P_{FA1} = false alarm probability of the first dwell

P_{FA2} = false alarm probability of the second dwell

q = number of cells to be searched

N = number of chips to be searched

$\frac{\Delta T_c}{T_c}$ = step size of search

Δf_c = code doppler in chips

K = penalty (in terms of τ_{D2}) for a false alarm with the second dwell time

Note again that the algebraic sign of the code doppler Δf_c is assigned in such a way as to be positive when the search is speeded up by the code doppler. Note also that with $\tau_{D1} = 0$ and $P_{FA1} = 1$, the equations for the double dwell system reduces to those obtained for the single dwell system.

2.2.1.1.3 Detection and False Alarm Probability Approximations

The mean acquisition time and variance depend both on the probability of detection P_D and the probability of false alarm P_{FA} . In the ensuing paragraphs, expressions for the probabilities of detection and false alarm for a fixed dwell integration period, following a square law detector, with an input consisting of signal and/or band-limited Gaussian noise, will be described. Following this, the modifications that must be made to account for the effects of modulation and a nonconstant detection probability, will be discussed. Although more detailed and exact derivations exist (see references [14] and [20]), the analysis presented here will follow the more simple derivations presented in reference [15].

1

MicroD

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R S C 1970, c C-30

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c C-30

**DIRECT SEQUENCE SSMA COMMUNICATIONS WITH SOME IC
REALIZATIONS**

by

Donald C. Kemdirim, B.Eng.

A thesis submitted to the
Faculty of Graduate Studies and Research
in partial fulfilment of the requirements
for the degree of
Master of Engineering

Ottawa-Carleton Institute for Electrical Engineering
— Faculty of Engineering
Department of Electronics
Carleton University

August 1988. _____

©Donald C. Kemdirim, 1988

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-46272-8

The undersigned hereby recommend to
the Faculty of Graduate Studies and Research
acceptance of the thesis,
**DIRECT SEQUENCE SSMA COMMUNICATIONS WITH SOME IC
REALIZATIONS**

submitted by Donald C. Kemdirim, B.Eng.,
in partial fulfilment of the requirements for the degree of
Master of Engineering

Thesis Supervisor

Chairman,
Department of Electronics

Faculty of Engineering
Department of Electronics
Carleton University
October, 1988

ABSTRACT

A direct sequence spread spectrum multiple access system has been discussed, built and tested. A spread spectrum multiple access system is not only useful in military communications but also has varied commercial applications. The attractiveness of making such systems portable is the motivation for the integrated circuit implementations of a few parts carried out here. Essentially, two transmitters were built, one with voice capability. Gold codes were used for the spreading sequence and PN acquisition was accomplished using a serial search technique. PN tracking was achieved with a tau-dither loop. A squaring loop and a differential encoding scheme (for carrier phase ambiguity) was employed in the demodulation process, which was followed by conversion to a voice output when desired. The system was found to perform quite well under both single and multiple access conditions. As a demonstration of the feasibility of implementing the system on chips, a few small blocks were put on ICs (using CMOS technology); these performed as expected and clearly demonstrated both the feasibility and attractiveness of a VLSI implementation of the system.

ACKNOWLEDGEMENTS

The author would like to express his deep and heartfelt appreciation for the patience, support and guidance provided by his thesis supervisor, Professor J.S. Wight. Without his unflagging encouragement and dedication, the completion of this thesis would have been very difficult.

The author would also like to thank Nagui Mikhail for his much valued technical assistance. The many hours he spent helping me assemble and debug the circuits will never be forgotten.

The assistance of Professor C.H. Chan in implementing the ICs through the Canadian Microelectronics Corporation is much appreciated. The assistance of Graham Bell and Mike Gagnon, who made life on the Metheus bearable, is also gratefully acknowledged. I would also like to thank Mike Kelly for some of his technical suggestions, Guy Asselin for facilitating the acquisition of components and the rest of the Department of Electronics staff, for their goodwill.

I would like to thank my friends, Georgina Wainwright and Real Perriard who never doubted that I could finish this thesis.

Finally and most importantly, I would like to thank my parents. Without their encouragement, support (emotional and financial), inspiration and faith in me, the completion of this endeavor would have been impossible.

Contents

ABSTRACT	iii
ACKNOWLEDGEMENTS	iv
TABLE OF CONTENTS	v
LIST OF FIGURES	viii
LIST OF TABLES	xvi
1 INTRODUCTION	1
1.1 INTRODUCTION	1
1.2 MOBILE COMMUNICATIONS	2
1.3 DS SSMA FOR GAUSSIAN CHANNELS	4
1.4 VLSI CMOS CIRCUITS	5
2 SPREAD SPECTRUM SYSTEMS THEORY	6
2.1 SPREADING TECHNIQUES	6
2.1.1 Direct Sequence	7
2.1.2 Direct Sequence (PN) Power Spectral Density	12
2.1.3 Frequency Hopping	17
2.1.4 Other Spreading Techniques	20
2.2 SYNCHRONIZATION	21
2.2.1 PN Code Acquisition	23
2.2.2 PN Code Tracking	45

2.2.3	Carrier Acquisition and Tracking	76
3	SYSTEM CONSIDERATIONS AND DESIGN	93
3.1	PERFORMANCE CONSIDERATIONS FOR A DS SSMA SYSTEM	93
3.2	SYSTEM PARAMETERS	100
3.3	THE TRANSMITTER	104
3.3.1	System Description	104
3.3.2	Clock and Encoding Circuits	106
3.3.3	Modifications for the Second Transmitter	108
3.4	THE RECEIVER	110
3.4.1	The Receiver Gold Code PN Generator	113
3.4.2	Input and Reference Bandpass Filters	113
3.4.3	The Acquisition Channel	115
3.4.4	The Tracking Channel	130
3.4.5	Demodulation Circuits	141
4	VLSI IMPLEMENTATION	155
4.1	IMPLEMENTATION PROCEDURE AND OBJECTIVES	156
4.1.1	The PN Gold Code Generator	157
4.1.2	The Searchlock Logic Unit	161
5	EXPERIMENTAL RESULTS	165
5.1	Transmitters and Signal Transmission	165
5.2	Receiver System	180
5.2.1	Receiver Gold Code Generator	180
5.2.2	Acquisition Channel Results	180
5.2.3	Tracking Channel Results	191
5.2.4	Demodulation Channel Results	198
5.2.5	IC Results	210

6 CONCLUSIONS	228
6.1 Summary	228
6.2 Future Work	229
REFERENCES	231
APPENDIX A	236
A-1.1 PN Gold Code Generator	236
A-1.2 Gold Code Generator Initial Contents	245
A-1.3 Continuously Variable Slope Delta (CVSD) Modulator	246
A-2.1 Theoretical Analysis	260
A-2.1.1 MOS Transistor Fundamentals	260
A-2.1.2 Basic MOS Concepts	269
A-2.2 CMOS Processing	294
A-2.2.1 Silicon Semiconductor Processing	295
A-2.2.2 CMOS1B 5 μ Process	297
A-2.2.3 Layout Design Rules	304

List of Figures

1.1	Multipath Fading in a Mobile Communication Environment, [3].	3
2.1	BPSK Direct Sequence Modulation.	8
2.2	Mathematically Convenient Representation of Figure 2.1. [14] . . .	9
2.3	(a) BPSK Direct Sequence Spread Spectrum Receiver; (b) BPSK DS Spreading and Despreading. [14]	11
2.4	Power Spectral Densities: (a) of data modulated carrier; (b) of data and spreading code modulated carrier.	13
2.5	Autocorrelation Function for PN Waveform. [15]	14
2.6	Power Spectral Density of a PN Waveform. [14]	16
2.7	Spectral Density of Data and Code Modulated Carrier for (a) $T_s > 2NT_c$; (b) $T_s \leq NT_c$. [13]	18
2.8	Frequency Hop Spread Spectrum System: (a) Transmitter; (b) Receiver.	19
2.9	Time-Frequency Relationship for a Chirp Spread Spectrum System. [13]	22
2.10	Block Diagram of a Single Dwell PN Acquisition System. [17] . . .	26
2.11	Generating Function Flow Graph. [15]	29
2.12	Double Dwell System. [15]	33
2.13	Algorithm for the Double Dwell System. [15]	34

2.14	Model for the Detection and False Alarm Probability Approximations. [15]	38
2.15	The Sequential Detector. [17]	46
2.16	Integrator Output: (a) Bias voltage absent; (b) Bias voltage present. [17]	47
2.17	Non-Coherent Delay-Lock Loop. [17]	49
2.18	Non-Coherent Tau-Dither Loop. [17]	50
2.19	Equivalent Loop Model for the Tau-Dither Loop. [17]	52
2.20	Autocorrelation Functions for: (a) Advanced PN code; (b) Retarded PN code. [27]	55
2.21	Discriminator Characteristic (S-curve).	61
2.22	Autocorrelation and Cross-Correlation Functions for $q_1(t)$ and $q_2(t)$. [17]	64
2.23	Acquisition Trajectories for a "One-delta" Noncoherent Loop: (a) Loop gain $g = \infty$, (b) Loop gain $g = 10$. [15]	74
2.24	Transient Response of the Normalized Timing Error $x(\omega_n t)$ for a "One-delta" Noncoherent Loop with $g = \infty$. [15]	75
2.25	Normalized Mean Slip Time vs Jitter. [15]	77
2.26	Basic Diagram of a Phase-Locked Loop	79
2.27	Loop Filter for a Second Order Loop with a Transfer Function Given by $F(j\omega) = 1 + j\omega\tau_2/1 + j\omega\tau_1$	85
2.28	Sinusoidal Phase Detector Characteristic. [25]	86
2.29	Squaring Loop. [25]	90
3.1	Phase Coded (BPSK) SSMA System Model. [26]	94
3.2	Bounds on the Probability for the Set of Gold Sequences of Degree 13 with Partial Period Correlation of 60 Chips. [27]	97
3.3	Estimated Relative-Signal Levels at Different Points in the System.	102
3.4	Block Diagram of a Voice Input DS Spread Spectrum Transmitter.	105

3.5	DS SSMA Transmitter Schematic (Voice Input)	107
3.6	Encoding and Decoding Blocks for Data Ambiguity Resolution: (a) encoding circuit, (b) decoding circuit. [30]	109
3.7	Second Transmitter Schematic (Data Input).	111
3.8	Block Diagram of Receiver System.	112
3.9	Receiver Gold Code Generator.	114
3.10	Reference and Input Signal Bandpass Filters: (a) Reference filter, (b) Input filter.	116
3.11	Actual Component Values for the Reference and Input Filters.	117
3.12	Acquisition Channel Schematic.	118
3.13	Acquisition Channel Envelope Detector Characteristic.	122
3.14	Integrate and Dump Filter:	123
3.15	Block Diagram of Searchlock Unit.	125
3.16	Schematic of Searchlock Logic Unit.	126
3.17	Algorithm Implemented by the Searchlock Unit.	127
3.18	Searchlock Unit in Terms of a State Diagram.	128
3.19	Derivation of the Clocks Used for the Searchlock Unit.	131
3.20	Tracking Channel Schematic.	132
3.21	Tracking Channel Envelope Detector Characteristic.	135
3.22	Dithering Logic Schematic	136
3.23	S-curve for the Tracking Channel.	137
3.24	Dithering Logic (With Early and Late Inputs Reversed).	138
3.25	Component Values for the Loop Filters.	143
3.26	Receiver Code Generator Schematic.	144
3.27	Receiver Code Generator VCO Characteristic.	145
3.28	Demodulation Channel Schematic.	146
3.29	Loop Filter for Squaring Loop; $F(j\omega) = 1 + j\omega\tau_2 / 1 + j\omega\tau_1$	151

3.30	BPSK Demodulation: (a) Synchronous bit detection; (b) Asynchronous bit detection	153
4.1	13-Stage PN Gold Code Generator Implemented on a Chip.	158
4.2	Gold Code Generator Sample Simulation.	160
4.3	Searchlock Logic Unit Implemented on a Chip.	162
4.4	Searchlock Logic Unit Sample Simulation.	163
5.1	Spread Spectrum Multiple Access System.	166
5.2	Voice Input DS SSMA Transmitter.	167
5.3	Data Input SSMA Transmitter.	168
5.4	Time Waveforms for the Transmitter 13-Stage Gold Code Generator.	169
5.5	Baseband Code Spectra for the 13-Stage Transmitter Gold Code Generator.	171
5.6	Transmitted Signal Spectrum.	172
5.7	Balanced Mixer Output With Both Carrier and Code Imbalance. [2]	173
5.8	System Configuration Using 880 MHz Carrier.	174
5.9	System Configuration Using 70.6 MHz Carrier.	175
5.10	Detailed 70.6 MHz Carrier System Configuration for One User.	176
5.11	Detailed 70.6 MHz System Configuration for Two Users.	177
5.12	Received Noise and Signal Traces: Noise - Top Trace; Signal - Bottom Trace ($SNR_i \approx -8.4dB$).	178
5.13	Received Signal and Noise Traces: Signal Plus Noise - Top Trace ($SNR_i \approx -8.4dB$); Signal Only - Bottom Trace.	179
5.14	Reference and RF Input Filter Responses: (a) Reference filter response; (b) RF input filter response.	181
5.15	Receiver System.	182
5.16	Circuit Board Containing the Receiver Gold Code Generator and the Generator Outputs: (a) Gold code generator implementation; (b) Clock input and the early, punctual and late outputs.	183

5.17	Acquisition Channel Circuit Board.	184
5.18	Acquisition Channel Filter Responses: (a) IF bandpass filter response; (b) Lowpass filter response.	185
5.19	Correlation Triangle: y axis: 50 mv/div; x axis: 0.5 ms/div; ($SNR_c = -0.6$ dB).	186
5.20	Searchlock Logic Unit Pulses: y axis: 5V/div; x axis: 0.1 ms/div; F_{A1} : Top trace; F_{B1} : Second trace; F : Third trace; F_{B2} : Bottom trace.	188
5.21	Acquisition Time Measurement.	190
5.22	Tracking Channel Circuit Board.	192
5.23	Measured S-curve From the Tracking Channel: y axis: 50 mv/div; x axis: 0.2 ms/div.	195
5.24	Loop Filter Switch in Times: P#00: Acquisition occurs; loop filter 1 applied; P#01: Loop filter 2 applied; P#02: Loop filter 3 applied; P#03: Loop filter 4 applied.	197
5.25	Demodulation Channel Circuit Boards: (a) Carrier recovery board; (b) Circuit board used for data recovery and conversion to analog voice output.	199
5.26	Filter Responses for the Demodulation Channel: (a) Input bandpass filter response; (b) Response for the narrowband filter following the squaring operation.	200
5.27	Despread Signals: (a) Despread unmodulated carrier; (b) Despread data modulated carrier.	201
5.28	Examples of the Traces Used in Obtaining the Signal and Noise Values of Table 5.3.	203
5.29	Effect of Noise on System Performance: (a) Output with 9.8 dB less noise input than in case b; (b) Noise input is 9.8 dB greater than in case a.	204
5.30	Output from the Data Recovery Section When a 1:1 Bit Pattern is Transmitted: (a) Output following the information band low pass filter; (b) Output after decoding has taken place.	206
5.31	BER Results for Single User Transmissions. (Note that $m = 10^{-3}$, $\mu = 10^{-6}$, $n = 10^{-9}$ and $p = 10^{-12}$ on the y axis).	208

5.32	BER Plot for the Case of Two Users. (Note that $m = 10^{-3}$ and $\mu = 10^{-6}$ on the y axis).	211
5.33	Picture of the Fabricated ICs.	212
5.34	Bonding Diagram for the Gold Code Generator.	213
5.35	Bonding Diagram for the Searchlock Logic Unit.	214
5.36	Time and Spectral Outputs for the IC Gold Code Generator: (a) Time waveforms; (b) Baseband code spectra.	215
5.37	Baseband Code Spectra for IC Gold Code Generator as Code Rate is Increased.	217
5.38	Picture Illustrating the defect in the "Late" Output of the IC Gold Code Generator.	219
5.39	Output Stage Added to Correct Defect Seen in Figure 5.38.	220
5.40	Early, Punctual and Late Outputs for the IC Code Generator with the Output Stage of Figure 5.39.	221
5.41	BER For Single User Obtained When the IC Gold Code Generator was Used. (Note that $m = 10^{-3}$, $\mu = 10^{-6}$ and $n = 10^{-9}$ on the y axis).	224
5.42	Algorithm Realized By The IC Searchlock Unit.	225
5.43	An Example of the Results Obtained From the IC Searchlock Unit: (a) Result with $V_{id} = 0$; (b) Result with $V_{id} = 1$.	227
A-1.1	n -Stage Shift Register With Linear Feedback	237
A-1.2	Gold Code Generator Arrangement. [15]	239
A-1.3	Balanced Gold Code of Length $2^{13}-1$.	243
A-1.4	HCMOS Realization of 13 Stage Gold Code Generator	244
A-1.5	Eight Level Quantization of Half a Cycle of a One Volt Peak Sinusoid. [40]	247
A-1.6	Delta Modulator. [42]	249
A-1.7	Delta Modulator Waveforms: (a) Modulator output; (b) Input waveform and staircase approximation. [42]	250
A-1.8	Delta Modulator Receiver	251

A-1.9 A General Block Diagram of a CVSD Encoder. [41]	253
A-1.10 CVSD Decoder. [41]	255
A-1.11 CVSD Waveforms. [41]	256
A-1.12 HC-5512C CVSD Monolithic Filter: (a) Functional diagram, (b) Performance characteristic. [44]	257
A-1.13 MC3418 CVSD Block Diagram. [43]	258
A-1.14 CVSD Configuration	259
A-2.1 MOS Transistor Structures: (a) NMOS transistor; (b) PMOS transistor. [45]	261
A-2.2 Relative Conduction of Enhancement Mode Transistors Versus the Gate to Source Voltage (V_{ds} fixed): (a) n-channel; (b) p-channel. [45]	263
A-2.3 Relative Conduction of Depletion Mode Transistors Versus the Gate to Source Voltage (V_{ds} fixed): (a) n-channel; (b) p-channel. [45]	264
A-2.4 Physical Structure of an n-Channel Enhancement Type Transistor. [45]	265
A-2.5 Creation of the Inversion Layer in an NMOS Transistor. [45]	267
A-2.6 NMOS Transistor Behaviour With Different Terminal Voltages. [45]	268
A-2.7 Physical Structure of a p-Channel Enhancement Type Transistor. [45]	270
A-2.8 V-I Characteristic for the n and p Enhancement Transistors. [45]	272
A-2.9 MOS Transistor Geometric Terms. [45]	274
A-2.10 CMOS Inverter: (a) transistor representation; (b) schematic representation; (c) logic function	275
A-2.11 CMOS NAND Gate: (a) transistor representation; (b) schematic representation; (c) logic function	277
A-2.12 n-input NAND Gate. [45]	278
A-2.13 CMOS NOR Gate: (a) transistor representation; (b) schematic representation; (c) logic function	279
A-2.14 n-input NOR Gate. [45]	280

A-2.15 CMOS Inverter DC Transfer Characteristic. [45]	282
A-2.16 Equivalent Circuit for Inverter Operation in Region B. [45]	283
A-2.17 Effect of Varying β_n/β_p on the Inverter Transfer Characteristic. [45]	287
A-2.18 Noise Margin Definitions. [45]	289
A-2.19 Switching Characteristics for a CMOS Inverter: (a) an inverter driving a second inverter; (b) illustration of the rise time (t_r), fall time (t_f) and delay time (t_d). [45]	291
A-2.20 CMOS Logic States Under Static Conditions: (a) $V_{in} = "0"$, n-device is "OFF", and p-device is "ON"; (b) $V_{in} = "1"$, n-device is "ON" and p-device is "OFF". [45]	293
A-2.21 Packaging Sequence. [50]	296
A-2.22 Patterning the Silicon Dioxide: (a) Bare wafer; (b) Oxidation; (c) Coat with photo resist; (d) Mask and expose; (e) Exposed resist; (f) Develop resist; (g) Etch; (h) Remove resist. [50]	298
A-2.23 CMOS1B Processing Steps for Simple Inverter: (a) p-well definition; (b) Formation of Si_3N_4 layer; (c) Active area definition. [51]	300
A-2.23 Continued: (d) p-implant into p-well regions; (e) n-implant into n-well regions; (f) Field oxide growth over regions not covered by the nitride - SiO_2 sandwich	302
A-2.23 Continued: (g) Deposition of gate polysilicon; (h) n^+ and p^+ area definitions	303
A-2.24 Complete CMOS Inverter Structure. [51]	305
A-2.25 CMOS1B Pad Frames. [51]	306

List of Tables

2.1	Closed Form Results for M_2 and M_4 [17]	68
3.1	Acquisition Time Calculations	124
3.2	Switch in Times for the Loop Filters	142
5.1	Acquisition Time Measurements SNR_i (Input Signal to Noise Ratio) = -8.7 dB, Code Doppler (Δf_c) ≈ 0	191
5.2	Acquisition Time Measurements SNR_i (Input Signal to Noise Ratio) = -0.9 dB, Code Doppler (Δf_c) ≈ 0	193
5.3	Despread SNR and Processing Gain for Single User Transmission. Input Signal Power = -1.2 dBm	205
5.4	BERs for Single User Transmission	207
5.5	BER for Multiple Access Environment. (Number of Users =2). $P_N=2.8$ dBm, $P_D=4.8$ dBm	210
5.6	BER Obtained for Single User Transmissions When the IC Gold Code Generator Was Used	223
A-2.1	The Three Regions of Operation for the n- and p-Transistors. [45]	284
A-2.2	CMOS1B Pad Frame Sizes and Dimensions. [51]	307

Chapter 1

INTRODUCTION

1.1 INTRODUCTION

Spread spectrum communications systems have gained greater prominence over the last fifteen years as a result of the advancements in VLSI technology. A spread spectrum system can be defined as a system in which a transmitted signal "occupies a bandwidth much in excess of the minimum necessary to send the information" [1]. The bandspreading is accomplished by using a code that is independent of the data. At the receiver, an identical code, synchronized with the received code, is required to despread and recover the information.¹

At first glance, it may seem illogical to use a technique that occupies more bandwidth than is necessary to transmit the information. Nonetheless, the benefits inherent in such a system justifies its use. Some of these benefits include:

- interference suppression (intentional or otherwise)
- sharing of a spectrum by many users at the same time (multiple accessing capability)

- low probability of intercept
- ranging or time delay measurements
- selective addressing
- combatting of multipath [1,2]

These advantages, however, can not be realized if the received code is not synchronized with the receiver local code. Since the synchronization circuits make up a large portion of the spread spectrum receiver, it would be desirable in many applications to reduce the space taken up by these circuits. In as much as a direct sequence spread spectrum system is one of the simplest and least expensive of the spread spectrum systems, part of this thesis will focus on the feasibility of putting some of the direct sequence synchronization subsystems on a chip.

1.2 MOBILE COMMUNICATIONS

As was earlier mentioned, spread spectrum can be used to combat multipath fading, which is one of the major problems that affect the performance of a mobile communications system. Variations in received signal strength resulting from multipath fading is a consequence of the fact that the mobile radio channel is characterized by multiple signal paths between the transmitter and receiver. The multiple signal paths is a result of buildings, terrain, and other structures in proximity to the receiver, which scatter the signal energy. The multiple reflected signals add according to their phase relationships to produce a roughly stationary standing wave pattern with nulls occurring approximately at half wavelengths of the RF frequency. Thus, a mobile receiver moving through this spatial interference pattern experiences random signal fading (Figure 1.1). Signal strength variations as high as 40 dB are quite possible [2,3,4,5].

Figure 1.1: Multipath Fading in a Mobile Communication Environment, [3].

The mobile communications channel is often represented in literature [4,5,6,7] as having a slowly varying mean (due to shadowing by man-made and natural structures) with a lognormal probability density function, upon which is superimposed a fast fading component (due to multipath) with a Rayleigh distribution. Based on the experimental data obtained from various cities and suburbs, comparisons have been made between several distribution functions (Rayleigh, Rice, lognormal and Nakagami) and the experimental distribution function. The best fit to the data, as can be seen from [4,8,9] have come from the Nakagami and Rician distribution.

If a Rician channel is thus assumed, then the channel would give rise to a major stable communication path and a number of weaker communication paths. If the stable component is of sufficient amplitude, then the effects of fading (caused by the other paths) on the performance of a spread spectrum direct sequence synchronization subsystem can be neglected (note that a Rayleigh distribution provides no strong stable component) [10]. Thus, under the stated conditions, direct sequence spread spectrum can be used for mobile radio and will offer not only protection against multipath fading but will also offer a multiple accessing capability, interference rejection, privacy, etc. The obvious advantage over a conventional system is that the DS system will combat multipath without any additional signal processing (space diversity, etc.) as would be required in a conventional system.

1.3 DS SSMA FOR GAUSSIAN CHANNELS

Direct-sequence spread-spectrum multiple access (DS/SSMA) is a form of code division multiple access communications. In this system, several users occupy the same bandwidth simultaneously. However, each user uses a distinct pseudo noise (PN) code sequence which allows only the receiver with an identically synchronized code sequence to receive the message. Any interfering signal or transmissions using different sequences are rejected.

Under a Gaussian channel, a DS/SSMA system can accommodate a lot more users than under a fading channel [11,12]. Moreover, the analysis for a fading channel is much more complex than for a Gaussian channel. As a result, the synchronizers that will be designed here will be based on the assumption of an additive white Gaussian noise (AWGN) channel. Some of the performance considerations for a DS/SSMA that would affect the design of the synchronization subsystem will be touched upon in a later section.

1.4 VLSI CMOS CIRCUITS

As was earlier mentioned, it is highly desirable to put some of the DS synchronization subsystem on a chip. Of the available IC families, it was decided that CMOS logic would be preferred because of its low power, high fanout (ability to drive other gates) and high noise immunity.

As an illustration of the feasibility of the VLSI implementation of the direct sequence synchronization subsystem, some parts have been implemented on CMOS ICs. This has been done using the Metheus-Computervision 700 series workstation provided by the Canadian Microelectronics Corporation. Using the standard cells developed by previous students, Ralph Mason and Richard Vallee, the circuits have been designed and then subsequently fabricated by Northern Telecom. Because the whole layout was done manually (placing and routing of standard cells), there was time to implement only a few of the synchronization subblocks.

Chapter 2

SPREAD SPECTRUM SYSTEMS THEORY

Before embarking on the design of the spread spectrum system, a consideration of the different kinds of spread spectrum systems and their characteristics will be given. This is followed by a detailed analysis of the synchronization subsystem of the direct sequence system.

2.1 SPREADING TECHNIQUES

There are two fundamental types of spread spectrum systems:

1. Direct or pseudorandom sequence systems: in this system the carrier is directly biphase modulated by a code sequence which has a much higher bit rate (chip rate) than the bandwidth of the information signal.
2. Frequency hoppers: here the carrier frequency is shifted in discrete increments from frequency to frequency in a pattern determined by a code sequence.

Other spreading techniques such as time hopping and chirp systems will be discussed later. Hybrids of the above systems are often used in order to capitalize upon the advantage of a particular method while avoiding the disadvantages [13].

A measure of performance improvement that is common to all SS systems is the "processing gain". It is defined as the difference between system interference rejection performance using spread spectrum techniques and interference rejection performance when not using spread spectrum, all other conditions being equal [14]. The processing gain is quite often given as the ratio

$$PG = \frac{R_c}{R_s} \quad (2.1 - 1)$$

where R_s is the information symbol rate and R_c is the spreading code's "chip" rate.

2.1.1 Direct Sequence

Direct sequence modulation is obtained by digitally adding (modulo 2) the binary output of a code generator with that of the binary message, which is then used to modulate a carrier. This is illustrated in Figure 2.1. The output of the code generator - the pseudonoise (PN) code - is obtained by use of a maximal length (ML) shift register by using the appropriate feedback taps. The length of the output sequence for a maximal length shift register is given by $2^n - 1$, where n is the number of stages. A list of maximal length sequences as well as a discussion of other types of PN code generators can be found in [2,14,15].

In order to better illustrate what goes on in a DS system, the transmission and reception process will be examined for the simplest form of DS spread spectrum system - DS BPSK (Direct Sequence Binary Phase Shift Keying). Figure 2.2 illustrates a more mathematically tractable way of viewing Figure 2.1 (note that

Figure 2.1: BPSK Direct Sequence Modulation.

The undersigned hereby recommend to
the Faculty of Graduate Studies and Research
acceptance of the thesis,
**DIRECT SEQUENCE SSMA COMMUNICATIONS WITH SOME IC
REALIZATIONS**

submitted by Donald C. Kemdirim, B.Eng.,
in partial fulfilment of the requirements for the degree of
Master of Engineering

Thesis Supervisor

Chairman,
Department of Electronics

Faculty of Engineering
Department of Electronics
Carleton University
October, 1988

$d(t)$ and $c(t)$ are the bipolar waveform representations of the data bits and the PN generator output, respectively, while $\sqrt{2P} \cos \omega_0 t$ represents the carrier signal). The output of the transmitter is then given (as shown) by:

$$S_t(t) = \underbrace{\sqrt{2P} \cos [\omega_0 t + \Theta_d(t)]}_{S_d(t)} \times c(t) \quad (2.1.1 - 1)$$

where $\Theta_d(t)$ represents the data phase modulation and $S_d(t) = \sqrt{2P} \cos [\omega_0 t + \Theta_d(t)]$. At the receiver, if there is a correlation with a code which is identical to the spreading code, whose estimate of the transmission delay τ_d is given by $\hat{\tau}_d$ as in Figure 2.3(a), the recovered signal becomes

$$S_R(t) = \sqrt{2P} c(t - \tau_d) c(t - \hat{\tau}_d) \cos [\omega_0 t + \Theta_d(t - \tau_d) + \phi] \quad (2.1.1 - 2)$$

where ϕ is random phase. Now, if the receiver estimate $\hat{\tau}_d$ is equal to τ_d , then $c(t - \tau_d) \times c(t - \hat{\tau}_d)$ will equal 1 since $c(t) = \pm 1$. The result will then be equal to

$$S_d(t) = \sqrt{2P} \cos [\omega_0 t + \Theta_d(t - \tau_d) + \phi] \quad (2.1.1 - 3)$$

which can then be demodulated by a conventional phase demodulator [14]. Figure 2.3(b) shows the BPSK DS spreading and despreading waveforms.

A lot of insight can be gained by considering the power spectra of the signals discussed above. The data modulated signal before spreading has a two sided power spectral density (PSD) given by [13,15]

$$S_d(f) = \frac{1}{2} P T_c \left\{ \text{sinc}^2 [(f - f_0) T_c] + \text{sinc}^2 [(f + f_0) T_c] \right\} \quad (2.1.1 - 4)$$

Figure 2.3: (a) BPSK Direct Sequence Spread Spectrum Receiver; (b) BPSK DS Spreading and Despreading. [14]

assuming that the data is random (T_s represents the data symbol period while f_0 represents the carrier frequency). For the spread waveform, this becomes [13,15]

$$S_t(f) = \frac{1}{2}PT_c \{ \text{sinc}^2[(f - f_0)T_c] + \text{sinc}^2[(f + f_0)T_c] \} \quad (2.1.1 - 5)$$

again assuming that the spreading code is truly random (the actual power spectrum of a pseudorandom code sequence will be discussed later). These spectrums are illustrated in Figure 2.4, where $1/2PT_s$ has been assumed equal to P_0 . Comparing the two PSDs, it can be seen that the spectrum of the transmitted signal has been spread and lowered by a factor equal to

$$M = \frac{T_s}{T_c} = \frac{R_c}{R_s} = PG \quad (2.1.1 - 6)$$

This gives it a lower probability of detection. Now, at the receiver (with a synchronized code), the transmitted signal is collapsed to the data bandwidth, whereas any interfering signal is spread by a factor of M . Thus, the transmitted information is afforded a power advantage (over an interfering signal) that is equal to a factor of M .

2.1.2 Direct Sequence (PN) Power Spectral Density

In the previous section, the power spectrum for the transmitted signal was obtained by assuming that both the information signal and the spreading code sequence were totally random. In practice, though, the spreading PN code sequence is periodic. To determine the effect of this periodicity, the true power spectral density (PSD) of a PN waveform will be found.

Figure 2.4: Power Spectral Densities: (a) of data modulated carrier; (b) of data and spreading code modulated carrier.

Figure 2.5: Autocorrelation Function for PN Waveform. [15]

In order to obtain the PSD of the PN waveform, it is more convenient to first determine its autocorrelation function, from which the PSD is obtained by taking its Fourier transform [15]. The autocorrelation function is given as [15]

$$R_{PN}(\tau) = \frac{1}{NT_c} \int_0^{NT_c} PN(t)PN(t+\tau)dt \quad (2.1.2-1)$$

where T_c is the chip time, N is the length of the code sequence, $PN(t)$ is the PN waveform and τ represents the time displacement between the two waveforms

being correlated. (While the autocorrelation indicates the similarity between two identical codes or waveforms that have a time displacement τ between them, the cross-correlation would indicate the same thing between two different codes or waveforms). $R_{PN}(\tau)$, which is periodic, with a period equal to NT_c is shown in Figure 2.5. It can be written as

$$R_{PN}(\tau) = -\frac{1}{N} + \frac{N+1}{N} \bigwedge_{T_c}(\tau) * \sum_{m=-\infty}^{\infty} \delta(\tau + NmT_c) \quad (2.1.2 - 2)$$

where the '*' denotes the convolution operation and $\bigwedge_{T_c}(\tau)$ is given by [15]

$$\bigwedge_{T_c}(\tau) = \begin{cases} 1 - \frac{|\tau|}{T_c}; & |\tau| \leq T_c \\ 0; & \text{elsewhere} \end{cases} \quad (2.1.2 - 3)$$

Looking at Figure 2.5, it is easy to see how the spread spectrum system rejects multipath. If the reflected signal is delayed (compared with the direct path signal which it is assumed that the SS system is tracking) by more than one chip, there will be no correlation between it and the local code reference. Thus, it is treated exactly the same as any other interfering signal.

