

Copyright © 1996, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

**WIDEBAND SPREAD-SPECTRUM DIGITAL
COMMUNICATIONS FOR PORTABLE
APPLICATIONS**

by

Samuel Wei Sheng

Memorandum No. UCB/ERL M96/81

19 December 1996

**WIDEBAND SPREAD-SPECTRUM DIGITAL
COMMUNICATIONS FOR PORTABLE
APPLICATIONS**

Copyright © 1996

by

Samuel Wei Sheng

Memorandum No. UCB/ERL M96/81

19 December 1996

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Abstract

Wideband Spread-Spectrum Digital Communications for Portable Applications

by

Samuel Wei Sheng

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Robert W. Brodersen, Chair

Over the past several years, wireless communications have seen dramatic advances in two distinct areas. On one hand, the demand for portable voiceband services has completely inundated the capacity of existing analog cellular communication systems, resulting in intense research efforts to improve performance and increase capacity through digital technologies. Such systems focus on wide-area narrowband communications, providing services to individual users in a portable fashion. On the other hand, the need for more flexible computer networks has led to the advent of wireless LAN's, such as the Motorola Altair. Such systems focus on local-area wideband communications, providing services to computers but usually consuming far too much power to be easily portable. However, the distinction between these two is rapidly blurring – as greater and greater mobile computing resources are placed in the hands of individuals, wireless technologies capable of providing wide-area, wideband services will clearly be needed.

This thesis addresses the design and implementation of an indoor wireless system to support multimedia communication, with emphasis on a broadband downlink capable of supporting digital video. In particular, the development of integrated analog RF front-end and baseband digital interface circuitry, as well as the system simulations driving the design, have been examined. An ultimate per-user data rate 2 Mbps is the target; to achieve the required capacity, a picocellular system architecture will be employed, using cells on the order of 5m in size.

In examining multiple access strategies for such an application, direct-sequence spread-spectrum, or code-division multiple access, possesses many advantages. For this system, a symbol rate of 1 Mbaud with a chipping rate of 64 Mchip/sec is used; for the spreading code, a Walsh-PN hybrid loosely based on the existing IS-95 digital cellular standard is chosen. Beyond providing multiple access, the ability to resolve multiple arrivals and detect adjacent channels in the digital baseband circuitry also affords many benefits that other systems, such as time-division or frequency-hop, cannot easily provide. Taking advantage of the broadcast mode transmission of the downlink, each cell is keyed to a pseudorandom pilot tone, which tremendously simplifies timing recovery and detection in the mobile

Lastly, the implementation of a high-performance, monolithic RF transceiver has been addressed. It is evident that an all-MOS RF system operating in the 1-2 GHz range is quite feasible, given technologies that are currently available. The benefits of integration are enormous: reduced parasitic effects, greater manufacturability, and minimized power requirements to drive off-chip loads. Likewise, by examining the basic architecture used in the transceiver, the underlying digital nature of the signal can be used to simplify the resulting circuit considerably: homodyne demodulation using passive sampling techniques is one important example of how this can be achieved. By taking advantage of these techniques, factors of 10-20x reduction in the power consumed by the analog RF circuitry in conventional designs can be achieved, at the expense of increased digital process. In some sense, the analog front-end has been reduced to its most basic functionality: amplification, filtering, frequency conversion, and analog-to-digital conversion.

Lastly, by exploiting dedicated parallel and pipelined techniques, a low-power spread-spectrum receiver has been designed in spite of the tremendous chipping rate of the system (64 Mchip/sec). By scaling the supply voltage, and employing multiple supplies into the chip, optimal voltages can be chosen to meet throughput requirements and minimize power consumption in the baseband digital logic. A prototype receiver has been developed; due to technology limitations, data rate is limited to 32 Mchip/sec; total power consumption of the prototype, implemented as a two custom ASIC solution, is 154 mW.



Table of Contents

CHAPTER 1.	Introduction	1
1.1.	Thesis Overview	4
CHAPTER 2.	Modulation, Multiple Access, and How Radio Waves Behave Indoors.	7
2.1.	Cellular Reuse and the Indoor Propagation Channel	7
2.1.1.	Cellular Systems	8
2.1.2.	The Indoor Picocellular Environment	12
2.1.3.	Statistical Characterization and Multipath Propagation.....	13
2.1.4.	Channel Modeling	17
2.2.	Modulation Techniques	20
2.2.1.	Quadrature Amplitude Modulation	23
2.2.2.	Continuous-Phase Modulation	25
2.2.3.	Modulation Schemes: Comparison and Analysis	27
2.3.	Multiple Access Strategies.....	30
2.3.1.	Time-Division Multiple Access	30
2.3.2.	Frequency-Hopped Spread-Spectrum Multiple Access	33
2.3.3.	Code-Division Multiple Access	34
2.4.	Toward an Indoor Transmission Scheme.....	40
CHAPTER 3.	System Overview: The Broadband CDMA Downlink	41
3.1.	Physical Layer Parameters	41
3.1.1.	Modulation	41
3.1.2.	Pulse Shaping and System Transmit Bandwidth.....	43
3.1.3.	Transmit Power.....	45
3.1.4.	Carrier Frequency	48
3.2.	Multiple-Access Parameters	49
3.2.1.	Processing Gain and Walsh Spreading Codes	49
3.2.2.	PN Overlay Code, the Pilot Tone, and Cell-to-Cell Code Reuse.....	51
3.2.3.	Channel Estimation and RAKE Reception.....	55
3.3.	An Open-Ended Architecture	61
3.4.	Appendix: Ptolemy Simulation	62
CHAPTER 4.	Transmit Architecture and The Baseband Modulator Chip.....	63
4.1.	The Baseband Transmit Modulator	64

4.1.1.	DQPSK coder	64
4.1.2.	Walsh Modulator	65
4.1.3.	PN Modulator	66
4.1.4.	Per-User Power Control	71
4.1.5.	The Pulse Shaping Filter	72
4.1.6.	Filter Implementation	75
4.2.	Results	82
4.3.	Digital-to-Analog Conversion and Frequency Translation	85
4.3.1.	RF Rejection Method.....	85
4.3.2.	Down-Up Method	86
4.3.3.	Digital Hilbert Transform.....	86
4.1.	Epilogue: The Infopad Base Station.....	88
4.2.	Appendix: Modulator Electrical Interface	92
4.2.1.	Pinout	92
4.2.2.	Microprocessor timing	97
4.2.3.	Clock timing	100
4.2.4.	Description of registers.....	101
CHAPTER 5.	The Receiver:	
	Analog RF Front-End	109
5.1.	Receiver Architectures.....	110
5.2.	Silicon CMOS for RF Analog.....	113
5.3.	Low-Noise Amplification.....	115
5.3.1.	System Noise Requirements	116
5.3.2.	RF LNA Design and Matching Filter Networks	117
5.3.3.	Broadband-Matched Topology	121
5.3.4.	Tuned-Matching Topology.....	126
5.4.	Sampling Demodulation.....	132
5.4.1.	Quadrature Recovery and the Choice of System Carrier Frequency.....	136
5.4.2.	Bandpass Antialiasing	139
5.4.3.	Frequency Stability and Incoherent Demodulation	140
5.5.	MOS Sampling Demodulator Design.....	143
5.5.1.	The Bottomplate Sampling Demodulator	146
5.5.2.	On-Chip Clock Buffering and Layout Issues.....	151
5.6.	From RF to Baseband.....	153
5.7.	Appendix: RF Chip-on-Board Design	156
5.7.1.	Electromagnetic Effects	156
5.7.2.	General Layout Issues.....	158
5.7.3.	Discrete Inductors and Capacitors.....	159
5.7.4.	Chip-On-Board Design.....	162
5.7.5.	RF Testboard	163

CHAPTER 6.	The Receiver:	
	Baseband Analog Processing.....	169
6.1.	System Dynamic Range and Linearity Requirements.....	170
6.2.	Variable-Gain Amplifier Topology	172
6.3.	Operational Amplifier Design.....	178
6.3.1.	Optimizations	182
6.4.	The Analog-to-Digital Converter.....	186
6.4.1.	Comparator Design.....	189
6.4.2.	A/ D Optimizations	197
6.5.	Results	201
6.6.	Appendix: Electrical Interface Specifications	212
CHAPTER 7.	The Receiver:	
	Baseband Spread-Spectrum DSP.....	221
7.1.	Receiver DSP Architecture.....	222
7.2.	The Matched-Filter Correlator	223
7.3.	Timing Recovery and Adjacent-Cell Detection.....	226
7.3.1.	Coarse Acquisition	226
7.3.2.	Fine Timing Acquisition and the Digital Phase-Locked Loop	228
7.3.3.	Carrier-Offset Compensation.....	237
7.3.4.	Adjacent-Cell Detection.....	241
7.4.	Channel Estimation.....	245
7.5.	Data Recovery	246
7.6.	Results	250
7.7.	Appendix: Receiver DSP Electrical Interface.....	253
7.7.1.	Input/Output Description	253
7.7.2.	Programmable Registers	255
7.7.3.	Observation Modes	256
7.7.4.	Test Modes	256
CHAPTER 8.	Conclusions and Future Work.....	259

List of Figures

CHAPTER 1.	1
Figure 1.1: Infopad System Overview	2
CHAPTER 2.	7
Figure 2.1: Cellular Transmission System	9
Figure 2.2: Typical Cellular Reuse Patterns	10
Figure 2.3: Normalized BW of Cellular System.....	12
Figure 2.4: Measured Path Loss Characteristic	15
Figure 2.5: Two simulated random sample channels	19
Figure 2.6: Simulated QPSK eye diagram, no channel ISI	21
Figure 2.7: Simulated QPSK eye diagram, with multipath channel.....	22
Figure 2.8: QAM Constellations (Cross-configuration).....	24
Figure 2.9: General CPM Modulator.....	26
Figure 2.10: Comparison of Modulation Schemes (AWGN only).....	28
Figure 2.11: Required SNR vs. Constellation Size.....	29
Figure 2.12: TDMA Framing/Control Strategy (GSM)	31
Figure 2.13: Comparison of TDMA and FH/SSMA.....	54
Figure 2.14: RAKE Receiver architecture (3rd order)	37
CHAPTER 3.	41
Table 3.1: System link parameters	62
Figure 3.1: DQPSK phase transition diagram	43
Figure 3.2: Raised-cosine pulse shape eye diagrams	46
Figure 3.3: Transmit filter complexity versus excess bandwidth.	47
Figure 3.4: Spectral usage, 500-1250 MHz	48
Figure 3.5: Sample Length-64 Walsh Functions.....	50
Figure 3.6: Dataflow diagram of the spread-spectrum signal synthesis.....	51
Figure 3.7: Realizable 32768-length PN sequences	52
Figure 3.8: Sample multipath channel for channel estimation.....	58
Figure 3.9: Multipath estimator output.....	59
Figure 3.10: Plot of simulated BER versus number of in-cell users.....	60
CHAPTER 4.	63
Figure 4.1: Base station transmitter architecture.....	64
Figure 4.2: DQPSK coder schematic	65
Figure 4.3: Walsh sequence generator (length 64)	66

Figure 4.4:	PN Sequence Generators (a) True 15-tap, length 32767 MLSR schematic (b) 16-tap, length 32768 subsequence schematic (c) 15-tap, length 32768 schematic	68
Figure 4.5:	Autocorrelation performance curves	69
Figure 4.6:	Autocorrelation performance under restriction to length 1024 subcorrelations	70
Figure 4.7:	Plot of SNR due to off-peak nonzero autocorrelation, as a function of subcorrelation length.....	71
Figure 4.8:	Filter response after truncation and tap quantization.....	73
Figure 4.9:	Filter quantization noise effects	74
Figure 4.10:	4-Parallel Implementation of the Transmit Filter (a) Parallel architecture, (b) System-level schematic of filter structure, (c) Tap assignments	96-97
Figure 4.11:	Symmetric Filter Minimization (a) Original transversal structure (b) Shared multipliers (c) Pipelined output adder	78
Figure 4.12:	Pipelined output adder filter structure.....	79
Figure 4.14:	Final minimized filter architecture	80
Figure 4.13:	(a) Premodulation to 64 MHz (b) Folding the multiplication into the 4 parallel filters	81
Figure 4.15:	Modulator Die Photo plus Performance Summary.....	83
Figure 4.16:	Measured modulator output	104
Figure 4.17:	SSB Upconversion Strategies. (a) Analog Hilbert Transform; (b) RF rejection filter (c) Down-Up Method; (d) Digital Hilbert Transform	107-108
Figure 4.18:	Multi-modulator combining schemes for the base station	91
Figure 4.19:	Modulator package pinout (209-pin PGA)	92
Table 4.1:	Modulator pinout arranged by functional group	113
Table 4.2:	Modulator pinout by pin ordering	115
Figure 4.20:	Physical circuitry of the microprocessor interface	98
Figure 4.21:	Timing of a microprocessor read cycle.....	99
Figure 4.22:	Timing of a microprocessor write cycle	99
Figure 4.23:	Timing of signals relative to the 64 MHz clock (ϕ_1)	101
Table 4,3:	Description of Registers	122
CHAPTER 5.	109
Figure 5.1:	Off-the-shelf analog front-end implementation (a) Schematic (b) Semicustom board implementation	111
Figure 5.2:	Monolithic RF Front-End Architecture.....	113
Figure 5.3:	Achievable device ft vs. MOSFET drawn L.....	115

Figure 5.4:	System noise figure versus constellation size, for constant BER= 10^{-4}	117
Figure 5.5:	Generalized RF amplifier topology.....	118
Figure 5.6:	Power matching to achieve gain from passive elements. (a) Prototype case (b) Power matched case	119
Figure 5.7:	Broadband-matched LNA topology. (a) Schematic (b) Measured response.....	122
Figure 5.8:	Tuned-matching LNA topology	126
Figure 5.9:	Tuned L-section, 50 ohm match	127
Figure 5.10:	SPICE-simulated input and output reflection coefficients	129
Figure 5.11:	Measured input and output reflection coefficients.....	131
Figure 5.12:	Measured and simulated amplifier S21, noise figure	132
Figure 5.13:	Sampling Demodulation.....	134
Figure 5.14:	Time-domain view of sampling demodulation	138
Figure 5.15:	RF ceramic filter attenuation and return loss characteristics	160
Figure 5.16:	Translation of phase noise into additive signal noise	142
Figure 5.17:	MOS Sample-and-Hold Characteristics	144
Figure 5.18:	Bottomplate sampling topology	146
Figure 5.19:	Variable-gain, bottomplate sampling demodulator topology	147
Figure 5.20:	Sampling demodulator, track-mode configuration.	148
Figure 5.21:	Track-mode linearized model of the sampling array.....	148
Figure 5.22:	On-chip capacitor structure	149
Figure 5.23:	Simulated distortion characteristics of the sampling demodulator	170
Figure 5.24:	Schematic and layout of differential input clock buffer	152
Figure 5.25:	On-chip nonoverlapping multiphase clock generator	153
Figure 5.26:	Standalone sampling demodulator testchip	154
Figure 5.27:	Board transmission line structure	156
Figure 5.28:	Available board fabrication materials	157
Figure 5.29:	Parasitics in surface-mount discrete capacitors, inductors	159
Figure 5.30:	Table of commercially-available surface-mount inductors.....	160
Figure 5.31:	Impact of inductive parasitics on discrete capacitors at RF.....	161
Figure 5.32:	Board-level bypass strategy	162
Figure 5.33:	Board fabrication parameters	164
Figure 5.34:	RF testboard layout.....	165
Figure 5.35:	Interleaved via layout geometry of the chip-on-board bonding lands.	166
Figure 5.36:	Fabricated board photograph	167

CHAPTER 6.	169
Figure 6.1: BER vs. Number of in-cell users, as a function of A/D bit quantization.....	191
Figure 6.2: Bottomplate sampling topology	172
Figure 6.3: Coupling of digital ringing into differential signal paths.	172
Figure 6.4: Non-overlapping clock phases used to control sample & hold.....	174
Figure 6.5: Two-port representation of the operational transconductance amplifier	175
Figure 6.6: Amplifier in hold mode (evaluation phase).....	175
Figure 6.7: Sample-and-hold topology for sampling onto C_1 as well as C_S	177
Figure 6.8: Final sample-and-hold topology.....	178
Figure 6.9: 2 bit-2 bit pipeline with small interstage gain error	179
Figure 6.10: AGC compensation of gain errors.	179
Figure 6.11: Telescopic cascode amplifier. All device sizes are in microns.	180
Figure 6.12: Bias generation circuitry.....	181
Figure 6.13: SHA switch sizings.....	184
Figure 6.14: VGA stage kickback transient	185
Figure 6.15: Modified 1 bit- 3 bit pipeline A/D architecture.....	187
Figure 6.16: Functional illustration of the 1b-3b pipeline converter.....	188
Figure 6.17: High-speed, low-offset comparator	190
Figure 6.18: Capacitor-based input-sampling network.....	190
Figure 6.19: Differential input sampling network. (a) Standard topology; (b) Modified topology employed by the 1b-3b pipeline A/D.....	192
Figure 6.20: Final high-speed differential comparator	193
Figure 6.21: Non-overlap clock phasing for the comparator.	194
Figure 6.22: Comparator during reset phase of operation ($\phi_1 = 0, \phi_2 = 1$).....	194
Figure 6.23: Comparator during non-overlap period ($\phi_1 = 0, \phi_2 = 0$)	195
Figure 6.24: Comparator at the beginning of the evaluation phase ($\phi_1 = 1, \phi_2 = 0$).....	196
Figure 6.25: Effect of common-mode shift on the modified input stage.....	197
Figure 6.26: Comparator device sizing	199
Figure 6.27: Single comparator transient analysis.....	201
Figure 6.28: 3-bit flash comparator layout	202
Figure 6.29: Comparator layout.....	203
Figure 6.30: 3-bit flash A/D converter testchip	204
Figure 6.31: Sampling demodulator testchip.....	205
Figure 6.32: Full analog front-end receiver testchip.....	206
Figure 6.33: Measured results for the analog front-end receiver testchip.....	207
Figure 6.34: FFT of baseband sampling demodulator output	208
Figure 6.35: 700 MHz received eye diagram.	210

Figure 6.36:	700 MHz received data, sliding correlation test.....	211
Figure 6.37:	Analog front-end receiver pinout	213
Figure 6.38:	Sampling demodulator testchip pinout.	217
CHAPTER 7.		221
Figure 7.1:	Receiver DSP Block Diagram.....	222
Figure 7.2:	Basic 64 MHz, length-64 correlator datapath.....	224
Figure 7.3:	Comparison of the power dissipation of two's-complement vs. sign-magnitude representations	226
Figure 7.4:	Block diagram from coarse lock acquisition.....	227
Figure 7.5:	Block diagram of a multipath correlator.	228
Figure 7.6:	(a) Correlator outputs as a function of time offset (b) Early/Late phase sensing in the delay-locked loop.....	229
Figure 7.7:	Mapping of analog front-end outputs	230
Figure 7.8:	Block diagram of DLL phase detector	232
Figure 7.9:	DLL VCO (data multiplexer) block diagram	233
Figure 7.10:	Variable-phase 64 MHz clock	234
Figure 7.11:	(a) 64 MHz variable-phase clock synthesizer. (b) State machine to control clock phase. (c) 4-phase clock generator (derives the four required clocks from CLKX) (d) Timing diagrams	236
Figure 7.12:	Block diagram of the input multiplexer.	238
Figure 7.13:	1024-long correlation output with 25 kHz carrier offset.	240
Figure 7.14:	Compensated magnitude correlator output.	241
Figure 7.15:	Comparison of the L_1 norm against the L_2 norm.....	242
Figure 7.16:	Block diagram of the adjacent-cell detector	243
Figure 7.17:	Block diagram of multipath estimator.	246
Figure 7.18:	Modified channel estimator output.....	247
Figure 7.19:	Modified RAKE receiver	248
Figure 7.20:	Conceptual overview of the DQPSK slicer.	249
Figure 7.21:	Die photo of the receiver DSP prototype testchip.	250
Figure 7.22:	Block diagram of the receiver prototype	251
Table 7.1:	Pinout of demodulator chip listed in order of pin number.....	273
Table 7.2:	Pinout for demodulator chip grouped by signal name.....	274
Figure 7.23:	Block diagram of two level registers along with timing diagram.	256
Table 7.3:	Describes the demodulator's four writable registers	276
Figure 7.24:	Receiver DSP output signals	257
Table 7.4:	Description of the operating modes for the demodulator	277

Acknowledgments

Otherwise known as the only place in the thesis where nontechnical information can be found. And no, I am not going to begin this with a pithy quotation, although something like “Exeunt Omnes -Shakespeare” would probably suffice, or “Exit: Stage Left - Snagglepuss”. When this project was started seven years ago, in some sense its definition and scope was an act of sheer naivete: Mt. Everest doesn't look that big from about 300 miles away. However, in spite of all of the twists and turns, this point has finally been reached: I can finally say that I've more or less done what I set out to do. If nothing else, I've learned more than I can possibly imagine, that's for sure. And maybe it took about 2 years longer than I thought it would, but at least we're here.

Perhaps it's obvious, but I'll say it anyhow: it would be impossible for any single human being to finish a project of this magnitude without some serious help by some seriously talented people. And lots of help I did get. Friends, mentors, and family: all of them played a very large role in making this happen, the impact of which isn't always apparent when reading the next couple of hundred pages or so. Thus, before any technobabble and cryptic three-letter acronyms are thrown out, I'm going to take a brief moment to thank all those who have been such a wonderful part of my life (and also to get a few final digs in for all the ribbing I've gotten over the years). So, without any further adieu:

Justin, Henry, Tetiana, Eric, Elaine, and Rob: we survived undergrad at Berkeley, we survived grad at Berkeley. I guess we'll survive just about anything after this. Allyn, Ryan, Chris, and Laura – my best friends from those halcyon days of childhood: may we stay friends for another twenty years – and twenty more after that. All I can say is, “Yes, I'm finally out. So there.” Drs. Wu, Wu, Lo, and Wokurka (otherwise known as Joyce, Allan, Eliza, and Laura): I think watching all of you has finally convinced me that grad school is a lot easier than med school – congrats on all of the M.D.'s! Jenny Wu: for simply being who she is. May you never stop smiling, lest the world be a much unhappier place to live. Andy Burstein, one of the best cubiclemates one could hope for. Grad school would not have been the same without his biting wit (as well as his unfailing support through cooking classes and with ext2spice). Tom Burd,

the second of the best cubiclemates one could hope for. I guess I did lose that bet, didn't I? Oh, and by the way, the position of tools manager is still open. Scarlett Wu: for having such a pragmatic view of the world. And for reminding me that there are more important things in life, like taking care of white teddy bears. Jim and Amy Tso: winners in the best roommates' category. May you two have many happy years together, and hopefully I won't keep Jim up any more playing video games. Peggy Huang – just remember: “Change is good.” I won't forget, so long as you don't. Edward “T” Chang: major thanks for all of the years. And I promise that we *will* end up trekking across Europe someday: count on it. And Jennie Chen: for all of the time, support, and care she's given me (not to mention being the third of the best cubiclemates one could hope for).

Kevin Zimmerman, sysadmin extraordinaire: for many, many interesting conversations, for unflinching rising above and beyond the call of duty, and for saving my tail more times than I can count. Peggie Brown and Tom Boot: if I could give each of them a penny for every time I troubled them for something, they'd probably be bazillionaires by now. Brian Richards: for all of the times I've troubled him with silly questions about board design, tools, layout, cifplot generation, and few million other things.

To the members, past and present, of “BJGroup” and “PRGgroup”: thanks for all of the support and encouragement over the years, and for the many discussions (technical and otherwise) I've had with all of you – if I had to choose, I'd have to say that they truly were the best part of being in graduate school. Greg Uehara and Cormac Conroy: both friends and mentors. You guys really taught me the practical ins and outs of analog design; I'm looking forward to working with you both again. Randy Allmon: the contributions he made during his short stay as an industrial visitor were astounding. Kevin Stone and Jim Peroulas: the digital duo. Thanks for all of the hard work on this project, and I can honestly say I hope we never have to pull another all-nighter getting a chip out again. Lastly, to the one person who made the most difference: Lapoe Lynn. All I can say is, I owe you one for saving my bacon. Big time. Without his dedication, discipline, and sheer ability, this project never would have been completed; of that, I am absolutely certain.

To Professor Paul Gray: a true gentleman and scholar. Thank you for all of the time spent fielding questions, teaching about analog design, and for the patience and interest you've shown in this project over the years.

Finally, special thanks must be given to the Fannie and John Hertz Foundation, for their generous financial support on this project.

See? I said there were a lot of people. However, there are a two very special dedications that need to be made. I'm really not certain that this will do them justice, but here goes:

To my family: Mom, Dad, Amy, and Susie. Words cannot begin to describe how thankful I am to be able to call the four of you "family." I love you all dearly.

And last, but hardly least, to Professor Robert Brodersen: for the vision, for his many years of guidance and advice, and for challenging me to do this project in the first place. What he has taught me goes far, far beyond the merely technical.

Samuel Sheng

28 July 1996

PS: And now for something completely different. The following is one of the all-time great grad student survival necessities, and I highly recommend it to anyone embarking upon the quest for a graduate degree [Fields92]:

2.5C all-purpose flour, 0.5 tsp baking soda, 0.25 tsp salt, 1c firmly packed dark brown sugar, 0.5c white sugar, 1c softened unsalted butter, 2 large eggs, 2 tsp vanilla extract, 2C semisweet chocolate chips. Preheat oven to 300° F; in a medium bowl, combine flour, soda, and salt, mixing well with a wire whisk. Set aside. In a large bowl, with an electric mixer blend sugars at medium speed; add butter to form a grainy paste, scraping down the sides of the bowl. Add eggs and vanilla extract, and mix at medium speed until light and fluffy. Add the flour mixture and chocolate chips, blend and low speed until just mixed. Do not overmix. Drop by rounded tablespoons onto an ungreased cookie sheet, 2 inches apart. Bake for 22-24 minutes until golden brown (cookies should still be quite soft). Transfer cookies immediately onto a cool surface with a spatula. Yields 3.5 dozen.

1 Introduction

Over the past several years, wireless communications have seen dramatic advances in two distinct areas. On one hand, the demand for portable voiceband services has resulted in intense research efforts to improve performance and increase capacity through digital transmission. Such systems focus on wide-area narrowband communications, providing low-bandwidth network services to individual users in a portable fashion. On the other hand, the need for more flexible computer networks has led to the advent of wireless LAN's such as the Motorola Altair. Such systems focus on local-area wideband communications, providing networking services to individual computers but usually not easily portable.

However, the distinction between these two systems is rapidly blurring. As laptop computers place mobile computing resources in the hands of individuals, wireless technologies capable of providing wide-area, wideband services will be clearly needed. With this merging of computation and communications, individual users will have instantaneous and portable access to fixed information networks via a lightweight mobile unit. Furthermore, users will be capable of transferring data to other users and accessing fixed computing resources without any constraints on where or when such access takes place. As shown in figure 1.1, the mobile unit will support a myriad of services, including full-motion digital video and high-quality audio, and combines the functionality of today's analog mobile telephones, radio pagers, and laptop personal computers.

Since portability places severe constraints on the physical weight of the terminal, the available battery power is quite limited. Thus, power minimization is crucial; power reduction in both the digital and analog hardware must be achieved. To this end, the terminal should only carry the minimum of computing resources necessary to support its functionality; user computa-

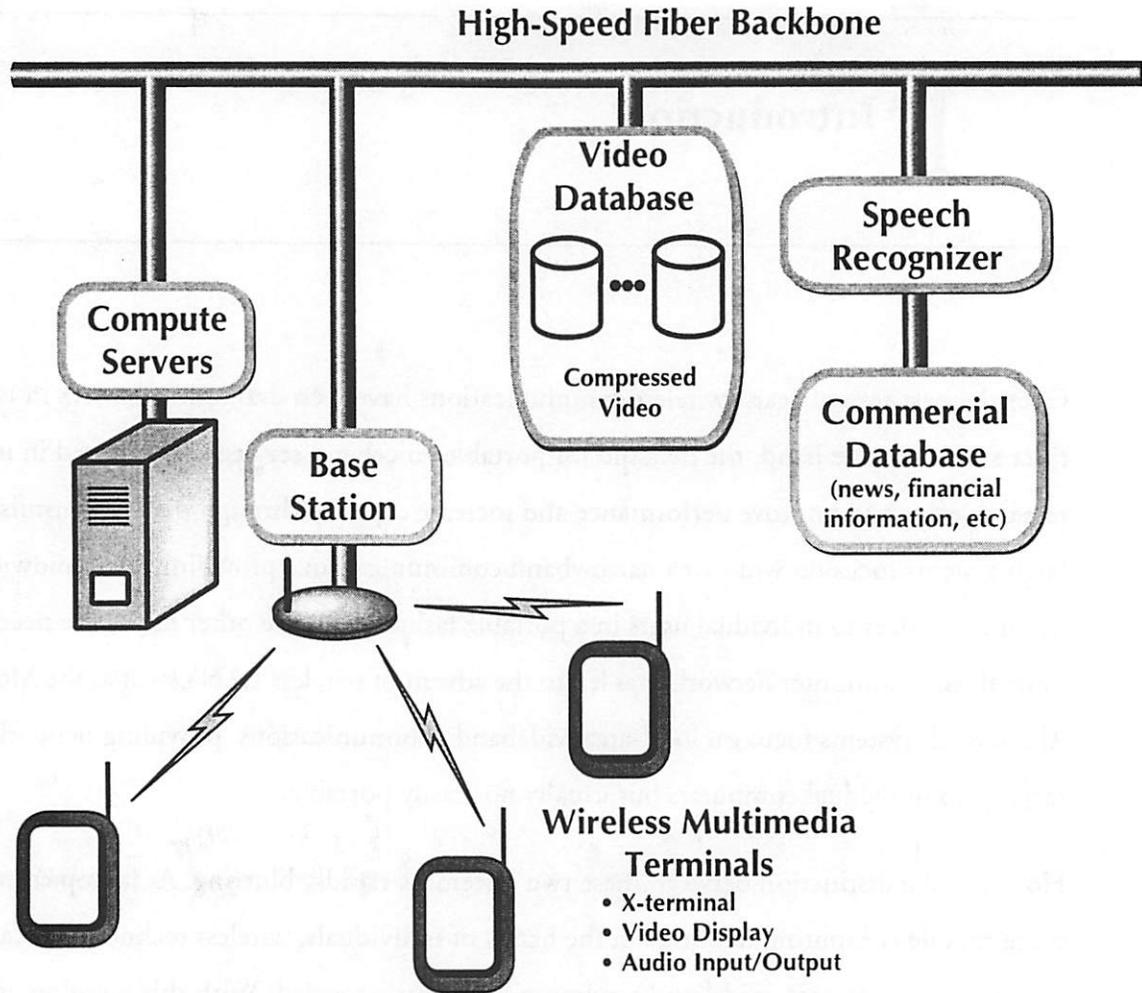


Figure 1.1: Infopad System Overview

tion should be mainly performed by large, non-portable computing facilities, with the high-speed wireless link serving as the terminal's means of accessing the fixed computation servers and data networks. Direct point-to-point wireless communication is not allowed; the link only provides the final interface into the wired data network, much as a conventional telephone handset serves as the link into the telephony system. Whereas the capability of moving massive amounts of digital data within networks already exists, the problem of easily getting data in and out of those networks is now addressed.

Although placing all computation services back in the wired network has immediate benefits in terms of reducing power consumption, it provides another advantage: data that is highly sensitive to corruption will not be transmitted over the wireless network. Existing distributed

computation environments are crucially dependent on the fact that data transmitted over the network has high integrity - i.e., bit-error rates on wired Ethernet are typically on the order of 1 per 10^{12} bits and further protection is gained by packet retransmit in the case of an error. However, on wireless networks this is not true; even after extensive error-correction coding, it is still difficult to attain error rates even remotely as low as this. User "computation" data, such as spreadsheets or simulation results, simply cannot be allowed to sustain any corruption. For wireless systems, this translates into an inordinate amount of transmission overhead in terms of coding and data retransmission to compensate. On the other hand, user "multimedia" information, such as voice and image data, is relatively tolerant of bit errors - an error in a single video frame or an audio sample will not significantly change the meaning or usefulness of the data. Thus, the portable unit described above is truly a terminal dedicated to multimedia personal communications, and not simply a notebook computer with a wireless LAN/modem attached to it.

With the shift from computation inside the mobile unit to communications outside, evidently the development of a wideband link capable of supporting the required user bandwidth becomes paramount. If the system is to support digital video, a minimum user data rate of 2 Mbps is required after error correction and coding, even utilizing the best data compression available today [MPEG90]. Due to these high bandwidth requirements, carrier frequencies must be placed in or above the low microwave bands, owing to spectrum congestion at frequencies below 1 GHz. Likewise, since one major objective of personal communications lies in granting each person individual access, a large number of users need to be accommodated by the system simultaneously, densely packed into a small physical area; continuous network access by 1 user every 4 to 10 square meters is typical for an office environment. Ostensibly, a microcellular or even picocellular reuse system must be employed to come remotely close to the required capacity [Lee89]; however, even with cellular reuse and a limited number of users per base station, the required aggregate channel data rates will be in the tens to hundreds of megabits per second.

It is extremely important to note that, in this system, broadband video data is only supported in the downlink to the mobile; the uplink is used only for low-rate (32 – 128 kbps/user) speech and control data. Although the system is still full-duplex, this asymmetry must be accounted for in the design, as the bandwidth requirements for maintaining the uplink are considerably less than those for the downlink. Several commercial solutions already exist [Prox94, Symb95] that can satisfy the limited requirements of the uplink; it is the downlink - especially the high-bandwidth, low-power receiver in the mobile - that presents the greatest research challenges.

Thus, the following constraints are imposed upon any transmission scheme that is to be used for such a personal communication system:

- The analog circuitry must perform reliably at variable carrier frequencies above 1 GHz, while supporting a sustained per-user data rate in excess of 2 Mbps.
- The system must support simultaneous access by a large number of users, within a small physical area.
- The mobile transceiver cannot consume excessive amounts of power, as the portable terminal must be powered by batteries, nor can it be excessively large.

Existing commercial solutions fall far short of this - the best wireless LAN's of today offer an aggregate channel data rate of 1 Mbps, with power consumption around 1 watt. The ultimate goal is to achieve a thousandfold improvement in the power-throughput ratio over existing hardware - for the receiver, 1 milliwatt per Mbps, while supporting a raw channel capacity in excess of 100 Mbps. The key contribution of this work has been the development of techniques by which such a thousandfold improvement can be achieved: system-level designs that inherently favor complex digital signal processing, and minimize analog complexity and power consumption.

1.1. Thesis Overview

In this thesis, the system design and implementation of the broadband downlink will be described, discussing issues ranging from low-level analog design, where a novel receiver structure using passband sampling is presented, up to the multiple access schemes utilized and its

impact on the transceiver system. Since such a system will first be utilized as a step beyond conventional wireless LAN's, an indoor microcellular environment will be of primary concern, although many of the techniques are applicable to both the outdoor and indoor environments.

In chapter 2, an overview of the system itself is presented, beginning with a brief description of the indoor radio-frequency (RF) environment and its various impairments, followed by a discussion of multiple access strategies. It will be shown that a design utilizing spread-spectrum, code-division multiple access is ideally suited for such a system, not only as a means to multiplex the various users, but also to relax hardware performance requirements and to ultimately minimize power consumption in the receiver. The actual system specification follows in chapter 3, documenting design decisions made on carrier frequency, channel bandwidth, modulation, multiple access, and spread coding, along with verification by system-level simulation.

In the remainder of the thesis, the actual hardware designed to implement the link will be examined. Chapter 4 examines the base-station transmitter, including a custom integrated circuit that performs user multiplexing, spectrum spreading, and channel pulse-shaping. The analog RF transmitter section is also described, consisting of a semicustom board design to upconvert the signal for transmission. Chapters 5 and 6 describe the custom analog circuit developed for the mobile receiver, consisting of a low-noise amplifier, quadrature homodyne demodulator that utilizes passband sampling to achieve frequency conversion, a multi-stage variable-gain amplifier, and an extremely high-speed analog-to-digital converter. Low power and high integration are emphasized here, in achieving a "single-chip" monolithic CMOS solution for the analog front-end – the tremendous impact of the use of code-division multiple access will be made apparent. Following this, chapter 7 examines the digital aspects of the receiver - the custom spread-spectrum digital demodulator circuit required to perform data and timing recovery, and the low-power techniques employed in its implementation. Lastly, conclusions and a description of future work are presented in chapter 8.

2 Modulation, Multiple Access, and How Radio Waves Behave Indoors.

In the design of any RF communications system, three distinct factors must be considered: the propagation environment in question, the multiplexing and modulation of user data streams, and the complexity of hardware required to implement the desired link. None of these three can be allowed to dominate; for example, implementation issues place constraints on answering “How many users?” and “How much bandwidth?”, whereas the choice of modulation and multiple access can have dramatic impact on the questions of “How many transistors?” or “How much power?”. To being answering these questions, the system design of the high-speed downlink will be discussed in this chapter and the next. First, a description of the indoor radio propagation environment will be presented, followed by a short exposition of the available multiple-access strategies, describing the advantages and disadvantages of each. Second, the actual system specification will be given – in terms of number of users per base-station, transmit bandwidth, modulation strategies – and a discussion of how this specification was arrived at.

It should be mentioned here that verification of the various design tradeoffs was done using the U.C. Berkeley Ptolemy simulator [Buck93]; details of the simulation code are contained in the appendix following this chapter.

2.1. Cellular Reuse and the Indoor Propagation Channel

Since the system must support independent full-motion digital video streams to each user, spectrum usage becomes of great concern, requiring a per-user data rate of 2 Mbps. Also, this data rate is clearly not needed on a continuous basis; when regular computation tasks are being performed, such as using a word processor or a spreadsheet, the screen changes only

slightly on a frame-by-frame basis and over only a small region, usually on the order of a single character or a few pixels. Hence, it is easily possible that the peak data rate required by a user is much larger than the overall time-average data rate, and minimizing overall system bandwidth consumption while supporting a large number of users accessing data simultaneously is of paramount importance. One technique, applied at the system level, is to utilize cellular networking techniques to achieve spatial frequency reuse.

The advantages in improved spectral efficiency afforded by cellular systems have long been known; having been employed extensively in present-day analog mobile radiotelephony, the large-scale cells utilized only exploit these advantages to a limited extent. By scaling down cell sizes, tremendous increases in spectral efficiency can be achieved. In this section, an analysis of these gains is presented, along with a description of the cellular concept as realized in an indoor environment. A statistical characterization of the indoor picocellular transmission environment and the corresponding model are then described.

2.1.1. Cellular Systems

As shown in figure 2.1, a simple cellular scheme consists of dividing the entire service area for the personal communication system into “cells” of radius R , with a single base station serving all mobile users within that cell; each cell utilizes its own distinct set of system resources, different from its surrounding neighbors. As an example of such resource allocation, conventional analog cellular, being a frequency division multiple-access system, allocates different frequency bands to each cell. There is a clear correlation between the method of resource allocation and the multiple access strategy, as we will see later: in time-division multiple access, each cell could be assigned different time slots, and in code-division multiple-access, each cell could be assigned different codes.

As users move from cell to cell, their transactions with the network are “handed off” from base station to base station, reconfiguring the network dynamically as the need arises. Such reconfiguration is thus accomplished locally, since the system itself is responsible for determining where and when handoffs occur. Clearly, the complexity involved with cellular net-

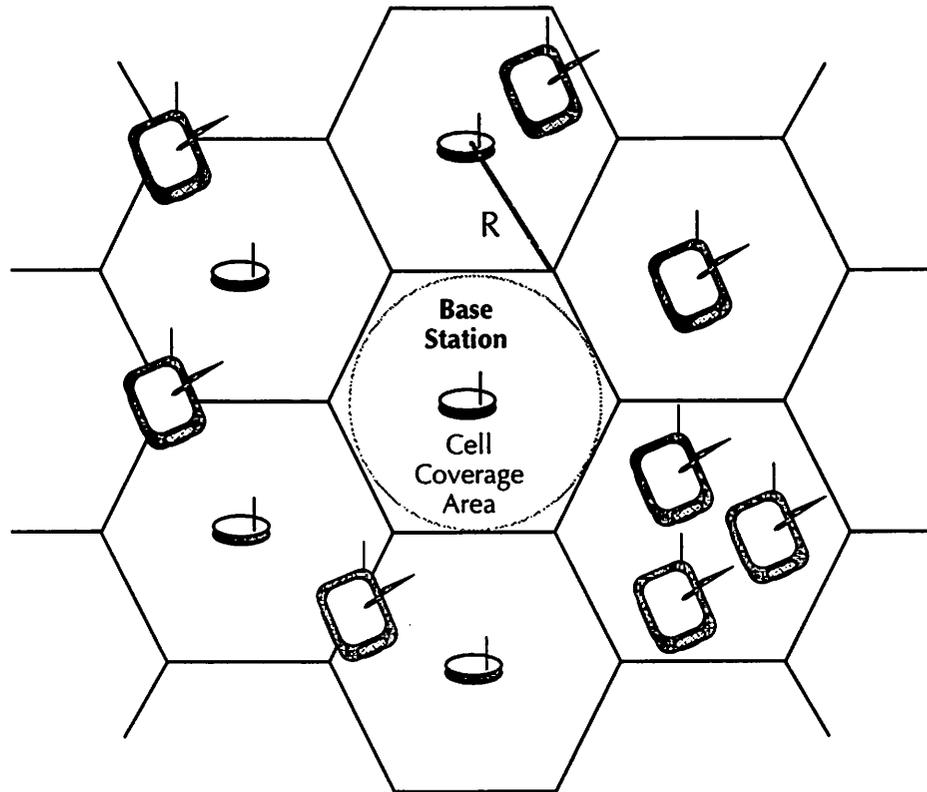


Figure 2.1: Cellular Transmission System

work control is much greater than that required for a classical “umbrella” scheme, with one base antenna for the entire service area, and much of the “intelligence” of the network must be deployed within the base station and the mobile transceivers.

The key advantage to cellular systems is that they allow the network to achieve spatial multiple access of the users. If two cells are separated by sufficient distance, each can use the same channel resources at the same time without resulting in disastrous cochannel interference.

Figure 2.2 shows several classical reuse patterns [Lee89]; such patterns are typically characterized by a reuse factor K , which represents the number of distinct sets of resources (either frequencies, or time slots, or codes) that need to be used to cover the entire service area.

From the point of view of spectrum usage, each user effectively consumes only B/N Hz of bandwidth, where B is the raw physical bandwidth needed to support transmission within a single cell, and N is the total number of users in the system. Hence, cellular systems are said

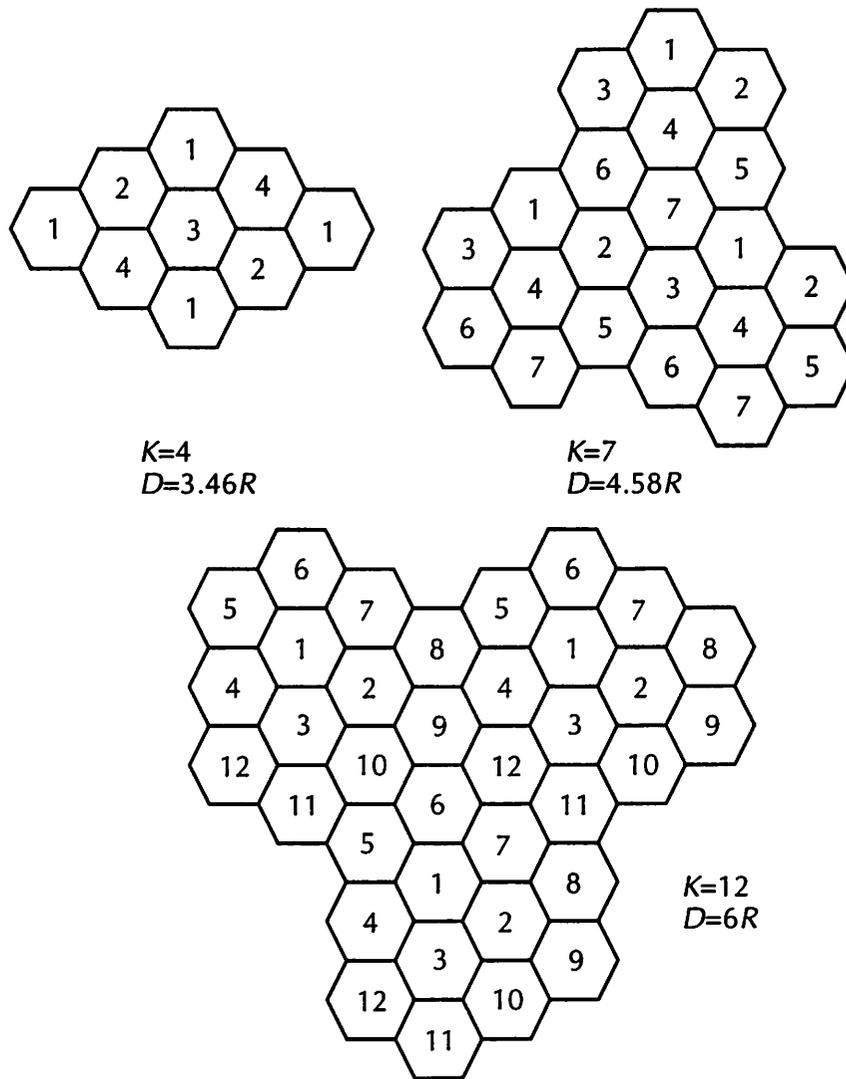


Figure 2.2: Typical Cellular Reuse Patterns

to be spectrally efficient. Clearly, minimizing the distance D between cells utilizing the same frequency yields the greatest frequency reuse, since the number of cells that can use the same frequency band in the service area is maximized, and hence the greatest gain in spectral efficiency is achieved. This frequency reuse distance is geometrically related to K and R by $D=R\sqrt{3k}$. In terms of system capacity, supposing that the total service area serving the N users is given by $\pi\rho^2$, then the number of cells is given by $(\rho/R)^2$ and hence the total bandwidth for the system has been reduced by a factor equal to $(R/\rho)^2K$ relative to an equivalent non-cellular system. As shown in figure 2.3, the normalized bandwidth required by a cellular sys-

tem is plotted as a function of R and K ; since it is quadratically dependent on R , it is of greatest benefit that the cell radius be reduced. Clearly, when $R=\rho$ and $K=1$, the original “umbrella” scheme is yielded.

Ostensibly, the ultimate limits on minimizing R (and hence D) lie in how much cochannel interference the system can tolerate, and the required complexity in network control. An important fact is that the level of cell-cell cochannel interference is independent of the scaling of R , since the transmit power in each cell scales with R and hence the relative interference stays constant. Thus, cochannel interference is only a function of K . Conversely, complexity in network control is only a function of R , since more handoffs will necessarily occur as R decreases. Both K and R are highly dependent on the transmission environment; however, these massive gains in spectral efficiency clearly will be needed to support the high-speed, high-bandwidth requirements that a personal communications system stipulates.

2.1.2. The Indoor Picocellular Environment

Within an indoor environment, it is no longer feasible to have only a single network transceiver station serving all of the terminals in the building. Due to the 5 to 15 dB attenuation through walls [Seidel91], the total microwave output power from all of the transmitters would have to be inordinately (and dangerously) high. However, this attenuation can be taken advantage of by a cellular network. Each room then naturally becomes its own cell; likewise, the cellular scheme now moves into three dimensions, since the floors also provide RF isolation. Since the cells are now extremely small, on the order of a two to five meters, this cellular strategy is commonly referred to as microcellular or picocellular networking. Hence, R is usually dictated by the size of the room, and K can be as low as 3 to 4, depending on how much attenuation is provided by the walls. If K is increased to 6 or 7, the assumption that co-channel interference is negligible becomes reasonable for most indoor office environments.

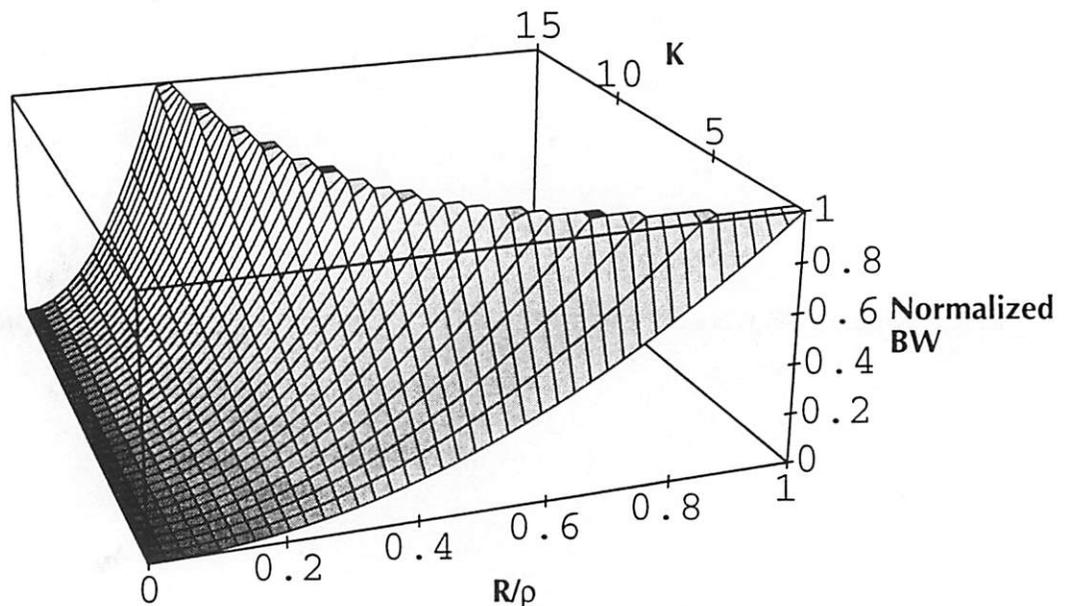


Figure 2.3: Normalized BW of Cellular System

Normalized with respect to the non-cellular case, assuming a constant number of users and data rate.

In light of the above considerations, the total amount of spectrum that will be consumed to provide the outlined services can now be addressed. After examination of the user density in a typical office environment¹, such as those found in modern buildings with open-area soft-partition cubicles, cells of 5 meter radius typically contain twelve to sixteen active users. A 2 Mbps data rate for full motion video (our worst-case requirement), using a linear DQPSK (differential quadrature phase-shift keying) modulation scheme and using design parameters from existing systems [Feher87, Raith91], would require a transmission bandwidth of approximately 1.6 MHz per user using a 30% excess bandwidth raised-cosine pulse shape and a 25% loss for packetization, equalizer training, and other overhead. Assuming that, of the 16 users in the cell, half of them are demanding the complete 2 Mbps data rate for full-motion digital video with the remainder utilizing 256 kbps² each for lower data rate applications, a picocellular system with $K=7$ would utilize approximately 100 MHz of bandwidth. Although 100 MHz is a considerable amount of spectrum, this is amortized over large numbers of people using this spectrum simultaneously within multiple buildings. Considering that the bandwidth of 100 MHz is designed to support full motion video and other multimedia network services for all users, this allocation of spectrum is not unreasonable given the level of service provided by the system, especially when compared to the spectrum allocated for existing systems such as NTSC television.

2.1.3. Statistical Characterization and Multipath Propagation

Unfortunately, the indoor environment also presents several additional transmission difficulties, the dominant one being the numerous reflections of the radio signal off walls, furniture, and even people. A large number of paths exist between transmitter and receiver; the received signal is thus corrupted by severe multipath distortion and inter-symbol interference, since these reflections can still be significant several symbols later. Also, in a picocellular net, it is likely that there is no direct line-of-sight propagation path due to shadowing, so we become critically dependent on the reflected waves. All of these effects, of course, vary slowly with time; even with the remote unit fixed in place during operation, the motion of people can

1. For example, the EECS graduate student research facility at this University.

2. 256 kbps is reflective of data rates used in existing X-terminals; this value may be considerably lower depending on the level of user activity.

cause significant variations in the environment. The critical statistical parameters to be determined are: the number of paths, the excess time delay for a single path, the path loss, and the time variation of the received signal, assuming a cell size of 2 to 10 meters. Using measured data collated from various sources [Seidel89, Seidel91, Ohrvi89], such a characterization of the indoor propagation environment at 1 GHz is determined.

First, the path loss in a multipath situation must be considered. It has been found experimentally [Ohrvi89] that the total received power at a particular distance d of a multipath profile can be modeled as a log-normal (normal in dB) distribution about a local mean path-loss law given by d^{-n} , valid for d approximately larger than λ , where λ is the carrier wavelength at 1 GHz = $c/f = 0.3\text{m}$. Values of n ranging from 1.5 to 6 have been reported, although typical values range from 1.5 to 3. Figure 2.4 shows a measured path loss characteristic [Ohrvi89], with the best-fit curve varying as $d^{-2.6}$. It has also been found that, for a single path component, the received power has a log-normal distribution about an exponential path-loss law $d^{-n(\tau)}$, where τ is the excess delay time defined as $t_{\text{arr}} - t_0$ and t_0 is the delay of the line-of-sight path, with an obstructed topology exhibiting greater attenuation than a line-of-sight one as a result of the extra path distances incurred by shadowing. Generally, the standard deviation of the log-normal distribution is insensitive to τ , and is approximately 4 dB for line-of-sight topologies, and 5 dB for obstructed ones.

The problem of the statistics of the number of paths and their arrival times at the receiver also needs to be considered. Clearly, receiver sensitivity here is critical, since more paths that can be resolved by a receiver with higher sensitivity. For low-sensitivity receivers the number of path has been measured to be approximately Poisson, with a mean of 4.4 paths and standard deviation of 2.1, for a receiver threshold of 30 dB below a 10λ received power reference. As receiver sensitivity increases, a Poisson distribution becomes a poorer model for the number of paths, and a normal distribution with a mean of 22.4 paths and standard deviation of 8.6 paths provides a good statistical model for a 48 dB threshold. The other path-dependent factor is the distribution of arrival times; having determined the behaviour of the number of

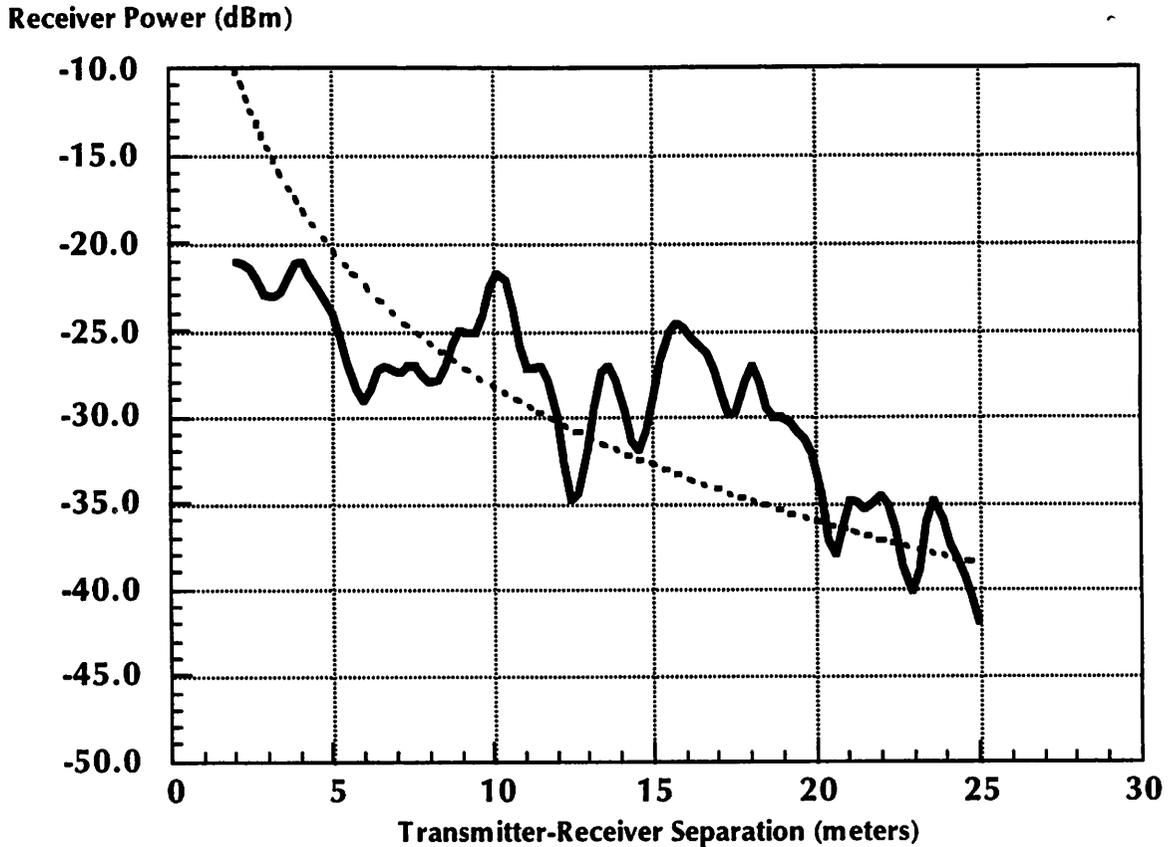


Figure 2.4: Measured Path Loss Characteristic

(from [Ohrvi89])

paths, the probability of a path having an excess arrival time τ needs to be characterized. From several measurements [Seidel89, Saleh87], the excess arrival times follow an exponential distribution, with a mean time of 30 nsec for transmitter-receiver separations below 10 meters.

One important parameter related to the path distribution is the total delay spread, or the total time between the first arrival and the last resolvable arrival due to multipath. The delay spread characterizes the signaling rate below which the fading will be “flat”; in other words, if the time for each transmitted bit is greater than the delay spread, there will be no intersymbol interference, since all of the reflections will have died away before the next bit occurs. In the frequency domain, this quantity is referred to as the coherence bandwidth, which can be intuitively viewed as the interval (in Hz) between consecutive fading nulls in the channel impulse response. The relationship between delay spread T_{del} and coherence bandwidth C_{BW} is that C_{BW} is approximately equal to $1/T_{del}$. For indoor channels, at a sensitivity of 48 below a

10λ threshold, T_{del} is approximately 30 nsec, corresponding to a coherence bandwidth of 33 MHz. As will be shown, it is desirable to either transmit at a data rate far below the coherence bandwidth (hence resulting in no intersymbol interference and flat fading), or at a data rate far above the coherence bandwidth, in the case of spread-spectrum.

Unfortunately, the difficulties presented by the multipath environment are further compounded by the fact that the environment is also time-varying, and hence results in fading, where the environment varies such that destructive interference occurs at the receiver and causes much more attenuation than predicted by path loss alone. Since it is assumed that the remote device is relatively stationary, the typical Rayleigh-distributed fading patterns found in analog mobile environments are not encountered. Instead, the primary cause of time-variation is the movement of people in the propagation path. For only a few people, the distribution has been found to be Rician more than Rayleigh; however, as the number of persons increases, the distribution becomes increasingly Rayleigh. From the measurements taken, a typical fluctuation results in a signal deviation 10% of the time in excess of 4 dB away from the median, and a deviation in excess of 8 dB 1% of the time [Ohrvi89].

Lastly, noise effects within the channel must be considered. From measurements taken of the 1-2 GHz frequency band, the additive white Gaussian noise present on the channel is essentially at the thermal noise floor (-110 dBm at 300°K over 1 MHz); hence, as far as Gaussian noise is concerned, the noise contribution is dominated by the noise at the input of the receiver. However, the low-microwave band also exhibits significant impulsive noise, resulting from RF interference from such sources as computer systems, power microwave cavities in cooking ovens, and common relay switches. It is extremely difficult to characterize these sources, since the presence and characteristics of such generators vary so widely, even within the same building. From some recent measurements, it has been shown that the average noise factor above the thermal noise floor for impulsive sources can be as high as 50dB [Black91].

2.1.4. Channel Modeling

For multipath transmission, an early time-varying, wideband model proposed by Turin is given as:

$$\sum_{k=0}^n a_k e^{jq_k s(\tau - \tau_k)} + n(\tau)$$

where $s(\tau)$ is the complex-valued, lowpass representation of the transmitted pulse, $n(\tau)$ is the noise component, q_k is the carrier phase shift, and a_k are the amplitudes of each arrival at time τ_k in the multipath profile. Since along each propagation path, the physical media is essentially “ideal” (i.e., represented by a simple attenuation constant), the impulse response of such a channel can thus be characterized by an expression of the form:

$$h(\tau) = \sum_{k=0}^n a_k \delta(\tau - \tau_k) e^{jq_k}$$

By using the statistical data described above, a model of this impulse response suitable for simulation can be developed.

The path strength coefficients a_k are set to the local mean value given by d^{-n} , where $n=2.6$, multiplied by the variation that is normally distributed in dB. Also, since of primary interest is a simulation model suitable for verifying the functionality and performance of the transceiver downlink, the time-variation of the channel is sufficiently slow that we can assume it is constant over the simulation time. For our model the path attenuations are thus fixed at the beginning of the simulation and remain invariant. Likewise, the number of paths and individual path delays are treated similarly. At the start of the simulation, the number of paths is randomly determined, which is normally distributed as described in the previous section. For each path, an exponentially distributed random value is assigned, corresponding to the path delay.

If long-term error performance simulations are required, such as large Monte-Carlo estimations of the bit-error rate, the model can be adapted for such purposes, by simply incorporating time variability into the model. The only consideration that needs to be made is that over

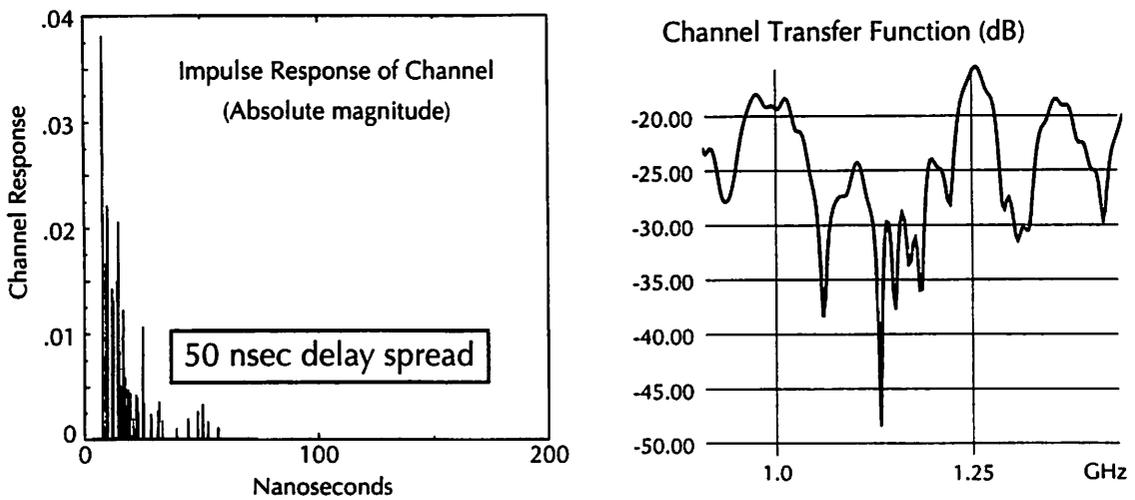
time, the amplitude and phase of each multipath component must be continuously related to the original randomly generated profile – the profile cannot be “randomly” updated. Statistically speaking, the resulting fading behaves as a Rician process; by folding in a time-series model, time variability can be accounted for.

Two Ptolemy simulation blocks, or “stars” have been developed to generate statistical multipath models that correspond to the above; one of these stars, `MPathImpulse.s`, models the channel as a real-valued FIR filter, whose impulse response is simply given by $\{a_k, k=0..n\}$. This star is useful for testing analog modulator and demodulator structures, since it gives the true passband response of the channel. The other star, `Multipath.s`, gives an equivalent baseband representation of the channel in terms of real and complex components. This star is useful for examining the baseband digital modulation and demodulation strategies; essentially, it encapsulates the analog passband modulation, channel effects, and passband demodulation into a single model. The key difference lies in the number of data points that the system must process in order to simulate a single transmitted symbol; in the case of `mPathImpulse.s`, the passband modulation is modeled, requiring an extremely fine timestep to capture the gigahertz band carrier, whereas with `multipath.s` the timestep needs to only capture the baseband phenomena, which is bandlimited to several megahertz.

