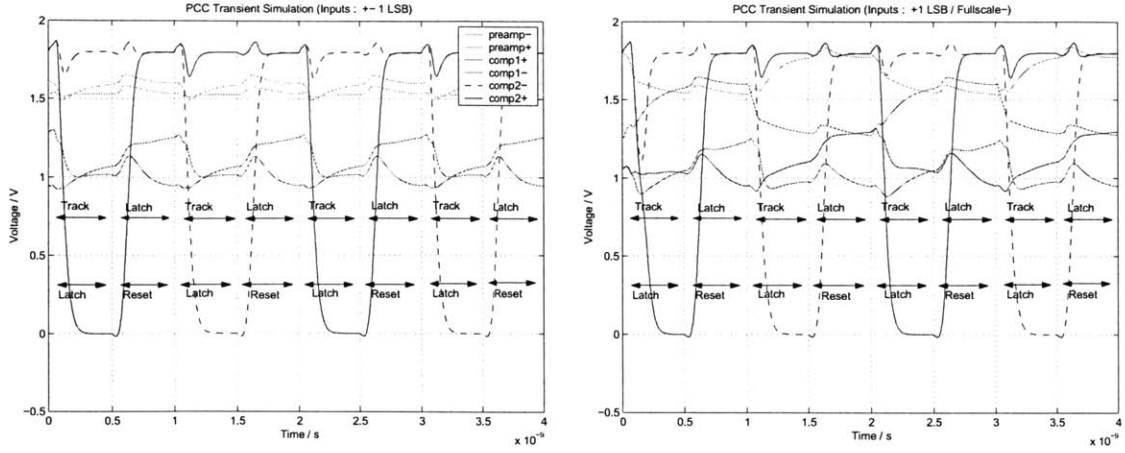


Figure 3-8: PCC Block

would like Q and Qbar outputs that are valid for an entire clock cycle. This can be done by following the StrongArm comparator with a stage that functions as a track and latch. NMOS switches connect the Q and Qbar nodes of the final cross-coupled inverter pair to Vdd and Gnd or vice versa depending on the StrongArm outputs that gate them. When the latter goes into reset mode and its outputs are both charged high, the switches are disconnected and positive feedback in the final cross-coupled inverter maintains the values of Q and Qbar. The overall circuit is also referred to as a regenerative amplifier-based flip-flop[2].

Two sets of simulation results are presented in Fig. 3-9, one with the preamplifier inputs toggling between +1 LSB (64mV) and -1 LSB (-64mV) from one clock cycle to the next, and the other based on inputs toggling between one LSB (64mV) and negative fullscale (-250mV). Outputs from the preamplifier, first and second comparators are overlaid on the same plot so as to clearly illustrate the timing relationships. The functional state of the first comparator is indicated by labels above the higher set of arrows. The lower set of arrows and their associated labels describe the second comparator's state.

It should be noted that in Fig. 3-9(b), the first comparator's outputs just barely change sign by the end of the 0.5ns tracking period in response to a 1 LSB input applied in the cycle *immediately after* a fullscale input of the opposite sign. If the small input were any less than 1 LSB, the first comparator's outputs would not cross



(a) Inputs : +1 LSB / -1 LSB

(b) Inputs : +1 LSB / Fullscale-

Figure 3-9: PCC Simulations

over on time and consequently, the second comparator's outputs would not change either. In other words, the ADC output would remain at the value determined by the fullscale. There is a trade-off between a comparator's speed and its overdrive recovery capability. The same circuit clocked at a slower rate, therefore, would be able to handle smaller inputs right after a large one. The PCC block was designed to have overdrive recovery that is just barely sufficient for 4-bit operation when clocked at 1GHz and with input signals up to 1 GHz.

Overall input-referred offset was characterized using Monte Carlo simulations. Following the norms of classical mismatch analysis, variations in each device's threshold voltage and W/L ratio were modeled as gaussian with standard deviations inversely proportional to the device area.

$$\sigma_{V_t} = \frac{A_{V_t}}{\sqrt{WL}} \quad (3.3)$$

$$\sigma_{\frac{\delta W}{W}} = \frac{A_{\beta}}{\sqrt{WL}} \quad (3.4)$$

Mismatch parameters were extrapolated from data for the TSMC 0.18 μm process. Based on these numbers, Monte Carlo simulations were carried out to determine the

input-referred offset contributed by each device independently. These were transient simulations in which a slow input ramp was applied, and the time at which the output toggles was used to extrapolate the circuit’s input-referred offset[3]. Threshold voltage variation was modeled by adding a voltage in series with the gates of the device in question. This voltage was parametrized as a gaussian random variable with zero mean and standard deviation A_{Vt} . The width of the device was also parametrized, with mean equal to the nominal designed value and standard deviation calculated from Eq. 3.4. The offsets due to each device were then combined in root-sum-squared fashion[7] to yield an estimate for the circuit’s overall input-referred offset⁸.

Offset and other key metrics of PCC performance are summarized in the table 3.2. 3σ value of the offset is 18 mV and thus, on the order of 0.25 LSB (16mV). The PCC unit far exceeds the requirements for 4-bit resolution. The bandwidth quoted is that of the preamplifier and first comparator cascaded. The latter was set to be in track mode for purposes of this bandwidth characterization. Although the overall bandwidth incorporating the latch stage of the first comparator as well as the second comparator is lower than the 1.25 GHz number quoted in the table below, the degradation is not severe[14] and this number serves as a reasonable estimate. The fact that it is well over the design objective of 1GHz is sufficient validation. Total power consumption is listed, together with a breakdown across the three sub-circuits⁹.

Table 3.2: PCC Performance

Metric	Value
BW	1.25 GHz
Offset (3σ)	18 mV
Total Power	1.30 mW
Preamp	0.36 mW
Comparator1	0.22 mW
Comparator2	0.72 mW

⁸It should be noted that this estimate is pessimistic since it assumes the variations of each device are totally uncorrelated. In a real integrated circuit, however, the devices in a small circuit are closely coupled especially if common-centroid techniques are used for the layout.

⁹Comparator2 consumes no static power, only dynamic.

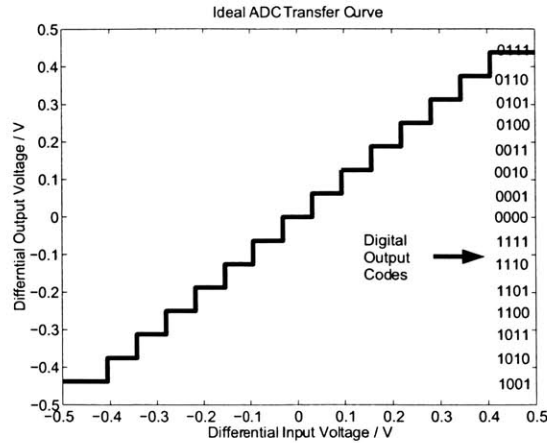


Figure 3-10: ADC Transfer Curve

3.4.3 Resistor Ladder

The desired ADC transfer curve¹⁰ is illustrated in Fig. 3-10. In order to realize this, 15 distinct reference voltages must be supplied to the bank of 15 PCC units. A resistor ladder is used for this purpose.

The use of a fully-differential preamplifier, however, necessitates the provision of differential reference voltages. Since the transition values or trip points of the ADC are symmetric about zero, this can be achieved with a single ladder. Fig. 3-11 illustrates a string of resistors with 16 taps. It should be noted that the two transition values in the middle are at -0.5 LSB and $+0.5$ LSB and the remaining transition values are at multiples of 0.5 LSB. Therefore, $R/2$ sections are needed at the ends of the ladder to provide this 0.5 LSB shift. The 16 taps are divided into 8 pairs which provide the set of 8 negative differential references. By simply cross-coupling wires from each pair, a corresponding set of 8 positive differential references are obtained. The highest positive reference is not needed, however, since a 4-bit ADC has $(2^N - 1)$ or 15 transition values. In fact, the most negative reference is

¹⁰The plot shows continuous to discrete conversion, with voltage units on both axes. In reality, the ADC generates digital output codes rather than voltages. Hence, the corresponding codes are labelled on the right hand side. They follow the twos-complement format as per the requirements of the receiver backend.

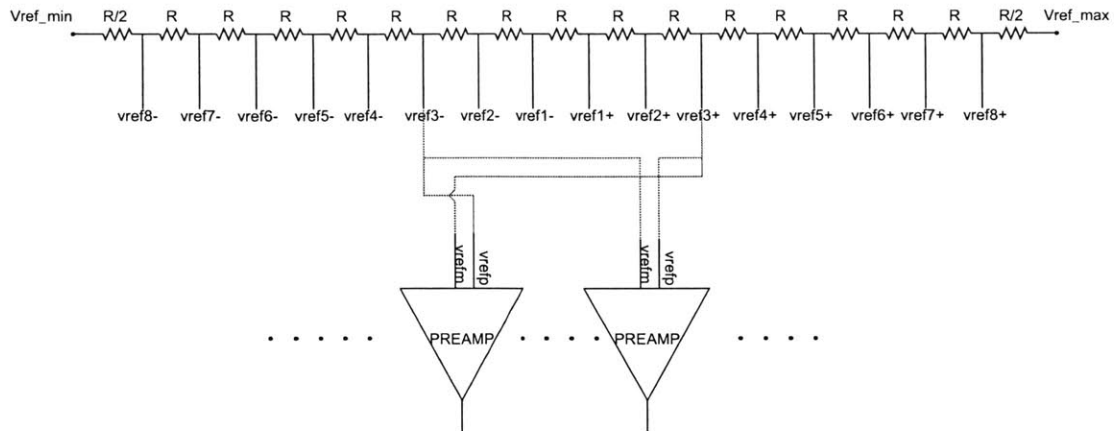


Figure 3-11: Resistor Ladder and Connections

not needed either for this application. Unlike a typical 4-bit ADC transfer curve, the one shown in Fig. 3-10 intentionally misses the lowest code value (-8 or 1000 in two's complement format). This was done in accordance with requirements of the receiver's backend signal processing, namely that maximum positive and negative excursions of the resulting digital signal be the same. By eliminating code 1000, there are exactly 7 positive and 7 negative codes. However, this specification was imposed late in the design phase and consequently, the most negative reference has in fact been used throughout the ADC¹¹.

The value of R was chosen to be $4K\Omega$. The reason for using large resistors was to reduce loading effects on the buffers supplying the top and bottom ladder voltages, $V_{ref,max}$ and $V_{ref,min}$. These buffers were originally assumed to be on-chip together with a set of voltage references that generate $V_{ref,max}$ and $V_{ref,min}$. However, due to time constraints these two nodes were pinned out and the two voltages supplied externally. The inclusion of bond wires in the signal path together with possible inductive effects in the long wires supplying references to the 4 ADC channels reinforced the need for large resistors. They provide damping and thus guard against ringing on the reference lines.

200fF ballast capacitors were placed at each of the resistor ladder taps to mitigate

¹¹The fix forcing the missing code was, instead, hardcoded into the digital section as will be discussed later

the effects of charge injection on to the preamplifier's reference inputs. The size of the associated voltage error is given by a capacitive divider relationship between C_{gd} of the devices in question and the ballast capacitors. Making them large relative to parasitics, therefore, addresses the problem.

As for thermal noise contributed by the resistor ladder, it gets low-pass filtered due to the capacitors at each tap and the bondpad capacitance at the ends of the ladder. The resulting noise, therefore, is related to kT/C rather than the values of the resistors themselves. The use of large capacitors as indicated above keeps thermal noise contributions small.

3.5 Digital Section

The outputs of the 15 PCC blocks in each FLASH ADC constitute a thermometer code. These 15 outputs must be converted to a 4-bit binary word. Since each ADC channel runs on a different clock phase, the next step is to re-align the four binary output words to the same clock edge. This facilitates their loading into a bank of registers as a 16-bit word with the same 1 GHz clock. The design of these two functional units will now be presented.

3.5.1 Thermometer to Binary Conversion

Placing a T/H before the preamplifiers in each FLASH ADC significantly reduces the probability of bubble errors. Nevertheless, the possibility of such errors one or two levels away from the ideal trip point cannot be completely eliminated. Path lengths from the T/H to each preamplifier differ and thus have different RC delays. The preamplifier input nodes will all settle to the same value eventually, but it should be noted that the system is being operated at high speed. The charging and discharging time constants are a fairly large fraction of each clock period. When the effects of clock skew across the bank of preamplifiers and comparators are considered together with the differing RC delays mentioned above, it is plausible that bubble errors could indeed creep in. We need some way of guarding against such errors.

The problem can be alleviated by using an intermediate gray code instead of converting directly from thermometer to binary. The inherent reason for the robustness of a gray code to bubble errors is that adjacent codewords differ in only one bit. An important ramification of this feature is that each thermometer input influences only one bit of the gray output. The error induced by bubbles one or two places away from the ideal trip point, therefore, is smaller than with direct thermometer to binary conversion. As always, there is a cost associated with this benefit, namely latency and power consumption. Extra logic is needed to go through the intermediate gray code and this needs to be pipelined to meet a 1 GHz throughput. The resulting latency is a non-issue because there is sufficient timing margin of several nanoseconds in the synchronization algorithms that have been designed for the UWB receiver. The power penalty is more important but can be reduced through the judicious choice of logic styles and flip-flop topology.