Now taking the Fourier transform of $R_{PN}(\tau)$ gives [15]

$$S_{PN}(f) = \frac{1}{N^2} \delta(f) + \frac{N+1}{N} \sum_{m=-\infty}^{\infty} \text{sinc}^2\left(\frac{m\pi}{N}\right) \delta\left(f + \frac{m}{NT_c}\right) \quad (2.1.2 - 4)$$

Figure 2.6 shows a sketch of this spectral density (note that the spectral lines are separated by $1/NT_c$). If the PN waveform, is used to biphase modulate a sinusoidal carrier having power P_c and frequency f_0 , the spectral density that results is given as [15]

Figure 2.6: Power Spectral Density of a PN Waveform. [14]

$$S_c(f) = \frac{P_c}{2} S_{PN}(f + f_0) + \frac{P_c}{2} S_{PN}(f - f_0) \quad (2.1.2 - 5)$$

Note that the result of this modulation is that the discrete spectrum of Figure 2.6 is shifted upward and downward by a frequency f_0 (the lines are still separated by $1/NT_c$). If, as is often the case, the carrier is also randomly modulated by data (in practice, the digital data is exclusively ored with the spreading sequence), then each one of the discrete lines in Figure 2.6 will be spread over the data bandwidth, resulting in a continuous spectrum. The resulting spectral shape however will depend on the relative size of $1/T_c$ and $\frac{1}{NT_c}$, where T_c is the data period. Figure 2.7 illustrates the cases for $T_c > 2NT_c$ and $T_c \leq NT_c$. It is obvious that for data concealment, the case of $T_c \leq NT_c$ is preferred. Thus, if $T_c \leq NT_c$, most of the power will not be restricted to only a few spectral components but is spread out such that its power spectral density is lower, flatter and more difficult to distinguish from thermal and environmental noise [16].

2.1.3 Frequency Hopping

As stated earlier, frequency hoppers (Figure 2.8) vary the carrier frequency from one frequency to another in a pattern determined by a code sequence. Frequency hoppers are often classified as either "fast hop" or "slow hop" systems. In the fast-hop system, the frequency hopping occurs at a rate that is greater than the message bit rate; while in the slow-hop system, the hop rate is less than the message bit rate [13].

Reception of frequency hopping signals is usually done on a noncoherent basis because coherent reception is more difficult to achieve. Thus, binary or M-ary FSK with noncoherent detection is commonly employed with FH spread spectrum systems [2,14,17]. Although FH systems can easily attain high spectrum spreading,

Figure 2.7: Spectral Density of Data and Code Modulated Carrier for (a) $T_b > 2NT_c$; (b) $T_b \leq NT_c$. [13]

Figure 2.8: Frequency Hop Spread Spectrum System: (a) Transmitter; (b) Receiver.

their complex frequency synthesizers and error correction coding (needed in some systems to correct for any simultaneous occupation of a frequency slot) complicate the ready implementation of these systems.

2.1.4 Other Spreading Techniques

Although the spreading techniques most often used are direct sequence and frequency hopping, some of the other available spreading techniques will be briefly discussed.

2.1.4.1 Time Hopping

In time hopping systems, the time of transmission is governed by a code sequence. The transmitter therefore turns on and off at pseudorandomly selected time slots in each time interval or frame. It only takes a little thought to see that the bandwidth (BW) of the information signal is not spread in the same fashion as the other forms of spread spectrum systems. The processing gain in this case depends on the number of time slots that each transmission interval or frame is divided into. Thus, if biphase modulation is used and the original transmission interval is divided into M time slots, then the duty cycle of the transmitter is reduced by M . This means that as a result of using time hopping, the transmitted signal BW has been increased by M .

It is important to note however that the above system can easily be jammed by a biphase modulated signal with a period equal to that of a time slot. The only advantage offered by time hopping is that since the interfering signal is on all the time, its power requirement exceeds that of the time hopper by a factor of M . Because of this relative vulnerability to interference, time hopping is most often used in conjunction with other spread spectrum systems.

2.1.4.2 Chirp

In a chirp spread spectrum system, the carrier is swept over a wideband during a given data interval. Note that in contrast to hopping from frequency to frequency as in FH systems, the frequency either increases or decreases in a continuous manner. Considering only linearly swept signals, the relationship between frequency and time is illustrated in Figure 2.9, resulting in a processing gain of BT .

Although the chirp technique is most often used in radar systems, it is also sometimes used in communication systems [13].

2.2 SYNCHRONIZATION

Synchronization is the most important part of a spread spectrum system. Without synchronization between the received and the identical local code, the transmitted information cannot be correctly recovered - it would be as if the receiver were operating without the correct code. Although code synchronization is necessary in all spread spectrum systems, only direct sequence synchronization methods will be addressed here.

There are two steps in the code synchronization process: acquisition (the local and received codes are phase aligned to within a fraction of a chip) and tracking (the local and received code phases are further aligned and maintained). In the following sections, the techniques used for direct sequence synchronization will be examined, following which the methods for carrier synchronization (needed to coherently demodulate the data) will be discussed.

Figure 2.9: Time-Frequency Relationship for a Chirp Spread Spectrum System. [13]

2.2.1 PN Code Acquisition

Spread spectrum systems employ noncoherent PN code acquisition techniques more often than coherent techniques. This is because PN coherent acquisition systems would require that carrier acquisition and tracking be accomplished beforehand. To do this (carrier acquisition and tracking) under the extremely low SNRs of SS systems is difficult. With noncoherent PN acquisition, not only is carrier acquisition and tracking performed afterwards, they are carried out under higher SNR conditions, which makes it easier.

PN code acquisition can be accomplished in many different ways. The methods which can be readily implemented on a chip include:

1. Maximum Likelihood Algorithm - this algorithm requires that the input PN signal be correlated with all possible code positions, with the one giving the highest correlation value being chosen. Note that since all code positions (or fractional code positions) must be examined before a decision is made, this technique would take either too much time (if the correlations are done in a serial fashion) or entail excessive hardware complexity (for parallel correlations).
2. Sequential Estimation - this technique estimates and loads the code generator with the first n received PN code chips (n is the number of stages in the code generator). Since the next combination of states of the code generator depends only on the present combination, if the initial estimates are correct, then the sequence produced by the estimates will be in phase with the received sequence and the signal could be successfully despread. If the initial estimates are incorrect, then another estimate is made. This process repeats itself until there is an indication that the signal has been successfully despread. This method is not suitable for low SNR conditions and is highly vulnerable to

noise and interfering signals. Moreover, it cannot be used for multiple access systems.

3. Digital Matched Filters - the digital matched filter digitally correlates the incoming sequence with a fixed finite segment (reference) of the PN waveform while the continuous time output is tested against a threshold to determine when the two are in synchronism. When the threshold is exceeded, the PN generator is started with the initial condition corresponding to the stored reference (which is now synchronized with the incoming sequence). Although this technique has very short acquisition times the achievable processing gain (time bandwidth product) is technologically limited [14,18].
4. Serial Search - in this method, also referred to as a sliding correlation, a search of different fractional code positions is performed while testing whether the continuous correlation of the received and local PN sequences exceed a certain threshold. In the past, this serial search was accomplished by linearly varying the time difference between the PN modulation on the incoming carrier and the PN waveform generated at the receiver (in essence, offsetting the clock frequencies for the incoming and local codes). In recent years, the trend has been to accomplish the time variation by a discrete stepping process where the phase of the local PN code is at uniform time increments advanced (or retarded) by a fraction of a chip. This latter method, when used with an integrate and dump filter and proper choice of variables, can produce smaller acquisition times. The advantage of the serial search method is that it is simple to implement, although it has a relatively long acquisition time (in comparison to the matched filter or sequential estimation) associated with it.

From an examination of the above techniques, it can be deduced that from the standpoint of multiple access operation, large processing gain and hardware complexity, the serial search method is the best. The disadvantage of having an

acquisition time that is longer than that of the sequential estimation or DMF techniques can be alleviated by applying certain acquisition time reduction methods (which will be discussed later). Thus, the serial search technique will be focused on in describing the acquisition process.

2.2.1.1 The Serial Search Technique

In the stepped serial search technique, the received and local PN waveforms are correlated for a certain period of time before a decision is made as to whether they are in phase or whether the local code should be stepped to a new code phase. Systems that correlate for one fixed examination period before making a decision are called "single dwell systems" while those that employ more than one fixed examination period before a decision is made are called "multiple dwell systems". By no means does this exhaust the types of serial search techniques available; for a more detailed examination, reference [17] should be consulted.

The focus here will be on single and double dwell systems, starting from the analysis of each to the advantage of the double dwell system over a single dwell system. Following this, the detection and false alarm probabilities of the above systems will be discussed. Finally, the methods that are employed to ensure that at the indicated point of acquisition, the local and received PN waveforms are in phase, will be examined.

2.2.1.1.1 The Single Dwell System

In the single dwell system (Figure 2.10), the received signal plus noise is correlated with an identical local PN code sequence, and is passed through a bandpass (pre-detection) filter. The filter output is then square-law envelope detected and integrated for a fixed time duration, τ_D (the dwell time), by an integrate and dump circuit. This is then compared to a predetermined threshold. If the threshold is not

Figure 2.10: Block Diagram of a Single Dwell PN Acquisition System. [17]

exceeded, the search control retards or advances the PN code generator so that the local code moves to a different code phase and the process is repeated [17,18].

The acquisition model for the above described system is considered in the following paragraphs. The number of possible code alignments or "cells" to be examined is given by q . Now, in practice, it is usually required that the PN code sequences be aligned to within one half of a code chip period ($T_c/2$) before relinquishing control to the tracking system. As a result, the local PN code is usually retarded or advanced in discrete steps of a one-half chip period in order to meet this requirement. In this particular situation, q would be equal to $2N$, where N is the number of chips in the PN code period. If the output of the integrate and dump circuit (I&D) is above the threshold ("hit"), the system enters a verification mode which can include a longer dwell time and entry into the code loop tracking mode. If this hit is a false alarm (code alignment indicated when the codes are not aligned), then there is a false alarm penalty time of $K\tau_D$ associated with this "false hit" (note that although $K\tau_D$ is usually mathematically modelled as being fixed it is in reality a random variable). However, if the hit corresponds to a true hit, then the system has acquired the signal and the search comes to an end [15,17].

As can be readily deduced from the above, the time (T_{acq}) to achieve acquisition (obtain a true hit) is a random variable and depends in general on the initial code phase position of the local PN generator relative to that of the received code. Because it is very difficult to obtain the probability density function of the time to acquire lock, the mean acquisition time, $\bar{T}_{acq}$ and the acquisition time variance, σ_{acq}^2 , are often used instead as a measure of system performance. In the initial determination of the acquisition time and variance, it will be assumed that there is no code doppler present in the received PN signal and that the detection probability P_D is constant. The assumption of constant P_D implies that there is one cell that corresponds to a "correct" code alignment. However, since the PN code correlation function is triangular from $-T_c$ to T_c , for a search conducted in increments of $T_c/2$

(as is most often the case), there are actually four cells which would give a non zero code correlation. Consequently, the results obtained with a constant P_D assumption must be appropriately modified [15,17].

A Markov chain can be defined as a sequence of random experiments whose outcomes depend at most upon the result of the immediately preceding experiment [19]. Now, in the acquisition process, the probability of acquiring lock during the searching of the n^{th} cell depends on the result of searching the $n - 1$ cell; for if acquisition had occurred at the $n - 1$ cell, then the process would have terminated. Consequently, the Markov chain model can be used to characterize the acquisition process.

The Markovian chain acquisition process can be described by state transition diagrams which can be transformed into a generating function flow graph. From the generating function flow graph, the moment generating function for the acquisition time can then be obtained, resulting in the determination of $\bar{T}_{acq}$ and σ_{acq} (for a more detailed discussion of generation function flow graphs, etc., reference [15] should be consulted).

Assuming that the number of cells to be searched is q , then the probability of having the signal present (true hit) in the k^{th} cell, given that it was not present in cells 1 through $k - 1$, is from [15]:

$$P_k = \frac{1}{q + 1 - k} \quad (2.2.1.1.1 - 1)$$

The generating function flow graph for the q -state (number of cells to be searched) Markov chain is shown in Figure 2.11. At each test position, one of two things can happen: (1) a false alarm can occur (an indication that acquisition has occurred when the PN codes are in reality misaligned) with probability P_{FA} , and a time penalty of $K\tau_D$ seconds or (2) no false alarm occurs with probability $(1 - P_{FA})$ and

Figure 2.11: Generating Function Flow Graph. [15]

a single dwell time of τ_D seconds. In addition to the above, at the true code phase position, either a correct detection can happen with probability P_D or no detection can take place with probability $(1 - P_D)$ [15,17].

The branches in the flow graph of Figure 2.11 are labelled with the product of the transition probability that is associated with going from the beginning to the end of each branch and an integer (including zero) power of a parameter z . This parameter z is used to mark time as one goes through the graph while its power represents the number of time units (dwell times) spent in traversing that branch. In addition, the sum of the branch probabilities emanating from each node is equal to unity [15,17] (refer to [15] or [17] for a more detailed discussion).

Using signal flow graph reduction techniques, reference [17] gives the moment generating function for the flow growth of Figure 2.11 as

$$U(z) = \frac{(1 - \beta)z}{1 - \beta z H^{q-1}(z)} \left[\frac{1}{q} \sum_{\ell=0}^{q-1} H^{\ell}(z) \right] \quad (2.2.1.1.1 - 2)$$

where $\beta = 1 - P_D$ and

$$H(z) = P_{FA} z^{K+1} + (1 - P_{FA})z$$

The mean acquisition time $\bar{T}_{acq}$ is next given by [17] as:

$$\bar{T}_{acq} = \frac{\partial \ln U(z^{r_D})}{\partial z} \Big|_{z=1} = \frac{2 + (2 - P_D)(q - 1)(1 + K P_{FA})}{2 P_D} \tau_D \quad (2.2.1.1.1 - 3)$$

For $q \gg 1$ (as is most often the case)

$$\bar{T}_{acq} = \frac{(2 - P_D)(1 + K P_{FA})}{2 P_D} (q \tau_D) \quad (2.2.1.1.1 - 4)$$

The variance of the acquisition time is given by [17]

$$\sigma_{acq}^2 = \left[\frac{d^2 U(z^D)}{dz^2} + \frac{du(z^D)}{dz} - \left(\frac{du(z^D)}{dz} \right)^2 \right]_{s=1} \quad (2.2.1.1.1 - 5)$$

For $q \gg 1$ and $K \ll q$, this is further reduced by [17] to

$$\sigma_{acq}^2 = \tau_D^2 (1 + K P_{FA})^2 q^2 \left(\frac{1}{12} + \frac{1}{P_D^2} - \frac{1}{P_D} \right) \quad (2.2.1.1.1 - 6)$$

In the presence of code doppler, the mean acquisition time and variance are given by [15,17]:

$$\bar{T}_{acq} = \frac{(2 - P_D)(1 + K P_{FA})(N \tau_D)}{2 P_D |\mu|} \quad (2.2.1.1.1 - 7a)$$

$$\sigma_{acq}^2 = \frac{N^2 (1 + K P_{FA})^2 \left(\frac{1}{12} - \frac{1}{P_D} + \frac{1}{P_D^2} \right)}{(\mu)^2} \quad (2.2.1.1.1 - 7b)$$

where $\mu = \frac{\Delta T_c}{T_c} + \tau_D \Delta f_c + K \tau_D \Delta f_c P_{FA}$

N = number of chips to be searched

$\frac{\Delta T_c}{T_c}$ = step size of search in fractions of a chip (typically 1/2)

Δf_c = code doppler in chips

$\Delta f_c \tau_D$ = PN code phase timing shift due to code doppler during dwell time

$\Delta f_c K \tau_D$ = code phase shift during hit verification

P_{FA} = probability of a false alarm

P_D = probability of detection

Note that the algebraic sign for the code doppler Δf_c must be assigned in such a way that it is positive when the code doppler speeds up the search process and vice versa. Also note that when $\Delta f_c = 0$ (i.e., no code doppler), the results reduce to those obtained for no doppler present (equations (2.2.1.1.1-4) and (2.2.1.1.1-6)).

2.2.1.1.2 The Double Dwell System

The double dwell system is very similar to the single dwell system, requiring only a slight generalization of the single dwell system. The double dwell system has two integration periods, one to quickly search the code phases (τ_{D1} small) and the second (larger τ_{D2} implying greater probability of detection) to give a better estimate as to whether an "in-sync" code position has been found [15]. Because most of the cells searched correspond to incorrect alignments, this ability to quickly eliminate them produces a reduction in acquisition time over the single dwell system.

Figures 2.12 and 2.13 show a typical double dwell system and its algorithm. The search starts with an integration of τ_{D1} seconds. If the threshold is not exceeded, the local code phase is shifted by half a chip, for example, and the process is repeated. If however the threshold is exceeded, the integration time is changed to τ_{D2} ($\tau_{D2} > \tau_{D1}$) without changing the code phase position. This dwell provides a higher probability of detection. If the second threshold is exceeded, the code loop (for tracking) is activated while a verification mode is initiated. In this mode, the integration time τ_{D3} is greater or equal to τ_{D2} . This verification will go on during tracking and will indicate when lock has been lost. Essentially, the idea is to detect a certain number of consecutive times that the integrator output does not exceed the threshold (this will be discussed further in a later section). Although Figures 2.12 and 2.13 indicate two different threshold detectors, these can in practice be implemented as one threshold detector with a variable integration time and a changeable threshold.

Figure 2.12: Double Dwell System. [15]

Figure 2.13: Algorithm for the Double Dwell System. [15]

From [15], the mean acquisition time and variance with no code doppler are given as

$$\bar{T}_{acq} = \frac{2 - P_D}{2P_D} [\tau_{D1} + \tau_{D2} P_{FA1} (1 + K P_{FA2})] q ; \quad q \gg 1 \quad (2.2.1.1.2 - 1)$$

$$\sigma_{acq}^2 = \tau_{D1}^2 + \tau_{D2}^2 P_{FA1}^2 (1 + K P_{FA2})^2 q^2 \left(\frac{1}{12} - \frac{1}{P_D} + \frac{1}{P_D^2} \right) \quad q \gg P_{FA2} K (K + 1) \quad (2.2.1.1.2 - 2)$$

In the presence of code doppler, this becomes

$$\bar{T}_{acq} = \frac{2 - P_D}{2P_D} \frac{[\tau_{D1} + \tau_{D2} P_{FA1} (1 + K P_{FA2})]}{|\mu|} \quad (2.2.1.1.2 - 3)$$

$$\sigma_{acq}^2 = \frac{N^2 [\tau_{D1} + \tau_{D2} P_{FA1} (1 + K P_{FA2})]^2 \left(\frac{1}{12} - \frac{1}{P_D} + \frac{1}{P_D^2} \right)}{\mu^2} \quad (2.2.1.1.2 - 4)$$

where

$$\mu = \frac{\Delta T_c}{T_c} + \tau_{D1} \Delta f_c + P_{FA1} \tau_{D2} \Delta f_c + P_{FA1} P_{FA2} K \tau_{D2} \Delta f_c$$

and the parameters used above are described as

τ_{D1} = first dwell time

τ_{D2} = second dwell time

$P_D = P_{D1} P_{D2}$, product of dwell one and two detection probabilities

P_{FA1} = false alarm probability of the first dwell

P_{FA2} = false alarm probability of the second dwell

q = number of cells to be searched

N = number of chips to be searched

$\frac{\Delta T_c}{T_c}$ = step size of search

Δf_c = code doppler in chips

K = penalty (in terms of τ_{D2}) for a false alarm with the second dwell time

Note again that the algebraic sign of the code doppler Δf_c is assigned in such a way as to be positive when the search is speeded up by the code doppler. Note also that with $\tau_{D1} = 0$ and $P_{FA1} = 1$, the equations for the double dwell system reduces to those obtained for the single dwell system.

2.2.1.1.3 Detection and False Alarm Probability Approximations

The mean acquisition time and variance depend both on the probability of detection P_D and the probability of false alarm P_{FA} . In the ensuing paragraphs, expressions for the probabilities of detection and false alarm for a fixed dwell integration period, following a square law detector, with an input consisting of signal and/or band-limited Gaussian noise, will be described. Following this, the modifications that must be made to account for the effects of modulation and a nonconstant detection probability, will be discussed. Although more detailed and exact derivations exist (see references [14] and [20]), the analysis presented here will follow the more simple derivations presented in reference [15].

The input signal $y(t)$ (Figure 2.14) is taken as

$$y(t) = \sqrt{2P} \sin \omega_0 t + n(t) \quad (2.2.1.1.3 - 1)$$

where the signal is a sine wave with power P and the noise process is a sample function of a White Gaussian noise process with a two sided noise spectral density of $N_0/2$. Therefore, with the signal present, the mean output of the square law detector (Figure 2.14) is given by the sum of the noise and signal powers as:

$$\mu = N_0 B + P = N_0 B(1 + SNR) \quad (2.2.1.1.3 - 2)$$

where B is the noise bandwidth of the bandpass filter (BPF), $SNR = P/N_0 B$ and N_0 is the one-sided noise spectral density. Thus, the output $z(t)$ of the integration after τ seconds, has a mean of

$$\mu_0 = N_0 B(1 + SNR)\tau \quad (2.2.1.1.3 - 3)$$

Since the two-sided spectral density at the origin out of the detector is given by

$$S(0) = N_0^2 B + 2PN_0 \quad (2.2.1.1.3 - 4)$$

and the transfer function for a τ second integration is given by

$$H(f) = \tau \frac{\sin \pi f \tau}{\pi f \tau} e^{-j\omega \tau/2} \quad (2.2.1.1.3 - 5)$$

so that the variance out of the integration is then

Figure 2.14: Model for the Detection and False Alarm Probability Approximations.
[15]

$$\sigma_0^2 \approx \int_{-\infty}^{\infty} S(0) \tau^2 \left(\frac{\sin \pi f \tau}{\pi f \tau} \right)^2 df \quad (2.2.1.1.3 - 6)$$

Making the assumption that $1/\tau \ll B$ means that only the spectrum near $f = 0$ is important in the evaluation of the integral of equation (2.2.1.1.3-6). With this assumption then, the evaluation results in

$$\sigma_0^2 = N_0^2 B(1 + 2SNR)\tau \quad (2.2.1.1.3 - 7)$$

(Note that $\left[\frac{\sin \pi f \tau}{\pi f \tau} \right]^2$ for f near zero is approximately equal to 1). Following the above steps for the case when the signal is absent (i.e. $P = 0$) results in the following:

$$\mu_0 = N_0 B \tau \quad \sigma_0^2 = N_0^2 B \tau ; \quad \text{signal absent}$$

$$\mu_0 = N_0 B(1 + SNR)\tau \quad \sigma_0^2 = N_0^2 B(1 + 2SNR)\tau ; \quad \text{signal present} \quad (2.2.1.1.3 - 8)$$

Now, if it is assumed that the BPF preceding the square law detector has sharp cutoff slopes, then the samples (coming out of the filter) that are separated by more than $1/B$ seconds are essentially independent [14, 15]. If, in addition $B\tau \gg 1$ ($\tau \gg 1/B$), then the output variable $z(t)$ will be the sum of a large number of independent variables and will, according to the central limit theorem, tend towards a Gaussian or normal random variable. Urkowitz [20] indicates that the Gaussian approximation is valid for $B\tau \geq 125$. For $B\tau < 125$, the more difficult analysis presented in references [20] and [14] can be used.

However, if Gaussian statistics are assumed, the probability of a false alarm (probability that $z(t)$ is above the threshold η , given that the signal is absent) is given by

$$\begin{aligned}
P_{FA} &= \int_{\eta}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left[\frac{-(z - \mu_0)}{2\sigma_0^2} \right] dz \\
&= \text{erfc} \left(\frac{\eta - \mu_0}{\sigma_0} \right) = \text{erfc} \left(\frac{\eta - N_0 B \tau}{N_0 \sqrt{B\tau}} \right) \\
&= \text{erfc}(\beta)
\end{aligned} \tag{2.2.1.1.3-9}$$

$$\text{where } \beta = \frac{\eta - N_0 B \tau}{N_0 \sqrt{B\tau}}$$

Similarly, the probability of detection (probability that $z(t)$ is above the threshold η , given that the signal is present) is given by

$$\begin{aligned}
P_D &= \int_{\eta}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left[\frac{-(z - \mu_0)^2}{2\sigma_0^2} \right] dz \\
&= \text{erfc} \left(\frac{\eta - \mu_0}{\sigma_0} \right) = \text{erfc} \left[\frac{\eta - N_0 B (1 + SNR) \tau}{N_0 \sqrt{B(1 + 2SNR) \tau}} \right] \\
&= \text{erfc} \left[\frac{\eta - N_0 B \tau}{N_0 \sqrt{B(1 + 2SNR) \tau}} - \frac{N_0 B (SNR) \tau}{N_0 \sqrt{B(1 + 2SNR) \tau}} \right] \\
P_D &= \text{erfc} \left[\frac{\beta - \sqrt{B\tau} SNR}{\sqrt{1 + 2SNR}} \right]
\end{aligned} \tag{2.2.1.1.3-10}$$

From the above equation, it can be seen that once the false alarm probability is known, so is β ; and if $B\tau$ and the SNR are known, P_D can then be found. When the signal is modulated by data, there will be some loss due to the bandpass filter. To account for this, the SNR is changed to

$$SNR = \frac{\alpha P}{N_0 B} \quad (2.2.1.1.3 - 11)$$

where α denotes the fraction of the modulated signal that is passed through the BPF.

As was earlier indicated, the calculation of the detection probability above assumed that only one cell in the entire search will give a "signal present" indication. However, since the PN correlation curve exists over ± 1 chip around the peak, a system that retards (or advances) the local PN code phase in, for example half-chip increments, could have up to four cells for which a nonzero correlation could be obtained [17].

If the maximum correlation value is taken to indicate the "signal present" position, the best condition (in terms of correlation values), for a half chip increment search, would be obtained when there are three cells with non zero correlation values and only one of these cells corresponded to the maximum correlation value. The worst case condition occurs when there are four cells with a non zero correlation value. Since systems are usually designed on the basis of worst case conditions, the latter case, with two maximum correlation points (signal present) will be used. These two positions correspond to points that are $\pm T_c/4$ from the correlation peak. Assuming normalized correlation values, these two points would have a value of 0.75 relative to the correlation peak. This therefore means that the SNR that would be used to calculate the single signal present point detection probability should be reduced by $10 \log_{10}(0.75)^2 = 2.5$ dB. In general, the amount of reduction depends on the distance from the correlation peak and is given by [17]

$$L = \left(1 - \frac{\tau}{T_c}\right)^2 \quad (2.2.1.1.3 - 12)$$

where τ represents the time misalignment from the correlation peak in chips. Thus,

the SNR that should be used in the above calculations (taking into account modulation and multiple signal present positions) is

$$SNR = \frac{\alpha LP}{N_0 B} \quad (2.2.1.1.3 - 13)$$

Note that since there are two signal present positions, the effective probability of detection increases and is given by [17] as

$$P'_{D1} = P_{D1} + (1 - P_{D1})P_{D1} = 2P_{D1} - P_{D1}^2 \quad (2.2.1.1.3 - 14)$$

where P_{D1} represents the probability of detecting signal present on the first correlation point and $(1 - P_{D1})P_{D1}$ is the joint probability of not detecting the signal on the first correlation point and detecting the signal present on the second correlation point (note that this applies to the search mode).

Another refinement to the detection probability results from the handover process between acquisition and tracking. Usually, after a "hit", the code tracking is turned on to pull the code into tight lock and maintain the synchronization. Often in low-SNR systems where both acquisition bandwidth (used to pull code into lock) and tracking bandwidth (smaller than acquisition bandwidth) are used, multiple code loop bandwidths are employed in order to soften the transition between acquisition and tracking modes. As a result, the probability of going from the acquisition mode to the final tracking loop bandwidth (without losing lock) will be less than one. At high SNRs, this handover probability is close to one. However, for low SNRs, the detection probability used in the acquisition formulas should be changed to

$$P'_D = P'_{D1} P_{D2} P_{H0} \quad (2.2.1.1.3 - 15)$$

Although this handover probability is very hard to estimate, references [15] and [21] indicate that for the S band shuttle system, it was experimentally found that at $C/N_0 = 51$ dB-Hz, the handover probability varied from 0.06 to 0.5, depending on the code doppler (C/N_0 denotes the carrier power-to noise density). In terms of the mean acquisition times given by the formulas above, this would imply an increase in acquisition time of 2 to 16 times for this situation ($C/N_0 = 51$ dB-Hz) if the handover probability is taken into account.

2.2.1.1.4 Searchlock Strategy - Acquisition Verification

The searchlock strategy (SLS) of a receiver controls the transition from the search mode to the lock mode (tracking). Not only does it verify the validity of a "hit" (before turning control over to the tracking loop), it also continuously monitors the receiver while in its lock mode to determine when a loss of synchronization has occurred (so that the search mode could be restarted) [17].

The lock detectors used in the SLS indicates both when the system has acquired lock and when the code tracking loop loses lock. There are two times that are of interest here: the mean time to indicate that the loop is out-of-lock (given that the signal is absent, i.e., false alarm) and the mean time to give an out of lock signal given that the signal is present (false dismissal). As was mentioned earlier, the idea is to detect a certain number of consecutive misses or below threshold values of the integrator before an out of lock condition is given. It can be seen that if the probability of declaring signal absent when the signal is indeed present is given by a $q < 1$, the probability of declaring it absent (when it is indeed present) n consecutive times would be q^n (which would be very small).

Holmes [15] has shown that the " n consecutive count" type of detector has a mean false dismissal time of

$$\bar{T}_f = \frac{1 + q + \dots + q^{n-1}}{q^n} \tau_D ; \quad q = 1 - p \quad (2.2.1.1.4 - 1)$$

where τ_D is the dwell time of the integrator, q is the probability that the integrator output is below the threshold and p is the probability that the integrator output is above the threshold. This formula holds both for when the signal is present and when it is absent where, of course, a different q and p apply for each case.

2.2.1.2 Methods of Acquisition Time Reduction

It was indicated in an earlier section that from the standpoint of multiple access operation, large processing gain and hardware complexity, the serial search method is the best of the acquisition techniques discussed. The disadvantage of long acquisition times can be alleviated by employing acquisition time reduction techniques, a few of which are briefly discussed in the following paragraphs.

The multiple dwell technique is one method of improving the acquisition time performance of the serial search technique (the double dwell system analyzed earlier is an example of a multiple dwell system). Acquisition time was reduced by quickly searching the code phases with a small integration time until a hit was obtained, where upon a longer integration time was used to give a better estimate.

In situations where the a priori probability density function of the region to be searched is not uniform, sweep strategies have been developed [17] that reduce the acquisition time by first searching the cells that would most likely contain the correct code phase. Another method for improving acquisition time performance is by using what is called sequential detection. In this method, a bias voltage is subtracted from the signal before integration so that when signal is present, the output of the integrator increases linearly while decreasing linearly when no signal is present. A choice of a small enough negative threshold will enable the system

to reject the incorrect cells much faster than in a conventional system (note that the negative threshold is crossed only when noise only is present). The sequential detector and its operation is illustrated in Figures 2.15 and 2.16.

Perhaps the most straight forward way to reduce the acquisition time is by use of parallel searching techniques. By dividing the entire search region into M subregions and simultaneously searching the M subregions, acquisition time can be reduced by up to a factor of approximately M . Note that in this and all of the above cases, this reduction in acquisition time was attained only with some unavoidable increase in system complexity.

2.2.2 PN Code Tracking

Once acquisition has occurred, the second part of the PN synchronization process, PN code tracking, is initiated. This part of the process tries to maintain the received and local PN codes in phase synchronism, thus allowing the received signal to be correctly recovered. As was earlier pointed out, the low SNRs associated with spread spectrum systems makes it hard to accomplish carrier acquisition and tracking before PN acquisition and tracking (as would be required for coherent PN acquisition and tracking systems). As a result, many systems employ noncoherent PN acquisition and tracking (which are performed before carrier acquisition and tracking).

There are basically two types of PN tracking loops: (1) the delay-locked loop (DLL) and (2) the tau-dither loop (TDL) - which is also sometimes called the time-shared loop. The other types of PN tracking loops are basically derived from modifications of the above two [17]. From the standpoint of performance (timing error variance, pull in or tracking range) and hardware simplicity, the tau-dither loop is quite attractive. Thus, the focus here will be on the TDL, although a brief description of the DLL operation will be given first.

Figure 2.15: The Sequential Detector. [17]

Figure 2.16: Integrator Output: (a) Bias voltage absent; (b) Bias voltage present.
[17]

2.2.2.1 The Delay-Locked Loop

In the non-coherent delay locked loop (DLL) (Figure 2.17), the input signal $x(t)$ is cross-correlated with advanced and retarded versions of the local PN code generator. The results are then bandpass filtered, square-law envelope detected and subtracted from each other to produce an error (discriminator) signal. The error signal $\xi(t)$ is applied to a loop filter and voltage controlled clock (VCC) which drives the PN code generator that generates the local PN code. The DLL basically maintains lock with the incoming code as follows: if the DLL's estimate of the transmission delay $\hat{\tau}_d$ is correct, then the difference of both arms will produce a null error signal $\xi(t)$. However, if the estimate is incorrect, the error signal produced would be nonzero (either positive or negative). Since the loop filter averages the error signal to produce a dc signal that drives the VCC, it can be seen that a nonzero error signal (incorrect estimate) speeds up or slows down the VCC (in order to correct the transmission delay estimate); while a null error signal does nothing to the VCC.

The advance and retard interval δ (as shown in Figure 2.17) is restricted to a range of $\delta \leq T_c$, where T_c is the length (in seconds) of a PN code chip. δ is commonly defined as T_c/N , where N is an integer greater than 1. This means that the advanced and retarded PN signals are $2T_c/N$ apart, although such a loop is often said to have a T_c/N correlator spacing. When $N = 2$ (i.e. $\delta = T_c/N$), the loop is commonly called a "one-delta" loop [17].