To give an idea of how severe the fading environment can be, the impulse responses (both in time and frequency) of two such randomly generated channels are shown in figure 2.5; they are quite similar to impulse responses measured in the field. The magnitude frequency response of the channel, $H(j\omega)$, is plotted in a 300 MHz bandwidth in the neighborhood of 1 GHz. In comparing the two, the dynamic range due to indoor fading can easily exceed 30 dB, consistent with measured results. Unless such nulls are combatted, signal degradation and possibly complete link breakdown will result, since the signal-to-noise ratio has been compromised by the null.

As an important application of this model, the effects of the multipath on an actual transmitted signal will be simulated, by generating an eye diagram of the received signal. Using a simple 4-PSK QAM modulation scheme with white input data, the effect of the multipath distortion is examined for a data throughput of 1 Mbps at a carrier of 1 GHz. As described

Simulated channel, in-phase primary arrivals at 1 GHz (strong peak)



Simulated channel, out-of-phase primary arrivals at 1 GHz (strong null)

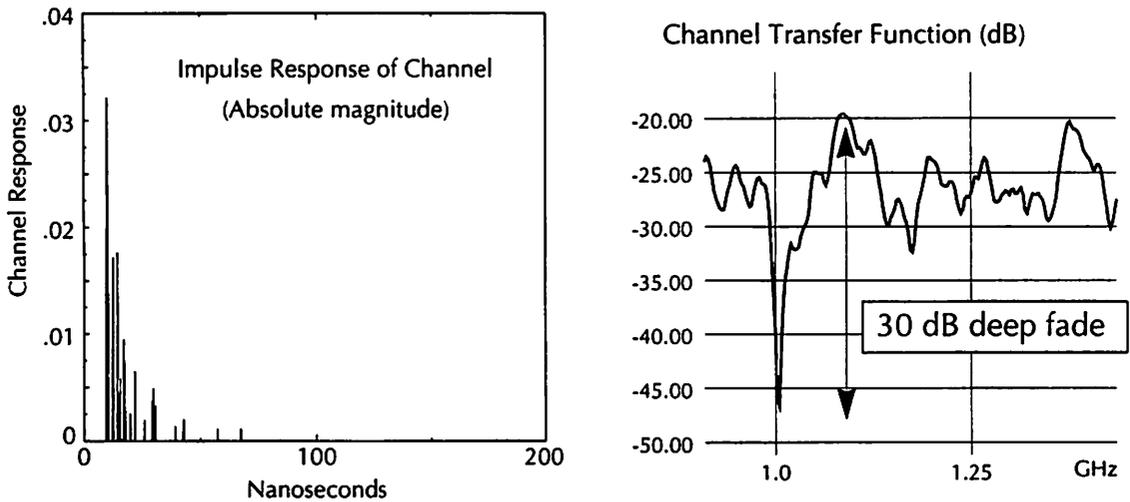


Figure 2.5: Two simulated random sample channels

(top) Strongly in-phase arrivals; (bottom) Strongly out-of-phase arrivals
Narrowband dynamic range > 30 dB
Delay spread ~ 50 nsec

above, the multipath star is used heavily in this type of simulation. A 50% excess-bandwidth raised-cosine filter response is assumed for the combined transmit and receive bandlimiting filters, with the filter partitioned equally between the transmitter and the receiver, i.e., the transmit filter has a frequency response equal to $\sqrt{H(j\omega)}$, where $H(j\omega)$ is the transfer function of the raised-cosine lowpass filter, and similarly for the receive filter. Also, the amplitude of the transmitted signal is normalized to unity, since only the relative loss is of real interest.

In figure 2.6, the ideal (no channel) eye diagram of the in-phase received signal is shown after a transmission of 100 symbols; as expected, the “eyes” are fully open, and the signal value there is precisely contained in the set $\{-1,0,1\}$, indicating zero distortion and perfect recoverability. The waveform in between the sampling points is the response of the raised -cosine filter; this also verifies that the filter responses were designed correctly, as the Nyquist zero-forcing criterion is clearly achieved. In figure 2.7, the channel is now inserted between the transmitter and receiver in the system, and the resulting eye diagram is shown. Demonstrating the key difficulty in transmitting within such an environment, the eye openings are now relatively closed due to the multipath distortion. The “zero” level has clearly split into two; this is to be expected, since when the in-phase symbol has value zero, the quadrature signal is one. Due to the random phase shift caused by the channel, the in-phase signal is corrupted by crossover distortion from the quadrature phase.

2.2. Modulation Techniques

When considering the design of the baseband signal modulation, it is important to note that this discussion is essentially restricted to the digital domain, and hence complex filtering and signal generation are not a problem. For example, simple frequency-shift keying, which uses different frequency tones to encode the data, traditionally utilizes a low-frequency analog voltage-controlled oscillator to generate the tones, and then modulates this signal up to the passband. However, settling time, temperature stability, phase noise, and other nonidealities in the baseband oscillator often complicate the system design. Today, a simple direct-digital frequency synthesizer can easily generate the baseband signals, without any of the difficulties encountered with analog designs. The use of digital baseband processing allows one to limit all of the analog hardware to the passband conversion circuitry, and opens the possibility of

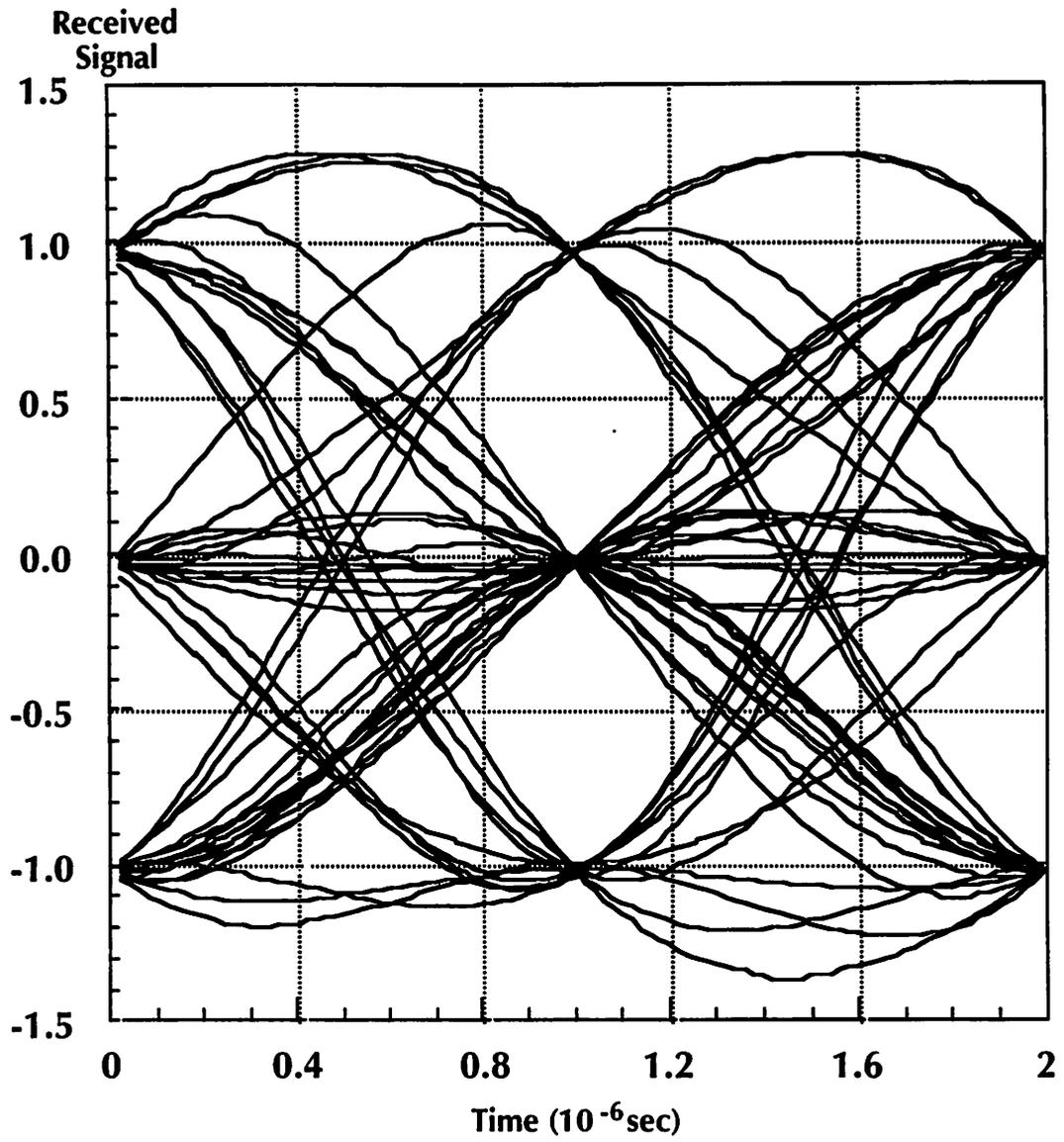


Figure 2.6: Simulated QPSK eye diagram, no channel ISI

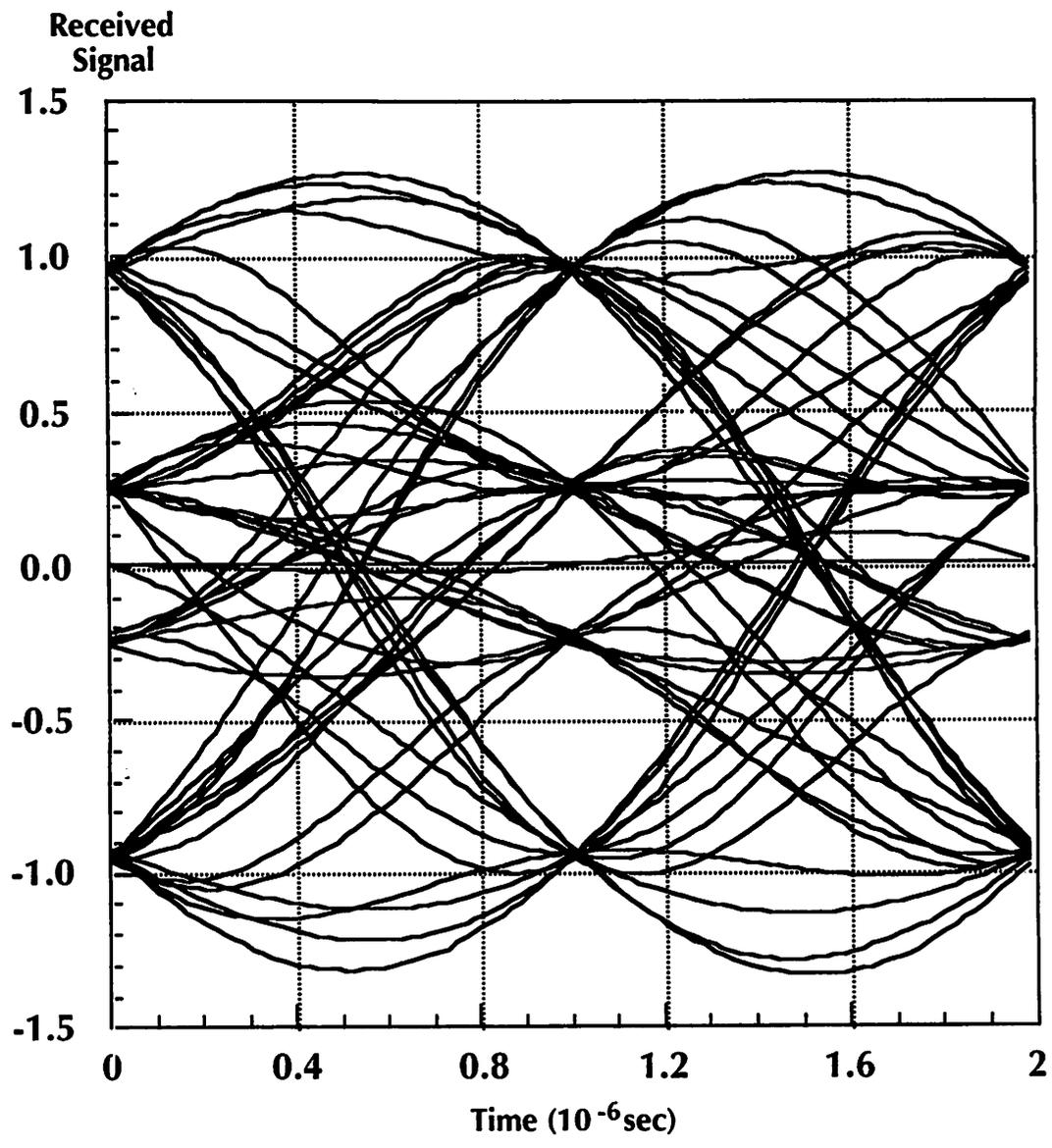


Figure 2.7: Simulated QPSK eye diagram, with multipath channel

simplifying the analog hardware at the expense of more complicated baseband processing. These concepts will be explored more fully in subsequent chapters; at this point, it is simply important to realize the complexity in modulation synthesis is no longer as great of a concern as it once was, given ready access to digital signal processing. However, the hardware in the receiver - required to recover the modulated signal - must still be minimized, given the need for low power and high performance.

2.2.1. Quadrature Amplitude Modulation

QAM is simply the digital version of classical analog AM, using modulation on both the in-phase and quadrature signals to achieve the same spectral efficiency as single-sideband modulation [Lee88]. Essentially, a sinusoidal carrier waveform is linearly modulated by bandlimited signal of the form:

$$m(t) = \sum_{k=0}^{\infty} A_{kT} g(t - kT)$$

where $\{A_{kT}\}$ are the complex-valued transmit symbols corresponding to the user data, and $g(t)$ is a bandlimited baseband pulse. This results in a transmitted signal $s(t)$ equal to:

$$s(t) = 2\text{Re}\{e^{j\omega_c t} m(t)\} = \sum_{k=0}^{\infty} 2\text{Re}\{(A_{kT} g(t - kT)) e^{j\omega_c t}\}$$

The $\{A_{kT}\}$ data symbols are chosen from a symbol constellation, as shown in figure 2.8; each group of bits to be transmitted is encoded into the complex-valued symbol shown. Clearly, the larger the constellation, the more bits are transmitted per symbol, and hence the required transmission bandwidth is reduced. However, since the distance between constellation points is reduced as the constellation size increases, the probability that a symbol error occurs also increases. Under additive white Gaussian noise conditions, the probability of a bit error for a QAM signal is approximately given by:

$$P_{\text{error}} = \left(\frac{\eta}{M\sqrt{2\pi}}\right) \int_q^{\infty} e^{-(x/2)^2} dx = \left(\frac{\eta}{M}\right) F_{\text{err}}(q) \leq \left(\frac{\eta}{M\sqrt{2\pi}}\right) \exp(-q^2/2)$$

where $q=(d_{\min}/2\sigma)$, d_{\min} is the minimum distance between two nearest-neighbor constellation points, σ is the noise power, M is the number of bits per symbol, and η is a degeneracy factor equal to the number of nearest neighbors at the minimum distance.

The primary advantage of QAM is that it is known to be spectrally efficient, which is ideal for a personal communications system with its severe bandwidth requirements; also, it is conceptually simple to equalize out the effects of the channel since the use of adaptive linear equalization is possible. However, the receiver hardware complexity immediately comes into

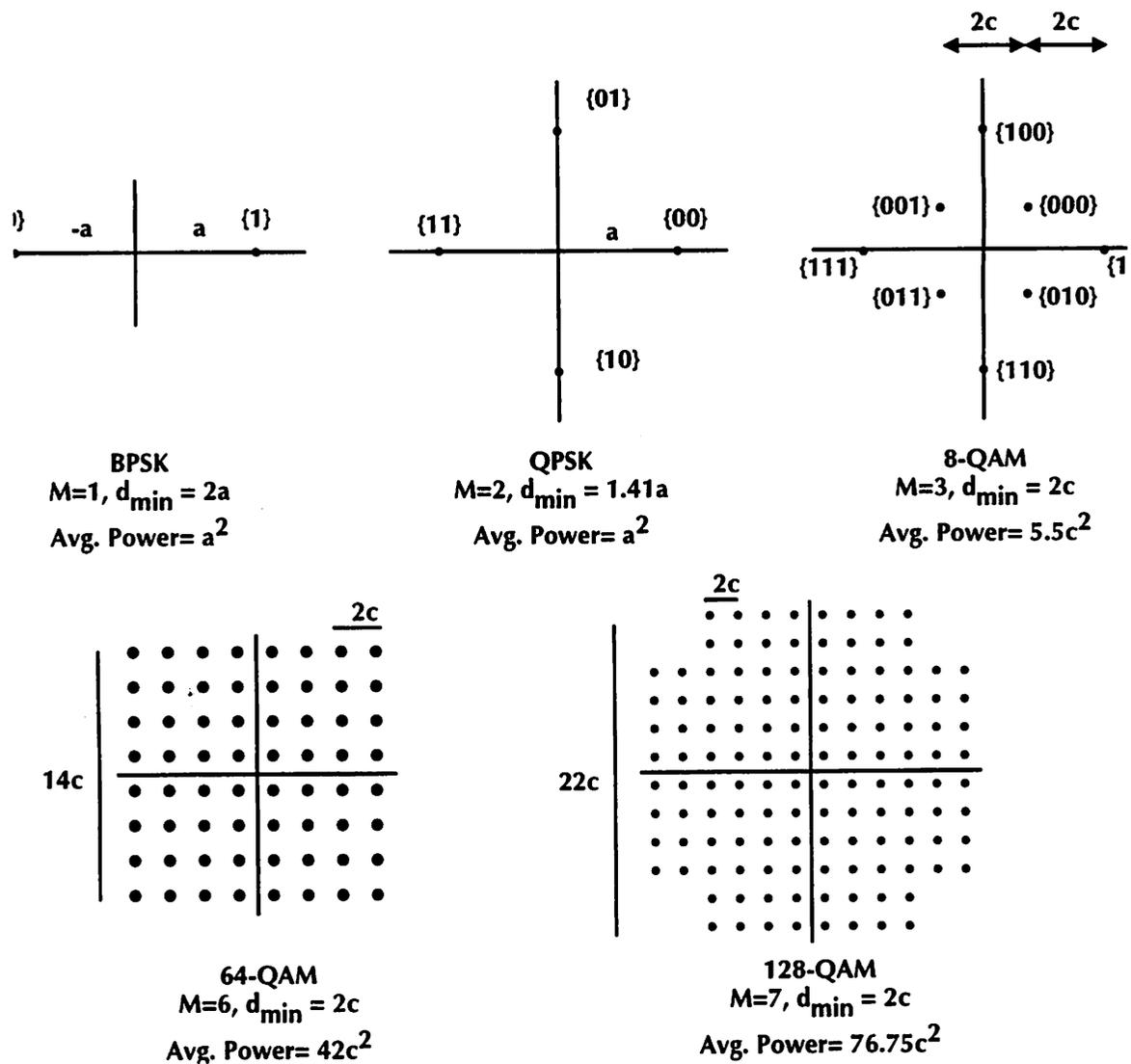


Figure 2.8: QAM Constellations (Cross-configuration)

question, since coherent demodulation has traditionally been utilized requiring phase- and carrier-recovery circuitry. For these high data rates at such high carrier frequencies, essential components such as rejection filtering, carrier recovery, automatic gain control, and voltage-controlled local oscillators become complex and expensive to implement. If QAM is to be employed, methods of minimizing the receiver complexity must be developed if the spectral efficiency of QAM is to be exploited; for example, the use of differential phase encoding, instead of absolute phase, obviates the need to perform coherent data recovery.

2.2.2. Continuous-Phase Modulation

Continuous-phase techniques present an interesting class of modulation schemes that have the desirable property that they can be incoherently detected with minimal hardware complexity. They are characterized by a constant amplitude envelope, with all of the information carried in the phase of the signal alone. The simplest form of CPM is FSK (frequency-shift keying), in which two separate frequencies are used to distinguish “0” and “1” data bits [Lee88]. The general form of the transmitted signal is given by [Ekel87]:

$$(s(t) = \sqrt{2(E_s R)} \cos(\omega_c t + \varphi(t, \alpha)), \varphi(t, \alpha) = 2\pi h \sum_{k=0}^{\infty} A_{kT} q(t - kT)$$

where $q(t) = \int_{-\infty}^t g(\tau) d\tau$, E_s is the energy per symbol, and A_{kT} are the transmitted data bits.

Since this is a constant envelope modulation scheme, CPM is power efficient at the transmitter, since a class-C nonlinear power amplifier can be used. Likewise, it is immune to channel amplitude nonlinearities; for some schemes the receiver can simply hard-limit the signal without any loss of information or signal degradation, obviating the need for high-precision automatic gain control as needed by linear modulation schemes. Another important advantage is that most CPM schemes can be demodulated incoherently, in that the absolute phase of the modulation carrier need not be known for reliable data recovery; for example, to demodulate FSK a simple band-pass filter bank can be used to detect the various frequency tones, thus recovering the data without need for complex phase recovery.

Conceptually, CPM schemes are the digital equivalent to analog FM, with the digital signal first being low-pass filtered before frequency modulation, as shown in figure 2.9. The name CPM comes from the requirement that the phase be continuous at all points, which implies that no “jumps” in phase are allowed, as in simple FSK. This continuity requirement improves the spectral efficiency significantly, and it is the bandwidth of the low-pass filter which determines the overall spectrum consumption of the system. A simple extension of FSK that meets this continuity requirement is minimum-shift keying (MSK) [ProakB83]; it does so by a careful choice of harmonically-related signaling frequencies and signal polarities. However, it is not spectrally efficient; one related “spectrally efficient” CPM scheme is GMSK, or Gaussian Minimum-Shift Keying¹ [Ekel87, Murota81]. The low-pass filter is specified to have a Gaussian response, and it has been shown that to maximize the spectral efficiency the bandwidth of the Gaussian filter should be set to approximately $0.25R$, where R is the bit rate. To compare with simple MSK, the 99% power containment bandwidth for GMSK can be shown to be $0.86R$ [Murota81], as opposed to $1.2R$ for MSK.

However, constant-envelope schemes also have several serious disadvantages. First, on the time-varying fading channel that the indoor environment presents, some form of equalization is usually required to neutralize the resulting intersymbol interference. Such equalization becomes extremely difficult with CPM schemes owing to the fact that the signal is not a linear function of the data, making adaptive techniques nearly impossible. Second, the spectral efficiency of CPM is not as good as that of QAM – given the severe bandwidth constraints in the

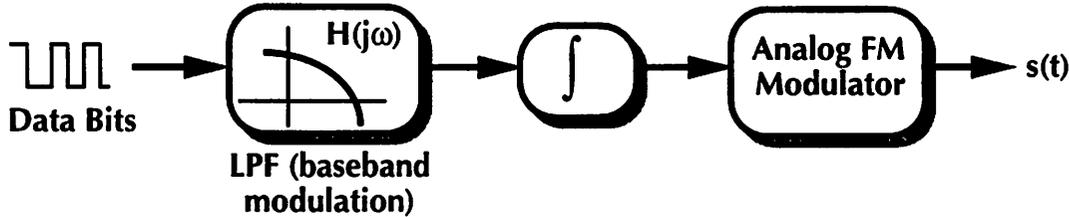


Figure 2.9: General CPM Modulator

1. GMSK has been of the object of significant interest recently, since it is the modulation scheme that has been specified for use in the European GSM digital cellular system.

desired link, this makes CPM intrinsically unattractive as a modulation format. The question immediately arises: QAM or CPM? A quantitative analysis, considering the particular characteristics of the downlink in question, is given below.

2.2.3. Modulation Schemes: Comparison and Analysis

In figure 2.10, a comparison of the important properties of several modulation schemes is shown; the modulation schemes listed represent the majority of the practical systems in use today. All the schemes are assumed to have the same transmit power P_{transmit} and bit rate R bits/sec, and subjected to additive white Gaussian noise with noise power σ^2 . For the class of linear schemes, the baseband pulse shape $g(t)$ is assumed to be a Nyquist raised-cosine pulse with 50% excess bandwidth.

From the table, it is immediately evident that linear QAM schemes are more spectrally efficient than nonlinear CPM ones, since Q is assumed to be constant. Even for a small constellation such as QPSK, the spectral efficiency already surpasses that of an optimal GMSK scheme. This argues strongly that for a high-data rate, wideband application like the personal communications system a linear modulation scheme should be utilized. To date, nonlinear modulation has been almost invariably favoured given its transmit power efficiency and incoherent reception capabilities. However, for a wideband microcellular system like the Infopad downlink, where transmit power is already minimal by necessity and bandwidth is critical, continuous-phase modulation is simply not practical. .

Considering the linear schemes, another evident point is that as the constellation size increases, the BER also increases dramatically, since the complementary error function is a strong function of its argument. Intuitively, one can understand this as the distance between the constellation points is decreasing since P_{transmit} is held constant. As the noise power also remains constant, the probability that a symbol is received incorrectly will increase. Plotted in figure 2.11 is a graph of the required signal-to-noise ratio $P_{\text{avg}}/(RN_o)$, which is the ratio of received energy per bit to the power spectral noise density at the receiver, versus the size of the constellation for constant BER; superimposed is the bandwidth decrease relative to the BPSK case for a fixed data rate. Due to the sensitivity of this curve to SNR, small constellations are normally used in RF systems, such as QPSK or BPSK. Although it seems that only a few dB

Modulation Scheme	Type	Bits/Symbol	99% Power BW (Hz)	Spectral Efficiency (Bps/Hz)	Probability of bit error (BER-1)	Demodulation
BPSK	Linear	1	1.5R	0.66	$F_{err}(\sqrt{Q/1.5})$	Coherent
QPSK	Linear	2	0.75R	1.33	$F_{err}(\sqrt{Q/1.5})$	Coherent
DQPSK	Linear	2	0.75R	1.33	$F_{err}(\sqrt{Q/3})$	Incoherent
8-PSK	Linear	3	0.5R	2	$F_{err}(\sqrt{Q/2.75})$	Coherent
8-QAM	Linear	4	0.35R	2.85	$F_{err}(\sqrt{Q/3.5})$	Coherent
128-QAM	Linear	7	0.21R	5	$F_{err}(\sqrt{Q/16.1})$	Coherent
MSK	CPM	1	1.2R	0.833	$F_{err}(\sqrt{Q/2})$	Incoherent
GMSK	CPM	1	0.86R	1.16	$F_{err}(\sqrt{Q/2.38})$	Incoherent

Figure 2.10: Comparison of Modulation Schemes (AWGN only)

$Q = P_{avg} / (RN_o)$ (E_b/N_o ratio, thermal noise power only)

N_o = Power Spectral Density of additive noise (W/Hz)

R = Bit rate (bps); P_{avg} = Average transmit power (W)

BPSK = Binary Phase-Shift Keying

QPSK = Quadrature Phase-Shift Keying (4-PSK)

DQPSK = Differential Phase-Shift Keying (same constellation as QPSK, data is encoded on phase difference).

MSK = Minimum-Shift Keying.

GMSK = Gaussian Minimum-Shift Keying (BW of Gaussian LPF=0.25R).

Data for MSK, GMSK from [Murota81]; the above neglects the constant multiplier for BER, since it makes little difference at high SNR.

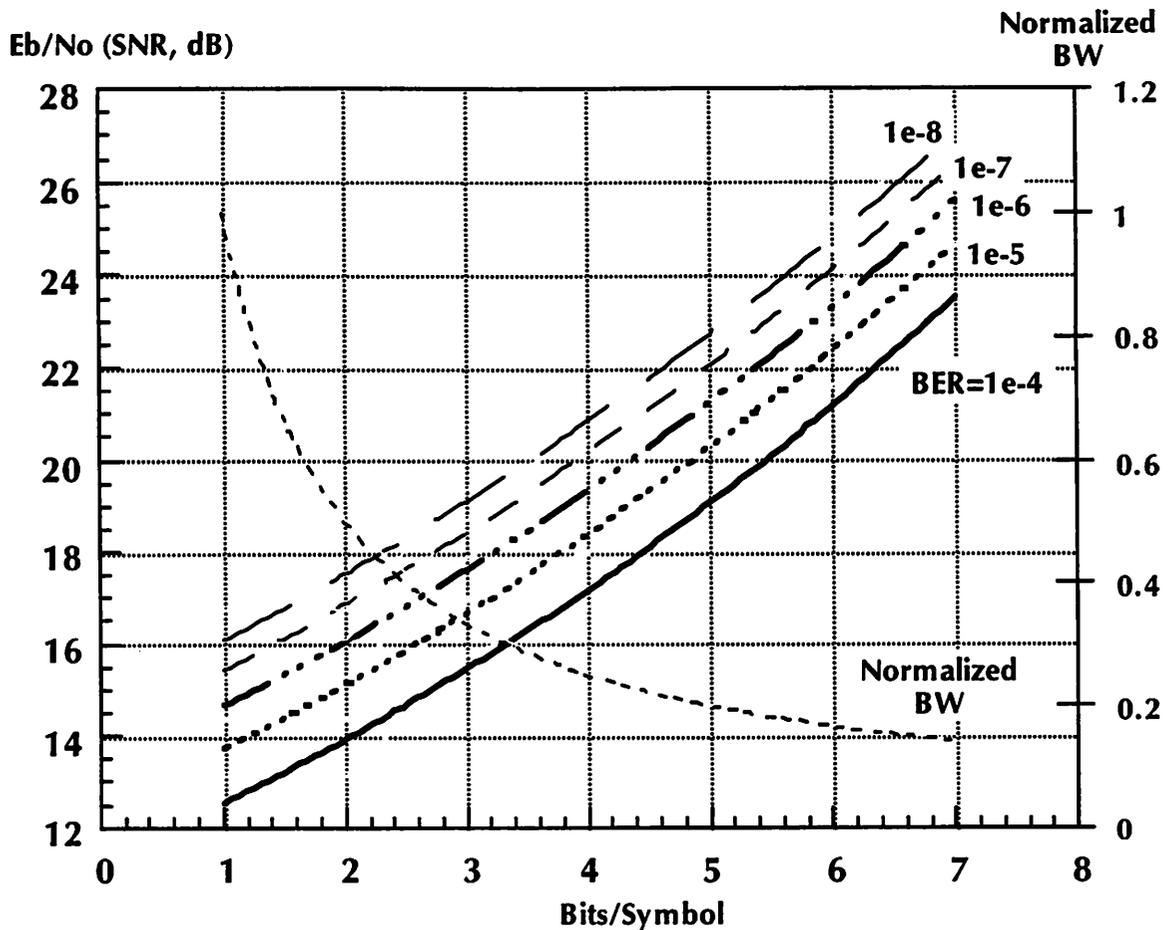


Figure 2.11: Required SNR vs. Constellation Size

in SNR can purchase a doubling in bandwidth (going from QPSK to 16-QAM), due to multipath fading, cochannel interference, and distortion and noise in the receiver, mobile wireless systems usually maintain a larger margin to preserve link under worst-case conditions. It may be possible that a larger constellation size, combined with error-correction coding (e.g., trellis coding), will provide greater spectral efficiency under the same fading conditions as a smaller constellation without error coding; however, this discussion is beyond the scope of this thesis.

2.3. Multiple Access Strategies

Multiple-access techniques are methods by which many users can access the system simultaneously, where users are separated from one other by allocating a small slice of the available resources to each user. For example, when cellular technology is employed, a form of spatial multiple access is achieved, in which two users separated by a sufficient distance can independently use the same frequencies at the same time. Another classical scheme is frequency-division multiple access (FDMA), where users are separated from each other by assigning a distinct frequency band to each. Hence, two sets of independent data can be transmitted at the same time at the same location, without resulting in interference between the two users. Below, a short discussion is presented for three other important schemes: time-division multiple access (TDMA), direct-sequence spread-spectrum (code-division) multiple access (DS/SSMA, or CDMA), and frequency-hopped spread-spectrum multiple access (FH/SSMA), which is a variant of FDMA.

Of importance is the fact that FDMA, TDMA, and direct-sequence CDMA can all be viewed as partitionings of the available signal space into N subsets, where N is the number of users that are to be multiplexed. Optimally, N is equal to the dimensionality of the signal space, which has been shown to be of dimension $2BT$, where B is the total available bandwidth for transmission, and T is the amount of time that is available for the N users to all transmit one symbol. For example, in FDMA each user of the N users transmits in a bandwidth of B/N , over the complete time interval T . As discussed below, for TDMA each user transmits in separate bursts of time of length T/N , over the complete bandwidth B , and for direct-sequence CDMA each user is assigned a one-dimensional subspace derived from one of $2BT$ orthogonal basis vectors for the signal space. The important realization here is that no one scheme is inherently superior to any other; it is only after considering the transmission environment and the required performance that advantages and disadvantages can be discussed.

2.3.1. Time-Division Multiple Access

In TDMA, the users are separated on a common communications medium by ensuring that they simply transmit at different times. Perhaps the most obvious form of multiple access, it is surprising simple in concept, and variants have long been used in satellite communications

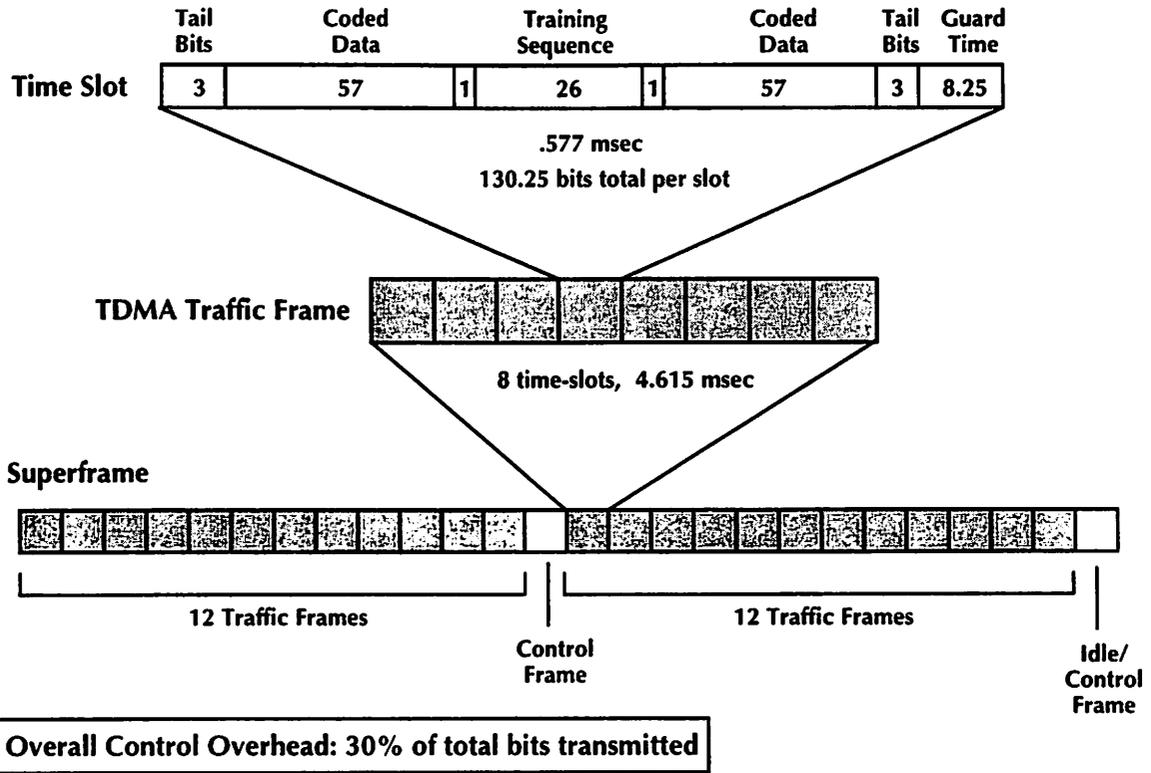


Figure 2.12: TDMA Framing/Control Strategy (GSM)

and wired local area networks such as Ethernet. Transmitted bursts of “low data rate” user information is multiplexed and interleaved into time-slices on a “high data rate” communications channel, as shown in figure 2.12. Received TDMA data is put through a demultiplexor to reverse the interleaving process, extracting only the time slices that belong to that user. The time slices of user data are typically placed into frames, where each frame contains overhead acquisition information for timing and carrier recovery (framing bits) in addition to the data itself. From this, a considerable amount of overhead is clearly expended in reacquiring synchronization and training the adaptive equalizers from frame-to-frame; for the system of figure 2.12, 1 in 3 of the bits transmitted is used only for control purposes.

TDMA multiplexing has been heavily favored for the past several years for use in second- and third- generation microcellular digital radiotelephony systems such as the European GSM project [Raith91], mainly because of its simplicity and ease of implementation. However,

problems have already been encountered with the use of TDMA on microcellular channels; to surmount the time-varying channel multipath distortion and fading necessitates the use of an adaptive decision-feedback equalizer (DFE) [Sven90].

The usual design rule is that the number of taps in the equalizer structure should provide a time-delay span approximately equal to the length of the channel impulse response that needs to be equalized. For low-data rate (100 kbps TDMA burst rate, 32 kbps average user rate) voice telephony applications, this implies a 3-7 tap adaptive equalizer¹ should be sufficient, where the number taps is the total of the forward filter and feedback filter [Proak83]. Interestingly, for the indoor picocellular channel that we are considering, a 3-7 tap structure would also suffice in spite of the higher data rate. From the analysis in section 2.1.2, a 100 MHz burst data rate is needed to support the necessary throughput to all users in the downlink, 1000x faster than the typical voiceband cellular channel. However, the cell size has also shrunk 1000x, with a commensurate compression in delay spread, thus implying that the equalizer structure will remain about the same complexity. The difference is how fast the receiver digital circuitry has to operate: the equalizer needs to function at the full 100 MHz bandwidth, which is not trivial even with today's technology. Furthermore, to enable these equalizers to converge sufficiently quickly, use of the so-called "fast" RLS, modified LMS, or the Kalman filter algorithm is needed. These algorithms are complex, and tend to be computationally intensive in the receiver. Given the complexity and difficulties that have been encountered in forcing even small adaptive decision feedback equalizers (DFE's) to converge quickly, it will be difficult to design an adaptive DFE that will meet the performance requirements of a downlink. The only mitigating factor is that the multipath profile is changing slowly with respect to time, allowing slower convergence rates, as opposed to the typical fast-fading pattern encountered with operation of cellular transceivers in a moving vehicle.

¹. These data values are for the European GSM system.

2.3.2. Frequency-Hopped Spread-Spectrum Multiple Access

A variant of FDMA, frequency-hopped spread spectrum allocates each user a small slice of the total available spectrum, where each slice is equal to the user's data bandwidth. Unlike pure FDMA, the carrier frequency at which each user is transmitting is changing with time, hence the name frequency-hopped. The spectrum is "spread" in the sense that the user ranges over the entire transmit band due to the hopping, but only occupies a small amount at any given instant.

Although it would seem that FH/SSMA would be subject to the same disadvantages that TDMA has, the transmitted bit rate for FH/SSMA is the same as the data rate, unlike TDMA. Hence, the fading is narrowband - the channel transfer profile is approximately flat over the bandwidth, and hence much less equalization is needed. In other words, relative to the symbol rate the time-span of the multipath is much shorter than that of TDMA, since we do not need to transmit in high-frequency bursts. In figure 2.13, the simulated deep-fade transfer function from figure 2.5 is plotted again, except with the bandwidth occupancy of the TDMA and FH/SSMA shown; the narrowbanding effect is evident. Of course, this narrowbanding advantage would also be realized with pure FDMA - no mention of the necessity of frequency hopping has been mentioned yet. Also shown in figure 2.13 is an example hopping pattern for FH/SSMA - on hop number 2, it drops directly into a 30 dB fading null. In all likelihood, with such a deep null, the narrowband signal will be completely engulfed by noise, resulting in link breakdown. It is the fact that it will quickly hop back out of the null that makes FH/SSMA a viable multiple access scheme at all.

The major disadvantage of FH/SSMA arises from the fact that it simply must hop in frequency, necessitating a frequency-agile wideband oscillator. Unfortunately, since the carrier must span the entire 100 MHz transmit bandwidth of the system at the gigahertz passband, direct-digital synthesis becomes impossible, thus requiring an agile analog synthesizer. The frequency settling time for accurate, low-phase noise phase-locked loops tends to be on the order of microseconds [Stir87], or tens of symbol periods, resulting in a considerable resynchronization delay every time a hop is performed. Likewise, all of the transceivers active within the cell

must be synchronized in their hopping scheme, otherwise disastrous cochannel interference will result if two transceivers randomly happen to occupy the same frequency band at the same time.

2.3.3. Code-Division Multiple Access

Another multiple-access scheme employs code division, where the users are allowed to transmit simultaneously in the same frequency band, without requiring them to be separated physically by a large distance. Instead, each user is assigned a code waveform from the set $\{p_j(t), j=1..M\}$, where the set represents an orthogonal basis for the signal space and each

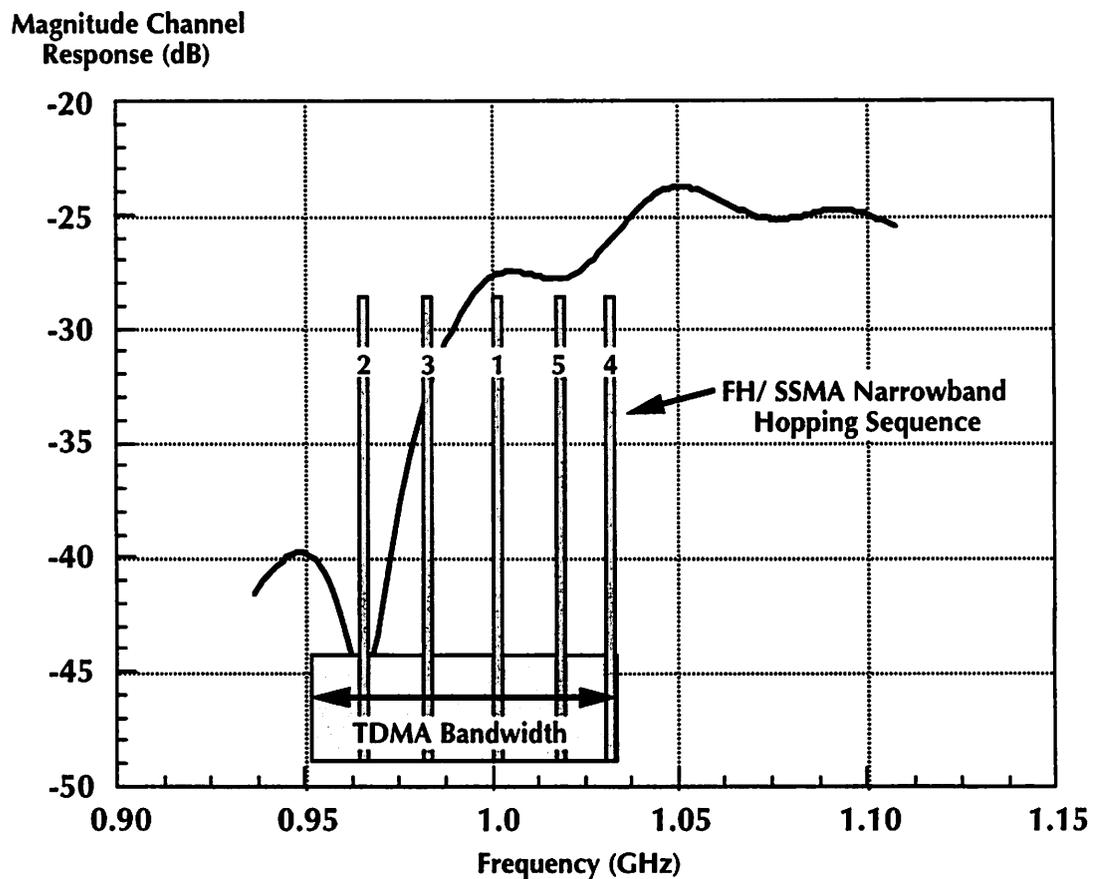


Figure 2.13: Comparison of TDMA and FH/SSMA under multipath conditions

$\{p_j(t)\}$ occupies B Hz of bandwidth. Since the set $\{p_j(t)\}$ is assumed to be orthogonal, where δ

$$\int_0^T p_k(t) p_j(t) dt = \delta(j, k)$$

is a Kronecker-delta functions assumption is also made that the autocorrelation function of each of the $p_j(t)$ is also a delta function, i.e.,

$$\int_{-\infty}^{\infty} p_j(\tau) p_j(t + \tau) d\tau = \delta(t)$$

Thus, the transmitted waveform for user j is then given by:

$$2\text{Re}\left\{ \sum_{k=0}^{\infty} A_{kT} p_j(t - kT) e^{j\omega_c t} \right\}$$

Since these signals are transmitted at the same time in the same frequency band, the received signal at any receiver is just the linear superposition of these over all N users.

Now, after the signals have traversed the channel and have been mixed down to baseband, the receiver is presented with the problem of extracting that particular user's data from the superposition of the data from many users and additive noise, i.e.,

$$\sum_{n=1}^N \sum_{k=0}^{\infty} A_{kT}^n p_n(t - kT) + n(t)$$

However, this is simply accomplished: the receiver knows *a priori* which pulse $p_j(t)$ was used to transmit. Hence, by simply correlating the received signal against the known $p_j(t)$, the correct data can be extracted, since:

$$\int_{kT}^{(k+1)T} \left\{ \sum_{n=1}^N \sum_{k=0}^{\infty} A_{kT}^n p_n(t - kT) + n(t) \right\} p_j(t - kT) dt = A_{kT}^j + \int_{kT}^{(k+1)T} p_j(t - kT) n(t) dt$$

As the noise signal is assumed to be nearly white, its power must be evenly spread out among each of the code waveforms, since these represent a basis for the signal space; hence, the pow-

er in the noise term after correlation has been reduced by approximately a factor of N . This is an elegant scheme for achieving multiple access, since it simply performs a correlation in the receiver against digital codes. In particular, the code space can be designed to consist of antipodal sequences of $\{+1, -1\}$ values. Such code sequences are easily synthesized in the transmitter; what is more important is that this implies that the multiplication required in the receiver correlator can be performed easily with a minimum of hardware expenditure, even at the speeds required by the downlink.

Essentially, direct-sequence CDMA utilizes pseudo-noise spread-spectrum techniques to achieve its ends, since the original user bandwidth has been expanded to occupy the entire bandwidth available. Alternatively, this can be viewed as multiple high-data rate symbols being sent in place of one user symbol, which effectively is a form of coding redundancy. This increase in error immunity is called the processing gain P_g , defined as the ratio of the spread spectrum to the original spectrum, and is exemplified by the apparent reduction in the noise power after the correlation despreading [Simon85, Cook83, Pick82]. However, beyond simply providing a multiplexing technique, the coding and processing gains of spread-spectrum also yield two other interesting properties: spread-spectrum signals are low-probability-of-intercept, since knowledge of the precise code used is required, and spread-spectrum signals tend to be naturally immune to multipath distortion [Lee88, ProakB83]. Both of these are excellent advantages considering the indoor environment that the transmission is taking place in, and the immunity to multipath is especially important, since it implies that the complex adaptive equalizer is not needed.

The immunity to multipath distortion can be understood by increased time-resolution of a spread-spectrum signal. Since the $p_j(\tau)$ spreading function is assumed to have a near delta-function autocorrelation, the arrival time of each symbol can be determined precisely, by using a "sliding correlator" to detect when the correlation is nonzero. For a single transmitted symbol at $t=kT$, this sliding correlation is given by:

$$\int_{-\infty}^{\infty} A_{kT} p(\tau - kT) p(\tau - t) dt$$

This is clearly nonzero only when $t=kT$, as expected. However, suppose the signal has been

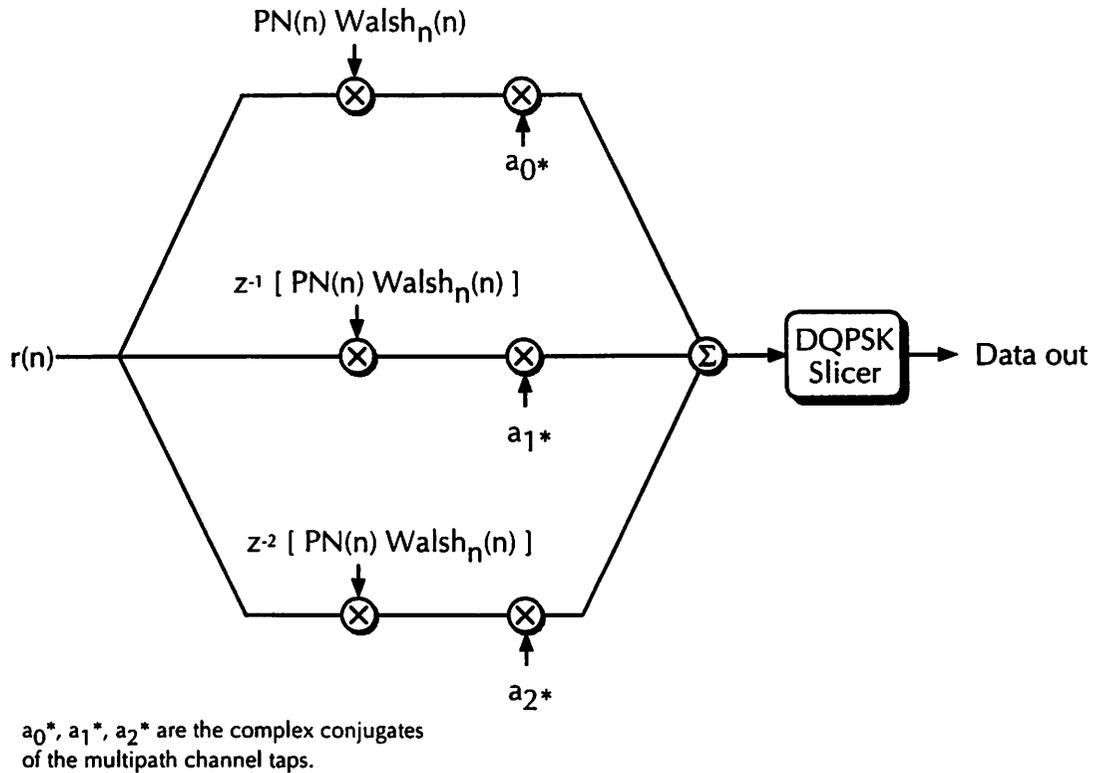


Figure 2.14: RAKE Receiver architecture (3rd order)

subject to multipath distortion. Then the sliding correlation yields a nonzero value not only at $t=kT$, but also at all of the multipath arrival times thereafter as well. Hence, the arrival times of the signal can be resolved precisely; it is only when a multipath arrival time happens to be a precise multiple of T is there interference. In fact, if multiple correlators are used, one correlator per multipath arrival, the data in the multipath itself can be resolved to create a form of time-diversity. Effectively, the multipath reflections are scaled and time-shifted copies of the transmitted information; by coherently combining these copies, the SNR of the system can be improved. This is the basis of the RAKE receiver [Dixon84]; a basic RAKE structure is shown in figure 2.14. For each multipath arrival a_k , a correlator is used to resolve the information available. By multiplying each estimate against a_k^* and summing, the estimates are phase coherent with respect to one another, and the receiver SNR has been increased by a factor equal to:

$$\text{SNR}_{\text{increase}}(\text{dB}) = 10 \log \sum_k \left| \frac{a_k}{a_0} \right|^2$$

Of course, this presupposes exact knowledge of the $\{a_k\}$ multipath coefficients. The specific performance improvements from employing RAKE combining will be considered in the next chapter.

Of course, this is assuming that the $p_j(t)$ are continuous signals that possess a delta-function autocorrelation. In practice, the $p_j(t)$ are digital signals, which take on only values in the set $\{\pm 1\}$, and are similar in nature to Walsh functions. The chip interval T_{chip} is the higher-frequency sampling rate for the $p_j(t)$ spreading function, and T_{chip} is related to the symbol period T by $T_{\text{chip}} = T/P_g$, where P_g is the processing gain given above, since a bandwidth spread of P_g yields the same increase in data rate. Thus, the multipath arrivals can be resolved to a resolution given by T_{chip} ; the number of resolvable arrivals is related to the delay spread T_{del} ; in particular, $N_{\text{resolvable}} = T_{\text{del}} / T_{\text{chip}}$. Clearly, the greater the spreading factor, the more paths can be resolved and the better the immunity.

An immediate question that arises is how many users can practically be supported in a PN spread-spectrum CDMA system. Assuming that error-control coding is utilized on top of simple spectrum spread and the waveforms are perfectly orthogonal, it has been shown [ProakB83] that the decoder error probability P_{error} is bounded by:

$$P_{\text{error}} < \left(\frac{M-1}{2} \right) F_{\text{err}} \left(\sqrt{\frac{2P_g R_c d_{\text{min}}}{(N-1) + \frac{\sigma}{P_{\text{avg}}}}} \right)$$

where M is the size of the code, P_g is the processing gain, R_c is the code rate, d_{min} is the minimum code distance, P_{avg} is the average power of the user signal, and σ represents the additive Gaussian noise power in the receiver. For a specified error probability P_{error} and N users, the design space for the required coding gain, processing gain, and receiver noise in the system is thus constrained.

Traditionally, the use of spread-spectrum has been hampered by an intrinsic set of problems: synchronization required in the receiver, efficiency in spreading code design, and near-far crosstalk. Surprisingly, each of these problems can be eliminated by the fact that the broadband downlink signal is being transmitted from a single point of origin: namely, the base station in each cell:

- In the above analysis, it was assumed that the correlation performed in the receiver is perfectly aligned with the incoming signal. If they are not synchronous, the scheme fails since the receiver only sees the partial correlation, which is designed to be small, and thus results in severe information loss. Since a base station is available, a pilot tone can be injected at the transmitter to aid the mobile timing recovery, avoiding many of the difficulties with complex matched-filter based algorithms.
- The design of the codes used, i.e., the set of orthogonal signaling waveforms, is another difficult issue that has received extensive discussion in the literature [Gold67, Sarwa80, Scholtz79]. Perfectly orthogonal codes with perfect autocorrelation are difficult to generate; hence, all practical codes always have some nonzero correlation between the $\{p_j(t)\}$ transmit waveforms. Without the ability to achieve code synchronization between all users, the available set of orthogonal codes is fairly limited; Gold and Kasami sequences [Kasami66, Gold67] are the best known candidates. The coding efficiency with such sequences is quite poor; for a spreading factor of 64, there are only 21 known Kasami sequences available, leading to an efficiency of 33%. However, with code synchronization between all users possible in the base station, the family of Walsh functions [Walsh73] can be employed, leading to an effective code efficiency of 100% (64 codes for a spreading factor of 64).
- Nonzero cross correlation in conventional spread-spectrum systems leads to near-far crosstalk problems, where a transmitter close to the receiver can corrupt the data from one further away despite code orthogonality, since the received power from the nearby transmitter is much higher, and hence the undesired correlation may mask out the actual signal. However, since all signals are originating from a single point, no near-far reception problems can occur.

Lastly, if direct-sequence CDMA is to be employed, small constellations in the QAM modulation are dictated. The use of larger constellations, while reducing symbol rates, significantly decreases the tolerable signal to noise ratio at the receiver. The situation with CDMA is

unique; as transmit power scales, so does the noise, since the vast portion of the noise interference comes from other users. Due to nonzero correlations and other nonideal effects, even greater spreading factors and processing gain will be required for large constellations, thus offsetting the decrease in symbol rate. It has been found that QPSK provides a reasonable constellation configuration for use in direct-sequence CDMA [Simon85, ProakB83].

2.4. Toward an Indoor Transmission Scheme

With the above discussions of multiplexing and modulation, the various advantages and disadvantages must be considered in light of the characteristics of the indoor picocellular propagation environment, and the data rates required by the personal communications system. The real advantage of CPM techniques, power efficiency in the transmitter, is negated by the fact that the distances involved with a picocellular scheme are small, with very low transmit power requirements. Given the poorer spectral efficiency of CPM, QAM is thus the logical choice for modulation, especially if differential phase encoding is employed to alleviate carrier recovery requirements. Under the same considerations, the equalization difficulties and overhead costs that have already been encountered with TDMA do not bode well for its use in a system that already requires user data rates of 2 Mbps. FH/SSMA is an attractive compromise between TDMA and direct-sequence CDMA, possessing multipath immunity while only requiring minimal equalization in the received signal; however, it has severe analog oscillator requirements in the receiver, as well as significant overhead incurred by the need to resynchronize after each hop. The natural immunity of CDMA to multipath, and the intrinsically digital nature of the spreading codes makes it an ideal choice for a broadband wireless indoor communications system; we will see later that the use of CDMA, with its ability to intrinsically reject additive noise, results in several significant reductions in analog hardware complexity, essentially for free.

Thus, with the transmission environment in mind, differential QAM (QPSK) coupled with code division multiple access is the candidate of choice for implementation in the downlink. The exact link specification, based on these decisions, will be discussed in the next chapter.

3 System Overview: The Broadband CDMA Downlink

From the preceding discussion, the ability of a synchronous CDMA system to “reject” additive noise and multipath interference, combined with potential reductions in implementation complexity, makes it the multiple access strategy of choice for the broadband downlink. It will become readily apparent that there is a tremendous amount of interdependency between the various parameters in the link design; it is the goal of this chapter not only to describe the parameters themselves, but also to make these interdependencies clear. In table 3.1, the link specification is given. The user layer parameters - a 2 Mbps per-user data rate and a 2-5 meter cell radius - are established by the functionality that we wish to achieve: an indoor mobile picocellular system, capable of providing ubiquitous network access for users with sufficient per-user bandwidth to sustain full-motion digital video. The direct consequences of this – the physical layer and multiple access parameters – are addressed in depth below.

3.1. Physical Layer Parameters

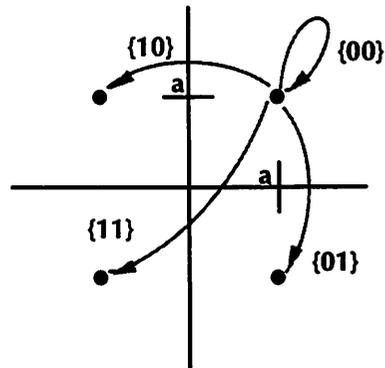
3.1.1. Modulation

As discussed in chapter 2, the spectral efficiency of QAM is extremely desirable; in light of the need for noise immunity to maximize performance of direct-sequence spread-spectrum systems, a constellation size of 2 bits/symbol (QPSK) will be used. The fact that the data is entirely encoded on phase, with a single amplitude level, greatly simplifies the receiver design in the sense that extremely accurate gain control is not needed; the automatic gain-control loop, comprising of a power detector plus a VGA, needs to be accurate only to a few dB. Even if only two amplitude levels were permitted, a few dB would be disastrous with regard to system performance, especially with multipath fading taken into account.

User Parameters	
Per-user data rate	2 Mbps
Reuse pattern	K=7
Reuse type	PN code isolation from cell-to-cell
Cell radius	2-5 meters
Physical Layer Parameters	
Modulation	Differential quadrature phase-shift keying
signaling rate	1 MHz
Carrier frequency	1.088 GHz
Transmit power	0 dBm
99% Power bandwidth	85 MHz
Transmit pulse shape	30% Excess bandwidth raised-cosine
Multiple Access Parameters	
Multiple Access	Direct-sequence, code division multiplex
Processing gain	64 (18 dB)
Aggregate downlink data rate	128 Mbps
Maximum users/base station	64
Spreading Code	64-length Walsh with pseudonoise overlay
Per-user power control	Linear 6-bits (36 dB)
Pilot Tone	32768-length pseudonoise sequence
Synchronization lock time	128 milliseconds, worst-case

Table 3.1: System link parameters

Furthermore, to simplify the receiver design further, we remove the need for absolute phase recovery by employing a differential phase encoding; as shown in figure 3.1, the data is encoded on transitions between symbols, not on the symbols themselves. So long as the channel phase is slowly varying between consecutive symbols, incoherent detection can be accomplished, without needing to know the absolute phase reference. In some sense, the system becomes self-referential; the phase reference for the next symbol is the previous received symbol. If incoherent detection is thus employed, a penalty for this is a 3 dB decrease in SNR, since the noise energy from two consecutive symbols is now introduced at the slicer. With an indoor environment, the channel phase rotation will definitely be varying slowly with respect to the signaling rate. Channel variation occurs with the movement of people and objects in the environment, on the order of seconds; the signaling rate will be at 1 MHz per user under DQPSK, thus eliminating the need to perform phase recovery.



$\pi/4$ -DQPSK,
 phase transitions in gray
 $M=2$
 Avg. Power = a^2

Figure 3.1: DQPSK phase transition diagram

The interesting phenomenon is that, with a sufficiently high signaling rate, the need for carrier recovery is similarly eliminated. Carrier recovery is needed to resolve frequency mismatch between the transmit local oscillator and the receive oscillator; by examining several families of crystal-referenced oscillators, the error can be discerned. Readily available, off-the-shelf crystal-based PLL-synthesized oscillators [Phil95, ZComm95] can be tolerated to a worst-case error between 1 part in 10^5 and 1 part in 10^6 . At 1 GHz, this translates to an error of approximately 10 kHz between transmit and receive oscillators. With narrowband systems, with a per-user signaling bandwidth of 20 kHz, it is obvious why carrier recovery is necessary. However, at a signaling bandwidth of 1 MHz, a 10 kHz error translates to a worst-case symbol-to-symbol differential phase offset of 1%, or about 3° . Compared to the interference energy due to multipath, thermal noise, and inter-cell interference, this is practically negligible. Thus, carrier recovery, which is expensive in terms of hardware and power, can be avoided in the receiver.

3.1.2. Pulse Shaping and System Transmit Bandwidth

Given that wireless systems are intrinsically bandlimited, the decision on what transmit pulse shape to use needs to be determined, and the system spectral usage. From the preliminary analysis in chapter 2, an aggregate bandwidth of 100 MHz is expected; the exact value will now be determined. From a $K=7$ reuse pattern with 8 users per cell requiring the full 2 Mbps

video data rate (as discussed in chapter 2), under DQPSK encoding (1 Mbaud/user) the net throughput for all users will be 56 Mbaud. Again, this required throughput will be independent of the multiple access strategy chosen; it is simply a function of the desired per-user throughput and the expected number of simultaneous links per base station. To provide some margin for excess, as well as simplify spreading code generation (see section 2.2) for direct sequence multiple access, an aggregate signaling rate of 64 Mbaud is chosen.

Since spectral efficiency is desirable in such a broadband system, minimizing the baud/Hz ratio becomes paramount, and is a direct function of the excess bandwidth (defined as the fraction above 1 baud/Hz necessary to reliably transmit the signal) of the chosen transmit pulse shape. A simple analog RC filter will not suffice: to achieve even reasonable ISI performance, it requires excess bandwidths well above 100%, which is not feasible for this system. Fortunately, complex DSP is available as a synthesis solution for such a pulse; there are no real limitations from implementation complexity, beyond that of available die size. Clearly, the pulse should achieve zero ISI; it has been shown [Lee88] that such pulses must satisfy the frequency-domain Nyquist zero-ISI criterion:

$$\frac{1}{T_{\text{baud}}} \sum_k P\left(j\omega - \frac{jk2\pi}{T_{\text{baud}}}\right) = 1$$

The family of raised-cosine pulses, given in the frequency domain as:

$$P(j\omega) = \begin{cases} 1, & 0 \leq |\omega| \leq (1 - \alpha)\frac{\pi}{T_{\text{baud}}} \\ 0.5\left(1 - \sin\left(\frac{T_{\text{baud}}}{2\alpha}\left(|\omega| - \frac{\pi}{T_{\text{baud}}}\right)\right)\right), & (1 - \alpha)\frac{\pi}{T_{\text{baud}}} \leq |\omega| \leq (1 + \alpha)\frac{\pi}{T_{\text{baud}}} \\ 0, & |\omega| > (1 + \alpha)(\pi/T_{\text{baud}}) \end{cases}$$

satisfies this, where α is the excess bandwidth and T_{baud} is the baud interval. Indeed, the eye diagram reveals that such pulses do satisfy the zero ISI criterion (figure 3.2). As an implementation issue, as the excess bandwidth approaches zero, the required number of taps in the filter increases rapidly, to minimize the effects of length truncation in the filter. A reasonable requirement is that the first out-of-band sidelobe due to truncation effects is below 40 dB, sim-

ilar to FCC specifications for other wireless transmission systems. With this in mind, a plot of FIR filter complexity (at a 4X oversampling ratio over T_{chip}) is shown in figure 3.3; from this, a choice of 30% excess bandwidth was made, being the largest number of taps that could be implemented in this process.