The following equations must be implemented by the logic for thermometer to gray code conversion. It should be noted that both polarities of thermometer code inputs are available to this block due to the use of fully-differential comparators in the analog section. G1 is the least significant bit (LSB) while G4 is the most significant bit (MSB) of the gray code outputs. T1 through T15 and $\overline{T1}$ and $\overline{T15}$ are the thermometer code inputs.

$$G1 = T1 \cdot \overline{T3} + T5 \cdot \overline{T7} + T9 \cdot \overline{T11} + T13 \cdot \overline{T15} \quad (3.5)$$

$$G2 = T2 \cdot \overline{T6} + T10 \cdot \overline{T14} \quad (3.6)$$

$$G3 = T4 \cdot \overline{T12} \quad (3.7)$$

$$G4 = T8 \quad (3.8)$$

Static CMOS is a reasonable choice for implementing the decoding logic since it is

robust, easy to design and has no static power consumption. Since latency is not an issue, the desired speed and throughput can be achieved using pipelining. Noting that a single level of static CMOS logic can only provide inverting functions, the above equations were re-formulated using NAND and NOR gates. This transformation is facilitated by the availability of complementary signals for the thermometer code inputs.

$$G1 = \overline{\overline{T1 \cdot \overline{T3}} \cdot \overline{\overline{T5 \cdot \overline{T7}}} \cdot \overline{\overline{T9 \cdot \overline{T11}}} \cdot \overline{\overline{T13 \cdot \overline{T15}}}} \quad (3.9)$$

$$G2 = \overline{\overline{T2 \cdot \overline{T6}} \cdot \overline{\overline{T10 \cdot \overline{T14}}}} \quad (3.10)$$

$$G3 = \overline{\overline{\overline{T4} + T12}} \quad (3.11)$$

The conversion from the intermediate gray code to the desired binary code can be accomplished with XOR operations :

$$B4 = G4 \quad (3.12)$$

$$B3 = G3 \oplus G4 \quad (3.13)$$

$$B2 = G2 \oplus G3 \oplus G4 = G2 \oplus B3 \quad (3.14)$$

$$B1 = G1 \oplus G2 \oplus G3 \oplus G4 = G1 \oplus B2 \quad (3.15)$$

Based on the above equations, the decoding circuit illustrated in Fig. 3-12 was designed. It comprises four levels of logic and employs seven 2-input NAND gates, one 4-input NAND gate, a 2-input NOR gate and three 2-input XOR gates. Pipelining is needed to ensure a throughput of 1 GHz, since the cumulative delay through 4 levels of logic is well over 1 nanosecond. Boundaries of the four pipeline stages are shown

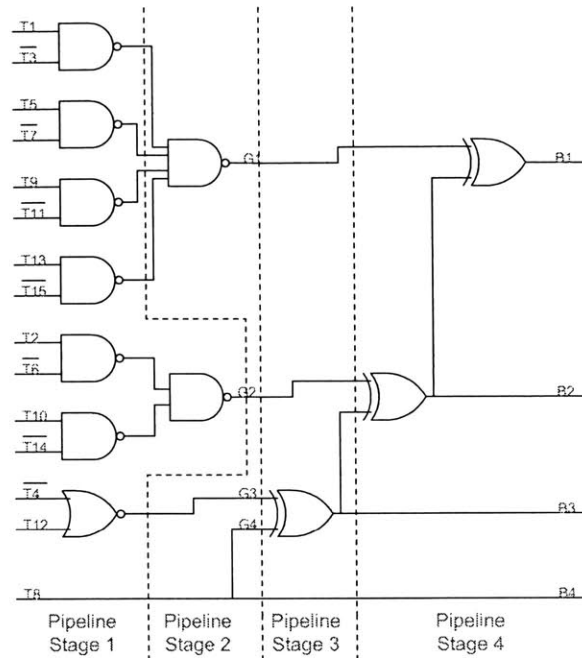


Figure 3-12: Decoding Logic

with dotted lines in Fig. 3-12. Flip-flops were inserted at these boundaries in order to register the intermediate signals.

Setup time constraints for a pipeline stage are governed by the following equation.

$$T_{clk} > t_{clk-Q} + t_{logic} + t_{skew} + t_{setup} \quad (3.16)$$

Dividing 4 levels of logic into 4 distinct pipeline stages leaves about 300ps of setup margin. Even assuming skew as large as 100ps (10% of the clock cycle), this is a conservative partition. The primary reason for leaving so much slack in setup margin was to allow the signal and clock lines to be routed in *opposite* directions. This deliberately introduces some additional skew that eliminates the possibility of hold time violations by delaying the clock of the preceding flip-flop. To further guard against hold time violations, buffers are placed in the fast sections of the critical path to ensure that the minimum delay condition is met. For instance, parts of Fig. 3-12 show a wire cutting across an entire pipeline stage with no logic between the

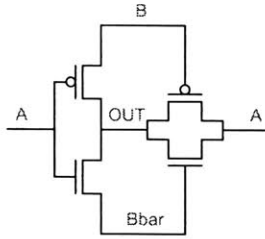


Figure 3-13: XOR Gate

boundaries where flip-flops stand. A buffer is inserted in those places to provide some minimum delay. With hold time concerns dismissed and setup time requirements comfortably met, we can be sure the decoding logic will run at any frequency up to 1GHz.

Circuit-level optimizations included sizing gates for minimum delay and ordering inputs to a NAND or NOR transistor stack such that slower inputs are closer to the output node. A regular static CMOS XOR gate contains 8 transistors instead of which the 4-transistor structure[22] depicted in Fig. 3-13 was employed. Noting that input B is more heavily loaded than input A, the slower inputs were assigned to A. The flip-flop was based on the classical configuration of master and slave transmission-gate latches as shown in Fig. 3-14. Aside from being robust, this circuit was found to have the lowest power consumption among a host of other topologies. It needs a complementary clock and although four clock phases are available in this time-interleaved ADC, local inverters were preferred to generate \overline{CLK} . This method ensures minimum skew between the two clock phases feeding the flip-flop and thus keeps its hold time low¹².

A final point that must be mentioned about the decoding logic is the need to disallow the code 1000 as per the symmetry requirements of the backend ie. equal number of positive and negative codes. Since this specification was established late in the design cycle, a simple fix was devised that required minimal change to the

¹²A small positive delay between CLK and \overline{CLK} creates a non-zero hold time, while a negative delay between them creates a transparent window around the falling edge of CLK that is potentially more problematic. Thus, \overline{CLK} is generated from CLK and not the other way around.

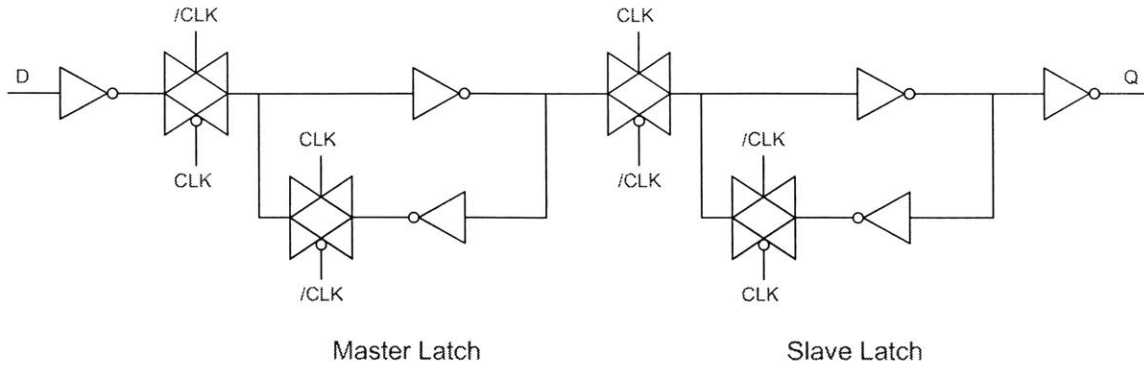


Figure 3-14: Flip-Flop

existing logic. Input T1, corresponding to the lowest level of the thermometer code, was set permanently high by tying it to Vdd.

3.5.2 Retiming

An N-channel time-interleaved ADC has an effective sampling rate that is N times the clock rate, f_{CLK} . However, the digital backend it feeds often cannot be run any higher than f_{CLK} . In fact, in our system the digital signal processing modules are run at a *tenth* of this speed. In order to interface between the high-speed ADC and the low-speed digital backend we must first re-time the 4 ADC output words, each 4 bits wide, to the same 1GHz clock phase. They can subsequently be treated as a monolithic 16-bit word that can be loaded in parallel into a bank of shift registers. The latter buffers up samples over 10 nanoseconds, and its 160 stored bits then get read together by the backend at 100MHz.

In this section, the re-timing circuit is presented. The four clock phases are denoted by ϕ_1 , ϕ_2 , ϕ_3 and ϕ_4 respectively. Since the clock frequency is 1GHz, the rising edges of adjacent phases are 250ps apart as shown in Fig. 3-15. Suppose ϕ_3 is the phase to which the rest of the ADC outputs must be aligned. Let us first consider following a flip-flop clocked on ϕ_1 with a flip-flop clocked on ϕ_3 . These two phases are 500ps apart. Since t_{clk-Q} is around 250ps at the slow-slow (SS) process corner, there is a potential for setup-time violations (assuming a worst-case skew between ϕ_1 and ϕ_3 of 150ps) making this a risky proposition. Instead, consider re-timing from ϕ_1 to

ϕ_3 in two steps, with ϕ_4 as the intermediate clock phase. The setup-time constraints are now considerably relaxed since there is 750ps between successive clock edges for adjacent flip-flops. Before deciding on this re-timing scheme, however, the potential for hold-time violations must be examined. If ϕ_1 is advanced by 125ps while ϕ_4 is delayed by 125ps, we have two successive flip-flops clocked on the same edge with no buffer in between. Since t_{clk-Q} is around 100ps at the fast-fast (FF) process corner, we have some protection against a hold-time violation. Also, if signal and clock are routed in opposite directions for the retiming block as well, the pathological case described will simply not occur in practice. Thus, the re-timing scheme meets both setup and hold time constraints and appears to be robust.

Re-timing for the other 3 clock phases is based on the same principle of cascading flip-flops and is depicted in Fig. 3-15. inA, inB, inC and inD are a subset of the ADC outputs over a 1 nanosecond interval, say the most-significant-bits (MSB's) from the 4 channels. The relationship between the clock phases is also depicted in the figure and the dotted lines refer to the sampling instants for each ADC. Thus, the signals inA, inB, inC and inD are generated in staggered fashion. They must be re-aligned to the *same* rising edge of ϕ_3 . This is achieved by cascading flip-flops and moving from each phase to ϕ_3 in multiple hops. Adjacent flip-flops are clocked with phases 270 degrees apart until ϕ_3 is reached, beyond which additional flip-flops clocked on ϕ_3 may be added for delay matching¹³. Clearly, the ADC clocked on ϕ_3 needs no retiming, but two flip-flops must still be placed in its path in order to align the final output outCr to the same clock edge as outAr and outBr. Likewise, retiming from ϕ_4 to ϕ_3 also needs only two flip-flops to achieve same-edge alignment.

3.6 External Interface

In this section, the interfacing of the ADC with other blocks in the UWB receiver, and with the outside world, will be discussed.

¹³Buffers have been inserted between adjacent flip-flops clocked on the same phase to guard against hold time violations.

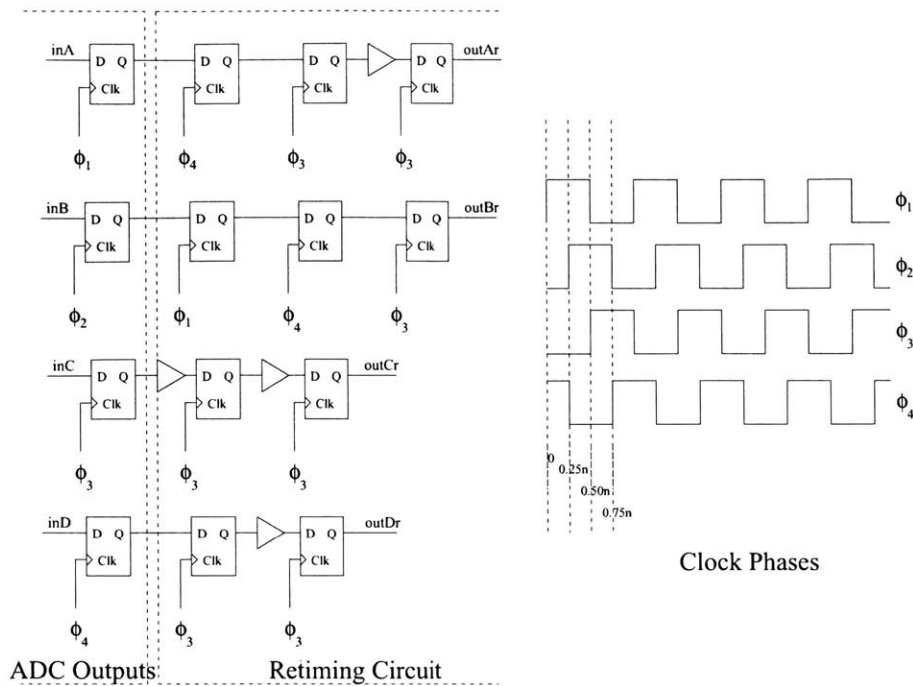


Figure 3-15: Retiming Block

3.6.1 Clocks

The phase-locked loop (PLL) providing four phases of a 1 GHz clock from a 31.25 MHz crystal reference was designed and implemented by Fred Lee, a member of this research group. Fig. 3-16 shows a block diagram of the Type II PLL, based on a popular design by J. Maneatis[15]. The voltage controlled oscillator (VCO) is a 4-stage ring oscillator. By simply tapping the outputs of each stage, the 4 desired clock phases are obtained. Differential inverters are used for each stage, and the outputs of the last one are cross-coupled to the inputs of the first to ensure oscillations with period equal to 4 inverter delays. Based on the charge pump output, the bias generator sets the bias current flowing through the differential inverters, thus controlling their delay and the resulting oscillation frequency. An important point of interest is that the higher phase noise of ring oscillators compared with LC-tank designs is not a problem for this application. Clock jitter has the same effect on receiver performance as additive white gaussian noise (AWGN) in the channel[1]. As