2.2.2.2 The Tau-Dither Loop

The tau-dither loop (TDL) of Figure 2.18 basically emulates the action of the DLL, using however, a lot less hardware. The operation of the TDL is as follows. The received signal $s(t)$ - Figure 2.18 - plus noise is alternately (rather than simultaneously as in the DLL) correlated with the early (advanced) and retarded (late)

Figure 2.17: Non-Coherent Delay-Lock Loop. [17]

Figure 2.18: Non-Coherent Tau-Dither Loop. [17]

versions of the locally generated PN code; ~~thus~~ the name dither. The result of the correlation is bandpass filtered, envelope detected and alternately inverted by the binary signal $q(t)$ to produce an output error signal that is used to drive the voltage controlled clock (VCC) through the loop filter $F(s)$. Note that as a result of the dithering or switching, the action of the DLL is emulated by the TDL which uses one less signal processing channel, albeit with a slightly degraded noise performance. However, since only a single input correlator is required, the problems of gain imbalance and other mismatches present in a two-channel loop like the DLL are avoided. Because the linear range and peak of the error (discriminator) signal is maximized for a one delta loop [22], the loop model and equations of operation discussed here will specifically be for the one delta loop. Note that the discussion closely follows reference [17] and is for the case of a biphas data modulation on the carrier.

When the dither frequency f_d is small compared to the noise bandwidth, B_H , of the bandpass filter (usually the case in practice), the TDL of Figure 2.18 can be described by the equivalent loop of Figure 2.19 (which is essentially a DLL) [17,22]. The received signal $x(t)$ of Figure 2.18 is the sum of the signal $s(t)$ and additive noise $n_i(t)$

$$x(t) = s(t) + n_i(t) \quad (2.2.2.2 - 1)$$

where

$$s(t) = \sqrt{2P_s} c(t - \tau_d) m(t - \tau_d) \cos[\omega_0 t + \theta(t)]$$

$$n_i(t) = \sqrt{2} \{ N_c(t) \cos[\omega_0 t + \theta(t)] - N_s(t) \sin[\omega_0 t + \theta(t)] \}$$

Figure 2.19: Equivalent Loop Model for the Tau-Dither Loop. [17]

where P_r is the received average signal power, $c(t - \tau_d)$ is the received PN signal with transmission delay, τ_d , $m(t - \tau_d)$ is the data modulation with the same transmission delay, ω_0 is the carrier radian frequency and $\theta(t) = \theta_0 + \Omega_0 t$ is an unknown carrier phase made up of a constant term and a term proportional to Doppler. $N_e(t)$ and $N_s(t)$ are roughly statistically independent, stationary, low-pass white Gaussian noise processes with single sided noise spectral density $N_0(W/Hz)$ and one-sided bandwidth $B_N \ll \omega_0/2\pi$. Now, $c(t - \hat{\tau}_d + T_c/2)$ and $c(t - \hat{\tau}_d - T_c/2)$ represent the advanced (early) and retarded (late) local PN signals, where $\hat{\tau}_d$ indicates the TDL's estimate of τ_d . Therefore, the cross-correlation (phase detector) outputs (Figure 2.19) are

$$\begin{aligned}
 e_{\pm}(t) = & \sqrt{2P_r K_m} m(t - \tau_d) \overline{c(t - \tau_d) c(t - \tau_d \pm T_c/2)} \cos [\omega_0 t + \theta(t)] \\
 & + \sqrt{2P_r K_m} m(t - \tau_d) [c(t - \tau_d) c(t - \hat{\tau}_d \pm T_c/2) \\
 & - \overline{c(t - \tau_d) c(t - \hat{\tau}_d \pm T_c/2)}] \cos [\omega_0 t + \theta(t)] \\
 & + K_m c(t - \hat{\tau}_d \pm T_c/2) n_i(t)
 \end{aligned} \tag{2.2.2.2-2}$$

where K_m denotes the phase detector gain and the overbar means statistical expectation. Note that the output of the phase detector consists of a desired signal term, a PN code self noise term (i.e., code induced noise term), and a noise term. For a long PN code period p , the autocorrelation functions for the early ($R_{PN+}(\epsilon_d)$) and late ($R_{PN-}(\epsilon_d)$) cases are:

$$R_{PN+}(\epsilon_d) \triangleq \overline{c(t - \tau_d)c(t - \hat{\tau}_d \pm T_c/2)} = \begin{cases} 0; & \epsilon_d \leq -3/2 \\ 3/2 + \epsilon_d; & -3/2 < \epsilon_d < -1/2 \\ 1/2 - \epsilon_d; & -1/2 < \epsilon_d < 1/2 \\ 0; & \epsilon_d > 1/2 \end{cases}$$

$$R_{PN-}(\epsilon_d) \triangleq \overline{c(t - \tau_d)c(t - \hat{\tau}_d - T_c/2)} = R_{PN+}(\epsilon_d - 1)$$

$$R_{PN\pm}(\epsilon_d) \triangleq R_{PN\pm}(\epsilon_d \pm np); \quad n = \pm 1, \pm 2, \pm 3, \dots \quad (2.2.2.2 - 3)$$

where $\epsilon_d = (\tau_d - \hat{\tau}_d)/T_c$ is the normalized transmission delay error. Note that the autocorrelation functions are periodic (with period p) although only one period is shown in Figure 2.20.

Letting $H_t(s)$ denote the low-pass equivalent of the bandpass filter transfer function $H(s)$ which is applied on $e_{\pm}(t)$, then $y_{\pm}(t)$ (Figure 2.19) can be written as

$$\begin{aligned} y_{\pm}(t) &= H_t(p)e_{\pm}(t) \\ &= \sqrt{2P_s K_m} \hat{m}(t - \tau_d) R_{PN\pm}(\epsilon_d) \cos[\omega_0 t + \theta(t)] \\ &\quad + \sqrt{2P_s K_m} \hat{s}_{c\pm}(t - \tau_d, \epsilon_d) \cos[\omega_0 t + \theta(t)] \\ &\quad + \sqrt{2K_m} \hat{N}_{c\pm}(t) \cos[\omega_0 t + \theta(t)] + \sqrt{2K_m} \hat{N}_{s\pm}(t) \sin[\omega_0 t + \theta(t)] \end{aligned} \quad (2.2.2.2-4)$$

Figure 2.20: Autocorrelation Functions for: (a) Advanced PN code; (b) Retarded PN code. [27]

where

$$\hat{m}(t) = H_\ell(p)m(t)$$

$$\hat{s}_{c\pm}(t, \epsilon_d) = H_\ell(p)[m(t)s_{c\pm}(t, \epsilon_d)]$$

$$s_{c\pm}(t - \tau_d, \epsilon_d) \triangleq c(t - \tau_d)c(t - \hat{\tau}_d \pm \epsilon_d) - \overline{c(t - \tau_d)c(t - \hat{\tau}_d \pm \epsilon_d)}$$

$$\hat{N}_{c\pm}(t) = H_\ell(p)[\hat{c}(t - \tau_d \pm \epsilon_d)N_c(t)]$$

$$\hat{N}_{s\pm}(t) = H_\ell(p)[c(t - \tau_d \pm \epsilon_d)N_s(t)]$$

and $H_\ell(p)f(t)$ denotes the Heaviside operator $H_\ell(p)$ operating on $f(t)$. Note that $H_\ell(p)f(t)$ denotes $h_\ell(t) * f(t)$, where $h_\ell(t)$ is the low pass equivalent of the bandpass filter impulse response ($p \triangleq d/dt$ and the symbol $\hat{\cdot}$ indicates the filtered version of the corresponding term).

Note that the term $s_{c\pm}(t - \tau_d, \epsilon_d)$ is used to represent the PN code self-noise process. When the single-sided loop bandwidth B_L is much less than the PN code chip rate (which is usually the case in practice), the effect of the PN code self-noise on loop performance can be ignored. Therefore, neglecting the self noise term in (2.2.2.2-4) and the second harmonic terms produced by the square law envelop detector (since the loop filter will remove them in any case), the input to the loop filter, $\xi(t)$, becomes

$$\hat{\xi}(t) \triangleq y_-^2(t)q_1(t) - y_+^2(t)q_2(t)$$

$$= P_s K_m^2 \hat{m}^2(t - \tau_d) [q_1(t) R_{PN-}^2(\epsilon_d) - q_2(t) R_{PN+}^2(\epsilon_d)]$$

$$+ K_m^2 n_e(t, \epsilon_d)$$

(2.2.2.2-5)

where

$$n_e(t, \epsilon_d) = q_1(t) [\hat{N}_{c-}^2(t) + \hat{N}_{s-}^2(t)] - q_2(t) [\hat{N}_{c+}^2(t) + \hat{N}_{s+}^2(t)]$$

$$+ 2\sqrt{P_s} \hat{m}(t - \tau_d) [q_1(t) R_{PN-}(\epsilon_d) \hat{N}_{c-}(t)$$

$$- q_2(t) R_{PN+}(\epsilon_d) \hat{N}_{c+}(t)]$$

(2.2.2.2-6)

Since $q(t) = q_1(t) - q_2(t)$ and $q_1(t) + q_2(t) = 1$, then $\xi(t)$ simplifies to

$$\xi(t) = P_s K_m^2 \hat{m}^2(t - \tau_d) D(t, \epsilon_d) + K_m^2 n_e(t, \epsilon_d)$$

(2.2.2.2 - 7)

where

$$D(t, \epsilon_d) \triangleq q_1(t)R_{PN-}^2(\epsilon_d) - q_2(t)R_{PN+}^2(\epsilon_d) = \begin{cases} 0; & \epsilon_d \leq -3/2 \\ -q_2(t)(3/2 + \epsilon_d)^2; & -3/2 < \epsilon_d \leq -1/2 \\ (1/4 + \epsilon_d^2)q(t) + \epsilon_d; & |\epsilon_d| \leq 1/2 \\ q_1(t)(3/2 - \epsilon_d)^2; & 1/2 \leq \epsilon_d \leq 3/2 \\ 0; & \epsilon_d \geq 3/2 \end{cases}$$

$$D(t, \epsilon_d) = D(t, \epsilon_d + np); n = \pm 1, \pm 2, \pm 3, \dots \quad (2.2.2.2-8)$$

The term $\hat{m}^2(t - \tau_d)D(t, \epsilon_d)$ in (2.2.2.2-7) can be decomposed into its mean value and modulation self-noise, that is

$$\begin{aligned} \hat{m}^2(t - \tau_d)D(t, \epsilon_d) &= \overline{\hat{m}^2(t - \tau_d)} D(t, \epsilon_d) \\ &+ [\hat{m}^2(t - \tau_d) - \overline{\hat{m}^2(t - \tau_d)}] D(t, \epsilon_d) \end{aligned} \quad (2.2.2.2-9)$$

where $\langle \rangle$ stands for time average. Furthermore, $\overline{\hat{m}^2(t - \tau_d)}$ can be written as

$$\overline{\hat{m}^2(t - \tau_d)} \triangleq M_2 = \int_{-\infty}^{\infty} S_m(f) |H_L(j2\pi f)|^2 df \quad (2.2.2.2 - 10)$$

where $S_m(f)$ is the power spectral density of the data modulation. Note that M_2 is the equivalent power reduction factor that results from the distortion suffered by the data modulation as it passes through the bandpass filter. Since the loop bandwidth, B_L is in practice chosen to be narrow with respect to the data symbol

rate, $R_s = 1/T_s$, and to the dither frequency, $f_d = 1/T_d$, several things can be concluded:

1. $\hat{m}^2(t - \tau_d)D(t, \epsilon_d)$ is averaged
2. Because $B_L \ll R_s$, the modulation self-noise can be neglected with little error
3. Because $B_L \ll f_d$, discrete lines at f_d will not get into the loop to deteriorate loop performance.

As a result of observations (1) and (2), $\hat{m}^2(t - \tau_d)D(t, \epsilon_d)$ can be replaced by $\overline{\hat{m}^2(t - \tau_d)D(t, \epsilon_d)}$. This can further be expressed as

$$\begin{aligned} \overline{\hat{m}^2(t - \tau_d)D(t, \epsilon_d)} &= \overline{\hat{m}^2(t - \tau_d)} \langle D(t, \epsilon_d) \rangle \\ &= M_2 \langle D(t, \epsilon_d) \rangle \end{aligned} \tag{2.2.2.2-11}$$

Since $\langle q(t) \rangle = 0$ and $\langle q_1(t) \rangle = \langle q_2(t) \rangle = 1/2$, then

$$\begin{aligned} D(\epsilon_d) &= \langle D(t, \epsilon_d) \rangle \\ &= \langle q_1(t)R_{PN-}^2(\epsilon_d) - q_2(t)R_{PN+}^2(\epsilon_d) \rangle \\ &= \langle q_1(t) \rangle R_{PN-}^2(\epsilon_d) - \langle q_2(t) \rangle R_{PN+}^2(\epsilon_d) \end{aligned} \tag{2.2.2.2-12}$$

$$\begin{aligned}
&= 1/2 [R_{PN-}^2(\epsilon_d) - R_{PN+}^2(\epsilon_d)] \\
&= \begin{cases} 0 ; & \epsilon_d \leq -3/2 \\ -1/2(3/2 + \epsilon_d)^2 ; & -3/2 \leq \epsilon_d \leq -1/2 \\ \epsilon_d ; & |\epsilon_d| \leq 1/2 \\ 1/2(3/2 - \epsilon_d)^2 ; & 1/2 \leq \epsilon_d \leq 3/2 \\ 0 ; & \epsilon_d \geq 3/2 \end{cases}
\end{aligned}$$

$$D(\epsilon_d) = D(\epsilon_d + np) \quad ; \quad n = \pm 1, \pm 2, \pm 3, \dots$$

Thus, equation (2.2.2.2-7) can be rewritten as

$$\xi(t) = P_s K_M^2 M_2 D(\epsilon_d) + K_M^2 n_e(t, \epsilon_d) \quad (2.2.2.2 - 13)$$

The discriminator characteristic (S-curve) $D(\epsilon_d)$ is shown in Figure 2.21. The instantaneous normalized delay estimate $\hat{\tau}_d/T_c$ of the PN code generator output can be written in terms of the error signal as

$$\frac{\hat{\tau}_d}{T_c} = K_{VCC} K_2 \frac{F(p)}{p} \xi(t) \quad (2.2.2.2 - 14)$$

where $F(p)$ is the loop filter transfer function, $p = \frac{d}{dt}$, K_{VCC} is the gain of the voltage controlled clock that drives the PN code generator and K_2 is the gain of a possible amplifier before the loop filter. Combining (2.2.2.2-7) and (2.2.2.2-14) and letting $K = K_2 K_m^2 K_{VCC}$, $\hat{\tau}_d/T_c$ becomes

$$\frac{\hat{\tau}_d}{T_c} = \frac{KF(p)}{p} [P_s \hat{m}^2(t - \tau_d) D(t, \epsilon_d) + n_e(t, \epsilon_d)]$$

Figure 2.21: Discriminator Characteristic (S-curve).

or

$$\epsilon_d = \frac{\tau_d}{T_c} - \frac{KF(p)}{p} [P_s \hat{m}^2(t - \tau_d) D(t, \epsilon_d) + n_e(t, \epsilon_d)] \quad (2.2.2.2 - 15)$$

Using the same reasoning used to change $\hat{m}^2(t - \tau_d) D(t, \epsilon_d)$ in equation (2.2.2.2-7) to $M_2 D(\epsilon_d)$ in equation (2.2.2.2-13), the $\hat{m}^2(t - \tau_d) D(t, \epsilon_d)$ term in (2.2.2.2-15) can similarly be changed to $M_2 D(\epsilon_d)$ so that (2.2.2.2-15) can be rewritten as

$$\frac{d\epsilon_d}{dt} = \frac{1}{T_c} \left(\frac{d\tau_d}{dt} \right) - KF(p) P_s M_2 \left[D(\epsilon_d) + \frac{n_e(t, \epsilon_d)}{P_s M_2} \right] \quad (2.2.2.2 - 16)$$

In order to get a statistical characterization of the equivalent additive noise, $n_e(t, \epsilon_d)$, its autocorrelation function must be obtained. This can be written as

$$R_e(\tau, \epsilon_d) \triangleq \overline{n_e(t, \epsilon_d) n_e(t + \tau, \epsilon_d)} \quad (2.2.2.2 - 17)$$

Now, if the auto and cross correlation functions of the switching waveforms are written as

$$R_{q1}(\tau) \triangleq \langle q_1(t) q_1(t + \tau) \rangle$$

$$R_{q2}(\tau) \triangleq \langle q_2(t) q_2(t + \tau) \rangle$$

$$R_{q1q2}(\tau) \triangleq \langle q_1(t) q_2(t + \tau) \rangle \quad (2.2.2.2 - 18)$$

then

$$R_{q1}(\tau) = R_{q2}(\tau) = 1/2 - R_{q1q2}(\tau) \triangleq R_q(\tau) \quad (2.2.2.2 - 19)$$

$R_q(\tau)$ is as shown in Figure 2.22. Reference [17] indicates that by substituting (2.2.2.2-6) into (2.2.2.2-17) and making use of (2.2.2.2-19), that (2.2.2.2-17) can be reduced to

$$\begin{aligned} R_s(\tau, \epsilon_d) = & 8R_N^2(\tau)R_q(\tau) + 8R_N^2(0)[2R_q(\tau) - 1/2] \\ & + 4P_s R_m(\tau)R_N(\tau)R_q(\tau)f(\epsilon_d) \end{aligned} \quad (2.2.2.2-20a)$$

where

$$R_N(\tau) = \frac{N_0}{2} \int_{-\infty}^{\infty} |H_L(j2\pi f)|^2 e^{j2\pi f\tau} df \quad (2.2.2.2 - 20b)$$

$$R_m(\tau) = \frac{N_0}{2} \int_{-\infty}^{\infty} S_m(f) |H_L(j2\pi f)|^2 e^{j2\pi f\tau} dt \quad (2.2.2.2 - 20c)$$

$$f(\epsilon_d) = \begin{cases} 0 ; & \epsilon_d \leq -3/2 \\ (3/2 + \epsilon_d)^2 ; & -3/2 < \epsilon_d \leq -1/2 \\ 1/2 + 2(\epsilon_d)^2 & | \epsilon_d | \leq 1/2 \\ (3/2 - \epsilon_d/2)^2 ; & 1/2 \leq \epsilon_d \leq 3/2 \\ 0 ; & \epsilon_d \geq 3/2 \end{cases} \quad (2.2.2.2-20d)$$

and $N_0/2$ is the two sided Gaussian noise spectral density. Since the TDL loop bandwidth B_L is normally chosen to be much smaller than the equivalent noise

Figure 2.22: Autocorrelation and Cross-Correlation Functions for $q_1(t)$ and $q_2(t)$.
[17]

bandwidth of $n_e(t, \epsilon_d)$, the equivalent single sided noise spectral density of $n_e(t, \epsilon_d)$ can therefore be written as [17]

$$N_e(\epsilon_d) = 2 \int_{-\infty}^{\infty} R_e(\tau, \epsilon_d) d\tau \quad (2.2.2.2 - 21)$$

According to reference [17], if equations (2.2.2.2-20a-d) are substituted into equation (2.2.2.2-21) and integrated in the frequency domain, $N_e(\epsilon_d)$ can be rewritten as

$$\begin{aligned} N_e(\epsilon_d) = & 16 \int_{-\infty}^{\infty} [S_{\hat{N}}(f) * S_{\hat{N}}(f)] S_q(-f) df \\ & + 8 P_s f(\epsilon_d) \int_{-\infty}^{\infty} [S_{\hat{m}}(f) * S_{\hat{m}}(f)] S_q(-f) df \end{aligned} \quad (2.2.2.2-22a)$$

where the asterisk indicates convolution and

$$S_{\hat{N}}(f) = \mathcal{F} \{R_N(\tau)\} = \frac{N_0}{2} |H_t(j2\pi f)|^2 \quad (2.2.2.2 - 22b)$$

$$S_{\hat{m}}(f) = \mathcal{F} \{R_m(\tau)\} = S_m(f) |H_t(j2\pi f)|^2 \quad (2.2.2.2 - 22c)$$

$$S_q(f) = \mathcal{F} \{R_q(\tau)\} = 1/4 \delta(f) + 1/4 \sum_{\substack{n=-\infty \\ n \text{ odd}}}^{\infty} \left(\frac{2}{n\pi}\right)^2 \delta\left(f - \frac{n}{T_d}\right) \quad (2.2.2.2 - 22d)$$

If equation (2.2.2.2-22b-c) are substituted into equation (2.2.2.2-22a), reference [17] states that N_{ϵ_d} can be expressed as

$$N(\epsilon_d) = \frac{P_s N_0}{2} \left[2M'_4 f(\epsilon_d) + \frac{2K'_L}{\rho_H} \right] \quad (2.2.2.2 - 23a)$$

where

$$M'_4 = M_4 + 2 \sum_{n=1,3,5,\dots}^{\infty} \left(\frac{2}{n\pi} \right)^2 M_{4n} K_{Ln} \quad (2.2.2.2 - 23b)$$

$$K'_L = K_L + 2 \sum_{n=1,3,5,\dots}^{\infty} \left(\frac{2}{n\pi} \right)^2 K_{Ln} \quad (2.2.2.2 - 23c)$$

$$M_{4n} \triangleq \int_{-\infty}^{\infty} S_m(f) |H_t(j2\pi f)|^2 |H_t[j2\pi(\frac{n}{T_d} - f)]|^2 df \quad (2.2.2.2 - 23d)$$

$$K_{Ln} = \frac{\int_{-\infty}^{\infty} |H_t(j2\pi f)|^2 |H_t[j2\pi(\frac{n}{T_d} - f)]|^2 df}{\int_{-\infty}^{\infty} |H_t(j2\pi f)|^2 df} \quad (2.2.2.2 - 23e)$$

$$K_L \triangleq \frac{\int_{-\infty}^{\infty} |H_t(j2\pi f)|^4 df}{\int_{-\infty}^{\infty} |H_t(j2\pi f)|^2 df} \quad (2.2.2.2 - 23f)$$

$$M_4 \triangleq \int_{-\infty}^{\infty} S_m(f) |H_t(j2\pi f)|^4 df \quad (2.2.2.2 - 23g)$$

$$\rho_H = \frac{P_s}{N_0 B_H} \quad (2.2.2.2 - 23h)$$

$$B_H = \int_{-\infty}^{\infty} |H_t(j2\pi f)|^2 df \quad (2.2.2.2 - 23i)$$

ρ_H gives the signal to noise ratio in the bandwidth B_H , which is the two-sided noise bandwidth of the equivalent low-pass filter $H_t(s)$. Recalling that $H_t(s)$ is the low-pass equivalent of the bandpass filter $H(s)$ (Figure 2.19), it can be seen that ρ_H is essentially the signal to noise ratio at the output of the bandpass filter.

When the equivalent loop signal to noise ratio is large and it is assumed that $\frac{d\epsilon_d}{dt} = 0$, the linear tracking performance for the TDL can be obtained from equation (2.2.2.2-16) with $D(\epsilon_d)$ replaced with ϵ_d (note that strictly speaking, the linear analysis is valid over $|\epsilon_d| \leq 1/2$ since $D(\epsilon_d) = \epsilon_d$ for $|\epsilon_d| \leq 1/2$). The mean squared tracking jitter can therefore be expressed as

$$\sigma_\epsilon^2 = \overline{\epsilon_d^2} = \frac{\overline{N_e(\epsilon_d)} B_L}{(P_s M_2)^2} \quad (2.2.2.2 - 24)$$

Using equations (2.2.2.2-20d) and (2.2.2.2-23a) in equation (2.2.2.2-24) results in

$$\sigma_\epsilon^2 = \frac{1}{2\rho} \left\{ \frac{M_4' + \frac{2K_L'}{\rho_H}}{M_2^2 \left[1 - \frac{2}{\rho} \left(\frac{M_4'}{M_2^2} \right) \right]} \right\} \quad (2.2.2.2 - 25)$$

where $\rho = P_s/N_0 B_L$ is the equivalent loop signal to noise ratio. Since ρ is assumed large, σ_ϵ^2 can be approximated as

$$\sigma_\epsilon^2 \approx \frac{1}{2\rho} \left\{ \frac{M_4' + \frac{2K_L'}{\rho_H}}{M_2^2} \right\} \triangleq \frac{1}{2\rho S_L} \quad (2.2.2.2 - 26)$$

where S_L is the squaring loss of the TDL, which is simply a ratio given by

$$S_L = \frac{\text{signal} \times \text{signal}}{\text{signal} \times \text{noise} + \text{noise} \times \text{noise}} \quad (2.2.2.2 - 27)$$

In terms of the data symbol signal to noise ratio $E_s/N_0 = P_s T_s/N_0$ and the ratio of the bandwidth B_H to data rate R_s , S_L is

$$S_L = \frac{M_2^2}{M_4' + K_L' [(2B_H/R_s)/(E_s/N_0)]} \quad (2.2.2.2 - 28)$$

Table 2.1: Closed Form Results for M_2 and M_4 [17]

Evaluation of M_2 for one- and two-pole Butterworth filters	
	$(\text{Signal} \times \text{Signal Distortion})^2, M_2 = H_2(f/2\pi) = \frac{1}{1 + (f/f_c)^2}$ $\text{NRZ: } S_{\text{NRZ}}(f) = f \frac{\sin \pi f T}{(\pi f T)}$
$n = 1$	$1 - \frac{1}{2B_H/R} [1 - \exp(-2B_H/R)]$
$n = 2$	$1 - \frac{1}{4B_H/R} [1 - \exp(-2B_H/R) \{ \cos(2B_H/R) + \sin(2B_H/R) \}]$
	Manchester Code: $S_M(f) = f \frac{\sin^2 \pi f T/2}{(\pi f T/2)}$
$n = 1$	$1 - \frac{1}{2B_H/R} [1 - 4 \exp(-B_H/R) + \exp(-2B_H/R)]$
$n = 2$	$1 - \frac{1}{4B_H/R} [3 - 4 \exp(-B_H/R) \{ \cos(B_H/R) + \sin(B_H/R) \} + \exp(-2B_H/R) \{ \cos(2B_H/R) + \sin(2B_H/R) \}]$

Evaluation of M_4 for one- and two-pole Butterworth filters	
	$\text{Signal} \times \text{Noise Distortion}, M_4 = H_4(f/2\pi) = \frac{1}{1 + 6(f/f_c)^2 + (f/f_c)^4}$ NRZ:
$n = 1$	$1 - \frac{3 - 4(1 + 2B_H/R) \exp(-2B_H/R)}{4B_H/R}$
$n = 2$	$1 - \frac{5 - \{4B_H/R \cos(2B_H/R) + 5[\cos(2B_H/R) + \sin(2B_H/R)]\} \exp(-2B_H/R)}{16B_H/R}$
	Manchester Code:
$n = 1$	$1 - \frac{9 - 4(3 + B_H/R) \exp(-B_H/R) + (3 + 2B_H/R) \exp(-2B_H/R)}{4B_H/R}$
$n = 2$	$1 - \frac{15 - \{8B_H/R \cos(B_H/R) + 2[\cos(B_H/R) + \sin(B_H/R)]\} \exp(-B_H/R) + \{4B_H/R \cos(2B_H/R) + 5[\cos(2B_H/R) + \sin(2B_H/R)]\} \exp(-2B_H/R)}{16B_H/R}$

Although Table 2.1 shows closed form results for M_2 and M_4 (which correspond to one and two pole Butterworth filters and NRZ coded data) and K_L and B_H for an n -pole Butterworth filter are given as

$$K_L = \frac{n-1}{n}$$

$$B_H = 2f_c \left[\frac{\left(\frac{\pi}{2n}\right)}{\sin\left(\frac{\pi}{2n}\right)} \right] \quad (2.2.2.2 - 29)$$

where f_c is the 3-dB cutoff frequency of the equivalent low-pass filter $H_L(j2\pi f)$, it is still difficult in general to evaluate the integrals for M_{4m} and K_{LM} . Therefore, the squaring loss S_L is not easily obtained.

Using equations (2.2.2.2-26) and (2.2.2.2-28), σ_e^2 can be written in terms of B_H/R_s and E_s/N_0 such that

$$\sigma_e^2 = \frac{1}{2\rho} \left[\frac{M_4' + K_L' \left(\frac{2B_H/R_s}{E_s/N_0} \right)}{M_2^2} \right] \quad (2.2.2.2 - 30)$$

It should be noted that for each value of (E_s/N_0) , there exists an optimum ratio of filter bandwidth B_H to data rate R_s (B_H/R_s) in the sense of minimizing the loop's tracking jitter. For low SNRs, Holmes [15] states that the optimum 3-dB bandwidth for NRZ symbols is about $1.4 R_s$.

Since it is difficult to evaluate the integrals for K_{2m} and M_{4n} and thus obtain σ_e^2 , it is at times more convenient to use the result for the TDL normalized mean-squared error obtained by Hartman [22], that is

$$\sigma_e^2 = \frac{1}{\rho} \left[0.905 + \frac{4B_H/R_s}{E_s/N_0} \left(0.453 - \frac{0.2}{B_H T_d} \right) \right] \quad (2.2.2.2 - 31)$$

In his analysis, Hartman ignored the bandlimiting effect of the bandpass filter and therefore obtained a result which, although simpler to work with, is about 0.9 dB optimistic [17].

Although the more difficult nonlinear analysis for the TDL tracking performance was not presented here, Simon [23] shows that it can be done using Fokker-Planck methods. In reference [23], it can be seen (through some plots), that for a first order loop, the linear and non linear analysis are practically equal for tracking jitter less than 0.1 and R_s/B_L greater than about 400.

Even though the TDL loop requires about 1 dB more signal-to-noise ratio than the DLL for equal rms tracking jitters [22], its hardware simplicity relative to the DLL (one signal processing arm completely eliminated), makes it very desirable. Moreover, it should be noted that problems such as gain imbalance and other mismatches present in a two-channel loop like the DLL (which would degrade the loop performance if not properly attended to), are not present in the TDL.

2.2.2.3 Loop Acquisition (Transient) and Mean Time to Lose Lock for the TDL

The following paragraphs will address the transient response of the TDL with an emphasis on its acquisition behavior when there is an offset between the incoming PN code and the local code and follows the analysis outlined in references [15] and [17]. In considering the transient response of the TDL, two important questions will be considered:

1. What is the maximum relative code rate offset (due to Doppler, clock instabilities, etc.) between the received and locally generated PN codes that would still allow the received signal to be acquired?

and ω_n is the radian natural frequency of the loop. Holmes [15] has shown that the linear closed loop transfer function for a second order loop with a filter transfer function of $F(s) = \frac{1+r_2s}{1+r_1s}$ can be written as

$$H(s) = \frac{GF(s)}{s + GF(s)} \approx \frac{1 + 2\xi s/\omega_n}{1 + 2\xi s/\omega_n + (s/\omega_n)^2} \quad (2.2.2.3 - 2)$$

where ξ is the loop's damping factor. Thus, $gF(s)$ can be written as

$$gF(s) = \frac{1 + 2\xi s/\omega_n}{1/g + s/\omega_n}$$

or

$$gF_0(p_0) = \frac{1 + 2\xi p_0}{1/g + p_0} \quad (2.2.2.3 - 3)$$

Substituting equation (2.2.2.3-3) into equation (2.2.2.3-1) results in

$$\frac{1}{g}\dot{y} - \frac{1}{g}\dot{x} + \bar{y} - \bar{x} = D(x) \quad (2.2.2.3 - 4)$$

From the fact that

$$\frac{\bar{x}}{\dot{x}} = \frac{dx}{dx} \quad (2.2.2.3 - 5)$$

equation (2.2.2.3-4) can be written as [15]

$$\frac{d\dot{x}}{dx} = \frac{-D(x) - 2\xi \left[\frac{d}{dx} D(x) \right] \dot{x} + \bar{y} - (1/g)(\dot{x} - \dot{y})}{\dot{x}} \quad (2.2.2.3 - 6)$$

Equation (2.2.2.3-6) represents the defining equation for the phase plane and for a constant search velocity or code rate offset ($\dot{y} = 0$), it reduces to [15]

$$\frac{d\dot{x}}{dx} = \frac{-D(x) - 2\xi \left[\frac{d}{dx} D(x) \right] \dot{x} - (1/g)(\dot{x} - \dot{y})}{\dot{x}} \quad (2.2.2.3 - 7)$$

The plots generated by this equation (for example, Figure 2.23) can be used to determine the maximum code rate offset $\dot{y}$ that will still allow a trajectory to reach the $(0, \dot{y}/g)$ point in the $(\dot{x}, x)$ plane - i.e., where the TDL phase and frequency locks. For $g > 10$, the loop gain g has little effect on the trajectories. This is clearly seen in Figure 2.23 for $g = 10$ and $g = \infty$. Assuming that $g = \infty$, it can be seen from Figure 2.23 that for $\xi = \frac{1}{\sqrt{2}}$, $\bar{y} = 0$, the maximum relative code rate offset (code drift) that will still result in acquisition is $\dot{y} = 1$, while the maximum value of $\dot{x}$ that results in acquisition is $\dot{x} \approx 1$. Thus,

$$\begin{aligned} \dot{x} &= \frac{1}{\omega_n} \frac{dx}{dt} \approx 1 \Rightarrow \frac{dx}{dt} = \omega_n \\ \dot{y} &= \frac{1}{\omega_n} \frac{dy}{dt} = 1 \Rightarrow \frac{dy}{dt} = \omega_n \end{aligned} \quad (2.2.2.3 - 8)$$

Therefore, the code loop frequency error, $\frac{dx}{dt}$ (chips/sec) and the relative code rate offset $\frac{dy}{dt}$ must be less than ω_n for successful code acquisition. The time to acquire lock is defined as the time it takes the normalized timing error x to settle to a very small value. From Figure 2.24, for $g = \infty$, this time can be approximated as $t = \frac{6}{\omega_n}$.

The mean time to lose lock is an important quantity to know. Usually the loop is said to have lost lock when the timing error exceeds $\pm 3/2$ chips (beyond this point, the cross correlations with the early and late signals is zero). In [15], Holmes has derived the mean time to lose lock for a first-order noncoherent tracking loop.

Figure 2.23: Acquisition Trajectories for a "One-delta" Noncoherent Loop: (a) Loop gain $g = \infty$, (b) Loop gain $g = 10$. [15]

Figure 2.24: Transient Response of the Normalized Timing Error $x(\omega_n t)$ for a "One-delta" Noncoherent Loop with $g = \infty$. [15]

Figure 2.25 illustrates the results for the DLL which is also applicable to the TDL if σ_e^2 is appropriately modified. Holmes [15] indicates that these results can be used for a second order loop by degrading the loop SNR by 1 dB (i.e., increasing σ_e^2 by 1 dB). From Figure 2.25 and the above considerations, for the case of the TDL, it can be seen that in general, for a tracking jitter (σ_e) less than 0.1, $\log 2B_L\bar{T}$ is high. This implies that $\bar{T}$, the mean time to lose lock or slip will also be high. However, for tracking jitters greater than 0.2, the mean time to lose lock would be very short, resulting in excessive loss of lock.