It should be mentioned that another criterion for selecting the excess bandwidth is given by the timing jitter sensitivity of the pulse shape (the horizontal “opening” of the eye) - timing sensitivity increases with decreasing excess bandwidth, as shown in figure 3.3. Although the eye width theoretically goes to zero for zero excess bandwidths [Lee88], the probability of this occurring is infinitesimally small given that the pulses have been coded using the combined Walsh-PN sequences. Empirically, it was found that 30% excess bandwidth provided more than sufficient timing jitter immunity: a 1.2 dB amplitude error given the $T_{\text{chip}}/8$ worst-case timing error in the receiver (chapter 7).

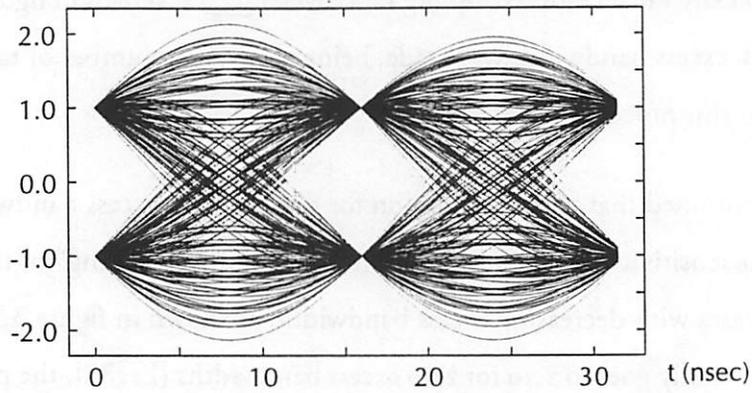
3.1.3. Transmit Power

Next, the allowed transmission power needs to be considered. Assuming that the condition of no cochannel interference is required, and a cellular reuse factor K of 7 with cells of 4 meter radius, the frequency reuse distance is thus 18.3 meters. Given that the 1 meter reference path loss at 1 GHz is approximately 30 dB for small dipole whip antennas, a conservative power loss through 5 walls of 20 dB, and using the median path loss coefficient of 2.6, the power in the nearest cochannel cell is given by the link-power equation

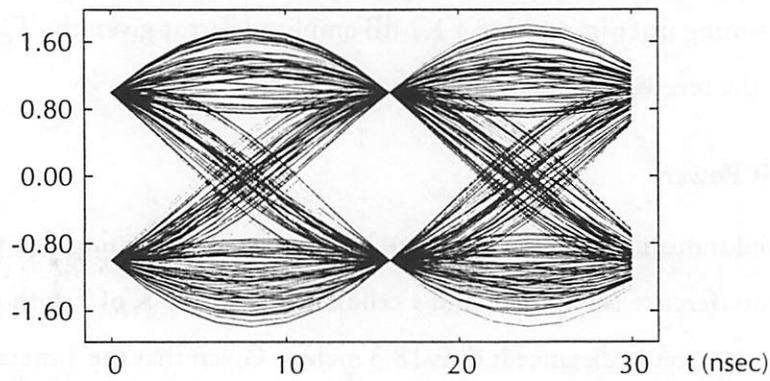
$$P_{\text{cochannel}} = P_{\text{transmit}} - 30 \text{ dB} - 20 \text{ dB} - 2.6 \cdot 10 \log (18.3/1)$$

Assuming a transmitter-receiver separation of 5 meters, the power of the received signal is equal to $P_{\text{transmit}} - 30\text{dB} - 2.6 \cdot 10 \log (4/1)$, or 34.6 dB above the cochannel interference power. Evidently, the isolation provided by the walls is significant in allowing such heavy frequency reuse; without the wall isolation, the cochannel interference power would be only 16 dB below the signal level. The design condition chosen is intended to make the cochannel interference power equal to the thermal noise power at the receiver, which as described below is limited by the performance of the input amplifier. In this way, neither the cochannel interference nor the receiver noise is allowed to dominate, yielding a maximum SNR utilizing the

10% Excess Bandwidth Raised-Cosine Eye Diagram



30% Excess Bandwidth Raised-Cosine Eye Diagram



100% Excess Bandwidth Raised-Cosine Eye Diagram

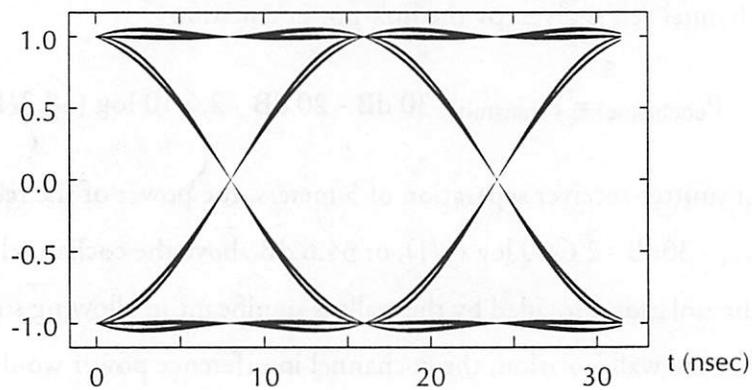


Figure 3.2: Raised-cosine pulse shape eye diagrams.

Note the decreasing sensitivity to timing jitter with increasing excess bandwidth.

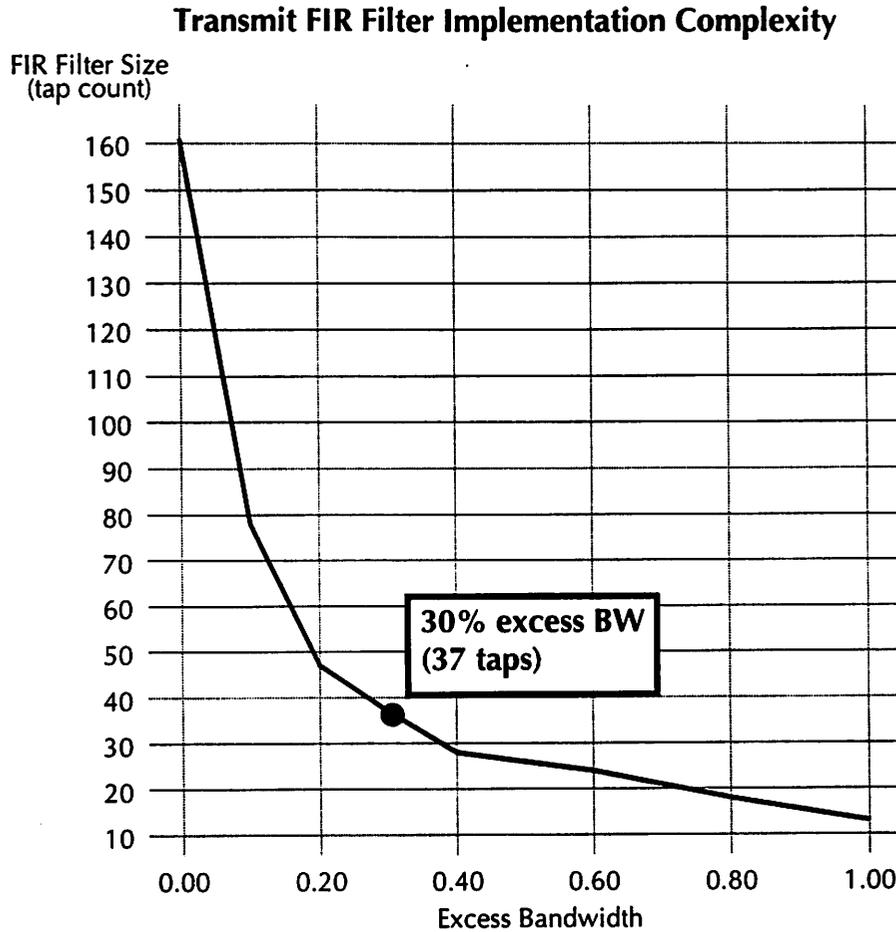


Figure 3.3: Transmit filter complexity versus excess bandwidth.

minimum transmission power. For a high-performance, low-noise amplifier, a noise figure as low as 5 dB can be realized, relative to a 50Ω load, resulting in a noise power of -87.5 dBm over a bandwidth of 1 MHz. Using the link power equation above, this implies that P_{transmit} is optimal at -4.1 dBm, or 0.33 mW; an upper transmit power limit is thus set at 0 dBm, to provide some margin for larger coverage areas. Even accounting for a ± 10 dB variation owing to the crudity of the wall-attenuation assumption, this is significantly smaller than the +30 dBm used in conventional analog cellular. As a rather striking demonstration of the impact of system design on hardware, this radically alleviates the design problems of achieving high-power RF amplification, which has been a critical issue in traditional RF systems.

3.1.4. Carrier Frequency

Lastly, a decision on carrier frequency needs to be made. Given that spread-spectrum is to be employed, a contiguous block of 85 MHz will need to be allocated, in a spectral band that does not already have significant energy. Given that the sheer channel capacity is already in question, clean spectrum will be needed; in-band interferers such as microwave ovens will simply eat into available capacity, disastrously so given the extremely low transmit power in the system. The size of the required spectral band immediately eliminates anything below 1 GHz; however, due to limitations in the available CMOS fabrication technology (a 0.8 micron HP/MOSIS process), the carrier frequency cannot significantly exceed this. In figure 3.4, the spectral energy is shown from 500 MHz to 1250 MHz. The significant received energy near 850 MHz is due to analog voiceband cellular; at 934 MHz, a nearby paging transmitter is

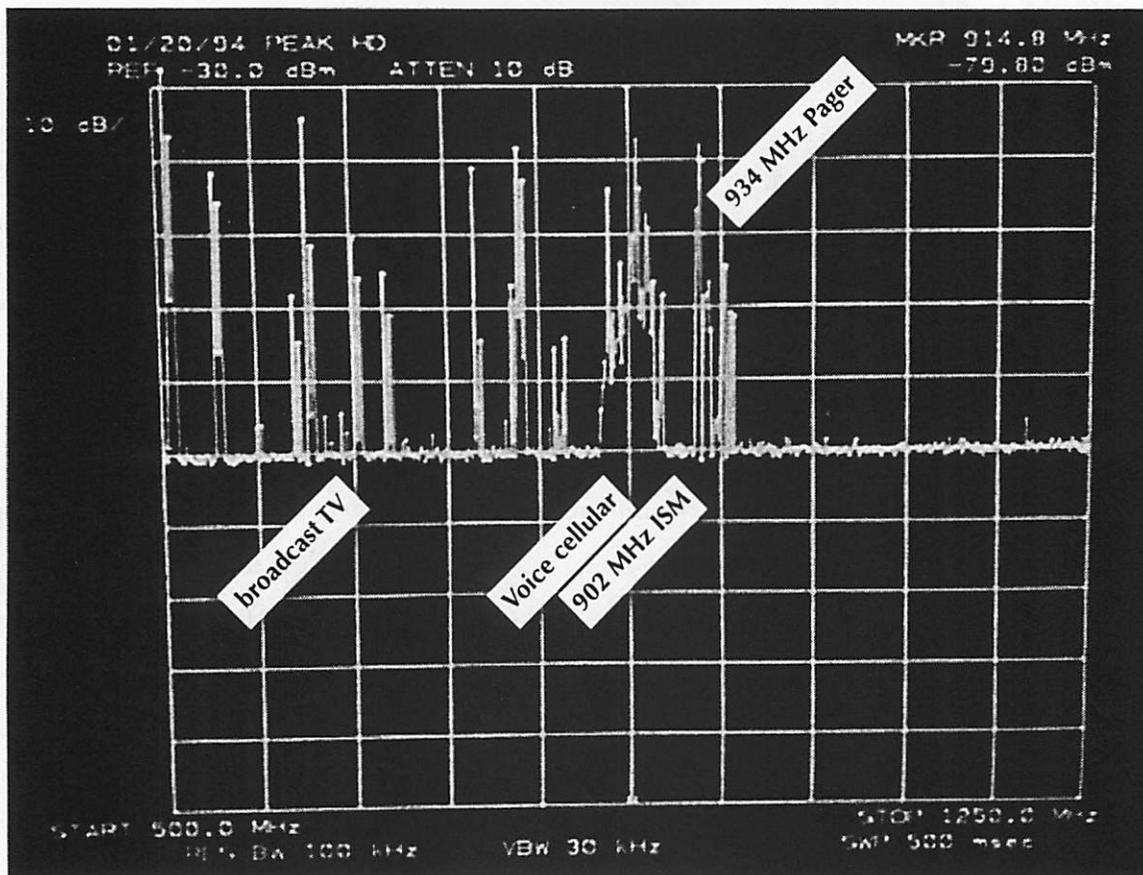


Figure 3.4: Spectral usage, 500-1250 MHz

emitting extremely strong narrowband carrier tones. Between 902-928 MHz, the open spectrum is the so-called ISM (industrial, scientific, medical) band – commercial microwave ovens radiate in this band, as well as digital cordless phones and wireless LAN's. Given that this is unregulated spectrum, the ISM band would appear to be suitable for operation; however, the number of potential local interferers, in addition to the small (26 MHz) size of the band, makes it unfeasible. Above 975 MHz, however, the spectrum is quite clean. Thus, we place the transmit carrier at 1.088 GHz, achievable by the available fabrication technologies. Obviously, FCC regulation needs to be considered if the system is to be deployed in the field; however, for lab experimentation purposes, the low 1 GHz band is ideal, especially with output power levels at or below 0 dBm. The specific value of 1.088 GHz will be made clear in chapter 5, where the use of sampling demodulation will be discussed and its need for a harmonic relationship between the baseband sampling frequency and the carrier.

3.2. Multiple-Access Parameters

3.2.1. Processing Gain and Walsh Spreading Codes

The key behind any direct-sequence CDMA system is the design of its spreading code; the performance of the code determines how many users can be effectively multiplexed, as well as how robust the system is in the presence of multipath fading. In this system, since the transmit signal comes from a single point-of-origin (the base station), code synchronization is possible between the various users, and this fact will be exploited to its fullest potential. The coding scheme described below – a hybrid PN-Walsh code – is a scaled version of that employed by the IS-95 digital cellular standard [Qual92]; it has been adapted to meet the particular data rate and transmission environment in question. From the system bandwidth analysis, each user bears a signaling rate of 1 Mbaud, with a minimum spread rate of 64 Mchip/sec (64 chips are transmitted per user bit). The required spreading factor may be higher than this, since coding inefficiencies, like those found in the Gold codes, may require a larger spreading factor to accommodate 64 users simultaneously.

Fortunately, a code with perfect efficiency - the Walsh functions - can be employed for this application, and thus the spreading rate for the system is 64 Mchip/sec. The Walsh functions are binary antipodal $\{+1, -1\}$ sequences that are designed, by definition, to be perfectly orthog-

onal to one another. In figure 3.5, several Walsh functions are shown; the Walsh number given for each corresponds to the frequency content, or “sequency”, of that function. For example, the zeroth Walsh function is a constant DC value, whereas the 64th Walsh function is alternating +1/-1, being of maximum frequency. Indeed, perfect efficiency is achieved: as shown in figure 3.6, each user’s data is spread by 64, and then all of the spread data streams are linearly combined.

The immediate question arises of why these codes are not used in all direct-sequence CDMA systems. The problem arises from the fact that these codes, while possessing perfect cross-correlation properties, are extremely poor in autocorrelation: they do not appear to be “white” noise. The desirable property of multipath rejection, which depends so heavily on autocorrelation performance, is completely lost. Furthermore, timing recovery in the receiver is exacer-

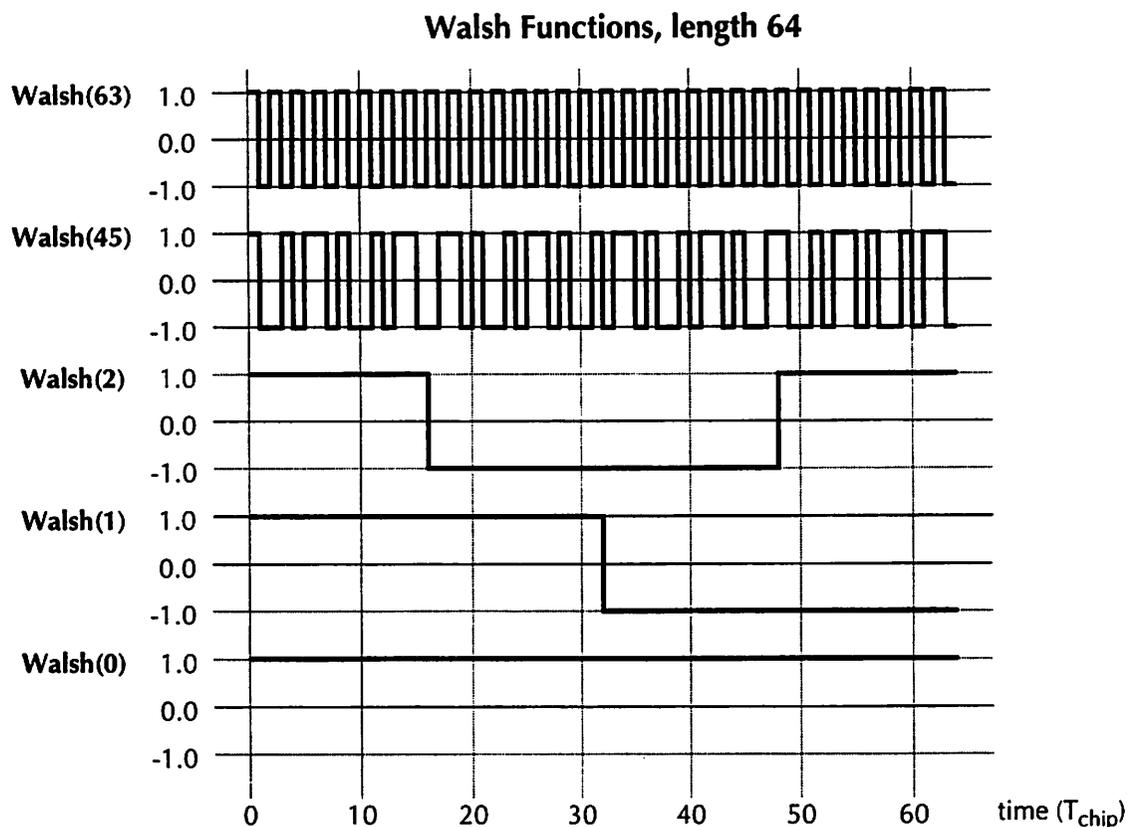


Figure 3.5: Sample Length-64 Walsh Functions

Illustrates the poor autocorrelation properties of the Walsh functions, as well as the need for code-to-code time synchronization (Walsh(2) and Walsh(1) are indistinguishable under time shift.)

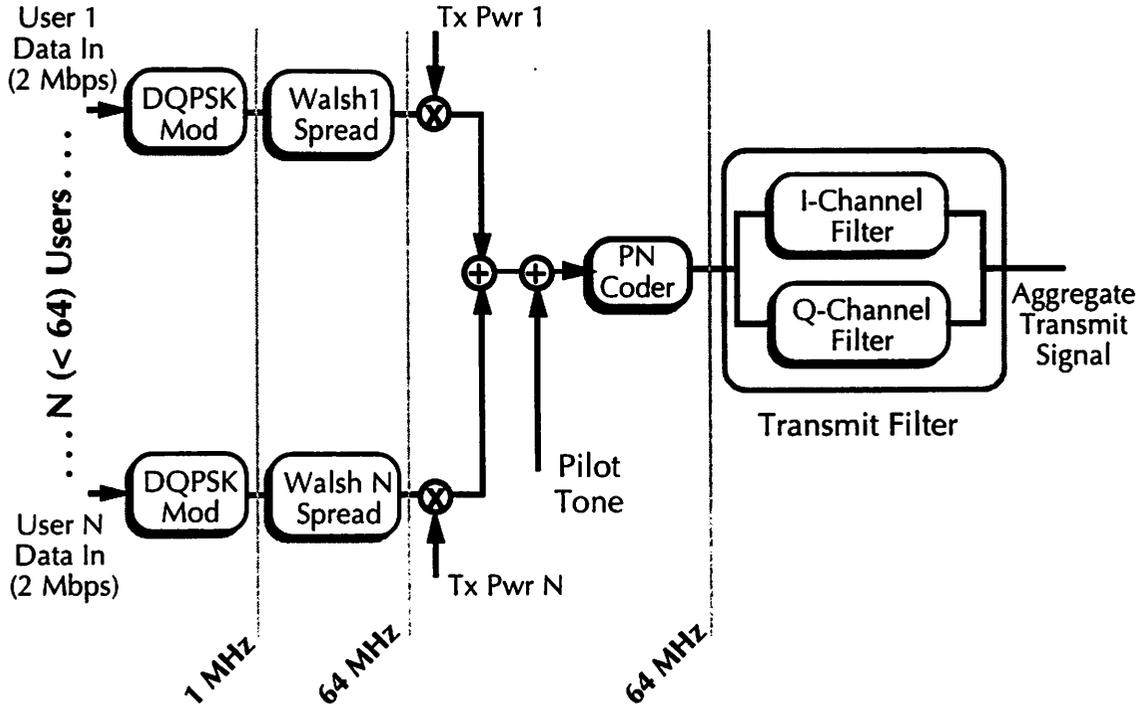


Figure 3.6: Dataflow diagram of the spread-spectrum signal synthesis. Signal bandwidth at each point is indicated.

bated, since it becomes difficult to determine when each users baseband bit-timing without employing large oversampling factors. Given that a base station is present in the downlink, capable of providing synchronization all of the user Walsh sequences by maintaining a pilot tone, these problems can be overcome.

3.2.2. PN Overlay Code, the Pilot Tone, and Cell-to-Cell Code Reuse

To restore autocorrelation performance to the overall code, a pseudonoise (PN) code is multiplied against the spread Walsh sequence on a chip-for-chip basis after all of the user Walsh sequences are combined, i.e., both the Walsh and PN sequences are time synchronous at 64 MHz (figure 3.6). The key property of PN sequences is that they appear to be “white” in nature: their power spectrum is flat, and their autocorrelation is nearly an impulse function (figure 3.7). Furthermore, the resulting spectrum of the concatenated Walsh/PN code is also white, to extremely good approximation. Intuitively, it is a similar situation as the case of sinusoidal modulation of white noise: the periodicity of the Walsh codes, spectrally concentrat-

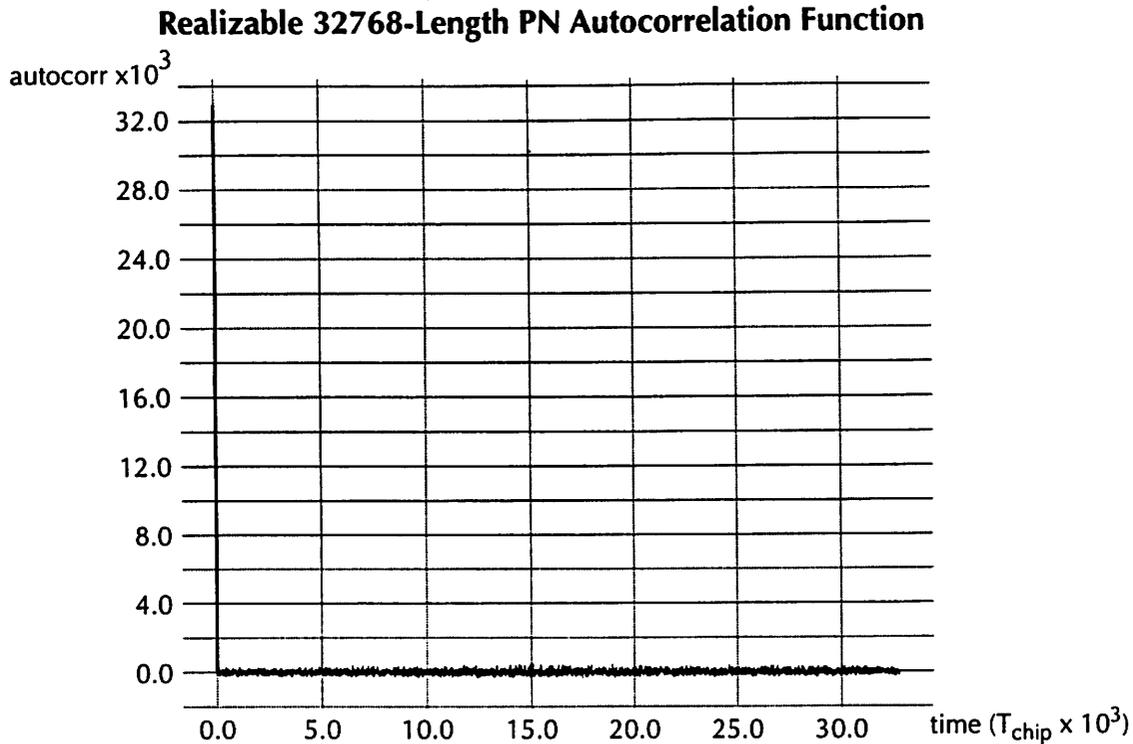


Figure 3.7: Realizable 32768-length PN sequences
(Implementation details of the PN code can be found in chapter 4)

ed, does not significantly affect the randomness of the spectrally flat PN code. Multiplication in the time domain by a periodic function results in a shift in frequency, and shifting a white spectrum results in the same. The result is an aggregate code that can simultaneously achieve 100% capacity, while still maintaining near-white spectral performance.

An important point here is that the length of the PN code and the length of the Walsh sequences need not be the same. The spreading factor is set by the Walsh sequence length: 64. On the other hand, the scrambling code can be significantly longer: as each user's Walsh sequence repeats from bit-to-bit, the scrambling sequence simply continues to progress. The reason that this is important is that longer PN sequences have better autocorrelation properties: the strength of the autocorrelation peak is proportional to the sequence length N . So long as the receiver has some means of synchronizing against the start point of the PN sequence, it can correctly "undo" the scrambling before correlating against the Walsh sequence to recover the user's data.

This synchronization can easily be achieved, given the PN sequence's autocorrelation properties, and becomes the key to the timing of the entire system. Since the transmit signal in each cell originates from a single point-of-origin, one channel can be expended as a timing tone, to aid in timing recovery in the receiver. The zeroth Walsh function – a constant DC - is designated as a pilot tone, with no data modulation. Examining the pilot tone signal path in figure 3.6, the contribution to the output will simply be the constant pilot tone multiplied by the PN sequence, which is effectively just the PN sequence. The receiver can simply search for the strong autocorrelation peak, and thus synchronize its local PN generator with that of the transmitter's. Since the PN sequence is being transmitted at the spread-spectrum chipping rate of 64 MHz, such synchronization also determines bit- and chip-level timing recovery for the receiver; if this pilot tone is continually tracked, any timing deviation can be monitored by chasing the correlation peak, using a delay-locked loop in the receiver. The benefit of a long PN sequence is clear: the effective processing gain on the pilot tone can be significantly larger than 18 dB, since correlation over intervals longer than 64 becomes possible. As the pilot tone is the key to timing for the entire receiver, immunity to multipath distortion and noise is paramount, and the larger effective processing gain greatly reduces the probability of false lock, or spoofing in the receiver.

The length of the PN sequence employed in this system is 32768, or 512 times longer than the Walsh sequence. Beyond the obvious answer of “as long as possible to achieve best code performance”, a number of factors contribute to this choice:

- Implementability:

The length of the PN sequence should be an integer multiple of 64, to give a direct relationship between the PN timing and the Walsh timing; furthermore, it should be a power of two, to employ shift-register techniques.

- Lock time:

Ostensibly, search-and-lock times for receiver synchronization are directly impacted by the length of the PN pilot tone. An infinitely long PN sequence takes an infinite amount of time to synchronize, as the receiver must step through each PN code phase to determine if lock has been achieved. One might conclude that

since lock is only measured at startup, a lock time on the order of seconds could be tolerated. However, if lock is lost, due to channel noise or a severe multipath fade, lock reestablishment time is also critical; a user cannot idly wait for seconds in the midst of a full-motion video stream, or a voice conversation. From the applications envisioned for the broadband downlink, a maximum worst-case lock time of 150 milliseconds was considered tolerable (five video frames at 30 fps).

- **Soft Handoff and Cell-to-Cell Code Reuse**

Up to this point, no mention has been made of cell-to-cell code performance, in terms of code reuse. While simple reallocation of Walsh codes is possible, it is far more efficient if the cells are simply designed to be orthogonal to one another. Each base station in the $K=7$ reuse pattern uses the same PN sequence; if a time offset is enforced between the autocorrelation peaks of each base station, the transmit streams from each base station become naturally orthogonal to one another, given the strong autocorrelation properties of the PN code. Furthermore, this time offset enables the mobile unit to detect nearby base stations: after the mobile synchronizes against the cell's pilot tone, another correlator is kept scanning for autocorrelation peaks. If peaks are detected at time offsets greater than the delay spread of the indoor channel, they must correspond to adjacent base stations. Furthermore, the detected energy in the offset peaks is an indication of how strong, or how close by, the other base stations are. This is the concept of soft handoff [Viter95]; the mobile make an assessment of the relative received power from adjacent base stations, and thus can assist the system in the decision to make a handoff. In order to achieve this soft handoff, the time offsets between base stations must be guaranteed, by some sort of timing tone on the backbone network. Due to network latency and delays, a synchronization of no better than 2 milliseconds can be achieved, assuming a delay through a single network bridge [Trew90]. At the 64 MHz chipping rate, it implies that a minimum of 4096 chip periods in the PN sequence is necessary to guarantee offset.

From the channel fading statistics, a conservative processing gain of 30 dB is needed for the pilot tone, implying a code length of at least 1024. However, 1024 does not provide enough time offset slots to accommodate the $K=7$ reuse pattern, given a 2 millisecond synchronization error between base stations, at 64 MHz a minimum code length of 28000 is needed; from rounding to the nearest power of two, a code length of 32768 is dictated.

The issue of lock time, however, remains. Assuming a single correlator in the receiver, correlating over the full 32768-long code across 32768 possible code phases, would require a worst-case time of 16 seconds to achieve lock, far too long to be useful. One obvious solution would be to use 100 parallel correlators to achieve the required lock time; however, this rather sorely violates any reasonable power and implementability constraint in the receiver. Some parallelism is quite possible - as will be discussed in chapter 8, four parallel correlators are naturally available in the receiver architecture for code search. The key to achieving lock time, however, is the realization that a sufficiently long subsequence of the PN sequence, statistically, should also be nearly white. Instead of correlating over the full 32768-long sequence, correlating over 1024-long subsequences is sufficient, provided that the spectral characteristics of the subsequences are indeed flat. As will be shown in chapter 4, the subsequence is nearly white (figure 4.4); the autocorrelation function exhibits a peak of 1024, as expected, with worst-case autocorrelation error performance 20 dB below the peak. Four parallel correlators, correlating over 1024-long subsequences across possible 32768 code phases, requires a worst-case time-to-lock of 128 milliseconds, which is well within acceptable parameters.

3.2.3. Channel Estimation and RAKE Reception

The pilot has even greater functionality than simply providing timing recovery for the mobile and mobile-assisted handoff: it can provide the channel estimates necessary to achieve RAKE reception. As described in chapter 2, direct-sequence spread-spectrum has the ability not only to reject multipath at the receiver, but to coherently combine detected energy from multipath arrivals to increase overall SNR performance with a RAKE combining algorithm. The key drawback from using a RAKE is the need to have accurate estimates of the impulse response

of the channel, since the relative phasing and amplitude of the arrivals are needed for coherence. In a pilot-tone assisted system, such estimation can be easily performed, without need for adaptive feedback techniques.

The first question that arises is the order, or size, of the RAKE. This is determined by the number of resolvable multipath arrivals, which is controlled by the spreading factor and the delay spread [Linn95], given as:

$$N_{\text{resolvable}} = 1 + \frac{T_{\text{del}}}{T_{\text{chip}}}$$

Intuitively, the time resolution of the spread-spectrum system can be no better than the chip-ping period T_{chip} ; the total number of resolvable multipath arrivals is simply the delay spread over T_{chip} . For the indoor broadband downlink, $T_{\text{chip}} = 16$ nsec, $T_{\text{del}} = 30$ nsec, implying that a third order RAKE is achievable.

To illustrate multipath resolution, a typical multipath profile for a channel is shown in figure 3.8, along with the transmitted and received waveform using just the pilot tone (no user Walsh data). As in chapter 2, this signal can be thought of as the linear combination of time-shifted and attenuated copies of the transmitted pilot tone $s(t)$, in particular:

$$r(t) = \sum_k a_{kT} s(t - kT)$$

where the $\{a_{nT}, n=0..2\}$ are the resolvable multipath coefficients at delay $t=nT$ relative to the first arrival. To resolve the n th multipath coefficient, a 1024-length correlation is performed using a copy of the pilot tone shifted by nT :

$$\begin{aligned} \text{MPath}(nT) &= \frac{1}{1024} \sum_{\tau = \tau - 1024T}^{\tau} s(\tau - nT) \sum_k a_{kT} s(\tau - kT) \\ &= a_{nT} + \frac{1}{1024} \sum_k a_k \text{err}((n - k)T) \end{aligned}$$

where $\text{err}(t)$ is the autocorrelation sidelobe noise at offset t (figure 4.4). To emphasize, this noise is due to the fact that we are taking correlations over a 1024-long subsequence of the 32768 PN code. Figure 3.9a also shows the resulting estimated output; the first and second estimates are quite good, whereas the third estimate is quite poor. This is to be expected; a_{2T} is quite small compared to the first two arrivals, and the estimate is being overwhelmed by intrinsic noise in the code. The 20 dB sidelobe performance of the PN subsequence implies that arrivals that are less than 0.1 in amplitude of the first arrival cannot be resolved. This performance is acceptable; from previous research [Teus94], multipath arrivals weaker than 10 dB relative to the primary arrival are insignificant in the RAKE coherent combining algorithm, given the quadratic nature of the SNR improvement. If improving the noise performance of the estimate is required, a simple moving-average filter is quite effective in minimizing the sidelobe noise, given that it is effectively white in nature. In figure 3.9b, a 4-tap MA filter has been applied to the multipath estimates; SNR has been globally improved by 6-10 dB, even estimating the last multipath arrival. It should be mentioned the conditions that the simulation was executed: no intercell interference was present, and a fixed frequency offset between the transmit and receive carriers was assumed (no carrier recovery is performed in the hardware, for power and complexity considerations, thus leading to a significant worst-case carrier offset). The estimation strategy is that described for the receiver DSP in chapter 7, using relative subcorrelations to compensate for the carrier offset; its output is the non-MA filtered estimate of figure 3.9.

Lastly, the performance benefit of employing a RAKE is illustrated in figure 3.10. Using the Ptolemy simulator, the two channels from figure 2.5 have been simulated for BER as a function of the number of users in the channel, with and without a RAKE combiner. Clearly, as the number of users increases, the multipath interference becomes more and more significant, as an individual user's energy becomes smaller proportional to the multipath reflections. Furthermore, in the case of a severe in-band null, the BER even for small numbers of users becomes unusable, without employing a RAKE combiner. The center curve is the case of the deep-fade channel with a RAKE combiner – for moderate numbers of users in the cell (less than 20), the performance has been dramatically increased, with a decrease in BER by over 4 orders of magnitude. As the number of users increases beyond 20, the RAKE becomes less

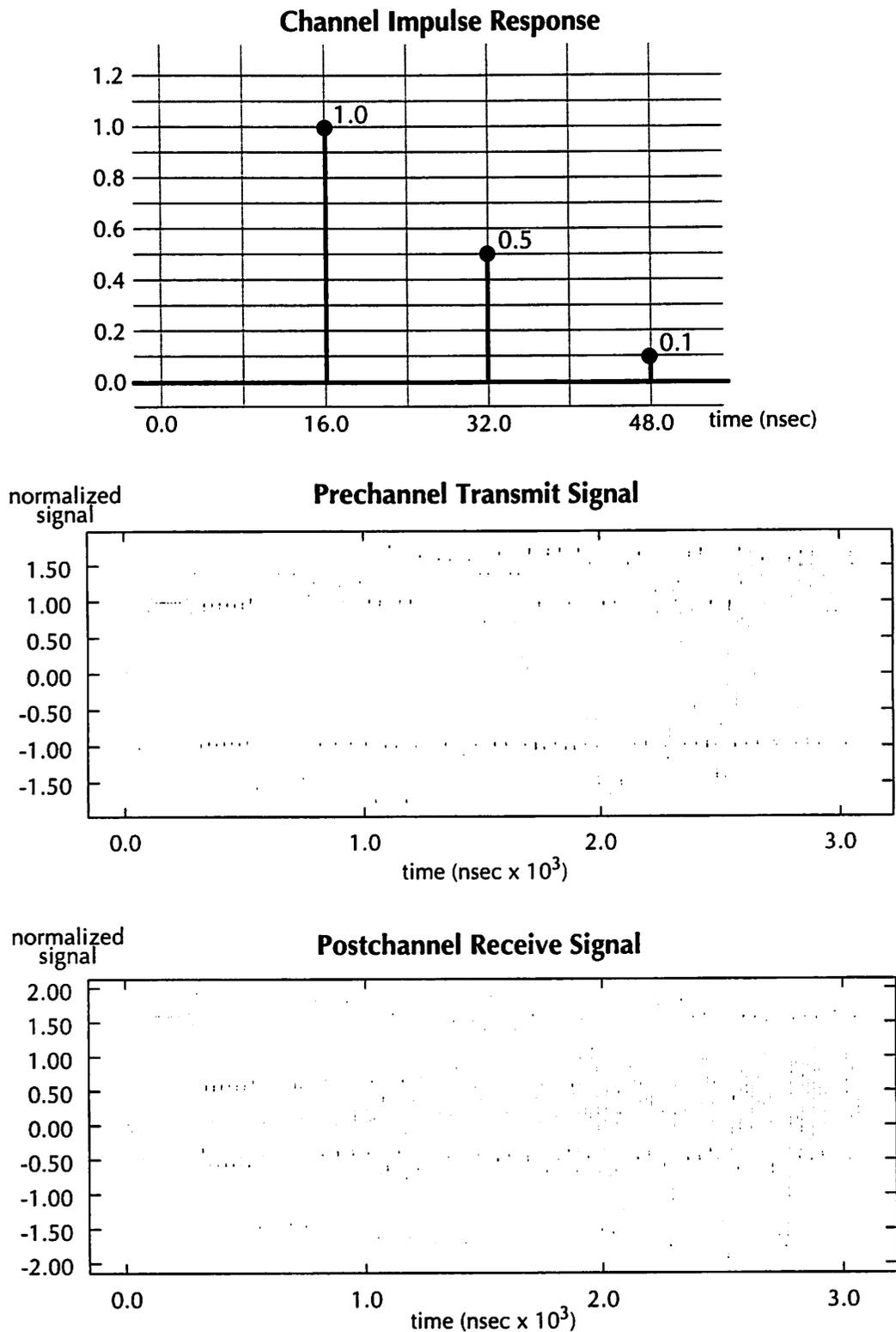


Figure 3.8: Sample multipath channel for channel estimation. ISI is fairly severe in this case.

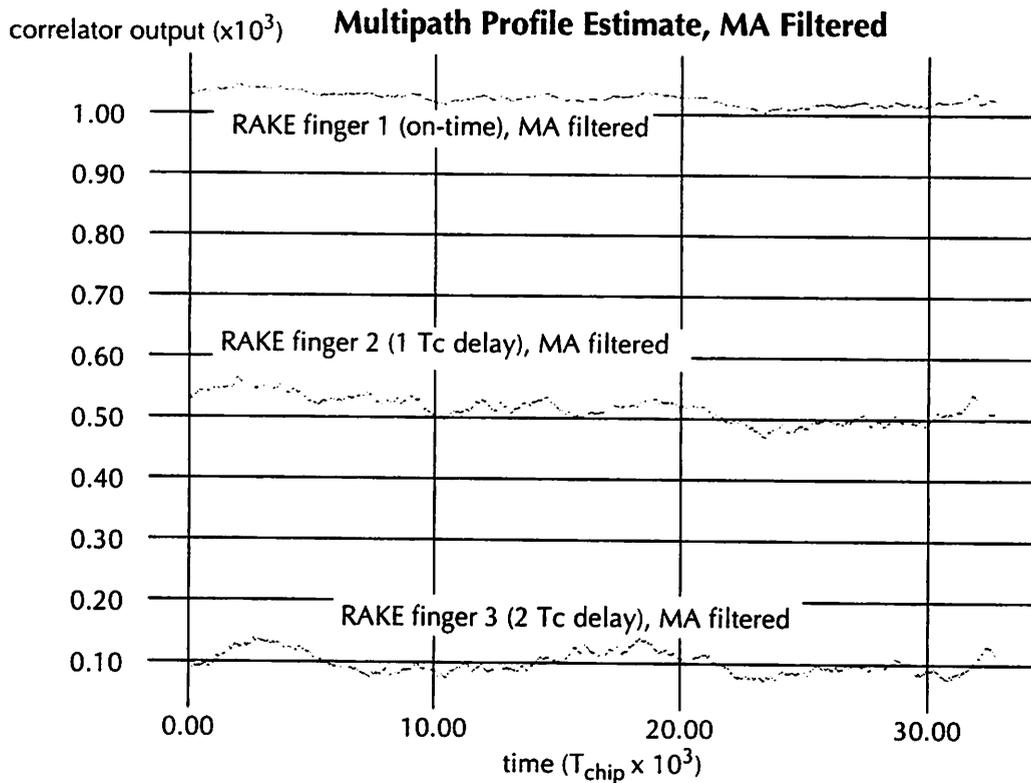
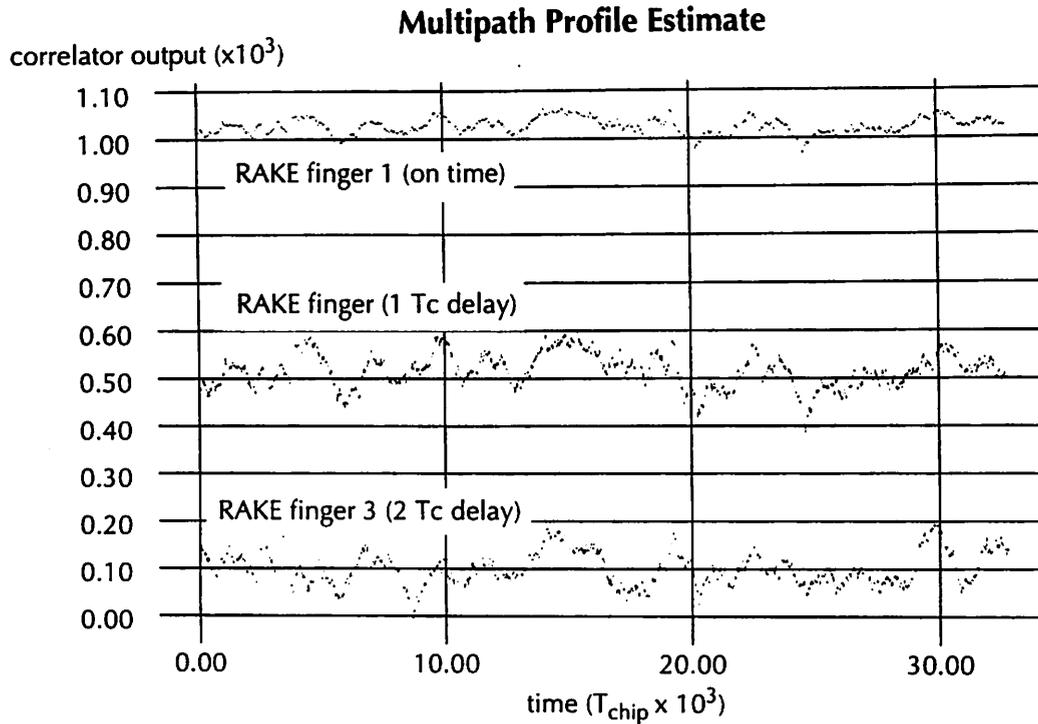


Figure 3.9: Multipath estimator output: (top) direct output; (bottom) MA filtered 1024-length subcorrelations, 32768-length sequence, Nominal outputs are 1024, 512, 102.4 MA filtering yields -6-10 dB SNR improvement.

and less effective, owing to increased noise in the channel estimate itself and loss of coherency. For the typical deployment conditions as described in section 2.1, 12 users will be serviced by each base station; a RAKE architecture will be critical in achieving channel capacity and performance.

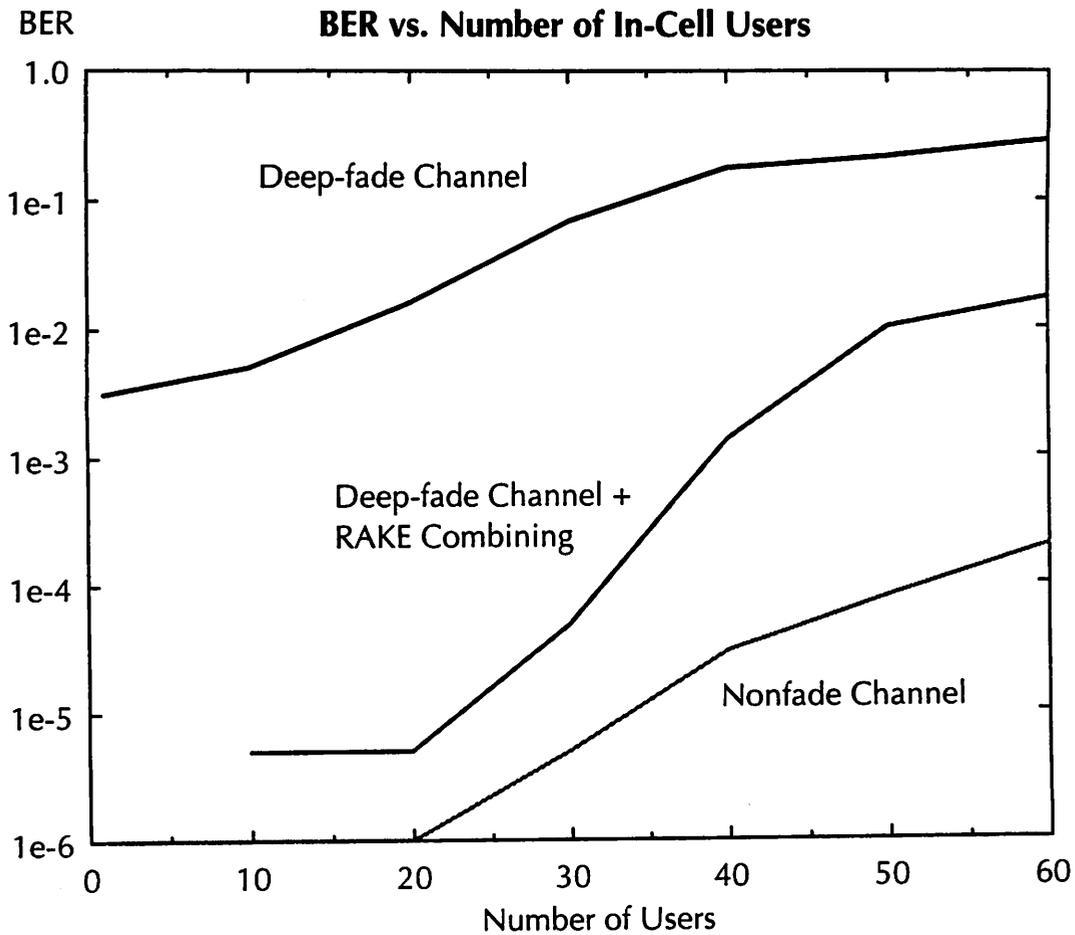


Figure 3.10: Plot of simulated BER versus number of in-cell users, no intercell interference, for the channel profiles shown in figure 2.4, including RAKE combining. It is clear that with no combining, under worst-case fading, self-interference severely limits the achievable BER performance for even 1 user.

3.3. An Open-Ended Architecture

As it stands, the system design and specification of the broadband downlink has achieved its stated goal: a system capable of providing high-speed, indoor picocellular data services to large number of users. However, a number of open questions have been raised to improve performance and capacity, to be answered in future research:

- A RAKE receiver is only truly effective with small numbers of users, when a good channel estimate can be found. Is there a better combining strategy than a simple RAKE? Can the estimates be used in some type of adaptive equalizer architecture that is superior in performance to a RAKE?
- Spread-spectrum easily accommodates users at the same data rate; can it be extended to easily support multiple data rates? The research behind the IS-95 digital cellular standard has shown that multiple data rates can be accommodated by modulating user power as a function of data rate; at this point, 6 bits of linear power control is provided in the specification, although this may be excessive. What is the optimal power-control algorithm as a function of the application?

However, the single most important question that remains is simply “Can it be built?” Can an implementation be realized in silicon CMOS, achieving the performance constraints of 64 MHz while remaining small enough consuming low enough power to be portable? It is to this question that the remainder of this thesis is dedicated.

3.4. Appendix: Ptolemy Simulation

In this chapter, system verification has been almost exclusively demonstrated using the U.C. Berkeley Ptolemy simulator [Buck93]. As a dataflow simulator, Ptolemy is extremely flexible, allowing the user to add new simulation blocks, or “stars” at will, as well as allowing for different simulation engines, or “domains”, to be employed. The two domains employed were the SDF (synchronous dataflow) domain, and the CGC (C-code generation) domain. The first employs Ptolemy itself as the simulation engine; simulations are run within Ptolemy itself. The second employs Ptolemy more as a CAD/software engineering tool; the block diagram is synthesized into an equivalent C-language source file, and then executed “outside” of Ptolemy. Both were extremely useful:

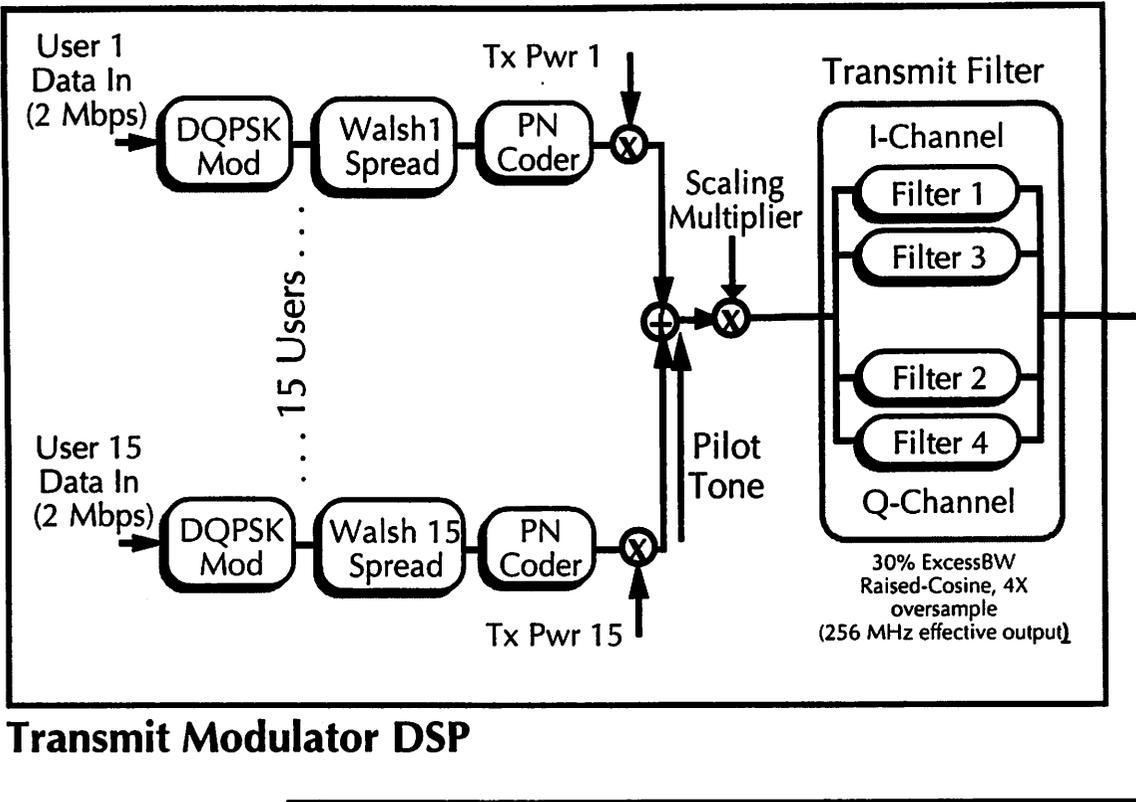
- SDF is best for “quick turn” simulations, such as simple filtering, FFT, etc., where the overhead of generating full C source code plus compilation time is not justified. For example, figure 3.3 was generated this way; multiple runs could be quickly generated with varying parameters, to determine the transmit filter complexity.
- CGC is best suited for long Monte-Carlo style simulation runs, such as those involving BER (figure 3.10), or multipath estimation experiments (figure 3.9). Compiled C-code is very efficient; large data structures need not be passed (as in SDF simulation), and the sheer overhead of the user interface need not be incurred.

It should be mentioned that, over the course of this project, the simulation efficiency of SDF has been improved dramatically by the authors of Ptolemy. It is, of course, up to the judgment of the engineer to decide which domain is best suited for what simulation task. What follows is a complete listing of available stars that were developed for simulating the spread-spectrum downlink; they can be acquired by contacting the author, or on the cd-rom that contains all of the design files for this thesis.

4 Transmit Architecture and The Baseband Modulator Chip

To begin the discussion of the system implementation, the base station transmitter will be described in this chapter. The key functional blocks that need to be implemented are the spread-spectrum digital signal processing (data spreading, PN scrambling, etc.), baseband pulse shaping, D/A conversion, filtering, and frequency upconversion to the 1.088 GHz carrier frequency. The hardware partitioning consists of a custom baseband modulator integrated circuit [Peroul94], along with a semicustom board to implement the analog A/D and upconversion circuitry [Yee96]. A board level solution suffices for the transmitter – since this is in the base station, size and power are not nearly as critical as in the receiver, nor is noise performance as extreme an issue.

In figure 4.1, the transmit architecture for the base station is shown. Not surprisingly, the baseband modulator circuit mimics the dataflow block diagram of figure 3.7, with a few subtle differences. The PN modulator, which provides autocorrelation and spectral whitening, has been shifted before the user combining adder. The pulse-shaping filter, which controls the transmit bandwidth, filter is implemented as 4-parallel subfilters instead of a single high-speed one. Choice of number representation is also crucial; the representation changes as the data flows through the chip, from bit input to pulse-shaped output. Each modulator IC only supports 15 users, which is the most common deployment case and minimizes die size; multiple modulators can be ganged in parallel to support more users per base station. A two-stage frequency upconversion is used, one in the digital domain, and one in the analog domain. The impact on hardware complexity of these decisions and others will become apparent throughout; each block will be discussed in turn, in terms of choice of architecture and implementation.



Transmit Modulator DSP

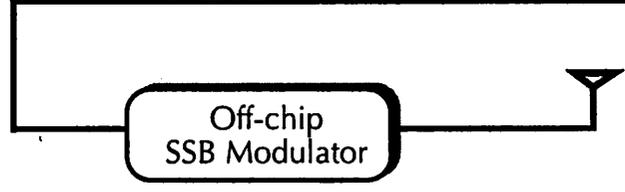


Figure 4.1: Base station transmitter architecture

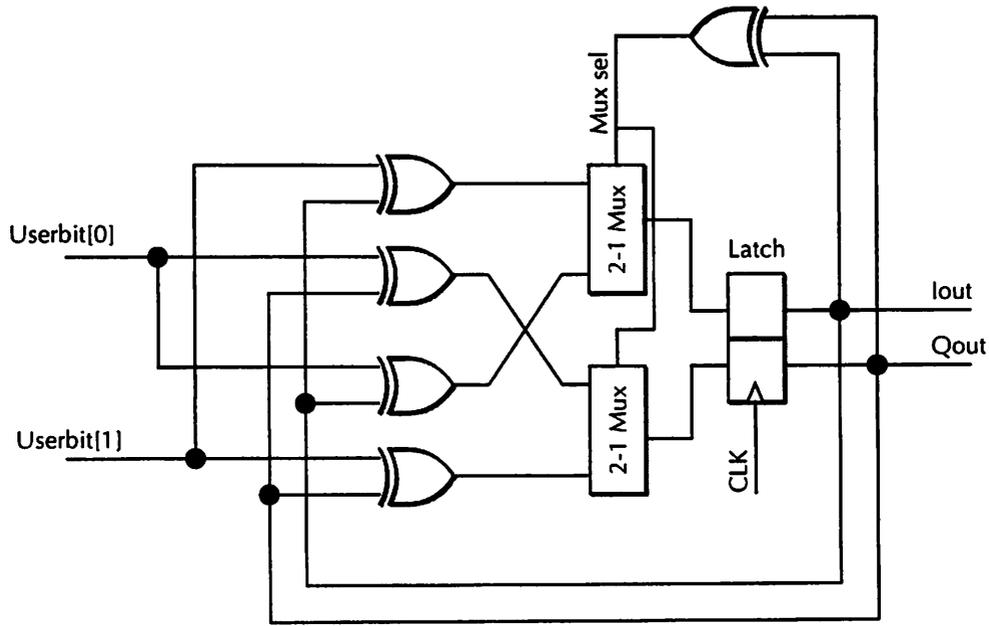
4.1. The Baseband Transmit Modulator

4.1.1. DQPSK coder

As the first block in the modulator, the DQPSK coder implements the bit-to-baud encoding shown in figure 3.1. Number representation here is extremely important: the possible values for either the in-phase or quadrature symbols are bi-state (+/- 1), and it is natural to map these two values into a single digital “0” or “1”. The representation chosen (which is standard in the communications literature) is:

- +1 maps into digital “0”
- 1 maps into digital “1”

It is akin to a “1-bit” signed magnitude. The ramifications of this will become apparent in the PN and Walsh modulator blocks: multiplies under 1-bit arithmetic can be isomorphically translated into a digital XOR (e.g., $-1 \cdot -1 = 1$, or “1” XOR “1” = “0”). Under this representation, the DQPSK coder implementation is shown in figure 4.2.



For inputs: If the input bit sequence is: 00011011,
then the inputs are presented as two-bit words: {00},{01},{10},{11}
leftmost bit is Userbit[1].

Figure 4.2: DQPSK coder schematic

4.1.2. Walsh Modulator

The encoded symbols are then spread, from the 1 Mbaud rate up to the full 64 MHz chipping rate. Since 64 individual Walsh codes need to be generated, a 6b-wide datapath is dictated; the architecture chosen is from [Beau84], which is based on Gray-code generators (figure 4.3). Each bit in the Walsh sequence, which is clocked at 64 MHz, is XOR’ed with the complex-valued user data symbol, which is clocked at 1 MHz, to effect spectrum spreading.

sequence of a 16-bit MLSR sequence would also seem to suffice; again, such a subsequence should be “nearly” white. Quantifying the term “nearly”, in terms of impact on system performance, becomes a crucial issue.

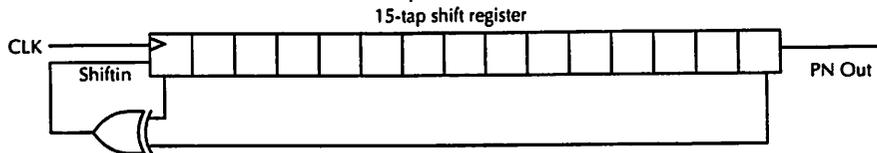
The implementations of these two possible choices are shown in figure 4.4; for comparison, the implementation of a simple 15-bit MLSR sequence generator is also shown. One of the key properties of MLSR generators is exploited here. Since the shift register is n -bits wide, there are 2^n possible words; the MLSR feedback guarantees that each word will occur in the sequence once and only once, with the exception of the all-zero word. In particular, the implementation keys off the “01111....111” word; this will occur one clock cycle before the all-one’s word (the extra delay is needed for control circuitry). In figure 4.4c, an extra clock cycle is added after the all-one’s word, to inject the extra bit; in figure 4.4b, the 32768-long subsequence is forced to terminate, by parallel loading a seed into the register. The seed is chosen such that, 32768 cycles later, the all-one’s word will occur once again.

In figure 4.5, the autocorrelation function of each of the generated sequences is shown. Again, for comparison, the autocorrelation function of the true 32767-long MLSR sequence is shown. The term “nearly” is now clear; compared to the peak autocorrelation value, the largest off-peak correlation is less than 512, out of 32768. Given that this is 36 dB down from the peak, from an SNR standpoint utilizing either the subsequence approach or the additional bit approach has negligible impact. Figure 4.4b was implemented; it has the property that the clock to the PN generator does not need to be gated, eliminating problems with clock skew and glitching.

Lastly, the performance of the code under 1024-long subcorrelations needs to be examined. By “subcorrelation,” it is meant that partial sums of the full correlation are taken; since subsequences are used, two parameters are necessary to specify the subcorrelation, i.e.:

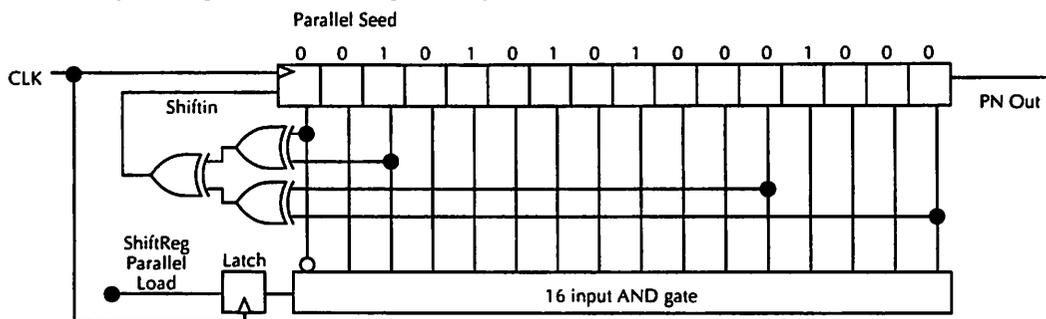
$$\text{subcorr}(\tau, \tau) = \sum_{k=0}^{1023} p_n((\tau - k) \bmod 1024) p_n((\tau - k) \bmod 1024)$$

Basic 15-tap shift register PN sequence generator



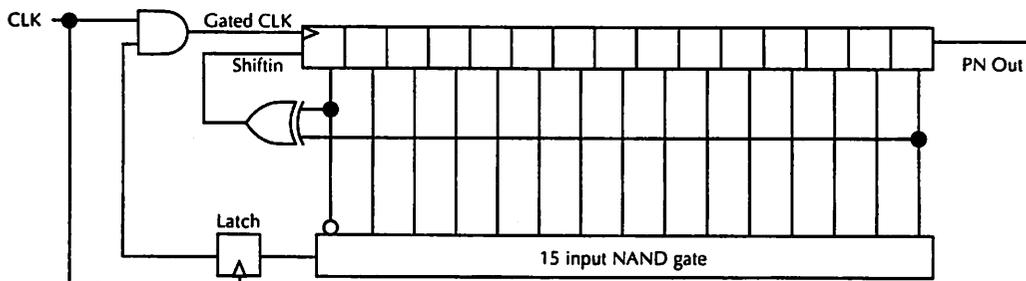
(a)

16-tap shift register base, 2^{15} -length subsequence



(b)

15-tap shift register base, additional bit added



(c)

Figure 4.4: PN Sequence Generators

- (a) True 15-tap, length 32767 MLSR schematic
 - (b) 16-tap, length 32768 subsequence schematic
 - (c) 15-tap, length 32768 schematic
- Figure 4.4b was implemented.