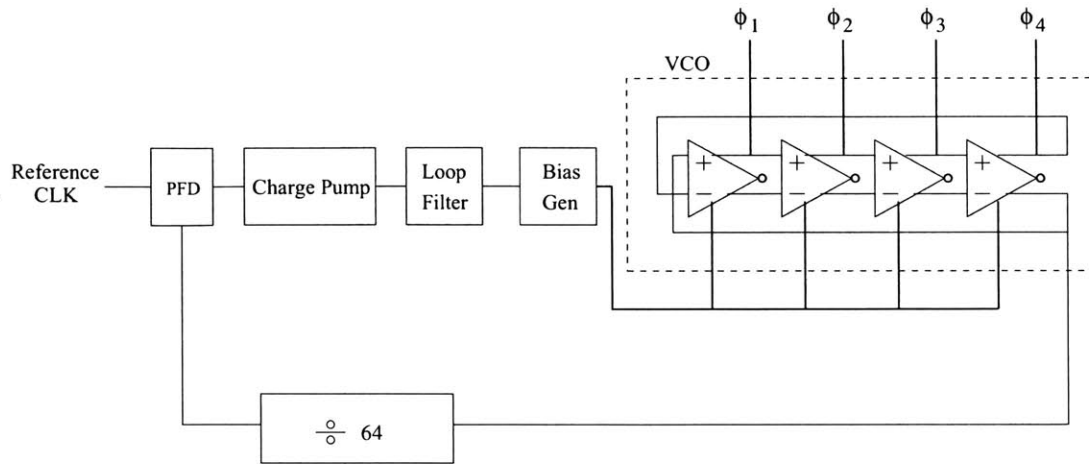


Figure 3-16: Phase-Locked Loop

explained in Chapter 2, there is considerable noise immunity in the system on account of its large processing gain. Therefore, the jitter specifications for the clock are fairly lax. It has been shown that the tolerable root-mean-square jitter is as high as 40ps (ie. 4% of the 1 ns clock period).

Clock drivers were needed to buffer up the PLL outputs and satisfy the drive requirements of the ADC. Simulations were performed to estimate the total amount of capacitive loading on the lines for each clock phase. This was found to be 6pF for ϕ_3 and less for other phases¹⁴. In order to drive this load, cascaded buffers were employed that were progressively sized up by a factor of 4 (starting from near-minimum-size inverters). There were 4 sets of cascaded buffers, one to drive each clock phase. There is large delay through these buffers and a potential cause for concern is delay mismatches across them that translate to skew between phases. The buffers span a large area across which significant process variations can occur. In order to address this concern, the buffer chains were simulated across all process corners and the delay spread was found to be at most 150 ps. Given the large setup margins the ADC was designed to meet, even in the re-timing logic, skews between clock phases on this order are tolerable.

¹⁴The discrepancy is due to the choice of ϕ_3 for re-timing.

3.6.2 Biasing

Several analog blocks in the ADC need bias currents. These include the buffers that precede and follow the T/H, preamplifiers and first comparators. Simple current mirrors were employed for the provision of these bias currents. In a FLASH ADC, an effort must be made to match operating points across the bank of $2^N - 1$ preamplifiers and comparators. Each channel in this design contains 30 current mirrors. Although the reference currents for each may be locally generated, it is simpler to inject a single external reference current through a package pin and have it distribute across all the current mirrors. However, if mismatches occur, the main reference current splits unequally. With the PCC bank occupying such a large area, mismatches are indeed likely unless large devices are used. Thus, the convenience of having a single reference current without sacrificing matching comes at the expense of large area and power drain in the current mirrors. In future implementations where low-power is a key design objective, it is advisable to use local reference currents generated using a constant g_m biasing[11] circuit.

3.6.3 Input Interface

Fig. 3-17 is a block diagram of the input interface to each ADC channel. Signal is applied *through* the low-noise amplifier (LNA) and unity-gain buffers¹⁵. There are two such buffers per ADC channel as shown, one for the positive signal and the other for its negative counterpart. Using a fully-differential buffer is preferable but two single-ended ones were chosen instead for the sake of circuit re-use and quick integration. This design decision and its implications will be examined in the Chapter 5.

The reason for inserting a gain stage prior to the ADC is reduced swing requirements at the package pin. Directly applying an RF signal with amplitude 500mV as required by the ADC is very difficult and perhaps only possible with a sophisticated package and thorough modeling of its parasitics. Applying a 10mV input at high speed is undoubtedly more tractable. The LNA then provides the necessary amplifi-

¹⁵The same circuit was inserted between the T/H and preamplifiers as stated earlier.

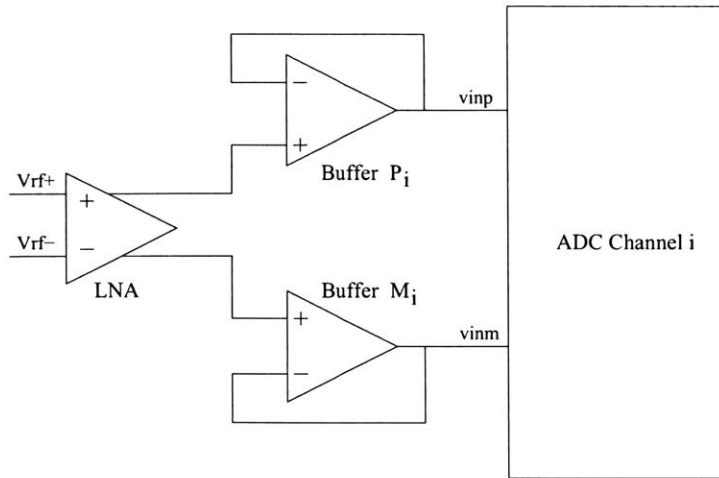


Figure 3-17: Input Interface

cation up to the ADC full-scale and the buffers provide additional drive-strength. As such, these blocks are integral components of a wireless receiver and it made sense to incorporate them into the testing framework of the ADC.

Another point worth mentioning is that there is no automatic gain control (AGC) prior to the ADC in this test-chip. In a practical system where the receive signal strength is not known a priori, an AGC is needed to ensure that the input signal amplitude matches the ADC's full-scale voltage. For a first prototype, however, this is not necessary. As long as the gain of the LNA and buffers is known precisely, an input signal of the correct amplitude can be applied to the former.

3.6.4 Test Interface

In light of the difficulty of bringing 1 GHz digital signals off-chip, a simple yet powerful methodology for testing the ADC[6] was chosen that requires only 1 out of every M samples to be acquired and post-processed. The details of this scheme will be outlined and discussed in Chapter 5. Here, the circuit for performing this decimation will be presented.

In order to sample the ADC outputs once every M clock cycles, an *Enable* signal

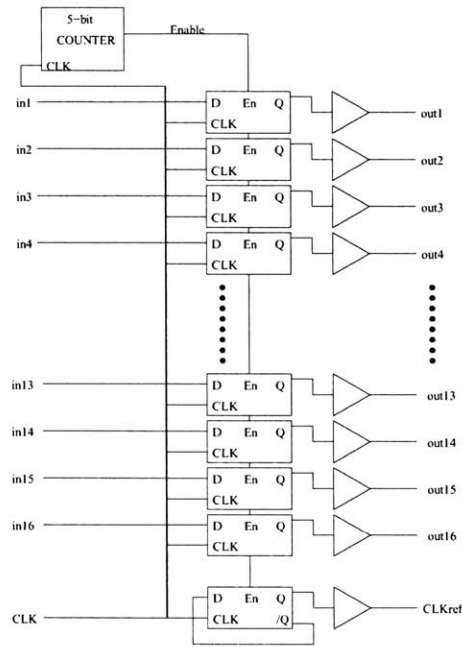


Figure 3-18: Test Interface

may be used that activates a bank of flip-flops only during the desired read cycle. This signal can be generated using a counter. The flip-flops designed for the digital section may be used here, with only a small modification to provide the enable feature. A 2:1 mux preceding the original flip-flop, with *Enable* as its select signal and with D and Q as its two inputs provides the desired functionality. The bank comprises a total of 17 flip-flops. Of these, 16 provide the decimated outputs of the 4 ADC channels, allowing each channel to be independently tested and characterized. The last flip-flop is configured to toggle once every M cycles, thus providing a divide-by-M reference clock (CLKref) that can be used as a trigger to view the decimated ADC outputs on an oscilloscope or logic analyzer.

A block diagram of the circuit is shown in Fig. 3-18. A value of 32 was chosen for the decimation factor M. The resulting frequency of CLKref and the decimated outputs is 31.25 MHz which is slow enough for these signals to be easily brought off-chip.

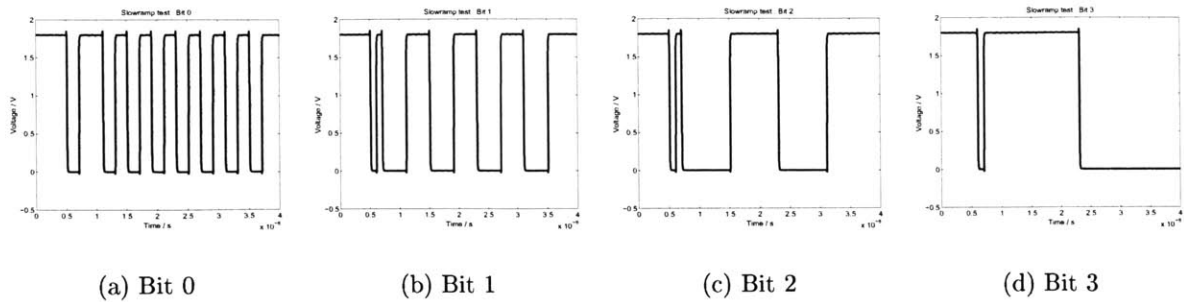


Figure 3-19: Slowramp Test

3.7 Top-Level Simulations

The various circuit blocks described in this chapter were combined and simulated as a whole to verify the chip’s functionality. Two tests were performed, both at a clock frequency of 1 GHz.

Slow Ramp Test

The ADC response to low-frequency inputs can be tested by applying a slow ramp input that covers the entire full-scale range. Fig. 3-19 shows transient plots of the 4 bits of the ADC channel under test. Since the converter has a latency of 10 clock cycles, the first 10 ns correspond to a startup transient. Hence, this period should be disregarded. Subsequently, the least significant bit (Bit 0) toggles fastest as expected. Successive bits toggle at half the frequency of the previous one. All 15 codes¹⁶ of the ADC are covered as the input ramp traverses the full-scale range.

High-Speed Test

A simple test that exercises the ADC’s response to fast inputs is the application of a 500 MHz square wave with amplitude equal to the converter’s full-scale voltage. The input thus toggles from one clock cycle to the next between the positive and negative ends of the input range. Accordingly, the ADC outputs toggle between the lowest and highest codes (1001 and 0111 respectively).

¹⁶Code 1000 does not occur as it is disallowed by design.

In order to test the ADC's robustness to process and environmental variations, both of the tests described above were carried out across process corners, temperatures from 0 to 70 degrees Celsius and power supply voltages from 1.6V to 2V. In all cases, the same response was observed.

Chapter 4

Chip Implementation

Translating a promising circuit design to working silicon at 1 GHz is a daunting task, especially for large mixed-signal chips. In a high-speed, time-interleaved ADC, there are a number of impairments that are a function of layout quality. Improper matching between devices leads to static offsets in the comparators. Noise coupling between the analog and digital sections of the chip causes dynamic offsets and bubble errors that degrade the effective number of bits (ENOB) achievable at high speeds. Long wires have large associated RC delays and become the main limiting factor on overall speed, particularly those used for clock distribution. Implementing the latter for a time-interleaved system adds another layer of complexity since skew between the various clock phases must be kept within tolerable limits. Careful layout is thus critical to ensure full functionality of an integrated circuit. Generally speaking, there are three key principles of good layout for mixed-signal systems : compaction, isolation of analog and digital, and matching.

Compaction

In fine linewidth process technologies, interconnect parasitics impose the biggest limitations on achievable bandwidth. Considerable effort must be directed, therefore, towards keeping wire lengths short. The principle of compaction must be applied across all levels of the hierarchy, from top-level floorplanning to the layout of individual blocks.

Isolation of Analog and Digital Circuits

Integrating sensitive analog circuits with noisy digital ones on the same die poses serious problems. Each time a digital gate switches, it injects charge on the digital power supply and the surrounding substrate[11]. This noise gets coupled on to analog circuits like the track-and-hold, preamplifiers and comparators and can seriously degrade the ADC's dynamic performance. In fact, it can even stand in the way of the chip's functionality if excessively large. Therefore, not only must the two sections be physically separated, extra isolation must be provided using thick guard rings. Also, it is imperative that separate power supplies be used for analog and digital. Last but not least, generous use should be made of substrate contacts and nwell ties to ensure a low-resistance path between the substrate or nwell surrounding a given circuit and its nearest ground and Vdd connections.