2.2.3 Carrier Acquisition and Tracking

After PN code synchronization has been achieved, carrier synchronization (in coherent demodulation systems) is initiated. The choice of coherent demodulation systems is usually motivated by its superior error rate performance.

The carrier synchronization process will be examined by first considering the phase-locked loop, the heart of a carrier synchronizer. Following this, the equations that characterize the acquisition and tracking performance of the synchronizer will be developed, paralleling the analysis done by Blanchard in reference [24]. In systems where the carrier is not readily accessible (suppressed carrier systems), the synchronizer must first regenerate the carrier before locking on and tracking it. In conclusion therefore, one particularly simple method of recovering the carrier in carrier suppressed systems will be discussed.

2.2.3.1 The Phase-Locked Loop

Although Blanchard [24] defines the phase-locked loop (PLL) as "a device by means of which the phase of a frequency modulated oscillator output signal is made to follow that of the input signal", it is perhaps functionally better described as a device

Figure 2.25: Normalized Mean Slip Time vs Jitter.. [15]

by means of which the "frequency" of a frequency modulated oscillator output signal is made to follow that of the input signal. The equivalence of these two definitions is clearer if you consider that frequency is the rate at which a signal phase changes with respect to time. Thus, as long as the input and oscillator frequencies are different, the PLL will continually try to match the oscillator rate of phase change with that of the input signal. If the input frequency is sufficiently close to the initial oscillator frequency, the oscillator frequency will eventually be identical to the input signal, except for a finite phase difference. In any case, considering the basic diagram of a phase-locked loop shown in Figure 2.26, the input signal, $y_i(t)$ and the VCO output signal $y_0(t)$ can be written as

$$y_i(t) = A \cos(\omega_i t + \theta_i) \quad (2.2.3.1 - 1a)$$

$$y_0(t) = B \cos(\omega_0 t + \theta_0) \quad (2.2.3.1 - 1b)$$

where ω_i is the angular frequency of the input signal and ω_0 is the VCO central angular frequency. For a sinusoidal phase detector, the output signal of the phase detector $u_1(t)$ would be given by

$$u_1(t) = K_1 \cos[(\omega_i - \omega_0)t + \theta_i - \phi_0] \quad (2.2.3.1 - 2)$$

where K_1 is the phase detector sensitivity. Since $y_i(t)$ and $y_0(t)$ are taken to be initially out of synchronism, $u_1(t)$ will be a sinusoid with a nonzero angular frequency equal to $(\omega_i - \omega_0)$. If the initial frequency difference is within the PLL's acquisition range, the VCO output signal $y_0(t)$ will eventually become synchronous with the input signal $y_i(t)$. When this happens, $\omega_0 = \omega_i$ and $y_0(t)$ can thus be written as

Figure 2.26: Basic Diagram of a Phase-Locked Loop

2

MicroD

$$y_0(t) \doteq B \cos(\omega_i t + \psi_0) \quad (2.2.3.1 - 3)$$

which implies that ϕ_0 could be viewed as a linear function of time, namely

$$\phi_0 = (\omega_i - \omega_0)t + \psi_0 \quad (2.2.3.1 - 4)$$

Substituting this expression for ϕ_0 into equation (2.2.3.1-2), it can be seen that when $\omega_0 = \omega_i$ (i.e., $y_i(t)$ and $y_0(t)$ are synchronous), the error signal $u_1(t)$ would have become a dc signal given by

$$u_1(t) = K_1 \cos(\theta_i - \psi_0) \quad (2.2.3.1 - 5)$$

The low-pass loop filter (which normally passes the dc component of u_1) will now pass the dc signal u_1 so that,

$$u_2 = K_2 u_1 = K_2 K_1 \cos(\theta_i - \psi_0) \quad (2.2.3.1 - 6)$$

where K_2 is the gain of a possible amplifier between the loop filter and the VCO (note that it is signal u_2 that either increases or decreases the VCO frequency). The instantaneous angular frequency ω_{inst} of the frequency modulated VCO is a linear function of the command signal u_2 around ω_0 and is given by

$$\omega_{inst} = \frac{d}{dt}(\omega_0 t + \phi_0) = \omega_0 + K_3 u_2 \quad (2.2.3.1 - 7)$$

where K_3 is the VCO modulation sensitivity. Equating the respective terms in equation (2.2.3.1-7) results in,

$$\frac{d\phi_0}{dt} = K_3 u_2 \quad (2.2.3.1-8)$$

Substituting equations (2.2.3.1-4) and (2.2.3.1-6) into equation (2.2.3.1-8) results in

$$\omega_i - \omega_0 = K_3 K_2 K_1 \cos(\theta_i - \psi_0) \quad (2.2.3.1-9)$$

which can be rewritten as

$$\psi_0 = \theta_i - \arccos \left(\frac{\omega_i - \omega_0}{K_3 K_2 K_1} \right) \quad (2.2.3.1-10)$$

Note that equation (2.2.3.1-10) relates the final phase differences (i.e., when synchronous) between y_i and y_0 , $(\theta_i - \psi_0)$ and the initial frequency difference.

Using the preceding development, the PLL operation can briefly be described as follows. To start with, signals y_i and y_0 are assumed nonsynchronous. Signal y_i has frequency ω_i and phase θ_i ; while y_0 has frequency ω_0 and phase ϕ_0 . After some time of operating, y_0 becomes synchronous with signal y_i (the loop is said to be in lock). This means that y_i and y_0 now have the same angular frequency ω_i but a difference in phase given by $(\theta_i - \psi_0)$, as equation (2.2.3.1-10) shows. It is this difference in phase that produces the phase detector dc component output that is used to modulate the VCO; that is, the command signal u_2 which can (from the above equations) be written as

$$u_2 = K_2 u_1 = \frac{\omega_i - \omega_0}{K_3} \quad (2.2.3.1-11)$$

This equation basically relates the final dc command signal u_2 to the initial frequency difference between y_i and y_0 .

Referring to equation (2.2.3.1-10), it can easily be seen that if the initial frequency difference $\omega_i - \omega_0$ is much smaller than the loop gain $K_1 K_2 K_3$, then equation (2.2.3.1-10) can be reduced to $\theta_i - \psi_0 \approx \text{Arc cos } 0 = \pi/2$, implying that y_i and y_0 would be practically in phase quadrature. It should be noted that despite the validity of the preceding approximation, strictly speaking however, phase quadrature actually corresponds to a zero initial frequency offset ($\omega_i = \omega_0$).

In order to obtain the general linearized transfer function of the PLL, the signals $y_i(t)$ and $y_0(t)$ are first represented as

$$y_i(t) = A \sin[\omega t + \phi_i(t)] \quad (2.2.3.1 - 12a)$$

$$y_0(t) = B \cos[\omega t + \phi_0(t)] \quad (2.2.3.1 - 12b)$$

Note that since the two signals do not necessarily have the same angular frequency, any difference in frequency is accounted for in $\phi_i(t) - \phi_0(t)$. Thus, if $\phi_i(t) = \phi_0(t)$, then y_i would be in phase quadrature with y_0 . Now, for a sinusoidal phase detector, the error signal u_1 can be written as

$$u_1(t) = K_1 \sin[\phi_i(t) - \phi_0(t)] \quad (2.2.3.1 - 13)$$

This means that $u_2(t)$ is then

$$u_2(t) = K_2 u_1(t) * f(t) \quad (2.2.3.1 - 14)$$

where $f(t)$ is the impulse response of the loop filter and the symbol $*$ denotes convolution. If equations (2.2.3.1-8), (2.2.3.1-13) and (2.2.3.1-14) are combined, the general time equation that governs the loop behaviour can be expressed as

For the noise equivalent bandwidths of other loops and filter types, reference [24] should be consulted.

2.2.3.3 The Squaring Loop

As was earlier mentioned, a conventional PLL cannot be used for suppressed carrier systems, since the carrier must first be regenerated before it is acquired and tracked. One of the simplest methods used for suppressed carrier tracking is the squaring loop shown in Figure 2.29. Assuming that binary signalling with NRZ waveforms (BPSK) is used, then the received signal (neglecting noise) could be written as [25]

$$V_r(t) = m(t) \sin(\omega_i t + \theta_i) \quad (2.2.3.3 - 1)$$

where ω_i is the carrier frequency, $m(t)$ represents the baseband data stream (± 1) and θ_i is the unknown signal phase. The operation of the squaring loop can thus be described as follows. The received signal is bandpass filtered, squared to remove the modulation, thereby resulting in

$$V_s = m^2(t) \sin^2(\omega_i t + \theta_i) = \frac{1}{2} m^2(t) [1 - \cos(2\omega_i + 2\theta_i)] \quad (2.2.3.3 - 2)$$

This result is bandpass filtered at $2\omega_i$ for PLL tracking. To obtain the coherent reference for demodulation, the VCO output is phase shifted 90° (to remove phase quadrature between phase detector input and VCO output) and then frequency halved.

Gardner [25] has analyzed the squaring loop performance and gives its VCO phase jitter σ_{ϕ}^2 (for a BPSK signal) as

Using equation (2.2.3.1-18b) in (2.2.3.1-17b), the error function of the PLL, $\Phi(j\omega)/\Phi_i(j\omega)$ can, after some simple algebraic manipulations, be written as

$$\frac{\Phi(j\omega)}{\Phi_i(j\omega)} = \frac{j\omega}{j\omega + KF(j\omega)} = 1 - H(j\omega) \quad (2.2.3.1 - 19)$$

2.2.3.2 Acquisition and Tracking

From the transfer and error functions of the PLL just derived, the acquisition and tracking performance for the phase-locked loop can subsequently be obtained. For a second order loop with a loop filter which has a transfer function given by $F(j\omega) = 1 + j\omega\tau_2/1 + j\omega\tau_1$ (Figure 2.27), Blanchard [24] has analyzed the natural acquisition of the input signal. The acquisition range obtained for such a loop if a sinusoidal phase detector, with a characteristic (dc error voltage versus phase error) as shown in Figure 2.28 is

$$\Omega_{acq} = 2\sqrt{K\xi\omega_n} \quad (2.2.3.2 - 1)$$

where Ω_{acq} is the acquisition range, K is the loop gain, ξ is the loop damping constant and ω_n is the loop natural angular frequency. The corresponding synchronization range Ω_{syn} is given as

$$\Omega_{syn} = K\pi/2 \quad (2.2.3.2 - 2)$$

By using phase plane techniques, an approximate time to achieve acquisition was obtained by Viterbi. This is stated in [24] as

$$T_{acq} \approx \frac{\Omega_0^2}{2\xi\omega_n^3} \quad (2.2.3.2 - 3)$$

Figure 2.27: Loop Filter for a Second Order Loop with a Transfer Function Given by $F(j\omega) = 1 + j\omega\tau_2 / 1 + j\omega\tau_1$

Figure 2.28: Sinusoidal Phase Detector Characteristic. [25]

where Ω_0 is the initial frequency difference (in radians/sec) between the input and the VCO central angular frequency. Note that for acquisition to occur, Ω_0 must be subject to the limits imposed by equation (2.2.3.2-1).

To obtain the tracking performance of the PLL, Blanchard [24] first expresses the input and output signals $y_i(t)$ and $y_o(t)$ as

$$y_i(t) = A \sin(\omega t + \theta_i) + n(t) \approx A \sin(\omega t + \theta_i + n'(t)) \quad (2.2.3.2 - 4a)$$

$$y_o(t) = B \cos[\omega + \varphi_o(t)] \quad (2.2.3.2 - 4b)$$

where $n'(t)$ is an equivalent modulation or representation of the influence of noise $n(t)$ on the signal and $\varphi_o(t)$, the instantaneous phase error, is the output signal of a filter with a transfer function $H(j\omega)$ given by equation (2.2.3.1-17) and driven by the process $n'(t)$ ($n'(t)$ is a stationary Gaussian noise with a mean value of zero and a uniform spectral density of N_0/A^2 (rad^2/Hz) in the band from $-W/2$ to $+W/2$).

From the above, it therefore follows that $\varphi_o(t)$ is a stationary Gaussian noise with a zero mean value and a power spectral density S_{φ_o} (in rad^2/Hz) from $-W/2$ to $W/2$ given by [24]

$$S_{\varphi_o}(f) = \frac{N_0}{A^2} |H(j2\pi f)|^2 \quad (2.2.3.2 - 5)$$

The variance consequently is written as

$$\sigma_{\varphi_o}^2 = \frac{N_0}{A^2} \int_{-W/2}^{+W/2} |H(j2\pi f)|^2 df \quad (2.2.3.2 - 6)$$

Because $|H(j2\pi f)|^2 \approx 0$ for $|f| > W/2$, the variance can thus be expressed as

$$\sigma_{\varphi_0}^2 = \frac{N_0}{A^2} \int_{-\infty}^{\infty} |H(j2\pi f)|^2 df \quad (2.2.3.2 - 7)$$

The noise equivalent bandwidth of a loop with a transfer function given by $H(j\omega)$ is expressed in [24] as

$$(2B_n) = \int_{-\infty}^{\infty} |H(j2\pi f)|^2 df \quad (2.2.3.2 - 8)$$

Note that the above definition is reserved for the case where the noise spectral density is double sided; while in the case of single sided noise density, B_n is used instead, where B_n is given by

$$B_n = \int_0^{\infty} |H(j2\pi f)|^2 df \quad (2.2.3.2 - 9)$$

Using the expression for $(2B_n)$, the variance can thus be written as

$$\sigma_{\varphi_0}^2 = \frac{N_0}{A^2} \int_{-\infty}^{\infty} |H(j2\pi f)|^2 df = \frac{N_0}{A^2} (2B_n) = \frac{N_0}{2P_s} (2B_n) = \frac{N_0 B_n}{P_s} \quad (2.2.3.2 - 10)$$

where P_s is the signal power. The equivalent SNR of the loop filter bandwidth can be expressed as

$$\rho_L = \frac{P_s}{N_0 B_n} = \frac{1}{\sigma_{\varphi_0}^2} \quad (2.2.3.2 - 11)$$

For the second order loop (where $F(j\omega)$ is a lowpass filter with phase lead correction - Figure 2.27), the noise equivalent bandwidth can be written as [24]

$$2B_n = \frac{\omega_n}{4\xi} (1 + 4\xi^2) \quad (2.2.3.2 - 12)$$

For the noise equivalent bandwidths of other loops and filter types, reference [24] should be consulted.

2.2.3.3 The Squaring Loop

As was earlier mentioned, a conventional PLL cannot be used for suppressed carrier systems, since the carrier must first be regenerated before it is acquired and tracked. One of the simplest methods used for suppressed carrier tracking is the squaring loop shown in Figure 2.29. Assuming that binary signalling with NRZ waveforms (BPSK) is used, then the received signal (neglecting noise) could be written as [25]

$$V_r(t) = m(t) \sin(\omega_i t + \theta_i) \quad (2.2.3.3 - 1)$$

where ω_i is the carrier frequency, $m(t)$ represents the baseband data stream (± 1) and θ_i is the unknown signal phase. The operation of the squaring loop can thus be described as follows. The received signal is bandpass filtered, squared to remove the modulation, thereby resulting in

$$V_s = m^2(t) \sin^2(\omega_i t + \theta_i) = \frac{1}{2} m^2(t) [1 - \cos(2\omega_i + 2\theta_i)] \quad (2.2.3.3 - 2)$$

This result is bandpass filtered at $2\omega_i$ for PLL tracking. To obtain the coherent reference for demodulation, the VCO output is phase shifted 90° (to remove phase quadrature between phase detector input and VCO output) and then frequency halved.

Gardner [25] has analyzed the squaring loop performance and gives its VCO phase jitter σ_{ϕ}^2 (for a BPSK signal) as

Figure 2.29: Squaring Loop. [25]

$$\begin{aligned}
\sigma_{\phi}^2 &= 4 \left(\frac{2B_n N_0}{A^2} \right) \left(1 + \frac{N_0 B_i}{A^2} \right) \\
&= 4 \left(\frac{B_n N_0}{P_s} \right) \left(1 + \frac{N_0 B_i}{2P_s} \right) \\
&= 4 \left(\frac{B_n N_0}{P_s} \right) \left(1 + \frac{1}{2\rho_i} \right) \\
&= 4 \left(\frac{B_n N_0}{P_s} \right) L_n \tag{2.2.3.3-3}
\end{aligned}$$

where B_i is the bandwidth of the input bandpass filter, $\rho_i = N_0 B_i / P_s$ is the signal to noise ratio in the input filter, N_0 is the spectral density of the white Gaussian input noise and L_n is the loss caused by noise intermodulation. From an examination of equations (2.2.3.2-10) and (2.2.3.3-3), it can be seen that the jitter variance of a squaring loop is $4(1 + 1/2\rho_i)$ times as large as that of an ordinary PLL tracking a pure carrier of the same amplitude in the same noise. The factor of 4 (factor of 2 for rms jitter) is due to the phase magnification effect of the frequency doubler since the amplitude of any phase disturbance at the fundamental frequency is doubled at the second harmonic. Because the phase magnification occurs only in the PLL at $2\omega_i$, it does not directly affect the recovered carrier phase at ω_i ; in fact, since the VCO frequency is divided by 2 in the squaring loop, the magnification is removed, so that the phase jitter of the carrier reference for demodulation will be

$$\sigma_{\phi}^2 = \frac{1}{4} \sigma_{\phi}^2 = \left(\frac{B_n N_0}{P_s} \right) L_n \tag{2.2.3.3 - 4}$$

Since the PLL however tracks the magnified phase, its locking ability is affected by the magnification so that a squaring loop could lose lock around a loop SNR of +6 dB whereas an ordinary PLL could hold lock down to 0 dB. The equations for the

• acquisition range, synchronization range and acquisition time derived previously for the conventional PLL still apply to the squaring loop with the understanding that the tracking is done at $2\omega_i$ and that any frequency uncertainty at ω_i is of course doubled at $2\omega_i$.

Chapter 3

SYSTEM CONSIDERATIONS AND DESIGN

Having presented a theoretical background for the spread spectrum system, the considerations that had to be taken into account in the system design (which naturally included the synchronizaton subsystem) will be discussed. The performance considerations for a DS SSMA system will first be briefly examined. After this, the different blocks of the transmitter and the receiver, plus the calculations used in their design, will be outlined.

3.1 PERFORMANCE CONSIDERATIONS FOR A DS SSMA SYSTEM

As was indicated in an earlier section, the DS SSMA system that will be designed here is based on the assumption of an additive white Gaussian noise channel. Considering the phase-coded (BPSK) SSMA system model of Figure 3.1 (also known as DS SSMA), one can see that for a completely time synchronized system with K_{users} , the time delays τ_k can be ignored (i.e., $\tau_k = 0$, for $k=1,2,\dots,K$). A time synchronous system however, would require a common timing reference for the K

Figure 3.1: Phase Coded (BPSK) SSMA System Model. [26]

transmitters as well as compensation for delays in the various transmission paths [26]. As Pursley [26] points out, this is generally not feasible. As a result, the system considered here will be asynchronous (nontime synchronous), that is, communication is assumed to be between independent users.

From the standpoint of error rate performance, the number of users is the limiting factor for a SSMA system. To understand why this is the case, a model for the effect of the other users in the receiver matched to the desired signal will be outlined, following the steps described in [27]. In laying out the model, it is assumed that there is perfect synchronization (phase, frequency, chip and bit epoch) between the received and local PN codes. It is further assumed that (1) there are many chips in each data bit period and (2) the PN code used is so long that $T_b \leq NT_c$, where T_b is the data period, N is the number of chips in one code period and T_c is the chip duration. When the code division sequences are samples of purely random sequences, reference [27] states that the above assumptions make it feasible to conclude that the effect of the other users can effectively be modelled as Gaussian random processes with power spectral densities that are essentially flat within the data bandwidth (recall from Section 2.1.2 that satisfying the condition that T_b be less or equal to NT_c allows the spectral density of the transmitted signals to be concealed in noise). Thus, a given receiver will therefore perceive the undesired transmissions as noise while despread the desired signal.

Because pseudorandom codes are used instead of purely random sequences, there is some partial-period cross correlation between the code sequence of the other users and that of the desired signal. The problem then is to determine if these partial-period correlations have a Gaussian distribution. Since it is very difficult to obtain a statistical description of the set of all partial-period cross correlation values for a class of code division sequences, estimates of the probability function were found instead. These estimates (upper and lower bounds) would be obtained from knowledge of the first L moments. Weber et al. [27] indicate that the moment

problem approach has already been applied to obtain upper and lower bounds on the probability functions for all cross and asynchronous autocorrelation values for sets of PN and Gold sequences. Because Gold codes have well bounded as well as lower cross-correlation values, they are very well suited for CDMA (code division multiple access) situations.

The results for the class of Gold codes of degree 13, where the partial-period correlation is over $M=60$ chips, is shown in Figure 3.2. These results are plotted on Gaussian probability paper, where the probability function for a Gaussian random variable is a straight line. The curves indicate the upper and lower bounds (of the probability function) using six and eight moments. Since the lines are close to a straight line, it can be concluded that for the set of Gold sequences of degree 13, the effect of the sum of the other users in the receiver matched to the desired signal can be effectively modeled as a Gaussian random variable [27]. Although the results on which the above conclusion is based does not include the effects of data (BPSK), Weber et al. [27] conjectured that if this effect were taken into consideration, the Gaussian assumption would be even more justified.

Thus, by assuming that the effect on the desired user - from the presence of other users - consist of additional broad band noise, Weber et al. [27] (with some other simplifying assumptions as outlined in [27]), obtained the following equation for the AWGN channel

$$(E_b/N_0)_n = \frac{E_b}{N_0 + \sum_{i=1}^{n-1} N_i} \quad (3.1 - 1)$$

where $(E_b/N_0)_n$ represents the actual signal energy per data bit to the one-sided power spectral density (PSD) for the receiver matched to the desired signal when n users are present; N_i is the equivalent one-sided PSD due to the i^{th} other user at the front end of a receiver matched to the desired signal and N_0 is the one-sided

Figure 3.2: Bounds on the Probability for the Set of Gold Sequences of Degree 13 with Partial Period Correlation of 60 Chips. [27]

noise PSD.

Note that N_i can be written as

$$N_i = \frac{P_i}{R_c} = P_i T_c = \alpha_i P_0 T_c \quad (3.1 - 2)$$

where

$\alpha_i = \frac{P_i}{P_0}$, is the power ratio of the i^{th} user to the desired user

P_i is the power of the i^{th} user

P_0 is the power of the desired user

R_c is the chip or code rate ($= 1/T_c$)

T_c is the chip time

Dividing the top and bottom of equation (3.1-1) by N_0 and replacing N_i by $\alpha_i P_0 T_c$ gives

$$\left(\frac{E_b}{N_0} \right)_n = \frac{(E_b/N_0)_1}{1 + \frac{1}{N_0} \sum_{i=1}^{n-1} \alpha_i P_0 T_c} \quad (3.1 - 3)$$

where $(E_b/N_0)_1 = E_b/N_0$ represents the bit energy to single sided noise density that would have been obtained for the single user case. By observing equation (3.1-3), it is clearly obvious that there is a reduction in $(E_b/N_0)_n$ because of the presence of the other users; in essence, $(E_b/N_0)_n < (E_b/N_0)_1$, for $n > 1$.

To maintain the same bit error probability when n users are present as is obtained when there is only a single user, the value of $(E_b/N_0)_n$ must be increased. Designating $(E_b/N_0)_R$ as the required single user E_b/N_0 necessary to make the n -user SNR, $(E_b/N_0)_n$ equal to $(E_b/N_0)_1$ results in [27]

$$(E_b/N_0)_n = \frac{(E_b/N_0)_R}{1 + \eta^{-1}(E_b/N_0)_R \left(\sum_{i=1}^{n-1} \alpha_i \right)} \quad (3.1 - 4)$$

where $\eta \triangleq T_s/T_c = R_c/R_s$ is defined as the ratio of the code rate R_c to the data rate R_s (note that T_s is the data period and that $P_0 T_c = P_0 T_s / \eta = E_b / \eta$). If $(E_b/N_0)_n$ is set equal to $(E_b/N_0)_1$, the required E_b/N_0 to achieve a specified (E_b/N_0) can then be solved as

$$(E_b/N_0)_R = \frac{(E_b/N_0)_1}{1 - \eta^{-1}(E_b/N_0)_1 \left(\sum_{i=1}^{n-1} \alpha_i \right)} \quad (3.1 - 5)$$

In terms of a degradation factor, DF (defined as the ratio of $(E_b/N_0)_R$ when n users are present to that needed to obtain a given probability of error when there is only a single user), equation (3.1-5) becomes [27]:

$$DF = \frac{(E_b/N_0)_R}{(E_b/N_0)_1} = \frac{1}{1 - \eta^{-1}(E_b/N_0)_1 \left(\sum_{i=1}^{n-1} \alpha_i \right)} \quad (3.1 - 6)$$

For n equal power users, $\alpha_i = 1$, for all i so that equation (3.1-6) becomes

$$DF = \frac{1}{1 - (n-1)\eta^{-1}(E_b/N_0)_1} \quad (3.1 - 7)$$

Only the case of equal power users will be considered here. For a consideration of unequal power users and also the improvement in performance attained by coding, reference [27] should be consulted.

To get an idea as to the maximum number of users that can be supported, equation (3.1-4) is rewritten for equal power users as

$$(E_b/N_0)_n = \frac{(E_b/N_0)_R}{1 + (n-1)\eta^{-1}(E_b/N_0)_R} \quad (3.1-8)$$

For large values of $(E_b/N_0)_R$, equation (3.1-7) becomes

$$\lim_{(E_b/N_0)_R \rightarrow \infty} (E_b/N_0)_n = \frac{\eta}{n-1} ; \quad n \geq 2 \quad (3.1-9)$$

If $(E_b/N_0)_n$ is set equal to $(E_b/N_0)_1$, it is then seen that the maximum number of users if an $(E_b/N_0)_1$ is desired, is given by

$$n = 1 + \frac{\eta}{(E_b/N_0)_1} \quad (3.1-10)$$

Only two users will be considered here, necessitating the implementation of two transmitters and one receiver.

3.2 SYSTEM PARAMETERS

In designing the spread spectrum system discussed below, several system parameters had to be considered. The choices and the reasons that led to them will now be discussed.

From the preceding discussions, it is clear that it is desirable to use as wide a spectrum as possible in spread spectrum transmissions. The FCC in the U.S.A. and the DOC in Canada have allocated a 40-MHz bandwidth in the 800 to 900 MHz frequency range for mobile radio telephone service. This 40 MHz band is separated into a 20 MHz base station transmit band from 870 to 890 MHz and a mobile radio 20 MHz transmit band from 825 to 845 MHz [3,28]. The SSMA system designed here had initially been expected to utilize the entire band from

870 to 890 MHz for transmission, therefore requiring a code rate of 10 MHz [for non-return-to-zero (NRZ) and BPSK modulation]. However, when the speed of the Gold code generator implemented on a chip was seen to be about 6 MHz, it was decided to operate at a code rate of 4 Mbits/sec, thus giving a spread bandwidth of 8 MHz. Since voice communications is considered here, a data rate of 64 Kbits/sec was considered necessary to achieve toll quality performance. This rate can be reduced to 32 Kbits/sec while maintaining the same performance by using a CVSD (continuously variable slope delta) modulator, as discussed later. Thus, with a 4 MHz chip rate and a 32 Kbits/sec data rate, N , the length of the code must then be greater or equal to 125 chips in order to conceal the spectral density of the transmitted signal in noise (Section 2.1.2). This means that the code generator must have $n \geq 7$ stages.

Although the code length N has little effect on the processing gain (if everything else such as chip rate and data rate are kept constant), it does however affect what Dixon [2] calls the index of discrimination, which is the difference between the autocorrelation when the code is correctly phased and when it is not. This parameter is important in detecting synchronization, and increases as N increases. Of course, the disadvantage in a large N is that it takes a longer time to attain synchronization. From a consideration of the above system characteristics, a 13 stage ($N=8191$) Gold code generator was chosen.

In order to carry out the calculations required for the acquisition, tracking and demodulation channels, the relative signal and noise levels at the input of the respective mixers must be approximated. The required signal levels can be obtained by working with the demodulation channel as shown in Figure 3.3. For voice communications, a bit error rate of at least 10^{-3} is usually required to attain an acceptable performance level. From Figure 6.36(b) of Holmes [15], it can be observed that with a bit energy to noise density ratio (E_b/N_0) of 7 dB, a $P_b \leq 10^{-3}$ is maintained for a steady state phase error (carrier tracking) of up to 3° and a σ_ϕ of

up to 9° (this assumes synchronous bit detection). Note that $\sigma_\phi = \sqrt{1/4\rho_s}$, where ρ_s is the equivalent loop SNR, which includes the squaring loss. Since asynchronous bit detection will be employed here, a degradation of about 1 dB is assumed, thus raising the required bit energy to noise density to about 8 dB. From [13], the signal to noise ratio could be expressed as

$$SNR = \frac{P_s}{N_0 B} = \frac{R E_b}{N_0 B} = \frac{E_b/N_0}{B/R} \quad (3.2 - 1)$$

where R is the bit rate, B is the bandwidth and P_s is the signal power. With a bit rate of 31.75 KHz and a noise bandwidth of 35.24 KHz, the signal to noise ratio at the output of the demodulation lowpass filter (Figure 3.3) becomes 7.5 dB. Now, it is further assumed that at most a 1 dB effective signal loss is introduced in the demodulation channel due to PN tracking error and that implementation losses due to non ideal components, modulation distortion, etc., would add up to a 4 dB loss. With a processing gain, PG of 21 dB [$10 \log(4 \times 10^6 / 31.75 \times 10^3)$], then the input signal to noise ratio (SNR_i) at the RF filter output can be calculated as

$$SNR_i + PG - 5dB = 7.5dB$$

$$SNR_i = -8.5dB \quad (3.2 - 2)$$

A final factor to consider is the multiple access situation. Since the system is to accomodate two users, the above results should be obtained with two users present. This means that values for the single user case would have to be a little higher in order to degrade to the above. From equations (3.1-5), (3.2-1) and (3.2.2), to attain a SNR_i of -8.5 dB when two users are present, a SNR_i of -8.2 dB must be available when only one user is present.

3.3 THE TRANSMITTER

Although two transmitters will be implemented, only the different parts of one of them will be examined in detail, as the two will be very similar. Any differences between these two transmitters will be addressed in the section called "Modifications for Second Transmitter".

3.3.1 System Description

Before examining the different blocks that make up the transmitter, a general description of the system will help in demonstrating how the different parts combine to function as a transmitter. Figure 3.4 shows a diagram of one of the implemented DS spread spectrum transmitters. The analog signal from the tape recorder/radio is passed through the CVSD filter and is then subsequently passed through the CVSD. The output of the CVSD constitutes the information data bits. The information bits are next passed into the encoder (the exclusive nor and D flip-flop combination). The purpose of the encoder is to resolve an ambiguity that is later encountered in the demodulation process.

The encoded data is now digitally added (modulo-2, i.e., exclusive-oring) to the pseudorandom output of the PN Gold code generator. Note that although the data rate is approximately 32 KHz, the output resulting from the digital mixing has a rate equal to the chip rate, in this case, 4 MHz. The digitally mixed signal is next level shifted, resulting in a bipolar signal. Using a coaxial line driver, the bipolar signal is used to modulate a carrier at f_c , thus producing a BPSK modulated signal. Note that the bandpass filter (BPF) indicated in Figure 3.4 was not used due to cost considerations.

Figure 3.4: Block Diagram of a Voice Input DS Spread Spectrum Transmitter.

The motivations for choosing a 4 MHz chip rate (for the code generator), and a carrier frequency of 880 MHz, etc., were touched upon in the section entitled "System Design Parameters". Looking at Figure 3.4, it is seen that the 31.75 KHz CVSD clock is divided down from the 4 MHz code generator clock. The reason for this is to ensure that the data transition times occur close to the same points as the transition times for the PN sequence binary symbols. (Ideally, the code and data clocks should be synchronized so that the code and data output transitions occur simultaneously. This was not rigorously pursued here). If this was not the case, it would be possible for anyone to read the data from a clean copy of the digitally mixed signal. As reference [29] indicates, this could be done by first estimating the PN sequence clock and then determining unscheduled transitions in the mixed sequence, which would be due to the data.

A discussion of the theory behind the Gold code generator, CVSD filter and CVSD modulator as well as the particular configurations implemented here (including the schematics) can be found in Appendices A-1.1 and A-1.3, respectively.

3.3.2 Clock and Encoding Circuits

Although not as important as the PN generator, there are other parts of the transmitter that nevertheless still deserve some discussion. These are the clock and encoding circuits.

Starting with the clock circuits, it should be recalled that the CVSD clock is divided down from the PN generator clock for reasons already discussed. This is done using the 74HC161 4-bit counters as the transmitter schematic of Figure 3.5 shows. Basically, the idea was to obtain approximately a 32 KHz output from a 4 MHz clock. In addition, a symmetrical output (50% duty cycle) was chosen since there is a minimum requirement on the CVSD clock width. Note that although a pulse rate of approximately 32 KHz would be obtained if one were to simply

detect each count of 125, this would not satisfy the CVSD minimum clock width requirement. Thus, instead of detecting the count of 125, the count of 63 was detected and subsequently used to toggle the output of a JK flip-flop (74HC112).

The 2 MHz clock for the CVSD filter is obtained by simply dividing the 4 MHz clock by the 74HC4017, a decade counter/divider that can divide the input frequency by up to a factor of 10. Note that the 4 MHz oscillator output had to be buffered to drive all the chips connected to it.

The encoding block (and its corresponding decoding block) consists of nor gates and D flip-flops as Figure 3.6 shows. As was earlier mentioned, its purpose is to resolve an ambiguity that crops up in demodulator circuits, which makes it impossible to decide whether a current bit is supposed to be a 1 or a 0 without more information. This ambiguity is fundamental to all phase shift modulation techniques. If the data is transmitted in N phases, there is an N phase ambiguity in the data recovery since the recovered carrier could have started in any of the N phases. To resolve the ambiguity in the present case, the 1, 0 streams are encoded in such a way that a zero produces a change in level, while a one maintains the same level. In this way a phase change will result if the data bit is a zero, while a one will produce no phase change. From this, the ambiguity is easily resolved. This type of encoding is generally referred to as differential encoding, which the circuit of Figure 3.6, obtained from reference [30], implements. Note that the resulting encoded data is also known as NRZ(I) or NRZ(S), where the I stands for inversion (i.e., level change) and S stands for space (i.e., zero).