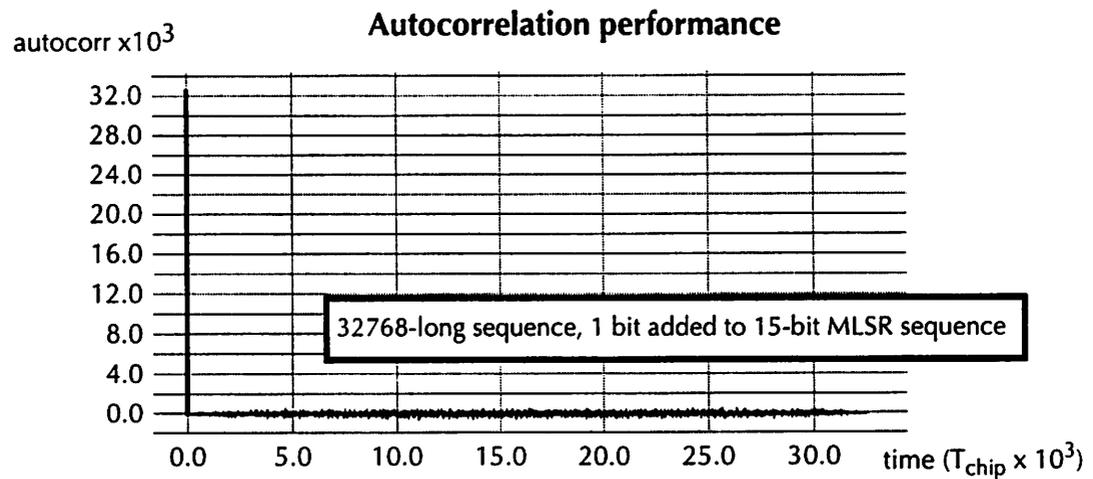
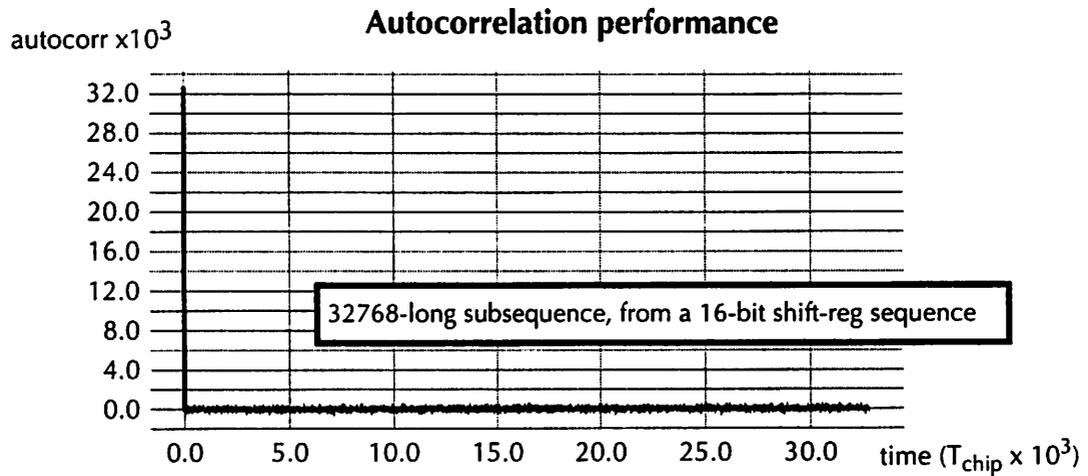
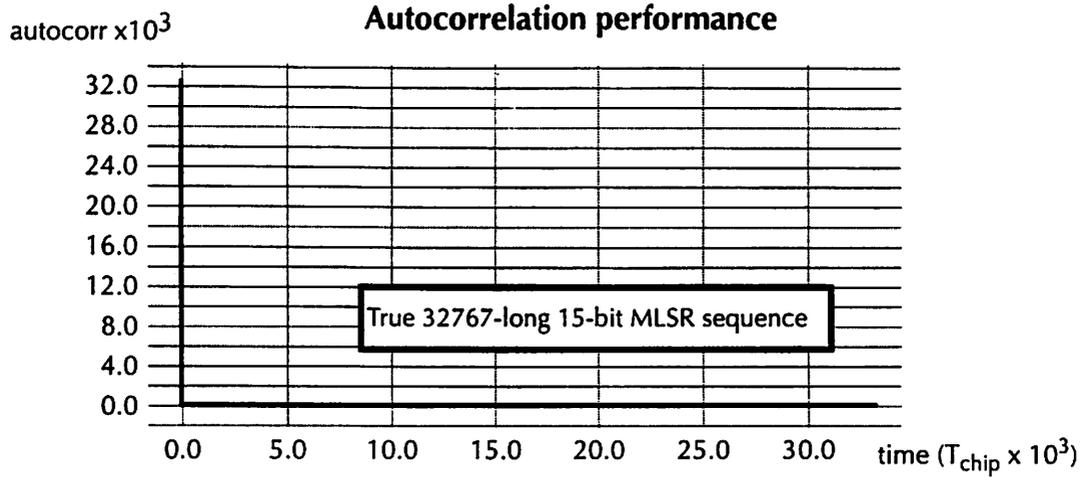


Figure 4.5: Autocorrelation performance curves of the architectures shown in figure 4.4.

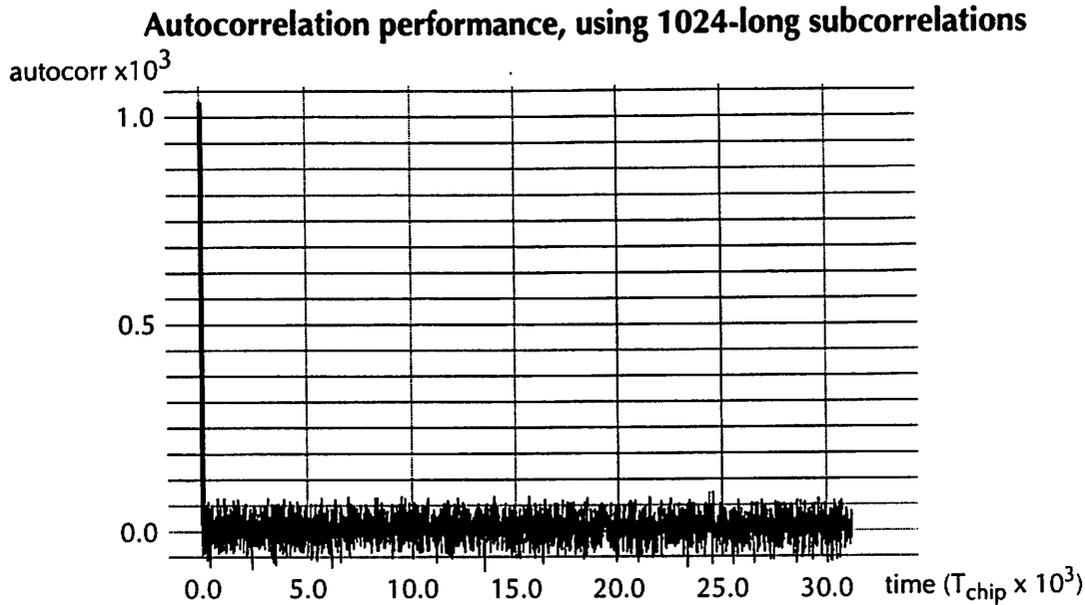


Figure 4.6: Autocorrelation performance under restriction to length 1024 subcorrelations (taken for $\tau=0$; the curve is similar for τ nonzero).

In figure 4.6, the plot of 1024-long subcorrelations of the full 32768-long sequence is shown for $\tau=0$; as expected, when the subsequences are synchronized ($t=\tau=0$), a peak value of 1024 is seen. However, the off-peak performance is crucial; when the receiver is searching for the pilot tone to indicate lock, tracking the peak for timing recovery, or measuring the multipath profile, it is this off-peak performance that determines the performance of these functions. From this, the SNR is approximately 20 dB, translating to 3 bits of dynamic range. The choice of 1024 was driven by the need to minimize search-and-lock times, plus the need to minimize hardware complexity in the receiver; under simulation with the multipath channel, 20 dB was the minimum value necessary to reliably achieve lock and timing recovery.

If greater SNR is needed, figure 4.7 displays the plot of SNR as a function of subcorrelation length – surprisingly, the SNR drops quite slowly, 6 dB for every doubling in the subcorrelation length. This translates to a doubling in the peak amplitude, which is expected; however, the off-peak noise remains fairly constant in amplitude past a length of 256. Hence, for every 6 dB of SNR performance in the autocorrelation code, a doubling of the receiver lock/acqui-

sition time is necessary, along with increasing the dynamic range by 1 bit. The moving average filter technique described in chapter 3 effectively increases the subcorrelation length by a factor equal to the number of taps.

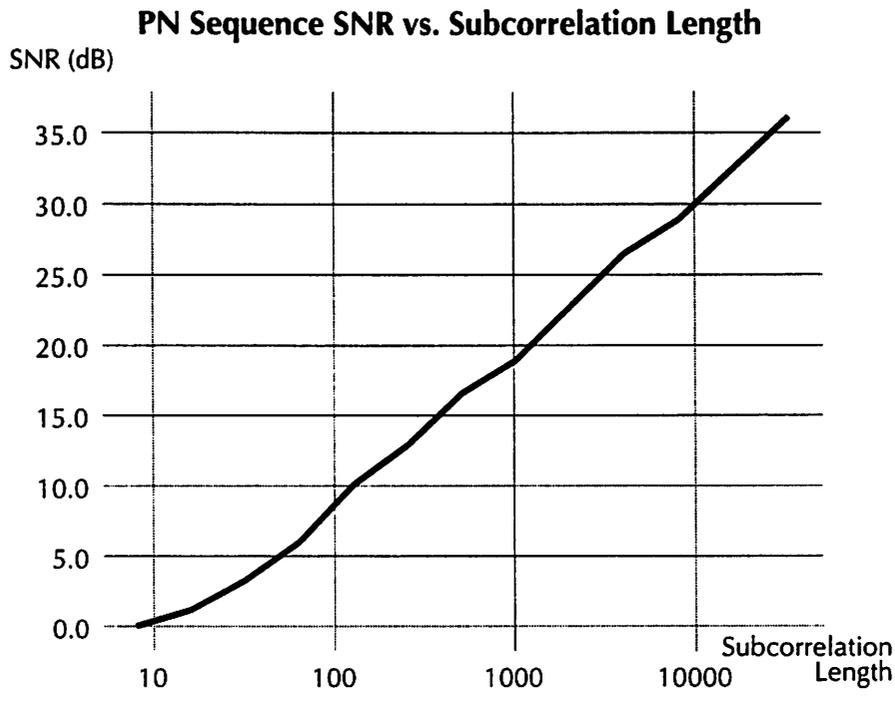


Figure 4.7: Plot of SNR due to off-peak nonzero autocorrelation, as a function of subcorrelation length.

4.1.4. Per-User Power Control

Lastly, the per-user power control needs to be considered. At this point, the choice of dynamic range is preliminary; research is currently underway to determine optimal power control algorithms as a function of the application, desired BER, number of users, etc.; 6 bits of dynamic range (36 dB) is allowed, modulating the user's power in a linear fashion.

Number representation here again becomes important. Given that the subsequent stage is the combiner, where all of the user transmit signals are linearly added, a two's complement number representation is desirable. Although the 1-bit signed magnitude representation has served well up to this point, the stream now needs to be translated into an equivalent 6-bit two's

complement representation. Instead of implementing a true multiplier, a simple multiplexer can be used: both the power control value and its negative are presented from registers to the input of the muxes, and the incoming 1-bit data on the I/Q lines selects the positive or negative value.

From this, it is clear why the PN modulation was moved before the combining and per-user power control: had it been left post-combiner, as in figure 3.7, the ± 1 multiply from the PN code would have required the two's complement negation of a 10-bit word at 64 MHz. By simply moving the PN modulator pre-power control, a simple XOR suffices.

4.1.5. The Pulse Shaping Filter

At this point, all 15 user data streams and a 6-bit pilot tone are linearly summed; the resulting 10-bit wide datastream is the baseband representation of the transmit signal. The remaining signal processing consists entirely of bandlimiting this signal, D/A conversion, and frequency translation into the 1 GHz transmit band. As described in section 3.1.2, a baseband digital bandlimiting filter is desirable to implement a 30%-excess bandwidth raised cosine pulse shape, for zero ISI at the expense of as little excess bandwidth as possible. Since the incoming data is complex-valued (in-phase and quadrature), two real-valued filters are needed to implement the equivalent complex-valued filter. Given the rather specialized shape of the desired transmit pulse, an IIR implementation becomes problematical; instead, a straightforward transversal FIR topology is used. As the impulse response will necessarily be time-truncated, a minimum out-of-band rejection of 40 dB is placed on the filter, which is typical of FCC emission-masking constraints and sets the 99% power containment bandwidth of the transmit signal. The out-of-band reject specification determines the number of taps needed in the filter structure, as well as truncation performance of the taps themselves.

In order to perform pulse-shaping on the 64 MHz stream, oversampling is needed to effectively interpolate the shape between consecutive data points. With the already-high data rate, as small of an oversampling factor is needed; in this case, 4x, with an effective filter output rate of 256 MHz. Under this oversampling constraint, from simulation the total number of

taps needed in the filter is 37 (figure 3.4); after Booth-encoding (canonical signed digit representation) of the tap coefficients and truncation, the resulting filter transfer function is shown (figure 4.8).

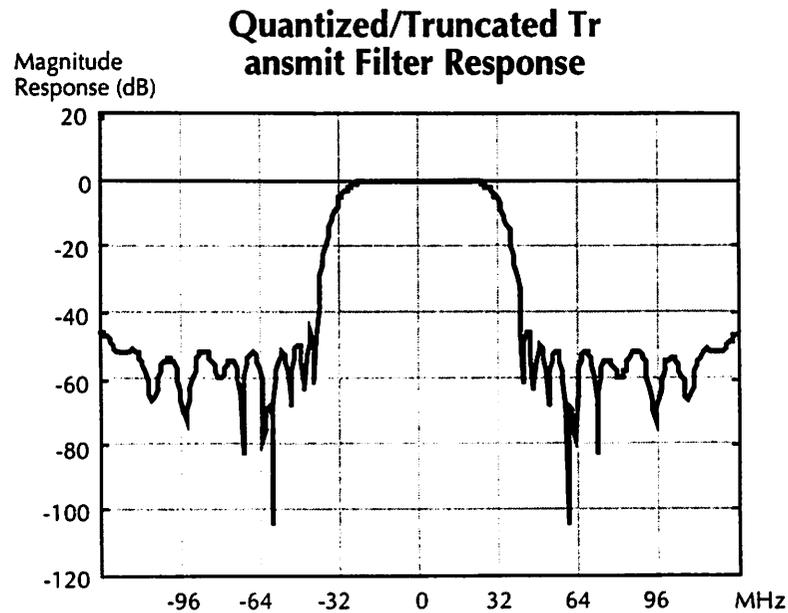


Figure 4.8: Filter response after truncation and tap quantization

One critical issue here is the dynamic range in the filter; given that digital-to-analog conversion needs to be done at the 256 MHz rate, a high-order DAC is out of the question; the highest-performance commercially available ECL DAC's can provide 10 bits of resolution at these speeds, and no more. Given that the input already has 10 bits of total dynamic range (15 users plus the pilot tone, with 6 bits of power control each), there is an obvious problem here: if all users are present, the full resolution of power control is not available, given truncation noise in the filter. Clearly, the worst-case situation occurs with 14 users and the pilot set to maximum power, and one user at the minimum power control value; the pulse shape of the single user will be nearly at the noise floor of the filter itself. In figure 4.9, this situation is illustrated graphically: since the data streams are orthogonal and uncorrelated, the total signal power is simply the aggregate sum of the individual signal powers. If the filter were to have sufficient truncation noise performance to maintain the 40 dB out-of-band pulse shape for the single user, 100 dB of dynamic range would be required, or over 16 bits. From simula-

tion, if some performance loss was allowed in terms of truncation noise injection, the single user could be recovered if the truncation noise floor were limited to 72 dB, or approximately 12 bits.

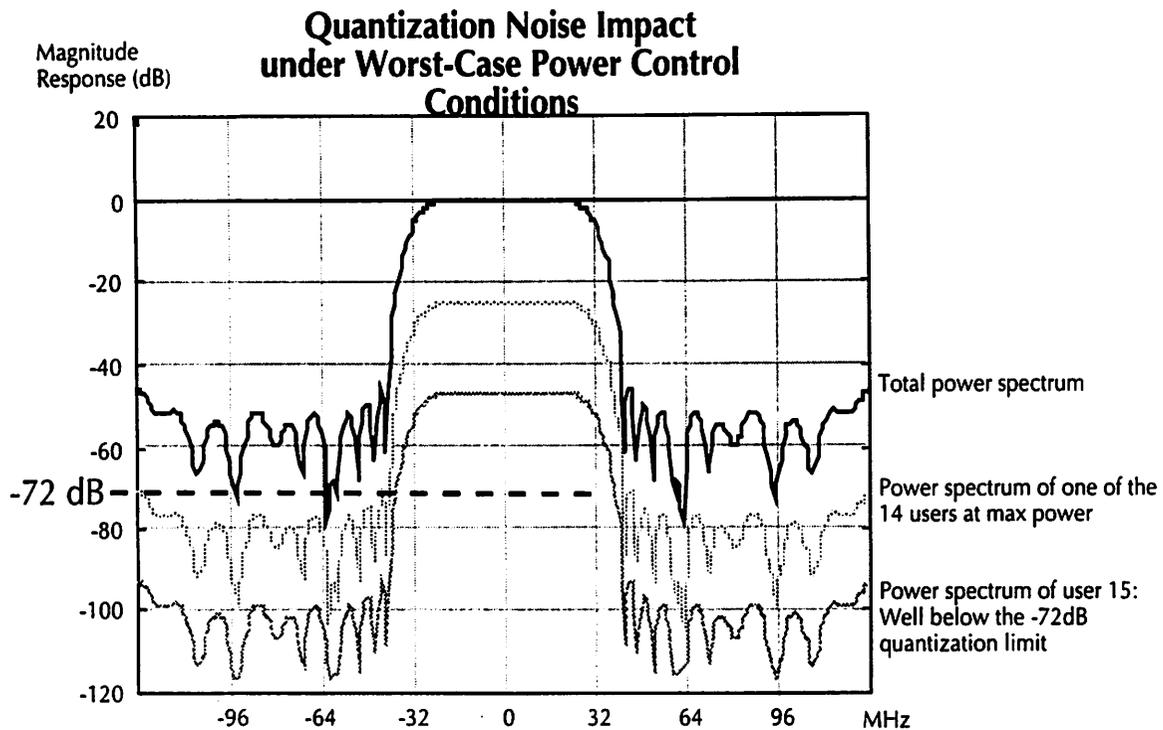


Figure 4.9: Filter quantization noise effects at maximum power control dynamic range (14 users with maximum power, one user with minimum power). The -72 dB mark indicates the minimum quantization performance required to recover user 15.

The ramifications of this, coupled with the 10 bit limitation in the DAC, imply that the least-significant three bits of the input signal to the filter are quantized away into the filter noise floor and are ineffective; if a user is set up such that its signal power is below -42 dB relative to the total signal power, that user will be subject to excessive BER at the receiver. To control this dynamic range issue, a “multiplier” is inserted before the filter as a form of aggregate transmit power control. The multiplier is implemented as a 4-bit shifter; although it is

only provides power control in fairly crude 6 dB steps, it achieves its desired function. The multiplier shift control and error condition check in the base station should be set according to the following equations:

$$\text{Shift} = \text{floor}\left(10 - \log_2\left(\sum_i P_i\right)\right)$$

$$\text{Error} = \bigcap_i \left(20 \log_{10}\left(\frac{P_i}{\sum_i P_i}\right)\right)$$

where the P_i 's are the users' power control settings (as 6-bit integers) and P_0 is the power in the pilot tone.

4.1.6. Filter Implementation

At this point, the physical implementation of the 256 MHz FIR pulse shaping filter needs to be discussed. Given the technology available (a 0.8 micron standard digital process), achieving these speeds can only be done with careful consideration of the architecture to be employed. A basic transversal structure, as shown in figure 4.10a, simply cannot achieve a 4 nsec operational cycle time, especially given that all multiplies and accumulates are done with data values of at least 10 bits. Parallelism and pipelining need to be taken into account.

4.1.6.1. Upsampling

The immediate optimization that occurs is consideration of the upsampling nature of the filter. The data samples are incident on the filter at 64 MHz; zero-interpolation is the natural means of upsampling the signal. The data stream should be thought of as a train of impulses being fed into a filter with the desired impulse response; the zero-interpolation simply describes the time in between each successive impulse. Thus, three of every four delay registers in the transversal filter structure of figure 4.10 will have zero samples at any given sample point, and their contribution to the output of the filter need not be computed. The implication is that the structure can be parallelized into four subfilters, running at 64 MHz, and their outputs interleaved with a multiplexer back to the full 256 MHz rate. It is important to note that the data is brought off-chip at 64 MHz, with the interleaving multiplexer implemented

using off-the-shelf ECL components. There are two reasons for this: first, driving the output pad ring of the chip at 256 MHz is difficult without incurring tremendous substrate injection and ringing, resulting in questionable signal integrity; second, bringing off the outputs in parallel allows more flexibility in the upconversion architecture, as described in section 4.2.

4.1.6.2. Tap Symmetry

From the coefficient assignment plot of figure 4.10, several more optimizations become readily apparent from the structure of the taps themselves. For example, the Nyquist criterion reduces filter 1 into a simple delay; all of the other tap coefficients are zero, reflective of the zero-ISI nature. Furthermore, the pulse shape is symmetric, implying that filter 3 is perfectly symmetric, and filters 2 and 4 are the time-reversed images of one another. An efficient implementation of symmetric filters can easily be achieved. As an example, in figure 4.11, a 6-tap symmetric filter is shown, along with two architectural minimizations. In the structure, there are 6 taps but only 3 different coefficients: A1, A2, and A3. Originally, when calculating the current output it implements the structure of figure 4.11a - a simple transversal filter.

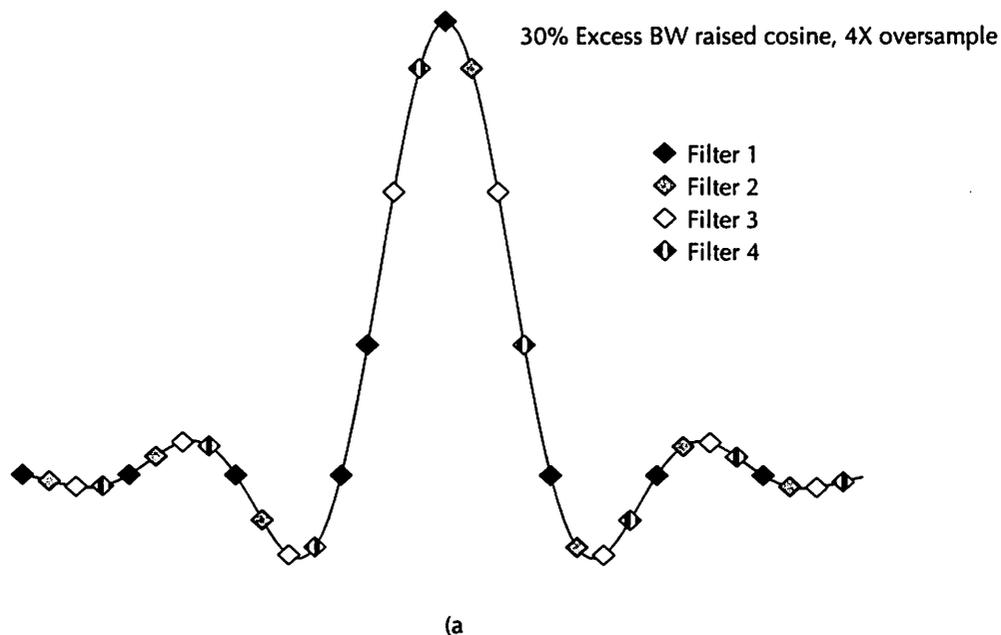


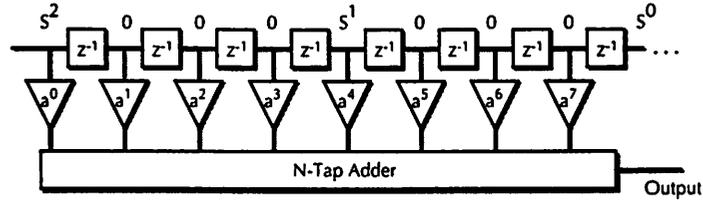
Figure 4.11: 4-Parallel Implementation of the Transmit Filter
 (c) Tap assignments (b) Parallel architecture (c) System-level schematic of filter structure

Output of Walsh+Pilot Tone Combiner at 64 MHz

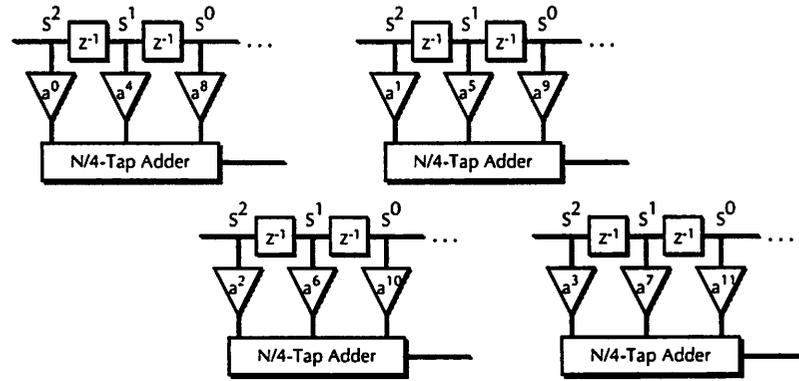
$s^0, s^1, s^2, s^3, s^4, s^5$
 4X Oversampled Output at 256 MHz

$s^0, 0, 0, 0, s^1, 0, 0, 0, s^2, 0, 0, 0, s^3, 0, 0, 0, s^4$

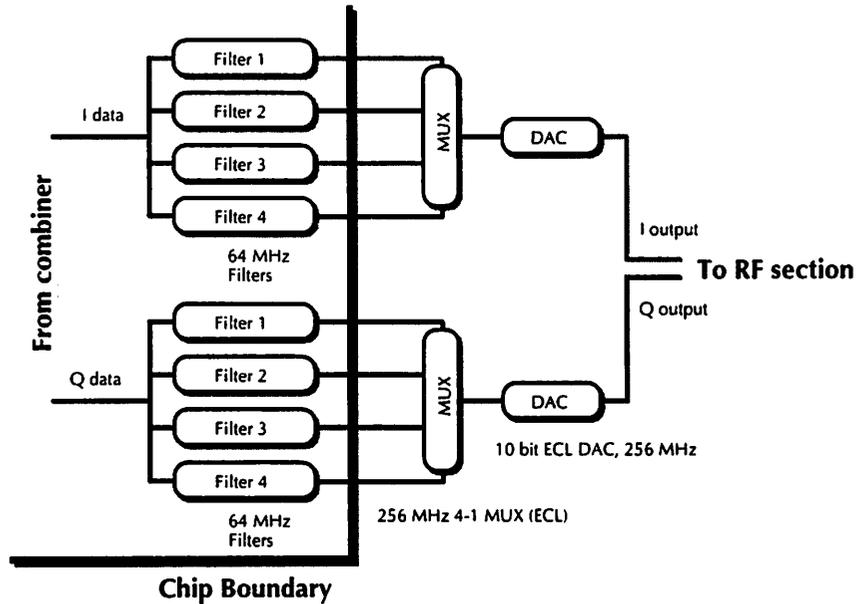
Filter Operation at 256 MHz



4-Parallel Filter Operation at 64 MHz



(b)



(c)

Figure 4.10: Continued.

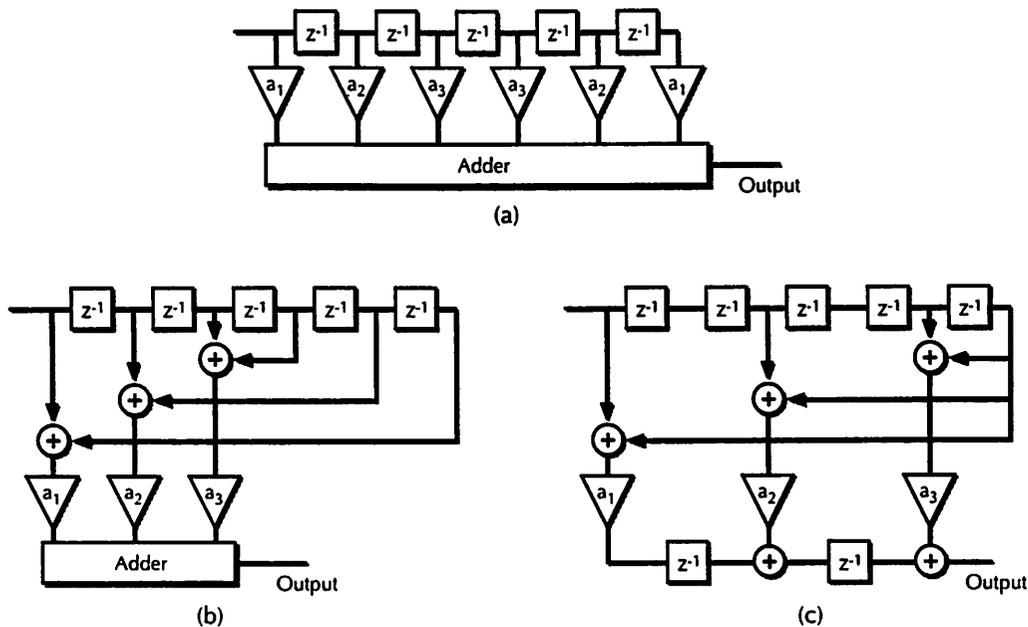


Figure 4.11: Symmetric Filter Minimization (filter 3)
 (a) Original transversal structure (b) Shared multipliers (c) Pipelined output adder

By grouping the multiplications with the same tap weights, a much simpler implementation is produced. At the expense of adding an adder in front of every multiplier, the structure can be implemented using half as many taps (figure 4.11b).

Lastly, the addition at the output of the filter needs to be pipelined, since it is impossible to combine all of the multiplier outputs in a single cycle. The final structure (which was implemented), is shown in figure 4.11c. The pipelining stages simplify the structure because they also force an extra delay for the input data in each tap. In the figure 4.11b, $x(n-5)$, $x(n-4)$, and $x(n-3)$ are fed back; from a layout and routing standpoint, this presents several difficulties, especially given the wide datapaths in question. In the final pipelined structure, only a single signal has to be routed back.

This structure can be verified by comparing figures 4.11b and 4.11c. In figure 4.11b, the input to the first multiplier is the sum of $x(n)$ and $x(n-5)$; i.e., it must add the current sample along with the sample that appeared five time periods ago. The second tap must add $x(n-1)$

and $x(n-4)$, combining samples with a delay of 3 between them. It can be seen that 4.11c is equivalent; the first tap adds $x(n)$ and $x(n-5)$, maintaining a time delta of 5 samples. Similarly, the second tap combines $x(n-2)$ and $x(n-5)$, maintaining a time delay of 3 samples.

Lastly, it should be mentioned that for filters 2 and 4, no similar optimization can be made, as the tap coefficients do not exhibit similar symmetry, beyond the fact that they are time-reflected versions of one another. However, these filter structures must still be pipelined, especially with regard to the output addition. Their structure is shown in figure 4.12.

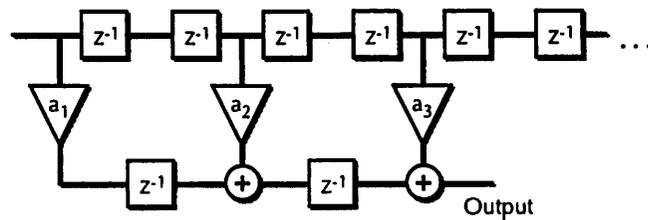


Figure 4.12: Pipelined output adder filter structure

4.1.6.3. $f_s/4$ Modulation

The final filter minimization employed is one that was originally developed in [Wong91]; it results in a factor of two reduction in hardware complexity, and is of sufficient merit that it should be mentioned here. The idea is to utilize a bandpass pulse shape instead of a lowpass one, with the bandpass centered at $1/4$ of the oversampling frequency f_s . The reason for this choice of center frequency is clear: a digital $f_s/4$ sinusoid sampled at f_s has a waveform that consists entirely of $+/-1$, or zero. The modulation can be folded into the FIR filter by simply multiplying the filter impulse response by the sinusoid; this is a lowpass-to-bandpass transformation (figure 4.13). By taking advantage of the fact that every other sample of the sinusoid is zero, half of the filter coefficients can be eliminated, resulting in a 2x reduction in hardware. In particular, given the parallel structure developed for the filter, two of the subfilters do not need to be implemented. The in-phase path will be modulated by a 64 MHz cosine wave,

eliminating subfilters 2 and 4, and the quadrature path modulated by a 64 MHz sine, eliminating subfilters 1 and 3. In comparison to figure 4.10b, the overall complex-valued, band-pass filter structure is shown in figure 4.14.

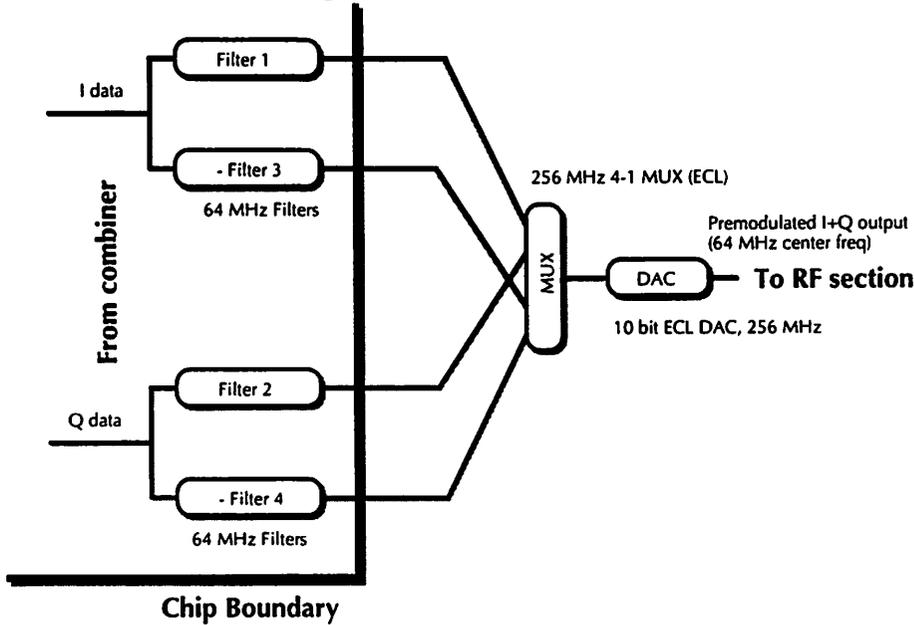


Figure 4.14: Final minimized filter architecture. One MUX/DAC can be eliminated due to the interleaved zero values in the addition.

The ramifications on the overall system design of this $f_s/4$ modulation need to be considered carefully. A conventional homodyne upconversion strategy simply multiplies the in-phase path by a cosine at the carrier $f_c = 1.088$ GHz, and the quadrature by a sine at the same frequency. Effectively, by introducing a digital modulation of $f_s/4$, a two-stage upconversion has been created, with the first stage at 64 MHz and the second stage at 1.024 GHz. The advantages of this are numerous:

- Reduction in digital filter complexity by a factor of 2.
- The digital data from the chip has no spectral energy at DC, being centered at 64 MHz. AC coupling can be employed in all of the analog stages that come after the modulator, especially minimizing factors such as DC offset and low frequency noise in the subsequent DAC.

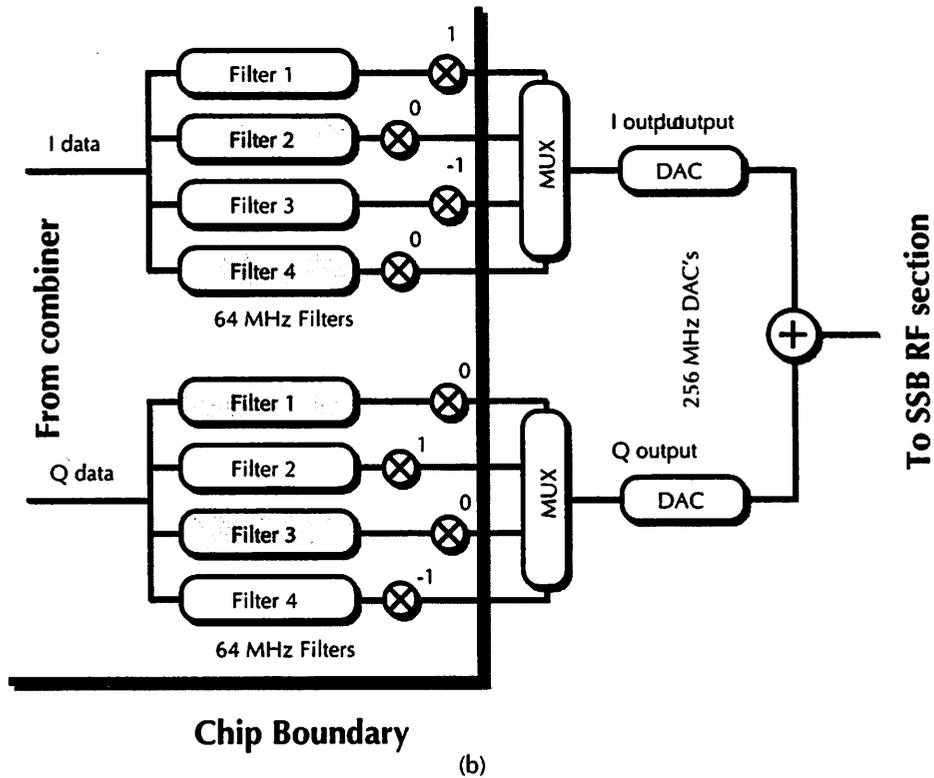
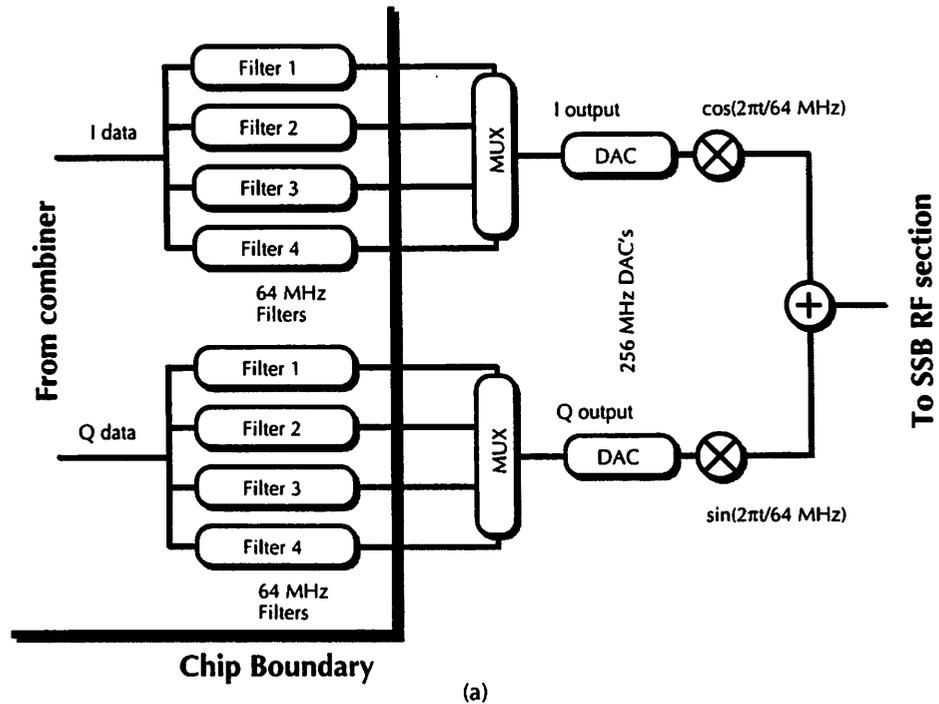


Figure 4.13: (a) Premodulation to 64 MHz (equivalent to a lowpass-bandpass pulse shape transform); (b) Folding the multiplication into the 4 parallel filters: filter complexity is halved (grey filters are unnecessary).

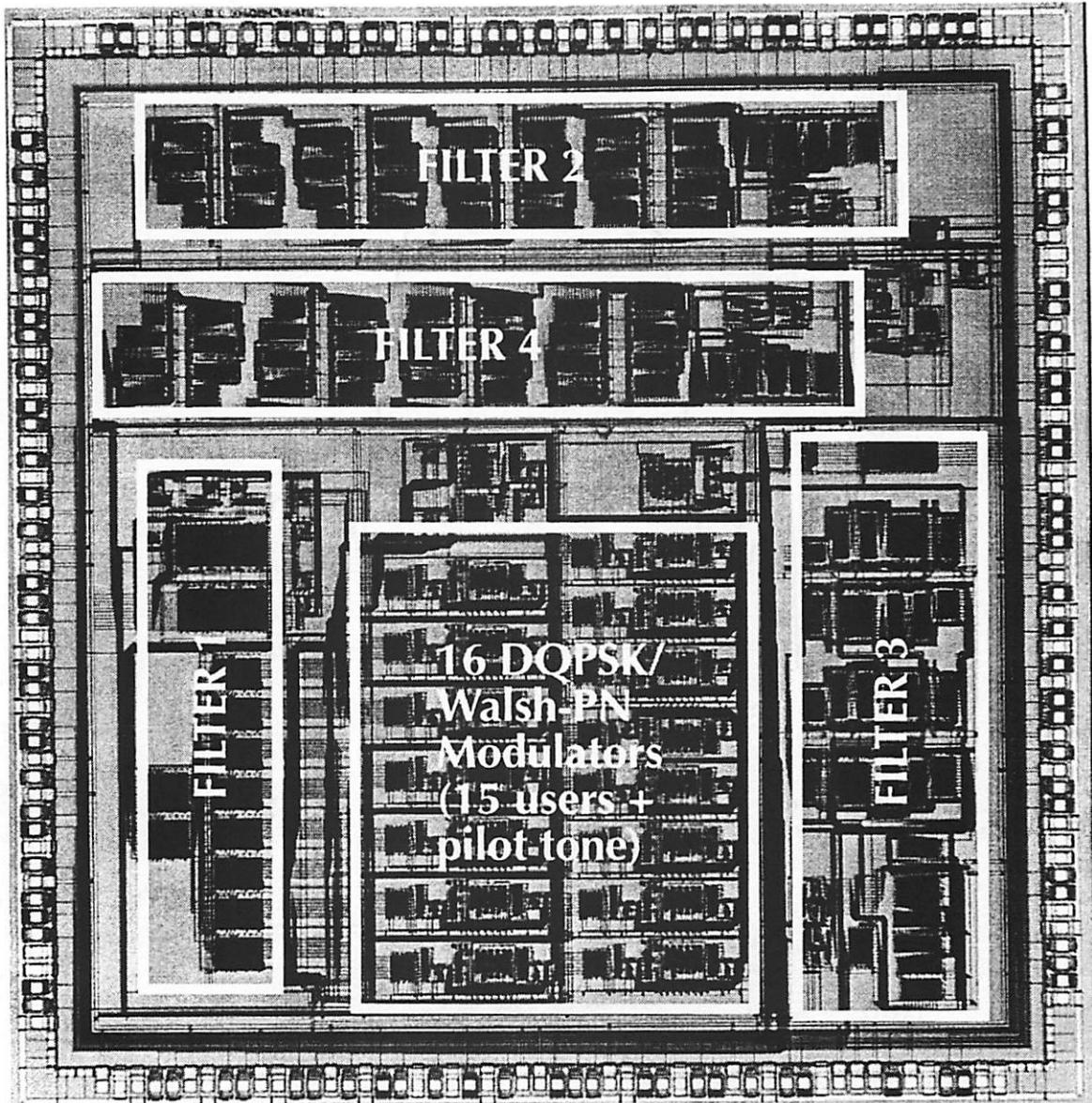
- As in superheterodyne receivers, a two-stage upconversion implies that the second oscillator will be out-of-band with respect to the transmit signal (1.024 GHz versus 1.088 GHz). Carrier feedthrough performance of the upconversion mixers has been reduced, since it will be filtered to some extent by the post-mixer spur suppression filter.

The problem that arises is the need to perform single-sideband modulation – simply multiplying by a cosine at 1.024 GHz is not sufficient, since both the positive and negative frequency images will be upconverted. The desired signal centered at 1.088 GHz will be synthesized; however, an unwanted image at 960 MHz will also be present; it is this negative frequency image that needs to be eliminated, though the use of single-sideband modulation. Section 4.3 will discuss the ramifications of this, in terms of hardware complexity of the D/A converter and modulator sections.

4.2. Results

The baseband modulator circuit has been fabricated in a standard digital 0.8 micron HP CMOS process; the die photo is shown in figure 4.15. Relevant measured die parameters are also listed there; the die has been fully tested to its full clock frequency of 64 MHz. The outputs, as designed, are effectively at 256 MHz, interleaved by a factor of 4. In figure 4.16a, the spectrum of the filter output is shown, for just the pilot tone as input. Since the pilot tone spectrum is white, the output will spectrally be proportional to the filter shape; it achieves the 40 dB out-of-band rejection specification, while maintaining a 30% excess bandwidth pulse shape. The reconstructed time-domain eye diagram is shown in figure 4.16b; it achieves perfect opening at the sample point, satisfying the Nyquist zero-ISI criteria.

One key result is the fact that this single die can support 15 users simultaneously, while consuming only 330 mW of power from a 3.3V rail. This includes all spectrum spreading, pulse shaping, combining, and per-user power control; the benefits of multiple access via digital coding are apparent here. In comparison, had a frequency-hopped system been employed, 15 separate digital oscillators would have been needed; from work in direct digital synthesis [Nichol91], the oscillators alone would have consumed 14.4 W of power.



Transmit Modulator Chip	
Area:	9.9mm x 10.0mm
Transistor Count:	112000
Clock rate:	64 MHz
Power/Supply:	330mW @ 3.3V

Figure 4.15: Modulator Die Photo plus Performance Summary

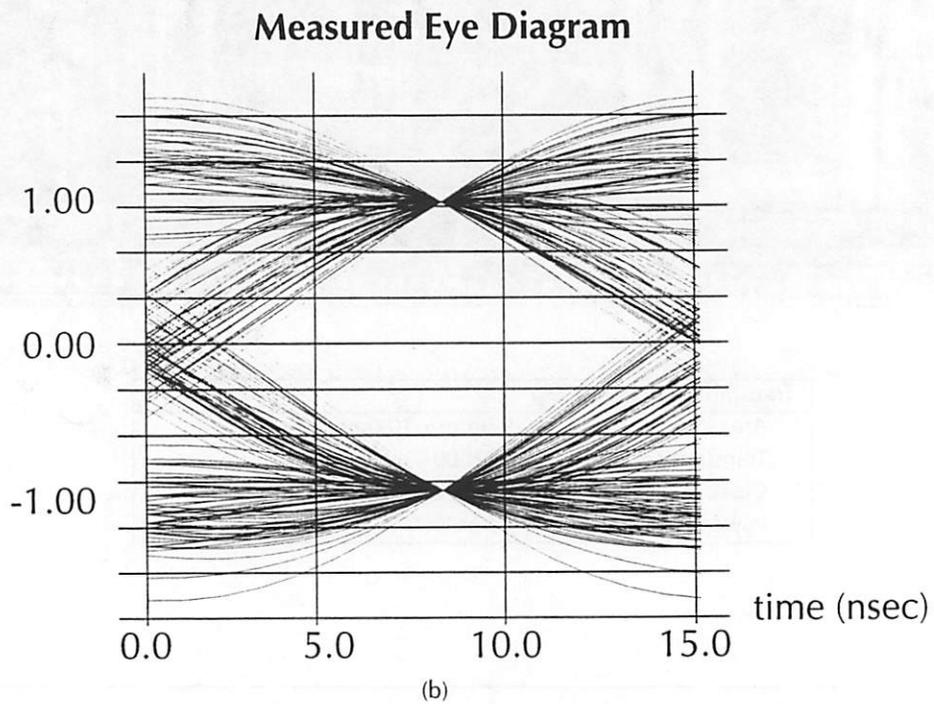
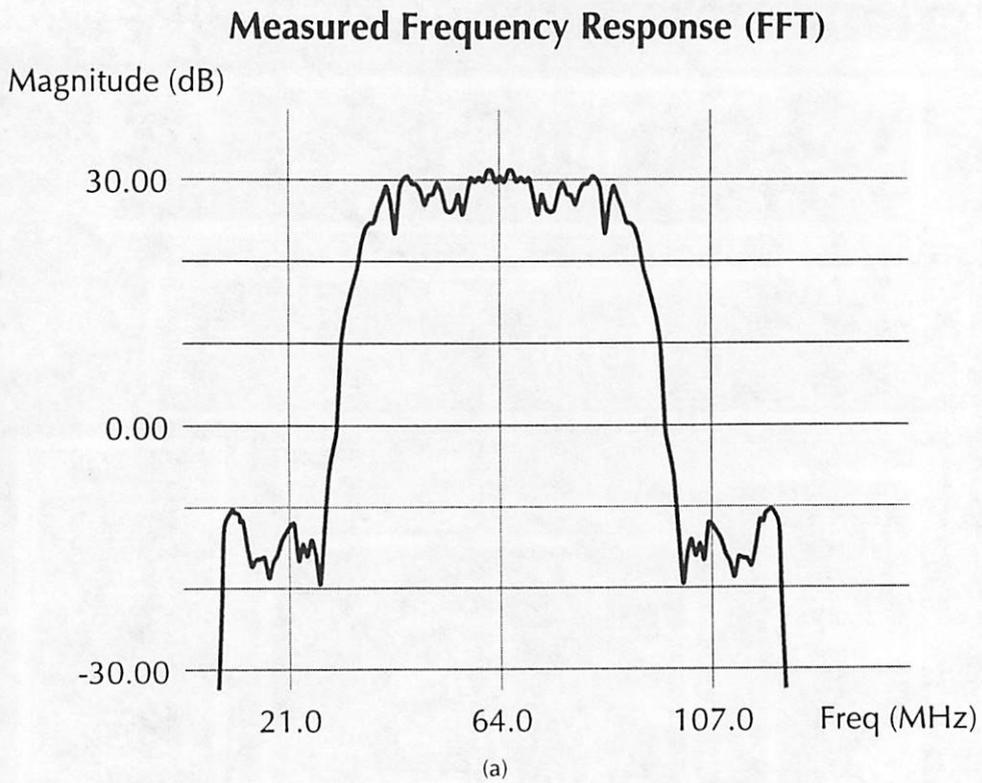


Figure 4.16: Measured modulator output. (a) FFT of die output (pilot tone only; output proportional to pulse shape); (b) reconstructed eye diagram

4.3. Digital-to-Analog Conversion and Frequency Translation

From the 4 parallel outputs of the modulator integrated circuit, they must be multiplexed back into a 256 MHz serial stream, passed through a state-of-the-art 10 bit ECL D/A converter, and then modulated to a carrier frequency by a 1024 MHz carrier. In this section, several architectures to perform the required analog signal processing are described, especially with regards to achieving broadband, single-sideband frequency upconversion. Again, single sideband is necessary since the first stage of upconversion has been performed in the digital domain – the output of the modulator is centered at 64 MHz, not DC.

The immediate question, of course, is why is single-sideband (SSB) conversion such an issue in this system; SSB has been employed in RF communications since the early 1960's. The key issue here is that a broadband SSB conversion is needed, in contrast to conventional narrow-band SSB systems. Conventional systems employ a 90° phase shift network [Koul91] to achieve an analog equivalent to a Hilbert transform (figure 4.17a); such phase shifters do not possess flat phase or amplitude response over the entire band of interest – 83 MHz wide centered at 64 MHz – and creation of a 90° phase shift network with such characteristics was found to be nearly impossible. Several alternatives are presented below.

4.3.1. RF Rejection Method

In figure 4.17b, the most obvious solution to the SSB upconversion problem is illustrated – simply filter out the unwanted negative frequency image at 960 MHz. Although this seems reasonable, the filter response is a severe issue: all isolation would come from the rolloff in 40 MHz of band (equivalently, slightly under 4% of the bandpass center at 1.088 GHz). From typical passive filter data sheets [Daden91], no more than about 20 dB of isolation could be reasonably expected, which is insufficient to meet out-of-band transmission rejection requirements.

4.3.2. Down-Up Method

Another possibility is to bring the modulated spectrum back to baseband, by digitally undoing the 64 MHz sinusoidal modulation (again, a simple ± 1 multiply) before the off-chip multiplexer, and then homodyne modulating the baseband signal up to 1.088 GHz (figure 4.17c). This is probably the least attractive solution, for the following reasons:

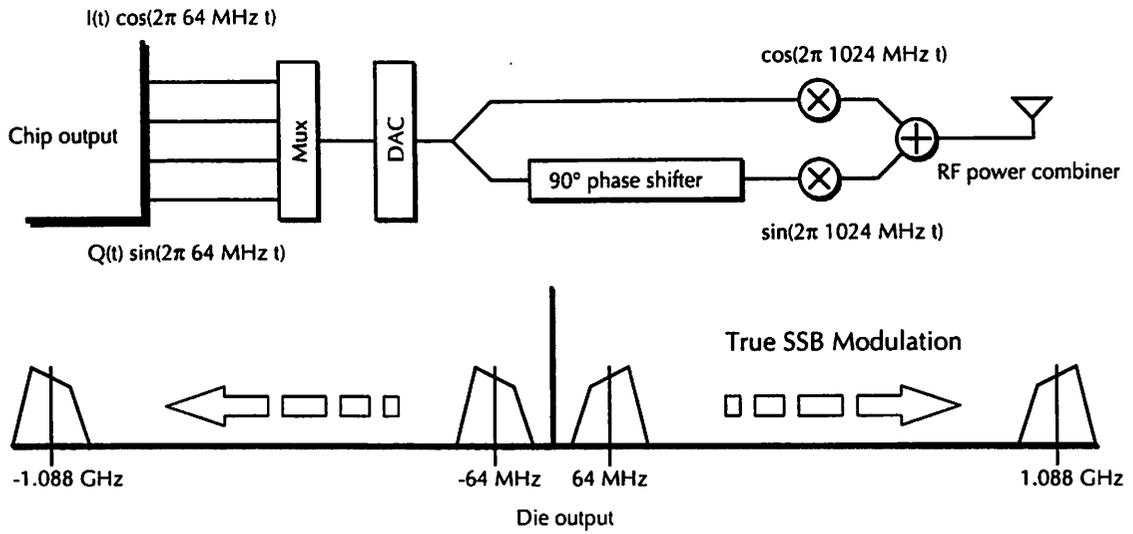
- Carrier feedthrough is now an issue, as well as DC coupling and offsets through the upconversion circuitry.
- A second DAC running at 128 MHz is now needed; furthermore DAC droop is now much more significant.

The advantage, of course, is that this is easily implementable, and is effective provided that the issues above can be overcome.

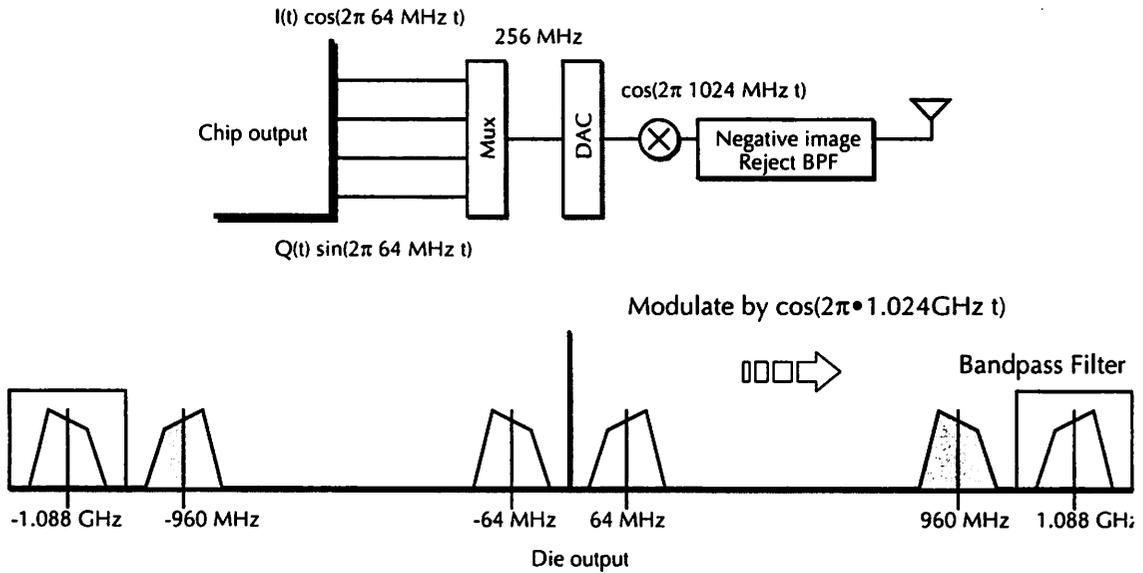
4.3.3. Digital Hilbert Transform

The last method that can be employed is to simply perform the Hilbert transform in the digital domain. Unlike analog 90° phase shifters, it is quite possible to perform the required phase shift while preserving both phase and amplitude flatness across the entire band. In particular, the existing outputs of the die are two signal streams, $I(t)\cos(\omega_s/4)t$ and $Q(t)\sin(\omega_s/4 t)$, where $I(t)$ and $Q(t)$ are the in-phase and quadrature signals respectively. By synthesizing $I(t) \sin(\omega_s/4 t)$ and $Q(t) \cos(\omega_s/4 t)$ on-chip, in the bandpass pulse-shaping fashion that the two existing signals are being generated, the true Hilbert transform can be taken, with no phase or amplitude distortion across the band. In some sense, the benefits of halving hardware complexity have disappeared, since eight filters are needed (four for each signal path). However, the benefits of AC coupling and minimized carrier injection still exist.

The shortcoming of this method is, of course, doubling the filter hardware on the transmit modulator die, as well as requiring a second DAC, as shown in figure 4.17e. Given that most of the existing 1 square centimeter die area is occupied by the current filter design, doubling the complexity becomes difficult from a strict area standpoint.

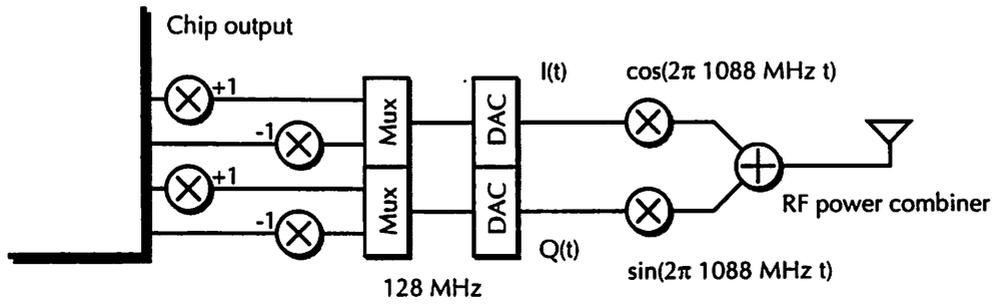


(a)

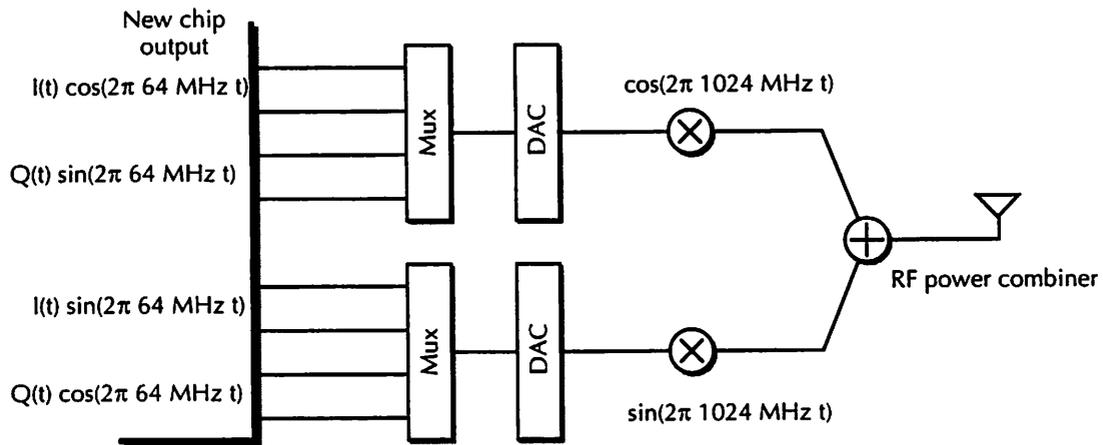


(b)

Figure 4.17: SSB Upconversion Strategies.
 (a) Analog Hilbert Transform; (b) RF rejection filter



(c)



(d)

Figure 4.17: SSB Upconversion Strategies (continued)
 (c) Down-Up Method; (d) Digital Hilbert Transform

4.1. Epilogue: The Infopad Base Station

The key research and development effort on the transmit side has been a fully custom base-band transmit modulator circuit, a single chip capable of synthesizing a broadband spread-spectrum signal multiplexing 15 users simultaneously. The complexity benefits of using spread-spectrum are apparent: the hardware necessary to synthesize each user's transmit signal

is easily implemented using basic digital blocks, as opposed to traditional frequency division systems that require one analog oscillator per user. For the remaining semicustom analog signal processing – D/A conversion and frequency translation – a preliminary prototype has been designed using the down-up modulation method, and is currently under test.

The final issue that remains is how multiple die can be ganged together to support more users than 15 in a single base station, if necessary. The key thing to note is that only one pilot tone should be present; if n die are to be used in parallel, one should be designated as a timing master and bears the pilot tone; the remaining $(n-1)$ should suppress pilot generation. The most obvious solution would be to take n modulator die, clocked by the same crystal oscillator, and perform a digital addition from the corresponding outputs of all n die (figure 4.18a). Provided their reset circuitry is released at the same instant, perfect PN and Walsh synchronization is guaranteed, since the combining of the n outputs is done in the digital domain. The key limitation to this method is the issue with dynamic range in the D/A converter: since multiple die outputs are being combined pre-DAC, the fixed 10-bit limitation of the DAC will end up reducing available dynamic range to each die. At the other extreme, another solution would be to simply drop multiple complete transmit boards, and power combine their RF 1 GHz outputs (figure 4.18b). The key drawback to this in synchronizing the die between each board – for the pilot tone system to work, the boards need to be synchronized to better than $0.125 T_{\text{chip}}$, or 2 nsec^1 , else the mis-synchronization would be disastrous. The optimal solution is shown in figure 4.18c; analog combining after the DAC, but before the frequency upconversion. Since the DAC's are clocked from a common frequency reference, provided careful skew control and layout are done on the board, synchronization is preserved between the n parallel signals; likewise, no dynamic range is lost in the DAC. The obvious drawback is

1. The $0.125 T_{\text{chip}}$ requirement arises from the worst-case delay-locked loop timing error in the receiver (chapter 7); it also arises from the timing tolerance in the eye diagram for a 30% excess bandwidth raised cosine (figure 3.3).

the need for multiple DAC's; clock coupling, radiation, and noise while preserving 10-bit linearity become challenging board design issues, especially with as many as four DAC's on each board.