Matching

In any differential circuit, mismatch between devices in opposite legs gives rise to an input-referred dc offset. Dynamic offsets due to the coupling of high-frequency noise can also occur. To reduce these errors, devices in the input pair or load pair of a differential amplifier must be carefully *matched* so that the ratio of their sizes is the same as designed and their threshold voltages and other process parameters are nearly identical. The basis of all techniques for doing so is to reduce susceptibility to directional processing gradients that exist in various fabrication steps like doping. These gradients introduce dependence of a device's parameters on its position and orientation on the actual die. An effective strategy for countering these effects, therefore, is to break up the pair of transistors that need to be matched into smaller segments and distribute them in a geometric pattern that evens out their dependence on the gradients. In other words, one must ensure that the two devices are affected in the same way and to the same degree. This can be accomplished using a methodology called common-centroid layout[9] in which segments of the two devices are arrayed in such a way that their axes of symmetry intersect at a single point. This implies that

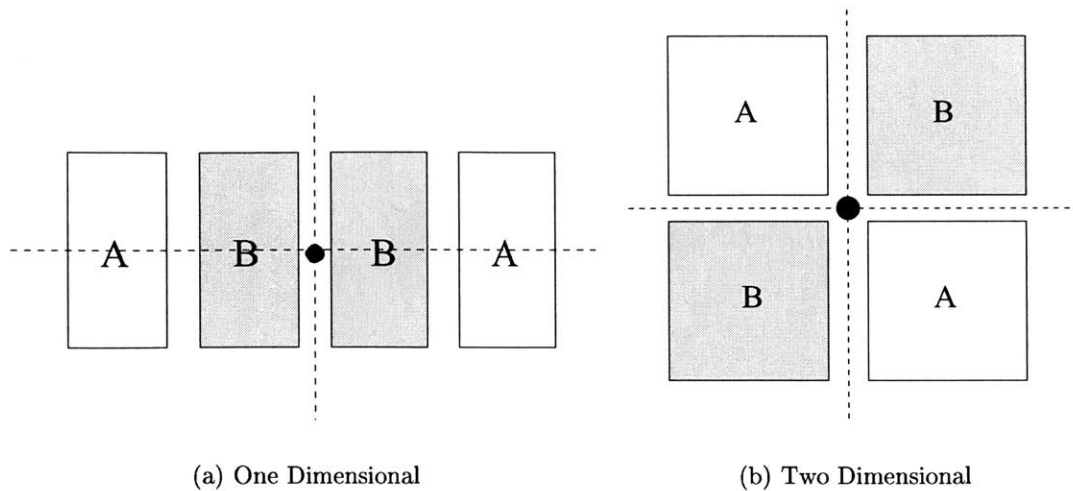


Figure 4-1: Common Centroid Layout

the centroids of each transistor coincide and herein lies the configuration's immunity to process gradients. Examples of common-centroid arrays are illustrated in Fig. 4-1. Both one-dimensional and two-dimensional topologies are possible as shown.

Better matching may be achieved by breaking up each device into even smaller segments. This ensures good *dispersion* of the two devices uniformly throughout the array, which reduces susceptibility to non-linear gradients. Fig. 4-2 shows schematic and layout views of a common-centroid transistor pair whose interdigitated fingers form the pattern ABBAABBAAB¹.

Layout, like design, is characterized by trade-offs between competing requirements. A more compact design comes at the expense of poorer matching. Not surprisingly, these trade-offs are more stringent in analog design. The first section of this chapter deals with layout of the ADC's analog circuits and highlights some of the compromises made. Layout of the digital decoding logic is then examined. Finally, top-level floorplanning is discussed and key details of the clock distribution network are presented.

¹The two devices share the same source terminal in this example which makes routing somewhat easier. This is not always the case, but is fairly common.

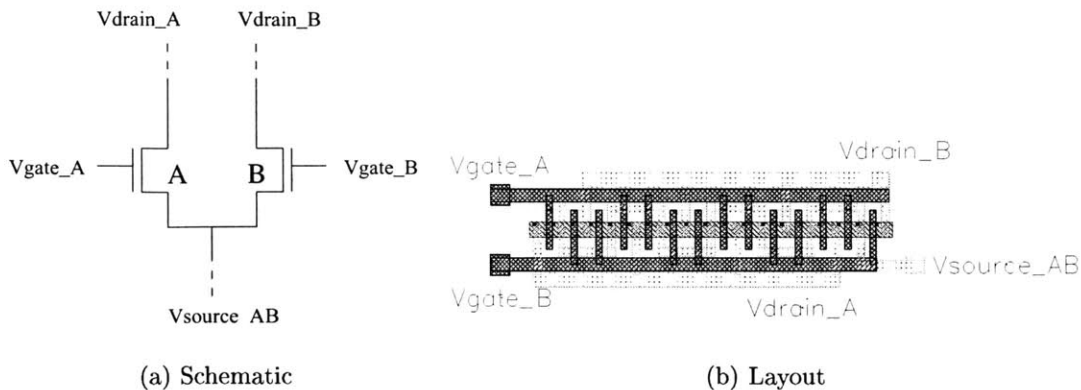


Figure 4-2: Interdigitated Common-Centroid Transistor Pair

4.1 Analog Section

Among the trade-offs in analog layout alluded to above, the most important one is between compaction and matching. Connecting together a single device's multiple segments or fingers involves considerable routing overhead, especially in 2-dimensional configurations. Wires take up precious space and their parasitic capacitance is significant in a fine linewidth process. The latter point also underscores the need to keep wire lengths balanced between the two devices in addition to matching the devices themselves. If only transistor matching is done but there is asymmetry in the wires feeding the device pair, its static offset is low but dynamic offset can be high. The space between these wires is also of critical importance in sensitive analog circuits and cannot be made too small. Cross-talk between adjacent or nearby wires can degrade performance and also limit the speed of preamplifiers and comparators. Finally, if we wish to minimize second-order effects of process gradients, an effort must be made to present the same external environment to each device in a matched pair. This can be done by inserting dummy transistors on either side of a one-dimensional common-centroid array or along the periphery of a two-dimensional one.

Clearly, a compromise must be reached between a well-matched layout and a compact one. Depending on the circuit and the role of a matched pair within it, priorities may be tilted in favor of one layout objective over the other. This is best

illustrated by comparing the layouts of the preamplifier and the second comparator.

Fig. 4-3 illustrates the layout of preamplifier's input *quad*. Since the design has two differential pairs, common-centroid principles were applied to four transistors instead of just two². A complex configuration with considerable routing overhead was chosen because mismatch in the input quad is the biggest contributor to overall input-referred offset.

The load pair and devices in the comparators that follow need not be matched to this level of accuracy since their contributions get divided down by the gain of preceding blocks. The bigger constraint for the comparators is speed and metastability resolution, particularly for the StrongArm latch since its outputs swing from rail to rail. This, in turn, requires the wiring parasitics to be small. Accordingly, the complexity of matching and its associated quality goes down along the signal path. Layout views of the regenerative pair in the first comparator and the second comparator's cross-coupled NMOS pair are shown in Fig. 4-4 and Fig. 4-5 respectively.

Simulations were carried out with extracted parasitics for the preamplifier and two comparators to ensure that these blocks still meet specifications despite the wiring overhead. They were then integrated in a compact fashion to generate the PCC unit. Fifteen copies of the latter were then aligned in a 4x4 array for each ADC channel. The layout view of the PCC bank is shown in Fig. 4-6. Ease of power, ground and clock routing was ensured by the structure's regularity. Thick power and ground buses, running both horizontally and vertically, were created to ensure small IR drops and sufficient current-handling capability to prevent electromigration. Both of these concerns are valid and it was important that they be addressed, given the 4.5mA of total bias current flowing through the PCC block.

Common-centroid techniques are also applied to the resistor ladder. Each 4K Ω resistor is realized as a series connection of four 1K Ω segments. A single row is then composed of segments of 4 resistors, interleaved in the configuration ABAB-CDCD-ABAB-CDCD. The complete ladder, comprising 16 resistors, is thus made up of four

²While mismatch between one *pair* and the other has little impact on static offsets, it can lead to dynamic offsets and must, therefore, be minimized.

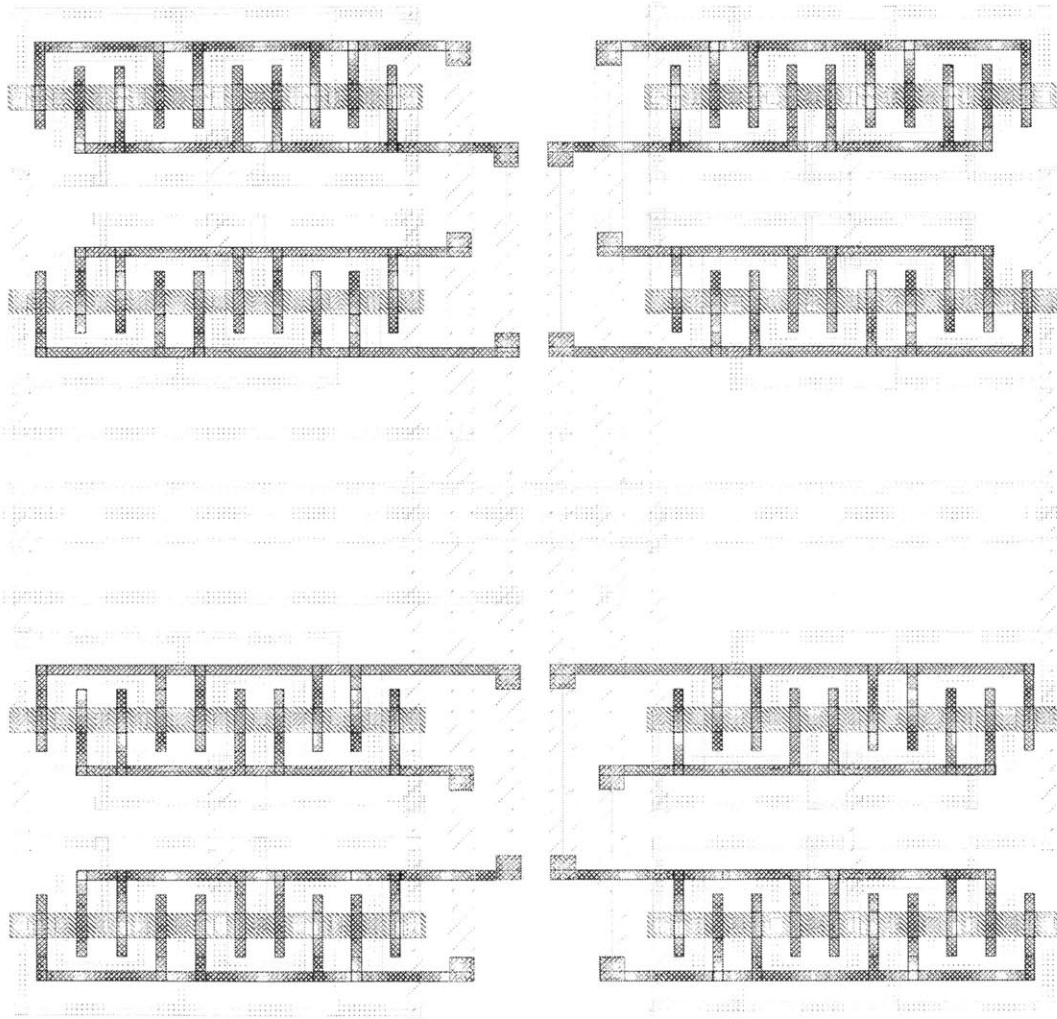


Figure 4-3: Common-Centroid Transistor Quad in Preamplifier

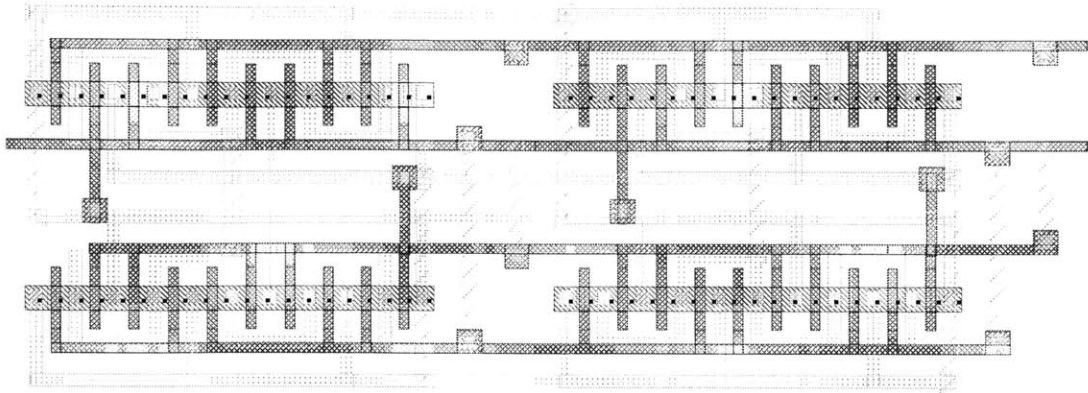


Figure 4-4: Regenerative Pair of First Comparator

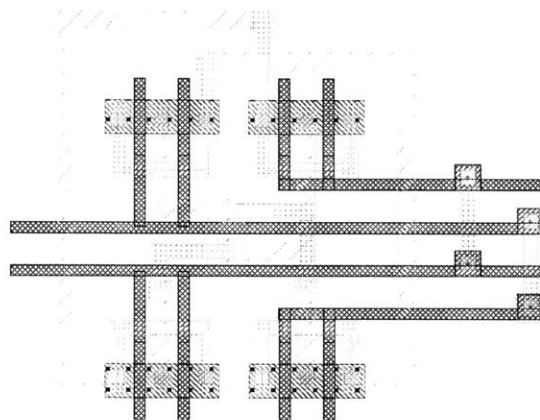


Figure 4-5: Cross-Coupled NMOS Pair of Second Comparator

	PCC15	PCC14	PCC13
PCC9	PCC10	PCC11	PCC12
PCC8	PCC7	PCC6	PCC5
PCC1	PCC2	PCC3	PCC4

Figure 4-6: PCC Bank

such rows.