3.3.3 Modifications for the Second Transmitter

Now that the major parts of the DS transmitter have been discussed in greater detail, several modifications made for the second transmitter will be outlined. Because the second transmitter was not designed for voice transmissions, there was

Figure 3.6: Encoding and Decoding Blocks for Data Ambiguity Resolution: (a) encoding circuit, (b) decoding circuit. [30]

no need for an analog input, a CVSD filter or a CVSD. Instead of speech signals, a pseudo-random bit sequence of length $2^6 - 1$ is transmitted. This pseudo-random sequence is generated by the block labeled "PN sequence generator" in Figure 3.7 and is clocked at approximately 32 KHz, as was the CVSD. Note that the division to obtain a 2 MHz clock for the CVSD filter is no longer required.

3.4 THE RECEIVER

Figure 3.8 shows a block diagram of the receiver system. Two things that should be noted are that (1) the input signal is first downconverted before it is filtered and (2) the modulated carrier reference is centered at approximately 70 MHz. The reason for this arrangement is that filters and amplifiers for the 800 MHz frequencies are quite expensive. By restricting the signals to be filtered to relatively lower frequencies, simple filters using capacitors and inductors can be constructed.

The use of a modulated carrier reference helps to avoid a potential problem with RF to IF feedthrough in nonideal mixers. The modulated carrier reference is filtered (in the same manner as the spread input RF signal) in an attempt to introduce the same amount of distortion as is introduced to the RF input by filtering. According to Holmes [15], this should help reduce the correlation loss that is introduced by filtering.

Since a detailed explanation of the theory of operation of a PN system has already been given in Chapter 2, the discussion here will focus on the design steps taken in implementing the different parts of the receiver. Starting with a brief description of the receiver Gold code generator and the 8.003 MHz bandpass filters for the input and modulated carrier reference filtering, the discussion will then proceed to an examination of the major system blocks, namely, the acquisition, the tracking and the demodulation channels.

Figure 3.8: Block Diagram of Receiver System.

3.4.1 The Receiver Gold Code PN Generator

The block diagram of the Gold code generator employed in the receiver appears in Figure 3.9. It can be noted that aside from the addition of some flip-flops and an inverter, this is essentially the same code generator utilized in the transmitters.

The additional hardware employed in the receiver generator enables it to produce three identical but phase shifted outputs. These three code sequences are the early (PN-E), the punctual (PN-P) and the late (PN-L) sequences (Figure 3.9). The on-time or punctual sequence is used in the correlations for the acquisition and demodulation channels. The early and late sequences (which are half a chip ahead and behind the punctual sequence, respectively) are used in the correlation for the tracking channel.

3.4.2 Input and Reference Bandpass Filters

In designing the 8.003 MHz input and modulated carrier reference bandpass filters (BPFs), a few issues had to be considered. From the discussion by Holmes [15], signal phase distortion would cause not only a correlation loss but would also cause the code tracking loop to track with a timing error bias. Thus, filters with linear phase responses (such as Bessel filters) should be used. In order to keep the hardware as simple as possible, only a second order bandpass filter was designed. This corresponds to a first order low pass equivalent.

In designing BPFs, the low pass equivalent is usually transformed to the BP equivalent. This transformation distorts the phase response of the filter, so that special procedures must be employed to allow the BPF to retain the same phase response (linear) possessed by its lowpass equivalent [31,32]. Note that these special techniques are not employed here. Rather, the approximately linear phase response obtained by transforming a first order lowpass Butterworth filter (which is the

Figure 3.9: Receiver Gold Code Generator.

Figure 3.10: Reference and Input Signal Bandpass Filters: (a) Reference filter, (b) Input filter.

Figure 3.10: Reference and Input Signal Bandpass Filters: (a) Reference filter, (b) Input filter.

(a). USED FOR THE ACQUISITION AND TRACKING REFERENCE FILTERS

(b). USED FOR THE INPUT AND DEMODULATION (REFERENCE) FILTERS

* NOTE THAT THE USE OF A 2-8pF TRIM CAPACITOR FOR FIGURE (b) WOULD HAVE BEEN PREFERABLE. THE 5.5-10pF TRIM CAPACITOR IN SERIES WITH THE 5pF CAPACITOR WAS USED AS A ROUGH APPROXIMATION. NOTE THAT THE RESPONSES OBTAINED FROM ALL FOUR FILTERS WERE PRACTICALLY IDENTICAL.

Figure 3.11: Actual Component Values for the Reference and Input Filters.

Figure 3.12: Acquisition Channel Schematic.

3.4.3.1 Filters

Two types of filters are used in the acquisition channel - an IF BPF and a lowpass filter. A 3-dB bandwidth of 67 KHz was chosen for this fourth order filter, corresponding approximately to a noise bandwidth of 74.4 KHz (this bandwidth will allow the signal energy to pass in the presence of some doppler). Note that larger bandwidths were tried in an attempt to keep the acquisition time low by keeping the dwell times small; however, because a lot more noise than expected was obtained on the circuit board, it was necessary to lower the bandwidth in order to increase the signal to noise ratio and thus allow the system to detect the signal present condition. The SNR at this filter is increased by about $10 \log \left(\frac{8 \times 10^6}{67 \times 10^3} \right) \approx 20.8 \text{ dB}$ over that at the input filter. This figure neglects correlation, implementation and modulation distortion losses which could reduce this increase down to about 17 dB or less.

Following conventional active filter design rules, a symmetrical response BPF with a center frequency of 600 KHz and 3-dB bandwidth of 67 KHz was designed. The fourth order lowpass Butterworth filter, following the envelope detector, filters out the double frequency terms. Its 3-dB bandwidth was designed for 67 KHz.

3.4.3.2 The Acquisition Channel Envelope Detector

The HP2835 diode is used as a square law envelope detector. The output voltage V_0 for a square law device can be related to its input voltage V_i by

$$V_0 = C_1 V_i^2 \quad (3.4.3.2 - 1)$$

where C_1 is a constant. Taking $20 \log_{10}$ of both sides results in

$$\begin{aligned}
20 \log_{10} V_0 &= 20 \log V_i^2 + 20 \log C_1 \\
&= 2 \{10 \log P_i\} + C_2 \\
&= 2 P_i(\text{dBm}) + C_2
\end{aligned}
\tag{3.4.3.2-2}$$

where C_2 is a constant and P_i is the input power. Equation (3.4.3.2-2) can be rearranged as

$$\log_{10} V_0 = 0.1 P_i(\text{dBm}) + C_3 \tag{3.4.3.2 - 3}$$

where C_3 is a constant. From equation (3.4.3.2-3), it seems that a square law device would have a slope of 0.1 if $(\log_{10} V_0)$ were plotted against $P_i(\text{dBm})$.

Now, if $(\log_{10} V_0)$ were plotted against $P_i(\text{dBm})$ for the HP2835 diode, it would be quite simple to find out in which region the diode behaved as a square law device by simply observing where the slope was approximately equal to 0.1; this then would be the region in which the diode should be made to operate in. From Figure 3.13, (which gives the characteristic for the acquisition channel envelope detector), it can be seen that its square law region lies in the area from $P_i = -19 \text{ dBm}$ to -8 dBm and is therefore the area in which the input signal to the diode must fall into.

3.4.3.3 The Integrate and Dump Filter

Following the square law envelope detection and filtering, the signal is then integrated by an integrate and dump filter. This is followed by some amplification so

that the integrated signal output is at a level that will allow the threshold detector voltage to be set at a reasonable value.

The component values for the integrate and dump filter (Figure 3.14) are determined from the acquisition variables. Considering the expression for the mean acquisition time (equation (2.2.1.1.2-3)), it can be seen that a lot of variables would have to be considered if the acquisition time is to be minimized. Since this system is a demonstration model, it was decided (after due consideration), to choose reasonable values for the parameters.

As has already been mentioned, the IF filter bandwidth for the acquisition channel had to be reduced to 74.4 KHz (from the original design value of 800 KHz). With $P_{FA1} = 10^{-6}$, $P_{FA2} = 10^{-6}$, $\tau_{D1} = 125\mu s$, $\tau_{D2} = 625\mu s$, $P_s/N_0 = 52.7dB$ Hz and using references [20] and Section 10.3.4 of [14] resulted in $P_{D1} \approx 0.75$, $P_{D2} \approx 1$. The rest of the parameters (see Section 2.2.1.1.2) were assumed as

$$\Delta f_c \leq \pm 300 \text{ chips/sec}$$

$$K = 5$$

$$N = 8191$$

$$\Delta T_c/T_c = 1/2$$

$$P_{HO} = 1$$

Using the equations outlined in Sections 2.2.1.1.2 and 2.2.1.1.4, the results in Table 3.1 were obtained. The resistor and capacitor values are determined from the dwell

* Note that $n=1E-3$

$V_{out}(mv)$	$P_{in}(dBm)$
0.05	-25.0
0.4	-24.0
1.3	-23.0
2.8	-22.0
3.8	-21.0
5.4	-20.0
7.6	-19.0
9.6	-18.0
12.4	-17.0
16	-16.0
20	-15.0
25	-14.0
31	-13.0
37	-12.0
45	-11.0
55	-10.0
64	-9.00
77	-8.00
89	-7.00
104	-6.00
122	-5.00
143	-4.00
165	-3.00
191	-2.00
220	-1.00
251	0.00

Figure 3.13: Acquisition Channel Envelope Detector Characteristic.

Figure 3.14: Integrate and Dump Filter.

Table 3.1: Acquisition Time Calculations

	Code Doppler (chips/s)		
	-300	0	+300
$\bar{T}_{acq}(sec)$	1.84	1.71	1.59
$\sigma_{acq}^2(sec)$	2.59	2.21	1.92

$\bar{T}_F$ (signal not present): 3.125 ms

$\bar{T}_F$ (signal present): mathematically infinite

time of $125\mu s$ ($\tau_{D1} = 125\mu s$). Note that in choosing the value for the capacitor, consideration must be given to the fact that a large capacitance would require a longer time to discharge.

3.4.3.4 The Searchlock Logic Unit

The searchlock logic unit, whose block diagram and schematic appear in Figures 3.15 and 3.16 respectively, coordinates the transition from acquisition to tracking, according to the algorithm of Figure 3.17. The design of the searchlock unit logic circuits begins with a description of the desired algorithm in terms of a state diagram (Figure 3.18); then, using conventional logic design procedures, the state diagram is subsequently transformed into logic circuits. In order to give an idea of how the circuits implement the algorithm of Figure 3.17, the different inputs and outputs

Figure 3.15: Block Diagram of Searchlock Unit.

Figure 3.16: Schematic of Searchlock Logic Unit.

Figure 3.17: Algorithm Implemented by the Searchlock Unit.

q0: ACQUISITION OR SEARCH MODE
FIRST DWELL PERIOD(125 US)

q1: VERIFY ACQUISITION & START TRACKING MODE
SWITCH TO SECOND DWELL PERIOD(625 US)

q2: TRACKING MODE

q3,q4,q5,q6: USED TO COUNT NUMBER OF CONSECUTIVE MISSES

Figure 3.18: Searchlock Unit in Terms of a State Diagram.

shown in Figure 3.15 will be briefly discussed.

The pulses F_{A2} and F_{B2} are used to time the dwell (integration) periods. While they are on, the integration capacitor discharges so that no integration occurs. During the times that they are low (off), the integrator integrates (refer to Chapter 2 for the acquisition channel operation). F_{A2} is five times as fast as F_{B2} so that its dwell period is 5 times as small. F_{A1} and F_{B1} are identical to F_{A2} and F_{B2} respectively, except that they occur slightly earlier. The reason for this is that the circuits driven by F_{A2} and F_{B2} depend on the state of an input (V_{td} - threshold level exceeded indicator) that must first be clocked in by F_{A1} or F_{B1} . Thus, during the first dwell time, F_{A1} clocks in the input V_{td} while F_{A2} is used to discharge the capacitor and to drive other circuits that determine the present state of the acquisition process. Once the threshold level is exceeded, F_{B1} and F_{B2} are used for a longer integration time (note that the threshold level is increased for the longer integration time).

Considering Figure 3.12, it can be seen that F_{A2} and F_{B2} are derived from F_{A2x} and F_{B2x} , respectively. Note that F_{A2x} could have been used directly; it was put through the same process as F_{B2x} for uniformity. The reason for the derivation process on F_{B2x} is as follows; because F_{B2x} is five times slower than F_{A2} , there are five F_{A2} pulses (dwell periods) between two F_{B2} pulses. Now, suppose that F_{B2x} is used instead of F_{B2} and that at the end of the period for the second F_{A2} pulse or dwell period, the threshold is exceeded. The system would automatically deselect the F_{A2} input while selecting the F_{B2x} input, at the same time raising the threshold level. Because this switch from a short to a longer integration period did not occur at the beginning of the longer dwell period (in fact, it occurred 2/5 into it), the new threshold will not be exceeded whether the signal is present or not. To get around this, the system was set up such that the first long dwell period (after switching to a longer integration period) is always ignored. This is done by inhibiting the first F_{B2x} pulse following the switch to a longer integration period. In this way, it is ensured

that no action is taken as a result of the shortened dwell period. In essence then, F_{B2} is identical to F_{B2x} , except for the pulse representing the shortened period that is ignored. In this manner the circuit operation remains faithful to the algorithm of Figure 3.17. Note that the four clock inputs, F_{A1} , F_{A2x} , F_{B1} and F_{B2x} are all derived from the same source as Figure 3.19 shows.

The V_{td} input is the threshold detector level indicator, which is clocked in by either F_{A1} or F_{B1} and is saved by a flip flop. The VCO input is at a frequency that is twice the required VCO frequency. The reason for this is that if the reference code has to be delayed by half a chip, all that is required is to inhibit this VCO input by one period (this inhibition is carried out by using F_{A2}). The output, "PN generator clock", is the VCO input signal divided by 2. The DISCH output is used to discharge the capacitor and is normally derived from F_{A2} or F_{B2} (F_{B2x} is used once only, when the first F_{B2} output following the switch to a longer integration time is inhibited). Finally, the LOCKIND output is used to indicate that acquisition has been attained. This initiates actions such as switching the acquisition threshold to the tracking threshold, starting the switching in of different tracking loops, changing the voltage applied to the VCO so that it is now driven by the tracking loop filter outputs, etc.

3.4.4 The Tracking Channel

The third row in Figure 3.8 and the schematic of Figure 3.20 both show the tracking channel. As in the acquisition channel cases, some of the constituent elements will now be further discussed.

Figure 3.19: Derivation of the Clocks Used for the Searchlock Unit.

Figure 3.20: Tracking Channel Schematic.

3.4.4.1 Filters

Three types of filters were used; the IF BPF, the lowpass filter (LPF) following the envelope detector and the loop filters. The loop filters will be discussed in the section entitled "Loop Bandwidth Control". Here a few words will be said about the BPF and the LPF.

The design of the IF BPF was similar to the acquisition channel IF BPF discussed earlier. The tracking channel BPF 3-dB bandwidth was chosen as 67 KHz. This choice was motivated by three factors:

1. The tracking jitter optimum bandwidth of $1.4 R_s$ can be varied over a 2 to 1 range with little degradation (R_s is the data rate).
2. Modulation loss decreases as bandwidth increases.
3. The signal energy has to be passed in the presence of doppler.

Although the bandwidths here are identical to the acquisition channel case, the signal to noise ratio following the IF BPF will not increase by the same amount (17 dB) over that at the RF filter output. This is because an additional loss of 6 dB must be factored in due to the fact that the signals that the tracking channel correlates with the incoming signal are half a chip ahead and behind, respectively.

The lowpass filter following the tracking channel envelope detector is identical in design and function to that of the acquisition channel LPF previously discussed.

3.4.4.2 Tracking Channel Envelope Detector

Similar to the acquisition channel, a HP2835 diode is used as a square law envelope detector. Figure 3.21, the tracking channel envelope detector characteristic, shows

that the square law region is from $P_i = -19$ dBm to $P_i = -8$ dBm.

3.4.4.3 Dithering Logic

In Chapter 2, the tau-dither loop was extensively examined. Recall that the input signal was alternately correlated with the early and late versions of the locally generated PN code, thereby giving rise to the name dither. The logic which is used to implement this action is shown in Figure 3.22. The 74HC153 multiplexer selects either the early or late signal according to the dithering signal $g(t)$. This dithering signal is also applied to the multiplier (Figure 3.20) preceding the multiple loop filters (this is after it has been voltage divided down to $\pm 1V$). Note that because of noise problems, it was necessary to filter this input to the multiplier.

Now, in order to realize a discriminator characteristic similar to that shown in Figure 2.21, it was decided that the signal, $g'(t)$, applied to the multiplier must be equal to $-1V$ when the early signal is selected while a $+1V$ should be applied when the late signal is selected. Choosing the multiplier components appropriately to realize a gain of approximately one ensured that the multiplier output, after being lowpass filtered would represent the difference between the squared late and early correlations ($X_L^2 - X_E^2$). Using Figure 2.21 and the envelope detector characteristic of Figure 3.21, the S-curve of Figure 3.23 was obtained. It should be noted that in order to get the system to track, it was necessary to reverse the early and late inputs to the 74HC153 of Figure 3.22, resulting therefore in the schematic of Figure 3.24. It is possible that the multiplier may have introduced a 180° phase shift in its output, thereby necessitating this switch of the early and late signal inputs.

* Note that $m=1E-3$

Vout(mv)	Pin(dBm)
0.0	-25.0
1.4	-24.0
2.5	-23.0
3.0	-22.0
4.3	-21.0
5.9	-20.0
8	-19.0
10	-18.0
13	-17.0
16	-16.0
19	-15.0
23	-14.0
28	-13.0
35	-12.0
42	-11.0
51	-10.0
51	-9.00
72	-8.00
84	-7.00
99	-6.00
113	-5.00
130	-4.00
148	-3.00
163	-2.00
191	-1.00
216	0.00

Figure 3.21: Tracking Channel Envelope Detector Characteristic.

Figure 3.22: Dithering Logic Schematic

* NOTE THAT ϵ_d HAS BEEN DENORMALIZED HERE
IN CONTRAST TO CHAPTER 2.

Figure 3.23: S-curve for the Tracking Channel.

Figure 3.24: Dithering Logic (With Early and Late Inputs Reversed).

3.4.4.4 Loop Bandwidth Control

As was discussed earlier in Chapter 2, multiple code loop bandwidths are usually employed in order to soften the transition between the acquisition and tracking modes (if different loop bandwidths are employed for acquisition and tracking). To determine the required tracking bandwidth, several factors must be considered. These are (refer to Chapter 2):

$B_L \leq 1/T_s$; so that modulation self noise can be ignored

$\sigma_e < 0.1$; so that loss of lock occurs infrequently

$B_L \ll 1/T_c$; so that code self noise can be ignored

$B_L \ll 1/T_d$; so that the effects of dithering can be ignored

$T_d > T_s$; to allow dithered signal to be passed by IF filter

In addition to the above, it is assumed that no more than 1 dB degradation should be introduced into the demodulation channel by the tracking jitter. The effect of the tracking jitter on the demodulation channel can be expressed in terms of a loss term L_T [21,33]

$$L_T = 10 \log(1 - 1.6\sigma_e) \quad (3.4.4.4 - 1)$$

In order to tolerate the errors introduced by use of nonideal elements, only a 0.25 dB degradation will be allowed in the design. This resulted in a $\sigma_e = 0.035$. Adhering to the constraints listed above, choosing $f_d = 1/T_d = 16 \text{ KHz}$ and using equation (2.2.2.2-31), resulted in

$$\sigma_e^2 \approx 1.01 \times 10^{-5} B_L \quad (3.4.4.4-2a)$$

Since the above expression was said (Chapter 2) to be optimistic by about 0.9 dB, a $\sigma_e = 0.03$ is used, resulting in

$$\sigma_e^2 = 1.01 \times 10^{-6} B_L \leq 9 \times 10^{-4} \quad (3.4.4.4-2b)$$

Thus, $B_L \leq 89$ Hz (a $B_L = 50$ Hz is used). Note that the use of $\sigma_e = 0.03$ satisfied all the constraints that were listed.

For code acquisition, it is assumed that the VCO will have a maximum code frequency error of 300 chips/sec at a code rate of 4 MHz. Thus, from equation (2.2.2.3-8)

$$\frac{1}{\omega_n} \frac{d}{dt}(\epsilon_d) \leq 1$$

$$\omega_n \geq 300 \quad (3.4.4.4 - 3)$$

For a second order loop filter whose transfer function is given by $(1 + \tau_2 s)/(1 + \tau_1 s)$ with a ξ of 0.707, B_L is approximately equal to $0.53\omega_n$. Thus, B_L becomes greater than or equal to 159 Hz. In order to tolerate nonideal components, $\omega_n = 1000 \text{ rad/s}$ ($B_L = 530 \text{ Hz}$) was used. Since the tracking required a B_L of 50 Hz ($\omega_n = 94 \text{ rad/s}$) while the code loop acquisition requires $B_L = 530 \text{ Hz}$ ($\omega_n = 1000 \text{ rad/s}$), multiple code loop bandwidths must be employed.

The stepping from 1000 rad/s to 94 rad/s was performed such that the ratio between the filter being switched from and the filter being switched to was always maintained at approximately 2.2. Since the loop gain K must be greater than the code loop frequency error, a K of 700 Hz was considered adequate. Using the following equations from reference [24],

$$\tau_1 = (R_1 + R_2)C = \frac{K}{\omega_n^2} \quad (3.4.4.4 - 4a)$$

$$\tau_2 = R_2 C \approx \frac{2\xi}{\omega_n} \quad (3.4.4.4 - 4b)$$

the component values for the loop filters were obtained as shown in Figure 3.25. Since the settling time between each bandwidth switch is about $t = 6/\omega_n$, the times at which each of the filters should be switched in were calculated and are presented in Table 3.2.

The 74LS629 was used as the receiver code generator (Figure 3.26) VCO (CLOCK). As the frequency versus voltage plot of Figure 3.27 shows, the frequency swing is not symmetrical; moreover, the VCO characteristic is linear only over certain portions of the curve. Assuming that the control voltage would swing from 1.80 V to 3.0 V resulted in a VCO sensitivity of $K_V = 3.34 \text{ KHz/V}$ for the voltage range of 1.8 V to 2.4 V and a VCO sensitivity of $K_V = 1.32 \text{ KHz/V}$ for 2.4 V to 3.0 V. Using the lower sensitivity of 1.32 KHz/V and noting from the S-curve of Figure 3.23 that the small error sensitivity is given approximately by $K_D = 0.14 \text{ V/chips}$, the gain of the amplifier between the multiplier output and the loop filters (Figure 3.20) can be found as

$$K_2 = \frac{K}{K_V \times K_D} = 3.8$$

3.4.5/ Demodulation Circuits

The top branch of Figure 3.8 and the schematic of Figure 3.28 show the demodulation channel. The result of the correlation between the incoming signal and the punctual code is filtered and amplified. From this, the carrier is recovered and is

Table 3.2: Switch in Times for the Loop Filters

LOOP FILTER	SWITCH FILTER IN AT
1	$t=0$
2	$t=6\text{ms}$
3	$t=19.2\text{ms}$
4	$t=48.2\text{ms}$

LOOP FILTER	ω_n (rad/sec)	P_1 (Hz)	t_1 (ms)	t_2 (ms)	C (uF)	R_2 (k-ohms)	R_1 (k-ohms)
1	1000	530	4.4	1.4	0.1	14	30
2	455	241	21.2	3.1	0.1	31	131
3	207	110	102.0	6.3	1	60	95.8
4	94	50	497.3	15.0	1	15	452.0

Figure 3.25:- Component Values for the Loop Filters.

Figure 3.26: Receiver Code Generator Schematic.

Figure 3.27: Receiver Code Generator VCO Characteristic.

Figure 3.28: continued

subsequently used to demodulate the information. This is followed by lowpass filtering, decoding of the data bits and conversion to analog form when speech signals are transmitted. The analog result is finally amplified and used to drive a speaker.

As the carrier recovery theory has been discussed in Chapter 2, the following sections will be devoted to discussing some of the design considerations and calculations that were necessary for the carrier acquisition and tracking circuits, as well as the circuits used to recover the data.

3.4.5.1 Carrier Acquisition and Tracking

The recovery or acquisition of the carrier depends to a large degree on the conditions that the system must operate under. If it is assumed that doppler and oscillator instabilities would not produce a carrier uncertainty of more than $\pm 2\text{KHz}$, then the squaring of the signal would make this uncertainty equal to $\pm 4\text{KHz}$. This means that an acquisition range of at least 4 KHz is required.

Since a steady state phase error of less than 3° is desired (refer to Section 3.2), using an expression from reference [24] for phase error, the required loop gain can be obtained as

$$K \geq \frac{\Omega_0}{\sin \phi_\infty} \quad (3.4.5.1 - 1)$$

where K is the loop gain in radians per second, Ω_0 is the initial frequency difference (4 KHz in this case) and ϕ_∞ is the steady state phase error. Thus, K must be greater than or equal to $4.8 \times 10^5 \text{rad/s}$. Using a K of $5 \times 10^5 \text{rad/s}$ in equation (2.2.3.2-1) results in an ω_n that must be greater than or equal to 446 rad/s.

As was discussed earlier, the phase jitter σ_ϕ of the recovered carrier should preferably be less than or equal to $9^\circ (0.157^\circ)$. Neglecting filtering effects, the phase

jitter for the recovered carrier is given by equation (2.2.3.3-4) and can be expressed as

$$\sigma_{\phi}^2 = \left[\frac{B_n}{(P_s/N_0)} \right] \left[1 + \frac{B_i}{2} \left(\frac{1}{P_s/N_0} \right) \right] \quad (3.4.5.1 - 2)$$

where P_s/N_0 is the signal power to noise density, B_n is the one-sided loop bandwidth and B_i is the input bandpass filter noise bandwidth. In the present case, this input BPF is the IF BPF (note that the BPFs that follow the squaring operations at the input and in the loop are there simply to pass the desired carrier term - Figure 3.28).

In situations where the acquisition range is greater than data related sidebands which are spaced $1/2T_b$ from the carrier ($1/T_b$ is the data rate), false lock is a major concern. In such cases, some measure of false lock protection is afforded by an appropriate choice of the input filter noise bandwidth. Chapter 5 of reference [15] and Chapter 11 of reference [25] discuss the topic of data related sidebands in greater detail. Although the acquisition range in the present case can be made less than the data related sidebands, it was deemed prudent to choose the input filter noise bandwidth such that adequate false lock protection would be provided in the event that the use of nonideal components increased the acquisition range. Thus, using a 3 dB bandwidth of 320 KHz (noise bandwidth of approximately 355.2 KHz) for the input filter ensures that the sideband levels relative to the carrier are low (note that this is about 10 times the data rate, which is 31.75 KHz).

The signal to noise ratio at the pre-squaring input filter is increased over that at the output of the RF input filter by about $10 \log(8 \times 10^6 / 320 \times 10^3) = 14 \text{ dB}$. If the implementation, correlation, modulation distortion and tracking jitter losses are factored in, the processing gain from the RF input filter to the pre-squaring filter could be reduced to a value approaching 9 dB. This therefore results in a P_s/N_0 of

56 dB-Hz (for an RF input SNR of -8.5 dB), which when substituted into equation (3.4.5.1-2), yields

$$\sigma_{\varphi}^2 = 3.63 \times 10^{-6} B_n \leq (0.157)^2$$

Thus B_n should be less than or equal to 6.88 KHz (this corresponds to a loop SNR of about 17.6 dB). For a loop filter as shown in Figure 2.27 with $\xi = 0.707$, then B_n is approximated by $0.53 \omega_n$. Therefore, ω_n is upper bounded (for tracking purposes) by 1.3×10^4 rad/s. As a result, the acquisition and tracking requirements indicate that ω_n should be such that

$$446 \text{ rad/s} \leq \omega_n \leq 1.3 \times 10^4 \text{ rad/s}$$

Consequently, a choice of $\omega_n = 2.34 \times 10^3 \text{ rad/s}$ should satisfy both requirements (note that with the above choices for K and ω_n , the implemented acquisition range becomes 9.15 KHz). The acquisition time was calculated (from equation (2.2.3.2-2)) as 34.9 ms (an initial frequency difference of 4 KHz was assumed). The values for the loop filter (lowpass filter with phase lead correction) components as shown in Figure 3.29 were calculated from equations (3.4.4.4-a and b) as

$$\tau_1 = \frac{K}{\omega_n^2} = 9.13 \times 10^{-2} \text{ s}$$

$$\tau_2 \approx \frac{2\xi}{\omega_n} = 6.04 \times 10^{-4} \text{ s}$$

Choosing $C = 1 \mu F$ resulted in

$$R_2 = 604 \Omega$$

Figure 3.29: Loop Filter for Squaring Loop; $F(j\omega) = 1 + j\omega\tau_2 / 1 + j\omega\tau_1$

$$R_1 = 90.7 K \Omega$$

The loop gain K , which was chosen as $5 \times 10^5 \text{ rad}$ can be written as

$$K = K_d K_2 K_{2VCO} \quad (3.4.5.1 - 3)$$

where K_d is the phase detector conversion gain, K_{2VCO} is the effective VCO sensitivity and K_2 is the amplifier gain between the phase detector and the VCO. The XR-2208 was used as the phase detector; its conversion gain is given as $2V/\text{radian}$. For the VCO, the XR-2206 was used. Components were appropriately chosen to give a VCO sensitivity, K_{1VCO} of approximately -4.1 KHz/V . Since the reference used in the squaring loop is the squared VCO frequency, then the effective VCO sensitivity is thus given by

$$K_{2VCO} = 2 \times K_{1VCO} = -8.2 \text{ KHz/V} \approx -5.2 \times 10^4 \text{ rad/V}$$

As a result, K_2 can be solved as -4.8 .

3.4.5.2 Data Recovery Circuits

Ideally, to recover the data, the block diagram of Figure 3.30(a) should be used. However, for asynchronous bit recovery as is implemented here, the integrator, sampler and threshold detector are not required. Rather, simple filtering as shown in Figure 3.30(b) is all that is needed. Note that if the synchronous bit recovery system of Figure 3.30(a) had been implemented, it would have been necessary to have bit synchronization - that is, synchronizing to the data bits so that the sampling is performed at the correct times.

Figure 3.30: BPSK Demodulation: (a) Synchronous bit detection; (b) Asynchronous bit detection

The circuits used in decoding the NRZ(I) encoded data has already been shown in Figure 3.6. The XR-2209 was used for the 16XCLK1. Note that the derived CLK1 (31.75 KHz) is in sync with the received data. Once the decoding is complete, the CVSD (discussed earlier) is used - in the case of voice transmission - to convert the data bits into analog form. The analog signal is subsequently filtered by the CVSD filter (discussed in Appendix A-1.3), amplified and finally used to drive a speaker, as the schematic of Figure 3.28 shows.

Chapter 4

VLSI IMPLEMENTATION

Many of the technological innovations of the past few decades owe their existence to small sized integrated circuits (ICs). Although both analog and digital circuits could be implemented on a chip, the circuits implemented for this thesis are all digital (if time had permitted, some analog circuits would have been tried). As was mentioned earlier, it was decided to use CMOS (Complementary Metal Oxide Semiconductor) logic due to its low power consumption, high fanout and high noise immunity.

The design process starts with a design in terms of discrete logic gates. Using some CMOS standard cells (layouts of standard logic gates) developed by previous students, the design is next translated into their CMOS equivalent layout. The layout equivalent is then used to generate masks that will subsequently be used to fabricate the circuits. The sections that follow will give a more detailed description of the steps followed in implementing the chosen circuits for this thesis. However, Appendices A-2.1 and A-2.2 provide a brief theoretical analysis of CMOS circuits and processing steps.

4.1 IMPLEMENTATION PROCEDURE AND OBJECTIVES

Due to time constraints, only digital circuits have been considered for implementation. These circuits were implemented on the Metheus-CV workstation, using standard cells developed by previous students. For a detailed description of the Metheus-CV workstation along with the design tools available on it for CMOS digital design, reference [34] should be consulted. Reference [35] contains a list of the available standard cells and simulations indicating both the logic and transient behaviour of the cells under certain conditions.

Ideally, a design sequence proceeds in the following manner. After obtaining a schematic of the circuit to be implemented, it is then entered into the computer using SCHEDIT, a schematic editor. Once this has been done, the circuit can be simulated with one of the available simulators including SPICE, MERCURY (Hg) and HILO. While SPICE is a general purpose circuit program that simulates the time domain as well as the frequency response of circuits, MERCURY performs switch level simulations of transistor level circuits and HILO is a gate level simulator.

Once the simulation is completed, PHLED (physical layout editor) or VALE, is used to place and route the required cells, according to the design rules. Following this, a design rule check (DRC) is done on the completed circuit. Once this layout stage is finished, PHLEX (physical layout extractor) and SCH2SIM are used to extract net lists from the layout and schematics respectively. The COMPARE operation is then used to verify that the layout corresponds to the schematic. Note that alternatively the schematic simulation could be skipped; PHLEX could be used to extract the required parameters from the layout and put in a form that can be handled directly by the Hg or SPICE simulators. Thus, the simulation can be done directly on the layout (one of the advantages here is that the interconnect

capacitances and resistances are included in the simulation). Following this latter approach, the PN Gold code generator and the Searchlock logic unit, discussed in Chapter 3, were implemented. The two sections that follow, point out some of the differences in the circuits discussed in Chapter 3 and the ones that were realized on ICs.

Before moving on to the next section however, a few things should be said about SPICE and MERCURY (Hg), the two available simulators that can run simulations directly on the layout. Several factors made it difficult to use SPICE for the circuits considered here. These include:

1. The limitation on the size of the circuit that SPICE can handle
2. The exorbitant amount of time it requires to run anything on SPICE
3. Convergence problems in circuits employing flip-flops

The only outstanding problem associated with MERCURY (Hg) is that it sometimes had difficulties handling circuits employing transmission gates, such as the standard cell exclusive-OR gate. However, for the speed and ability to see in a relatively short time (compared to SPICE) if the circuits are functionally correct, MERCURY was employed. Note however that the results obtained are quite approximate and should not be overly relied on.