Given that a solution has been developed to support multiple users, perform pulse shaping, and convert the signal into the 1 GHz transmit band, the remaining problem, which is also the most difficult one of the entire system, lies in the implementation of the mobile receiver. From a tiny broadband received signal off the antenna, the receiver must amplify this signal, perform analog-to-digital conversion, and provide sufficient digital signal processing capability to recover and RAKE combine the multipath arrivals. The next two chapters deal with a single-chip solution for the analog RF receiver front-end, while chapter 7 describes the inverse of the modulator chip: the digital spread-spectrum receiver.

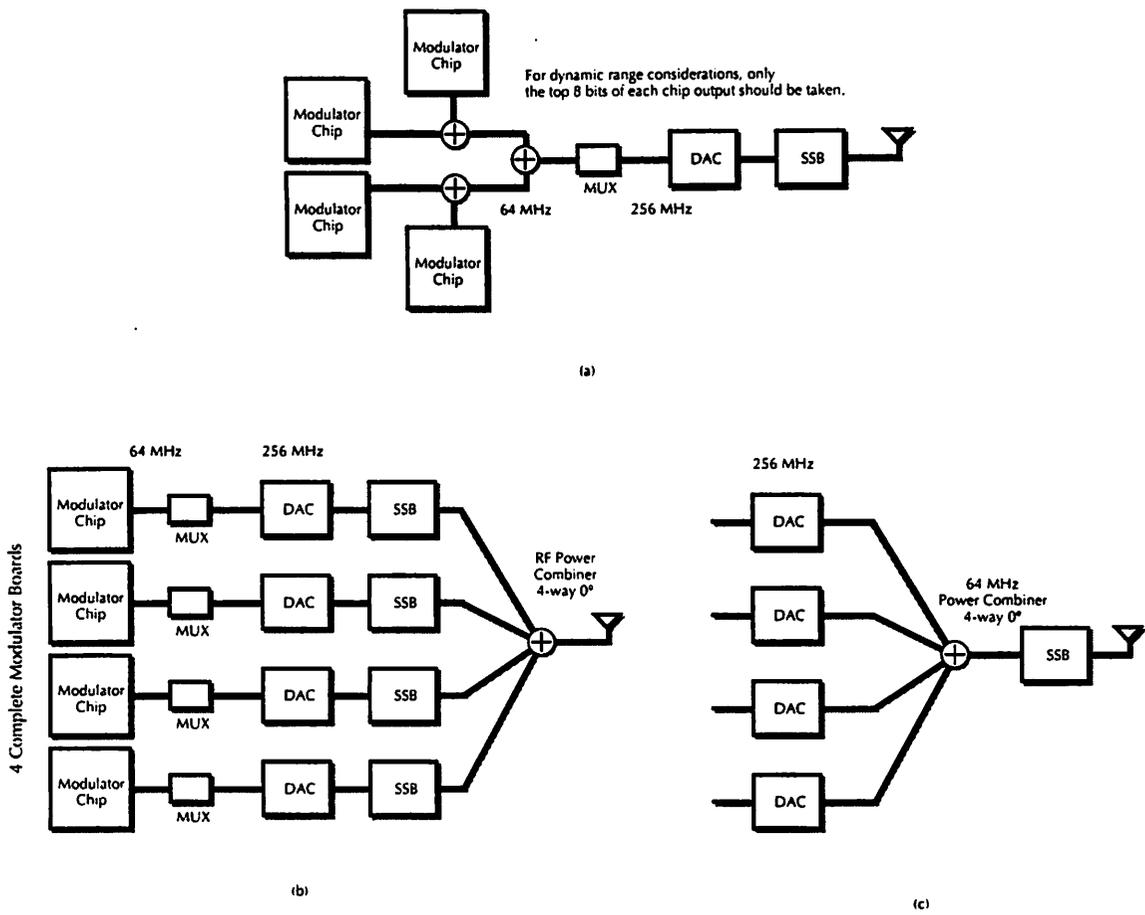


Figure 4.18: Multi-modulator combining schemes for the base station

4.2. Appendix: Modulator Electrical Interface

This section is meant to be used as a programming reference for the chip and describes all the internal registers, the read and write timing, reset procedures, recommended usage, etc. In addition, some of the layout of the internal blocks will also be described.

4.2.1. Pinout

The pads on the die are numbered from 1 to 208 starting in the upper right corner and going counterclockwise. The die were mounted in a Kyocera 208 pin PGA whose pinout is given in figure 4.17. Note that this is the bottom view and, for reference, pad 108 appears as the upper left pin when looking from the bottom.

	A	B	C	D	E	F	G	H	J	K	L	M	N	P	R	S	T	
17	108	112	113	117	119	122	124	128	132	133	137	139	142	144	147	149	160	17
16	97	103	107	109	114	118	123	127	129	134	138	143	148	151	153	155	161	16
15	95	101	104	106	111	116	121	126	131	136	140	145	150	154	156	159	165	15
14	92	99	102	105	110	115	120	125	130	135	141	146	152	157	158	161	169	14
13	90	96	98	100										162	163	166	171	13
12	87	91	93	94										167	168	170	179	12
11	85	86	88	89										172	173	175	175	11
10	81	82	84	83										177	178	179	180	10
9	80	77	79	78										182	183	181	184	9
8	76	75	74	73										187	188	186	185	8
7	72	71	69	68										193	192	190	189	7
6	70	66	64	63										198	197	195	194	6
5	67	62	59	58	209									204	202	200	199	5
4	65	57	54	53	48	42	37	31	26	21	16	11	6	1	206	203	195	4
3	61	55	52	50	46	41	36	32	27	22	17	12	7	2	208	205	199	3
2	60	51	49	47	44	39	34	30	25	23	19	14	10	5	3	207	201	2
1	56	45	43	40	38	35	33	29	28	24	20	18	15	13	9	8	4	1

Figure 4.19: Modulator package pinout (209-pin PGA)

The following tables describe the functionality and pin assignments of the die interface. The first column indicates the pad number as described above. The next column tells whether this is an input or an output pad. The next name indicates the full name of the signal that is connected to that pad. The last string on a given line is an alias for the name. For example, databus_pins_in[0] is also referred to as di0.

Microprocessor Interface	208 inaddressbus_pins[0] a0
5 inaddressbus_pins[1] a1	15 inaddressbus_pins[2] a2
18 inaddressbus_pins[3] a3	20 inaddressbus_pins[4] a4
24 inaddressbus_pins[5] a5	28 inaddressbus_pins[6] a6
171 indatabus_pins_in[0] di0	169 indatabus_pins_in[1] di1
159 indatabus_pins_in[2] di2	153 indatabus_pins_in[3] di3
144 indatabus_pins_in[4] di4	142 indatabus_pins_in[5] di5
139 indatabus_pins_in[6] di6	137 indatabus_pins_in[7] di7
194 outdatabus_pins_out[0] do0	191 outdatabus_pins_out[1] do1
189 outdatabus_pins_out[2] do2	185 outdatabus_pins_out[3] do3
184 outdatabus_pins_out[4] do4	180 outdatabus_pins_out[5] do5
176 outdatabus_pins_out[6] do6	174 outdatabus_pins_out[7] do7
196 indb_write_endbwe	205 incs_lcs1
99 inwritewr	109 inreadrd
Parallel outputs	
59 outi1out[0] i10	64outi1out[1] i11
69 outi1out[2] i12	74outi1out[3] i13
79 outi1out[4] i14	84outi1out[5] i15
88 outi1out[6] i16	93outi1out[7] i17
98 outi1out[8] i18	102outi1out[9] i19
17 outi2out[0] i20	22outi2out[1] i21
26outi2out[2] i22	32outi2out[3] i23
36outi2out[4] i24	41outi2out[5] i25
46outi2out[6] i26	50outi2out[7] i27
54outi2out[8] i28	178outq1out[0] q10
183outq1out[1] q11	188outq1out[2] q12
192outq1out[3] q13	197outq1out[4] q14
202outq1out[5] q15	206outq1out[6] q16
2outq1out[7] q17	7outq1out[8] q18
12outq1out[9] q19	131outq2out[0] q20
136outq2out[1] q21	140outq2out[2] q22
145outq2out[3] q23	150outq2out[4] q24
154outq2out[5] q25	158outq2out[6] q26
163outq2out[7] q27	168outq2out[8] q28
173outq2out[9] q29	
User Data	132ininMSB[1] im1
124ininMSB[2] im2	119ininMSB[3] im3
113ininMSB[4] im4	95ininMSB[5] im5
90ininMSB[6] im6	85ininMSB[7] im7
80ininMSB[8] im8	72ininMSB[9] im9
67ininMSB[10] im10	61ininMSB[11] im11
45ininMSB[12] im12	40ininMSB[13] im13
35ininMSB[14] im14	29ininMSB[15] im15

Table 4.1: Modulator pinout arranged by functional group

133ininLSB[1] i11	128ininLSB[2] i12
122ininLSB[3] i13	117ininLSB[4] i14
103ininLSB[5] i15	92ininLSB[6] i16
87ininLSB[7] i17	81ininLSB[8] i18
76ininLSB[9] i19	70ininLSB[10] i110
65ininLSB[11] i111	51ininLSB[12] i112
43ininLSB[13] i113	38ininLSB[14] i114
33ininLSB[15] i115	
Test signals	164intestloadmulttstlm
201intestloadPCwtstlp	4intestDCpulsetstdc
160intestmodetestmode	
Signals for the PN generator	108instopPN_lstpPN
97inload_PNldPN	126outPN_allones_lPNa
8outPNsigPNsig	60inPNreset_lPNrst
Misc signals	
116inphi1phi1	112instopWalsh_lstpww
121outdataclockdc	56inresetrst
3outloadPCwlp	13outloadmultlm
Core VDD	
147Vdd	114Vdd
129Vdd	148Vdd
105Vdd	130Vdd
96Vdd	89Vdd
172Vdd	175Vdd
77Vdd	186Vdd
62Vdd	200Vdd
37Vdd	55Vdd
44Vdd	30Vdd
14Vdd	
Core GND	149GND
118GND	134GND
151GND	110GND
141GND	91GND
170GND	78GND
182GND	181GND
75GND	195GND
57GND	25GND
49GND	39GND
27GND	10GND
Padring VDD	
123Vdd	138Vdd
101Vdd	106Vdd
156Vdd	165Vdd
125Vdd	146Vdd
152Vdd	100Vdd
162Vdd	166Vdd
94Vdd	82Vdd
177Vdd	73Vdd
68Vdd	190Vdd
198Vdd	204Vdd
53Vdd	42Vdd
21Vdd	6Vdd

Table 4.1: Modulator pinout arranged by functional group

52Vdd	199Vdd
34Vdd	23Vdd
9Vdd	71Vdd
Padring GND	
107GND	127GND
143GND	155GND
104GND	111GND
115GND	120GND
135GND	157GND
161GND	167GND
86GND	83GND
179GND	187GND
193GND	66GND
63GND	58GND
48GND	31GND
16GND	11GND
1GND	203GND
47GND	19GND
207GND	

Table 4.1: Modulator pinout arranged by functional group

1GND	2q1out [7]
3loadPCw	4testDCpulse
5addressbus_pins [1]	6Vdd
7q1out [8]	8Pnsig
9Vdd	10GND
11GND	12q1out [9]
13loadmult	14Vdd
15addressbus_pins [2]	16GND
17i2out [0]	18addressbus_pins [3]
19GND	20addressbus_pins [4]
21Vdd	22i2out [1]
23Vdd	24addressbus_pins [5]
25GND	26i2out [2]
27GND	28addressbus_pins [6]
29inMSB [15]	30Vdd
31GND	32i2out [3]
33inLSB [15]	34Vdd
35inMSB [14]	36i2out [4]
37Vdd	38inLSB [14]
39GND	40inMSB [13]
41i2out [5]	42Vdd
43inLSB [13]	44Vdd
45inMSB [12]	46i2out [6]
47GND	48GND
49GND	50i2out [7]
51inLSB [12]	52Vdd
53Vdd	54i2out [8]
55Vdd	56reset
57GND	58GND

Table 4.2: Modulator pinout by pin ordering

59i1out[0]	60Preset_1
61inMSB[11]	62Vdd
63GND	64i1out[1]
65inLSB[11]	66GND
67inMSB[10]	68Vdd
69i1out[2]	70inLSB[10]
71Vdd	72inMSB[9]
73Vdd	74i1out[3]
75GND	76inLSB[9]
77Vdd	78GND
79i1out[4]	80inMSB[8]
81inLSB[8]	82Vdd
83GND	84i1out[5]
85inMSB[7]	86GND
87inLSB[7]	88i1out[6]
89Vdd	90inMSB[6]
91GND	92inLSB[6]
93i1out[7]	94Vdd
95inMSB[5]	96Vdd
97load_PN	98i1out[8]
99write	100Vdd
101Vdd	102i1out[9]
103inLSB[5]	104GND
105Vdd	106Vdd
107GND	108stopPN_1
109read	110GND
111GND	112stopWalsh_1
113inMSB[4]	114Vdd
115GND	116phi1
117inLSB[4]	118GND
119inMSB[3]	120GND
121dataclock	122inLSB[3]
123Vdd	124inMSB[2]
125Vdd	126PN_allones_1
127GND	128inLSB[2]
129Vdd	130Vdd
131q2out[0]	132inMSB[1]
133inLSB[1]	134GND
135GND	136q2out[1]
137databus_pins_in[7]	138Vdd
139databus_pins_in[6]	140q2out[2]
141GND	142databus_pins_in[5]
143GND	144databus_pins_in[4]
145q2out[3]	146Vdd
147Vdd	148Vdd
149GND	150q2out[4]
151GND	152Vdd
153databus_pins_in[3]	154q2out[5]
155GND	156Vdd
157GND	158q2out[6]
159databus_pins_in[2]	160testmode
161GND	162Vdd

Table 4.2: Modulator pinout by pin ordering

163q2out[7]	164testloadmult
165Vdd	166Vdd
167GND	168q2out[8]
169databus_pins_in[1]	170GND
171databus_pins_in[0]	172Vdd
173q2out[9]	174databus_pins_out[7]
175Vdd	176databus_pins_out[6]
177Vdd	178q1out[0]
179GND	180databus_pins_out[5]
181GND	182GND
183q1out[1]	184databus_pins_out[4]
185databus_pins_out[3]	186Vdd
187GND	188q1out[2]
189databus_pins_out[2]	190Vdd
191databus_pins_out[1]	192q1out[3]
193GND	194databus_pins_out[0]
195GND	196db_write_en
197q1out[4]	198Vdd
199Vdd	200Vdd
201testloadPCw	202q1out[5]
203GND	204Vdd
205cs_1	206q1out[6]
207GND	208addressbus_pins[0]

Table 4.2: Modulator pinout by pin ordering

4.2.2. Microprocessor timing

The next thing that has to be described is the electrical interface to the chip and the relative timing of some of the control signals. Probably the most important of these timing diagrams is the one that describes how the various registers in the chip are written to and read from.

At the time that the chip was designed, bi-directional pads were not available in this technology and hence, the data bus was implemented with two separate sets of eight lines. One set was for input (databus_in) and the other was for output (databus_out). When reading from various registers from within the chip, one will put the address on the address bus and observe the data on databus_out. When writing, the programmer must present the data on databus_in and then perform a write. The wiring of this interface is shown below.

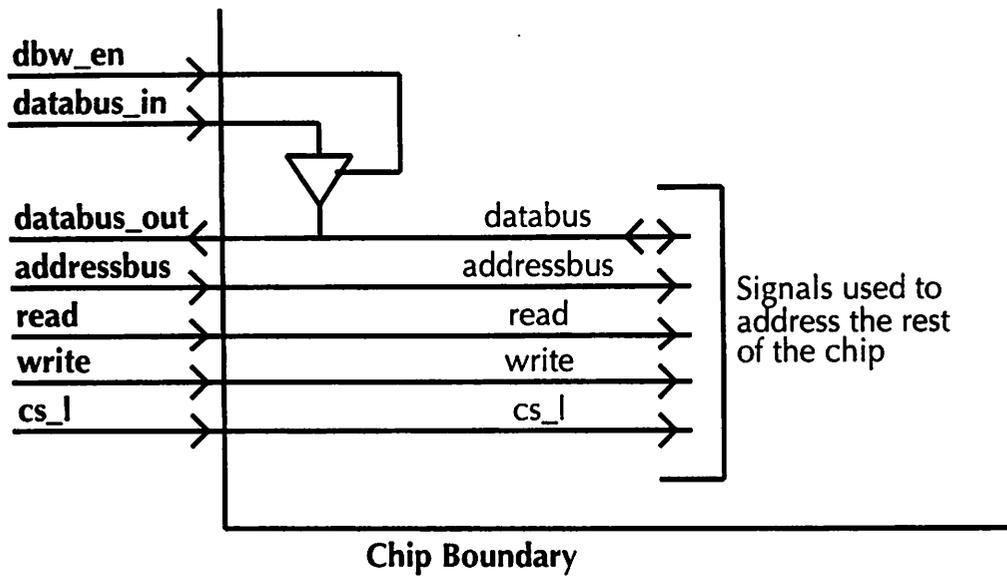


Figure 4.20: Physical circuitry of the microprocessor interface

As can be seen, the internal data bus is a bi-directional line. All the internal registers have tristates that enable them to conditionally drive the internal data bus lines. The `dbw_en` signal is used to drive the internal data bus with the number that has been placed on the `databus_in` line. As a side effect of the above wiring, when the `dbw_en` signal is high, whatever is placed on the `databus_in` lines appears on the `databus_out` lines (assuming that there is no contention between the `dbw_en` tristate and another tristate that is on the data bus).

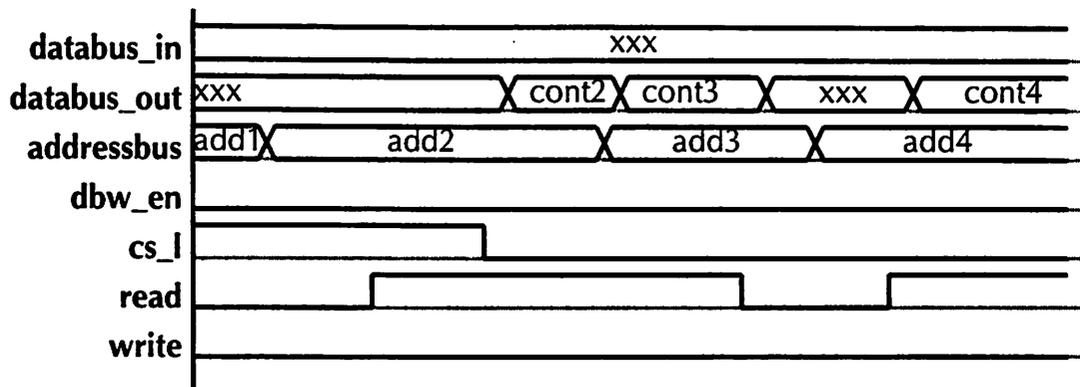


Figure 4.21: Timing of a microprocessor read cycle

The above timing diagram describes the procedure whereby a constant is read from one of the internal registers. The idea is fairly standard in that you place the address on the address lines, lower the write signal, and raise the read signal. The contents of the register being addressed will appear on the databus_out lines.

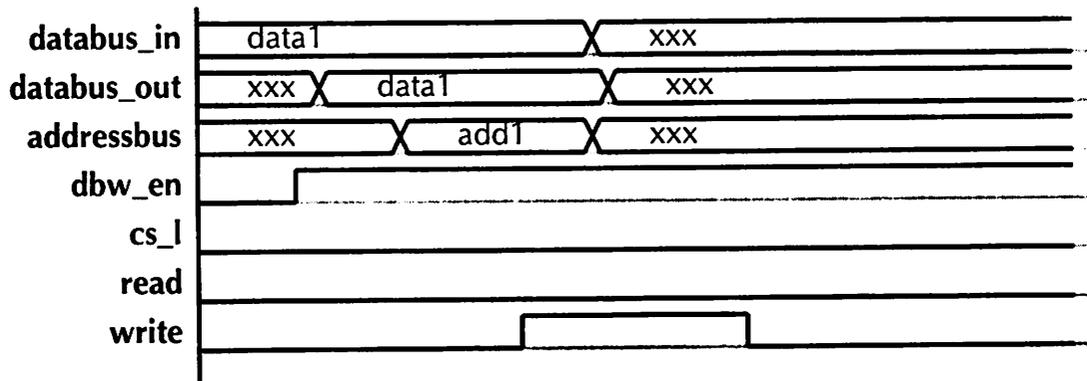


Figure 4.22: Timing of a microprocessor write cycle

This timing diagram shows how a write cycle is expected. Two things must happen and may be done in any order. One of those things is that the data to be written must be placed on the databus_in lines and the dbw_en line must be brought high. The second thing that must be

done is that the address to be written must be placed on the address bus. After these steps have been performed, the write can proceed. This is accomplished by raising the write line. The data is stored on the positive edge of the write line and hence, after the positive edge, the data on the databus_in lines, and the address on the address bus, can become unstable.

4.2.3. Clock timing

This section describes the timing of the input data, the output data, and the data clock signal that comes from the chip. Basically, when this chip is in use, it will be in a free running mode. The programmer will have performed various writes that tell the chip which users are active and what their power control values are. After this is done, the chip will just keep running based on the 64 MHz clock that comes into it.

When the programmer wants to change a particular user's power control value, he will do so by performing writes as indicated in the previous section. The thing to note is that these writes are asynchronous with respect to the 64 MHz clock. The system designer does not have to worry about the state of the 64 MHz clock when performing a write. Synchronization is performed internally through various means such as double buffering various constants.

Hence, this section does not include any timing diagrams that include a 64 MHz clock and any microprocessor signals (read, write, cs_l, etc.). This section only deals with the timing of the output data, the input data, and the data clock signal relative to the 64 MHz clock (which is called phi1).

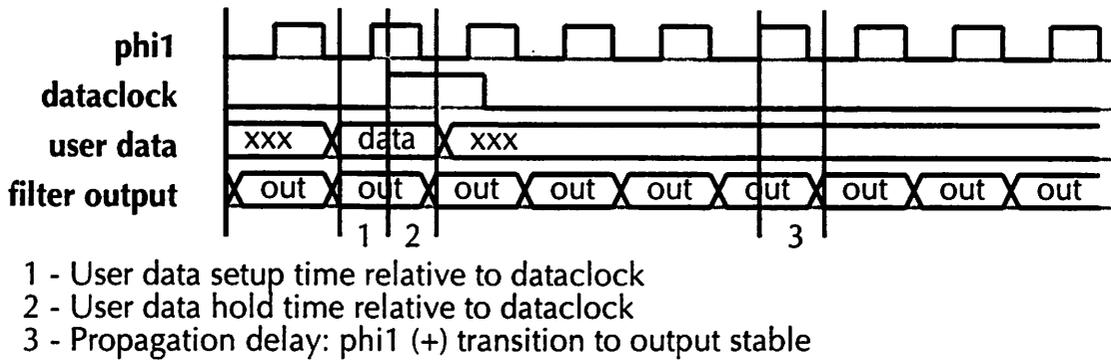


Figure 4.23: Timing of signals relative to the 64 MHz clock (phi1)

With all that aside, the above figure is the timing diagram. The data clock signal is a 1 MHz clock that is generated by dividing the 64 MHz clock down by 64. This is the clock that is used to time the user data going into the chip. For instance, it is anticipated that the data coming into the chip will be coming from a shift register. The positive edge of the data clock line can be used to indicate that a shift must be performed.

Internally, the data is latched on the positive edge of this line and hence, by the time the data clock line goes high, the data has already been strobed. In other words, the delay induced in the propagation of the data clock signal from inside the chip to the outside is greater than the hold time for the internal registers.

4.2.4. Description of registers

The following table is a comprehensive list of all the registers in the chip along with a brief description of the use of the particular register. More detail on each of these registers can be found after the table. Note that a star (*) indicates a register or a bit of a register that is only useful for debugging the chip and will not be used in the normal operation of the chip.

Address		Bits	R/W	Description	
Start	End				
Name					
0000000	0001111	PC##	7*	R	qbit signal internal to the given spreading block
			6*	R	ibit signal internal to the given spreading block
			5:0	W	Store the power control value that is to be loaded later.
			5:0	R	Read the current power control value. This will be the same as the value written only after a powercontrol load has been initiated.
0010000	0011111	WALSH##	7*	R	PNXORWalsh signal from within spread block
			6*	R	Walsh signal from within spreading block
			5:0	W	Store the Walsh constant that is to be loaded later.
			5:0	R	Read the current Walsh constant. This will be the same as the value written to these bits only after a Walsh load has been initiated.
1000000		PNseedms8	7:0	R/W	Read and write the MS 8 bits of the PN seed. The PN seed is loaded upon the rising edge of the loadPN pin.
1000001		PNseedls8	7:0	R/W	Read and write the LS 8 bits of the PN seed. The PN seed is loaded upon the rising edge of the loadPN pin.
1000010		multconst	7	R/W	0 indicates normal multiplier operation. 1* indicates that the output of the multiplier should be the constant contained in the altoutput register. This is used for debugging so that the programmer can control the data going into the filters.
			3:0	R/W	Stores the multiplication constant. The input to the filters will be the output of the combiner times this multiplication constant.
1000011		multiinMS8	7:0*	R	Observe the MS 8 bits of the I data coming out of the combiner and going into the multiplier.
1000100		multiinLS8	7:0*	R	Observe the LS 8 bits of the I data coming out of the combiner and going into the multiplier.
1000101		altoutMS8	7:0*	W	Store the MS 8 bits of the alternate output register of the multiplier. Used in conjunction with the multconst register at location 1000010.
			7:0*	R	Read the MS 8 bits of the Q output of the multiplier.
1000110		altoutLS8	7:0*	W	Store the LS 8 bits of the alternate output register of the multiplier. Used in conjunction with the register at location 1000010.
			7:0*	R	Read the LS 8 bits of the Q output of the multiplier.
1000111		multQinMS8	7:0*	R	Observe the MS 8 bits of the Q data coming out of the combiner and going into the multiplier.
1001000		multQinLS8	7:0*	R	Observe the LS 8 bits of the Q data coming out of the combiner and going to the multiplier.
1001001		spreadIout	6:0*	R	Read the I output of spread block 1
1001010		spreadIQout	6:0*	R	Read the Q output of spread block 1
1001011		debugR1	7*	R	muxout signal of the dcgen block
			6*	R	state signal of the dcgen block
			5:4*	R	state signal of the gen1234 block
			3:0*	R	Contents of the phi registers of the gen1234 block
1001100		loadREQ	1*	R	muxout signal of the dcgen block
			0*	R	state signal of the dcgen block
			N/A	W	A write to this register (the value written is irrelevant) indicates that the power control, Walsh, and multiplication constants should be loaded on the next rollover of the PN generator.
1001101		multPCWGat	7*	R	Read the status of the next pending signal in the genload block.
			6*	R	Read the status of the pending signal in the genload block.
			5*	R	Read the status of the write_decD3 signal in the genload block.

Table 4.3: Description of Registers

		4*	R	Read the status of the write_decD2 signal in the gen-load block.
		3:2	R/W	Set the gating control of the loadmult signal: 00 - No gating 01 - Gate on philret1 10 - Gate on philret2 11 - Gate on philret3
		1:0	R/W	Set the gating control of the loadPCw signal: 00 - No gating 01 - Gate on philret1 10 - Gate on philret2 11 - Gate on philret3
101110	clksource_dcgating	7:4*	R	Read the LS 4 bits of the counter in the dcgen block.
		3:2	R/W	Set the clock source of this block: 00 - phil 01 - philret1 10 - philret2 11 - philret3
		1:0	R/W	Set the gating control of the dataclock signal: 00 - No gating 01 - Gate on philret1 10 - Gate on philret2 11 - Gate on philret3
1001111	multIoutMS3	2:0*	R	Read the MS 3 bits of the I output of the multiplier.
1010000	multIoutLS8	7:0*	R	Read the LS 8 bits of the I output of the multiplier.
1010001	skewReg	7:0*	R	This register is used to give the chip tester a general view of the timing of the clock signal in various places of the chip. By reading this register and toggling the phil pin, the chip tester can get a general feel as to how long the clock takes to get to various places from within the chip.
		7*	R	Observe the dataclock signal.
		6*	R	Observe the philret3 signal.
		5*	R	Observe the philret2 signal.
		4*	R	Observe the philret1 signal.
		3*	R	Observe the qfir2return signal.
		2*	R	Observe the qfir1return signal.
		1*	R	Observe the ifir1return signal.
		0*	R	Observe the phil signal.
1010010	q1t5LS8	7:0*	R	Read LS 8 of output of q1_qtap5
1010011	q1t6LS8	7:0*	R	Read LS 8 of output of q1_qtap6
1010100	q1t7LS8	7:0*	R	Read LS 8 of output of q1_qtap7
1010101	q1t8LS8	7:0*	R	Read LS 8 of output of q1_qtap8
1010110	q2t5LS8	7:0*	R	Read LS 8 of output of q2_qtap5
1010111	q2t6LS8	7:0*	R	Read LS 8 of output of q2_qtap6
1011000	q2t7LS8	7:0*	R	Read LS 8 of output of q2_qtap7
1011001	q2t8LS8	7:0*	R	Read LS 8 of output of q2_qtap8
1011010	i1t1LS5	7:0*	R	Read LS 5 of output of i1_itap1
1011011	i1t2LS8	7:0*	R	Read output of i1_itap2
1011100	i1t3LS8	7:0*	R	Read LS 8 of output of i1_itap3
1011101	i1t4LS8	7:0*	R	Read LS 8 of output of i1_itap4

Table 4.3: Description of Registers

4.2.4.1. Power control registers PC00-PC15

This is where the power control values for all the users are stored. User 0 (the pilot tone) has his power control value stored in location 0000000 (PC00). User 1 is located at 0000001 (PC01), etc. Because it is desired for the power control settings for the various users to change

simultaneously, all of these registers are double buffered. A write to a particular user's power control register simply specifies the next power control value that user should use. It does not instantly change the power control value for that user.

The idea is that the system programmer will go in, write all the power control constants for the users, and then tell the system to load them in at a later time. This time has been determined to be the next occurrence of the PN generator rolling over. This provides an update frequency of about 2000 times a second. The programmer indicates that the power control values should be loaded the next time the PN generator rolls over by performing a write to register 1001100 (loadREQ).

4.2.4.2. Walsh registers WALSH00-WALSH15

Put simply, these registers behave the same way as the power control registers. The system programmer must write to these registers to update the Walsh constants. After this is done, the system programmer indicates that the new constants should be loaded by performing a write to the same register described above: 1001100 (loadREQ). Please note that it is not necessary to update the power control constants and then update the Walsh constants. In fact, the programmer can update the power control values, update the Walsh constants, and also update the multiplication constant before writing to loadREQ.

4.2.4.3. PN seed registers PNseedMS8, PNseedLS8

This set of registers does not have a complicated double buffering scheme since the PN seed will only be set once and never be touched again. The programmer must write to both of these registers in order to set the 16 bit PN seed. Once this is done, the loadPN pin on the chip must go through a positive transition. Once this is done, the PN generator has been initialized and will start to generate the PN sequence.

4.2.4.4. Multiplier registers

Since the number of users and their power control values will vary widely, a multiplier was added between the combiner and the filter inputs so as to ensure that the data going to the filters would be at full scale. The reason for doing this is that the filters will insert a fixed about

of noise into the system. Regardless of the magnitude of the data coming into the filters, the power of the noise will be the same. Therefore, by increasing the magnitude of the data coming into the filters, one effectively increases the signal to truncation noise ratio of the signal coming out of the filters.

Multiplication constant multconst

The first register to be described and really the only one that the system programmer will have to use is multconst. This register stores the multiplication constant and determines whether the multiplier is in test mode or not. The least significant 4 bits of this register store the multiplication constant and are double buffered just like the power control and Walsh registers. The system programmer must update the multiplication constant and then indicate to the chip that the new constant should be loaded in the next time that the PN generator rolls over.

A note must be made as to how this multiplication constant should be chosen. Each user has a six bit power control constant. If all 16 users are transmitting at the same time, then the sum of all their outputs will be an 11 bit number (the power control value is a positive number and hence must be appended by a sign bit before all users are added together. This gives 16 seven bit numbers that must be added and produces an 11 bit result).

If all the users are transmitting at full power, then the inputs to the filters would be at full scale. In this case, the multiplication constant should be set to 1. Any other values would saturate the multiplier and have unexpected results. The multiplier becomes useful when there are fewer users transmitting or when the ones that are transmitting are at a lower power control value. The idea is that the multiplier is used to bring up the data to full scale. If only half of the users are transmitting at full power control, then the input to the filters would be at half scale. In this case, the multiplier constant should be set to 2. Another corner case to consider is when there is only one user in the system. In this case, this user's power control should be set to full scale and the multiplier set to 1111.

Alternate output registers: altoutMS8, altoutLS8

The most significant bit of the above register is used to override the output of the multiplier

and force a given value onto the outputs. This is done to enable the programmer to debug the filters that appear after the multiplier. By controlling the output of the multiplier, the programmer can control the data going into the filters and hence test to make sure that the filters are operating properly.

If this MSB is 0, then the multiplier will be operating normally and the output will be the input times the multiplication constant. If this bit is 1, then both the I output and the Q output will be the value stored in the alternate output registers: 1000101 (altoutMS8) and 1000110 (altoutLS8). The first register contains the most significant 3 bits of the alternate output and the second contains the lower 8 bits. (Both the I and the Q outputs are 11 bits wide.)

Reading these registers can be used to observe the Q output of the multiplier. The most significant 8 bits can be read from altoutMS8 and the least significant 8 bits can be read from altoutLS8.

Observing multiplier input: multIinMS8, multIinLS8, multQinMS8, multQinLS8

The data that goes to the multiplier comes from the output of the adder that combines all the streams. Therefore, these registers allow the programmer to observe both the output of the adder and the data that is going to the multiplier. They can be used in conjunction with the multIoutMS8, multIoutLS8, altoutMS8, and altoutLS8 registers to determine if the multiplier is operating properly. These registers can be used to observe the output and the multIinMS8, multIinLS8, multQinMS8, and multQinLS8 registers can be used to observe the input to the multiplier.

4.2.4.5. Observing one of the spread blocks: spread1Iout, spread1Qout

The spreading block for user 1 is the only spread block whose output can be observed. This is the block that does the DQPSK encoding, Walsh multiplication (spreading), PN multiplication, and finally the power control multiplication. The I and Q output of this block can be observed in the spread1Iout and spread1Qout registers. Note that only bits 6:0 of these registers are defined. Bit 7 is undefined but currently set to zero.

4.2.4.6. Debugging register 1: debugR1

Although this register has the designation debugR1, this is the only register that is dedicated solely to miscellaneous debugging signals. There is no debugR2 or debugR3. This register contains the value of signals at various test points from within the dcgen and gen1234 blocks. This register only has meaning when used in combination with the source SDL file to debug these blocks at a very low level.

4.2.4.7. Load request register: loadREQ

This is the register that was mentioned earlier in the description of the power control, Walsh, and multiplication registers. The system programmer will write the new values of these constants and then initiate a load request. A load request is initiated by writing to the loadREQ register. The actual value written has no effect since it is the actual write operation that indicates that a load should be performed. After a write is performed to this register, the dcgen block within the clkgen block will wait for the PN generator to roll over. Once it sees that the PN generator has rolled over, it will assert the loadPCw, and loadmult signals to load the new power control, Walsh, and multiplication constants.

4.2.4.8. Clock timing registers: multPCwGat, clksource_dcgating

These registers are used primarily to control the clock source of the clock generator and the gating of the individual clock signals. A definition of the clock source and the clock gating can be found in section 3.5.

The source of the clock generator is programmed by writing to bits 3:2 of the clksource_dcgating register. The gating of the dataclock signal is controlled by bits 1:0 of this register. Bits 3:2 of the multPCwGat register control the gating of the loadmult signal. Bits 1:0 of this register control the gating of the loadPCw signal.

The other bits of these registers have various meanings and are only to be used for low level debugging. multPCwGat contains various critical signals in the genload block and clksource_dcgating contains the LS4 bits of the counter in the dcgen block.

4.2.4.9. Observing clock skew: skewReg

This register was mentioned briefly in section 3.5 where the clock network was described. This register allows the system programmer to observe the clock signal at various points from within the chip. This allows the programmer to determine the relative clock skew of various points from within the chip but cannot be used as an absolute measure of this skew. The reason that it is not an absolute measure is that by observing the clock through the register, the programmer is observing the skew through both a tristate and an output driver. There is no guarantee that the delay through all 8 tristates and output drivers is the same.

4.2.4.10. Filter debugging registers: q_t_LS8, i1_t_LS8

These registers are used to observe the signal coming out of some of the filter taps. The notation indicates the filter number and the tap number. For instance, q1t5LS8 indicates that this register allows the programmer to observe the least significant 8 bits of tap 5 of filter q1.

5 The Receiver: Analog RF Front-End

The challenge presented by the mobile receiver – low-power gigahertz-band analog performance in addition to digital signal processing in the hundreds of megahertz – is the greatest barrier that needs to be surmounted, if the vision of a broadband wireless terminal is ever to be achieved. As an architectural issue, one major design goal is that the analog hardware be simplified as much as possible; since the carrier frequencies are above 1 GHz, the complexity and difficulties in implementation imply that simplifying the circuitry or relaxing the required analog performance should be paramount. Given the tremendous levels of digital computation achievable by today's scaled MOS technologies, as much as possible of the required signal processing should be implemented at baseband, in the digital domain. Many analog transceivers used in today's digital communications systems do not make use of the fact that the underlying signal is digital, and simply utilize a conventional analog transceiver after converting the digital bit stream into an analog baseband signal. Use of such techniques as sampling demodulation and homodyne receiver architectures all present new methods in developing high-performance demodulators, which take advantage of the fact that digital data – spread-spectrum digital data – is being transmitted.

Concomitant with the goal of simplification is the desire for monolithic integration of as much of the analog circuitry as possible. Traditionally, the realm of gigahertz-band RF front-end circuitry has been dominated by discrete gallium-arsenide MESFET's using stripline filters. Such discrete-element designs consume significant amount of area on a circuit board; given that separate packages need to be utilized for each component, trace lines must be placed carefully to minimize cross-coupling, and termination resistors are required to minimize transmission-line effects from the trace. Likewise, discrete designs are poor from a power con-

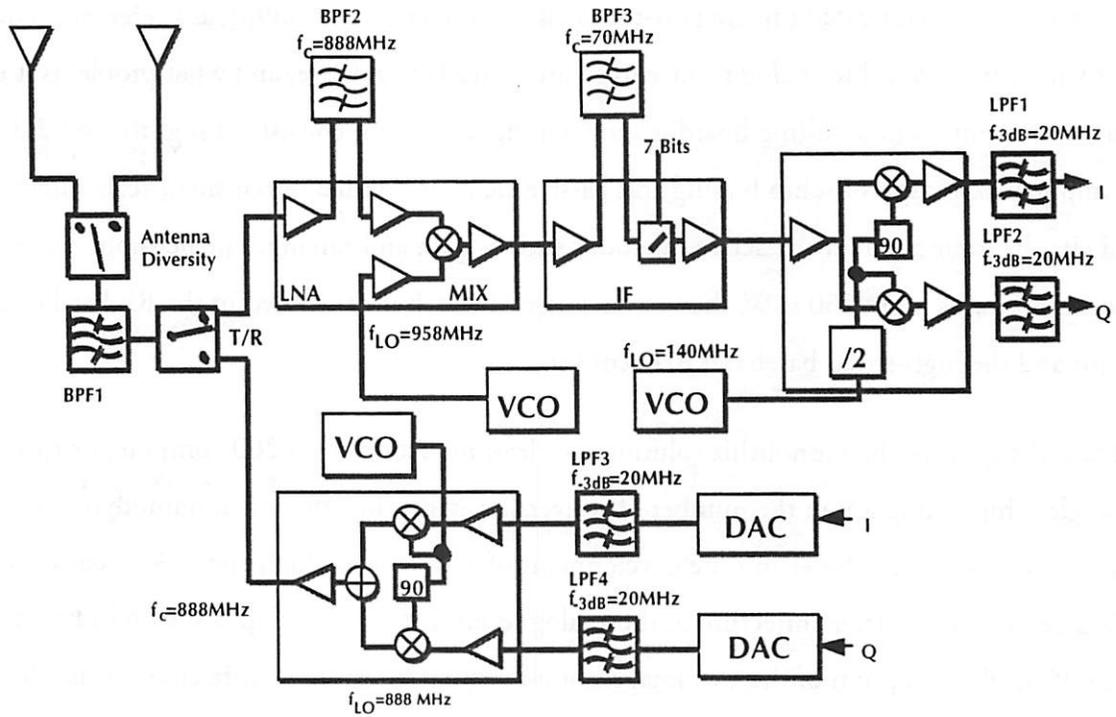
sumption standpoint, since the power required to drive high-speed analog signals across interconnect lines is directly proportional to the amount of parasitic leakage that must be overcome. Likewise, the requirements of proper termination and impedance control are necessitated at the board level; current must be expended in driving matching 50Ω loads. Since the parasitics for on-chip interconnects are at least an order-of-magnitude lower than board-level interconnects, and short on-chip propagation distances imply that termination is unnecessary, the power savings realized by monolithic circuit integration are considerable [Bakog90]. To date, most of the applications that utilize gigahertz and higher frequencies have been mainly for fixed, nonportable installations such as ground satellite transceivers, and hence power and size have not been a problem. For portable applications, low power and small size are of paramount importance, thus demanding that highly integrated analog technologies be exploited.

5.1. Receiver Architectures

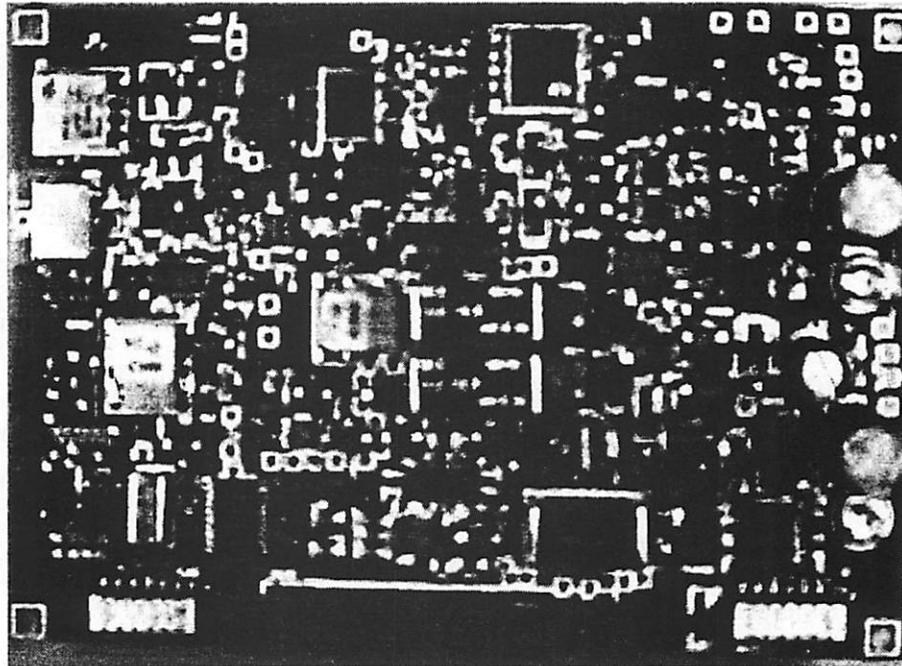
In figure 5.1a, a conventional superheterodyne RF front-end, potentially suitable for use in this system, is shown. Schematically, anti-image filtering is performed, followed by low-noise amplification and two conversion stages of mixing and filtering. The power breakdown for this schematic is as follows:

Receiver	Power (mW)
LNA/Mixer	65
IF VGA	60
Demodulator	30
RF+IF Osc	110
Buffers	100
A/D Converter: 8b 50MS/sec	770
Total Rx Power	1130

Transmitter	Power (mW)
Modulator	120
Buffers	45
D/A Converters: (8b, 40 MS/sec)	165
Total Tx Power	330



(a)



(b)

Figure 5.1: Off-the-shelf analog front-end implementation
 (a) Schematic (b) Semicustom board implementation

This design was fabricated utilizing off-the-shelf components [Behzad95], as a reference point for what the monolithic analog front-end solution needs to achieve, and what problems it must overcome. The resulting board is shown in figure 5.1b; it consists of slightly over 200 components, mostly off-chip biasing and passive elements, with a dozen small-scale integrated circuits to implement the active functions such as gain and mixing. Furthermore, power consumption is high: 750 mW, the vast majority of which is consumed in the RF local oscillator and the high-speed baseband A/D converters.

Thus, the goals of the monolithic solution are clear: reduction from 200 components to a “single” chip; reduction in the number of conversion stages, i.e., the use of homodyne conversion instead of heterodyne; and the development of a low-power, high-speed A/D converter. As a preview, the final architecture of the analog receiver front-end chip is shown in figure 5.2. With the exception of the two image/noise-rejection filters, the entire circuit is implemented on a single die, consuming a total of 108 mW. Frequency downconversion is achieved in a single stage; furthermore, it is achieved not with a conventional local oscillator-mixer configuration, but by directly subsampling the RF signal, using the harmonics of the sampling operation to achieve conversion. Power has been dramatically reduced by the architectural elimination of several components, most notably the RF and IF oscillators. In some sense, this architecture is the absolute minimum functionality an analog front-end needs to achieve: amplification, frequency translation, and A/D conversion. It depends entirely on the subsequent receiver DSP to achieve user channel selection, frequency offset compensation, and timing recovery – functions that traditionally have been done in the analog domain. To begin the discussion of this architecture, the RF aspects – low noise amplification, filtering, and homodyne sampling downconversion – will be examined in this chapter. The following chapter will examine the baseband analog issues – primarily that of high-speed A/D conversion and variable-gain amplification.

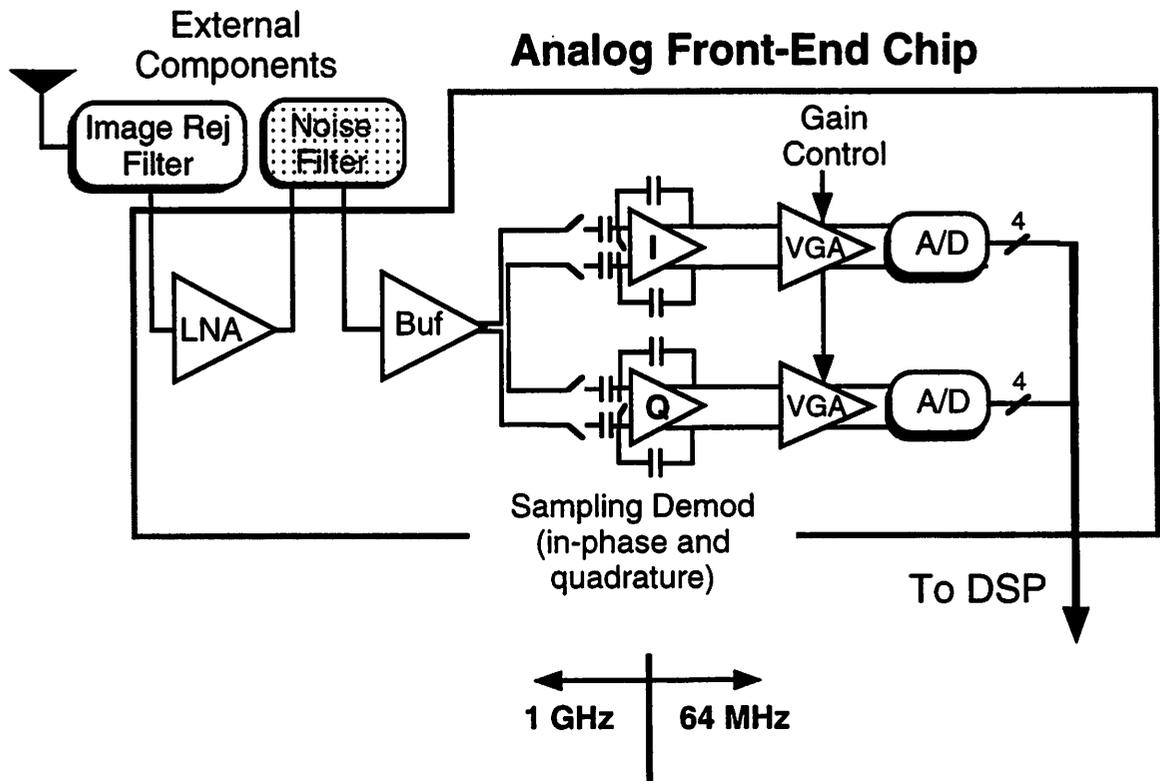


Figure 5.2: Monolithic RF Front-End Architecture

5.2. Silicon CMOS for RF Analog

Given the advances in silicon processing and technology of the past decade, digital technologies have seen a revolutionary breakthrough in both performance and size through the use of device scaling, especially in the arena of silicon complementary MOSFET's. For micro- and millimeter-wave IC technologies, gallium arsenide has been used almost exclusively to date, despite the extra cost and processing difficulties, simply because silicon devices have not possessed the necessary performance. However, the same benefits derived from scaling are reflected in analog applications as well; increases in digital switching speed commensurately yield increased analog bandwidths.

While MOS has been regarded in the past as a “slow” technology, the scaling of device feature sizes has increased MOS performance dramatically over the past decade, and will continue to do so as devices are reduced into the deep-submicron regime. In particular,

$$f_t = \frac{g_m}{C_{\text{gate}}} = \left(\mu_0 C_{\text{ox}} \left(\frac{W}{L} \right) (V_{\text{gs}} - V_t) \right) / (W L C_{\text{ox}}) = \frac{\mu_0}{L^2} (V_{\text{gs}} - V_t)$$

where V_{gs} = gate bias voltage, V_t = device threshold voltage, C_{ox} = oxide capacitance, and μ_0 = channel mobility; W and L are the width and length of the device channel, respectively.

As the channel length L is reduced by a factor λ , the f_t thus increases as λ^2 . This is borne out well by the performance scaling of MOS devices seen over the past decade. In 1980, a conventional MOS device was fabricated with a minimum drawn length of 10μ , and yielded a corresponding f_t of around 80 MHz; today, a 0.8μ channel length NMOS device can achieve an f_t of 10 GHz, over 100 times larger, and certainly practical for operation at 1 GHz. In figure 5.3, the f_t of MOS devices as a function of channel length is shown. Below 1μ , short-channel nonidealities begin to dominate, such as velocity saturation in the channel, which correspondingly degrades the transconductance g_m . Assuming complete channel velocity saturation at a carrier velocity v_{sat} , it can be shown [Watts89] that g_m and f_t are now given by $g_m = C_{\text{ox}} W v_{\text{sat}}$, and $f_t = (v_{\text{sat}} / L)$. Even under this extreme assumption, the analog device performance still scales as $1/L$ – yielding potentially 100 GHz f_t devices at a drawn L of 0.2μ . Given the additional benefit of reduction in parasitics afforded by integration, gigahertz-band monolithic CMOS circuitry is both desirable and feasible.

With the current generation of silicon technologies, the achievable f_t of conventional silicon CMOS is at performance levels required for effective operation in the low-microwave regime. In particular, with the technology available for fabrication at this time (a 0.8μ digital HP CMOS process), operation at 1 GHz can be achieved, and the choice of system carrier frequency was set by this. In future versions of this system, when non-experimental deployment needs to be considered, operation at 2.4 GHz or even 5 GHz will be desirable, necessitating even greater device scaling.

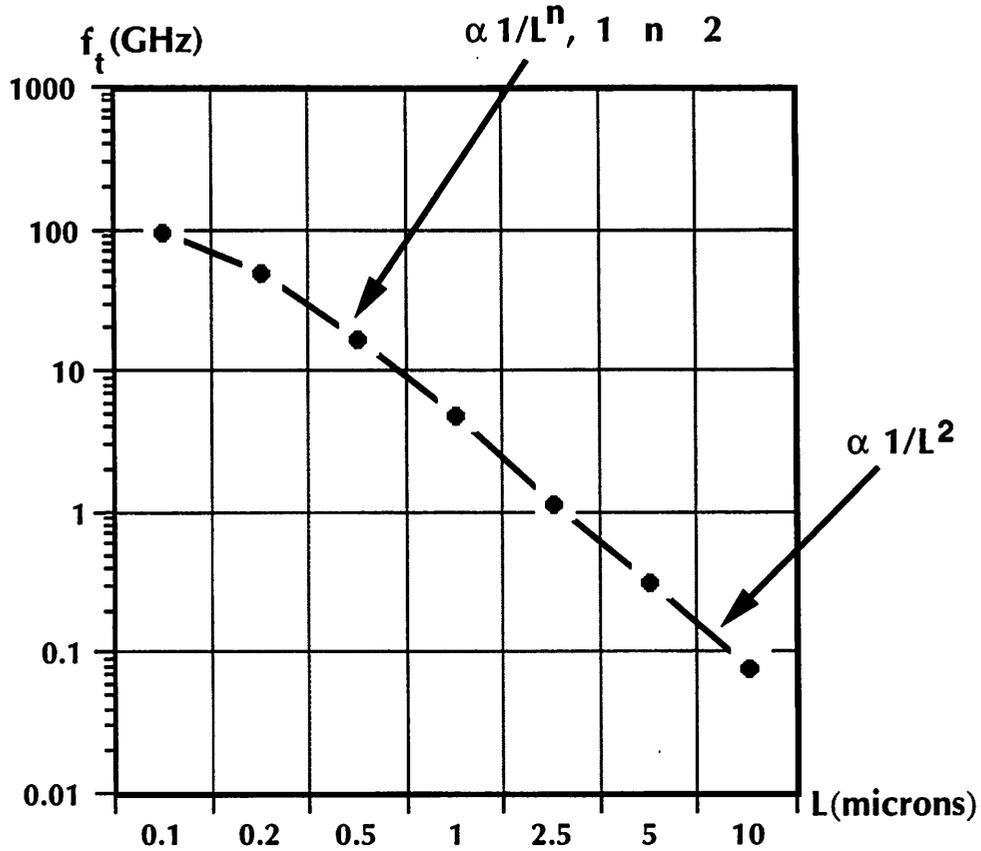


Figure 5.3: Achievable device f_t vs. MOSFET drawn L

5.3. Low-Noise Amplification

Ostensibly, one key component of any receiver chain is the low-noise amplifier coming off the antenna; since the signal is comparatively weak, good gain and noise performance is necessary. The overall noise figure of the receiver is given as:

$$NF_{\text{receiver}} = \left(\frac{1}{G_{\text{LNA}}} \right)^2 (NF_{\text{mixer}} - 1) + NF_{\text{LNA}}$$

where NF_{mixer} is the aggregate input-referred noise figure of the subsequent amplifiers and mixing stages, and G_{LNA} , NF_{LNA} are the gain and noise figure of the LNA itself. All of the noise figures in the above expression are in linear units, not in dB. The noise of all subsequent stages is reduced by the gain of the low-noise amplifier, requiring high gain; the noise of the amplifier is injected directly into the received signal, requiring low noise. There are well-

known tradeoffs in amplifiers between noise and gain; often, one is achieved at the expense of the other [Pozar90]. The inherent issues and compromises between these two will be examined below.

5.3.1. System Noise Requirements

The overall required system noise performance can be determined by examination of the transmission environment, and simulation of the system in the presence of additive noise. For the receiver, assuming a microcellular transmit power of 0 dBm and 0 dBi antenna gains (omnidirectional antennas), a 1 meter path-loss reference of 30 dB, and a path-loss coefficient of 2.6, the spatial-average received power at the boundary of a 4 meter cell is approximately -60 dBm; under worst-case fade conditions, a fading margin of -10 dB is added to this¹, implying that the receiver noise performance must recover of a -70 dBm signal, or a zero-to-peak voltage of 0.1 mV (.07 mV RMS) across a 50 Ω load.

Using these values, a plot of required noise figure as a function of QAM constellation size is shown in figure 5.4, for a constant BER = 10^{-4} under the assumption that the primary noise source is due to receiver noise and not intercell interference. Two opposing factors are seen here: the decreasing bandwidth (and hence noise power) as constellation size increases, and the decreasing distance between constellation points; the first serves to lower the BER; the second increases it. Surprisingly, the peak occurs for an extremely small constellation size – for constellations larger than this the NF must drop off quickly to maintain the BER. For the broadband CDMA system in question, a 7 dB noise figure can be tolerated, which is good from the standpoint of both CMOS implementation and low power consumption, as will be shown below. In actuality, an even higher noise figure than this may be usable: the assumption that intercell interference is negligible is not true under all conditions, especially at the cell boundaries where cells overlap. In a deep fade, if the dominant noise element is from in-

1. Typical narrowband fading margins are much larger than this; however, due to the power-averaging effect of spread-spectrum across the band and the indoor Rician fading statistics, the fading margin is much smaller. Intuitively, the peaks and deep nulls are averaged out if the transmit band is much larger than the coherence bandwidth [Viter95].

tercell interference, low-noise design in the amplifier will count for nought; error-correction coding, soft-handoff, and microcell diversity will play a far greater role in maintaining the wireless link. However, as a conservative design goal, a noise figure of 7 dB is set.

5.3.2. RF LNA Design and Matching Filter Networks

Low-noise amplifier design, and RF amplifier design in general, can be approached from one of two methods: broadband or narrowband. The broadband methodology is akin to the design of op-amps – stability, gain, and noise performance are considered across all possible frequencies, using Bode/Nyquist plots. The narrowband methodology, on the other hand, deals with the stability, gain, and noise performance of the circuit at a single frequency f_c , and the analysis is valid for a small range of frequencies around f_c – design by S-parameters and network analysis falls into this category [Gonz84]. Obviously, for the same topology, the two

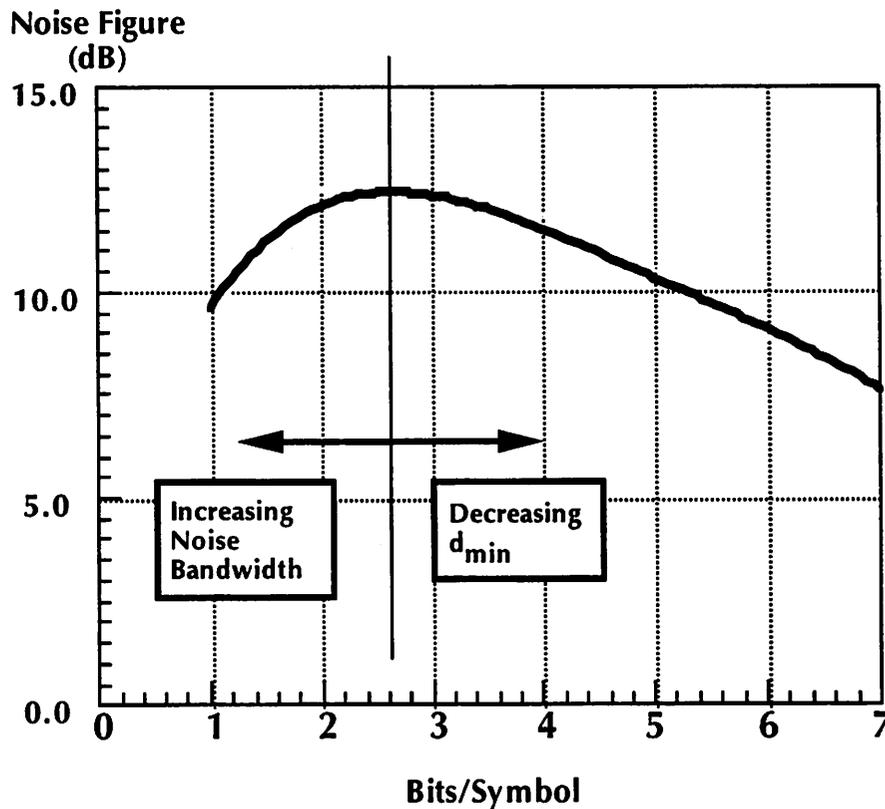


Figure 5.4: System noise figure versus constellation size, for constant BER= 10^{-4}

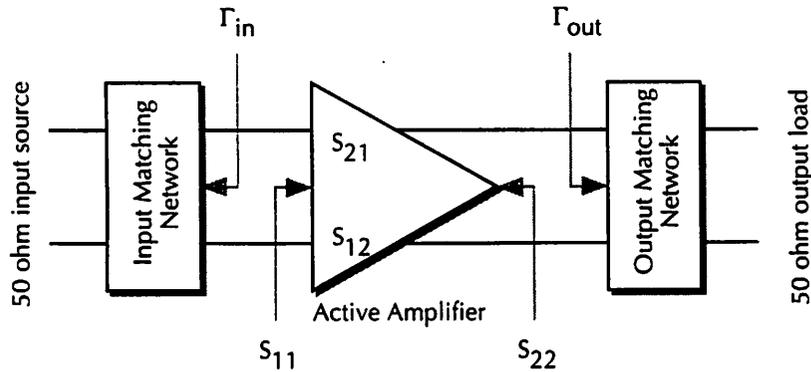


Figure 5.5: Generalized RF amplifier topology

methods must come to the same conclusion; however, when dealing with RF tuned systems, the band of interest is narrow and centered about a carrier frequency, and hence the S-parameter method is usually of greater utility.

As far as matching is concerned, it is obvious why an input match to 50 ohms is necessary in the LNA: it will ultimately be connected to the off-chip filters and the receiver antenna. However, one immediate question that arises is why an output match to 50 ohms is necessary in the low-noise amplifier. There are several reasons for this additional requirement. First, interfacing to standard test equipment is vastly simplified, since such equipment assumes 50 ohm connections at all test ports. Second, it provided additional flexibility to the architecture; such an LNA could work in a standalone mode at 1 GHz, as opposed to being incapable of driving any off-chip load. Lastly, and most importantly, the antialiasing requirements of the sampling demodulator (section 5.4) demand a second filter stage after the LNA, not only to provide additional image rejection, but to filter the broadband white thermal noise of the LNA itself. For these reasons, an output 50 ohm match was a critical design aspect of the LNA topology; as will be shown below, this also proved to be one of the most difficult to meet.

The general topology of any RF amplifier can be broken down into three stages: an input matching network, an active transducer (the amplifier proper), and an output matching network (figure 5.5). The input and output networks are passive, consisting of striplines, induc-

tors, capacitors, and resistors; their role is to achieve optimal gain and/or noise performance in the system by transforming the source/load impedances into values that are matched to the those of the transducer. For the case of maximizing gain, maximum overall power transfer is desired. Looking back from the input of the amplifier, the input matching network should exhibit a reflection coefficient Γ_{IN} equal to S_{11}^* ; i.e., the transformed source impedance should be the complex conjugate of the amplifier's input impedance (under the assumption that S_{12} is small). Likewise, the same is true for the output: Γ_{OUT} should equal S_{22}^* for maximum power transfer.