4.2 Digital Section

Circuits operating faster than 500 MHz are best laid out by hand rather than relying on a synthesis tool. With greater control over wiring, location of terminals, aspect ratio and orientation of each constituent block, one can ensure that delay constraints are met throughout the signal path. Since it is critical that wires be kept short, compaction is the most important principle underlying the layout of fast digital circuits.

It should be noted, however, that compaction is more important at higher levels of the hierarchy. Wire length must be above $100\mu\text{m}$ to start approaching typical device parasitics. Local wiring thus has a much smaller impact on total achievable speed than in the distribution of global signals like CLK. Accordingly, effort towards compaction was focused on regularity within the top-level structure and choosing the right aspect ratios of constituent blocks. *Standard cell* layout is a common methodology that helps in achieving these objectives. The idea is to use the same height for each transistor, simply adding more fingers and elongating the structure if a wider device is desired. Gates and flip-flops composed of such standard cells fit well together and yield compact circuits.

The pipelined decoder described in Chapter 3 was laid out compactly through careful floorplanning of its constituent gates and flip-flops. Its layout view is shown in Fig. 4-7.

4.3 Complete ADC Layout

Top-level layout of the ADC was based on a simple floorplan. The 4 channels of the ADC were partitioned and within each one, the analog and digital sections were physically separated and guard rings were inserted around the former for substrate isolation. Also, separate power supplies were used for the two sections.

Clock distribution was one of the key challenges in this final phase of layout. The

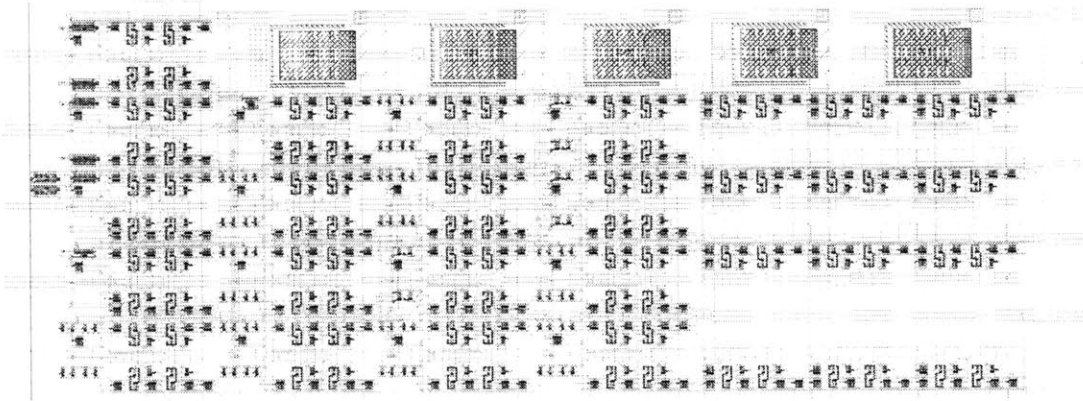


Figure 4-7: Layout of Decoding Logic

ADC is clocked at 1 GHz and each channel presents 6pF of load capacitance to its clock lines. Furthermore, with 4-way interleaving, adjacent phases are just 250ps apart. Clock lines for each phase must, therefore, be balanced in their length and loading to keep skew within tolerable limits³. Devising a balanced routing strategy for the four clock phases, however, is quite complicated. Since the dimensions of each channel are large, even a small amount of asymmetry in the clock distribution tree translates to large differences in wire length. And although each channel has a distinct sampling clock, its retiming circuits require other phases as well. Assuming we wish to re-align outputs to ϕ_3 , this phase must be fed to all the channels. Intermediate phases are also required, as described in this circuit's description in Chapter 3.

Consider two possible clock distribution networks, as shown in Fig. 4-8. Topology B is a better choice since it achieves better balancing between the clock phases. A caveat is that its total wire length is larger, and hence stronger clock drivers are needed. These buffers take up space, have large dynamic power consumption and are noisy. The objective of small skew between phases is paramount, however, and this cost had to be borne.

The relative directionality of clock and signal in the ADC was also an important

³An alternative is to use active skew compensation circuitry to manage this problem, but at the expense of additional complexity and cost. Hence this option was dismissed

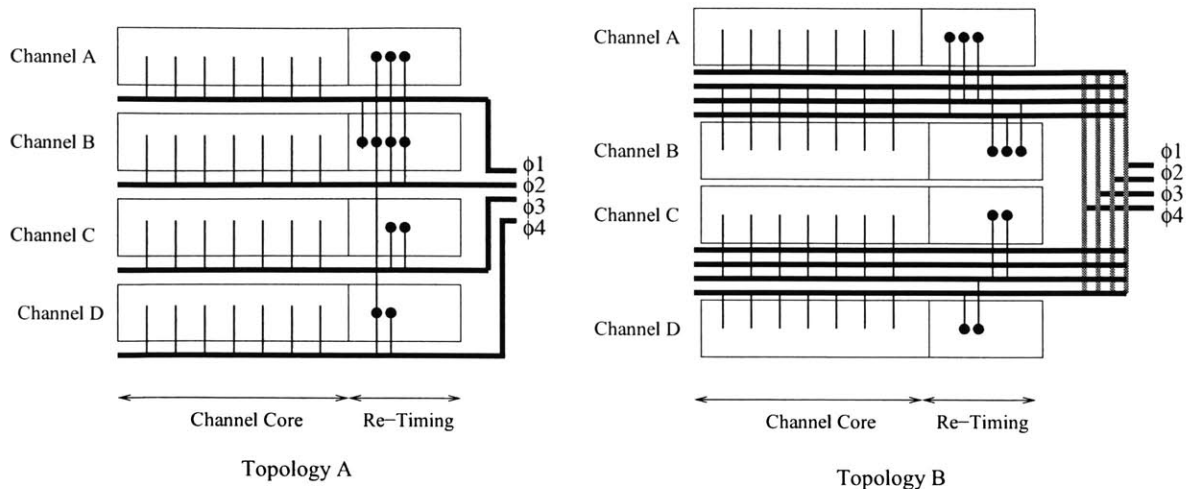


Figure 4-8: Clock Distribution Options

design decision. Since there was sufficient margin in meeting setup time requirements at all flip-flops in the chip, signal and clock were routed in *opposite* directions. This significantly reduces the possibility of hold time violations. The latter are of concern in some of the retiming circuits, as explained in the previous chapter.

Routing of other signals common to the four ADC channels was simplified by the regularity of their placement. These included inputs from the LNA, numerous bias voltages and a line carrying bias current and reference voltages from the resistor ladder. The latter was inserted in the top-left of the ADC and occupied a very small area. The fact that total wire length from the ladder to the reference inputs in the four channels is not matched is hardly a problem since these are dc signals driving high-impedance gates. The different resistances and capacitances of the path lengths are irrelevant, therefore. Power and ground routing in the top-level was simple, and comprised a large number of vertical bars, each $30\mu\text{m}$ thick. This guarantees that electromigration risk is minimal and that IR drops are small. The latter was also ensured through the generous use of contacts to the power and ground buses.

Layout of the entire ADC is shown in Fig. 4-9. The four channels run from top to bottom. In each, the analog section is isolated from the digital decoding logic using a thick guard ring. The T/H and the pair of buffers that drive the preamplifiers are

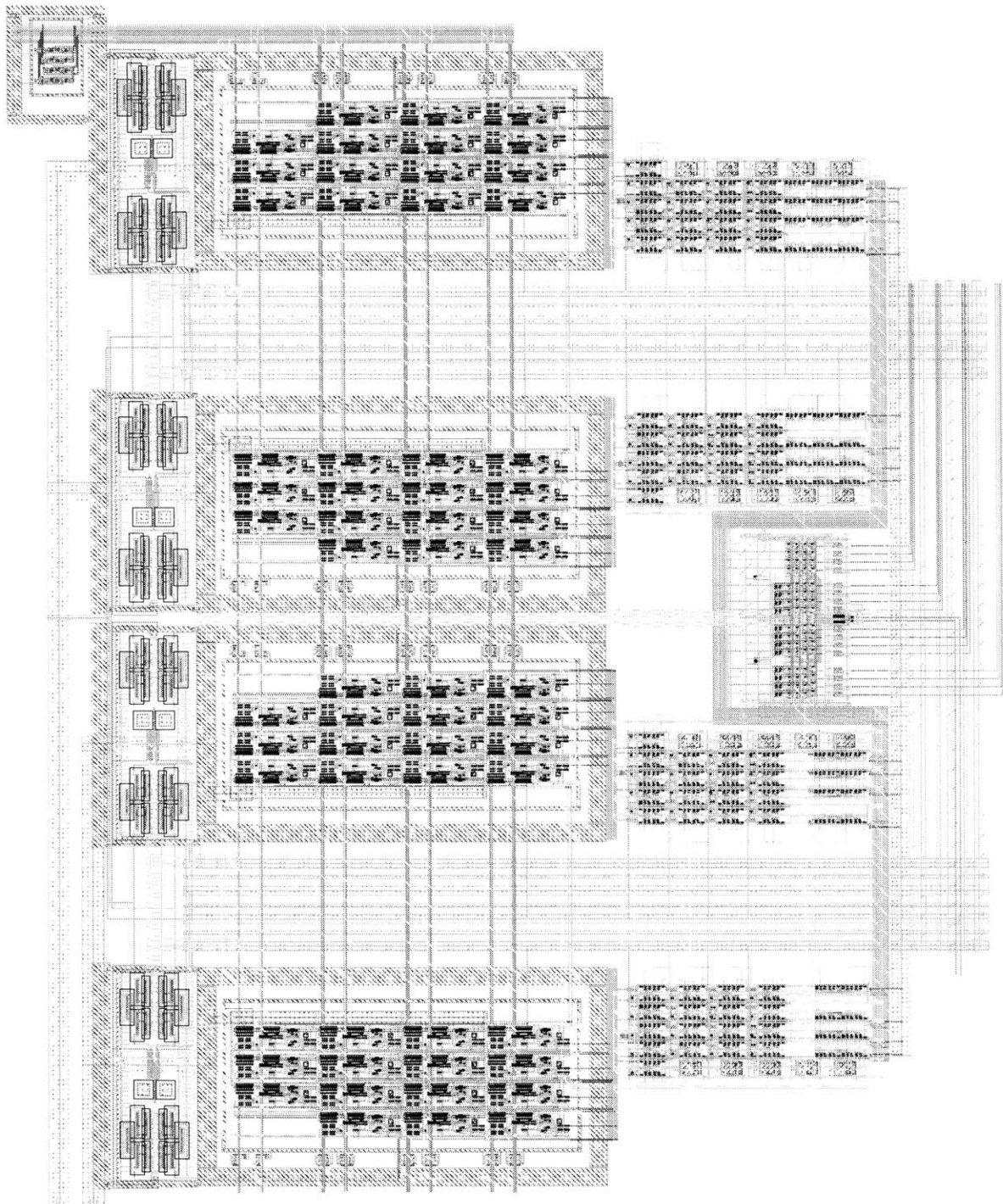


Figure 4-9: Layout of Complete ADC

located to the left of each channel. The resistor ladder sits in the top right-hand corner next to channel 1. The long, thick wires of the clock distribution network are clearly discernible.

4.4 Top-Level Simulations

Simulations were performed to verify the functionality of the ADC channels with full extracted parasitics. The channels passed the slowramp test at a clock frequency of 1 GHz at the typical corner. The high-speed test, however, revealed limitations in the ADC's dynamic performance.

- With a clock frequency of 800 MHz and an input frequency of 400 MHz, the outputs toggled between code values +5 and -5 instead of hitting +7 and -7, the top and bottom of the code range. When the full-scale voltage was reduced from 1V peak-to-peak to 5/7th of this value and the test repeated, the outputs were now found to reach +7 and -7. This confirms the presence of a bandwidth limitation that appears to attenuate the signal before it is processed by the preamplifier bank.

- With a clock frequency of 1 GHz, an input frequency of 500 MHz and the regular full-scale of 1V peak-to-peak, the outputs toggled between code values +4 and -3. Thus, a frequency-dependent asymmetry is observed in addition to bandwidth limitation.

Both these effects were absent in simulations performed without top-level interconnect parasitics. Inspecting the network of wires serving each ADC channel, it is readily apparent that the ones between the T/H and the PCC bank are the most problematic. They are long, thin wires with a maximum length of 1 mm and a width of 0.5 μm . A rough calculation of the wire's R and C values based on process information provided by the fabricator confirms that these parasitics dominate over those of the preamplifiers they drive. The same set of long wires are likely to be responsible for the frequency-dependent asymmetry described above. Path lengths from the T/H to preamplifiers 1 and 15 are 650 μm and 1 mm respectively. If the bandwidth limitation does indeed come from these wires as suggested above, this large a difference in path

lengths would explain the asymmetry in ADC outputs seen at high input frequencies. Greater RC delays in the signal path to PCC_1 mean that the lowermost code values are less likely than the uppermost ones, despite a symmetric applied input.

This simulation could not be performed prior to the tape-out due to time constraints. Back of the envelope calculations for the time constant associated with these long wires were presumably optimistic and did not fully take fringing effects into account, resulting in the problem escaping notice. The above observations are useful, nevertheless, in making sense of similar effects seen in dynamic testing of the actual chip, as discussed in the following chapter.