4.1.1 The PN Gold Code Generator

The schematic of the PN Gold code generator implemented on a chip (1324 transistors) appears in Figure 4.1 (this is intended to be used in the receiver system). It utilized all of an A size pad frame with a working area of 4040×4040 microns (see Section A-2.2.3). Although the PN Gold code generator realized with discrete

Figure 4.1: 13-Stage PN Gold Code Generator Implemented on a Chip.

components (Figure 3.9) is functionally equivalent, there are a few noteworthy differences that must be addressed.

In the first case, since the clock signal is gated - that is, it is allowed to propagate only under certain conditions - the output of the gate could not directly drive all the flip-flops connected to it. A buffer therefore had to be used (Figure 4.1). The 26-input nor gate of Figure 4.1 is used to detect an all zero condition, which then initiates a reload (an all zero condition is not allowed). Note that the correct action should be to detect the all zero condition for register 2 only, as register 1 (the registers referred to here have 13 stages) is allowed to be all zero (see Appendix A). This was realized after the design had already been sent out for fabrication. However, in the realization with discrete components, this oversight was corrected. The final difference in the PN generators of Figure 3.9 and Figure 4.1 is in the selection of the initial conditions to achieve one-zero balance. In Figure 4.1 (IC realization), only the initial contents of register 1 is variable, whereas in Figure 3.9, both registers 1 and 2's initial contents can be varied.

Because an error was made in the division to obtain the initial conditions (for register 2) that would give a balanced Gold code, and since there is no method to correct this on the chip, the Gold codes (from the chip) will not have one-zero balance. On the other hand, because the discrete component realization has switches to vary these initial conditions, it will be possible not only to obtain a balanced Gold code, but also codes that will match the sequences produced from the chip. This will facilitate the testing of the chip in the spread spectrum system. A sample of the simulations carried out appears in Figure 4.2. C_{in} in the Figure represents the clock input which is inhibited as long as "reset" is high or for the duration that the all zero conditions exists, as indicated by n26 going high. C_{out} represents the actual clock (obtained from C_{in}) that is used to drive the register, while C_{nand} is an intermediate clock output and should be the inverse of C_{out} as is obtained. The variable latch (which has an inverse of latchb) is used to latch in

Figure 4.2: Gold Code Generator Sample Simulation.

or set the initial contents of the registers, while l1 and l7 represent the actual low signals that are used to set the outputs of the second and eighth D flip-flops to an initial high value (the flip-flops are numbered 0 to 12). Note that the D flip-flop outputs O1 and O7 have been initially set to a high by the latch signal (after an all zero condition was detected) and they remain this way until the first positive edge of clock pulse C_{out} , whereupon the contents of the flip-flop or register preceeding the O1 and O7 registers (in this case, lows) are shifted into the O1 and O7 registers respectively (a register in this case is regarded as having one stage). As can be seen, the early (PN-E), punctual (PN-P) and late (PN-L) are half a chip apart.

4.1.2 The Searchlock Logic Unit

The searchlock logic unit that was implemented on a chip (310 transistors) appears in Figure 4.3; its operation is as was discussed in Chapter 3. A B sized pad frame (4510×2250 microns maximum outside dimensions - see Table A-2.2) was needed to implement this circuit. It is to be noted that this system would have been practically identical to the discretely implemented one of Figure 3.16 had not a few minor changes been necessary. In trying to get the system of Figure 3.16 working, it became necessary to change a few gates. This problem was not detected with the Mercury simulator probably because of its problems with exclusive-OR gates.

In anycase, a sample of the simulation results from the layout of the schematic of Figure 4.2 appears in Figure 4.4. Most of the variables appearing in Figure 4.4 have already been discussed at some point before. However, some still require a few comments. The variables y_1 , y_2 and y_0 represent the state in the acquisition process that the system is in, while xck is the clock used to clock in the threshold level exceeded indicator, V_{td} , and x represents the clocked in value (xb is the inverse of x). Note that xck changes when the switch was made to a longer integration time - that is, F_{B1} was selected when V_{td} became a 1.

Figure 4.3: Searchlock Logic Unit Implemented on a Chip.

Figure 4.4: Searchlock Logic Unit Sample Simulation.

There are a few discrepancies that should be mentioned. The first is the area with a cross in it. This represents an undefined value; the correct value should be a low. During simulation, it was suspected that the problem arose from the Mercury simulator; this was later confirmed when the fabricated chip did not show this problem. The second discrepancy comes from the LOCKIND (lock indicator) output. The only time it should be low is when y_0 , y_1 and y_2 are all either zero or one. The glitch that appears in the LOCKIND trace came at a time when y_0 , y_1 and y_2 were all momentarily zero; it was assumed that Mercury could not handle such transitions.

Chapter 5

EXPERIMENTAL RESULTS

The direct sequence spread spectrum multiple access system (transmitters and receiver) of Chapter 3 was implemented as shown in Figure 5.1. Performance measurements on the system were made under different input signal to noise ratios. Observations and tests were carried out on the constituent parts of both the transmitters and receiver. These parts included the Gold code generators, the input and reference filters, the acquisition, tracking and demodulation channels. In addition, the processing gain as well as the BER for single and double users were measured. Finally, performance measurements for the ICs were also obtained.

5.1 Transmitters and Signal Transmission

The voice input spread spectrum transmitter appears in Figure 5.2. The data input spread spectrum transmitter is shown in Figure 5.3. Note that in both, a switch is provided to allow either a transmission of the data modulated carrier or just the carrier with no data modulation.

The transmitter Gold code generator produces time waveforms as shown in Figure 5.4. Note that this shows only a very small segment of the 8191 chip long

Figure 5.1: Spread Spectrum Multiple Access System.

Figure 5.2: Voice Input DS SSMA Transmitter.

Figure 5.3: Data Input SSMA Transmitter.

Figure 5.4: Time Waveforms for the Transmitter 13-Stage Gold Code Generator.

code. The corresponding baseband code spectra is shown in Figure 5.5. Notice the low levels of the stationary signals at the nulls, in comparison to the peak, which indicates good code balance. This is desirable because if the stationary signals at the nulls were high, they could cause problems for the acquisition circuits.

Using a carrier of 880 MHz, the transmitted signal was observed to have a spectrum as shown in Figure 5.6 (obtained from a HP8569B spectrum analyzer). It can be seen that both good code and carrier suppression is attained especially when Figure 5.6 is compared to Figure 5.7, an example of a balanced mixer output with both carrier and code imbalance.

Although Figure 5.6 shows the transmitted signal at 880 MHz, it was necessary to move it to 70.6 MHz instead. This was due to the fact that the frequency synthesizer that was being used for the 880 MHz carrier was not very reliable. Thus, instead of using the desired configuration of Figure 5.8, that of Figure 5.9 had to be utilized. Figures 5.10 and 5.11 show the actual configurations used (for the one and two user cases) in greater detail. The input signal to noise ratios for the measurements made were arrived at by measuring the noise and signal powers at the output of the RF input filter (Figures 5.10 and 5.11). Figures 5.12 and 5.13 show samples of the measured data. The top trace of Figure 5.12 shows the received noise alone (9.6 dBm) while the bottom trace shows the received noise free signal (1.2 dBm). The bottom trace of Figure 5.13 is identical to that of Figure 5.12 while the top curve represents the received signal plus noise trace. Note that due to some power combiner loss, there is a slight drop in the noise plus signal trace. By considering Figures 5.12 and 5.13, it can be seen that not only is the signal spectral density lower than the noise density, the noise plus signal trace looks no different from the noise trace alone. As a result, spread spectrum signals can be well concealed in noise.

Figure 5.5: Baseband Code Spectra for the 13-Stage Transmitter Gold Code Generator.

CTR 880.2 MHz SPAN 2 MHz RES BW 100 kHz VF .3
REF -2 dBm 10 dB ATTEN 40 dB SWP .5 sec

Figure 5.6: Transmitted Signal Spectrum.

Figure 5.7: Balanced Mixer Output With Both Carrier and Code Imbalance. [2]

Figure 5.8: System Configuration Using 880 MHz Carrier.

Figure 5.9: System Configuration Using 70.6 MHz Carrier.

Figure 5.10: Detailed 70.6 MHz Carrier System Configuration for One User.

3

1.0

1.1

1.25

1.4

1.6

1.8

2.0

2.2

2.5

2.8

3.2

3.6

4.0

4.5

5.0

5.6

6.3

7.1

8.0

9.0

MICRO

Figure 5.11: Detailed 70.6 MHz System Configuration for Two Users.

Figure 5.12: Received Noise and Signal Traces: Noise - Top Trace; Signal - Bottom Trace ($SNR_i \approx -8.4dB$).

Figure 5.13: Received Signal and Noise Traces: Signal Plus Noise - Top Trace ($SNR_i \approx -8.4dB$); Signal Only - Bottom Trace.

5.2 Receiver System

A typical response for the reference filters discussed in Chapter 3 appears in Figure 5.14(a) while that of the RF input filter appears in Figure 5.14(b). Note that the responses are practically identical, as was required in the discussion of Chapter 3.

In the following sections, the rest of the receiver sections (Figure 5.15), and their performance will be discussed.

5.2.1 Receiver Gold Code Generator

The receiver Gold code generator was implemented on part of the circuit board of Figure 5.16(a). The outputs appear in Figure 5.16(b). Note that, "E" stands for early, "P" for punctual, "L" for late and "CLK" for clock. From the 4 MHz clock input (bottom trace), it can be seen that the early sequence is half a chip ahead of the punctual sequence which is in turn half a chip ahead of the late sequence, as was expected from Chapter 3.

5.2.2 Acquisition Channel Results

The board containing most of the acquisition channel components appears in Figure 5.17. Figure 5.18(a) shows the acquisition channel IF bandpass filter response while Figure 5.18(b) shows the response of the lowpass filter following the envelope detector (these measurements were made on the HP3585 spectrum analyzer). As is seen, the 3-dB bandwidth for the BPF is approximately 67 KHz as was designed; the lowpass filter response, though not as flat in the passband as was desired, was satisfactory.

The correlation triangle appears in Figure 5.19. This was measured right after

Figure 5.14: Reference and RF Input Filter Responses: (a) Reference filter response; (b) RF input filter response.

Figure 5.15: Receiver System.

(a)

(b)

Figure 5.16: Circuit Board Containing the Receiver Gold Code Generator and the Generator Outputs: (a) Gold code generator implementation; (b) Clock input and the early, punctual and late outputs.

Figure 3.17: Acquisition Channel Circuit Board.

Figure 5.18: Acquisition Channel Filter Responses: (a) IF bandpass filter response; (b) Lowpass filter response.

Figure 5.19: Correlation Triangle: y axis: 50 mv/div; x axis: 0.5 ms/div; ($SNR_i = -0.6$ dB).

the diode envelope detector of the acquisition channel. In theory, the maximum width of the correlation triangle should be equal to 0.5 ms, the time required to slide the receiver local code 2 chips past the received code (note that the sliding is done in half chip increments). However, from Figure 5.19, a base width of about 1.1 ms is obtained for the correlation triangle. It is conjectured that this discrepancy could be due to the input capacitance of the lowpass filter following the envelope detector. This conclusion is suggested by the exponential kind of decay exhibited by the triangular pulse.

In Chapter 3, the searchlock logic unit (which is on the circuit board of Figure 5.16(a)) was discussed in some detail. Particular attention was paid to the four pulses used to time the integration periods, cause the sliding of the two codes past each other, etc. These pulses appear in Figure 5.20. Notice that as was desired in Chapter 3, F_{A1} (top trace) is five times as fast as F_{B1} (second trace) while F_{A2} (third trace) is five times as fast as F_{B2} (bottom trace). Although the four traces had pulses about 135ns wide, only F_{A2} is strictly required to be this width. This is because the F_{A2} pulse is used in the inhibition of the code generator clock, which allows the local code to slide past the received code.

It should be mentioned that the acquisition thresholds predicted by theory could not be used because of the unexpected amount of noise and interference obtained on the circuit boards. As a result, the thresholds (for dwell periods one and two) had to be determined experimentally. The values that were satisfactorily used were

$$V_{T1} = 1.0V \quad V_{T2} = 2.5V$$

where V_{T1} represents the threshold value that had to be exceeded for the first dwell period, while V_{T2} represents the same for the second dwell period.

The acquisition times were measured with a logic analyzer. Basically, the local

Figure 5.20: Searchlock Logic Unit Pulses: y axis: 5V/div; x axis: 0.1 ms/div;
 F_{A1} : Top trace; F_{B1} : Second trace; F_{A2} : Third trace; F_{B2} : Bottom trace.

code generator was reset by an active low signal, which also inhibited the input clock. Once the reset signal goes high again (i.e., when the reset push button switch is released), the local code generator is reloaded and the clock is then allowed to operate the generator. The logic analyzer was set up to trigger on the lock indication signal; the time from the release of the reset button to the lock indication signal then indicates the acquisition time, as is illustrated in Figure 5.21. The top trace (P#00) indicates the reset signal while the second trace (P#01) indicates the lock indication signal. The acquisition time for this particular trial (1288 ms) is given in the top right hand corner.

Although detailed acquisition time measurements relating different acquisition times to the input signal to noise ratio for given code doppler rates could be made, time constraints allowed only for the measurements of Tables 5.1 and 5.2. The calculated mean acquisition times and variances were obtained by using the following expressions for sampled data, extracted from reference [36].

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (5.2.2 - 1(a))$$

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (5.2.2 - 1(b))$$

Here, $\bar{x}$ represents the average of n data values, while s^2 represents the variance. Although a larger sample space would produce more accurate results, 32 sample points (Tables 5.1 and 5.2) were deemed satisfactory for the present analysis.

Comparing the theoretical results of Table 3.1 with those of Table 5.1 shows fairly close agreement between the acquisition time and variance calculations. Note however that the theoretical results for Table 3.1 were computed assuming an SNR_i of -8.5 dB while the SNR_i in the case of Table 5.1 was -8.7 dB. A comparison of Tables 5.1 and 5.2 seems to indicate that a higher input signal to noise ratio

Figure 5.21: Acquisition Time Measurement.

Table 5.1: Acquisition Time Measurements SNR_i (Input Signal to Noise Ratio) = -8.7 dB, Code Doppler (Δf_c) ≈ 0

Trial	Time(s)	Trial	Time(s)
1	2.6	17	0.897
2	1.374	18	3.068
3	1.236	19	3.705
4	0.437	20	3.755
5	3.439	21	2.234
6	1.846	22	0.855
7	2.556	23	2.228
8	1.313	24	2.064
9	0.534	25	2.187
10	0.987	26	3.720
11	4.260	27	3.984
12	1.245	28	7.713
13	0.405	29	0.867
14	1.917	30	2.731
15	1.916	31	1.101
16	1.272	32	3.555

$$\bar{T}_{acq} = 2.25 \text{ seconds}$$

$$\sigma_{acq}^2 = 2.291 \text{ seconds}$$

(SNR) results in a shorter acquisition time. This agrees with theory in that a higher input SNR implies a higher probability of detection and from the equations of section 2.2.1.1.2, if everything else is kept constant, this would result in a shorter acquisition time.

5.2.3 Tracking Channel Results

The tracking channel circuit board is shown in Figure 5.22. The IF bandpass and lowpass filter responses were identical to the acquisition channel filter responses

Figure 5.22: Tracking Channel Circuit Board.

Table 5.2: Acquisition Time Measurements SNR_i (Input Signal to Noise Ratio) = -0.9 dB, Code Doppler (Δf_c) ≈ 0

Trial	Time(s)	Trial	Time(s)
1	0.470	17	1.960
2	1.030	18	1.820
3	1.400	19	1.130
4	1.210	20	1.020
5	1.690	21	1.750
6	0.510	22	1.770
7	1.970	23	0.740
8	0.170	24	1.940
9	2.050	25	0.060
10	0.680	26	0.750
11	0.020	27	0.300
12	1.050	28	0.620
13	1.610	29	0.450
14	0.520	30	0.450
15	1.350	31	0.380
16	0.100	32	1.540

$$\bar{T}_{acq} = 1.016 \text{ seconds}$$

$$\sigma_{acq}^2 = 0.423 \text{ seconds}$$

of Figure 5.18. The tracking loop was observed to maintain lock for code dopplers greater than the assumed 300 Hz of Chapter 3.

In order to obtain the S curve, it was necessary to ensure that the two codes (reference and the received code) must slide past the "in-phase" position. This was accomplished by making sure that the first threshold was so high that it could not be exceeded. Since the first dwell period was approximately equal to $125\mu s$, the dithering frequency had to be greater or equal to 16 KHz (corresponding to a square wave frequency of 8 KHz) in order to obtain the complete S-curve characteristic. Adhering to the above, the S-curve of Figure 5.23 was obtained. This picture was taken after the amplifier which follows the multiplier used in summing the early and late correlation signals; it was then filtered by a 2 KHz sharp cutoff filter which smoothed out the spikes, etc., introduced by the switching action.

From Chapter 2, the theoretical S-curve (discriminator characteristic) extends from $-3/2T_c$ to $3/2T_c$. Thus, the total width is given by $3T_c$. Since the search is done in half chip increments or steps and about $125\mu s$ is spent at each step, then the base width (horizontal axis) of the S-curve should be given by

$$3T_c \times \frac{2\text{steps}}{T_c} \times \frac{125\mu s}{\text{step}} = 0.75ms$$

From an examination of Figure 5.23, it can be seen that the experimental S-curve has a base width of about 0.84 ms. The 12% discrepancy could be attributable to filtering effects.

The slope of the S curve of Figure 5.23 near $\epsilon_d = 0$ is 0.1875 V/chip (note that the center of the oscilloscope does not represent $\epsilon_d = 0$). For an amplifier gain K_2 of 2.3 (note that the designed gain was for 3.8; the gain of 2.3 was accidentally left on when the picture was taken), the experimental K_D becomes

Figure 5.23: Measured S-curve From the Tracking Channel: y axis: 50 mv/div; x axis: 0.2 ms/div.

$$K_D = \frac{0.1875}{K_2} = 0.082V/chip$$

Comparing with the deduced K_D of 0.14 V/chip from Chapter 3, it can be seen that there is quite a discrepancy, which could be caused by a decrease in the signal level at the envelope detector output (the signal level at the detector output was not measured when the photograph was taken). However, the effect of a lower K_D can easily be corrected for by increasing the amplifier gain K_2 .

The loop bandwidth control, discussed in some detail in Chapter 3, performed quite well. The picture of Figure 5.24 shows the times that the loop filters are switched in. The first trace (P#00) indicates when acquisition occurs (the first filter is immediately applied); the second, third and fourth traces show when the second, third and fourth loop filters are switched in. Note the time in the right hand corner of the picture; this represents the time it takes from the moment of acquisition to the moment when the fourth loop filter is applied. A comparison with the designed value of 48.2 ms shows that it meets the requirements fairly well.

Finally, although the system was able to acquire lock for input signal to noise ratios down to -12.1 dB, it was able to maintain lock or track for input signal to noise ratios as low as -13.8 dB. The unexpected high noise levels encountered on the circuit boards affected the acquisition and tracking circuits. As a result the system lost lock more often than was expected. The time between loss of lock varied from a few minutes to up to 30 minutes, depending on the input signal to noise ratio, with the higher input signal to noise ratios maintaining the system in lock for longer periods (note that the system automatically reacquires lock following any loss of lock, the mean time to reacquire depending of course on the particular input signal to noise ratio). The use of an automatic gain control (decided against due to time constraints) may have helped in combatting the effects of the noise, by sensing the variations in the input noise level and appropriately compensating for it.

Figure 5.24: Loop Filter Switch in Times: P#00: Acquisition occurs; loop filter 1 applied; P#01: Loop filter 2 applied; P#02: Loop filter 3 applied; P#03: Loop filter 4 applied.

5.2.4 Demodulation Channel Results

Figures 5.25(a) and 5.25(b) show pictures of the demodulation channel circuit boards. The board of Figure 5.25(a) is basically used to recover the suppressed carrier, while the board of Figure 5.25(b) is used for data recovery and conversion to analog voice output. The response for the demodulation channel input bandpass filter appears in Figure 5.26(a) while that of the narrow band bandpass filter following the squaring operation is shown in Figure 5.26(b).

The carrier acquisition system was able to acquire lock for signals in the range 593.5 KHz to 607 KHz. This compares favorably to the theoretical value of $600 \text{ KHz} \pm 9.1 \text{ KHz}$ obtained from the parameters arrived at in section 3.4.5.1. The despread unmodulated carrier following the demodulation BPF appears in Figure 5.27(a). The corresponding despread data modulated output is depicted in Figure 5.27(b). The total signal power in the $\left(\frac{\sin x}{x}\right)^2$ trace of Figure 5.27(b) can be represented by

$$P_s = P_0 R_s \quad (5.2.4 - 1)$$

where P_s is the total signal power, P_0 is the maximum spectral density of the signal (at $x = 0$) and R_s is the data rate. Using the above relation, the total signal power in Figure 5.27(b) was obtained as -5 dBm. The 0.4 dB discrepancy observed by comparing to the carrier power in Figure 5.27(a) could be attributed to modulation distortion (note that the data modulated output of Figure 5.27(b) employs a 63-bit pseudorandom sequence).

As was discussed in Chapter 3, a demodulation BPF bandwidth about 10 times the data rate was used in order to ensure there would be no chance of false lock. With this in mind and accounting for degrading factors, it was decided that the processing gain from the output of the RF input filter to the demodulation BPF output would drop to about 9 dB. In retrospect, a demodulation BPF about three

(a)

(b)

Figure 5.25: Demodulation Channel Circuit Boards: (a) Carrier recovery board; (b) Circuit board used for data recovery and conversion to analog voice output.

(a)

(b)

Figure 5.26: Filter Responses for the Demodulation Channel: (a) Input bandpass filter response; (b) Response for the narrowband filter following the squaring operation.

Figure 5.27: Despread Signals: (a) Despread unmodulated carrier; (b) Despread data modulated carrier.

or four times the size of the data rate may have been just as satisfactory. In any case, the processing gain in the information bandwidth (which would determine the bit error rate) would still be given by 21dB losses.

The despread SNR and processing gain following the demodulation input band-pass filter appears in Table 5.3. Figure 5.28 shows some of the traces for the despread data spectrum that was used in obtaining the values of Table 5.3. Note that in extracting the noise density (used in calculating the output noise power) from these traces, the noise floor was taken to be in the region from 700 KHz to 800 KHz. The reason for not measuring everything at one point was that for high noise inputs, the output signal levels were practically obscured by noise near 700 KHz; while for low noise inputs, the noise density measurements had to be made somewhat farther away since the signal levels were more prominent.

As has already been indicated, it was assumed that implementation, correlation, modulation distortion and tracking jitter losses would reduce the processing gain from its theoretical value of 14 dB to about 9 dB . Considering Table 5.3, it can be seen that this was somewhat pessimistic as the degradation ranged from 1.9 dB to 4.2 dB . In considering the values of Table 5.3, it is advisable to give about a 0.5 dB leeway for any inaccuracy that might have been introduced in extracting the data from traces such as is depicted in Figure 5.28.

It was observed that as the input signal to noise ratio was decreased (noise input was increased), the output signal power decreased even though the input signal power was constant. This is illustrated in Figure 5.29. Both Figures 5.29(a) and 5.29(b) have the same input signal power although the input noise in Figure 5.29(b) is 9.8 dB greater than that in Figure 5.29(a). Note that the despread carrier power for Figure 5.29(a) is 3.8 dB greater than that of Figure 5.29(b). It

Figure 5.28: Examples of the Traces Used in Obtaining the Signal and Noise Values of Table 5.3.

Figure 5.29: Effect of Noise on System Performance: (a) Output with 9.8 dB less noise input than in case b; (b) Noise input is 9.8 dB greater than in case a.

Table 5.3: Despread SNR and Processing Gain for Single User Transmission. Input Signal Power = -1.2 dBm

<i>SNR_i</i> (dB)	Output Noise Power (dBm)	Output Signal Power (dBm)	Output SNR (dB)	Processing Gain (dB)
-13.8	-6.8	-10.8	-4	9.8
-12.1	-7.3	-8.8	-1.5	10.6
-10.3	-8.3	-7.8	0.5	10.8
-8.4	-8.8	-7.3	1.5	9.9
-6.6	-10.3	-6.8	3.5	10.1
-4.6	-11.3	-5.3	6	10.6
-2.6	-12.3	-4.8	7.5	10.1
-0.6	-15.3	-3.8	11.5	12.1
1.4	-16.8	-3.8	13.0	11.6
3.4	-19.3	-3.8	15.5	12.1

is conjectured that the reason for this decrease is that because of the higher input noise, the tracking circuit performance is degraded, therefore resulting in a higher signal loss. This, in addition to the introduction of errors from the data extraction process (already mentioned) could explain the variation in processing gain observed in Table 5.3.

Figure 5.30(a) shows the output of the information band lowpass filter when a 1:1 bit pattern is transmitted. Note that at this point, data demodulation has already been accomplished. The result of amplifying this analog output, followed by hardlimiting and decoding is the clean bit stream of Figure 5.30(b).

Since the observed degradation in processing gain (at the BPF output) for the measured bit error rates ranged from 3.2 to 4.2 dB, it was decided to assume an average degradation of 3.7 dB in obtaining the signal to noise ratios required in calculating the expected bit error rates. If a further 1.3 dB is factored in for asynchronous bit detection and any additional losses from the demodulation process,

(a)

(b)

Figure 5.30: Output from the Data Recovery Section When a 1:1 Bit Pattern is Transmitted: (a) Output following the information band low pass filter; (b) Output after decoding has taken place.

Table 5.4: BERs for Single User Transmission

SNR_i (dB)	z (dB)	BER (Expected)	BER (Experimental)
-13.8	2.2	3.44×10^{-2}	2.3×10^{-1}
-12.1	3.9	1.32×10^{-2}	1.4×10^{-1}
-10.3	5.7	3.26×10^{-3}	6.5×10^{-2}
-8.4	7.6	3.4×10^{-4}	1.2×10^{-2}
-6.6	9.4	1.5×10^{-5}	8.4×10^{-4}
-4.6	11.4	7.4×10^{-8}	2.3×10^{-6}
-2.6	13.4	1.9×10^{-11}	2.0×10^{-6}

a 5 dB signal loss is arrived at. The theoretical probability of error for a binary coherent PSK system can be written as [40,42]

$$P_E = \frac{1}{2} \operatorname{erfc}(\sqrt{2}) = \operatorname{Erfc}(\sqrt{2z}) \quad (5.2.4 - 2)$$

where

$$\operatorname{erfc}(u) \triangleq 1 - \frac{2}{\sqrt{\pi}} \int_0^u e^{-y^2} dy \text{ and } \operatorname{Erfc}(u) \triangleq \int_u^\infty \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy$$

and z is the ratio of the signal power to the noise power in the information or signal bandwidth. Using a processing gain of 16 dB (21 dB-5 dB assumed losses), the expected information bandwidth SNR and BER of Table 5.4 were obtained. The experimental BER (obtained with the use of a HP1645A data error analyzer) is also shown. It should be noted that noise bursts (especially at low SNRs) made it a little difficult to measure the BER; the average of several readings had to be used. The results are plotted in Figure 5.31.

Considering Figure 5.31, it can be seen that for a given BER, a difference of

Figure 5.31: BER Results for Single User Transmissions. (Note that $m = 10^{-3}$, $\mu = 10^{-6}$, $n = 10^{-9}$ and $p = 10^{-12}$ on the y axis).

about 3-dB (on the average) exists between the expected and experimental values. Since the loss due to despreading (correlation loss), PN tracking jitter, modulation distortion and implementation losses were shown (for the SNR_i involved) in Table 5.3 to have an average value of 3.7 dB, it then stands to reason that the 1.3 dB SNR_i degradation assumed for asynchronous bit detection and the demodulation process was too low. Adding the observed 3 dB difference to the 5 dB loss already assumed would result in an effective processing gain of 13 dB (21 dB-8 dB) from the RF filter output to the information bandwidth. Since many of the parts used in the system were makeshift in nature, it is expected that with much better constructed parts, careful placing and shielding of components and the use of printed circuit boards (instead of wirewrapping and soldering), the SNR degradation could be easily reduced to about 4 dB. The use of synchronous bit detection would further reduce this degradation.

In evaluating the performance of speech transmissions, it was observed that at an SNR_i of -8.4 dB the voice quality was barely acceptable; however, at an SNR_i of -6.6 dB, the voice quality was quite good and got even better as the input signal to noise ratio was increased (noise input reduced). In assessing the multiple access performance for the voice transmitter, the power for the data input transmitter (which was considered as the unwanted transmission) was varied while the noise input was kept constant. It was observed that an increase in the power for the unwanted transmission tended to degrade the quality of the voice receptions (and vice versa) much in the same way that an increase in noise power degraded its performance.

Although the observations on the multiple accessing performance agree with the equations of Chapter 3, a more quantitative analysis would have been more instructive. However, due mainly to time constraints and the makeshift nature of the system, it was decided that a qualitative study showing general trends would be adequate. In Table 5.5, the effect of varying the power of the undesired signal (in

Table 5.5: BER for Multiple Access Environment. (Number of Users = 2). $P_N = 2.8$ dBm, $P_D = 4.8$ dBm

P_u (dBm)	P_D/P_u (dB)	BER
10.6	-5.8	2.5×10^{-1}
9.5	-4.7	1.1×10^{-2}
8.5	-3.7	7.5×10^{-3}
7.6	-2.8	3.6×10^{-3}
6.9	-2.1	6.0×10^{-4}
5.9	-1.1	2.5×10^{-4}
4.8	0.0	1.0×10^{-4}
3.8	1.0	7.0×10^{-5}
2.8	2.0	5.0×10^{-6}
1.8	3.0	4.0×10^{-6}

this case, the voice transmission which is designated P_u) on the BER is presented. Note that in this study, the noise power (P_N) and the desired signal power, are kept constant at 2.8 dBm and 4.8 dBm respectively. Figure 5.32 illustrates the effect of varying P_u ; note that the BER decreases with an increase in P_u just as it does in Figure 5.31 with an increase in noise power.

5.2.5 IC Results

A picture of the fabricated ICs appears in Figure 5.33. Figures 5.34 and 5.35 illustrate the bonding diagrams for the Gold code generator and searchlock logic units.

The time waveform of the Gold code generator pseudorandom output at 4 MHz is shown in Figure 5.36(a) while its baseband code spectra at the same frequency appears in Figure 5.36(b). Note the high spikes at the nulls - this indicates code imbalance due to the mistake already discussed in Chapter 4. Compare this to an

Figure 5.32: BER Plot for the Case of Two Users. (Note that $m = 10^{-3}$ and $\mu = 10^{-5}$ on the y axis).

Figure 5.33: Picture of the Fabricated ICs.

PIN NUMBER	FUNCTION
1	D0
2	D1
3	D2
4	D3
5	D4
8	D5
9	D6
10	VCC
26	GND
27	PN-E
28	PN-L
30	PN-P
34	reset
35	C_{in} (PN CLOCK INPUT)
36	D7
37	D8
38	D9
39	D10
40	D11

* The rest of the pins are unused.

Figure 5.34: Bonding Diagram for the Gold Code Generator.

PIN NUMBER	FUNCTION
2	y1
3	y2
14	DISCH
15	FB2
18	FA2
19	FB1
20	FA1
21	LOCKIND
23	GND
26	PN generator clock
27	VCO
28	V _{id}
34	VCC
40	y0

* The rest of the pins are unused.

Figure 5.35: Bonding Diagram for the Searchlock Logic Unit.

(a)

(b)

Figure 5.36: Time and Spectral Outputs for the IC Gold Code Generator: (a) Time waveforms; (b) Baseband code spectra.

output with code balance as is shown in Figure 5.5. The baseband code spectra was observed as the clock frequency was increased. Figure 5.37 shows the results.

Although serious distortion seems to begin only after 6 MHz, this does not mean that all the bits produced are correct for frequencies less than 6 MHz. Beyond 6.78 MHz, the baseband code spectra is totally distorted; the result in the time domain is an output that remains high all the time (note that speeds of up to 10 MHz were attainable in simulation).

In testing the three outputs of the code generator (the early, punctual and late outputs), it was observed that the late output was defective as seen in Figure 5.38, which has a clock frequency of about 120 KHz. It was surmised (after re-examining the layout) that one of the gates used in the output stage was defective. Thus, an output stage as shown in Figure 5.39 was provided off chip. This produced the results shown in Figure 5.40 (the time at the top right hand corner of the pictures corresponds to the highlighted portions of the traces). Note from these figures that the reason the time difference between the early and punctual sequences is slightly larger than the difference between the punctual and late sequences is because the clock does not have exactly a 50% duty cycle.

In using the IC Gold code generator in the system, it was more difficult to attain and maintain lock than in the case where the discretely implemented code generator was used. The longest that the system was able to maintain lock was about two minutes. It is possible that the reason for this malfunction could be due to noise problems on the chip, which could introduce errors into the bit sequence produced by the IC code generator, thereby making it hard to acquire and maintain lock. After reconsidering the layout of the code generator, it was felt that the limitation on the speed and any noise problems could be due to two factors, these being (1) insufficient drive capability for some sections; (2) the use of a series connection for the power and ground connections (a parallel connection would have been better).

Figure 5.37: Baseband Code Spectra for IC Gold Code Generator as Code Rate is Increased.

Figure 5.38: Picture Illustrating the defect in the “Late” Output of the IC Gold Code Generator.

Figure 5.38: Picture Illustrating the defect in the "Late" Output of the IC Gold Code Generator.

Figure 5.39: Output Stage Added to Correct Defect Seen in Figure 5.38.

Figure 5.40: Early, Punctual and Late Outputs for the IC Code Generator with the Output Stage of Figure 5.39.

Figure 5.40: continued

Table 5.6: BER Obtained for Single User Transmissions When the IC Gold Code Generator Was Used

SNR_i (dB)	BER (Experimental)
-13.8	2.5×10^{-1}
-12.1	1.5×10^{-1}
-10.3	9.1×10^{-2}
-8.4	6.4×10^{-3}
-6.6	6.9×10^{-4}
-4.6	7.2×10^{-6}

The BER shown in Table 5.6 and plotted in Figure 5.41 represent an average of those measurements that were made while the system was able to maintain lock. The same expected values used in Figure 5.31 are also used here. Note that the obtained BER of Figure 5.41 is slightly worse than that of Figure 5.31.