It is useful to gain some intuition as to how the matching networks, being entirely passive, can add “gain” into the system, and why maximizing power transfer is tantamount to maximizing gain. As an example, consider an ideal transducer, with finite gain $S_{21} = 10$, an input impedance of $500\text{ k}\Omega$ and a $50\text{ }\Omega$ output impedance, constant for all frequencies. As shown

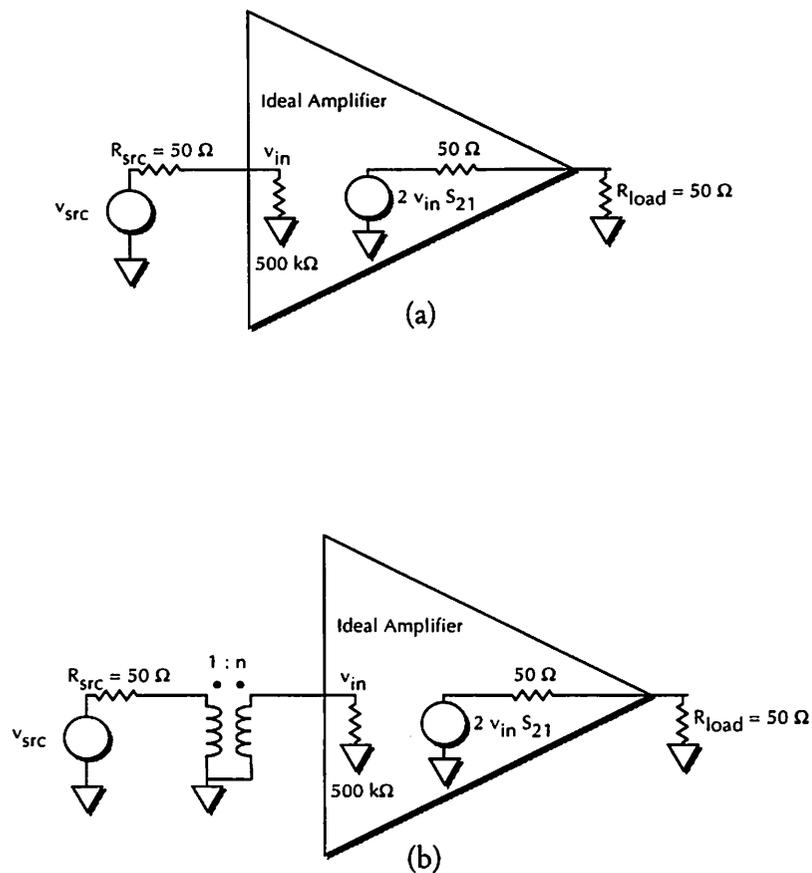


Figure 5.6: Power matching to achieve gain from passive elements. (a) Prototype case (b) Power matched case

in figure 5.6a, this amplifier could easily be driven directly by the source impedance, resulting in an overall gain equal to S_{21} . One might mistakenly equate this with the maximum gain possible – it is not. The input is accepting effectively zero power: the S_{11} of the amplifier is equal to 1.0. In figure 5.6b, the amplifier plus its ideal matching network is shown: a transformer with turns ratio equal to 100; looking back from the amplifier, it now sees an effective impedance equal to 500 k Ω . The overall gain is now:

$$(S_{21} \cdot n) \frac{1}{1 + \frac{n^2 50\Omega}{500k\Omega}} = 500$$

representing a boost by 34 dB. The numerator is the effective voltage gain of the transformer times the transducer gain; the denominator comes from the power now being absorbed by the 500 k Ω input resistance (resistor divider with the transformed 50 Ω). This is also easily seen to be the maximum gain; if the turns ratio is larger than 100, the voltage divider effect destroys any further gain. Intuitively, there are two signal-bearing variables: the voltage and the current. In the case of figure 5.6a, no current signal is being employed; the source can provide a significant current signal, but the amplifier cannot accept it. In the case of 5.6b, the voltage and current signals are both being employed by proper adjustment of the input impedance: this is the concept of power matching.

Thus, there is a maximum achievable transducer power gain [Pozar90], which is controlled not only by the S_{21} of the amplifier, but also the S_{11} and S_{22} as well. With power matched input and output networks, it is given by:

$$\text{Gain}_{\text{transducer, max}} = \frac{1}{1 - |S_{11}|^2} |S_{21}|^2 \frac{1}{1 - |S_{22}|^2}$$

under the assumption that the amplifier exhibits good reverse isolation (S_{12} is zero). In the previous example, an ideal broadband matching element was used: a transformer. Although it is possible to utilize transformers at RF, it is difficult to achieve any reasonable turns ratio, due to the interwinding capacitance limiting the frequency response of the coils. Broadband match is preferably implemented using active elements in feedback to achieve controlled impedances. However, given that the signal is inherently narrowband, another means of imped-

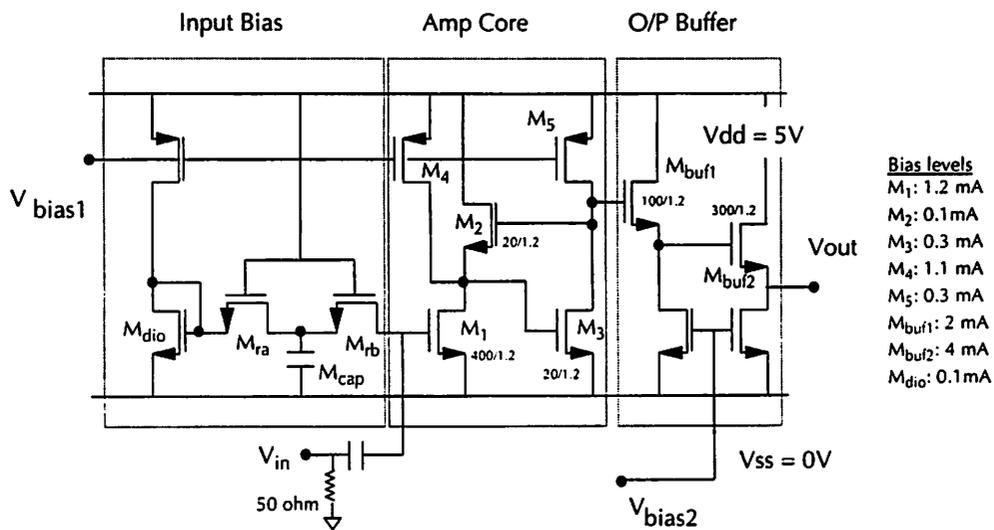
ance matching is available: transformation using resonant circuits. Furthermore, narrowband resonant matching has another advantage: the required input and output noise limiting filters can be folded in as part of the matching network. Both broadband and narrowband matching techniques will be examined below; it becomes rapidly apparent that if proper impedance matching is needed at both input and output, broadband design is simply not useful from either a low-power or high-performance standpoint for the available fabrication technologies.

5.3.3. Broadband-Matched Topology

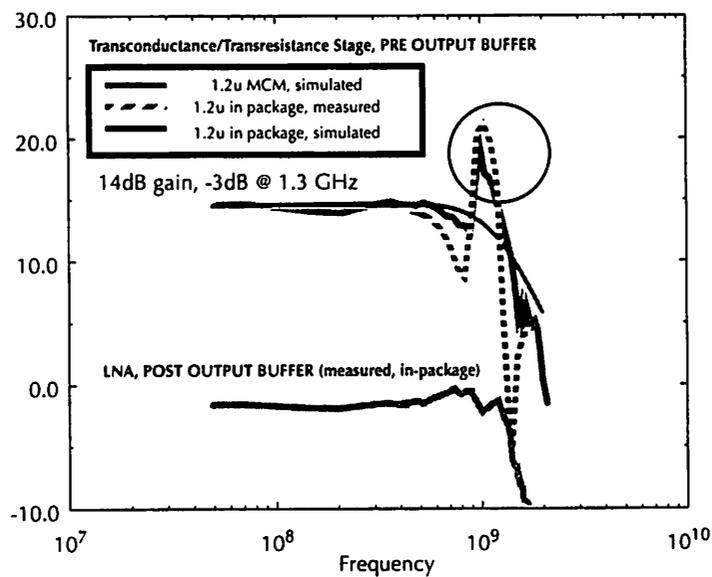
First, a low-noise RF CMOS broadband design will be examined, as a basis for a tuned design, and to explore the advantages and disadvantages inherent in such a design. A broadband design is desirable, since it can be used at any frequency in its amplification band independent of the implementation of the noise-rejection filters; the filters and amplifiers become decoupled design problems. Likewise, it should be mentioned that the original intent of this design was as a low-noise amplifier for a multichip module [MCM91], in which 50Ω match at the output was unnecessary due to the extremely short transmission distances within the multichip and the ability to formulate filters within the multichip structure itself. Given the research issues inherent in the fabrication of the multichip, decoupling the filter and amplifier designs was critical.

The proposed amplifier topology (figure 5.7a) is similar to what is done in conventional GaAs MMIC designs [LNA81]; it is a transconductance device (M_1) driving a tightly-coupled transresistance feedback pair (M_2 - M_3). The target fabrication technology is a 1.2 micron non-double poly HP digital process, and all results below are with respect to this. Although the output resistance of the transresistance stage is low (effectively $1/g_{m3} = 1.2k\Omega$), it cannot be made sufficiently low to achieve a 50Ω match; this would require M_3 to be resized to a W/L of $300\mu/1.2\mu$, with a higher bias current of 5 mA. Both of these are undesirable, since the higher W/L results in severe loading on the previous stage as well as vastly increasing the power consumption of the amp. Furthermore, given that the ultimate intent of the transconductance-transresistance design is for a multichip module, such matching is unnecessary as well. For testing purposes, a pair of source followers are needed to match this impedance to the 50Ω load; the output followers would be removed for the multichip implementation after the

core amplifier design is characterized. Two stages are used: the first buffer stage is smaller than the second, to minimize loading on the transresistance pair; the second is designed such that $1/g_m$ is equal to 50Ω



(a)



(b)

Figure 5.7: Broadband-matched LNA topology.
 (a) Schematic (b) Measured response. The top 3 curves are for the transconductance/transresistance stage by itself (“MCM-mode”); the bottom curve is the amplifier with the matching output buffers.

First, the DC bias of the entire topology must be considered, with all bias voltages being internally generated. The most interesting bias problem is establishing the base V_{gs} of M_1 , and hence its desired drain current of 1.2 mA. Given that the RF antenna input is capacitively coupled into the amplifier at this point, it is important to prevent attenuation of the RF signal, as well as prevent the input signal from being injected into the current bias chains. This is accomplished by using an R-C-R filter, formed by M_{ra} - M_{cap} - M_{rb} . M_{ra} and M_{rb} are run in triode, with a $W/L = 2.4\mu/5\mu$; their V_{DS} is zero at DC, since no current can flow through them. Thus, they behave as resistors, with resistance equal to $1/[k'(W/L)(V_{gs}-V_t)] = 8\text{ k}\Omega$. The center capacitor M_{cap} is simply a $100\mu/100\mu$ FET gate, to maximize capacitance per unit area. With respect to the diode chain, the R-C-R network forms a lowpass filter, preventing any RF signal at the input from being injected into the bias references, while permitting the DC bias established by M_{dio} to be mirrored in M_1 . This is the integrated equivalent of an RF bias tee, without the need for inductor chokes. To complete the biasing of the amplifier, standard cascoded PMOS current sources set up the bias in M_2 and M_3 (0.1 mA and 0.3 mA respectively); it is important to note that the current in M_2 is the difference between the M_1 bias and the current source. Non-cascoded current sources are used to bias the output source followers to 2mA and 4mA respectively, since headroom here is a serious issue.

As far as gain performance is concerned, the transresistance pair can be viewed as a degenerate shunt-shunt feedback, and from standard analysis has the following characteristics:

$$\text{Input resistance} = 1/(g_{m2} g_{m3} R_{o3})$$

$$\text{Output resistance} = 1/g_{m3}$$

$$\text{Transresistance} = 1/g_{m2}$$

where R_{o3} is the aggregate resistance seen at the drain of M_3 , being the parallel combination of g_{ds3} and the output resistance of the biasing PMOS cascode current source.

Thus, the overall transfer function of the amplifier can then be expressed approximately as:

$$S_{21}(\omega) = \left(\frac{1}{1 + j\omega(C_1 R_1)} \right) g_{m1} \times \left(\frac{g_{m3} - j\omega C_2}{(g_{m3} - j\omega C_2)(g_{m2} + j\omega C_2) + (g_{m2} + j\omega C_3) \left(\frac{1}{R_{op}} - j\omega C_4 \right)} \right) \times 0.5 \left(\frac{1}{1 + \chi_{buf1}} \right) \left(\frac{1}{1 + \chi_{buf2}} \right) \left(\frac{g_{mbuf1}}{g_{mbuf1} + 0.5j\omega C_{\pi buf2}} \right)$$

where:

$$\begin{aligned} C_1 &= C_{\pi 1} + C_{bondingpad} + C_{\mu 1} \\ C_2 &= C_{\pi 2} + C_{\mu 3} \\ C_3 &= C_{PD1} + C_{PD4} + C_{\pi 3} + C_{\mu 1} + C_{\pi 2} + C_{\mu 3} \\ C_4 &= C_{PD3} + C_{\mu 2} + 0.5C_{\pi, buf1} + C_{PD5} + C_{\mu 5} + C_{\mu 3} \\ R_{op} &= r_{o3} // r_{o5} \end{aligned}$$

The χ parameters are the body effect loss parameters for the FET devices, and it should be noted that capacitances C_1 through C_4 are the lumped parasitics at each node in the circuit. $C_{PD,S}$ are the parasitic drain/source capacitances, respectively, and the approximations made here are that the package parasitics are small, the source followers have operational cutoff near f_t , and the Miller effect on M_1 is negligible (since it sees the low impedance shunt feedback input). This agrees with the low-frequency gain, given by:

$$S_{21, LF} = 0.5 \left(\frac{g_{m1}}{g_{m2}} \right) \left(\frac{1}{1 + \chi_{buf1}} \right) \left(\frac{1}{1 + \chi_{buf2}} \right)$$

The -3 dB point can be extracted from this:

$$\omega_{-3dB} = \frac{g_{m3} g_{m2} R_{op}}{(g_{m3} - g_{m2}) R_{op} C_2 + g_{m2} R_{op} C_4 + C_{PD1}}$$

which is equal to:

$$\frac{1}{\frac{C_4}{g_{m3}} + \frac{C_{PD1}}{g_{m3}g_{m2}R_{op}} + \left(\frac{1}{g_{m2}} - \frac{1}{g_{m3}}\right)C_2}$$

Intuitively, this agrees with the results from zero-valued time constant analysis, with a correction for the pole splitting introduced by C_2 shunted across the gate-drain of M_3 . It is important to note that the frequency rolloff is controlled primarily by the parasitic capacitances at the gate of M_3 and at the drain of M_1 (both encapsulated in the C_{PD1} term). M_2 and M_3 are necessarily smaller devices, to minimize the impact of their parasitic capacitance; also, it is desirable that M_2 be small, since the amplifier gain increases with decreasing W in M_2 . However, M_1 is necessarily a very large device ($400\mu/1.2\mu$), to maximize gain as well as to minimize noise; thus, the parasitic perimeter and area capacitance at its drain determines the overall frequency response of the amplifier. In order to control this capacitance, a ring transistor structure is used in M_1 to achieve a large- W device while reducing drain area capacitance and eliminating drain perimeter cap. For the device sizes shown in figure 5.7a, a -3 dB cutoff was designed 1.3 GHz. In figure 5.7b, the simulated and measured frequency response of the amplifier is shown. The upper 3 curves are the gain up through the transresistance stage; the lower one is the aggregate frequency response of the amplifier, including the loss through the output buffers and packaging effects. The 3 curves for just the transresistance stage consist of the SPICE-simulated frequency response assuming an MCM package, the SPICE-simulated frequency response assuming the 52-pin PLCC package that the amplifier was measured in, and the actual measured data for the gain through the transresistance stage; agreement between the SPICE simulation and the measured amplifier response is quite good. The limitations of the output stages are apparent: an enormous amount of gain loss is incurred through body effect, coupled with the resistive dividers at the input and output (a total of nearly 20 dB of loss). The 1.2 micron process, being limited in both achievable g_m and hence f_t , simply cannot achieve a broadband 50Ω active output match without consuming excessive amounts of power as well as signal loss.

As a broadband amplifier for use in a multichip module, without the limiting output stage, the transconductance-transresistance topology could have achieved its projected +12 dB of gain, and would have been quite usable as the first stage of the receiver chain. However, the multichip module fabrication ultimately proved to be unfeasible, owing to implementation and cost factors. Given the overhead in power and performance required to broadband-match an off-chip, 50 ohm load, a narrowband redesign is called for. The amplifier core is reasonably sound, but a better form of matching – through the use of tuned elements – is needed.

5.3.4. Tuned-Matching Topology

As a second iteration, the output stage of the untuned design is replaced by a common source stage, with an inductor choke plus tuning at the output network. Simply tuning the output resistance of the broadband stage is hardly optimal; the body effect loss of the output stages will still be present. Instead, a common source output stage is used; in and of itself, it provides another stage of transconductance gain, in addition to the fact that it results in an untuned S_{22} that is nearly equal to unity, and from the maximum unilateral transducer gain equation (section 5.3.2) this maximizes the overall S_{21} gain of the stage. Likewise, input match is simi-

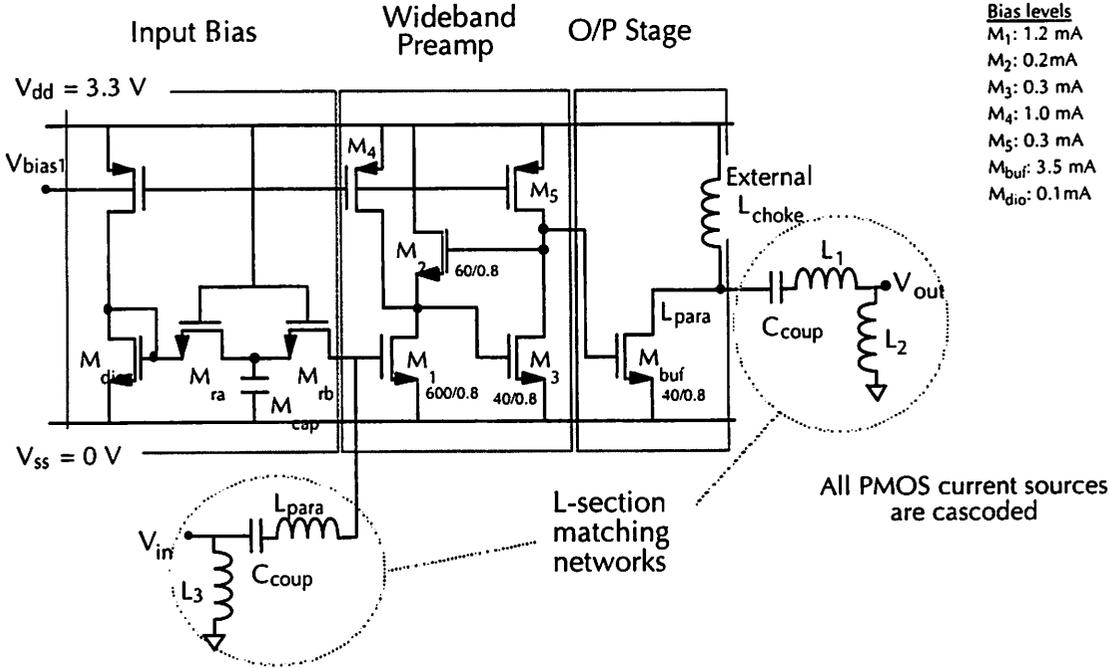


Figure 5.8: Tuned-matching LNA topology

larly implemented; instead of wasting signal power by using a broadband 50 ohm match at the input, proper termination is achieved by matching against the input bias device M_{rb} , which appears to be a passive resistor. The final configuration is shown in figure 5, along with device sizings and bias levels; the matching networks will be described in detail below.

Basically, the input and output matching networks need to match the large input and output resistances to 50 ohms, as shown in figure 5.8, while resonating the respective input and output capacitances. Interestingly, both input and output are quite similar: a large R shunted with a parallel C , plus a small series parasitic L due to the chip-on-board packaging. The chip-on-board parasitic inductance is extremely small, consisting of a bondwire contributing approximately 2 nH; specific details about the chip-on-board design can be found in the appendix to this chapter. The problem of matching such a network will be approached from two perspectives: use of standard lumped-element analysis, and the use of the Smith chart.

To examine the matching problem from the lumped-element approach, the standard L-section matching network is shown in figure 5.9. L_1 implicitly lumps the bondwire parasitic, as well as the parasitic inductance of a series AC coupling capacitor. The effective Z_{in} of the network, looking in from the 50 ohm side, is equal to:

$$R_{eq} = \frac{j\omega L_2(j\omega L_1 - \omega^2 L_1 R C + R)}{R(1 + (-\omega^2(L_1 + L_2))C) + j\omega(L_1 + L_2)}$$

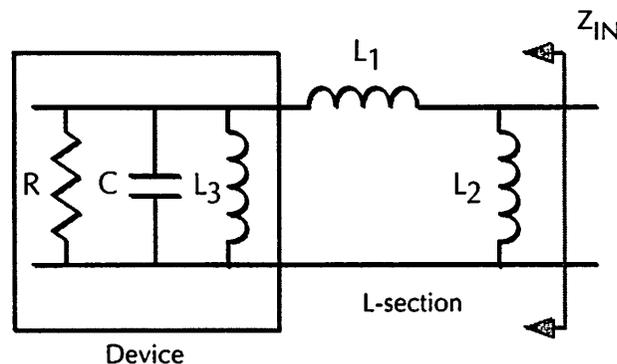


Figure 5.9: Tuned L-section, 50 ohm match
(valid for L_3 an RF choke; R large compared to 50 ohms)

The resonance point needs to be set to the desired center frequency; i.e., $1/\sqrt{C(L_1 + L_2)}$ is set equal to 1.088 GHz. At this point, R_{eq} is approximately equal to:

$$R\left(\frac{L_2}{L_1 + L_2}\right)^2 = R_{eq}$$

Surprisingly, this behaves quite similarly to a transformer, with turns ratio equal to $L_1/(L_1+L_2)$. Setting R_{eq} equal to 50Ω yields two equations in the two unknowns L_1 and L_2 .

From a Smith chart standpoint, SPICE provides the input/output reflection coefficient plots as shown in figure 5.10. From the reflection coefficients at 1.088 GHz, the desired locus to transform this to the origin of the Smith chart (i.e., properly matched to 50 ohm) is also indicated. As before, this is equivalent to a series L, followed by a shunt L to ground. Both the lumped-element method and the Smith chart yield the following values for input and output matching networks, using the 0.8μ extracted layout parameters to determine R and C:

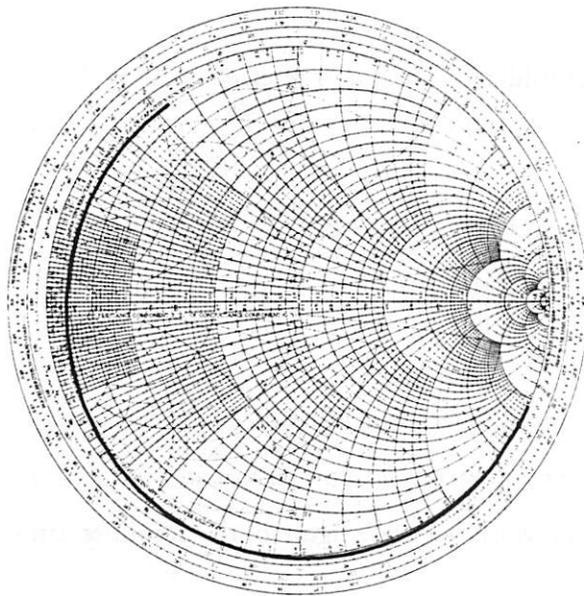
Port	R	C	L_1	L_2
Input	1.8k Ω	3.6pF	5nH	1nH
Output	3.1k Ω	1.1pF	16.6nH	2.4nH

With this in mind, the gain at resonance is approximately given by:

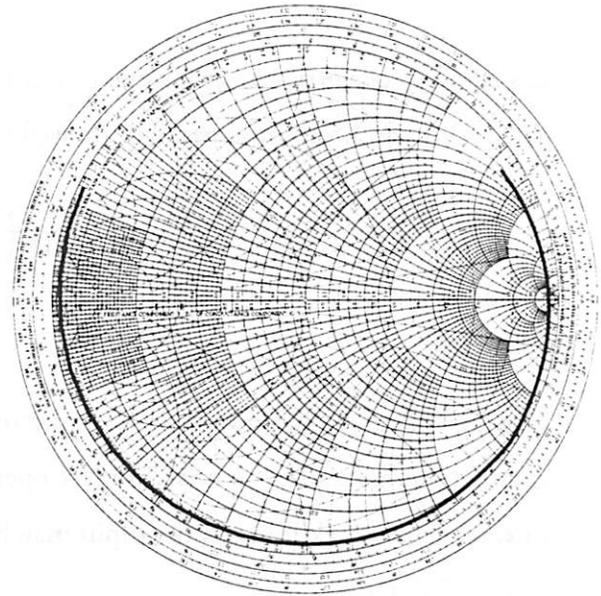
$$S_{21}(\omega_{res}) = \left(\frac{g_{m3} - j\omega_{res}C_2}{(g_{m3} - j\omega_{res}C_2)(g_{m2} + j\omega_{res}C_2) + (g_{m2} + j\omega_{res}C_3)\left(\frac{1}{R_{op}} - j\omega_{res}C_4\right)} \right) \times$$

$$0.5g_{mbuf}r_{obuf}\left(\frac{L_{2out}}{L_{1out} + L_{2out}}\right)\left(\frac{L_{1in} + L_{2in}}{L_{2in}}\right)$$

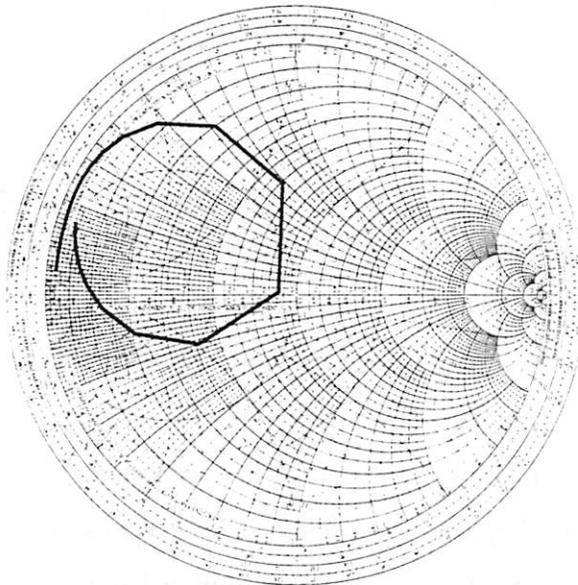
It is similar to that of the broadband case, with the final g_{mbuf} arising from the gain of the output stage. Since the device is tuned, the basic frequency-dependent terms in the equation are from the preamp circuit, which bears the only “untuned” nodes in the circuit.



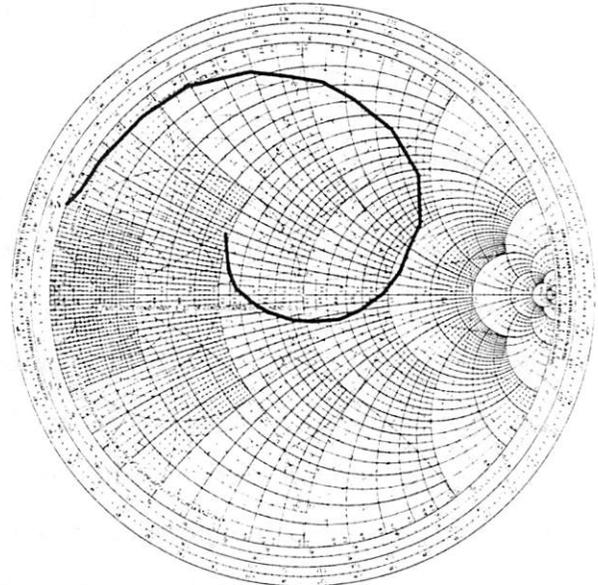
Simulated S_{11} , untuned



Simulated S_{22} , untuned



Simulated S_{11} , tuned



Simulated S_{22} , tuned

Figure 5.10: SPICE-simulated input and output reflection coefficients, tuned and untuned. Plotted from 200 MHz to 1.2 GHz

Lastly, noise performance at resonance must be considered. The input referred noise at the gate of M_1 can be approximated quite well by:

$$v_{ineq}^2 = 4kT \left(\frac{2}{3} \right) \left(\frac{1}{g_{m1}} + \frac{g_{m2}}{2} + \frac{g_{m4}}{2} \right)$$

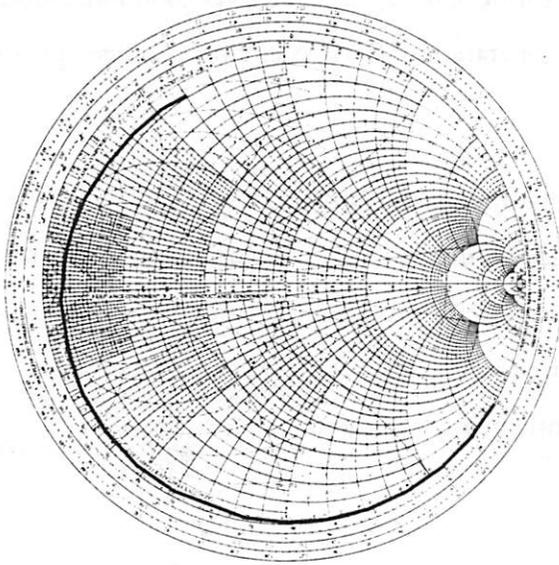
As expected, given the extremely large size of the input device, the noise is dominated by its channel thermal noise; likewise, at these operating frequencies, $1/f$ noise is not an issue. From this, reflecting back through the input matching network the noise figure of the amplifier can be calculated as:

$$NF(dB) = 10 \log \left(2 + 4 \left(\frac{L_{2in}}{L_{1in} + L_{2in}} \right)^2 \left(\frac{1}{50\Omega} \right) \left(\frac{2}{3} \right) \left(\frac{1}{g_{m1}} + \frac{g_{m2}}{2} + \frac{g_{m4}}{2} \right) \right)$$

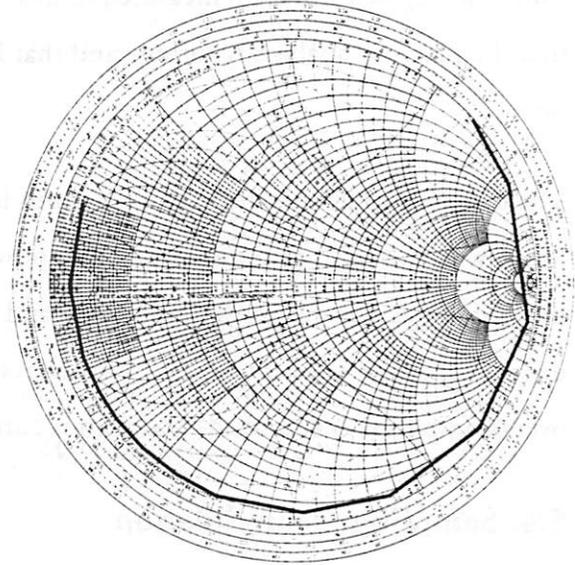
Measured Results

The design shown in figure 5.8 was fabricated and tested as part of the monolithic front-end receiver test chip; instead of the HP 1.2u process, a more advanced 0.8u HP digital process became available and was utilized. The measured S_{11} and S_{22} plots are shown in figure 5.11; they agree well with the simulated S_{11} and S_{22} plots of figure 5.10. From this, the design was tuned using off-chip L-sections; tuning of S_{11} was highly exacerbated by the need for small, high-tolerance surface mount inductors to tune the large input capacitance, leading to an imperfect match. The measured S_{21} is shown in figure 5.12, along with the simulated and measured noise figures of the amplifier; the IM_3 intercept is -10 dBm for this topology.

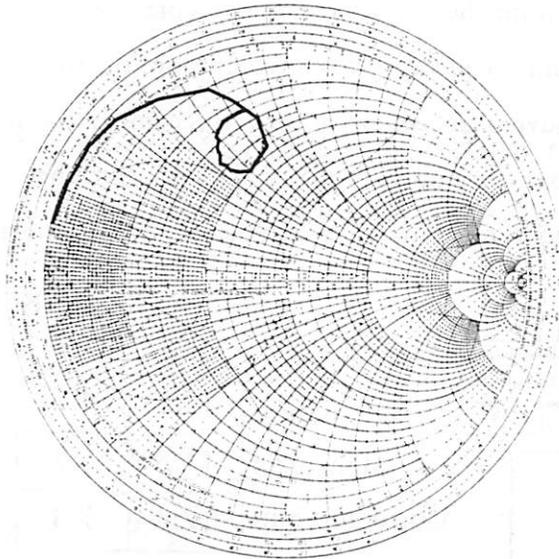
After the die was tested, it became apparent that the noise figure was approximately 3 dB high, and the gain was 3 dB low. relative to SPICE predictions. The discrepancy was tracked down to the input transconductance device of the LNA; the use of the ring structure resulted in an effective device W that was approximately one-half of what SPICE and hand calculation were utilizing, due to corner fringing effects. Using a corrected W model for the LNA, SPICE



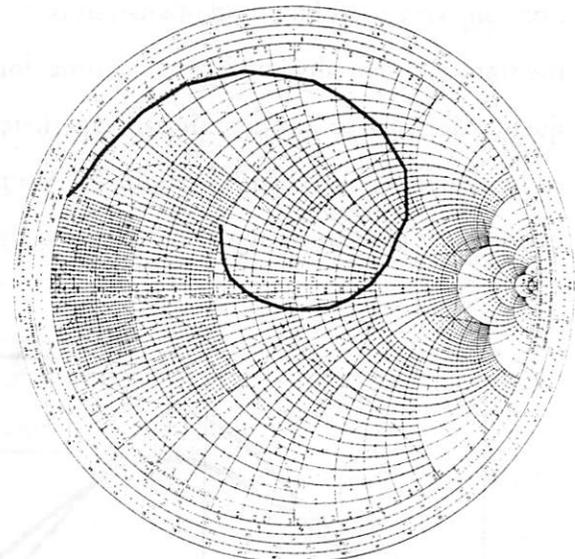
Measured S_{11} , untuned



Measured S_{22} , untuned



Measured S_{11} , tuned



Measured S_{22} , tuned

Figure 5.11: Measured input and output reflection coefficients, tuned and untuned. Plotted from 200 MHz to 1.2 GHz

simulation agreed well with measured results. Regardless, the agreement gives good indication that the previous analysis is correct, and that RF operation of CMOS circuits is indeed possible.

The conclusion that can be drawn from this is that the preamp structure provides very little in the way of overall gain, relative to that provided by the matching network. In some sense, most of the gain is due to the large input and output resistances; proper tuning allows one to deliver maximum power to these resistors, resulting in maximum voltage gain through the overall amplifier. A corrected-W design is currently being fabricated.

5.4. Sampling Demodulation

Beyond amplification, the second major function required by any receiver front-end is that of frequency conversion of the received RF signal to baseband. When considering demodulation, one very useful method of analysis is to examine the frequency domain representation of the signals; essentially, mixing in the time domain is equivalent to convolution in the frequency domain. In a conventional (superheterodyne) frequency conversion scheme, illustrated by the upper signal path in figure 5.13, a mixer driven by a local oscillator equal to the carrier ω_c is used to bring the bandlimited signal from RF down to baseband (of course, in a

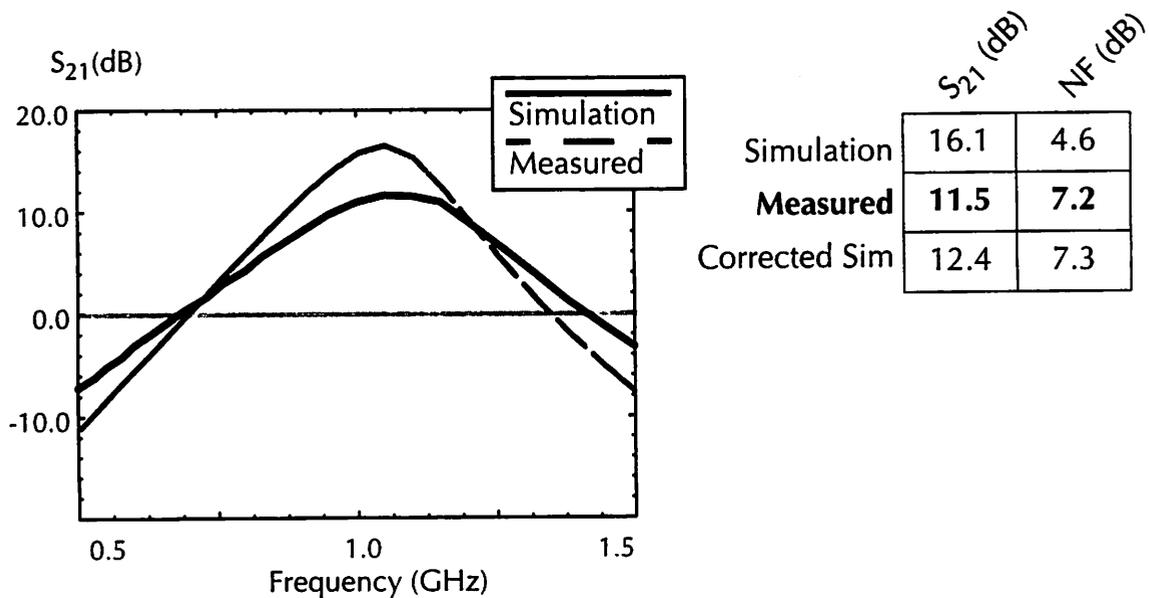


Figure 5.12: Measured and simulated amplifier S_{21} , noise figure

superheterodyne scheme, several stages of mixing may be employed). The resulting baseband signal is then sampled at its Nyquist rate ω_{samp} . Mathematically, if $F(\omega)$ is the original (double-sided) baseband spectrum, bandlimited to $\omega_b = \omega_{\text{samp}}/2$, then the RF received signal can be expressed as:

$$\frac{1}{2}(F(\omega - \omega_c) + F(\omega + \omega_c)) = F_{\text{mod}}(\omega) \quad (\text{eq. 1})$$

After superheterodyne mixing and Nyquist-rate sampling, the resulting discrete-time signal has aliases at *all* multiples of the sampling frequency ω_{samp} , given by:

$$\left(\frac{2\pi}{\omega_{\text{samp}}}\right) \sum_{i=-\infty}^{\infty} F(\omega - i\omega_{\text{samp}})$$

However, mixing followed by sampling need not be the only solution for achieving this. Suppose that the received signal $F_{\text{mod}}(\omega)$ (eq. 1) is bandpass subsampled at the Nyquist rate ω_{samp} of the baseband signal. After such sampling, the spectrum of the resulting discrete-time signal has a transform equal to:

$$\frac{1}{2} \sum_{i=-\infty}^{\infty} (F(\omega - \omega_{\text{carrier}} - i\omega_{\text{samp}}) + F(\omega + \omega_{\text{carrier}} - i\omega_{\text{samp}}))$$

Making the important stipulation that $\omega_{\text{carrier}} = k \cdot \omega_{\text{samp}}$, where k is an integer, this sum can be simplified to:

$$\left(\frac{2\pi}{\omega_{\text{samp}}}\right) \sum_{i=-\infty}^{\infty} F(\omega - i\omega_{\text{samp}})$$

which is precisely the same result achieved by the original superheterodyne demodulation structure. The entire demodulation step has been reduced to a single sampling operation, sampled at the Nyquist rate for the *baseband* signal. A graphical depiction of this process is shown in figure 5.13; the path indicated by the solid black arrows shows the result of each

step of a conventional mixer-sampler structure, and the single path indicated by the large gray arrow is the result of bandpass sampling the modulated signal. Clearly, both will yield the same aliased spectrum provided $\omega_{\text{carrier}} = k \cdot \omega_{\text{sample}}$.

Another interpretation of this result can be seen as follows. The frequency-domain equivalent of sampling is the convolution by a stream of Dirac delta functions, i.e., the effect of sampling the modulated signal with transform $F_{\text{mod}}(\omega)$ yields a signal with a Fourier transform equal to:

$$\left(\frac{\omega_{\text{sample}}}{2\pi} \right) \int_{-\infty}^{\infty} \left\{ F_{\text{mod}}(\omega - \Omega) \sum_{i=-\infty}^{\infty} \delta(\Omega - i\omega_{\text{sample}}) \right\} d\Omega$$

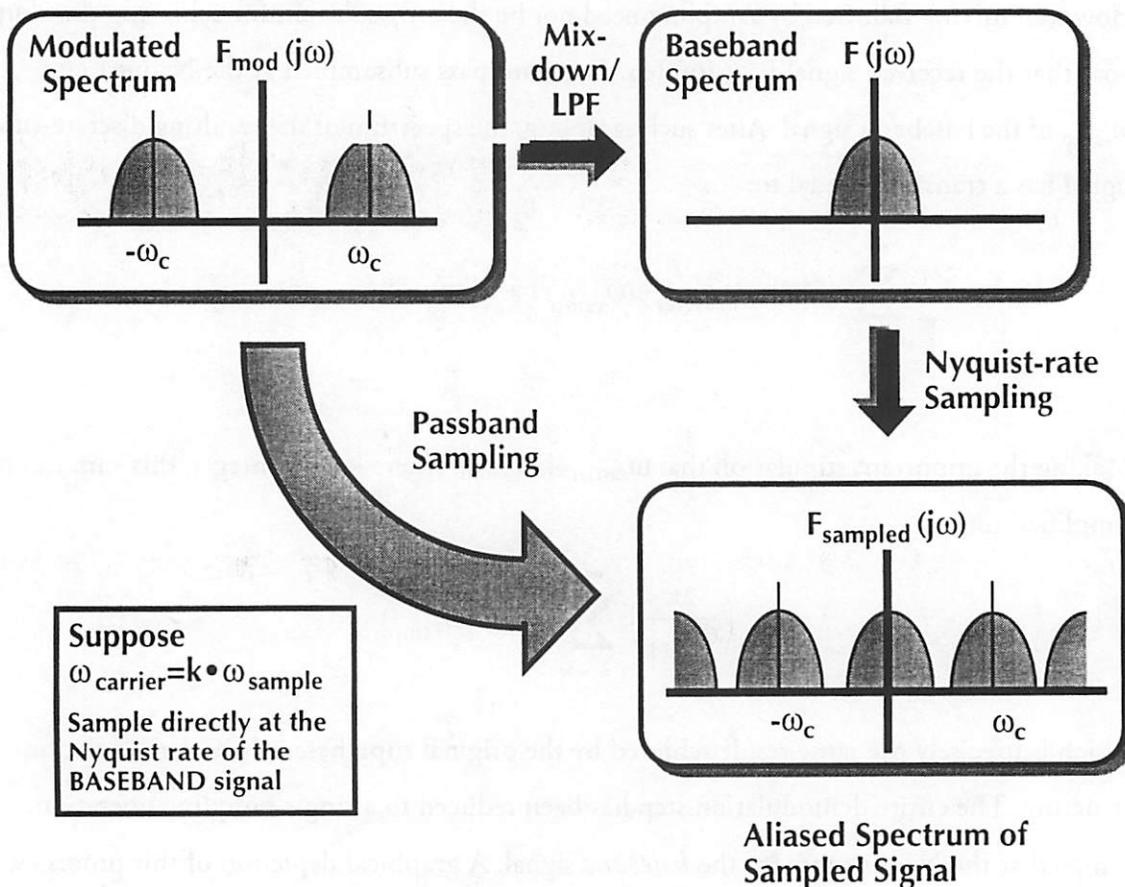


Figure 5.13: Sampling Demodulation

Alternatively, this is the same as multiplication in the time domain of the received signal by a sum of equal-power cosines at all harmonics of the sample rate; using the following identity:

$$\sum_{i=-\infty}^{\infty} \delta\left(t - 2\pi \frac{i}{\omega_{\text{samp}}}\right) = \left(\frac{\omega_{\text{samp}}}{\pi}\right) \sum_{i=0}^{\infty} \cos(i\omega_{\text{samp}}t)$$

sampling $F_{\text{mod}}(\omega)$ can be equated to:

$$f_{\text{mod}}(t) \times \sum_{i=0}^{\infty} \cos(i\omega_{\text{samp}}t)$$

Provided proper bandpass anti-alias filtering is done to isolate the modulated signal at $\omega_{\text{carrier}} = k \cdot \omega_{\text{samp}}$, the required demodulation-and-sampling is thus yielded. Essentially, the mixer has been shown to be redundant, since the sampling operation itself can easily perform the same mixing function.

Of importance is the fact that the sampling rate does *not* change frequencies, even as the carrier changes. The k in the above stipulation was not specified; so long as the carrier frequency is an integer multiple of the baseband sampling rate, this method works. Hence, a voltage-controlled oscillator is not needed here, and a high accuracy crystal oscillator can be used instead. The hardware cost is minimal: an accurate switch, implementable using standard MOS technologies, and a fixed-frequency crystal oscillator, which is the same oscillator as used to reference the transmitter synthesizer PLL. In terms of power, the demodulator is passive, dissipating energy only in the crystal oscillator and the resistive channel of the MOS device, and is certainly much lower than an equivalent full mixer/oscillator configuration.

Theoretically, the sampling demodulator is an excellent implementation for the demodulator in terms of hardware and power, utilizing an active device as a simple passive switch. However, before any judgment of real usefulness is passed, its practicality for use in a transceiver for personal communications must be assessed: it must be able to demodulate a signal centered at $\omega_c = 1.088$ GHz, while sampling in excess of $\omega_{\text{samp}} = 128$ MHz for timing recovery. The above analysis was performed symbolically; the specific choice of carrier frequency and sampling rate will be examined below.

5.4.1. Quadrature Recovery and the Choice of System Carrier Frequency

Since the sampling introduces mixing with equal-power cosines, the in-phase signal is easily recovered, as in the above derivation. However, the quadrature phase requires a sine wave in the mixing operation to successfully recover the information. Clearly, some more work is needed than just the simple sampling step, since the quadrature data will be lost. One immediate method is to utilize an additional sampling operation, shifted in phase by $\pi/2$. By doing so, the time-domain representation of this additional sampling operation is:

$$f_{\text{received}}(t) \times \left\{ \left\{ \sum_{i=0}^{\infty} (-1)^i \cos(2i\omega_{\text{samp}} t) \right\} + \left\{ \sum_{i=0}^{\infty} (-1)^i \sin((2i+1)\omega_{\text{samp}} t) \right\} \right\}$$

The proper mixing is yielded if the condition is made that $\omega_{\text{carrier}} = k \cdot \omega_{\text{samp}}$, and k is odd. Otherwise, the second sampling step simply yields the in-phase signal again. Thus, use of such a system with quadrature modulation requires only a second sampling unit offset by a simple 90° phase shift.

If the parameters of the system are considered - a 64 Mchip/sec DQPSK spread-spectrum signal, requiring a minimum sampling rate of 128 MHz to perform timing recovery - the above stipulation that k be odd places some severe restrictions on the choice of carrier frequency. In particular, near 1 GHz the only possibilities are 896 MHz and 1.152 GHz. Eliminating this restriction is critical, since neither of these frequencies is usable for the system. Instead, a carrier of 1.088 GHz is chosen, with a pair of interleaved 128 MHz baseband sampling switches. Since interleaving is performed, there is a 180° phase relationship between the two switches, instead of 90° as described above. Given that a noninteger, 180° phase relationship is established between the carrier and the baseband sampling rate, the impact of this on the demodulation process must be considered.

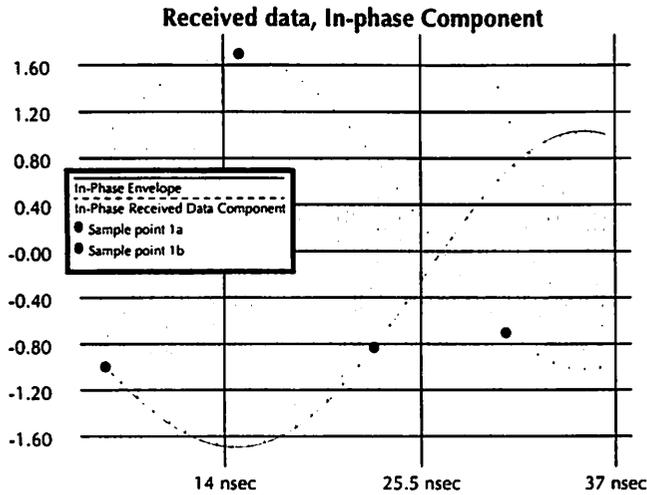
Because of the interleaving, the aggregate effective sampling rate of the pair is 256 MHz; relative to 1.088 GHz, the demodulated output should have a frequency offset equal to $1.088 - 1.024 \text{ GHz} = 64 \text{ MHz}$. As with the premodulation of the FIR filter in chapter 4, a 64 MHz cosine wave sampled at 256 MHz yields a sequence $\{\dots +1, 0, -1, 0, \dots\}$, whereas a 64 MHz

sine wave sampled at 256 MHz yields a sequence { ... 0, +1, 0, -1, ... }. Thus, the interleaving is correct: due to the zero samples, one sampler output will have entirely in-phase information, with alternating sign bits due to the +/-1, and the other will have entire quadrature information. In figure 5.14, the demodulation process is illustrated graphically. The received modulated in-phase and quadrature data streams are shown, with the dotted lines indicate the envelope of the baseband data and the sample points marked for each demodulator. Clearly, one sampler recovers the in-phase data, whereas the other recovers the quadrature. The sign alternation, or demodulation from 64 MHz down to baseband, can be removed digitally by simply toggling the sign bit after A/D conversion.

There are two important points that need to be made:

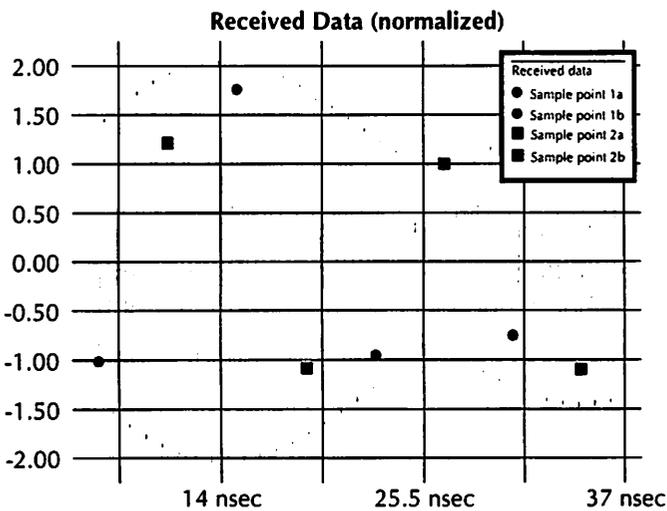
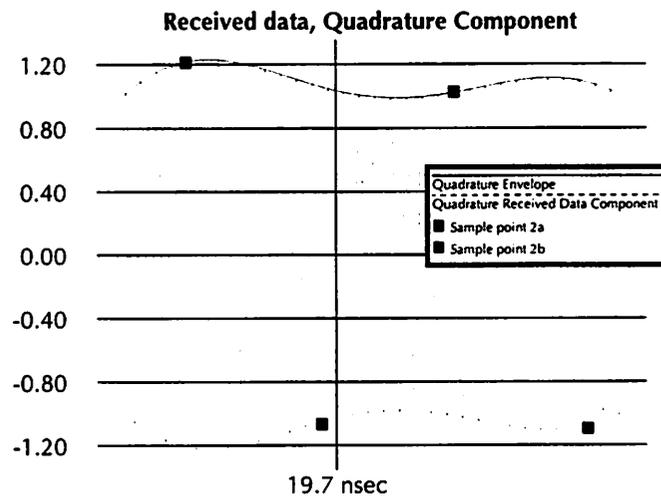
- Since the final 64 MHz demodulation is done in the digital domain, there is clearly no DC information content in the analog signal. Circuit impairments such as $1/f$ noise and DC offsets in the analog circuitry can be filtered out without affecting the recovered signal.
- The demodulator will have a pair of interleaved sampling switches, one for in-phase and one for quadrature. However, their behaviour is symmetric: there is nothing intrinsic in the architecture to determine that one sampling path is “in-phase” or one sampling path is “quadrature”. Such assignments can only be made after timing recovery has been performed in the digital CDMA receiver. This fact is utilized heavily in the design of the digital CDMA receiver timing recovery loop: in particular, no timing-recovery feedback is necessary between the digital and analog sections of the receiver, resulting in several significant hardware reductions (chapter 7).

Lastly, some intuition should be forthcoming, given the elegance of the result. In actuality, 1.088 GHz does maintain an odd-integer, 90° phase relationship with the baseband signal: it is an odd multiple of 64 MHz. If samples 1a and 2a of figure 5.14 are considered exclusively as the output, it is apparent that these will simply be the original data signal, in-phase and quadrature. The addition of output samples 1b and 2b provides in-phase and quadrature data



(a) In-phase received data component; I-sampling demodulation points are marked at 1/128 MHz

(b) Quadrature received data component; Q-sampling demodulation points are marked at 1/128 MHz



(c) Received data: both I/Q sampling instants are marked (interleaved 128 MHz, effective 256 MHz)

Figure 5.14: Time-domain view of sampling demodulation. Note the recovery of both in-phase and quadrature data components, and the timing relationships between samples 1a and 2a.

outputs at another sampling phase for timing recovery; because of the carrier modulation, those other samples will be 180° relative to samples 1a and 2a, requiring a simple digital sign toggle to correct.

5.4.2. Bandpass Antialiasing

Furthermore, careful consideration needs to be given to the image rejection filtering at RF. Since the received signal is being subsampled, this image rejection filtering now performs an antialiasing function above and beyond the typical anti-image function normally found in RF transceiver systems. The need for this is clear: the demodulator will perform frequency conversion at any integer multiple of ω_{sample} , not just the desired one. The 128 MHz oversampled interleaving described in the previous section has a second benefit: as in baseband analog systems, oversampling decreases the antialiasing requirements in the system: the nearest frequency images that need to be filtered from the system are a 960 MHz and 1.216 GHz, as opposed to 1.024 GHz and 1.152 MHz. Likewise, the broadband thermal noise introduced by the LNA will be spectrally folded if it is not rejected before the sampling switch. As shown in figure 5.2, two filter stages are used, on either side of the LNA. Not only do two filtering stages increase rejection, but the post-LNA placement of the second filter rejects the out-of-band thermal noise introduced by the LNA itself. Unfortunately, since neither multichip module packaging nor on-chip inductors were available, the second filter stage must be off-chip, nominally matched to a 50 ohm impedance level from the LNA. Given that a typical RF ceramic filter response [Murata93] has the characteristics as shown in figure 5.15, allowing an extra 64 MHz in the antialias rolloff translates to another 30 dB (for two stages) in adjacent channel rejection at the first alias frequency, critical if the receiver is to coexist with high-power cellular phones in the 800 MHz band, pagers in the 930 MHz band, and cordless telephones between 902 and 928 MHz.

Lastly, the overall system design provides additional rejection filtering against interference from aliased images. As described in chapter 2, the intrinsic property of spread-spectrum is that of rejecting in-band additive noise: exactly the problem posed by spectral aliasing. With respect to the signal post-correlation, the effective impact of the aliased images has been lessened by the processing gain. In other words, the filter skirts in figure 5.15 are effectively

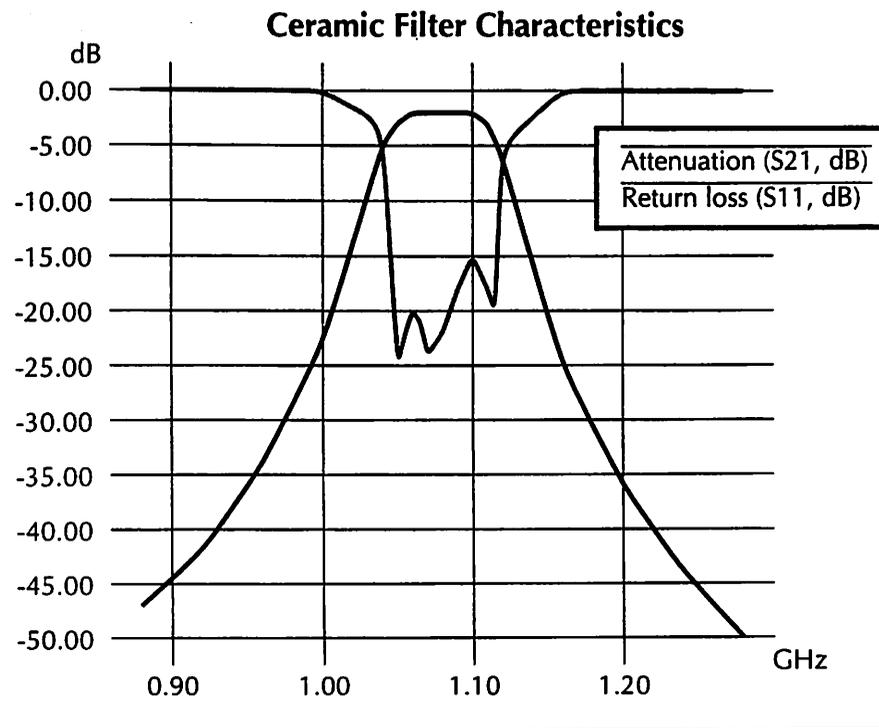


Figure 5.15: RF ceramic filter attenuation and return loss characteristics [Murata93]

18 dB lower than what would be found in a non-spread-spectrum system employing sampling demodulation. The ultimate choice of using spread-spectrum in this system was hinged upon the additional filtering gain provided with respect to aliasing in the sampling demodulator: another example of the impact that coupling analog, digital, and system design had on the overall wireless system architecture.

5.4.3. Frequency Stability and Incoherent Demodulation

To ease the hardware requirements, the necessity of zero phase and frequency offsets must be carefully considered. If possible, carrier recovery should be avoided at all costs, which leads to the requirement that very accurate frequency references are needed in both the transmitter and receiver, and immunity to nonzero phase offsets.

First, the assumption that $\omega_{\text{carrier}} = k \cdot \omega_{\text{samp}}$ in the transmitter must be examined. Surprisingly, this is not a difficult requirement for the hardware; as discussed above, a digital PLL synthesizer automatically achieves this condition, and does so with extreme accuracy when

referenced to a crystal oscillator. Second, although the accuracy of crystals is extremely good (typically 1 part in 10^5), it is not perfect: the transmitter and receiver will be frequency (and phase) offset by some amount. From the data sheets on the crystals to be used in this system [RFM90], worst-case frequency mismatch between transmitter and receiver is specified as 25 kHz; from physical measurements on pairs of crystals, this is in actuality quite pessimistic: in the sample set available, the worst-case measured offset was actually 3 kHz. Using the 25 kHz value, for a baud rate of 1 MHz this translates to a worst-case phase error of 9° from baud-to-baud, assuming no carrier recovery in the receiver whatsoever - a very small error, considering the impediments of intercell and intracell interference in the overall system.

The simplest solution to this problem is to utilize a phase-differential QAM scheme, which encodes the transmitted information on the transition between constellation points instead of the constellation points themselves - hence the choice of DQPSK for the system modulation format. For voiceband systems, carrier recovery is absolutely necessary: 25 kHz is an enormous offset relative to a 32 kHz voice channel. For high-speed data communications, 25 kHz is minuscule compared to the 1 Mbaud channel signalling rate, and simplistic solutions such as differential phase encoding are viable, leading to both power and complexity minimization. This is an example of how a digital encoding at baseband allows the elimination of costly analog hardware; removal of the absolute phase dependence requirement is achieved with only a minor expense in digital processing, and no expense in bandwidth.

Lastly, the phase noise of the crystal oscillator itself must be considered. Such jitter can be modeled as a phase error in the recovered/sampled output signal, given that it arises from random fluctuations in both the transmitter and receiver crystal output frequencies. As a first-order analysis, supposing that the signal to be sampled is given by $A_v \cdot \cos(2\pi \cdot 1.024 \cdot 10^9 t)$, sampling at 128 MHz with zero phase difference should yield a constant (DC) output. To compute the impact of phase noise on the recovered signal, the worst-case will be at the point of maximum slew, or at the zero crossing of the sinusoid (figure 5.16). Theoretically, the output should be constant zero; instead, it yields some random output noise signal, whose amplitude can be approximated as being linearly related to the jitter, resulting in an SNR due to jitter noise equal to $20 \log\left(\frac{2}{\omega_c \tau_{\text{jitter}}}\right)$.

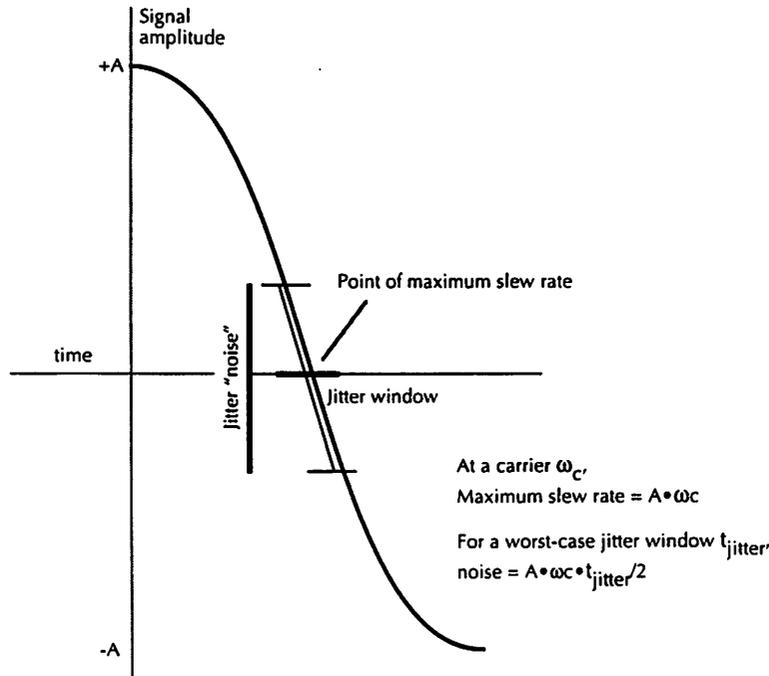


Figure 5.16: Translation of phase noise into additive signal noise in the demodulation process.

Thus, a worst-case estimate on the allowable jitter window can be realized. For a sampling jitter SNR of 30 dB¹, at 1.024 GHz with a 128 MHz sampling rate, this translates to an allowable jitter of +/- 28 psec about the true sampling point, or approximately 1 part in 300.

Although it seems like a simple matter for a crystal oscillator to achieve this, the design challenge arises in maintaining this accuracy through the necessary on-chip clock buffering, especially in the presence of supply bounce, substrate injection, and capacitive coupling effects.

1. The linearity of the A/D converter is only required to be 4 bits from system simulations, due to the spread-spectrum nature of the signal; see chapter 6. The sampling jitter SNR is chosen to be 6 dB beyond this, hence the 30 dB requirement.

5.5. MOS Sampling Demodulator Design

Of course, the sampling described in the above theory is an idealization; no device can ever achieve “delta-function” precision in sampling, since it takes a finite amount of time for the sampling switch to close. In this section, the practical aspects of designing such a demodulator will be addressed, from the perspective of switch track bandwidth, noise performance, and distortion. The sample-and-hold (S/H) itself must have be able to track input waveforms with a bandwidth exceeding the carrier frequency, which lies between 1 and 2 GHz, For a silicon MOS sample-and-hold, there are several significant nonidealities that must be considered. In figure 5.17, a basic MOS S/H is displayed, along with a plot of a signal to be sampled and the output of the S/H. Essentially, the MOS device is a simple on-off switch, passing charge to the sampling capacitor when in track mode, and preserving the charge when in hold mode.

First, when the S/H is in track mode, the switch is closed and the voltage across the sampling capacitor should ideally equal the input voltage. However, the drain-source resistance R_{ds} of the MOS device is nonzero, equal to:

$$R_{ds} = \frac{1}{\mu C_{ox} \left(\frac{W}{L} \right) (V_{gs} - V_t)}$$

where the device parameters are the same as described in the discussion of f_t above. The effect of R_{ds} is to create an R-C lowpass filter with C_{sample} , hence limiting the bandwidth that the system can track accurately. Likewise, the acquisition time t_{acq} needed to switch from hold to track mode is determined by this lowpass filter, and thus $t_{acq} = (1/\text{Track BW}) = R_{ds} \cdot C$. For a sampling demodulator, this must be well in excess of the desired carrier bandwidth, requiring that both R_{ds} and C be significantly smaller than what would normally be done in a sampling array. Both of these implications are rather dire. A small R_{ds} implies that the sampling switch needs to be large for a fixed supply voltage, exacerbating the charge injection and feedthrough effects described below. What is more important is that a small C results in a proportionally larger kT/C noise, impacting the effective noise figure of the demodulator. Both of these will be discussed in detail with regard to the final design of the sampling demodulator.