4.5 Chip-Level Integration

The ADC was finally integrated with a low-noise amplifier, a 1 GHz phase-locked loop and the test interface as described in Chapter 3. The resulting chip, referred to as a UWB receiver front-end, was taped out on December 9 through MOSIS using the TSMC 0.18 μm mixed-mode process. A die micrograph of the finished chip is shown in Fig. 4-10. Its dimensions are 3.2mm x 2.2mm. The chip was encapsulated in a QFN 48-pin package. Power supplies for each block were pinned out separately and most grounds were connected using downbonds to a ground plane underneath the die, referred to as the *paddle*.

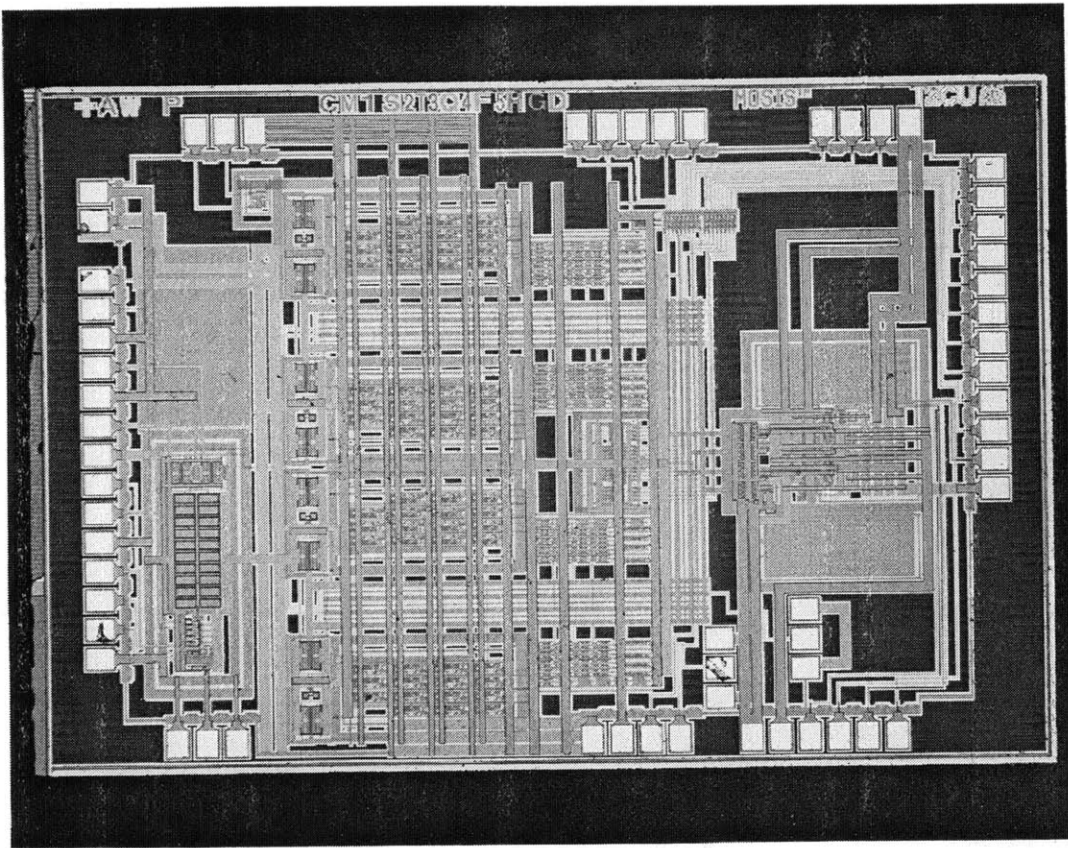


Figure 4-10: Die Photograph

Chapter 5

Testing and Analysis

The test strategy employed for the ADC is based on a popular methodology devised by Joey Dornberg, Hae-Seung Lee and David Hodges[6]. Its key concepts are briefly reviewed in the first section of this chapter. Details of the actual test setup, post-processing and measured results are then presented. Finally, ideas are proposed for improving the testability, performance and power consumption of future designs based on this architecture.

5.1 ADC Test Methodology

The underlying principle of this methodology is statistical characterization using a large pool of random samples of a precisely known periodic input signal. Assuming its frequency is not harmonically related to the clock, the sampling instants are asynchronous relative to the input signal. A large pool of ADC output samples is taken, based on which a histogram is constructed showing the number of occurrences of the 2^N digital output codes of the N-bit converter. Ensuring that all of these 2^N codes are hit requires setting the input signal's amplitude to be equal to the full-scale voltage of the ADC. The *code density* information captured by the histogram can be used to compute the $2^N - 1$ transition levels. The linearity, gain and offset errors of the ADC can then be easily determined. This provides full characterization in the amplitude domain.

There are several possible types of input signals that may be applied, a prime candidate being a ramp or triangle wave as used in the simulations described in Chapter 3. However, the reliability of ADC testing is limited by the precision of its input source. Commercial signal generators are able to produce a pure sine-wave with greater precision than a ramp, making the former a better choice for testing purposes[6].

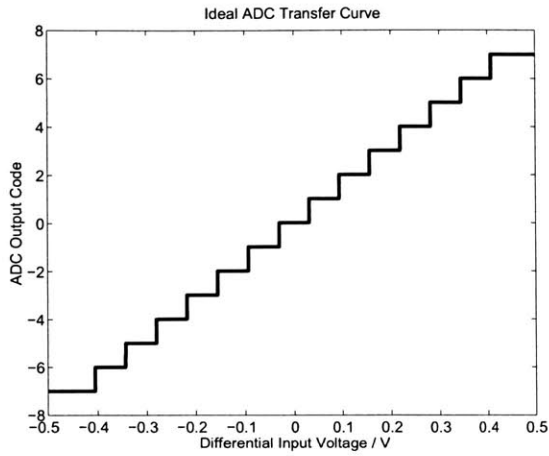
The basis of the histogram test is random sampling of the sine wave, which explains the need for the input frequency to be harmonically unrelated to the sampling clock. Random sampling vastly simplifies the problem of testing a high-speed ADC since it obviates the need to view every successive sample generated. Every Mth sample is sufficient, provided a large enough pool of such samples is gathered. The minimum number of samples N_t necessary for β bit precision and $100(1 - \alpha)$ percent confidence is given by Eq. 5.1 below[6]. The parameter $Z_{\alpha/2}$ can be found using a table of the standard normal distribution function.

$$N_t \geq \frac{Z_{\alpha/2}^2 \pi 2^{n-1}}{\beta^2} \quad (5.1)$$

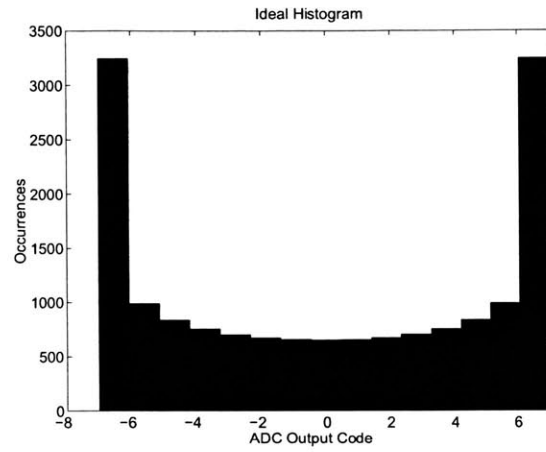
In order to determine the differential nonlinearity for a 4-bit converter to within 0.10 bits and with 99 percent confidence, 16750 samples are needed.

A decimation factor of 32 was used in the chip's test interface, as described in Chapter 3. Thus, the bank of flip-flops in this interface is enabled once every 32 clock cycles. Accordingly, the 16 output bits have a maximum frequency of $f_{clk}/32$. This speed is low enough that the signal can be easily brought off-chip. The pins for these decimated outputs and a divided-down clock signal called *clkref* are connected to a row of in-line headers on the printed circuit board to which logic analyzer probes are attached for data acquisition.

Post-processing is carried out as follows. The data acquired is parsed from a stream of bits into a vector of code values ranging from -7 to +7, corresponding to the 15 possible codes of this ADC. The lowest code value of -8 corresponding to code 1000 is disallowed in this design as explained in Chapter 3. A histogram $H(i)$ is then

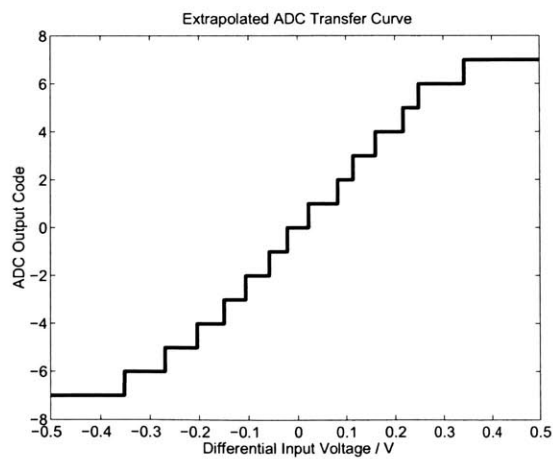


(a) ADC Transfer Curve

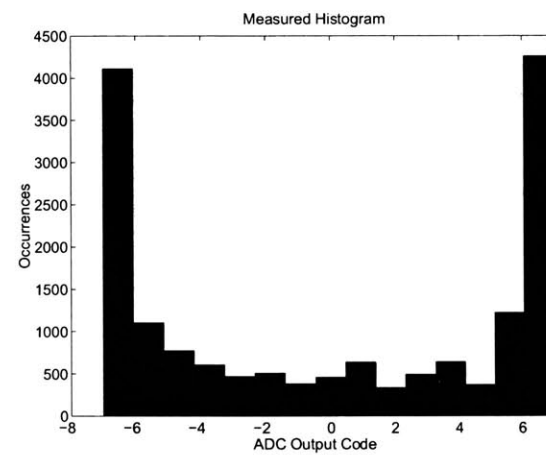


(b) Code Density Histogram

Figure 5-1: Ideal 4-bit ADC



(a) ADC Transfer Curve



(b) Code Density Histogram

Figure 5-2: Non-Ideal 4-bit ADC

constructed showing the number of occurrences of each code i . Fig. 5-1 and Fig. 5-2 show the histograms that would be obtained for *ideal* and *non-ideal* 4-bit transfer curves. Based on $H(i)$, a *cumulative* histogram CH_i is then obtained that lists the total number of occurrences of code values 0 through i . The transition voltages of the ADC transfer curve can be extrapolated[6] from CH_i using Eq. 5.2. N_s denotes the size of the sample pool. The transition voltages V_t are normalized such that the full range of transitions is ± 1 .

$$V_t = -\cos\left(\frac{\pi CH_i}{N_s}\right) \quad (5.2)$$

With this set of transition voltages, the ADC can be completely characterized. There are two common measures of a converter's effective accuracy :

(1) The maximum voltage offset, expressed in units of LSB, between the extrapolated transition values in Eq. 5.2 and the ideal ones, is the basis for a converter's *absolute accuracy*. This measure takes into account gain and offset errors as well the ADC's intrinsic linearity.

$$2^{N_{abs}} = \frac{16}{Offset_{max}} \quad (5.3)$$

(2) The ADC's differential non-linearity (DNL) and integral non-linearity (INL) provide a measure of *relative accuracy* and also are a snapshot of how that accuracy varies across the code range. The steps in an ideal ADC transfer curve as shown in Fig. 5-1(a) are 1 LSB wide but they may deviate from this value in a real converter. DNL errors quantify this deviation and are thus also measured in LSB. INL errors, on the other hand, indicate the deviation of the transfer curve from a straight line. DNL and INL plots are indicative of the performance of the ADC across its code range. Calculation of these values from an extrapolated transfer curve follows standard techniques[11]. It should be mentioned that prior to computing DNL and INL values, the gain and offset errors in the extrapolated transfer curve are removed. This is common practice[21] and provides an idea of how precisely controlled the converter's transfer response is. The maximum INL value may be used to determine the

ADC's *relative accuracy*.

$$2^{N_{rel}} = \frac{16}{INL_{max}} \quad (5.4)$$

5.2 Chip Measurements

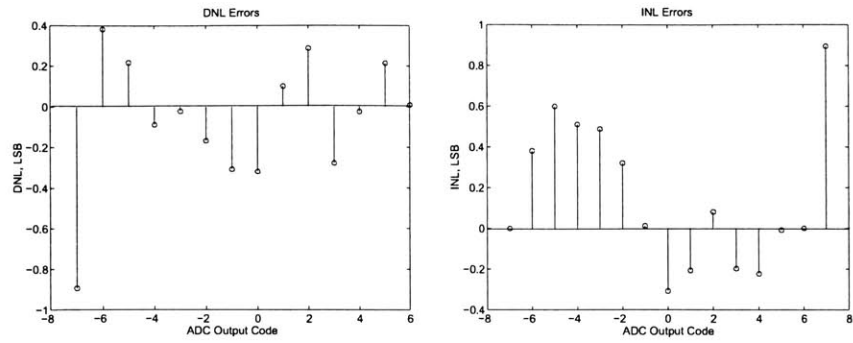
Static Performance

The procedure described above was applied to the ADC under test. It should be mentioned, however, that direct access to the ADC input was not available. The input interface comprises a low-noise amplifier (LNA) and two unity-gain buffers per channel, as described in Chapter 3. An extra *replica* buffer, identical to the ones driving the ADC, was used to route the LNA output to a package pin for external probing. Signal from the latter was used to infer the amplitude of the ADC input and ensure that it matches the full-scale voltage. At low frequencies, this inference can be safely made since the buffers are unity-gain.