In terms of power consumption, the IC code generator was observed to draw less than 5 mA while the one implemented with discrete components drew more than 50 mA of current at 4 MHz. This clearly shows the attractiveness of implementing the system blocks on ICs.

As was already indicated in Chapter 4, it was necessary to change a few gates in the searchlock logic unit in order for it to perform as desired. As the changes were made after the designs had been sent out, the chip design does not reflect these changes and could not therefore be used in the receiver system. Nevertheless, it was tested to see if it functioned as was designed. The state diagram of Figure 5.42 shows the algorithm realized by the IC searchlock unit and conforms to what was expected in light of the errors mentioned earlier.

Other tests were performed to ensure that the rest of the circuits were working. These included tests to ensure that the pulses were being properly selected, the lock

Figure 5.41: BER For Single User Obtained When the IC Gold Code Generator was Used. (Note that $m = 10^{-3}$, $\mu = 10^{-6}$ and $n = 10^{-9}$ on the y axis).

q0: ACQUISITION OR SEARCH MODE
FIRST DWELL PERIOD(125 US)

q1: VERIFY ACQUISITION & START TRACKING MODE
SWITCH TO SECOND DWELL PERIOD(625 US)

q2: TRACKING MODE

* NOTE THAT ONLY TWO CONSECUTIVE MISSES ARE NEEDED HERE BEFORE THE SYSTEM RETURNS TO THE SEARCH MODE.
IN THE DESIRED CASE OF FIGURE 3.18, WHICH WAS ACCURATELY IMPLEMENTED WITH DISCRETE COMPONENTS, FIVE CONSECUTIVE MISSES ARE NEEDED TO RETURN THE SYSTEM TO THE SEARCH MODE.

Figure 5.42: Algorithm Realized By The IC Searchlock Unit.

indicator was working, the VCO input was being properly divided and inhibited. An example of the latter appears in Figure 5.43. In Figure 5.43(a), the threshold level exceed indicator, V_{td} , is zero and thus the 8 MHz input clock (top trace) is correctly both inhibited and divided by 2 to produce the approximately 3.996 MHz output (bottom trace) to be used to drive the PN generator. When V_{td} is a one, there is no inhibition of the 8 MHz input (Figure 5.43(b)) although it is still divided down to 4 MHz to be used to drive the PN generator. Note the difference in the bottom traces of Figure 5.43(a) and (b). The ghostly lines in the bottom trace of Figure 5.43(a) indicates the slight frequency deviation caused by the inhibition of the 8 MHz input; note that this is not present in the bottom trace of Figure 5.43(b).

Although there were some flaws noted for the two implemented ICs, they all can easily be corrected. When it is considered, however, that this was only the first iteration, the flaws look quite insignificant.

(a)

(b)

Figure 5.43: An Example of the Results Obtained From the IC Searchlock Unit:
(a) Result with $V_{td} = 0$; (b) Result with $V_{td} = 1$.

Chapter 6

CONCLUSIONS

6.1 Summary

A theoretical and a hardware realization of a direct sequence spread spectrum multiple access system has been presented. In addition to providing speech capability, an IC implementation of a few of the system blocks was carried out. A detailed theoretical background was first provided, followed by a concise and informative description of the considerations and design steps required to realize the system. The resulting system was observed to work extremely well, especially when the makeshift nature of some of the components used is considered.

Although PN acquisition was achieved only down to -12.1 dB and lock maintained down to -13.8 dB, it is felt that with the use of printed circuit boards and some shielding, the high noise levels observed on the circuit boards would be greatly diminished, resulting in a better acquisition and tracking performance. The carrier acquisition and tracking loops performed quite well. An SNR degradation of about 8 dB was observed from the RF filter output to the information band lowpass filter.

While this is a high value, it can be easily and dramatically reduced by methods already cited.

The BER performance for both the single and multiple access situations were quite good considering the quality of the components used. Although no quantitative comparisons to the equations of Chapter 3 were made (for the multiple access case), the general trend of results is nevertheless instructive.

The implemented chips were tested and functioned pretty much as expected (considering the mistakes that were not spotted in time). Despite their flaws (which could easily be corrected in a second iteration), these chips clearly demonstrated the feasibility of implementing the system on CMOS ICs.

6.2 Future Work

The work presented here by no means exhausts the subject. To begin with, the system performance could be greatly improved by reducing the amount of noise on the boards by, as has already been indicated, careful layout of components, shielding and utilization of a printed circuit board. The use of an automatic gain control circuit may also help in combating the deleterious effects of the noise.

Secondly, it would be very interesting to see how well the system performs in a field trial where antennas and a real channel is used. The system performance (including the acquisition and tracking systems) under a multipath fading channel would be highly instructive in deciding how suited the system is for mobile communications. Trying more than two users and attempting to quantitatively compare the results with the multiple access equations of Chapter 3 would be worthwhile.

Finally, it would be highly desirable to see more of the system implemented on ICs. This is the only route to go in terms of making spread spectrum transceivers portable.

REFERENCES

- [1] Pickholtz, R.L., et al., "Theory of Spread Spectrum Communications - A Tutorial", *IEEE Trans. on Communications*, Vol.COM-30, pp.855-884, May 1982.
- [2] Dixon, R.C., *Spread Spectrum Systems*, Wiley, New York, 1984.
- [3] Lee, William C., *Mobile Communications Engineering*, McGraw-Hill, New York, 1982.
- [4] Knight, T.A., *Fading Characteristics of the Land Mobile Radio Channel at 820 MHz*, M.Eng. Thesis, Department of Systems and Computer Engineering, Carleton University, Ottawa, 1985.
- [5] Caples, Edgar L., et al., "A UHF Channel Simulator for Digital Mobile Radio", *IEEE Trans. on Communications*, Vol.VT-29, No.2, May 1980.
- [6] Cooper, G. and Nettleton, R., "A Spread-Spectrum Technique for High-Capacity Mobile Communications", *IEEE Trans. on Vehicular Technology*, Vol.VT-27, No.4, November 1978.
- [7] Lee, W. and Yeh, Y., "On the Estimation of the Second Order Statistics of Log Normal Fading in Mobile Radio Environment", *IEEE Trans. on Communications*, Vol.COM-22, June 1974.
- [8] Suzuki, H., "A Statistical Model for Urban Radio Propagation", *IEEE Trans. on Communications*, Vol.COM-25, No.7, July 1977.

- [9] Aulin, T., "Characteristics of a Digital Mobile Radio Channel", *IEEE Trans. on Vehicular Technology*, Vol.VT-30, No.2, May 1981.
- [10] Borth, D.E. and Pursley, M.B., "Analysis of Direct-Sequence Spread-Spectrum Multiple Access Communication Over Rician Fading Channels", *IEEE Trans. on Communications*, Vol.COM-27, No.10, October 1979.
- [11] Turin, George L., "The Effects of Multipath and Fading on the Performance of Direct-Sequence CDMA Systems", *IEEE Trans. on Vehicular Technology*, Vol.VT-33, No.3, August 1984.
- [12] Chester, S. Gardner and Orr, John A., "Fading Effects on the Performance of a Spread Spectrum Multiple Access Communication System", *IEEE Trans. on Communications*, Vol.COM-27, No.1, January 1979.
- [13] Cooper, G.R. and McGillen, C.D., *Modern Communications and Spread Spectrum*, McGraw-Hill, New York, 1986.
- [14] Ziemer, R.E. and Peterson, R.L. *Digital Communications and Spread Spectrum Systems*, Macmillan Publishing Company, New York, 1985.
- [15] Holmes, Jack K., *Coherent Spread Spectrum Systems*, Wiley and Sons, New York, 1982.
- [16] Skaug, R. and Hjelmstad, J.F., *Spread Spectrum in Communications*, Peter Peregrinus Ltd. on behalf of IEE, London, 1985.
- [17] Simon, Marvin K., et al., *Spread Spectrum Communications*, Vol.3, Computer Science Press, Rockville, MD., 1985.
- [18] Rappaport, S.S. and Grieco, D.M., "Spread-Spectrum Signal Acquisition: Methods and Technology", *IEEE Communications Magazine*, Vol.22, No.6, 1984.

- [19] Guenther, W.C., *Concepts of Probability*, McGraw-Hill, Inc., New York, 1968.
- [20] Urkowitz, H., "Energy Detection of Unknown Deterministic Signals", *Proceedings of the IEEE*, Vol.55, No.4, April 1967.
- [21] Alem, W., et al., "Spread Spectrum Acquisition and Tracking Performance for Shuttle Communication Links", *IEEE Trans. on Communications*, Vol.COM-26, No.11, November 1978.
- [22] Hartman, H.P., "Analysis of a Dithering Loop for PN Code Tracking", *IEEE Trans. on Aerospace and Electronic Systems*, Vol.AES-10, No.1, January 1974.
- [23] Simon, M.K., "Noncoherent Pseudonoise Code Tracking Performance of Spread Spectrum Receivers", *IEEE Trans. on Communications*, Vol.COM-25, No.3, March 1972.
- [24] Blanchard, Alan, *Phase-Locked Loops*, Wiley-Interscience, New York, 1976.
- [25] Gardner, F., *Phaselock Techniques*, Wiley-Interscience, New York, 1979.
- [26] Pursley, Michael B., "Performance Evaluation for Phase-Coded Spread-Spectrum Multiple-Access Communication - Part 1: System Analysis", *IEEE Trans. on Communications*, Vol.COM-25, No.8, August 1977.
- [27] Weber, Charles L., et al., "Performance Considerations of Code Division Multiple-Access Systems", *IEEE Trans. on Vehicular Technology*, Vol.VT-30, No.1, February 1981.
- [28] Government of Canada, Department of Communications, "Technical Requirements for Cellular Radiotelephone Systems Operating in the Bands 825-845 MHz and 870-890 MHz", *Standard Radio System Plan*, Issue 1, September 1985.
- [29] Simon, M.K., et al., *Spread Spectrum Communications*, Vol.2, Computer Science Press, Rockville, MD., 1985.

- [30] Bierman, H., *Circuits and Software for Electronics Engineers*, McGraw-Hill Book Co., 1983.
- [31] Temes, G.C. and Mitra, S.K., *Modern Filter Theory and Design*, Wiley-Interscience, New York, 1973.
- [32] Blinchikoff, H.J., and Zverev, A.I., *Filtering in the Time and Frequency Domain*, John Wiley and Sons, Toronto, 1976.
- [33] De Couvreur, G.A., "Effect of Random Synchronization Errors in PN and PSK Systems", *IEEE Trans. on Aerospace and Electronic Systems*, Vol.AES-6, No.1, pp.98-100, January 1970.
- [34] Sheedy, K.A., *Workstation Preparation for Digital Integrated Circuit Design With CMOS Standard Cells*, Department of Electronics, Carleton University, Ottawa, Ontario, December 1985.
- [35] Mason, R.D. and Vallee, R.E., *CMOS1B Standard Cell Library*, Version 1.1, Department of Electronics, Carleton University, Ottawa, Ontario, June 1985.
- [36] Miller, Irwin and Freund, John E., *Probability and Statistics for Engineers*, Prentice-Hall, Englewood Cliffs, N.J., 1977.
- [37] Proakis, John G., *Digital Communications*, McGraw-Hill, New York, 1983.
- [38] Peterson, W. and Weldon, E., *Error Correcting Codes*, 2nd Edition, MIT Press, Cambridge, Mass., 1972.
- [39] Sarwate, D. and Pursley, M., "Crosscorrelation Properties of Pseudorandom and Related Sequences", *Proceedings of the IEEE*, Vol.68, No.5, May 1980.
- [40] Stremler, F.G., *Introduction to Communication Systems*, Addison-Wesley, Reading, MA., 1977.

- [41] Freeman, R.L., *Telecommunication Transmission Handbook*, 2nd Edition, Wiley and Sons, Toronto, 1981.
- [42] Ziemer, R.E. and Tranter, W.H., *Principles of Communications: - Systems, Modulation and Noise*, 2nd Edition, Houghton Mifflin Co., Boston, 1985.
- [43] Motorola Inc., *Linear and Interface Integrated Circuits*, Phoenix, Arizona, 1985.
- [44] Harris Corporation, *Analog Data Book*, Melbourne, Florida, 1982.
- [45] Weste, N. and Eshraghian, K., *Principles of CMOS VLSI Design: A Systems Perspective*, Addison-Wesley Publishing Co., Reading, MA., 1985.
- [46] Sedra, A.S. and Smith, K.C., *Microelectronic Circuits*, CBS College Publishing, New York, 1982.
- [47] Mavor, J., Jack, M.A., and Denyer, P.B., *Introduction to MOS LSI Design*, Addison-Wesley Publishers Limited, Don Mills, Ontario, 1983.
- [48] Chan, C., *Introduction to CMOS IC Design*, Department of Electronics, Carleton University, Ottawa, Ontario, 1984.
- [49] Hodges, D.A. and Jackson, H.G., *Analysis and Design of Digital Integrated Circuits*, McGraw-Hill, Toronto, 1983.
- [50] Mead, C. and Conway, L., *Introduction to VLSI Systems*, Addison-Wesley, Don Mills, Ontario, 1980.
- [51] Canadian Microelectronics Corporation, *Guide to the Integrated Circuit Implementation Services of the Canadian Microelectronics Corporation*, CMC, Kingston, Ontario, 1986.

APPENDIX A

A-1.1 PN Gold Code Generator

Since Gold codes have well bounded, as well as low cross-correlations values, they are often used in multiple access situations. Because Gold codes can be obtained by the addition of two different linear maximal length sequences (also known as m sequences), a brief discussion of m sequences will first be carried out. For a more detailed examination, references [2] and [15] should be consulted.

A linear maximal sequence is generated by an n stage shift register with linear feedback, as is illustrated in Figure A-1.1. The generated m sequence has length equal to $L = 2^n - 1$ bits, with each period containing 2^{n-1} ones and $2^{n-1} - 1$ zeros. Note that whether a sequence is linear or nonlinear depends on whether the feedback function is linear [2,15]. In the cases discussed here, the feedback function is expressed as a mod 2 sum which is linear. Although nonlinear sequences offer greater security, they are not pursued here since they are more difficult to generate and work with and also since the application being considered here does not require them.

Depending on the particular feedback connections or taps used in implementing the shift register generator, different sequences are obtained. Dixon [2] and Holmes [15] show tables that have the number of m sequences available from certain register lengths and some of the feedback taps that would result in m -sequences (note that in generating the m sequence, the all-zero state, for the shift registers, must not be

Figure A-1.1: n -Stage Shift Register With Linear Feedback.

allowed to occur, or else, the output will always remain at zero).

It is possible to represent a linear code by a polynomial of the form

$$f(x) = 1 + A_1x + A_2x^2 + A_3x^3 + \cdots + A_nx^n \quad (A-1.1-1)$$

Each term of this characteristic polynomial (except for the first, i.e., 1) corresponds to a stage of an n -stage binary shift register. Each of the coefficients $(A_1, A_2, \dots, A_n)$ is either 0 or 1 such that the feedback connections in the code generator are specified by the terms in the polynomial whose coefficient is a 1. For instance, a code generator with a characteristic polynomial given by $f(x) = 1 + x + x^2 + x^3 + x^7$ would have seven stages with feedback taken from the first, second, third and seventh stages¹. This connection could also be specified as (7, 3, 2, 1) [2].

Although it is possible to use m sequences for multiple access applications, the number of suitable sequences are far smaller than what can be obtained with Gold code generators; in addition, the peak cross-correlations for Gold codes are generally smaller than those for m sequences [37]. Section 8.2.3 of reference [37] goes into more detail. Now, as was indicated earlier, Gold codes are produced by modulo-2 addition of a pair of maximal linear sequences as is illustrated in Figure A-1.2. The generated codes have three possible cross-correlations $(R_x(k))$ values which are [37,15] $\{-1, -t(n), t(n) - 2\}$ where

$$t(n) = \begin{cases} 2^{(n+1)/2} + 1; & n \text{ odd} \\ 2^{(n+2)/2} + 1; & n \text{ even, } n \neq 0 \pmod{4} \end{cases} \quad (A-1.1-2)$$

Thus, it can be seen that the cross-correlation values $R_x(k)$ are such that

¹Holmes [15] indicates that some authors have shown that only $(2n - 1)$ bits observed without error is required to determine the feedback connections needed in reproducing any linear m sequence. However, Holmes goes on to point out that since most maximal length sequences have a very low chip energy-to-noise spectral density, observations without error are very unlikely.

Figure A-1.2: Gold Code Generator Arrangement. [15]

$$|R_z(k)| \leq \begin{cases} 2^{(n+1)/2} + 1; & n \text{ odd} \\ 2^{(n+2)/2} + 1; & n \text{ even, } n \neq 0 \pmod{4} \end{cases} \quad (\text{A-1.1-3})$$

In addition, it is possible to generate $2^n + 1$ linear code sequences of length $L = 2^n - 1$ (note that the maximum autocorrelation value would be equal to $2^n - 1$) [15]. In spread spectrum applications, it is desirable to use balanced codes, in which an almost equal number of ones and zeros occur (for m sequences and for balanced Gold codes, the number of ones exceed the number of zeros by 1). The reason for this code balance is that it allows the DC component in a code or code modulated signal to be neglected. In other words, one-zero balance can limit the degree of carrier suppression, as carrier suppression is dependent on the symmetry of the modulating signal [2].

Proakis [37] defines a "preferred pair" as two m sequences of length L with a periodic cross-correlation function that takes on the possible values $\{-1, -t(n), t(n) - 2\}$. The balanced Gold sequences are generated by using these preferred pairs. The steps, as well as the theorems by which preferred pairs are picked and Gold codes are generated are very well described in reference [15]. Using the table in Appendix C of reference [38] and following the steps outlined in reference [15] the first polynomial code of the preferred pair for a 13 stage ($L = 2^{13} - 1$) was found as

$$\underbrace{1}_{\text{(root)}} \underbrace{20033}_{\text{octal}} = \underbrace{10000000011011}_{\text{binary}} \quad (\text{A-1.1-4})$$

Thus, $f(x)$, which would represent register 2 (Figure A-1.2), would be given by

$$f(x) = x^{13} + x^4 + x^3 + x + 1 \quad (\text{A-1.1-5})$$

(The "System Design Parameters" section indicates the motivations behind the se-

lection of a 13 stage generator.) Following the steps discussed in reference [15], the polynomial root of the second code that completes the preferred pair is $2^{(n-1)/2} + 1 = 2^{(13-1)/2} + 1 = 2^6 + 1 = 65$. Thus, the second code $g(x)$ (register 1) is given by

$$\underbrace{65}_{\text{(root)}} \quad \underbrace{3 \ 3 \ 3 \ 4 \ 3}_{\text{octal}} \quad \underbrace{1 \ 1 \ 0 \ 1 \ 1 \ 0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 1 \ 1}_{\text{binary}}$$

$$g(x) = x^{13} + x^{12} + x^{10} + x^9 + x^7 + x^6 + x^5 + x + 1 \quad (A - 1.1 - 6)$$

Now, in order to obtain a balanced code, the registers must have certain initial conditions. Register 1 must be such that the entry in the first stage is zero as shown below

$$[x \ x \ x \ x \ x \ x \ x \ x \ x \ x \ x \ x \ 0] \quad (A - 1.1 - 7)$$

where the symbol "x" represents don't care conditions (i.e., 0 or 1). The initial condition for register 2 is found by doing the differentiation $\frac{d}{dx} [xf(x)]$, with the coefficients interpreted mod 2. The result of this differentiation is then divided by $f(x)$ as is illustrated below.

$$g_I(x) = \frac{d}{dx} [xf(x)] = \frac{d}{dx} (x^{14} + x^5 + x^4 + x^2 + x) = x^4 + 1 \quad (A - 1.1 - 8a)$$

$$\frac{g_I(x)}{f(x)} = \frac{1 + x^4}{1 + x + x^3 + x^4 + x^{13}} \quad (A - 1.1 - 8b)$$

Carrying out this division as shown in Appendix A-1.2 results in the following initial condition for register 2

$$[0\ 1\ 1\ 0\ 1\ 1\ 0\ 1\ 1\ 0\ 1\ 1\ 1]$$

(A - 1.1 - 9)

Now, to generate different Gold sequences, all that is required is to use different initial conditions for register 1 (i.e., changing the "x" values around) and add it to register 2 as Figure A-1.3 illustrates. Note that since register 1 contains $n - 1$ don't care conditions, only 2^{n-1} Gold sequences will be balanced, out of the $2^n + 1$ possible Gold sequences that can be generated. In this particular case, where $n = 13$, this corresponds to 4096 balanced Gold sequences out of the total 8193 Gold sequences. It should be noted that although Gold codes are most often used for CDMA applications, there are other codes which are also appropriate for CDMA applications, such as Kasami sequences. For more detail on other suitable sequences, the paper by Sarwate and Pursley [39] should be examined.

After the basic block diagram of the PN Gold code generator had been obtained (Figure A-1.3), it was then realized using high speed CMOS chips. The resulting schematic is given in Figure A-1.4. Considering the initial contents of register 1 equation (A-1.1-7), it can be seen that for balanced Gold codes, the all zero state is valid for this register. However, since the initial state for register 2 is nonzero, the all zero state will not appear. In fact, it should not be allowed to occur. The schematic of Figure A-1.4 thus includes a check by which an all zero condition for register 2 is used to reload both registers 1 and 2 with their respective initial conditions (a low on pin 9 of the 74HC195s). Additionally, if a restart of the sequence is desired, the generator can be reset (low on pin 1 of 74HC195s), producing an all zero condition in registers 1 and 2. The all zero condition of register 2 is next sensed and is then used to reload both registers 1 and 2.

Figure A-1.3: Balanced Gold Code of Length $2^{13} - 1$.

Figure A-1.4: HCMOS Realization of 13 Stage Gold Code Generator.

A-1.2 Gold Code Generator Initial Contents

The division which was required to obtain the initial contents of register 2 as discussed in Section A-1.1 is as follows:

$$\begin{array}{r}
 \begin{array}{cccccccccccccccc}
 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\
 1 & + & x & + & x^2 & + & x^4 & + & x^5 & + & x^7 & + & x^8 & + & x^{10} & + & x^{11} & + & x^{14}
 \end{array} \\
 1 + x + x^3 + x^4 + x^{13} \overline{) 1 + x^4} \\
 \underline{1 + x^4 + x + x^3 + x^{13}} \\
 x + x^3 + x^{13} \\
 x \quad \quad \quad + x^2 + x^4 + x^5 + x^{14} \\
 \hline
 x^2 + x^3 + x^4 + x^5 + x^{13} + x^{14} \\
 \underline{x^2 + x^3 \quad \quad \quad + x^5 \quad \quad \quad + x^6 + x^{15}} \\
 x^4 + x^6 + x^{13} + x^{14} + x^{15} \\
 \underline{x^4 \quad \quad \quad + x^5 + x^7 + x^8 + x^{11}} \\
 x^5 + x^6 + x^7 + x^8 + x^{13} + x^{14} + x^{15} + x^{17} \\
 \underline{x^5 + x^6 \quad \quad \quad + x^8 \quad \quad \quad + x^9 + x^{18}} \\
 x^7 + x^9 + x^{13} + x^{14} + x^{15} + x^{17} + x^{18} \\
 \underline{x^7 \quad \quad \quad + x^8 + x^{10} + x^{11} + x^{20}} \\
 x^8 + x^9 + x^{10} + x^{11} + x^{13} + x^{14} + x^{15} + x^{17} + x^{18} + x^{20} \\
 \underline{x^8 + x^9 \quad \quad \quad + x^{11} \quad \quad \quad + x^{12} + x^{21}} \\
 x^{10} + x^{12} + x^{13} + x^{14} + x^{15} + x^{17} + x^{18} + x^{20} + x^{21} \\
 \underline{x^{10} \quad \quad \quad + x^{13} + x^{14} \quad \quad \quad + x^{11} + x^{23}} \\
 x^{11} + x^{12} + x^{15} + x^{17} + x^{18} + x^{20} + x^{21} + x^{23} \\
 \underline{x^{11} + x^{12} + x^{15} \quad \quad \quad + x^{14} + x^{24}} \\
 x^{14} + x^{17} + x^{18} + x^{20} + x^{21} + x^{23} + x^{24}
 \end{array}$$

Following the steps discussed in reference [15], the initial contents for register 2 can now be determined from the above results as

[0 1 1 0 1 1 0 1 1 0 1 1 1]

A-1.3 Continuously Variable Slope Delta (CVSD) Modulator

In spread spectrum systems, the analog information signal is normally digitized first before the modulation takes place. The digitization process involves three major steps: (1) sampling, (2) quantization, and (3) binary (digital) representation of the quantized signal. The sampling must be done at a rate that is at least twice as high as the highest frequency component in the sampled signal. This is known as the Nyquist rate and allows the sampled signal to be recovered without distortion caused by aliasing. The aliasing error comes about from the fact that sampling introduces spectral replicas about multiples of the sampling frequency. If these replicas should overlap, then the original signal cannot be retrieved (by filtering) without distortion. Thus, the Nyquist criterion ensures that the spectral replicas are sufficiently separated from the desired spectral components to allow its recovery without distortion [40].

Following the sampling of the analog signal, it is next quantized or divided up into a number of discrete levels. Thus at each sample time, the level that is closest to the sampled signal is used to represent it (Figure A-1.5). This procedure naturally introduces some fluctuations about the true value, which is regarded as quantization noise. Of course, an increase in the number of quantization levels assigned will tend to decrease the quantization noise. To complete the digitization process, each level is now assigned a binary digit [40].

From the above, it is obvious that the higher the quantization level, the closer to the original the recovered signal will be. On the otherhand, a higher quantization level means that a greater number of bits is required to represent each level, consequently resulting in a higher data transmission bandwidth. Conflicting with this result however, is the desirability in spread spectrum systems to keep the data rate low, such that higher processing gains are possible. For telephone quality performance, a data rate (for the digitized voice signal) of 64 kilobits per

Figure A-1.5: Eight Level Quantization of Half a Cycle of a One Volt Peak Sinusoid. [40]

second is normally required. This comes from the fact that despite having frequency components from 20 Hz to 20 KHz, speech signals have most of their energy and intelligibility content in the range of 300 to 3400 Hz (the standard bandwidth of a voice channel) [41]. Thus, following the Nyquist criterion, the voice signal, bandlimited to 3.4 KHz, is sampled at about 8 KHz. This is followed by a 256 level quantization that results in an 8 bit binary representation. Thus, if there are 8000 samples per second, each represented by 8 bits, then the voice signal would have a 64,000 bits per second data rate.

By using a continuously variable slope delta modulator (CVSD), it is possible to attain telephone quality speech performance at a reduced data rate. Moreover, intelligibility is maintained with a bit error rate (BER) as high as 10^{-2} . In order to understand the functioning of a CVSD, it is very instructive to first examine the operation of a basic delta modulator. The delta modulator of Figure A-1.6 is composed of a comparator, a pulse generator and an integrator. The input signal $m(t)$ is compared to the integrator output. If $m(t)$ is greater than the integrator output, the comparator output will be positive. On the other hand, if $m(t)$ is less than the feedback signal, the result will be negative. These positive and negative outputs are used to generate positive and negative pulses as shown in Figure A-1.7(a) which are integrated and fed back to the comparator. From Figure A-1.7(b), it can be seen that the integrator output is a stepped approximation to the continuous input signal $m(t)$. Thus, a similar integrator at the receiver (Figure A-1.8), followed by a lowpass filter (for smoothing), would produce an output very close to the original signal, $m(t)$ [40,41,42].

With a little thought, it is not hard to see that the output pulses are indicative of the slope (i.e., derivative) of the analog input signal - a positive pulse for a positive slope and a negative pulse for a negative slope (remember that the comparison is made relative to the signal value in the previous sampling instant). Therefore, if the input signal is increasing (positive slope), the output will be a

Figure A-1.6: Delta Modulator. [42]

(a)

(b)

Figure A-1.7: Delta Modulator Waveforms: (a) Modulator output; (b) Input waveform and staircase approximation . [42]

Figure A-1.8: Delta Modulator Receiver.

series of positive pulses while the integrator output will be a positive going stepped (staircase) function. However, if the input signal is decreasing (negative slope), the output will be a series of negative pulses so that the integrator output will be a negative going stepped function. For a steady-state signal (0 slope), an alternating positive-negative pulse pattern results as the output; the integrator output will consequently have a zero value. Note that the integrator output increases or decreases by a certain voltage step that is a function of the integrator gain. This can be seen by considering the equation for the voltage change in an ideal current to voltage integrator [2]

$$\Delta V_{oc} = \frac{1}{C} \int I_p dt = \frac{I_p}{C} \int df \quad (A - 1.3 - 1)$$

where I_p is the current pulse level. In a simple delta modulator as discussed above, the fixed integrator gain which results in a fixed step size can be a disadvantage. This can be seen if the signal to be transmitted has a significant dynamic range. If a large and/or rapid change occurs, the fixed step size may not be large enough to adequately follow the intended waveform, thus introducing distortion (this problem is usually referred to as slope overload). On the otherhand, if the changes are slow and/or small, a smaller step size may be needed to effectively follow the input signal.

The solution to the problem of varying signal rates can be solved with a CVSD modulator. These modulators (Figure A-1.9) adapt to signal level changes by using an algorithm (level detect algorithm) which monitors the past few outputs of the CVSD modulator in a shift register of either 3 or 4 bits. The shift register is monitored until it contains all 1's or all 0's. When this condition (called coincidence) occurs, it indicates that the integrator gain or step size is too small. This coincidence output (low when coincidence occurs and high otherwise) is subsequently used to charge a single pole lowpass filter (for voice codecs like the CVSD, this lowpass filter is quite often called a syllabic filter and has typical time constant values of

Figure A-1.9: A General Block Diagram of a CVSD Encoder. [41]

6 ms to 50 ms). The lowpass or syllabic filter output voltage is used to control the integrator gain through a pulse amplitude modulator whose other input is a bit (from the CVSD digital output), which controls whether the integrator slope should be positive or negative [41,43]. Note that in the diagram of Figure A-1.9, the slope magnitude control is the device by which the integrator time constant is varied and the slope polarity switch ensures that a negative pulse is integrated for a zero output while a positive pulse is integrated for a one output. Figures A-1.10 and A-1.11 show a CVSD decoder and CVSD waveforms respectively.

Now that the operation of the CVSD modulator has been examined, the method by which it is utilized will be outlined. The input voice signal is first bandlimited by the HC-5512C CVSD monolithic filter, whose functional diagram appears in Figure A-1.12(a). In this bandlimiting operation, only the transmit filter block is used, with the typical performance characteristic shown in Figure A-1.12(b). Note that signals above 3.4 KHz and below 200 Hz are rejected. Although the recommended clock frequency for the filter is 2.048 MHz, it was much easier to obtain a 2 MHz signal. The use of a 2 MHz clock is not expected to appreciably affect the performance of the filter.

In using the MC3418 CVSD modulator (4 bit level detect algorithm), whose block diagram appears in Figure A-1.13, several design requirements must be considered. These include selection of clock rate, step size for high level signals and the minimum allowable step size. With the MC3418 CVSD modulator, commercial telephone quality speech performance can be attained at 32 Kbits/sec and above. Following the directions and recommendations outlined in reference [43], the component values used in the implementation of the CVSD modulator were obtained (just about the same component values were also used at the receiver). The configuration for the CVSD and its filter appear in Figure A-1.14. Note that the clock input to the MC3418 is inverted. This is because the digital output appears on the falling edge of the clock, so that an inverter was needed to synchronize the digital output with

Figure A-1.10: CVSD Decoder. [41]

Figure A-1.11: CVSD Waveforms. [41]

(a)

(b)

Figure A-1.12- HC-5512C CVSD Monolithic Filter: (a) Functional diagram, (b) Performance characteristic. [44]

Figure A-1.13: MC3418 CVSD Block Diagram. [43]

Figure A-1.14: CVSD Configuration.

the digital stream coming from the PN generator, where the digital output appears at the rising edge of the clock.

A-2.1 Theoretical Analysis

In order to better appreciate how the CMOS circuits implemented here function, a brief examination of MOS transistor fundamentals and other basic CMOS concepts will now be given.

A-2.1.1 MOS Transistor Fundamentals

CMOS circuits consist of two types of transistors - n-types (NMOS) and p-type (PMOS) transistors. The circuits are fabricated in silicon by using negatively doped silicon (rich in electrons) for the NMOS and positively doped silicon (rich in holes) for the PMOS. Figures A-2.1(a) and A-2.1(b) show the typical physical structures and circuit symbols for both NMOS and PMOS transistors [45].

Considering Figure A-2.1(a), it can be seen that the n-transistor consists of a p-type silicon separating two diffused areas of n-type silicon. The section separating these two regions is topped with an insulator (usually silicon dioxide - SiO_2), followed by a conducting electrode called the Gate. The p-transistor has a similar structure (Figure A-2.1(b)), except that an n-type silicon separates two p-type diffused areas. The two n-type diffused areas in the NMOS transistor (p-type for PMOS transistor) are called the Drain and the Source. Since the two terminals are physically equivalent and thus interchangeable, the name assignment depends on the direction of current flow. The MOS transistors can be basically viewed as simple on/off switches, with the gate controlling the flow of electrical current between the drain and the source, and thus determining when the transistor is on or off.

The operation of the MOS transistor is in a large measure affected by the threshold voltage. This is defined as the voltage difference that must exist between the gate and the source before the transistor turns on and begins to conduct. An

Figure A-2.1: MOS Transistor Structures: (a) NMOS transistor; (b) PMOS transistor. [45]

NMOS transistor which has a positive threshold voltage ($V_{tn} > 0$) is called an Enhancement type while those that have a negative threshold ($V_{tn} < 0$) are called depletion type transistors. On the otherhand, PMOS transistors with a threshold voltage V_{tp} greater than zero are called depletion type devices while those with V_{tp} less than zero are called enhancement type devices. Figures A-2.2 and A-2.3 illustrate the relative conduction of NMOS and PMOS enhancement and depletion type transistors against the difference in gate and source voltages. These graphs are plotted in terms of the source to drain current (I_{ds}) and the gate to source voltage (V_{gs}), for a fixed drain to source voltage, V_{ds} . Since most CMOS ICs at the present use the enhancement type transistor [45], this is the type of transistor that shall be discussed here.