Second, as the switch moves from track mode to hold mode at the sampling instant, two mechanisms enter to degrade the signal: feedthrough from the sampling clock, and charge injection from the switch itself. Feedthrough from the sampling clock is a result of the gate overlap capacitance C_{ovl} and the sampling capacitor forming a divider; if the sampling clock edge is falling from a supply voltage V_{dd} to ground, this results in a sampling error equal to:

$$\Delta V_{\text{feedthrough}} = -\left(\frac{C_{\text{overlap}}}{C_{\text{sample}}}\right)V_{dd}$$

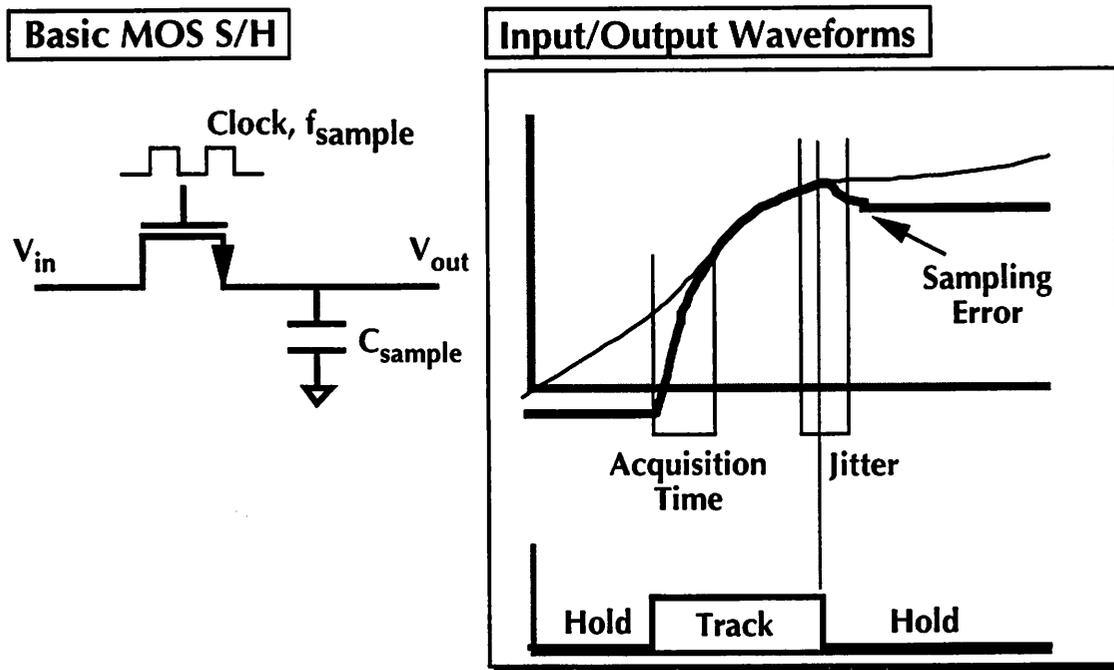


Figure 5.17: MOS Sample-and-Hold Characteristics

Likewise, charge stored in the channel of the MOS switch must be removed, and hence is injected directly onto the sampling capacitor. Assuming that the charge divides evenly between the source and drain of the switch, which is reasonable given the fast switching edge necessary for demodulation, this results in a sampling error ΔV equal to:

$$\Delta V_{\text{charge}} = \frac{-Q_{\text{chan}}}{2C_{\text{sample}}} = \frac{-WLC_{\text{ox}}(V_{\text{dd}} - V_{\text{in}} - V_{\text{t}})}{2C_{\text{sample}}}$$

It is important to note that both of these effects are inversely proportional to C_{sample} , and proportional to W . Given that W needs to be large for a small R_{ds} , and C_{sample} needs to be small to achieve the track bandwidth, the system must be able to tolerate a large ΔV at the sampling instant.

Of course, the above discussion is highly dependent on the fabrication technology. Ignoring the constant sampling error due to clock feedthrough for the moment, the above analysis yields the classical figure-of-merit for a MOS device [Gray89], known as the speed-accuracy product, equal to $\Delta V_{\text{charge}} \cdot \tau_{\text{acq}}$, and can be simplified to $2L^2/\mu$. Since minimizing this is the key goal, and like f_r is proportional to the square of the channel length, as devices scale the performance of MOS devices improves dramatically. For 4-bit accuracy on a 1 V swing, a track bandwidth of 1.8 GHz, and $\mu = 200 \text{ V}/(\text{cm}^2 \text{-sec})$, a minimum drawn L of 0.25 microns is required.

Obviously, a 0.25 micron process is not readily available at this time, and would thus seem to preclude the use of direct subsampling as a means of RF demodulation from a carrier of 1 GHz. However, the above result is quite pessimistic: it assumes that the charge injection pedestal ΔV onto the capacitor is entirely signal dependent, with no simple means of elimination. An improved sampling switch topology is shown in figure 5.18; a technique known as bottom-plate sampling is employed to remove the signal dependence in the error term. The operation can be understood as follows: At first, S_1 and S_2 are closed, allowing the input signal V_{in} to be sampled onto the capacitor C_S . When S_1 and S_2 are opened, an amount of charge proportional to the signal voltage and to the size of capacitor C_S is left floating on C_S exactly as in the previous circuit. However, if S_2 is opened slightly before S_1 , then the S_2

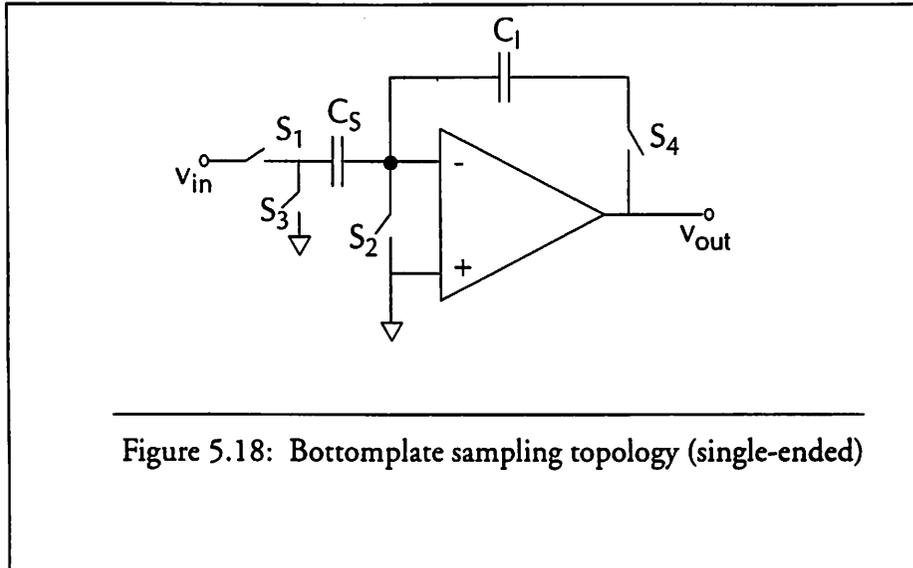


Figure 5.18: Bottomplate sampling topology (single-ended)

switch determines the sampling instant. Since both the drain and source of transistor S_2 are at ground, no signal dependent charge injection is introduced. When S_1 is subsequently turned off, the charge in the channel sees an open circuit on the other side of C_S and therefore all the signal-dependent charge that is injected must flow to the other side of the transistor to the input source (and again no error is introduced). Thus, the injection pedestal can be made constant; by employing differential signalling, this constant offset simply becomes a harmless common-mode shift; it also eliminates the constant offset due to clock feedthrough. In effect, the performance limits of such a differential bottomplate sampling topology are established by second-order effects, most notably differential mismatch [Lin90]. The switch accuracy does remain quadratically dependent on L : however, it is vastly improved over the simple topology of figure 5.17 by the differential bottomplate nature of the switch. Given careful layout, a sampling demodulator operating at 128 MHz, demodulating a 1.088 GHz carrier signal is quite possible with the 0.8 micron technology available today.

5.5.1. The Bottomplate Sampling Demodulator

The sampling demodulator that was designed and fabricated is shown in figure 5.19. Also shown is the operational transconductance amplifier (OTA), required to effect charge transfer from the sampling capacitors and buffering into the next stage. Furthermore, the sampling capacitor has been divided into two separate capacitors, C_{int} and C_{smp} , with C_{int} serving a dual role as the integrating capacitor for the OTA, and C_{smp} can be switched in and out of the switch array by FET switches. This “splitting” of the sampling capacitor allows a variable-

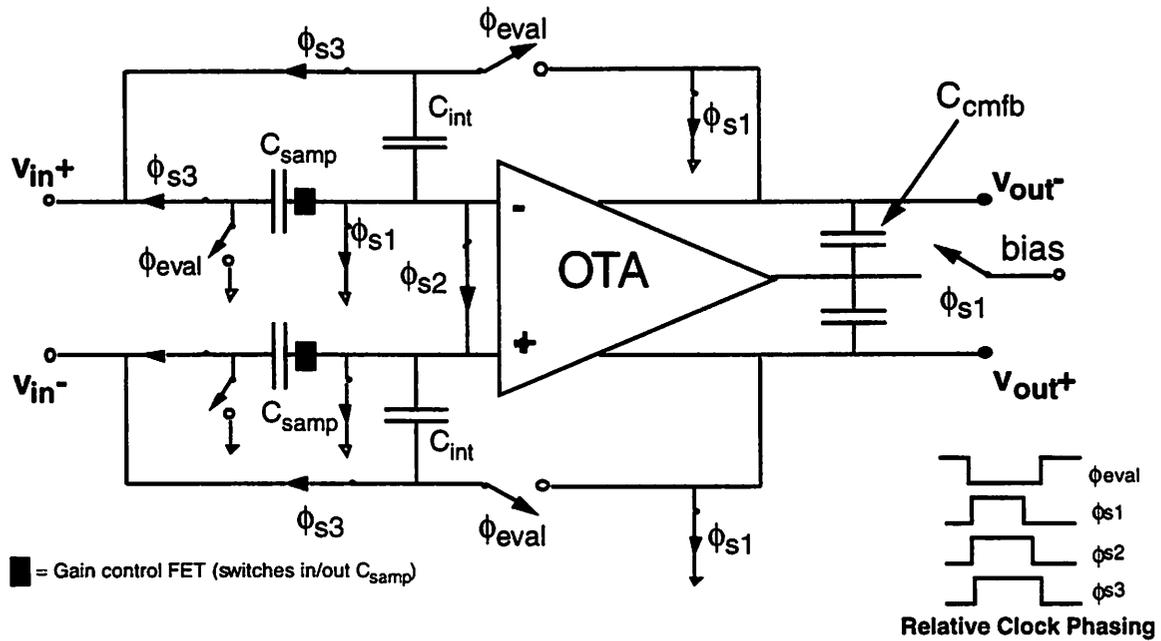


Figure 5.19: Variable-gain, bottomplate sampling demodulator topology

gain function to be achieved by the sampling switch; with C_S present, the conversion gain of the mixer is +3 dB, without C_S , the conversion gain is 0 dB. Since the issues of variable-gain amplification, frequency response, and settling characteristics of the OTA are really baseband analog issue, discussion of its design is relegated to the next chapter; the design of the switch array is the principle focus of the following discussion, with regard to track bandwidth, mixer noise figure, and linearity.

To determine the track-mode bandwidth of the switch, the track mode configuration of the sampling demodulator is shown in figure 5.20. Both capacitors are assumed to be present, giving the worst-case bandwidth condition due to the additional parasitics involved, and the equivalent linearized model is shown in figure 5.21. From extracted device parameters, the -3 dB bandwidth (voltage transfer onto the capacitors) is 2.1 GHz, both by hand calculations and SPICE simulation: the oversized track bandwidth is to account for extraction errors (especially with regard to analog parameters such as g_m), as well as provide some margin for phase balance between the each leg of the differential switch array. The penalty paid for the

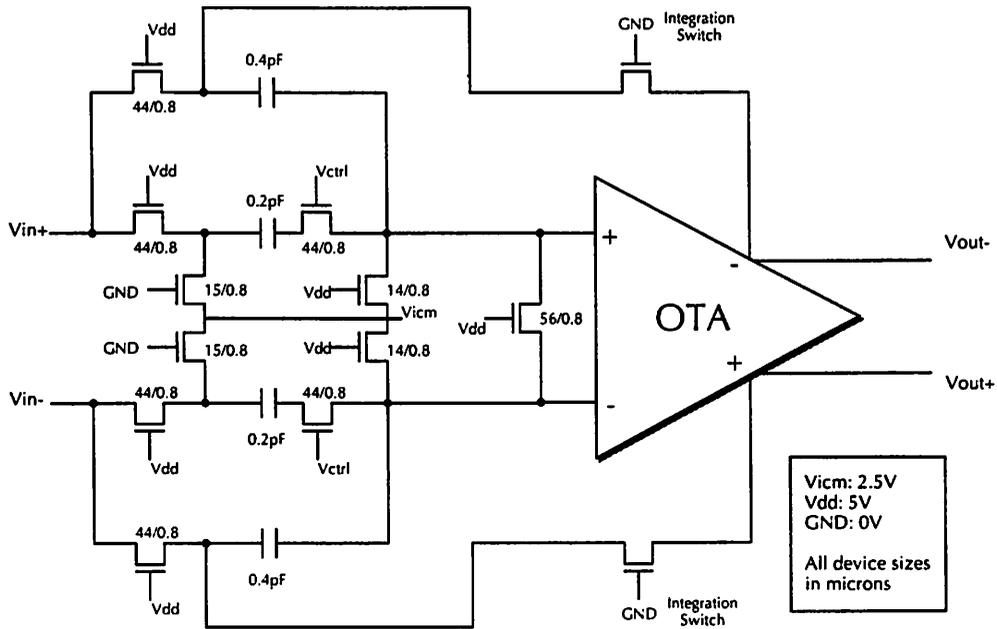
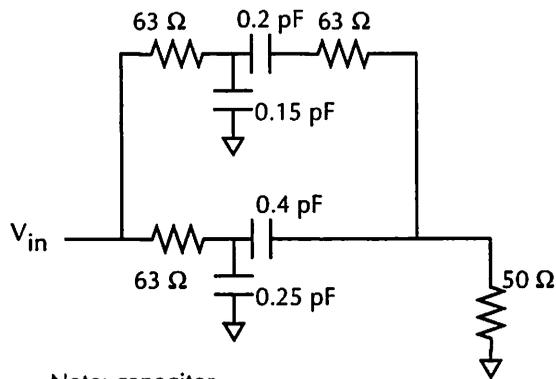


Figure 5.20: Sampling demodulator, track-mode configuration.



Note: capacitor and switch parasitics are included.

Figure 5.21: Track-mode linearized model of the sampling array (equivalent half-circuit)

lack of a second polysilicon layer is evident here; the capacitor is formed by a metal1-metal2-metal3 stacked structure (figure 5.22), with the parasitic capacitor associated with the metal1-metal3 top-plate being approximately half of the main capacitance. As a result, the design value for C must be significantly smaller than what would be possible in a double-poly process, impacting the kT/C performance dramatically; the design values for the integrating and variable-gain sampling capacitors are 0.4 and 0.2 pF respectively.

The noise figure of the mixer is thus straightforward to determine: the noise mechanism is simply the well-known kT/C noise of the switch itself [Gray89]. Given that the capacitors need to be extremely small to admit the 1 GHz received signal, it is not surprising that the noise figure of the mixer is poor. For the design values of $C_S + C_I = 0.6 \text{ pF}^1$, this results in an equivalent total mean-square voltage noise equal to $6.9 \cdot 10^{-9} \overline{v^2}$. Conservatively assuming a total signal bandwidth of 128 MHz (to account for filter rolloff), the kT/C noise can be referenced to the total in-band noise of a 50Ω resistor, leading to an equivalent noise figure of 18.1 dB. To compensate for this, the architecture of the receiver in figure 5.2 places an additional 16 dB low-noise buffer stage just before the demodulator. The extreme noise figure resulting from kT/C noise is perhaps the key disadvantage of employing a direct-sampling scheme for RF downconversion - the total broadband noise in the switch is admitted into the

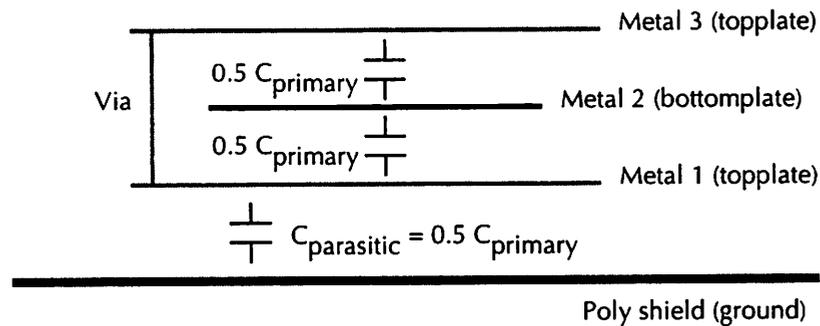


Figure 5.22: On-chip capacitor structure

HP 0.8u non-double-poly process:
 Area cap (M1-sub,M1-M2, M2-M3): 0.035 fF/sq.micron;
 Perim cap (all metal layers): 0.049 fF/micron.

1. This is assuming the weakest signal situation (maximum gain), where both C_S and C_I will be employed in the demodulator.

system. If some means were available to filter out the switch thermal noise, potentially by using an on-chip inductor to realize a bandpass network, the noise figure of the mixer could be reduced tremendously. Alternatively, with CMOS technology scaling, the R_{ds} of the switch could be made commensurately smaller, allowing larger C 's to be used in the sampling operation.

Lastly, distortion in the sampling demodulator needs to be considered. Surprisingly, the demodulator is quite linear; the linearized model of figure 5.21 applies over a large input range. Intuitively, the demodulator reduces to a FET switch and a capacitor; a high track bandwidth implies that the operating frequency is such that very little voltage is lost across the nonlinear resistance of the switch. Thus, until the input signal is large enough that the switches themselves begin to turn off, the demodulator should behave in a extremely linear fashion. The capacitor of figure 5.22 is a strictly linear device; the only circuit nonlinearity is that of the sampling switches themselves. The above analysis agrees with simulation, as shown in figure 5.23: the simulated input referred IP_3 of the modulator is 20.4 dB.

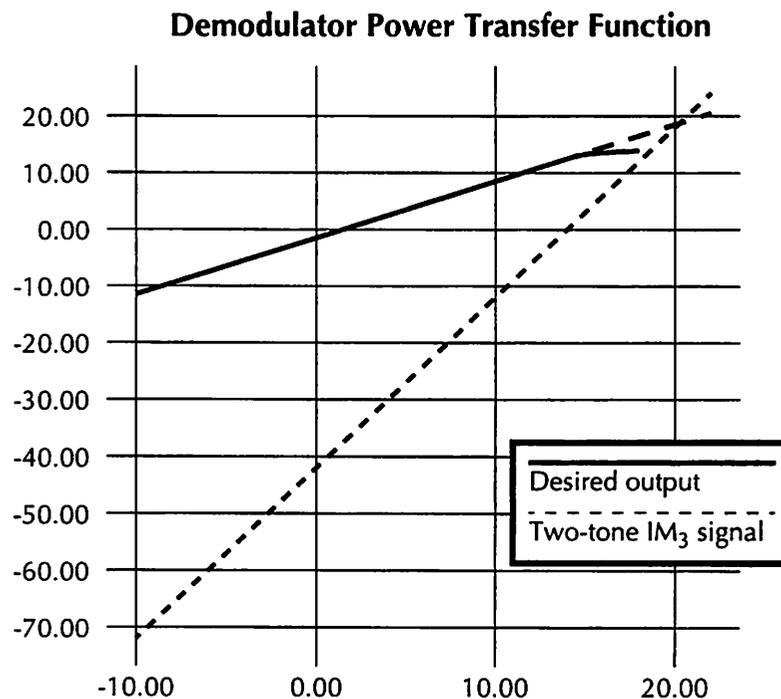


Figure 5.23: Simulated distortion characteristics of the sampling demodulator. Two equal-power input tones were used for the IM_3 test.

5.5.2. On-Chip Clock Buffering and Layout Issues

Lastly, the issue of how the sampling clock is generated must be addressed, with respect to the 28 psec jitter accuracy needed in the sampling operation. As described previously, this requirement can easily be met by the off-chip crystal oscillator; however, preserving this phase-noise performance on-chip is critical. In particular, proper buffering, coupled with careful layout considerations, is required to insure that the clock edges remain as jitter-free as possible, as well as minimizing substrate injection from the clock edges themselves.

First, for system-level considerations, a 128 MHz off-chip clock oscillator is chosen; sampling is done on both rising and falling edges, with the in-phase demodulator clocked on the rising edge and the quadrature demodulator clocked on the falling edge. Furthermore, the off-chip oscillator signal is chosen to be a differential sinusoid instead of a square-wave pulse. The differential nature of the signal minimizes radiation and coupling at the board level; the sinusoidal nature minimizes the harmonic energy at 1 GHz, important given that a high-sensitivity LNA is on the same die. Of course, this signal must be converted to a square wave before it can be utilized on-chip, requiring a differential analog input clock buffer. The clock buffer for the chip is shown in figure 5.24; it is embedded as part of the clock input pads. The center pad is used to bias the differential buffer; the nominal external bias for this node is 1 mA, for operation at 128 MHz. Likewise, the three sampling edges required by the differential bottomplate sample/hold must be synthesized from the converted clock signal, by the NAND-based nonoverlapping generator circuit shown in figure 5.25. Again, symmetric layout is employed, with signals switching differentially between the phi1 and phi2 clock phases.

In addition to the buffering, the layout of the sampling switches themselves is crucial.

Figure 5.26 shows the die fabricated to test the sampling demodulator by itself (a portion of the full analog test chip described in chapter 6), along with several important features of the layout. First, the clock lines are run down the centerline of the design, with the in-phase and quadrature demodulator blocks on either side. This digital “no-man’s land” is heavily shielded, with grounded metal-3 and poly shields above and below to minimize coupling through the air, as well as coupling into the substrate by the clock lines. Second, to guarantee the high-speed falling edge on the sampling switches, very large local clock buffers are placed next to

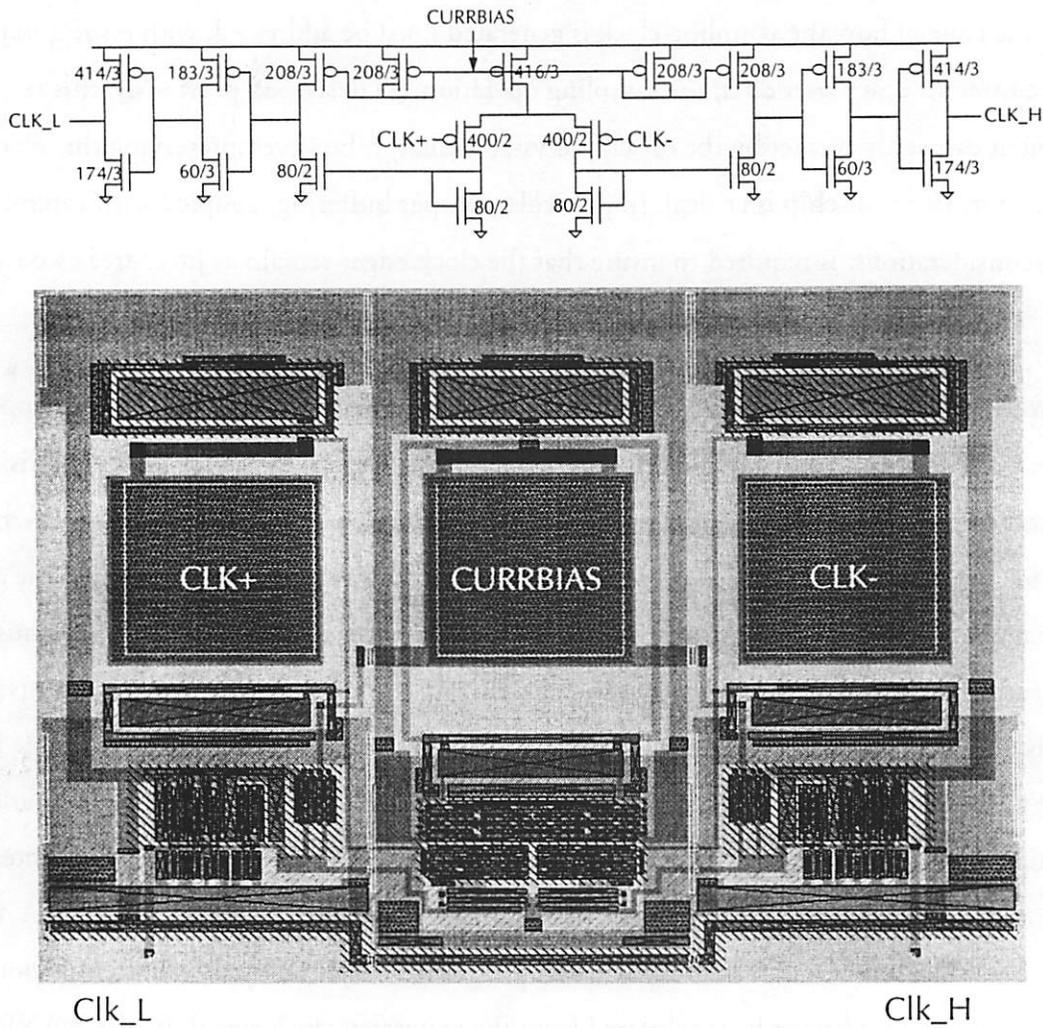


Figure 5.24: Schematic and layout of differential input clock buffer. Device sizes are in lambda (0.5 micron = 1 lambda)

each switch array; the sizing of the final inverter is a $78\mu/1\mu$ PMOS, $30\mu/1\mu$ NMOS. A simulated 10-90 fall time of 110 psec (40 psec to switch turn-off) is produced by these local buffers, necessary to accurately sample the gigahertz-band input signal.

Furthermore, large grounded guard rings are placed around all of the fast digital switching devices, as to capture substrate injection as early as possible. Although ground-lead inductance going off-chip typically limits the effectiveness of the grounded guard rings, all effort was

made to minimize this inductance, most notably by the presence of a backside substrate ground contact and direct vias to the ground plane in the chip-on-board packaging (see the appendix to this chapter). Lastly, large on-chip MOSFET-gate bypass capacitors (100 pF) are placed near each demodulator, to minimize the impact of supply bounce on the jitter performance of the clock buffers.

5.6. From RF to Baseband

In this chapter, the RF characteristics of a custom, integrated CMOS analog receiver front-end, suitable for use in the mobile end of the broadband CDMA downlink. Beyond achieving functionality normally relegated to silicon BJT's or GaAsFET's, such as low-noise amplification, the cornerstone of this new integrated architecture is the use of sampling demodulation to achieve frequency conversion, where the harmonics of a high-speed, low-jitter baseband clock edge are used to mix down the received signal. Although such a demodulator may appear to be a simple CMOS differential bottomplate sampling switch, the particular characteristics required to sample a 1 GHz modulated signal adds several layers of complexity and consideration to issues such as track bandwidth, charge injection, and switch linearity.

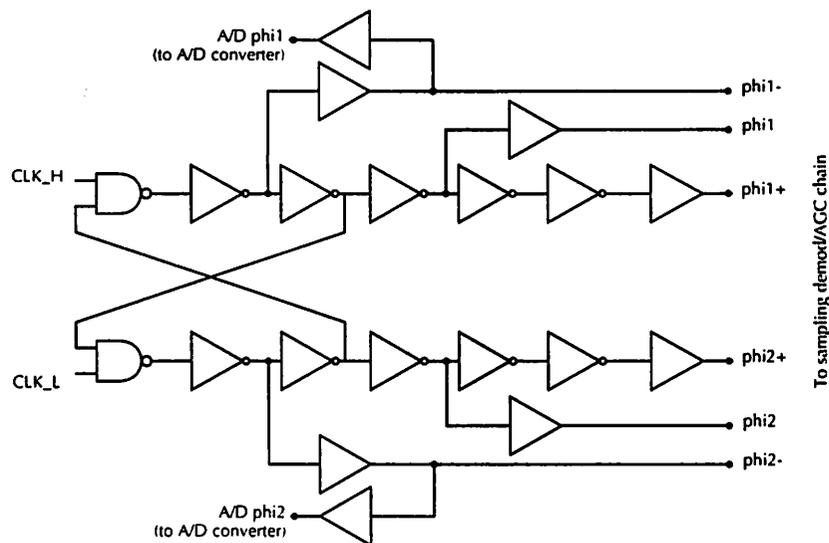


Figure 5.25: On-chip nonoverlapping multiphase clock generator (3 edges, both rising and falling phases)

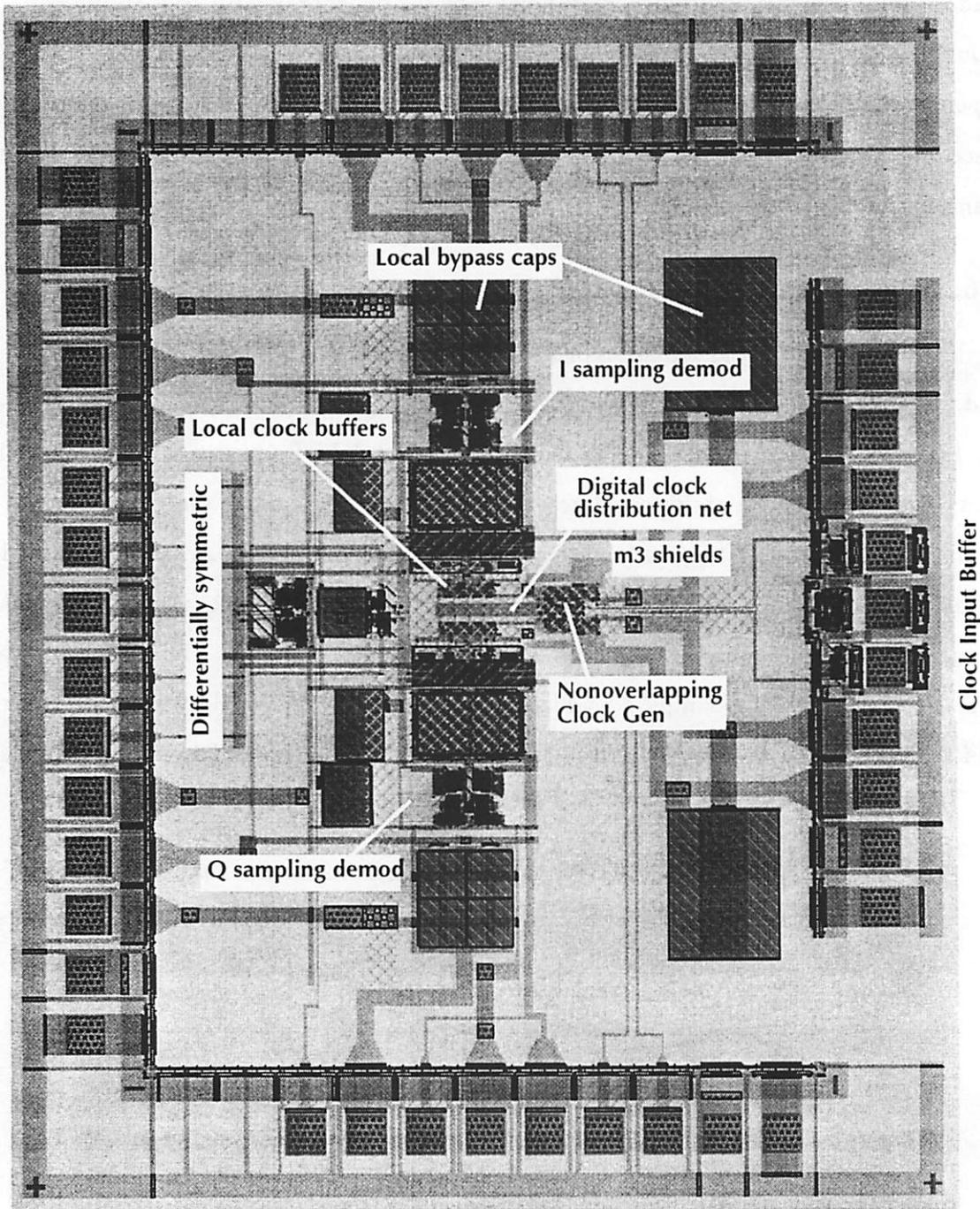


Figure 5.26: Standalone sampling demodulator testchip, illustrating several points on die layout. The layout of this die reflects that of the full analog front-end testchip, as discussed in chapter 6.

With the opening of the sampling switch, the signal has traversed the boundary from the continuous-time, gigahertz-band domain, into the discrete-time, baseband domain, where switched-capacitor techniques are dominant. The next chapter describes the second half of the analog front-end: the analog hardware required to perform baseband variable-gain amplification, and *A/D* conversion, completing the list of functions that need to be performed in the analog domain. Although the results for the prototype LNA have been discussed in this chapter, the testchip results for the complete demodulator will be deferred until the end of chapter 6, where the entire analog receive chain can be examined.

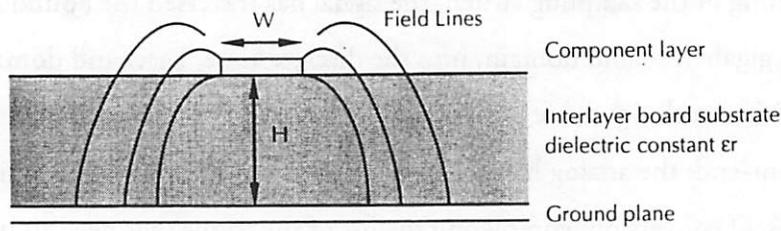


Figure 5.27: Board transmission line structure

5.7. Appendix: RF Chip-on-Board Design

This appendix is intended to be an overview of the board design process, especially in consideration of gigahertz-band operation and the need for chip-on-board packaging for parasitic control. It is not intended as a tutorial; it is assumed that the reader has some passing knowledge of electromagnetism and board design.

5.7.1. Electromagnetic Effects

To provide a good 50Ω matched conductor on the board, microstrip and stripline structures must be employed. Creation of microstrips is relatively easy to do in a multilayer board; if a ground plane is placed beneath an RF trace, a microstrip is naturally created, as shown in figure 5.27. Since the electromagnetic fields are half in air and half in the board material, an effective dielectric constant $\epsilon_{r, \text{eff}}$ is calculated with respect to the substrate ϵ_r :

$$\epsilon_{r, \text{eff}} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \frac{1}{\sqrt{1 + 12 \frac{H}{W}}}$$

Since wave propagation in microstrips can be approximated as quasi-TEM, an effective inductance and capacitance per unit length can be assigned, with an equivalent characteristic impedance Z_o equal to:

$$Z_o = \frac{60}{\sqrt{\epsilon_{r, \text{eff}}}} \ln \left(\frac{8H}{W} + \frac{W}{4H} \right)$$

The key thing to note is that Z_o is effectively controlled by the W/H ratio of the stripline; the larger the W/H , the lower the impedance. Furthermore, the smaller the ϵ_r , the higher the impedance; alternatively, for a smaller ϵ_r a larger W/H ratio will be needed to keep the same Z_o .

Figure 5.28 lists typical materials that are used in board fabrication. Although it has significantly inferior loss tangent and higher ϵ_r , FR4 is the material of choice: Duroid and cyanate ester is very brittle, having little water content, and it is difficult to make boards of any significant size. Also, at least for the boards that have been designed to test the custom integrated circuits, the traces are fairly short (around 5 cm maximum length), and there are only a few RF traces (since the circuits are highly integrated). Thus, signal loss due to skin effect and nonzero board substrate loss tangent is negligible; from a practical standpoint, multilayer FR4 is used for all of the board designs.

As far as choice of Z_o , 50Ω is usually used for the following reasons:

- Standard coax cabling is readily available at 50Ω
- All available test equipment will terminate and match in 50Ω
- 50Ω is realizable on the board: to achieve 50Ω in FR4, a W of 39 mils over an H of 19 mils ($W/H = 2.0$) is needed. This allows tolerance in the board fabrication process, which is typically ± 1 mil on H and ± 0.5 mil on W . For example, to achieve $Z_o=200\Omega$, $W/H = .027$ is needed; for $W=10$ mils, an H of 370 mils is needed, far too thick to be practical, as well as resulting in higher-order transmission modes along the line.

Material	ϵ_r	loss tangent/ dissipation factor
FR4	4.5	0.019
Duroid	2.20	0.0009
63N cyanate ester	3.3	0.009

Figure 5.28: Available board fabrication materials

5.7.2. General Layout Issues

A few rules work well:

- Keep RF traces as short as possible, and minimize the number of corners on RF traces; each corner will radiate. If trace widths need to change, tapering of the trace is effective in preventing reflections on the line; this is known as a tapered transformer [Pozar90].
- Use supply planes whenever possible. For example, the RF test board described below uses 3 supply planes: V_{dd} (analog), V_{dd} (digital), V_{icm} (common mode bias voltage). Not only does this minimize supply bounce/ripple by providing a physically short, low-impedance return path, but also simplifies routing on the board tremendously. Separate digital and analog supplies to minimize coupling, either by splitting the planes, or by using separate planes.
- The order of the board layers is important: make certain that the second layer of the board is a ground plane, otherwise the microstrip structure does not exist. Even if the second layer were a supply plane, supply planes tend to have more noise than ground planes, due to nonzero source resistance and inductance. Shield the RF traces from any digital switching lines by the digital lines on the solder side (bottom) of the board. By placing several supply planes between the critical analog traces and the digital output lines, better shielding is achieved (see board layer listing in figure 5.33).
- Route the RF traces by hand before any sort of automated routing. This is certainly CAD-tool dependent, but as a general rule hand routing will result in shorter RF traces and fewer corners.
- Use coaxial connectors to bring the supply voltage on-board. RF signals tend to couple into everything unless shielded; a long unshielded wire coming from the bench power supply is a perfect antenna, and can result in excessive RF noise coming from the supply.
- Capacitor bypass DC supply and control pins as much as possible. As discussed below, the choice of capacitor size varies critically with the application.

5.7.3. Discrete Inductors and Capacitors

Due to high-frequency effects, special consideration must be given to passive inductors and capacitors for use in RF boards. Not only are such devices needed for supply bypassing and AC coupling purposes, but they also become necessary for filter design: at 1 GHz, the wavelength (30 cm free-space) is still large, making striplines unattractive as a means of filtering while maintaining a small form factor. Fortunately, at 1 GHz, discrete passives can still be used; parasitic effects have not become dominant. The key issue is the self-resonance frequency of the devices: capacitors form a series tank circuit with the bondwire inductance of the package, whereas surface-mount spiral inductors have an intrinsic self-capacitance from winding turn to winding turn (figure 5.29).

Figure 5.30 lists several commercially available surface-mount inductors, along with their self-resonant frequencies (SRF). For operation at 1 GHz, an SRF in excess of 2 GHz is needed, which limits the usable inductance value for designs to around 10 nH. Another issue is that of inductor tolerance: small values of L are also not possible, owing to the parasitic inductance introduced by the solder ball on the surface mount board itself (on the order of 0.25 - 0.5 nH of inductance). Due to this, and the limiting parasitics involved with getting a signal on-and-off of a silicon chip, on-chip inductors are highly desirable. Unfortunately, on-chip inductors were not realizable in the available silicon fabrication technology, and thus significant effort was expended in “hand-tweaking” the board to achieve the accurate inductances.

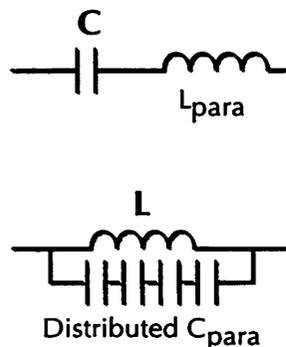


Figure 5.29: Parasitics in surface-mount discrete capacitors, inductors

L (nH)	Nominal Q @ 500 MHz	Series R (ohms)	Self-Res. Freq (GHz)
4.7	20	1	2.0
6.8	20	1	2.0
10.0	20	1	1.0
15.0	20	2	1.0
22.0	20	2	1.0
33.0	10	5	0.850

Figure 5.30: Table of commercially-available surface-mount inductors

An even more critical issue is the choice of capacitors, since they provide both AC signal coupling and DC bypass for the supply and control lines. The usual “low-frequency” design mentality is to use as large of a capacitor as possible for both functions; at RF, this is not viable owing to the limitations placed by the SRF. Shown in figure 5.31 is a plot of measured SRF as a function of capacitance value, for three families of capacitors (1206, 0804, 0603 size form factors). The dependence illustrated here is interesting; since $SRF = 1/\sqrt{L_{para}C}$, one would expect a square-root dependence on C and similarly dependent on the package size (controlling L_{para}). From the measured plot, size has little impact, whereas the SRF is inverse-linearly dependent on the capacitance value. The reason for this lies in the fabrication of the surface mount capacitor itself: to achieve high values of C, many layers are wound within the package itself, thus leading to larger parasitic L's. Thus, the parasitic L in the SRF is controlled by the capacitance value, not by the physical package size.

Thus, using a large capacitor is hardly optimal, since at 1 GHz the capacitor will appear to be strongly inductive. However, for both AC coupling and DC HF-bypassing, the capacitor needs to appear as a zero impedance. Since AC coupling applications only require a zero impedance at the operating frequency, the parasitic series resonance can fortuitously be used to

achieve this, since it results in a near-zero impedance at the SRF, as shown in figure 5.31b. Furthermore, the Q is fairly low; SRF tolerance is not critical; thus, by choosing capacitors with an SRF near the operating frequency of 1 GHz, a good AC couple can be achieved. For the RF testboard described below, all AC coupling capacitors were 0603 form factor, 10 pF capacitors. DC bypass applications, on the other hand, are more difficult. Although the SRF-

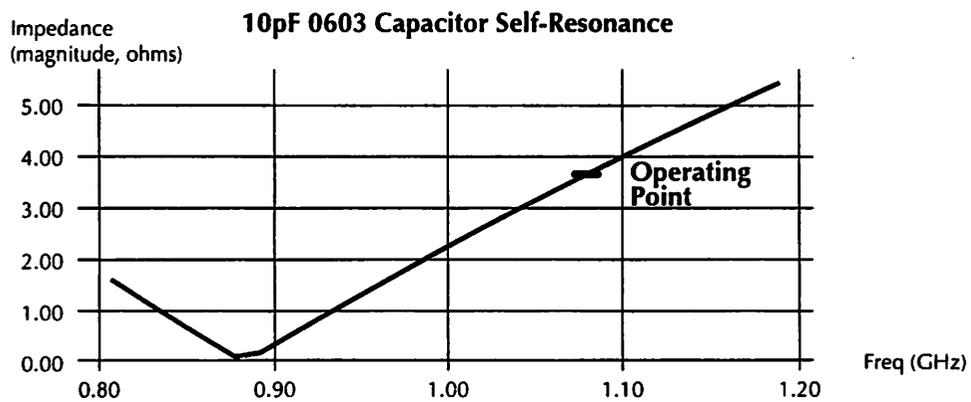
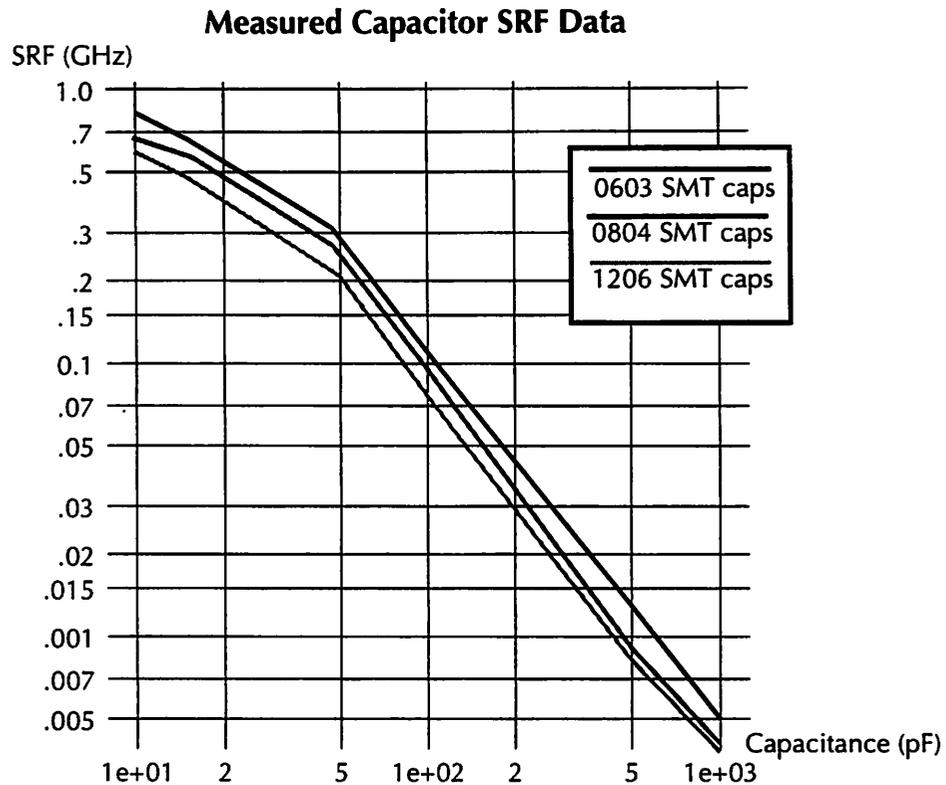


Figure 5.31: Impact of inductive parasitics on discrete capacitors at RF.

technique will provide a low impedance for a range of frequencies at RF, bypassing at lower frequencies is still necessary. Thus, two or more capacitors are usually used in parallel as shown in figure 5.32, with one capacitor providing low-frequency bypass and the second providing RF bypass. However, care must be taken: there is a parallel LC formed by the parasitic L of C_1 and the actual capacitance of C_2 . This implies that, at some resonant frequency, the bypass network itself is a nearly open circuit, and will be subject to ringing at that frequency. Damping of this parallel LC with a shunt R is necessary to prevent this.

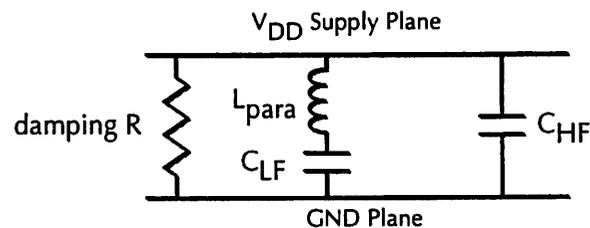


Figure 5.32: Board-level bypass strategy. Typical values are:
 $C_{LF} = 0.1 \mu\text{F}$, $C_{HF} = 10 \text{ pF}$, $R = 500 \Omega$

5.7.4. Chip-On-Board Design

For large pin count die, the packaging itself limits circuit performance tremendously: package inductance, plus large lead inductance, sets fairly low upper bounds on the maximum operating frequency, especially when the circuit input capacitance is large.

- High integration implies high lead count, and hence large packages. Since the entire analog section was to be integrated into a single die, the RF sections plus the baseband analog processing required a 132-pin CQFP package to accommodate all of the necessary signal, control, and supply./ground lines. Such packages were measured to have a worst-case package + bondwire inductance of 20 nH, and a lead frame capacitance of approximately 2 pF.

- For CMOS, a large input capacitance is nearly unavoidable, especially in light of the 0.8 μ m process available: the LNA input device, a 600 μ m/0.8 μ m NMOS, exhibits a C_{gs} of 0.96 pF, while the bonding pad results in another 0.5 pF of additional input capacitance.

The result of this is that the packaged version of the design limits the LNA maximum operating frequency to approximately 600 MHz (the self-resonance point of the effective series L-C). Furthermore, supply and ground lead inductance effects are similar; substrate injection and on-chip coupling cannot be minimized at 1 GHz, since the board ground plane is separated from chip ground by the package inductance.

Obviously, a better packaging technology is needed. Multichip modules are certainly one solution; however, a viable manufacturing process could not be obtained. “Good” high frequency packages have 50 ohm controlled microstrips for the leads, as well as local supply bypass capacitors on the physical package itself; unfortunately, these packages are prohibitively expensive. The final solution is particularly elegant: eliminate the package [Lau94]. Direct die-attach to the surface mount board is quite possible, easily integrated into the regular surface-mount process, and inexpensive. The details of the chip-on-board process used will be described below.

5.7.5. RF Testboard

The following documents the surface mount board that was developed for testing the entire monolithic analog front-end, as described at the end of chapter 6. The physical fabrication specifications of the testboard are as listed in figure 5.33. In figure 5.34a, the board layout is shown, with several salient features marked. A magnified view of the bonding site for the analog front-end testchip is shown in figure 5.34b; for bondability, a 4 mil¹ wide line is used for the board bonding lands, with the minimum allowable 3 mil space between them. Likewise, these board bonding lands are pitch-matched one-to-one against the pads on the die. This makes bonding particularly simple, as well as making certain that every pad has the same

1. From contact with various board fabrication/assembly houses, 4 mils is the minimum width that can be reliably bonded with technology available at this time.

(minimum) bondwire length and associated parasitic inductance. With a 7 mil pitch, this determined the pad pitch on the die of 150 microns. Furthermore, it is not sufficient just to be able to bond to the board; the signal must be able to connect to other components on the board; thus, an interleaved set of vias (figure 5.35) is incorporated on the board to provide connection points. In particular, these vias can directly contact the ground and supply planes, providing very high-performance, low inductance (< 1 nH) connections. The geometry indicated in figure 5.35 represent the minimum spacings allowable by our board fabrication technology.

As far as the actual die attach site goes, it comprises of a single square plane of gold-copper metal; with a die size of 280 mils by 280 mils; the actual bonding site has a 20 mil overhang on all sides to allow some tolerance in die attach (attach area is 320 mil by 320 mil). Further-

1) BOARD FILM SET DEFINITION

 Layer1 | Component Signal, SEE BELOW
 Layer2 | Power Plane GND, 1oz CU
 Layer3 | Power Plane 1 (analog VDD), 1oz CU
 Layer4 | Power Plane 2 (analog VICM), 1oz CU
 Layer5 | Power Plane 3 (digital VDD), 1oz CU
 Layer6 | Signal 4 (digital signals only), 1oz CU
 Layer7 | Ground Plane (digital GND), 1oz CU
 Layer8 | Solder Side (digital signals only), 1oz CU
 SMC | Solder Mask, Component Side
 SMS | Solder Mask, Solder Side
 SILKC | Silk Screen, Component Side

2) SIZE: 5500 mils x 4700 mils

3) MINIMUM LINE/WIDTH SPACING: 3 mils

4) Laminate is FR-4 (Green)

5) Overall board thickness is 0.075 +/- 0.008

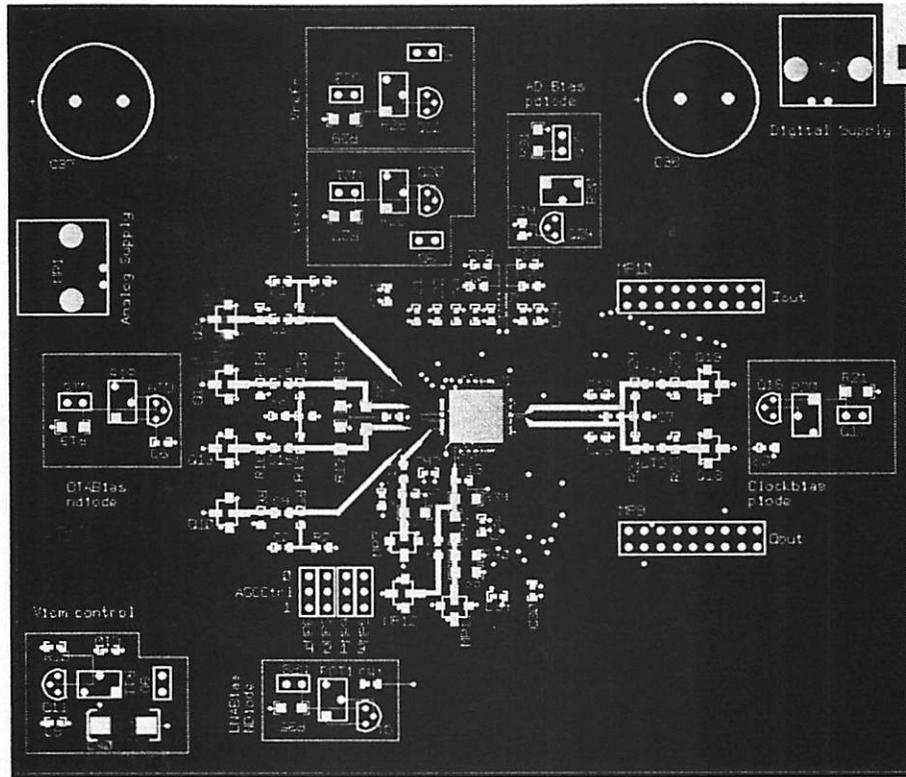
6) Comments/Special Handling

This is a chip-on-board design; since the die needs to be attached and bonded directly to the component layer, the comp layer needs to be plated with wirebondable gold on the component layer, 50 microinches thick with 150 microinches of nickel, 0.25oz foil

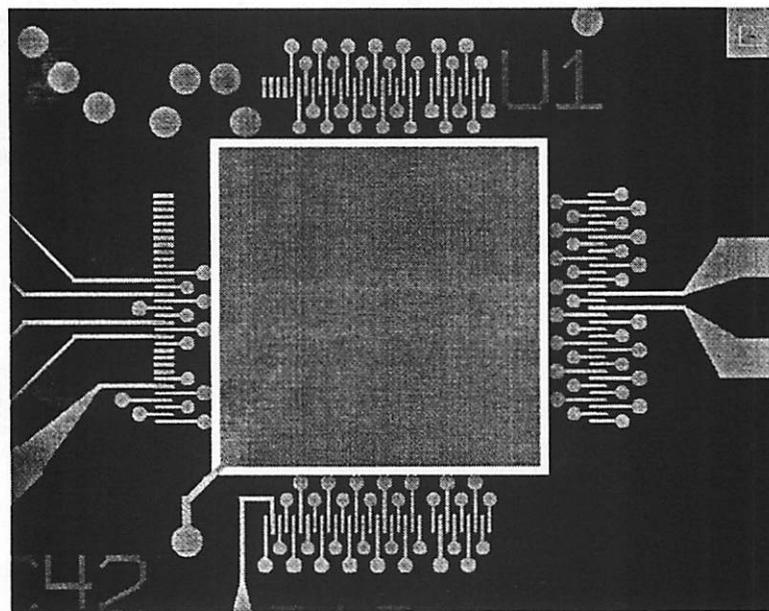
Also, layers 2, 3, 4, 5, and 7 are gnd/power planes; hence, the gerbers are negative images. Interlayer spacing between component layer and powerplane GND should be 21 mils, all other spacings should be minimum.

All dimensions are in English units unless otherwise specified

Figure 5.33: Board fabrication parameters
 (as submitted to the fabrication facility)



(a)



(b)

Figure 5.34: RF testboard layout. (a) Layout; (b) close-up of the chip-on-board bonding area. (Test jig for the analog front-end prototype die)

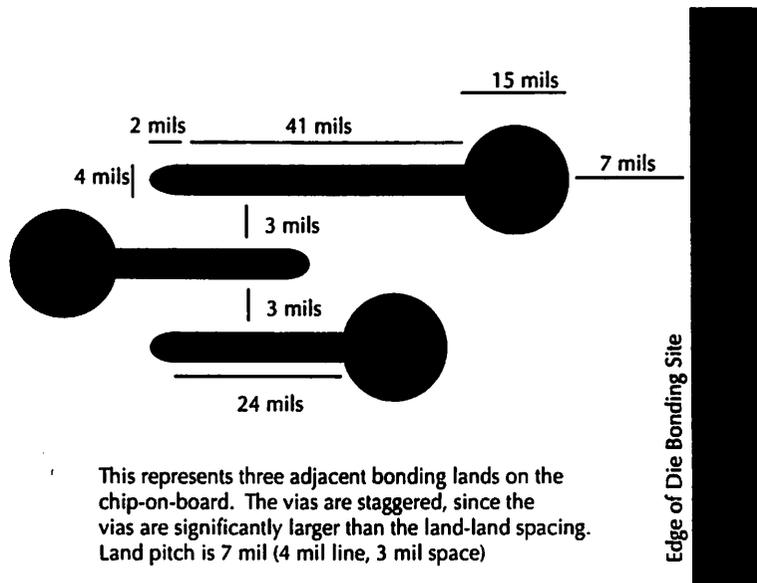
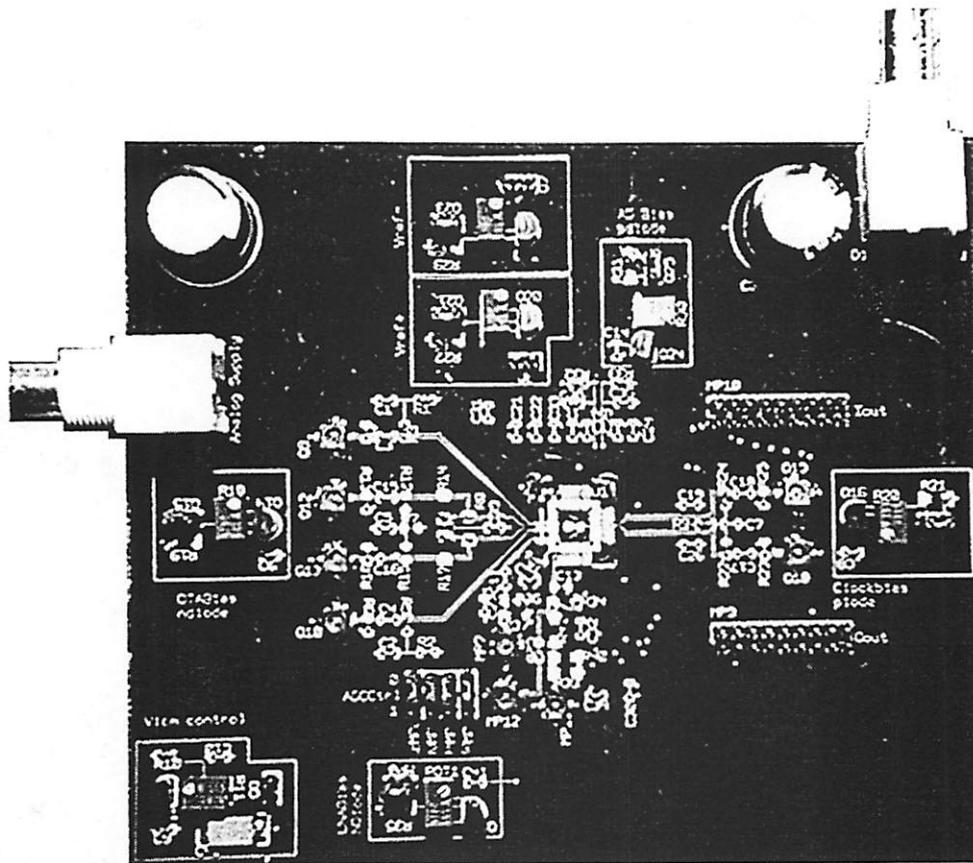


Figure 5.35: Interleaved via layout geometry of the chip-on-board bonding lands.

more, a direct ground via is connected at one corner of the bonding site; this provides a solid ground directly underneath the die; by using conductive epoxy to attach the die, a good back-side contact can be formed between the ground plane and the silicon die substrate. This is critical in minimizing the effects of on-chip substrate injection, since a path to a solid low-impedance ground is provided through the back-side. The fabricated board photo is shown in figure 5.36.



Blowup of
chip-on-board
die attach region

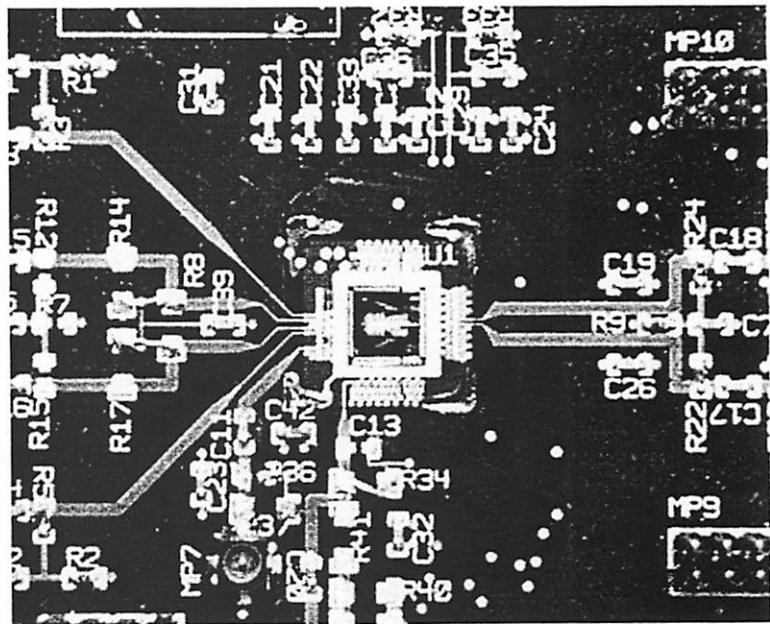


Figure 5.36: Fabricated board photograph

6 The Receiver: Baseband Analog Processing

With RF amplification and demodulation performed, the remaining functionality of the receiver front-end consists of automatic gain control, and analog-to-digital conversion. As discussed in chapter 5, the ability to leverage off complex digital processing results in an architecture that favors moving the analog-to-digital converter as early in the receive signal path as possible: in effect, creating an “A-to-D front-end”. Of course, this implies that an extremely high-speed, high-performance A/D converter is necessary; concomitant with this is the need for variable-gain capability in the analog domain. The explanation for this requirement is simple: the mobility of the terminal within a fading environment implies large and rapidly changing shifts in the received signal amplitude. A moving terminal may be physically very close to the transmitter (and therefore receiving a strong signal), and then move very far away (where it receives a weak signal); it may also suddenly move into a fading null, or a location where there is no line-of-sight path between the base station transmitter and the mobile.

Therefore, implementation of an automatic gain control circuit (AGC) is required, to automatically adjust the gain of the receive path so that the signal presented to the A/D converter circuitry appears to be of constant power regardless of the actual signal size at the antenna. In other words, the AGC and the A/D combine to detect a signal with a wide dynamic range. While it is true that bits in the A/D may be traded off for gain in the AGC (i.e., increased resolution in the A/D, beyond the minimum required for SNR, allows it to detect a smaller signal), higher resolution converters rapidly become prohibitively costly, in terms of power, area and complexity: given that a conversion rate of 128 MHz is needed for timing recovery of the baseband spread-spectrum signal, and with the limited 0.8 micron CMOS technology available, 4-5 bits of resolution is all that can be achieved by the A/D converter. Furthermore, the

digital processing that follows the A/D (the spread spectrum demodulator in this case) must also process any additional bits added to the A/D – further increasing power and area. Every AGC contains two critical blocks – a variable gain amplifier (VGA) and the power detector circuit that feeds back the control signal used to adjust the gain of the VGA. At the time of this writing, the power detection loop for the AGC block has not been implemented. However, the VGA has been designed and implemented with the eventual addition of the detector and the control loop in mind. Given that a sampling demodulator is used to convert the continuous-time RF signal to a discrete-time baseband signal, the architecture and implementation described below is a significant departure from traditional VGA and A/D designs for wireless systems.

Thus, to complement the discussion of the RF circuitry in chapter 5, this chapter is devoted to discussion of the design and implementation of the baseband variable-gain amplification and analog-to-digital conversion circuitry required in the analog front-end; more detailed documentation of the circuits discussed below can be found in [Lynn95].

6.1. System Dynamic Range and Linearity Requirements

Of course, the immediate question that arises is, “How many bits?” To answer this, system-level simulations of the overall receiver were performed, with respect to the impact on quantization noise on the ultimate system BER. Shown in figure 6.1 is the BER curve as a function of users, under statistically severe fading conditions and varying bit-quantization in the A/D converter. Clearly, no improvement in the curve is seen for greater than 4 bits of resolution: given that the desired user’s signal is already “buried” in the multipath noise of other users, beyond 4 bits the quantization noise ceases to be significant. Thus, a baseband linearity requirement of 4 bits is established for the entire discussion of this chapter; as will be seen, even 4 bits of linearity is quite difficult to achieve given the technology limitations.