A 50 MHz sine-wave input was employed. Since this frequency is well below the designed bandwidth, this set of tests essentially capture the ADC's *static* performance. A clock frequency of 385 MHz was used, which corresponds to an overall sampling rate of 1.54 GSPS. Functionality at the designed clock speed of 1 GHz could not be verified due to a problem with the on-chip test interface. Its suspected cause will be addressed later in this chapter.

Following the procedure described earlier, the histograms obtained for each channel were used to extrapolate transfer curves, based on which offset voltages, DNL and INL errors were computed. Plots of the latter are shown for each of the 4 ADC channels in Fig. 5-3 through Fig. 5-6. In arriving at these plots, 64K samples were acquired and processed per channel instead of the number derived earlier (16K). This was done to account for some of the non-idealities in the test setup, such as potential offsets in the buffers and non-linearity in the LNA. These non-idealities will be discussed shortly.

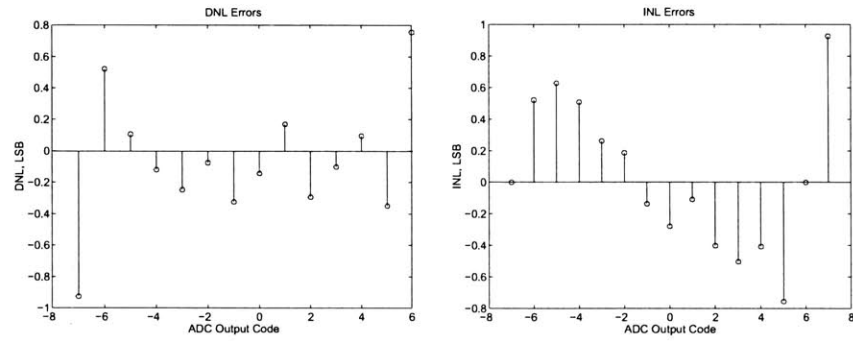
Maximum INL values and offset voltages are shown in Table 5.1 together with the



(a) DNL Plot

(b) INL Plot

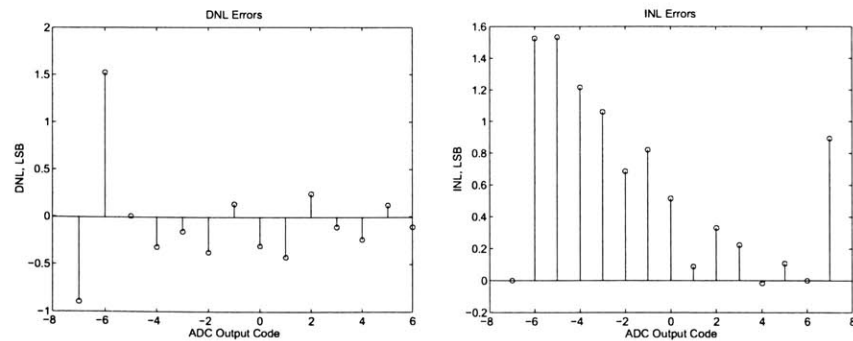
Figure 5-3: Channel 1 Performance



(a) DNL Plot

(b) INL Plot

Figure 5-4: Channel 2 Performance



(a) DNL Plot

(b) INL Plot

Figure 5-5: Channel 3 Performance

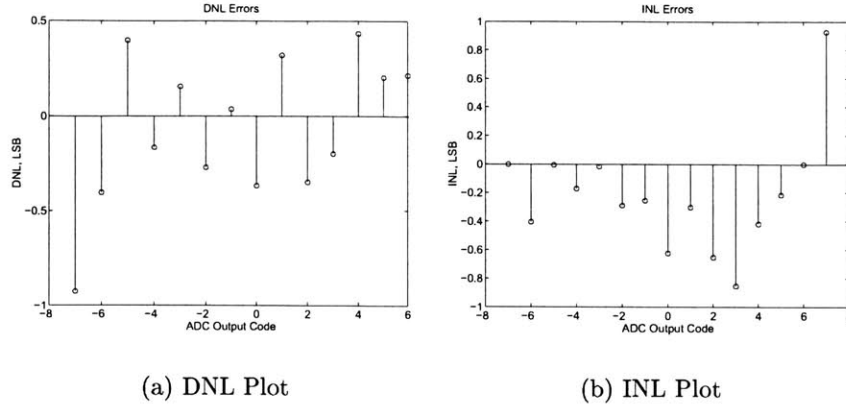


Figure 5-6: Channel 4 Performance

relative and absolute accuracies computed using Eq. 5.4 and Eq. 5.3 respectively.

Table 5.1: Accuracy of Channels

	INL_{max}	Relative Accuracy	$Offset_{max}$	Absolute Accuracy
Channel 1	0.9 LSB	4 bits	1.3 LSB	3.6 bits
Channel 2	0.9 LSB	4 bits	1.5 LSB	3.4 bits
Channel 3	1.5 LSB	4 bits	2.1 LSB	2.9 bits
Channel 4	0.9 LSB	3.4 bits	1.3 LSB	3.6 bits
Overall	1.1 LSB	3.9 bits	1.6 LSB	3.3 bits

Based on these results and an analysis of the test conditions, the following conclusions can be made :

- Since the ADC input is fed indirectly, mismatches and non-linearity in the LNA and buffers are embedded within the measures of absolute accuracy above. In fact, the output of the replica buffer reveals third-harmonic distortion, 20 dB below the fundamental sine-wave applied. Since these blocks are integral components of the wireless receiver being designed, a holistic measure of front-end distortion certainly has value. The relative accuracy measures, meanwhile, present a picture of the ADC's own linearity which was the focus of this design effort. Three out of 4 channels achieve 4 bits of relative accuracy.
- Channel 3 has considerably worse performance than the rest. There are

two possible reasons for this. One is large offsets and missing codes in the ADC channel itself, perhaps due to larger stress gradients exerted on that part of the die. Alternatively, the offset may have been introduced by mismatches within the buffer pair P_i and M_i driving this channel or in the buffer pair between the T/H and PCC bank. This offset error was perhaps too large to be cancelled out during DNL and INL computation, which might explain why the relative accuracy of Channel 3 is markedly lower also. However, since input was not directly applied to the ADC, there is no way of ascertaining which is the dominant cause of Channel 3's poor performance.

- In most applications, mismatches between channels in a time-interleaved ADC pose problems and often need to be calibrated out. If one channel has considerably worse non-linearity than the rest, tones appear in the output spectrum since one in every 4 samples has poor resolution. However, in this application, the 4 samples generated by the ADC in 1 clock period are *added* together in the multiply-accumulate units of the correlator and subsequently treated as a single monolithic sample. Thus, errors across the channels are averaged. The effective ADC resolution is given by the average performance of the channels instead of being limited by the worst-case. This is an important distinction, rooted in the target application for which the ADC was designed. Therefore, the 3.9 bits of relative accuracy and 3.3 bits of absolute accuracy are indeed representative of the performance achieved by the ADC.

Dynamic Testing

In order to characterize the ADC's *dynamic* behaviour, the same methodology must be applied at higher input frequencies. However, indirect access to the ADC input poses a major problem for reliable dynamic testing, namely knowing the exact amplitude of the signal fed to the ADC. The validity of the test methodology described is *contingent* upon the amplitude of the input signal matching the full-scale voltage of the ADC. At low frequencies, this can be guaranteed since the buffers driving the ADC are operating in the dc portion of their transfer characteristic and thus, have unity gain. At higher frequencies, however, the situation is different for two reasons :

- Bandwidth roll-off can occur throughout the signal path; at the package

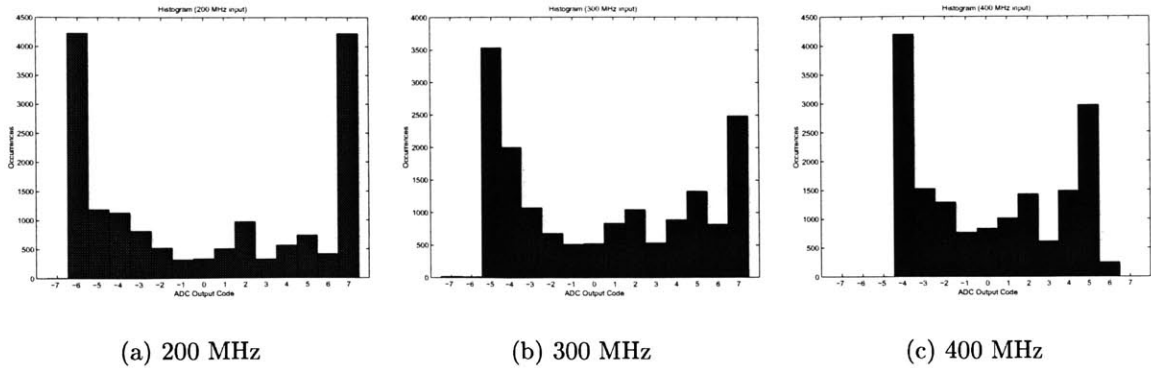


Figure 5-7: Dynamic Performance

interface, and in the LNA and buffers. In fact, this suspicion was borne out by LNA testing. Bandwidth of this block appears to be on the order of 500 MHz.

- Mismatches between the two buffers P_i and M_i driving each channel and in the pair between the T/H and PCC bank lead to an asymmetry in the input signal applied to the PCC bank, since positive and negative parts of the signal end up being amplified by different amounts. This asymmetry worsens as input frequency increases since differences in path lengths and the environment surrounding each buffer become important.

Therefore, there is no way to ensure that the signal fed to the ADC has amplitude equal to its full-scale voltage or that it has zero mean. Consequently, only a qualitative assessment of dynamic performance is possible for this chip.

Fig. 5-7 shows histograms of channel 1 of the ADC for three input frequencies, 200 MHz and 300 MHz and 400 MHz. The following observations and conclusions can be made about the ADC's dynamic performance :

- The range of codes covered gets smaller with input frequency. However, all histograms have the same general shape, resembling the bimodal distribution of a sine-wave. Certain detailed features match as well, like the high bumps at code values 1 and 2 that imply large differential non-linearity at these positions. It appears, therefore, that the shrinking span of codes is primarily due to bandwidth limitations that attenuate the signal being fed to the preamplifiers, pushing it below the ADC's

full-scale voltage as frequency goes up. The number of missing codes increases from 1 to 5 in going from 200 MHz up to 400 MHz. This corresponds to an attenuation of 25% or about 2.5 dB. As mentioned earlier, there is no way to ascertain exactly how much of this bandwidth limitation comes from the ADC front-end itself (T/H and buffers driving the preamplifier). In other words, the bandwidth limitations of the package interface, LNA and pre-ADC buffers are embedded within this measure.

- The lowest code in the 200 MHz histogram is missing. As frequency increases, more missing codes appear at the low end of the code range than at the high end. There are two possible sources of these dynamic offsets.

(a) Mismatches in path length and the environment surrounding the buffers P_i and M_i driving $vinp$ and $vinm$ that result in the differential signal being fed to the ADC having non-zero mean. Similar mismatches could also occur in the buffers following the T/H that drive the bank of preamplifiers.

(b) Path length variation across the bank of 15 PCC units in each channel. These units snake around as shown in Fig. 4-6. Input from the T/H, meanwhile, comes from the top. Consequently, PCC units at the top have smaller path lengths and smaller associated RC delays. The use of thin wires exacerbates these differences. This hypothesis is reinforced by the simulations carried out and reported at the end of Chapter 4.

- Although a quantitative measure of ADC performance at these high input frequencies cannot be obtained as explained, it is apparent that the converter supports at least 2 effective bits of resolution across this frequency range. This contention is supported by all the histograms in Fig. 5-7 having the correct shape, with the presence of two peaks at the extremities, and a general downswing followed by an upswing in going from one end of the code range to the other. It should be noted that some loss in resolution is expected given the attenuation of the signal below full-scale.

High Clock Speeds

As mentioned before, the chip could only be tested at 40% of the designed clock speed of 1 GHz. The suspected reasons for this are briefly outlined here. Although the PLL

is able to generate oscillations at 1 GHz as verified by a spectrum analyzer probing its power supply pin, the test interface that decimates the ADC outputs by a factor of 32 does not appear to work at this speed.

Recall that this block contains a bank of 17 flip-flops that is enabled once every 32 cycles, as described in Chapter 3. Thus, it produces the 16 decimated outputs of the ADC and a divided-down clock, referred to as CLK_{ref} . The latter is a useful diagnostic signal. For frequencies up to 400 MHz, CLK_{ref} oscillates at $f_{CLK}/32$. Beyond this frequency, however, it stays on one of the rails, glitching occasionally. The clock input to this block appears to be fine, however. Probing the PLL power supply pin shows a clean tone at the expected frequency.

The point of failure in the test interface is believed to be the 5-bit counter that generates the enable signal. This circuit was synthesized instead of being laid out by hand (largely due to time constraints prior to the tape-out) and is the weak link in terms of timing robustness. Although it was simulated to have a delay of around 0.5 ns at the slow corner, this number may be higher in practice considering the power supply voltage it actually sees is presumably less than the applied 1.8V due to IR drops, ringing and noise coupling on to the supply line. If the delay approaches a clock period, it can lead to setup-time violations in the bank of flip-flops. If it slightly exceeds a clock period, hold-time errors can occur. In fact, if noise and ringing on the power supply line are large, both types of timing violations may occur alternately over the course of successive clock cycles. The hypothesis, therefore, is that the counter fails as the clock period gets smaller because its delay becomes too large in proportion to the latter, causing timing violations in the flip-flops that follow. This hypothesis is supported by the fact that the minimum frequency of failure can be raised slightly by increasing the supply voltage.