A-2.1.1.1 NMOS Enhancement Transistor

The physical structure of an n-channel enhancement type transistor appears in Figure A-2.4. It is made up of a moderately doped p-type silicon substrate into which two heavily doped n^+ regions (the source and drain) have been diffused [45]. The narrow part of the p-type substrate between the drain and source is called the channel and is covered by a thin insulating layer of silicon dioxide (SiO_2) which is called the gate oxide. A polycrystalline silicon (polysilicon) electrode referred to as the gate is used to cover the oxide layer (note that polysilicon is silicon that is not made up of a single crystal). Because of the insulating properties of the oxide, there is virtually no gate current, while its high dielectric strength allows high gate fields to be applied [45].

The operation of the NMOS enhancement transistor is as described in the following paragraphs. With zero gate to source voltage ($V_{gs} = 0$), no current will flow no matter what the voltage between the drain and source (V_{ds}) is and the transistor is said to be in its "cut-off" mode. The reason for zero current flow is because the drain and source are effectively insulated from each other by the two

Figure A-2.2: Relative Conduction of Enhancement Mode Transistors Versus the Gate to Source Voltage (V_d , fixed): (a) n-channel; (b) p-channel. [45]

Figure A-2.3: Relative Conduction of Depletion Mode Transistors Versus the Gate to Source Voltage (V_d , fixed): (a) n-channel; (b) p-channel. [45]

Figure A-2.4: Physical Structure of an n-Channel Enhancement Type Transistor.
[45]

reverse biased p-n junctions indicated by the diode symbols in Figure A-2.4. When a voltage applied to the gate is positive with respect to the source and substrate, an electric field E is produced across the substrate, which attracts electrons towards the gate while repelling holes (note that the substrate is usually maintained at the same potential as the source; in the case of NMOS transistors, this is ground while for PMOS transistors, it is V_{DD}). If the applied gate voltage is sufficiently large ($\geq V_{tn}$), the area under the gate changes from p-type to n-type because of the accumulation of attracted electrons, thereby providing a conduction path from the source to the drain. The channel created due to the inversion of the p-type silicon under the gate is called the inversion layer. Figure A-2.5 illustrates some of the concepts discussed above. Note that this inversion layer (or n-channel) will only exist so long as the applied electric field exists.

Now, with $V_{gs} \geq V_{tn}$, if a voltage $V_{ds} > 0$ is applied, the horizontal and vertical components of the electric field due to V_{ds} and V_{gs} interact, resulting in the conduction of electrons from source to drain. As V_{ds} is increased, the channel characteristics begin to change. At the source end of the channel the full gate voltage is still used to invert the channel. On the otherhand, only the difference between the gate and the drain voltage, V_{gd} , is effective in inverting the channel at the drain end of the channel. The result of this is, as long as $V_{gs} - V_{tn}$ is greater than V_{ds} or $V_{gd} > V_{tn}$ (V_{gd} is the gate to drain voltage), the channel current remains a function of both the gate and drain voltages and the transistor stays in what is termed its linear, resistive, or unsaturated region. However, if $V_{gs} - V_{tn} < V_{ds}$ or $V_{gd} < V_{tn}$, then the channel becomes pinched-off; that is, it no longer reaches the drain (see Figure A-2.6). Under this condition, the transistor is said to be in its pinch-off or saturated region, where the magnitude of the drain current is controlled by the gate voltage and is almost independent of the drain voltage (note that the voltage across the pinched off region tends to remain at $V_{gs} - V_{tn}$ no matter what the value of V_{ds} is). The conduction in this saturated state is brought about by the drift mechanism

Figure A-2.5: Creation of the Inversion Layer in an NMOS Transistor. [45]

Figure A-2.6: NMOS Transistor Behaviour With Different Terminal Voltages. [45]

of electrons being attracted by the positive drain voltage. Electrons leaving the channel pass into the drain depletion region (empty of mobile carriers) and are thus accelerated towards the drain. Note that because the depletion region does not have mobile carriers the channel is insulated from the substrate, thus allowing no significant current to flow through the substrate.

A-2.1.1.2 PMOS Enhancement Transistor

As Figure A-2.7 shows, the PMOS enhancement mode transistor is very similar to the NMOS enhancement transistor. In the PMOS transistor, the n- and p-type regions are reversed (relative to the NMOS case); this results in the requirement for a negative gate voltage (with respect to the source) to operate the transistor. Applying a negative gate to source voltage, V_{gs} , will draw holes into the area under the gate. If this gate to source voltage is sufficiently negative (i.e., less than V_{tp} , where V_{tp} is the PMOS threshold voltage, which is a negative quantity), then the channel under the gate will change from n-type to p-type, thereby providing a conduction path between the source and drain. Note that in the PMOS case, conduction is carried on by holes (as opposed to electrons for NMOS); holes are swept from the positive source, across the channel, to the drain, by a negative voltage (with respect to the source). Since the operation of the PMOS enhancement transistor parallels in all other respects what has already been described for the NMOS case, it will not be repeated.

A-2.1.2 Basic MOS Concepts

Now that a basic description of MOS transistor operation has been given, it would be enlightening to briefly examine the equations and parameters that govern the way it functions.

The ideal first order equations [45,46] that characterize both the NMOS and PMOS enhancement transistors are given in equations (A-2.1.2-1) and (A-2.1.2-2)

Figure A-2.1: Physical Structure of a p-Channel Enhancement Type Transistor.
[45]

respectively.

$$I_{ds,n} = \begin{cases} 0; & V_{gs} - V_{tn} \leq 0 & \text{cut-off} & (a) \\ \beta_n \left[(V_{gs} - V_{tn})V_{ds} - \frac{V_{ds}^2}{2} \right]; & 0 < V_{ds} < V_{gs} - V_{tn} & \text{linear} & (b) \\ \frac{\beta_n}{2} (V_{gs} - V_{tn})^2; & 0 < V_{gs} - V_{tn} < V_{ds} & \text{saturation} & (c) \end{cases} \quad (A-2.1.2-1)$$

$$I_{sd,p} = -I_{ds,p} = \begin{cases} 0; & V_{gs} - V_{tp} > 0 & \text{cut-off} & (a) \\ \beta_p \left[(V_{gs} - V_{tp})V_{ds} - \frac{V_{ds}^2}{2} \right]; & 0 > V_{ds} > V_{gs} - V_{tp} & \text{linear} & (b) \\ \frac{\beta_p}{2} (V_{gs} - V_{tp})^2; & 0 > V_{gs} - V_{tp} > V_{ds} & \text{saturation} & (c) \end{cases} \quad (A-2.1.2-2)$$

where β_n and β_p are the MOS transistor gain factors for the n and p transistors respectively. Figure A-2.8 shows the V-I characteristics for both the n - and p -enhancement transistors.

β_n and β_p , the gain factors, can be expressed as [45]

$$\beta_n = \frac{\mu_n \xi}{t_{ox}} \left(\frac{W}{L} \right) \quad (A-2.1.2-3a)$$

$$\beta_p = \frac{\mu_p \xi}{t_{ox}} \left(\frac{W}{L} \right) \quad (A-2.1.2-3b)$$

where μ_n and μ_p are the effective surface mobility of the electrons and holes in the respective channels, ξ is the permittivity of the gate insulator, t_{ox} is the thickness of

Figure A-2.8: V-I Characteristic for the n and p Enhancement Transistors [45]

the gate insulator, W is the width of the channel and L is the length of the channel (Figure A-2.9). Note that the terms $(\mu_n \xi / t_{ox})$ and $(\mu_p \xi / t_{ox})$ depend on the process used in the fabrication stage and accounts for factors such as doping density and gate oxide thickness. It is often referred to as the process gain factor, K_p , and has typical values on the order of 10 to 30 μ_A/V^2 . The mobilities μ_n and μ_p indicate the ease with which carriers can drift in the substrate material and can be expressed as

$$\mu_n \text{ or } \mu_p = \frac{\text{average carrier drift velocity}}{\text{electric field}} \quad (A - 2.1.2 - 4)$$

The unit for the mobility is given as $cm^2/V\text{-sec}$ if the velocity is in cm/sec . and the electric field is in V/cm . The ratio, W/L , often called the aspect ratio, depends on the actual layout of the transistor and as such is under the control of the IC designer.

As the discussion on MOS characteristics covered here was of necessity quite brief, a more detailed discussion can be obtained from reference [45]. The next two sections will briefly examine some basic CMOS logic circuits and will give a transient analysis of the CMOS inverter which exhibits all the essential features of CMOS logic gates except for the function.

A-2.1.2.1 CMOS Logic Circuits

An examination of some basic CMOS logic circuits is helpful in illustrating how CMOS circuits are realized. Figure A-2.10 shows both the transistor and schematic representations for a CMOS inverter. When $V_A = V_{DD} > V_{tn}$, the NMOS transistor is on while the PMOS transistor is off. The result is that V_Y will be very close to ground potential. On the otherhand, when $V_A = 0 < V_{tn} < V_{tp} + V_{DD}$, the NMOS transistor will be off, while the PMOS transistor will be on, resulting in a V_Y that is very close to V_{DD} . Thus, V_Y is the inverse of V_A .

4

of/de

4

MicroD

Figure A-2.9: MOS Transistor Geometric Terms. [45]

Figure A-2.10: CMOS Inverter: (a) transistor representation; (b) schematic representation; (c) logic function

Figure A-2.11 shows a CMOS NAND gate. In this case, V_Y will be near ground potential only if both V_A and V_B are close to V_{DD} ; otherwise, V_Y will be near V_{DD} . Note that to have larger input nand gates, all that is required is to place an NMOS transistor in series on the n side and a PMOS transistor in parallel on the p side, for each additional input to the gate, as shown in Figure A-2.12. The CMOS NOR gate of Figure A-2.13 will ideally have a high output only if both V_A and V_B are low, remaining low if this is not the case. To increase its fan in (the number of inputs), an n-transistor is placed in parallel on the n side and a p-transistor is placed in series on the p side, for each additional input to the gate (Figure A-2.14).

An important characteristic of not only the above CMOS circuits, but of all CMOS circuits is their low static power dissipation. This is because under constant or static input conditions, there is never a path from V_{DD} to V_{SS} (ground) for any combination of inputs, thereby resulting in a low power dissipation (the power consumption of CMOS circuits will be addressed in a later section). Another characteristic of note possessed by CMOS logic circuits is the fact that there is no need to ratio the aspect ratios of the load and drive transistors in order to obtain the correct logic levels (in the circuits discussed above, the p-transistors were the "load" transistors while the n-transistors functioned as the "drive" transistors). The reason for not requiring ratioing is the fact that the p-transistor, which is a good conveyer of highs (ones) and a poor conveyer of lows (zeros) is used solely to pass highs in CMOS circuits. Similarly, the n-transistor which is a good conveyer of lows and a poor conveyer of highs, is used solely to pass lows. Thus, fully restored logic levels are attained.

A-2.1.2.2 Voltage Transfer and Switching Characteristics of the CMOS Inverter.

As was indicated earlier, the CMOS inverter exhibits all of the essential features of CMOS logic gates. Thus, the voltage transfer and switching characteristics for

(c)

V_A	V_B	V_Y
0V	0V	5V
0V	5V	5V
5V	0V	5V
5V	5V	0V

Figure A-2.11: CMOS NAND Gate: (a) transistor representation; (b) schematic representation; (c) logic function

Figure A-2.12: n-input NAND Gate. [45]

Figure A-2.13: CMOS NOR Gate: (a) transistor representation; (b) schematic representation; (c) logic function.

Figure A-2.14: n-input NOR Gate. [45]

the CMOS inverter will be examined in order to better understand how CMOS gates work.

In order to find the variation in output voltage (V_0) for changes in the input voltage (V_{in}) for the CMOS inverter, it can be thought of as having five regions of operation [45]; where the n- and p-transistors operate in either the cut-off region, the linear region or the saturation region as is shown in Table A-2.1. The five regions of operation (A, B, C, D, E) for the CMOS inverter is illustrated in Figure A-2.15 which shows the CMOS inverter DC transfer characteristic. In region A, $0 \leq V_{in} \leq V_{tn}$, so that the n-transistor is cut-off ($I_{ds,n} = 0$), whereas the p-transistor is in its linear region. However, since $I_{ds,n} = -I_{ds,p}$, $I_{ds,p}$ is also zero. Thus, $V_{ds,p} = 0$. Since $V_{ds,p} = V_0 - V_{DD}$, then

$$V_0 = V_{DD} \quad (A - 2.1.2.2 - 1)$$

In region B, $V_{tn} \leq V_{in} \leq V_{DD}/2$ (see Figures A-2.15 and A-2.16) and the p-transistor goes into its linear region, while the n-transistor goes into saturation (the n-transistor now acts like a current source, while the p-transistor acts like a resistor). Since the n-device is in saturation, the current $I_{ds,n}$ can be written from equation (A-2.1.2-1c) as

$$I_{ds,n} = \beta_n \frac{(V_{in} - V_{tn})^2}{2} \quad (A - 2.1.2.2 - 2)$$

where $V_{gs} = V_{in}$. For the p-transistor, $V_{gs} = V_{in} - V_{DD}$ while $V_{ds} = V_0 - V_{DD}$. Thus, from equation (A-2.1.2-2b)

Figure A-2.15: CMOS Inverter DC Transfer Characteristic. [45]

Figure A-2.16: Equivalent Circuit for Inverter Operation in Region B. [45]

Table A-2.1: The Three Regions of Operation for the n- and p-Transistors. [45]

	CUTOFF	LINEAR	SATURATION
p-device	$V_{gs} > V_{th}$	$V_{ds} < V_{gs} - V_{th}$ $V_{ds} < V_{th} + V_{dd}$	$V_{ds} > V_{gs} - V_{th}$ $V_{ds} > V_{th} + V_{dd}$
	$V_{gs} < V_{th}$	$V_{ds} < V_{th}$ $V_{ds} < V_{th} + V_{dd}$	$V_{ds} > V_{th}$ $V_{ds} > V_{th} + V_{dd}$
n-device	$V_{gs} > V_{th}$	$V_{ds} < V_{gs} - V_{th}$ $V_{ds} < V_{th} + V_{dd}$	$V_{ds} > V_{gs} - V_{th}$ $V_{ds} > V_{th} + V_{dd}$
	$V_{gs} < V_{th}$	$V_{ds} < V_{th}$ $V_{ds} < V_{th} + V_{dd}$	$V_{ds} > V_{th}$ $V_{ds} > V_{th} + V_{dd}$

$$I_{sd_p} = -I_{ds_p} = \beta_p \left[(V_{in} - V_{DD} - V_{tp})(V_0 - V_{DD}) - \frac{1}{2} (V_0 - V_{DD})^2 \right] \quad (A-2.1.2.2-3)$$

Since

$$I_{ds_n} = -I_{ds_p} = I_{sd_p} \quad (A-2.1.2.2-4)$$

the output voltage V_0 can be written in terms of the input voltage V_{in} by substituting (A-2.1.2.2-2) and (A-2.1.2.2-3) into (A-2.1.2.2-4), thus resulting in [45]

$$V_0 = (V_{in} - V_{tp}) + \left[(V_{in} - V_{tp})^2 - 2 \left(V_{in} - \frac{V_{DD}}{2} - V_{tp} \right) V_{DD} - \frac{\beta_n}{\beta_p} (V_{in} - V_{tp})^2 \right]^{1/2} \quad (A-2.1.2.2-5)$$

Next, considering region C, it is seen that both the n- and p-transistors are saturated. Noting again that $V_{gs} = V_{in}$ for the n transistor and $V_{gs} = V_{in} - V_{DD}$ for the p-transistor, and substituting equations (A-2.1.2-1c) and (A-2.1.2-2c) into equation (A-2.1.2.2-4), V_{in} can be written as [45]

$$V_{in} = \frac{V_{DD} + V_{tp} + V_{tn} \sqrt{\frac{\beta_n}{\beta_p}}}{1 + \sqrt{\frac{\beta_n}{\beta_p}}} \quad (A-2.1.2.2-6)$$

This can be interpreted as implying that region C exists for only one value of V_{in} which depends on V_{tp} , V_{tn} , β_n and β_p . To obtain the corresponding output voltage V_0 , Table A-2.1 should be consulted. Here, the output voltage for the n- and p-channel transistors in saturation are given as [45]

$$\text{n-channel: } V_0 > V_{in} - V_{tn} \quad (A-2.1.2.2-7a)$$

$$p\text{-channel: } V_0 < V_{in} - V_{tp} \quad (A-2.1.2.2-7b)$$

Combining equations (A-2.1.2.2-7a) and (A-2.1.2.2-7b), results in

$$V_{in} - V_{tn} < V_0 < V_{in} - V_{tp} \quad (A-2.1.2.2-8)$$

Note that if V_{in} is made equal to $-V_{tp}$, then $V_0 = V_{in}$ in region C, with the value of V_{in} determined by the ratio β_n/β_p (see equation (A-2.1.2.2-6)). This voltage, at which $V_0 = V_{in}$, is usually referred to as the gate threshold, V_{inv} and is the input voltage below which the output rises and above which the output falls [45,47]. Figure A-2.17 shows the effect of varying the ratio β_n/β_p on the inverter transfer characteristic.

In region D (Figure A-2.15), the input voltage is such that $V_{DD}/2 < V_{in} \leq V_{DD} + V_{tp}$. This means that the p-transistor is in saturation while the n-transistor is in the linear region. Using $V_{gs} = V_{in} - V_{DD}$ in equation (A-2.1.2-2c) for the p device and $V_{gs} = V_{in}$ in equation (A-2.1.2-1b) for the n device, and substituting these equations (A-2.1.2-1b and A-2.1.2-2c) into (A-2.1.2.2-4) results in [45]

$$V_0 = (V_{in} - V_{tn}) - \left[(V_{in} - V_{tn})^2 - \frac{\beta_p}{\beta_n} (V_{in} - V_{DD} - V_{tp})^2 \right]^{1/2} \quad (A-2.1.2.2-9)$$

Finally, region E has $V_{in} \geq V_{DD} + V_{tp}$. Thus, the p-device is cut-off ($I_{sdp} = 0$) as $V_{gs} = V_{in} - V_{DD}$ is greater or equal to V_{tp} . The n-device on the otherhand is in its linear region ($V_{gs} > V_{tn}$) so that the output voltage is given by

$$V_0 = 0 \quad (A-2.1.2.2-10)$$

Figure A-2.17: Effect of Varying β_n/β_p on the Inverter Transfer Characteristic.
[45]

Note that this comes about due to the fact that the on resistance of the n-transistor is negligible compared to the equivalent resistance of the cut-off p-transistor.

One final feature of Figure A-2.15 that should be considered is the dotted line in the middle of the plot. This represents the time (during transition) that both transistors are momentarily ON, resulting in a short pulse of current being drawn from the power supply.

A parameter that is closely linked to the input-output voltage characteristics is the noise margin [45], which can be thought of as the amount of noise voltage that can be allowed at the input of a gate without affecting the output. This noise margin (or noise immunity) is usually specified in terms of the LOW noise margin (NM_L) and the HIGH noise margin (NM_H) as given below [45]

$$NM_L = |V_{IL_{max}} - V_{OL_{max}}| \quad (A - 2.1.2.2 - 11a)$$

$$NM_H = |V_{OH_{min}} - V_{IH_{min}}| \quad (A - 2.1.2.2 - 11b)$$

where

$V_{IH_{min}}$ = the minimum voltage recognized as a HIGH input

$V_{IL_{max}}$ = the maximum voltage recognized as a LOW input

$V_{OH_{min}}$ = the minimum voltage recognized as a HIGH output

$V_{OL_{max}}$ = the maximum voltage recognized as a LOW output

Figure A-2.18 clearly illustrates the above definitions. The LOW and HIGH noise margins of equation (A-2.1.2.2-11) can also be written in terms of V_{in} , V_{tp} and V_{DD} [45].

Figure A-2.18: Noise Margin Definitions. [45]

$$NM_H = \frac{3V_{DD} + 5|V_{tp}| - 3V_{tn}}{8} \quad (A - 2.1.2.2 - 12a)$$

$$NM_L = \frac{3V_{DD} - 3|V_{tp}| + 5V_{tn}}{8} \quad (A - 2.1.2.2 - 12b)$$

The switching speed of a CMOS gate as Weste and Eshraghian [45] point out, is limited by the time it takes to charge and discharge the load capacitance, C_L . An input signal results in an output transition that will either charge C_L towards V_{DD} or discharge C_L towards V_{SS} (ground). Figure A-2.19a shows one CMOS gate driving another, with C_L representing the capacitive load that comes from the input of the driven gate, the output of the driving gate and the routing between them. Figure A-2.19b shows the input and output waveforms illustrating the meanings of the rise time (t_r), the fall time (t_f), and the delay time (t_d) which are defined below [45]:

- Rise time - the time required for a waveform to rise from 10 percent to 90 percent of its steady state value;
- Fall time - the time required for a waveform to fall from 90 percent to 10 percent of its steady state value.
- Delay time - the time between the 50 percent level on the input transition and the 50 percent level on the output transition; in otherwords, the time that it takes a logic transition to pass from input to output.

Weste and Eshraghian [45] have calculated the fall and rise times (for $V_{tn} = |V_{tp}| \approx 0.2V_{DD}$) as

$$t_f \approx 4 \frac{C_L}{\beta_n V_{DD}} \quad (A - 2.1.2.2 - 13a)$$

Figure A-2.19: Switching Characteristic for a CMOS Inverter: (a) an inverter driving a second inverter; (b) illustration of the rise time (t_r), fall time (t_f) and delay time (t_d). [45]

$$t_r \approx 4 \frac{C_L}{\beta_p V_{DD}} \quad (A - 2.1.2.2 - 13b)$$

Note that since the carrier mobility μ_n for the n-transistor is approximately twice that for the p-transistor (μ_p) [45], then for equally sized transistors, $\beta_n \approx 2\beta_p$ (refer to equation (A-2.1.2-3)). Thus, for equally sized n- and p-transistors [45]

$$t_f = \frac{t_r}{2} \quad (A - 2.1.2.2 - 14)$$

The delay time depends on the output rise and fall times as indicated below [45]

$$t_{dr} = \frac{t_r}{2} \quad (A - 2.1.2.2 - 15a)$$

$$t_{df} = \frac{t_f}{2} \quad (A - 2.1.2.2 - 15b)$$

where t_{dr} and t_{df} indicate the delay for rising and falling transitions, respectively. The average gate delay is therefore given by [45]

$$t_{d(avg)} = \frac{t_{df} + t_{dr}}{2} = \frac{t_r + t_f}{4} \quad (A - 2.1.2.2 - 16)$$

A-2.1.2.3 Power Dissipation of CMOS Circuits

The total power dissipation in CMOS circuits can be obtained by summing two dissipation components - static and dynamic. Static dissipation is due to leakage current while dynamic dissipation is caused by switching transient currents and the charging and discharging of load capacitances [45].

Figure A-2.20 shows the two logic states that can exist for a CMOS inverter under static conditions. When the input is a zero, the n- transistor will be "OFF"

Figure A-2.20: CMOS Logic States Under Static Conditions: (a) $V_{in} = "0"$, n-device is "OFF" and p-device is "ON"; (b) $V_{in} = "1"$, n-device is "ON" and p-device is "OFF". [45]

while the p-transistor will be "ON". The output voltage will thus be equal to V_{DD} (i.e. a logic 1). When the input is a "1", the n-transistor will be "ON" while the p-transistor will be "OFF", resulting in an output voltage of 0 volts (V_{SS}) or a logic zero. Thus, since no current flows into the gate and there is never a current path from V_{DD} to V_{SS} (ground) in any of the input-output combinations, the resultant steady state or quiescent power is ideally zero. However, due to reverse bias leakage current between the diffusion region and the substrate, there is a small static power dissipation, P_s . For an inverter operating at 5 volts, the typical static power dissipation is between 1 and 2 nano-watts [45].

As was indicated in an earlier section, both the n- and p-transistors are "ON" for a short period of time during a transition from either "0" to a "1" or vice versa, thereby resulting in a short current pulse from V_{DD} to V_{SS} . This transient current is generally much smaller than that required to charge and discharge the output capacitive load. Assuming that the rise and fall times of a square wave input to a gate are much smaller than the repetition period t_p of the input, the average dynamic power, P_d , dissipated during switching has been expressed by [45,48] as

$$P_d = \frac{C_L V_{DD}^2}{t_p} = C_L V_{DD}^2 f_p \quad (A - 2.1.2.3 - 1)$$

where $f_p = 1/t_p$. Note that although the average dissipated dynamic power, P_d , is proportional to the switching frequency (and independent of the device parameters), only a fraction of the gates in a circuit change state (and thus dissipate power) on each clock cycle [45,49].

A-2.2 CMOS Processing

In the following sections, a brief description of silicon semiconductor processing will first be presented in order to give a better appreciation of the succeeding discussion of the specific CMOS processing technology used in the fabrication of the

circuits implemented in this thesis. The final section will briefly examine the layout design rules that must be followed to enhance the probability of correct fabrication and operation of the circuits.

A-2.2.1 Silicon Semiconductor Processing

The manufacture of modern ICs starts with pure disk like slices or wafers of single crystal silicon, which have diameters that vary from 75 mm to 150 mm and a thickness that is less than 1 mm (after processing, each wafer is sawed into hundreds or thousands of identical rectangular chips, each containing the circuit realized on it - Figure A-2.21). In order to create the required circuits, impurity atoms (called dopants) must be introduced or diffused into different parts of the silicon wafer, followed by a superimposition of different insulating and conducting materials. The dopants either supply free electrons or holes; those that supply free electrons are called donors while those that supply holes are called acceptors (they accept electrons already in the silicon, thus leaving an excess of holes). If the resulting doped silicon contains a majority of holes (acceptors), it is known as p-type while one that contains a majority of electrons (donors) is known as n-type [45].

In order to obtain the selective diffusion of dopant atoms, silicon dioxide (SiO_2) which acts as a barrier against doping impurities, is first grown on the wafer. Next SiO_2 is removed in the areas where it is desired that dopant atoms pass into the wafer to form n- or p-type silicon. Finally, the exposed silicon is subjected to the desired dopant. The same process used in selectively removing the oxide is also used in patterning or forming the remaining layers. Consequently, the steps involved in the patterning of silicon dioxide will next be examined.

The silicon dioxide patterning process starts by realizing the pattern to be transferred to the surface of the silicon wafer on a mask, a transparent material coated with a thin layer of opaque material. Parts of the opaque material are

Accretion

CHIP

Monolayer

Figure A-2.21: Packaging Sequence. [50]

removed, leaving opaque material on the mask, in the precise pattern required on the silicon surface. Next, the oxide covered silicon is coated with photoresist (PR), an acid resistant coating that is sensitive to light. The patterned mask is used to cover the wafer, which is then exposed to ultraviolet light or low energy x-rays. The radiation is stopped in places where the mask has opaque material on its surface, while breaking down the molecular structure of the resist into smaller molecules in the exposed regions. Thus, the exposed resist can be removed with a suitable solvent, while the unexposed resist remains. The silicon dioxide that was below the now dissolved resist is next etched away with a solution that will not affect the unexposed resist, hence protecting the silicon dioxide under it. Finally, the remaining resist can be removed with a suitable solution, leaving only SiO_2 on the silicon wafer, in the desired pattern. Figure A-2.22 illustrates the steps discussed above. Note that the above discussion assumed positive resist being used. With negative resist, the resist material that would remain after exposure and developing would correspond to the transparent areas [45,50].

A similar sequence of steps as described above is used in the formation of the various layers that make up the devices on the IC, with etching down to the required level. It should be noted that amorphous silicon (silicon not having a carefully arranged lattice structure) which is commonly known as polysilicon, is usually used as both the gate electrode and as interconnects in silicon ICs. A more detailed examination of silicon semiconductor processing can be found in [45] and [50], as this thesis will now concentrate on describing the CMOS technology (the CMOS1B process) used in the fabrication of the circuits implemented here.

A-2.2.2 CMOS1B 5 μ Process

The CMOS1B process is a p-well CMOS process that has a 5-micron (5 micrometers) minimum feature size [51]. A p-well process normally starts with a moderately doped n-type substrate or wafer; a p-type well for the n-channel transistors is next

Figure A-2.22: Patterning the Silicon Dioxide: (a) Bare wafer; (b) Oxidation; (c) Coat with photo resist; (d) Mask and expose; (e) Exposed resist; (f) Develop resist; (g) Etch; (h) Remove resist. [50]

created, while the p-channel transistors are built in the original n-substrate. In the following discussion, the different mask levels which reflect the different processing steps required in a CMOS1B process will be described.

Although a simple inverter will be used in illustrating these processing steps, it should be noted that the same steps apply to other CMOS devices as well.

The process descriptions that follow are paraphrased from [51]:

- **p-well mask:** This first mask defines the p-well, which will contain the n-channel transistors (Figure A-2.23(a)). Following the creation of the p-well, a thin layer of silicon dioxide (approximately 0.1 microns thick) is grown, over which a layer of silicon nitride (Si_3N_4) is formed (Figure A-2.23(b)). These dielectric layers are used to mask the active regions from p- and n-type implantations into the non-active regions (the active regions contain the source and drain regions and the transistor channels). The SiO_2 layer also acts as a stress relief between the silicon substrate and the Si_3N_4 layer.
- **Device well mask:** The device well mask leaves the dielectric sandwich (Si_3N_4 and SiO_2) only over the areas that will become drains, sources, channel regions and diffusion interconnects (active region) in the p-wells or in the n substrate (Figure A-2.23(c)). Other names for this mask are [45]: thinox, active area, island and mesa.
- **p-guard mask:** This mask is used to accomplish a p-implant into p-well regions that will eventually be covered by field oxide (field oxide refers to the SiO_2 that covers the field or silicon area between transistors - i.e., the non-active regions). Note that the nitride layer masks the active channels from receiving the implant. The result of this implant is to set the threshold voltage for the silicon under the field oxide to be greater than the voltage for the silicon under the gate oxide. This step allows the silicon surface to be inverted under the

Figure A-2.23: CMOS1B Processing Steps for Simple Inverter: (a) p-well definition; (b) Formation of Si_3N_4 layer; (c) Active area definition. [51]

gate but not in the regions between transistors, thereby providing isolation for the devices. The above discussion is illustrated in Figure A-2.23(d).

- **n-guard mask:** This mask defines the regions in the n-substrate that are to receive an n-implant to set the threshold for the non-active regions in the n-substrate. As in the p-guard case, the nitride layer masks the active channels (Figure A-2.23(e)). Following the above step, the field oxide is grown in regions not covered by the nitride - S_iO_2 sandwich, resulting in Figure A-2.23(f).
- **Polysilicon mask:** This is used to define the gate regions for both p- and n-devices as well as poly interconnections on the field oxide (Figure A-2.23(g)). The whole surface is initially covered with polysilicon, then etched such that the polysilicon is left only in the required areas. Note that a low-dose implant of boron is carried out (prior to the deposition of the gate polysilicon) in order to set the p-channel transistor threshold voltages.
- **n^+ mask:** The n^+ mask is used to indicate those areas to be doped n^+ ; namely, the source and drains of the n-channel devices in the p-well and the n^+ regions (substrate contacts) in the n-substrate (Figure A-2.23(h)). Note that the use of polysilicon for the gate regions results in properly aligned source, drain and gate structures.
- **p^+ mask:** As in the n^+ case, the p^+ mask indicates those regions to be doped p^+ ; that is, the source and drains of the p-channel devices in the n-substrate and the p^+ regions (substrate contacts) in the p-well (Figure A-2.23(h)).
- **Contact mask:** A thick layer of S_iO_2 is now deposited over the entire surface; the contact mask is subsequently used to etch out the S_iO_2 in the regions where contact is desired, thereby providing windows by which the different levels can make contact (by the next step).

Figure A-2.23 continued: (d) p-implant into p-well regions; (e) n-implant into n-well regions; (f) Field oxide growth over regions not covered by the nitride - SiO_2 sandwich

Figure A-2.23 continued: (g) Deposition of gate polysilicon; (h) n^+ and p^+ area definitions.

- **Metal mask:** A layer of aluminum is next evaporated over the whole surface, following which a metal mask is used to define the metal lines (used for interconnects and contacts).
- **Passivation mask:** This final step is used to protect the silicon surface from contaminants that could modify circuit behaviour in unwanted ways. A layer of pyrox is deposited over the entire surface and the passivation mask is used to etch out openings over the chip bonding pads to allow for wire bonding (this connects the pads on the chip to the leads or pins on the package).

The complete CMOS inverter structure is shown in Figure A-2.24.

A-2.2.3 Layout Design Rules

As was previously noted, following the layout design rules enhances the probability of correct fabrication and operation of the circuits. The design rules for the CMOS1B process can be found in pages 6 to 9 of reference [51].

One other constraint that should be mentioned is the layout area constraint. In order to simplify the process of merging designs into a multiproject chip, the Canadian Microelectronics Corporation developed four standard pad frames, with different available enclosed areas. The submitted design must fit within one of these pad frames, the characteristics of which are outlined in Table A-2.2 and Figure A-2.25.

Size A

Size B

Size C

Size D

Size E: adjustable length

Figure A-2.25: CMOS1B Pad Frames. [51]

Table A-2.2: CMOS1B Pad Frame Sizes and Dimensions. [51]

Size A:	full size (40 pads)	
	maximum outside dimensions (microns)	4510 x 4510
	working area	4040 x 4040
Size B:	half size (32 pads)	
	maximum outside dimensions (microns)	4510 x 2250
	working area	
Size C:	quarter size (24 pads)	
	maximum outside dimensions (microns)	2250 x 2250
	working area	1780 x 1780
Size D:	one-sixth size (18 pads)	
	maximum outside dimensions (microns)	1490 x 2250
	working area	1020 x 1780
Size E:	body cell (2 pads)	
	maximum outside dimensions (microns)	1120 x 300
	working area	650 x 300
	end cap cell (2 pads)	
	maximum outside dimensions (microns)	1120 x 225
	total number of pads	$2(n) + 4$
	total outside dimensions (microns)	$1120 \times [300(n) + 450]$
	total working area	$650 \times 300(n)$
	where n = # body cells	

END

1 | 7 • 0 7 • | 8 9

FIN