Power Consumption

The ADC draws 241 mW of power from a 1.8V supply at a sampling rate of 1.54 GSPS (ie. clock frequency of 385 MHz). Table 5.2 shows the breakdown across the various functional blocks.

Table 5.2: ADC Power Consumption

ADC_{analog}	135 mW
$ADC_{digital}$	11 mW
PLL_{analog}	51 mW
$PLL_{digital}$	44 mW
Total	241 mW

5.3 Proposed Improvements

Some changes are proposed below to improve the testability and performance of future designs based on this architecture.

5.3.1 Ensuring Full Testability

The two main limitations that precluded full testing of the ADC can be circumvented as follows :

- Custom layout of the 5-bit counter should ensure that its propagation delay is well under 1 ns. For further protection, a negative-edge triggered register may be inserted between the counter and the bank of flip-flops that it feeds. The enable signal is thus synchronized to half a clock cycle before the rising edge on which this bank of flip-flops is activated. Even if the propagation delay of the counter is larger than anticipated, failure will occur only at specific frequencies and not consistently above a certain threshold. This is because with a large counter delay, the enable signal will now simply go high on the *next* cycle instead of getting dangerously close to the rising clock edge and causing timing violations.

- Provision of a mux that allows a choice between routing input through the LNA and doing so externally. Although only low-frequency inputs may be applied through a package pin with an amplitude of 500mV, this at least ensures more precise characterization of the ADC's static performance. Thoroughly testing dynamic performance requires an LNA and buffers with better linearity, improved modeling of package parasitics and better matching to transmission lines at the RF input pin.

5.3.2 Improving Dynamic Performance

The bandwidth limitation of the connections between the T/H and preamplifiers can be alleviated by making those wires thicker. As for the frequency-dependent asymmetry in the ADC's transfer characteristic, this can be fixed as follows :

- (1) Wherever a pair of single-ended unity-gain buffers are used to separately drive positive and negative signal paths, one fully-differential buffer should be used instead. Routing for the positive and negative signal paths should be closely matched.
- (2) The elimination of one comparator per PCC unit should be possible using faster process technologies as discussed in the next chapter. This will make the entire bank of 15 such units smaller, resulting in less path length variation across them.

Chapter 6

Conclusion

Ultra-wideband radio is a nascent technology with a number of compelling applications like high data-rate wireless communication and precision locationing. Fully-digital UWB receivers offer numerous benefits over their analog counterparts, including low cost and the flexibility to support multiple modulation schemes and bit rates and vary these parameters dynamically. Their feasibility, however, hinges upon the ability to do analog-digital conversion near the antenna at several gigasamples/sec. This is possible only if the desired resolution is low because of a fundamental trade-off between speed and precision.

In this thesis, detailed analysis of the impact of quantization noise in UWB receivers is presented that reveals the sufficiency of 4 bits of resolution for reliable detection. This result stems from the fact that a UWB system is power-limited and relies on its large processing gain to extract a signal buried in noise. That extraction takes place in the correlators after A/D conversion, and until that point in the receiver, the signal is immersed in additive white gaussian noise (AWGN) and interference from other radios. Quantization noise contributed by the ADC, therefore, is relatively small and has minimal impact on performance even at resolutions as low as 3 and 4 bits.

Data conversion at several gigasamples/sec (GSPS) is nevertheless challenging, even at low resolutions. The feasibility of this problem was investigated through the design and implementation of a 4-bit, 4 GSPS ADC. A standard CMOS process

(0.18 μm TSMC) was chosen as the technology platform for this design effort since the target application is a digital radio. A time-interleaved architecture was chosen, with 4 FLASH channels each running at 1 GHz using offset clocks, yielding an effective sampling rate of 4 GSPS. Key attributes in the design of the analog section include the use of fully-differential preamplifiers for protection against kickback and power supply noise, two latching comparators for good metastability resolution and a T/H preceding each ADC to reduce dynamic offsets. The desired resolution is achieved through proper device sizing in the preamplifier and comparators. The digital section features an intermediate Gray code for protection against bubble errors. In order to maintain a 1 GHz throughput, the multiple levels of logic in the decoder are pipelined. A re-timing block is used to synchronize outputs from the 4 ADC channels to the same clock phase. The design is fully functional at 4 GSPS, supports an input bandwidth of over 1 GHz and comfortably meets the offset requirements for 4 bits of resolution.

Layout of such a fast and large system was challenging and entailed making trade-offs between competing requirements like matching, compaction and isolation. Close attention had to be paid to the layout of the preamplifiers and comparators. Another performance-critical block that required care was the clock distribution network. Keeping clock lengths and loads balanced across the 4 phases is important to prevent large skews creeping in between them. Simulations of each ADC channel with full extracted parasitics verify functionality at 1 GHz clock frequency (4 GSPS sampling rate), but with sampling bandwidth reduced to around 400 MHz due to the use of long, thin wires between the T/H and preamplifiers.

The designed ADC was fabricated as part of an ultra-wideband front-end receiver chip. The data converter achieves 3.9 bits of relative accuracy and 3.3 bits of absolute accuracy at a sampling rate of 1.54 GSPS and with a 50 MHz input. Bandwidth roll-off was observed at input frequencies around 400 MHz, consistent with the results of extracted simulations. However, due to limitations of the test setup, dynamic performance could not be quantitatively assessed. Also, testing at clock frequencies higher than 385 MHz was not possible due to problems with the on-chip test interface. The suspected cause has been identified and a solution proposed for future designs.

It should be mentioned that the ADC's overall performance is given by the average of its 4 channels rather than being limited by the worst-case. This is a result of the post-processing applied to the samples from each channel. They are added together and subsequently treated as a single sample which *averages* the errors in the 4 channels. This application-specific feature greatly reduces the need for calibration across the channels as is required in most time-interleaved ADC's.

Future Work

The chip draws 241 mW from a 1.8V supply to support the 1.5 GSPS sampling rate. It was not explicitly designed for low-power consumption, but this will clearly be a priority for digital UWB radios to be practical and competitive with other architectures. Future research in the area of high-speed data conversion for UWB is likely to be focused largely on reducing power consumption. To this end, some circuit and system level approaches are proposed for designs based on this architecture.

Fewer Comparators

Two comparators were used in this design. The first is a track-and-latch stage that is fast on account of its limited output swing. However, it consumes roughly the same static power as the preamplifier and eliminating it would provide significant savings. The primary reason why this was not an option in the current chip is metastability resolution. The StrongArm latch on its own was found lacking in its ability to handle very small inputs. In future generations of process technology, however, this limitation will likely disappear. As explained in Chapter 3, the latch-mode time constant τ is inversely related to f_T and thus scales with shrinking gate lengths. This in turn implies that smaller inputs can be amplified rail-to-rail in the same amount of time or less.

Duty Cycling

UWB is a pulsed system with low duty-cycle signalling. This distinctive feature of UWB should be exploited by shutting down the receiver between pulses. As described in Chapter 2, the correlator essentially integrates the incoming pulse train over narrow windows centered about each pulse. Since the position of the pulses is not known

a priori, the receiver first needs to establish synchronization with the transmitter. During this period, all samples coming out of the ADC must be processed[1]. Once coarse acquisition has been achieved, however, there is no use for samples corresponding to the region between pulses. The ADC can be shut down during these *dead-zone* periods, and thus only kept active during narrow windows.

A limitation on how small the windows can be made, however, is the latency of the ADC. Its decoding and re-timing logic is pipelined and there is a delay of 10 clock cycles between the sampling of data and the availability of its corresponding digital output code from the ADC. It must be kept on for at least this duration, therefore, beyond the expected position of the pulse. This limitation can be alleviated in two ways. Firstly, as gate delays get smaller with improving process technology, fewer pipeline stages may be used thus reducing the overall latency. Secondly, the analog section may be shut down while the decoding logic is processing the thermometer code. In the current chip, analog power is dominant and there is large power saving associated with this semi-shutdown state. This concept can also be applied over a finer grain. Pipeline stages can be sequentially powered down as the last sample goes through.

Dynamic Scaling of ADC Resolution

It has been shown in Chapter 2 that for a certain range of signal-to-interference values (SIR), 2 bits *improves* performance over 3 and 4 bits. If the backend DSP can be designed to detect such a condition, the ADC resolution can be scaled down to squeeze out a few extra dB of performance. The simplest way of scaling down resolution is simply to throw away LSB's, but a clever scheme that shuts down certain comparator banks can be devised to reduce power consumption. If, on the other hand, performance requirements fall and a lower data rate is acceptable, a single comparator can be used for 1-bit A/D conversion. There is an exponential relationship between power consumption and resolution for a FLASH ADC, so the potential power saving would be huge. Dynamic ADC scaling would allow for operation at the right power/performance point, making it a key feature of a reconfigurable UWB software radio.

This work demonstrates that the problem of high-speed data conversion for digital ultra-wideband receivers is tractable in CMOS within a reasonable power budget. It remains an interesting area of research, nevertheless, with further technology scaling and aggressive power targets posing some major challenges.

Bibliography

- [1] R. Blazquez. "Design of a Synchronization Subsystem for an Ultra-Wideband Radio". *S.M. Thesis, Massachusetts Institute of Technology*, 2003.
- [2] Chandrakasan, A.P., Bowhill, W.J. and Fox, F., editor. *Design of High-Performance Microprocessor Circuits*. IEEE Press, 2001.
- [3] Choi, M. and Abidi, A.A. "A 6-b 1.3-Gsample/s A/D Converter in 0.35 μ m CMOS". *IEEE Journal of Solid-State Circuits*, 36(12):1847–1858, Dec 2001.
- [4] Federal Communications Commission. "First Report and Order, In the matter of - Revision of Part 15 of the Commission's Rules Regarding Ultra-Wideband Transmission Systems". ET Docket 98-153, April 2002.
- [5] D. Dickson and P. Jett. "An Application-Specific Integrated Circuit Implementation of a Multiple Correlator for UWB Radio Applications". *Military Communications Conference Proceedings (MILCOM), IEEE*, 2:1207–1210, 1999.
- [6] Dornberg, J., Lee, H. S. and Hodges, D. A. "Full-Speed Testing of A/D Converters". *IEEE Journal of Solid-State Circuits*, SC-19(6):22–26, Dec 1984.
- [7] W. Ellersick. "Data Converters for High Speed CMOS Links". *PhD Dissertation, Stanford University*, August 2001.
- [8] Ellersick, W., Chih-Kong Ken Yang, Horowitz, W. and Dally, W. "GAD: A 12Gs/s CMOS 4-bit A/D converter for an equalized multi-level link". *Symposium on VLSI Circuits, Digest of Technical Papers*, pages 49–52, 1999.

- [9] A. Hastings. *The Art of Analog Layout*. Prentice Hall, 2001.
- [10] Xtreme Spectrum Inc. "Trinity for High-Speed Wireless Applications". Product Brief, June 2002.
- [11] D. A. Johns and K. Martin. *Analog Integrated Circuit Design*. John Wiley and Sons, Inc., 1997.
- [12] C.J. Le Martret and G.B. Giannakis. "All-digital PAM impulse-radio for multiple-access through frequency-selective multipath". *Global Telecommunications Conference*, pages 77–81, 2000.
- [13] C.J. Le Martret and G.B. Giannakis. "All-digital PPM impulse radio for multiple-access through frequency-selective multipath". *Sensor Array and Multichannel Signal Processing Workshop*, pages 22–26, 2000.
- [14] T.H. Lee. *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge University Press, 1998.
- [15] J. G. Maneatis. "Low-Jitter Process-Independent DLL and PLL Based on Self-Biased Techniques". *Journal of Solid-State Circuits*, 31(11):1723–1732, 1996.
- [16] Newaskar, P., Blazquez, R. and Chandrakasan, A.P. "A/D Precision Requirements for Digital Ultra-Wideband Radio Receivers". *IEEE Workshop on Signal Processing Systems*, September 2002.
- [17] I. O'Donnell and R. Brodersen. "Imperceptible UWB". Poster Session, Berkeley Wireless Research Center, Jan 2002.
- [18] A. Petroff and P. Withington. "PulsON Technology Overview (White Paper)". Time Domain Corporation, July 2001.
- [19] Poulton, K. et al. "An 8-GSample/s 8-bit ADC System". *Symposium on VLSI Circuits, Digest of Technical Papers*, 1997.
- [20] J.G. Proakis. *Digital Communications*. McGraw Hill, Inc., 1983.

- [21] Maxim Integrated Products. "INL/DNL Measurements for High-Speed Analog-to-Digital Converters (ADCs)". Application Note 283, Sep 2000.
- [22] J.M. Rabaey. *Digital Integrated Circuits*. Prentice Hall, 1996.
- [23] B. Razavi. *Principles of Data Conversion System Design*. IEEE Press, 1995.
- [24] B. Razavi. *RF Microelectronics*. Prentice Hall, 1998.
- [25] Salkintzis, A.K., Hong Nie, Mathiopoulos, P.T. "ADC and DSP challenges in the development of software radio base stations. *IEEE Personal Communications*, 6(4):47-55, August 1999.
- [26] K. Siwiak and A. Petroff. "A Path Link Model for Ultra-Wideband Pulse Transmissions". *Proceedings of the IEEE Semi-Annual Vehicular Technology Conference, Rhodes, Greece.*, 2001.