However, it seems counterintuitive that a system that can intrinsically accommodate 64 users only requires 4 bits of resolution in the A/D converter itself for proper operation. The answer lies in the processing gain provided by the spread-spectrum decorrelation: in some sense, high-speed A/D conversion is traded against resolution requirements. More specifically, the

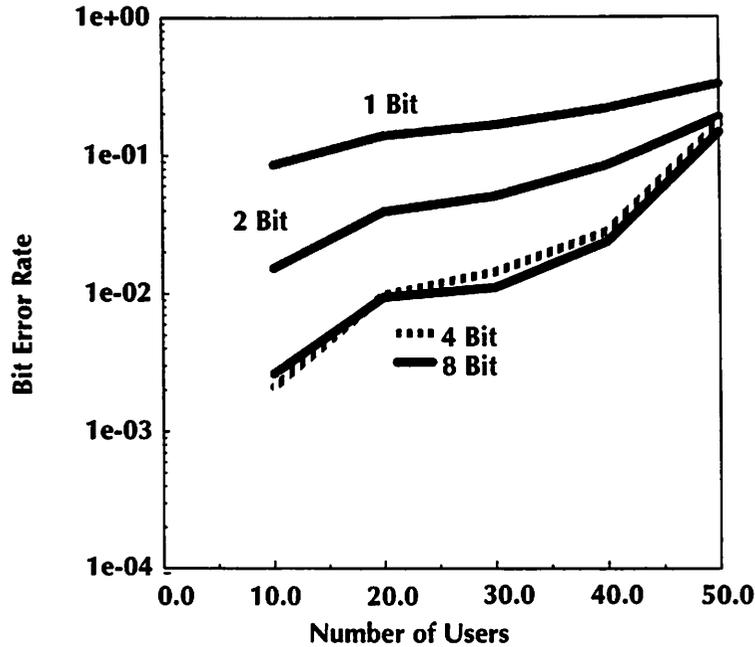


Figure 6.1: BER vs. Number of in-cell users, as a function of A/D bit quantization (extreme fade condition, no error correction)

18 dB of processing gain provided by the spreading translates into another 3 bits of effective resolution at the output of the spread-spectrum correlators in the digital back-end. Ultimately, with respect 1 MHz baud decisions being performed by the receiver, a 7-bit converter is being employed.

Lastly, the overall system dynamic range, from channel measurements and simulations discussed in chapter 2, is 45 dB. Part of the VGA functionality is embedded in the RF elements of the front-end: the LNA itself can be turned off or on, providing an extremely coarse two-setting (+16 dB or -2 dB) form of gain control. Likewise, a second “off-on” RF buffer stage is provided at the input to the sampling demodulator, again providing a coarse form (+17 dB or -2 dB) of gain control. The remaining 12 dB, however, must be much more precise than this: the variable gain topology described below is to implement this fine-grain control, being digitally controllable from 0 to 12 dB in 3 dB steps.

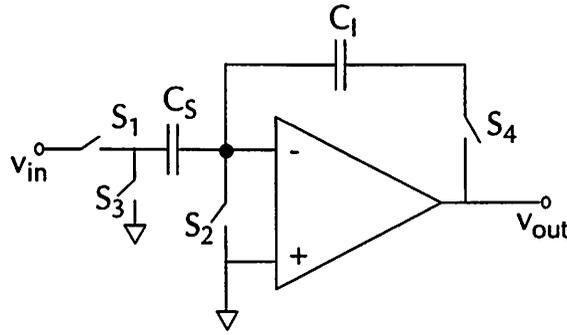


Figure 6.2: Bottomplate sampling topology (single-ended)

6.2. Variable-Gain Amplifier Topology

As mentioned before, the design of the receiver VGA is a significant departure from traditional designs. Because of the inherent sampling incorporated into the mixing operation, a multi-stage cascade of discrete-time amplifiers is used to replace a standard continuous-time VGA. Therefore, what was once an amplifier design problem involving a trade-off between gain and bandwidth becomes a sample-and-hold design requiring an op-amp and a set of switched capacitors that can settle to four bits of accuracy within one clock period. As described in chapter 5, figure 6.2 shows a basic bottomplate sample-and-hold amplifier: however, such a simplistic design is inadequate for practical use. First, the design must be made differential. A differential circuit is necessary not only to reject the charge injection from the switches, but also to reject common-mode noise coupling into the analog signal path. Figure 6.3 shows how digital switching can cause common mode ringing on analog lines. If the sample is taken

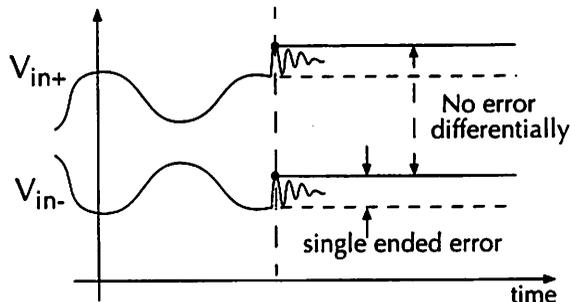


Figure 6.3: Coupling of digital ringing into differential signal paths.

single-ended, a large error would result – an error often larger than the signal itself. However, a differential design with careful layout, with balanced parasitic capacitances, can maintain a small differential signal amidst such large common mode noise.

When performing bottom plate sampling, the operational amplifier shown in figure 6.2 is used to buffer and drive the output to the correct value, proportional to the input voltage. However, settling is a critical issue: when the circuit is in track mode, the output no longer follows the input voltage. Therefore, when the hold phase arrives, the output must settle to the correct output voltage from some reset value, nominally zero. If the amplifier is appropriately designed with adequate phase margin, the settling curve follows an exponential curve with a single RC time constant. Typically, only half of the clock cycle time is dedicated to the hold period (the other half goes to the tracking period). Therefore, the output has half of one cycle ($T/2$) to exponentially approach its final value. Unfortunately, an infinite amount of time is required for an exponential curve to reach its final exact value, given by

$V_{out} = A(1 - e^{-t/\tau})$, where A is the final output value and τ is the time constant of the circuit. Therefore, there will always be an error introduced into the signal due to incomplete settling. Fortunately, the situation is not quite as grim as it may seem. The settling error, $Ae^{-t/\tau}$, is linearly proportional to the final value A if τ is constant. In other words, if given the same amount of time to settle, the circuit will always settle to the same percentage of its final value. Therefore, the gain of the circuit will always be reduced by the constant factor $e^{-t/\tau}$. As long as τ is constant, this gain compression does not introduce a real error into the signal. The actual exact value of the gain is rather unimportant since the negative feedback from the AGC control loop will try to force the output of the VGA to be constant regardless of the gain of each stage, although the compression still causes a loss of gain in the signal path. However, a real error does enter the signal path when the allowed time, t , is not constant. Timing jitter in the sampling instant provided by the edge of a clock causes t to vary somewhat from sample to sample, resulting in an error that is proportional to the jitter variation, D_t , and to the slope of the settling curve at the sampling instant. Given enough time or a very fast circuit, this error can be quite negligible (since the slope of the curve decreases with time). Unfortunately, the receiver design has neither of these luxuries. The 128 Msample per second requirement translates into a 7.8 nanosecond clock cycle time. At most half of this (and in reality, much

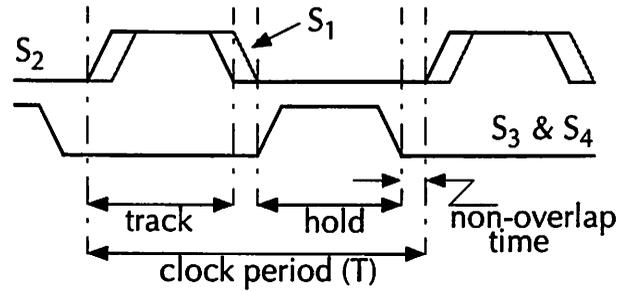


Figure 6.4: Non-overlapping clock phases used to control sample & hold.

less than half) can be used for settling time. Figure 6.4 shows the clock phases for the circuit in figure 6.2. The hold period (and therefore, the settling time) is reduced in length by the non-overlapping period between phases (necessary for proper sampling) as well as by the finite rise and fall time of the clock's edges. The hold period for the receiver's clocks is about 2.5 ns. In order to both avoid degrading the SNR of the signal, as well as to avoid significant gain attenuation, the settling error in the receiver VGA should be kept smaller than half the LSB of the 4 bit converter. In other words, the error must be less than 1 part in 32 (3.125%) of the signal level:

$$e^{-t/\tau} = 0.03125$$

$$t/\tau = \ln\left(\frac{1}{0.03125}\right) \cong 3.47$$

The above show that settling to this level of accuracy requires approximately 3.5 settling time constants. With $t = 2.5$ ns, this gives τ on the order of 0.7 ns. However, since the VGA design consists of more than one stage, each settling error will add to the others, resulting in a larger error. For example, in order to keep the error lower than one-half LSB for a four stage VGA, each stage would need to settle to less than 1 part in 128 (~0.8%) of the exact value. The result is a τ on the order of 0.5 ns, or over 310 MHz of bandwidth.

Therefore, a careful design of the amplifier is required with settling time (and therefore, closed loop bandwidth) as a primary design constraint. One of the main limiting factors in the design of a high speed sample-and-hold amplifier (SHA) is the operational amplifier used

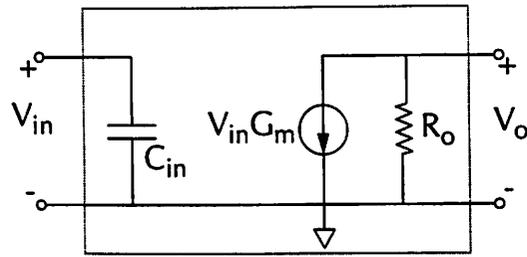


Figure 6.5: Two-port representation of the operational transconductance amplifier

to drive the output to its final value. In order to achieve the fastest possible circuit, a single-stage transconductance topology has been chosen for the op-amp. The details of the design of this operational transconductance amplifier (OTA) are described later in this chapter; however, for the purposes of this section, it is sufficient that the OTA can be modeled as a G_m transconductance with a high output impedance R_o , as illustrated in figure 6.5.

Figure 6.6 shows the SHA during the evaluation phase of its operation. C_I encloses the OTA in a series-shunt feedback loop with feedback factor:

$$f = \frac{C_I}{C_I + C_S + C_p + C_{in}}$$

where C_p is the parasitic capacitance at the summing node, C_{in} is the input capacitance of the op-amp, and C_L is the output load being driven. Given that the open loop gain of the amplifier is $G_m R_o$, the closed loop gain and bandwidth of the circuit can be determined. Including

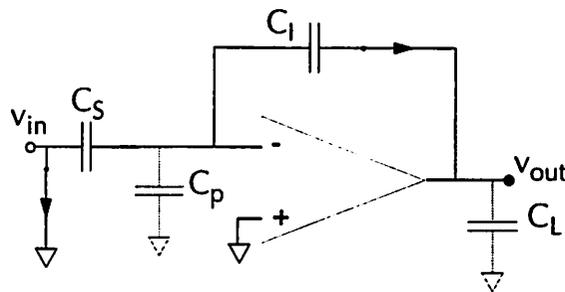


Figure 6.6: Amplifier in hold mode (evaluation phase).

the capacitive divider before the input to the OTA, the closed loop gain becomes:

$$A_{CL} = \left(\frac{C_S}{C_S + C_P + C_{in} + C_I} \right) \left(\frac{A_{OL}}{1 + A_{OL}f} \right) = \left(\frac{C_S}{C_S + C_P + C_{in} + C_I} \right) \left(\frac{G_m R_o f}{1 + G_m R_o f} \right) \frac{1}{f}$$

$G_m R_o f$ is the loop gain T of the circuit and f is the feedback factor as above. Substitution yields the expression:

$$A_{CL} = \left(\frac{C_S}{C_S + C_P + C_{in} + C_I} \right) \left(\frac{C_I + C_S + C_P + C_{in}}{C_I} \right) \left(\frac{T}{1 + T} \right) = \frac{C_S}{C_I} \left(\frac{T}{1 + T} \right)$$

The loop gain term in the above equations is close to unity for large values of T . However, in an AGC, a small reduction in the gain is quite irrelevant since the overall gain in the receiver is affected very little by this variation because of the negative feedback of the control loop. Therefore, lower values of T are acceptable, and the closed loop gain of the amplifier is close to (but not exactly) C_S / C_I . The settling time at the output node is determined by the output impedance of the closed-loop amplifier and by the output load. Therefore, the RC time constant at the output of the amplifier is given by:

$$\tau = \left(\frac{R_o}{G_m R_o f} \right) C_L = \frac{C_L}{G_m f}$$

Clearly, in order to maximize the speed of the amplifier, the feedback factor, f , should be made as large as possible, and the transconductance, G_m , should also be increased if possible. G_m is an op-amp parameter, and its optimization is discussed below. However, the feedback factor relates directly to C_S and C_I . In order to achieve a gain greater than one, the closed loop gain of this topology, C_S / C_I requires that C_S be larger than C_I , reducing the feedback factor by the same increase in gain. An alternative topology is shown in figure 6.7 in which the integrating capacitor is not only used to close the feedback loop around the OTA, but is also used as a second sampling capacitor to capture a sample of charge from the input. During the track mode (as shown in figure 6.7) the two capacitors are shorted together in parallel, effectively making one large sampling capacitor. When the hold, or integration phase arrives, the switches shown in the figure are reversed and the charge on C_S is transferred to C_I . The

closed loop gain of the circuit is now $\frac{C_S + C_I}{C_I}$. In other words, the new topology achieves a larger loop gain for the same feedback factor and bandwidth (a gain of one has been added to the old C_S / C_I gain term). Alternatively, the sampling capacitor can be made smaller for the same closed loop gain, but higher bandwidth. (e.g. C_S must equal C_I to effect a gain of two in the new topology, but in the previous design C_S had to be twice as large as C_I – decreasing the feedback factor.) The final design of the SHA for the VGA is shown in figure 6.8. The center sampling switch across the op-amp’s inputs and an extra clock phase, ϕ_{s2} , have been added to provide better charge injection matching from the sampling operation (charge injection from the center switch will be less dependent on geometry matching than charge from two separate switches). Also shown in figure 6.8 is the control for changing the gain of this stage. When the indicated switch is closed, then the circuit is essentially the same as the circuit in figure 6.7 whose operation was just described as having a closed-loop gain of $\frac{C_S + C_I}{C_I}$. However, when the switch is opened, C_S is removed from the circuit. The closed-loop gain of the circuit now becomes C_I / C_I . Therefore, each stage of the final VGA consists of a SHA providing a gain of either $\frac{C_S + C_I}{C_I}$, or a gain of unity. This capacitive ratio can be chosen to be any value desired; however, increasing C_S also decreases the feedback factor and therefore the bandwidth. Due to the extremely high speed requirements, stages implementing only 3 dB each have been cascaded together to provide 12 dB of total gain. Three dB of gain (or approximately 1.4x) requires a C_S equal to a little less than half of C_I .

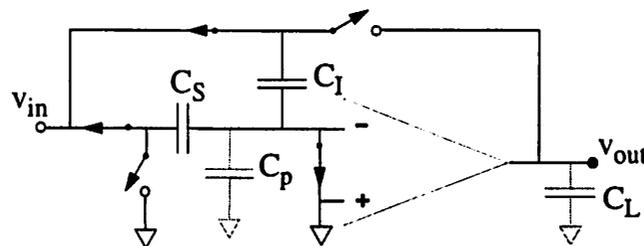


Figure 6.7: Sample-and-hold topology for sampling onto C_I as well as C_S .

Finally, the sample-and-hold design, being fully differential, requires common mode feedback to ensure that the common-mode voltage at the output remains fixed. Figure 6.8 shows the capacitive common-mode feedback that has been added with a switch for reset during the sampling phase. This feedback is essentially the same technique used when the C_1 loop is closed (except, of course that it is common mode), since the capacitors feed back the common mode output to the tail current source of the OTA, as will be seen in the next section.

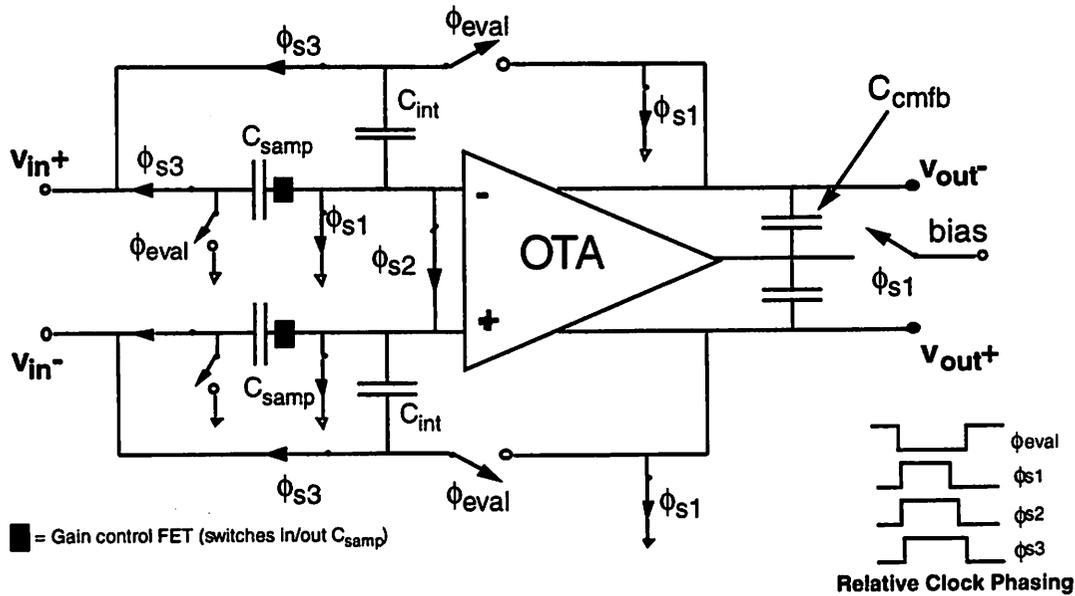


Figure 6.8: Final sample-and-hold topology

6.3. Operational Amplifier Design

The heart of the sample-and-hold amplifiers described in the last section is the operational amplifier. Clearly, the high-speed requirements of the system dictate a simple, high-bandwidth design for the OTA. While most CMOS sample-and-hold amplifiers (for use in a pipeline A/D for example) use two-stage amplifiers to achieve higher gain, the receiver VGA design does not have this requirement. Interstage gain amplifiers in a pipeline A/D require gain accuracy on the order of the resolution of the entire converter (in other words, each stage of an N-bit pipeline A/D must have gain precisely controlled to one part in 2^N) [Conroy94]. However, the negative feedback provided by the overall control loop compensates for any error in the gain of individual stages of the AGC. Therefore, in the trade-off between gain and

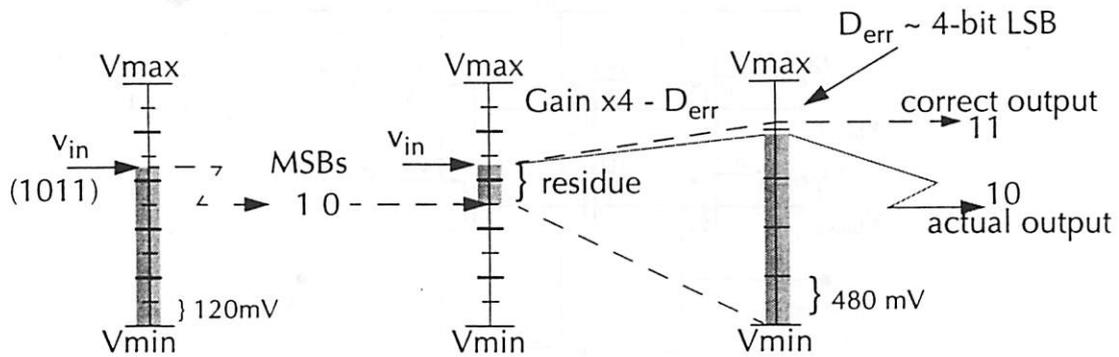


Figure 6.9: 2 bit-2 bit pipeline with small interstage gain error introduced

bandwidth, a lower-gain-but-higher-bandwidth design has been chosen. The effects of gain compression from finite OTA gain are illustrated in figure 6.9 for a pipeline stage and in figure 6.10 for the receiver VGA.

A single-stage telescopic cascode topology is the fastest known op-amp topology available in CMOS technology. The use of common gate cascode transistors eliminates the Miller effect at the inputs, and the low impedance seen looking into the source of a cascode transistor means the circuit is essentially a single-pole system. While a folded cascode has the advantage of increased headroom capability, the inclusion of PMOS transistors in a folded design adversely affects the non-dominant poles. This lower PMOS f_T limits the bandwidth when feedback is applied. Therefore, the telescopic cascode circuit in figure 6.11 has been designed

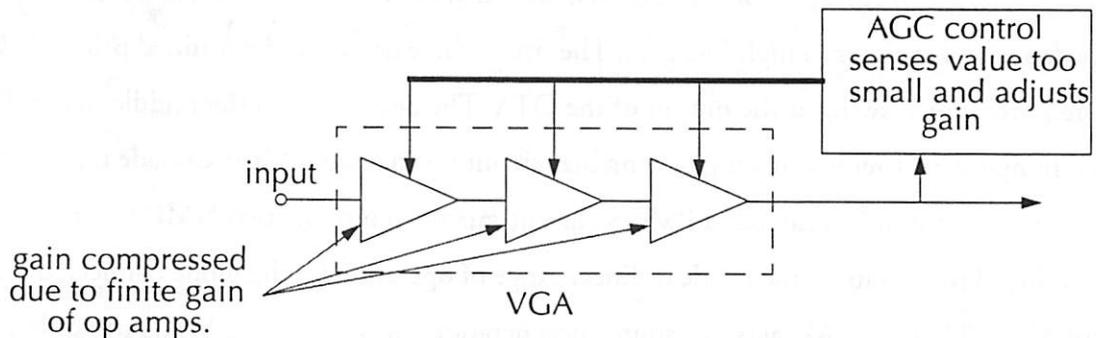


Figure 6.10: AGC compensation of gain errors in the feedforward receiver path.

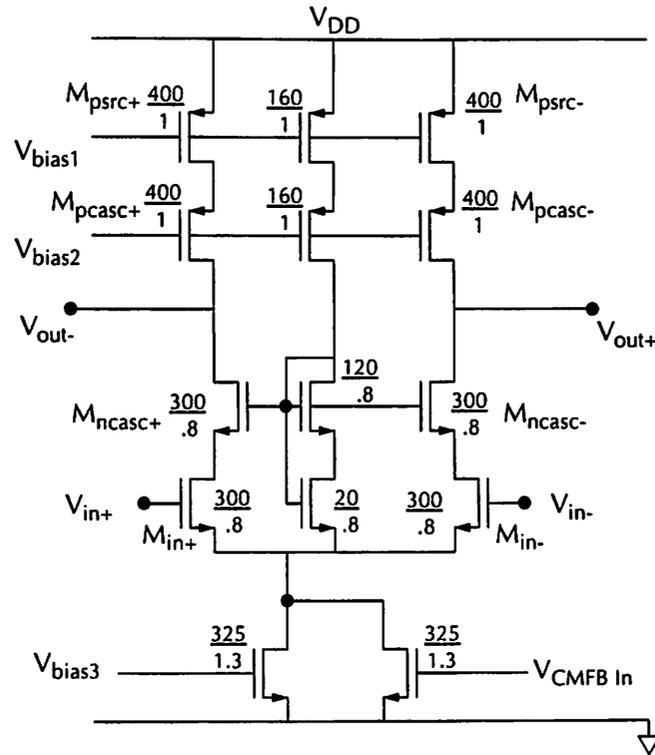


Figure 6.11: Telescopic cascode amplifier. All device sizes are in microns.

for use as an operational transconductance amplifier [Uehara93]. The input transistors and cascode transistors (M_{in+} , M_{in-} , M_{ncasc+} , and M_{ncasc-}) are n-type devices to maximize the bandwidth of the amplifier. The width of the input devices is determined by a settling time optimization analysis presented later in this chapter. PMOS transistors M_{psrc+} , M_{psrc-} , M_{pcasc+} , and M_{pcasc-} form a pair of active current source loads in order to achieve a high output impedance and therefore a high DC gain. The size of these devices is determined primarily by the required signal swing at the output of the OTA. The devices down the middle of the diagram in figure 6.11 represent a high-swing bias circuit for the two NMOS cascode transistors. M_3 and M_4 are simply a cascoded PMOS current mirror to bias the two NMOS transistors. M_1 is forced to operate in the triode or linear range of operation by the diode connected transistor M_2 . Therefore, M_1 acts as a source degeneration for M_2 , and the W/L ratio of M_1 is chosen so that the voltage drop from drain to source will match the desired V_{ds} across the two input devices. This value, V_{ds_input} should be as low as possible without pushing the input de-

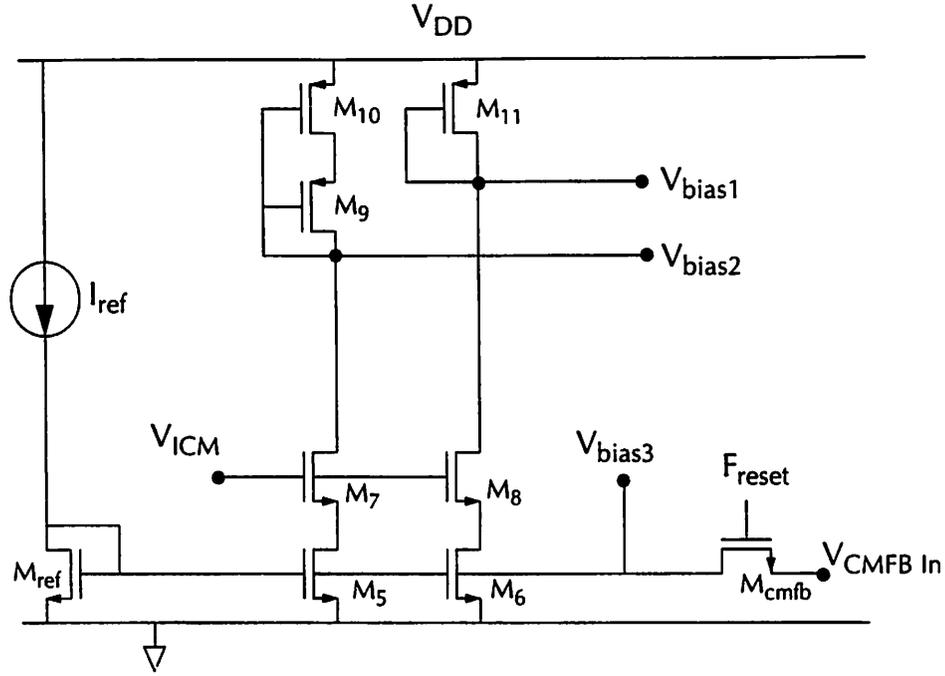


Figure 6.12: Bias generation circuitry for the OTA (shared between all OTA stages)

vices into the triode region of operation. Therefore, the final bias voltage at the gates of the NMOS cascode devices is designed so that the input transistors M_{in+} and M_{in-} have drain to source voltage: $V_{ds_input} = V_{ds_1} = V_{dsat} + V_{margin}$. A very aggressive margin of 150 mV has been chosen so that the available headroom for the signal to swing in the negative direction is maximized as long as M_2 is designed properly to match the V_{gs} drop of the cascode transistors. Therefore, the design equations for M_1 and M_2 are:

$$W/L_2 = \left(\frac{W/L_4}{W/L_{psrc}} \right) (W/L_{ncasc})$$

$$I_{D1} = \mu C_{ox} (W/L_1) \left((V_{gs1} - V_T) V_{ds1} - \frac{1}{2} V_{ds1}^2 \right)$$

The above can be solved for W/L_1 since I_{D1} , $V_{gs} - V_T$, and the desired V_{ds1} are known. The bias circuitry for the PMOS current sources have been designed with the same technique, as shown in figure 6.12. Figure 6.12 also shows the tail current source bias generated from an off-chip current source flowing into an NMOS diode. This diode serves as a reference for all

currents used by the OTA's. The tail current source of the amplifier simply mirrors the current reference, and transistors M_5 , M_6 , M_7 and M_8 also mirror the reference current for use in generating bias for the PMOS current source loads. M_7 and M_8 help to match the current being mirrored into the PMOS current sources with the current flowing in the tail source of the OTA by matching the V_{ds} across M_5 and M_6 to the expected V_{ds} across the tail source (this is critical for matching the currents through the devices – the extremely poor λ of the process means I_{ds} depends strongly on V_{ds}). The tail current source, formed from transistors M_{tail1} and M_{tail2} , is split into two parts so that half of the current source can be used for common mode feedback during the evaluation phase of operation (during the reset phase, both M_{tail1} and M_{tail2} are connected to the reference voltage). The use of only half the transistor in the common mode feedback loop brings down the loop gain around the loop, and helps ensure stability. It also helps protect the circuit from charge injection and ringing from turning on and off the reset switch M_{cmfb} .

6.3.1. Optimizations

Almost any wireless system is designed with the expectation that the received signal will be small. The loss in signal power suffered from transmission through the air usually means that the signal seen by the antenna is significantly smaller than the baseband circuitry (the A/D) can detect. Invariably, amplifiers must be inserted in the receive path of the signal in order to gain the signal up to a detectable level. At the time of this writing, a typical A/D converter might expect an input signal that had been amplified to a maximum range of about +/- 1 volt or even larger. However, sustaining such a large signal swing, especially in a switched-capacitor circuit, requires extra power and area. Also, slewing may begin to hinder the speed of operation for the circuit, and headroom limitations in the OTA may begin to cause distortion or clipping unless a wide-output swing topology is chosen (e.g. a folded cascode). Therefore, it is clearly advantageous to apply only enough gain to the signal so that it meets the minimum power level so that the A/D can accurately resolve it. Adding more gain than this minimum amount can help relax the accuracy requirements for the A/D, but might incur all the potential headaches just outlined. Therefore, in a somewhat backwards design approach, the signal swing at the input of the A/D has been chosen to just meet the minimum detectable signal

given the estimated offset characteristics of the A/D (allowing, of course, for some extra margin). The entire receive path has thus been designed for a maximum signal swing of +/-250 mV (differentially, -0.5 V to +0.5 V).

Figure 6.12 shows the relative sizing of the devices used to implement the OTA; the design process that yields the device sizing shown proceeds as follows:

Recall that $g_m = \frac{2I_D}{(V_{gs} - V_T)}$ for an MOS transistor. Therefore, in order to achieve maximum g_m (and therefore, maximum speed) for a constrained I_D , the current density¹ of the device, $V_{gs} - V_T$, should be minimized without forcing the transistor into the subthreshold region of operation. For the OTA used in the receive chain, a $V_{gs} - V_T = 150$ mV has been chosen. Once the current density has been fixed, the drain current through the device, $I_D = \frac{\mu C_{ox} W}{2L}(V_{gs} - V_T)^2$, becomes proportional to the transistor's W/L ratio. Increasing the width of the input devices certainly improves the g_m of the OTA, but unfortunately the input and output parasitic capacitances also increase – resulting in loss of feedback factor and an increased load that needs to be driven by the OTA. The following equation shows τ as a function of the input device sizes without taking the dependence of part of C_{in} on W/L.

$$\tau = \frac{C_L}{G_m f} = \frac{C_L(C_I + C_S + C_p + C_{in})}{C_I[\mu C_{ox}(W/L)(V_{gs} - V_T)]}$$

By taking the increased self-loading of a large device into account, the above becomes:

$$\tau = \frac{C_L(C_I + C_S + C_p + WLC_{ox})}{C_I[\mu C_{ox}(W/L)(V_{gs} - V_T)]}$$

Taking the derivative with respect to W, and setting the result equal to zero, results in a W/L that corresponds to a local minimum for τ [Conroy94]. The result of the optimization performed for this process is the two 300/0.8 micron input devices presented in figure 6.12. The

1. Current density may be somewhat of a misnomer here since $V_{gs} - V_T$ does not have the units of current per unit area. It does, however, refer to the fact that the $V_{gs} - V_T$ of a MOS transistor is proportional to the square root of the drain current, I_D , divided by the W/L ratio of the device.

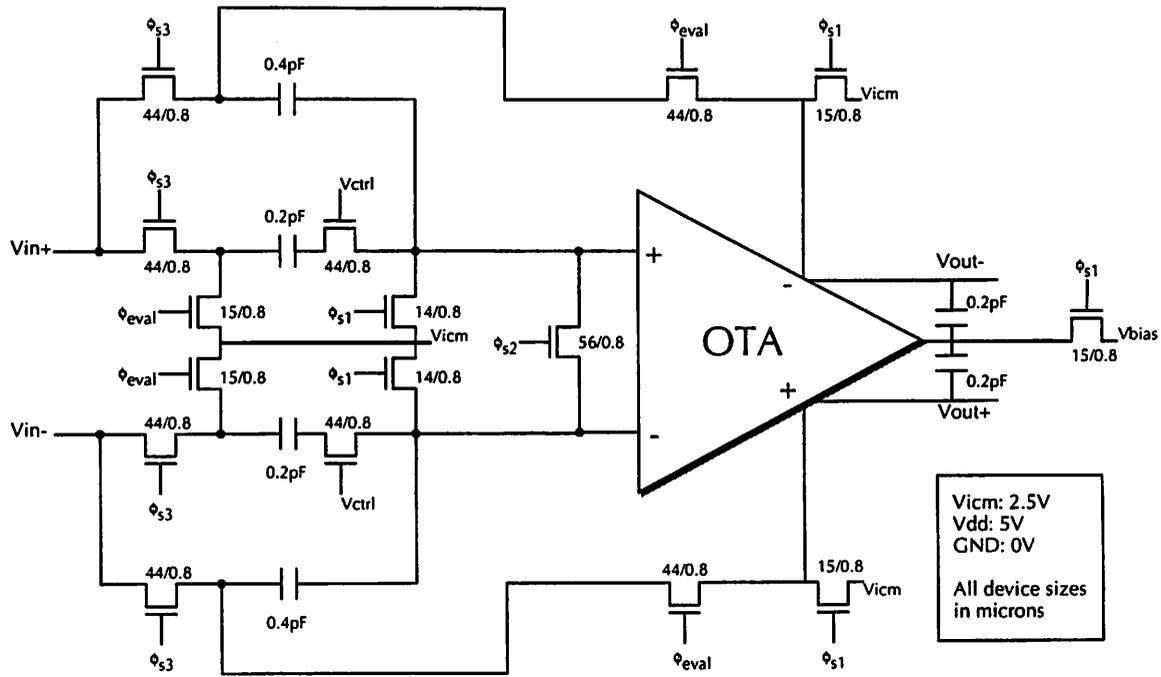


Figure 6.13: SHA switch sizings
(see chapter 5 for discussion of input array, and figure 6.11 for the OTA schematic)

desired current level in the input devices can now be determined from the relation:

$$I_D = \frac{\mu C_{ox} W}{2 L} (V_{gs} - V_T)^2$$

. From this analysis, a total tail current of 2 mA has been chosen. The sizes of all other devices in the circuit are simply chosen to meet the headroom required from the output signal swing. Finally, the tail current source has been designed with a non-minimum channel length in order to increase its output impedance (and therefore, the CMRR).

The designs discussed in the previous sections have been used to implement a small four-stage VGA. Each stage introduces a gain of either 0 dB or 3 dB into the signal path – resulting in a net controllable gain of 0 to 12 dB in 3 dB increments. The C_I and C_S used are 400 and 200 fF each respectively, and the transistor sizes are as shown in figure 6.11 and figure 6.13. The results of this prototype are documented at the end of this chapter.

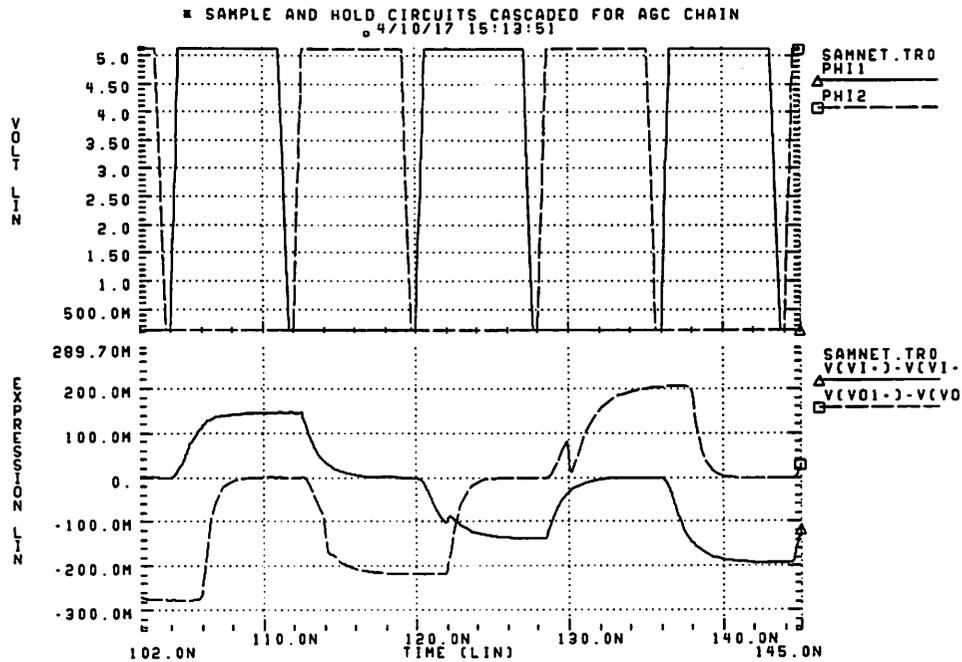


Figure 6.14: VGA stage transient analysis showing reduction in settling time due to signal-dependent kickback.

SPICE simulations of the proposed VGA design (extracted from actual layout) revealed that the sample-and-hold stage met the speed and accuracy requirements dictated by the system while consuming only 10 mW of power per stage. However, when two or more stages are cascaded together, a kickback noise problem was encountered. The problem stems from the large parasitic capacitance associated with the bottom (closest to substrate) plate of the integrating capacitor C_I . At the end of each evaluation phase, C_I and its parasitic C_{Ip} , have charge equal to C_{Vout} stored on them. As illustrated in figure 6.7, when the next track phase arrives, C_I is switched to the input in order to sample the next value. Unfortunately, the charge stored on C_I and C_{Ip} is kicked back to the previous stage. The result is a signal dependent signal excursion at the beginning of the previous stage's settling curve. These kickback excursions severely reduced the amount of time available to the circuit to settle to its final value. A transient analysis of two cascaded stages of the VGA is shown in figure 6.14. The clock rate has been slowed to 64 MHz because of the severe error introduced into the signal at full speed due to

incomplete settling. Simulations show the VGA stages settling to the required accuracy when being clocked at 1/2 the original rate (64 MHz); however, measured test results prove this to be extremely pessimistic (90 MHz was the measured maximum frequency; see section 6.5).

6.4. The Analog-to-Digital Converter

Of all the architectures for high speed A/D converters¹, the flash architecture is the fastest. An N-bit flash converter performs conversion simply by placing 2^N comparators in parallel and determining where, among the 2^N levels, the input lies. For a resolution as small as 4 bits, a flash architecture is a feasible option for a 128 Msample/sec converter. Unfortunately, the amount of hardware and power required for an N-bit flash A/D increases exponentially as 2^N , making a full flash design less attractive because of the extremely low-power requirements of the system. However, even with a very high speed comparator design, exploration of subranging and pipeline A/D converters revealed that the 7.8 ns cycle time was insufficient for performing the extra functions required in these architectures. Furthermore, the addition of a high speed sample-and-hold amplifier (SHA) in between stages of a pipeline converter increases the power consumption due to the static current in the operational amplifier. In fact, the increase in power consumption added by the insertion of a SHA in a pipeline far outweighs the power saved by reducing the number of comparators from 15 to 8 for a 4 bit converter². The conclusion that might be drawn from this analysis is that the correct architecture to choose for the A/D converter topology is a flash. Indeed, for a stand-alone 4-bit CMOS converter operating at 128 Msamples/sec, the flash design would almost certainly be the architecture of choice. However, by merging the A/D with the VGA, further improvements are possible. The final A/D design for the CDMA receiver pipelines one of the four bits, resulting in a 1-bit to 3-bit pipeline converter. This design is depicted in figure 6.15. Since the receiver uses a discrete-time AGC immediately before the A/D converter, each sample-and-hold am-

1. Although certain other A/D topologies, including delta-sigma and successive-approximation converters, may have high clock rates, the term "high speed" here refers specifically to a class of converters called 'Nyquist rate' converters. Unlike oversampled converters (e.g. delta-sigma), this class of converters processes one data sample per clock period, achieving the fastest possible data conversion rate for a given clock speed.

2. Note that for higher resolution converters, the addition of a pipeline stage might very well reduce the overall power consumption of the A/D since the number of comparators increases exponentially with the number of bits, N.

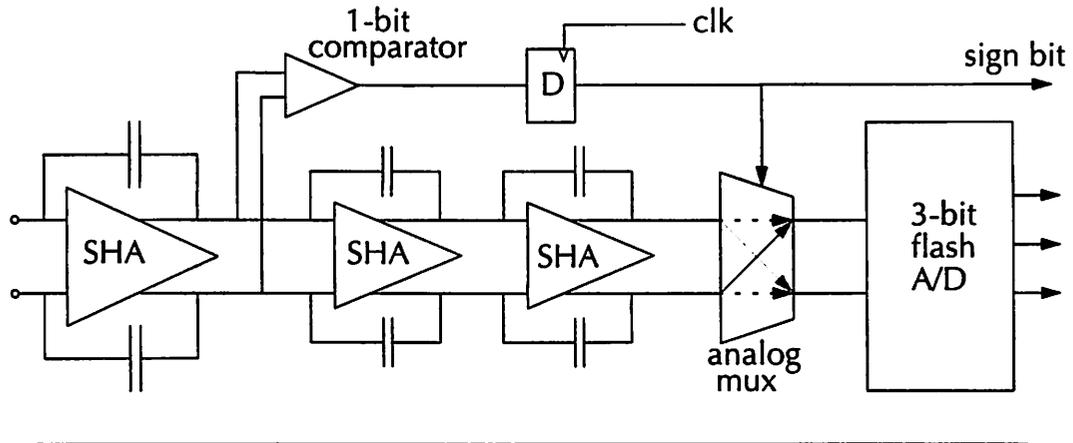


Figure 6.15: Modified 1 bit- 3 bit pipeline A/D architecture

plifier in the AGC looks exactly like the interstage gain amplifier of a pipeline A/D. Therefore, the hardware to perform most of the pipelining function is already in place and could be used by the A/D at no extra cost. The AGC consists of several stages of gain, some of which must be small (for a reasonably fine gain resolution). Therefore, the interstage gain in the pipeline can be made from multiple sample-and-hold amplifiers, allowing the function of generating of a residue (input signal minus analog version of MSB output) to be pipelined itself. It was found that by placing two stages of 3 dB gain (and hence two stages of delay) in between the first comparator and the remaining 3-bit flash A/D, the residue generation function could be split into two stages, thus permitting enough time to generate a residue at 128 MHz.

The receive path is fully differential; by pipelining a single bit of the A/D, the first stage (MSB) comparison becomes essentially a detection of the polarity of the incoming signal. Therefore, the 1-bit MSB subtraction function in a regular pipeline can be changed to a polarity switch (just cross-connected pass gates) based on the results of the first comparator. This concept is illustrated in figure 6.16, and really represents an absolute value function on the input signal, or analog rectification. This seemingly insignificant difference between 1/2 range subtraction and polarity swapping has a surprising effect – since the magnitude of the signal has been maintained, the interstage amplifiers of the pipeline can sustain the gain compression described previously without any detrimental effect. In other words, since the absolute

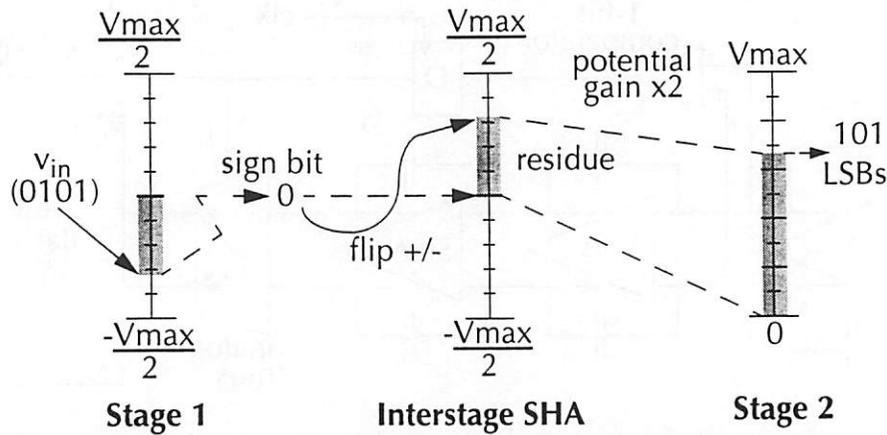


Figure 6.16: Functional illustration of the 1b-3b pipeline converter

value function must be performed by the AGC control block in any case, the polarity flip does not add a non-linearity into the AGC feedback loop – allowing the negative feedback to reject any gain variation in the amplifiers.

It should be mentioned here that the topology in figure 6.15 is not quite a true pipeline A/D. A true 1-bit to 3-bit pipeline topology would require a gain of exactly 2 after the first comparator stage. The two SHA's are really two variable gain stages (0 or 3 dB each) in the AGC. Therefore the total interstage gain between the two comparator stages can be either 0 dB (no gain), 3 dB, or 6 dB (gain of 2) based on the RSSI (received signal strength indicator) detected by the AGC. Indeed, since a fair amount of gain droop can be tolerated in the system, the finite DC gain of the OTA's may cause the total gain to be even less than 3 or 6 dB. Fortunately, the variability of the interstage gain does not have a significant impact on the design. Since the feedback effect of the AGC control loop tries to force the signal at the output of the AGC to be of constant power, the reference ladder for the last stage of the A/D is fixed. The interstage sample-and-hold amplifiers between the two stages may be set to provide no gain at times. When this situation exists, offset requirements of the first stage comparator are identical to those of the last stage. However, when 6 dB of gain exists between the two stages, the signal (and the allowable offset) at the first stage is half that seen in the last stage. As with a traditional pipeline, overranging comparators and digital correction could be implemented to

fix any error made by excessive offset in the first stage. However, as will be described below, the design of the single sign-bit comparator has a smaller offset characteristic than those of the other comparators. Thus, even when the signal swing at the input to the first stage of the pipeline is 1/2 of full range, the converter performs without incorrect codes even without the benefit of digital correction.

6.4.1. Comparator Design

The heart of any analog to digital conversion circuit is the comparator. The comparator performs the quantization function of the A/D by making a decision about the input signal relative to some fixed reference. The most basic of converters, the flash A/D, simply compares the input to each of the 2^N possible discrete values between the maximum allowed input and zero. Each of the comparators decides if the input is larger or smaller than one of the reference levels, resulting in a code of '1's and '0's that can be decoded into a digital word representing the input's magnitude. In this situation, the factor that fundamentally limits the resolution of the A/D converter is the minimum resolvable signal that the comparator can correctly make a decision upon. This key parameter is determined by several different characteristics of the comparator design, including speed of regeneration, overload recovery, and random offset. The primary difficulty in achieving high accuracy in a comparator is the inherent random offset associated with any differential structure. A differential pair ideally has a differential output of zero only when the inputs are exactly equal. However, since the devices and parasitics on each side of the layout can not match each other exactly, this is never the case. The input offset voltage, V_{os} , is the input voltage that compensates for these non-idealities, and brings the outputs to zero. V_{os} is dependent on process variations, temperature gradients, and geometry mismatches of the design. For a simple MOS differential pair, V_{os} is given by [Gray84]:

$$V_{os} = \frac{\Delta V_T}{V_T} + \frac{(V_{gs} - V_T)}{2} \left[\frac{\Delta(W/L)}{W/L} + \frac{\Delta Load}{Load} \right]$$

An error in the comparator decision can be made if $|V_{in} - V_{ref}|$ is less than or equal to $|V_{os}|$. As always, a trade-off exists between speed and accuracy. Therefore, the high speed throughput required of the A/D converters in the receiver prohibits the use of most techniques for

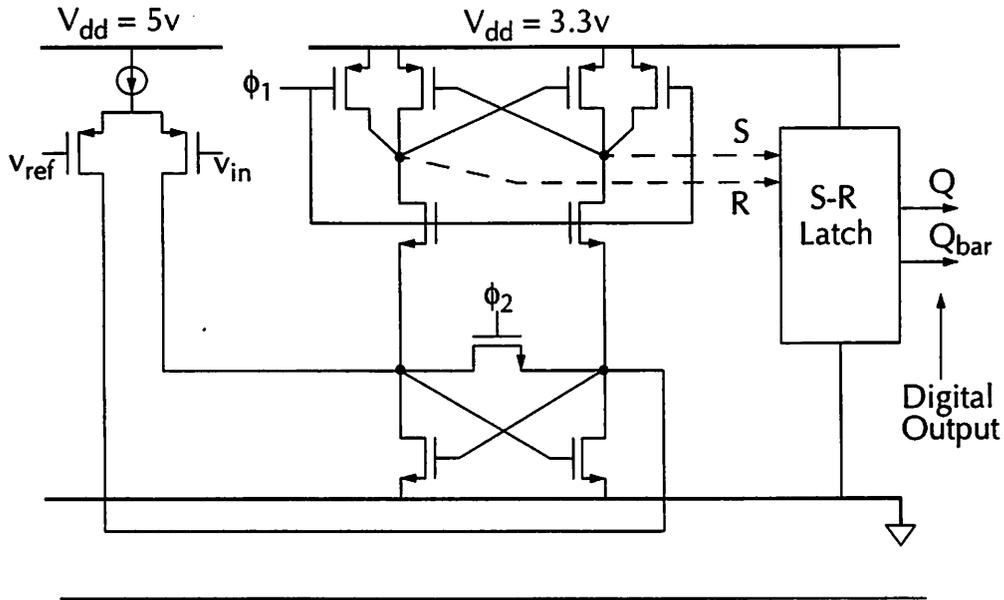


Figure 6.17: High-speed, low-offset comparator (after [Yin92])

combating the offset problem (e.g. offset cancellation, large preamplification). However, a comparator topology has been proposed that achieves a very good combination of speed and accuracy and is shown in figure 6.17 (after [Yin92]).

This comparator topology has a single-ended input and therefore needs to be modified for the receiver's differential signal path. This could be accomplished through the use of a capacitive input sampling network included immediately preceding the comparator, such as the one shown in figure 6.18. Much like the sample-and-hold amplifiers, the network captures a sample of charge proportional to $v_{in+} - v_{icm}$ on the sampling capacitors during ϕ_2 . When ϕ_1 closes, one side of C_{S+} is shorted to v_{ref+} , forcing the other side of the capacitor to

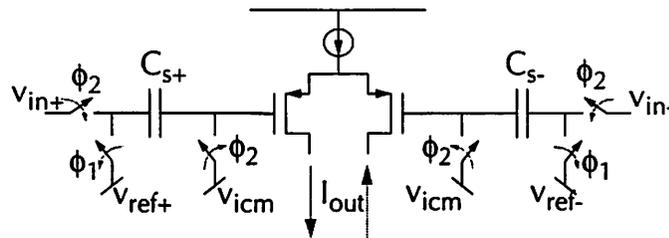


Figure 6.18: Capacitor-based input-sampling network

$v_{in+} - v_{icm} - v_{ref+}$. Similarly, the other side of the circuit generates $v_{in-} - v_{icm} - v_{ref-}$. Since v_{icm} is a common mode voltage, the differential pair generates a differential output current proportional to $(v_{in+} - v_{in-}) - (v_{ref+} - v_{ref-})$. Therefore, the inclusion of this additional circuit has a two-fold purpose. Not only does it provide an efficient differential-to-single ended conversion for the comparator, but it also performs a rudimentary sample-and-hold operation that might be useful if the A/D is not preceded by an active sample-and-hold circuit. The extra sampling operation provided by the input network of figure 6.18 is certainly not needed in the receiver, since the input signal has already been sampled and held. Furthermore, a network that depends on large, very precisely matched capacitors becomes very unattractive in a standard CMOS process like the one available. Not only does the addition of large capacitors severely affect the input bandwidth of the A/D, it also becomes very costly in terms of area when the process used does not support a second layer of polysilicon. As discussed in chapter 5, precision capacitors for the receiver have been created using the dielectric between the first-to-second and second-to-third metal layers. These capacitors consume so much area, and incur so much unwanted parasitics, that the prospect of creating $2 \cdot 2^N$ such capacitors for an N-bit flash converter is extremely unattractive. Therefore, a modification of the original comparator has resulted in a design capable of accepting a fully differential input, and comparing it to a differential reference without the benefit of an input sampling network. As shown in figure 6.20, the original PMOS differential pair input has been replaced with the modified double-differential pair topology shown in figure 6.19b. The output of the circuit is given by the difference between the two currents I_{out_r} and I_{out_l} :

$$I_{out} = \left(\frac{-g_m}{2} \right) ((v_{in-} - v_{ref-}) + (v_{ref+} - v_{in+})) - \left(\frac{-g_m}{2} \right) ((v_{ref-} - v_{in-}) + (v_{in+} - v_{ref+}))$$

Therefore, if the bias currents match exactly, the output current is proportional to $(v_{in+} - v_{in-}) - (v_{ref+} - v_{ref-})$, which is exactly the output generated with the input sampling network. Notice that this topology does not use the differential pairs in the traditional manner – v_{in+} is not being compared directly with v_{in-} . A more standard topology is shown in figure 6.19a. The differential output current is identical to that of the output of the modified circuit, but there is a significant difference between the two. Suppose a full range input signal

is being compared to the maximum reference (i.e., $v_{in+} - v_{in-}$ is at its maximum value, as is $v_{ref+} - v_{ref-}$). In order for the differential pairs to stay in the linear range of operation (i_{out} proportional to differential input voltage), the $V_{gs} - V_T$ bias for the input devices must be larger than one-half the differential input. In other words, if the devices in the differential pair were biased with a $V_{gs} - V_T = 200\text{mV}$, and 500mV were to be applied across the input terminals, then one side of the differential pair would be completely shut off, while the other would carry the entire current from the tail source. Unfortunately, when trading off bandwidth for power, $V_{gs} - V_T$ should be kept as small as possible to achieve the highest bandwidth for the smallest power (since $g_m = \frac{2I_D}{(V_{gs} - V_T)}$, this is really a matter of maximizing g_m for a given current level). Therefore, making the $V_{gs} - V_T$ bias as large as the full scale input voltage swing adversely affects either the speed or the power consumption of the circuit; likewise, a large $V_{gs} - V_T$ also reduces the available headroom on the input stage. However, the proposed design in figure 6.19b does not have this problem. Since v_{in+} is paired with v_{ref+} , and v_{in-} is paired with v_{ref-} , both differential pairs will be well within their linear range of operation even when both the input and the reference are near their maximum values. While it is true that

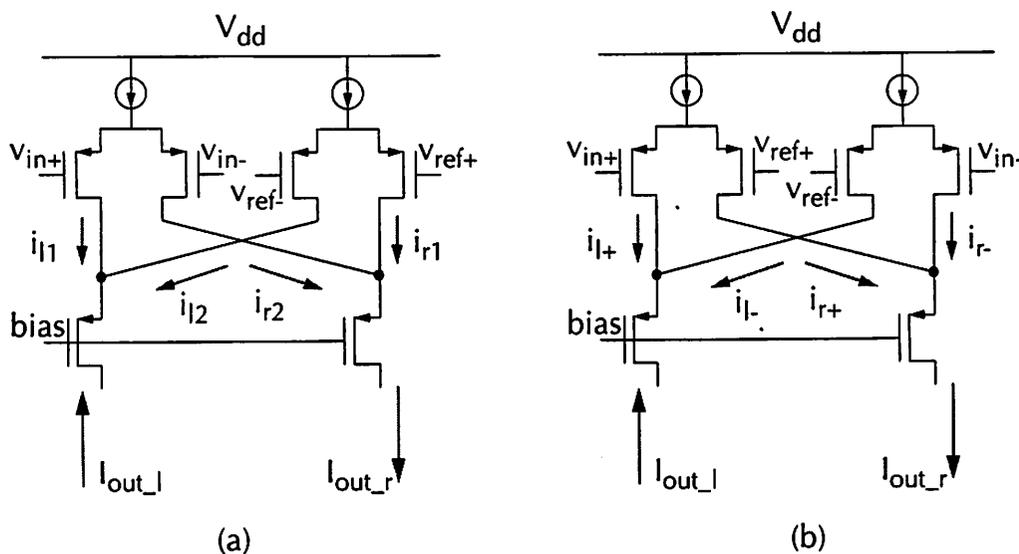


Figure 6.19: Fully-differential input sampling network. (a) Standard topology; (b) Modified topology employed by the 1b-3b pipeline A/D. Note the change in inputs

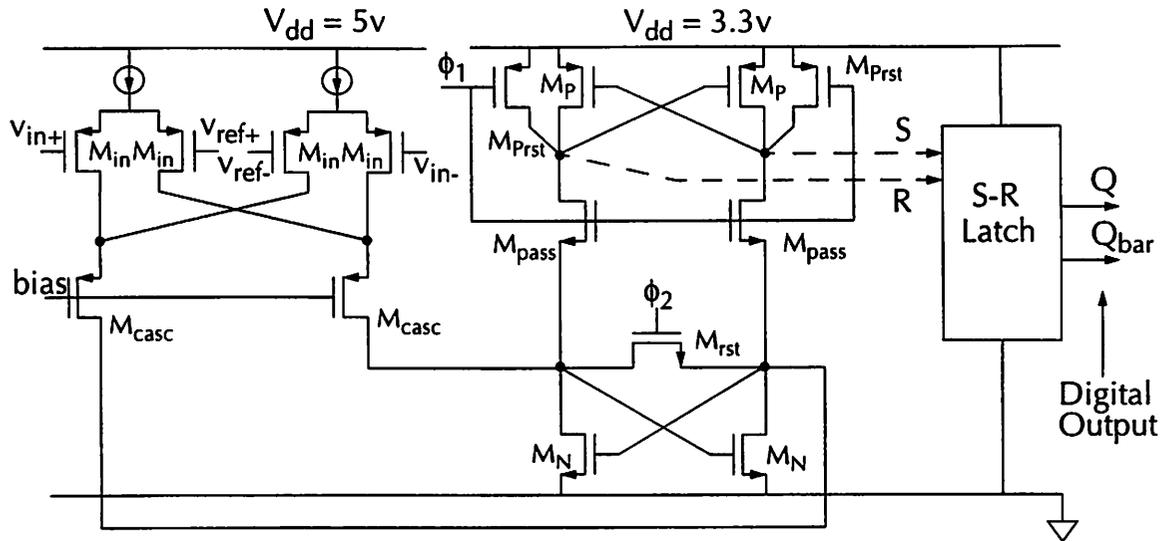


Figure 6.20: Final high-speed differential comparator

the differential pair will become unbalanced when the input differs greatly from the reference voltage, this is of little importance because a comparator is only interested in the signal when it is near the crossing point with the reference.

The final comparator topology is shown in figure 6.20. As described above, the input stage performs a conversion from a doubly-differential voltage to a single differential current. PMOS cascode transistors have been added to the input stage to both improve the output impedance, as well as to eliminate the Miller effect at the input. The cascode devices also help to improve the regeneration speed of the comparator since the parasitic capacitance on the drains of the cascodes can be reduced by undersizing the devices (the input devices must necessarily be larger to achieve a higher g_m). Therefore, the input stage can be regarded as a single-pole voltage-to-current converter with a high output impedance. Keeping this in mind, the operation of the rest of the circuit can now be analyzed. The comparator requires two clock phases, as shown in figure 6.21. The two phases must be non-overlapping, a fact which is common for many A/D converters, but is especially important for the operation of this particular topology. The clock generation circuitry is slaved off the multiphase clock generator used for the sampling demodulator (section 5.5.2).

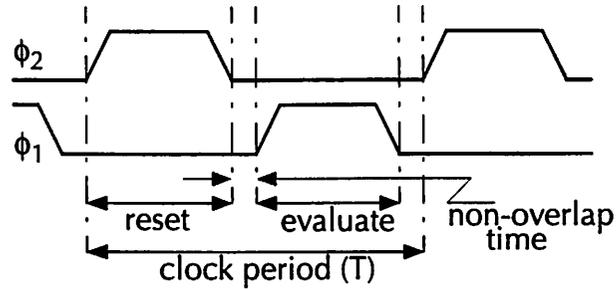


Figure 6.21: Non-overlap clock phasing for the comparator.

The first phase of operation can be called the reset phase and is illustrated in figure 6.22. The output nodes, which go to the S-R latch, are reset to V_{dd} . During this time, the S-R latch maintains its last value. Nodes 1 and 2 are shorted together by an NMOS switch with on resistance R_{on} , providing a good overload recovery for circuit¹. Although M_1 and M_2 are connected in a positive feedback loop, the loop gain is forced to be less than 1 by the reset switch

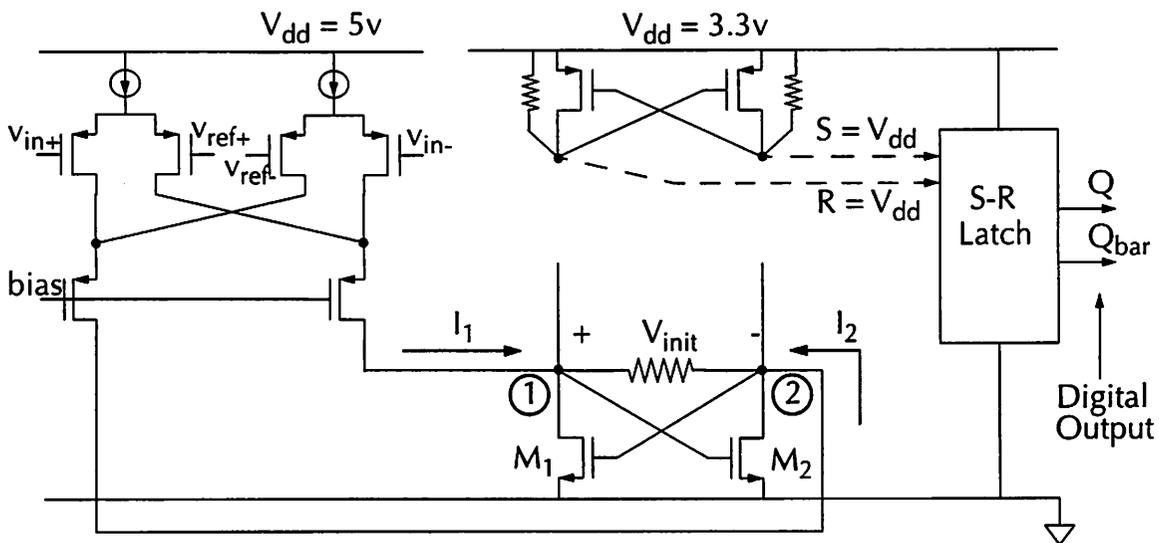


Figure 6.22: Comparator during reset phase of operation ($\phi_1 = 0, \phi_2 = 1$)

1. Overload recovery refers to a comparator's ability to recover from evaluating a sample and then correctly evaluate a subsequent value.

– making the two transistors look diode connected. However, the input stage is acting as a differential current source, forcing current down M_1 and M_2 . The differential current flowing down the two devices will create a differential voltage at the two drain nodes, with each device acting as a $1/g_m$ impedance load. Therefore, the voltage shown as v_{init} in the figure is proportional to the comparator's input voltage by the g_m ratio of the input devices to the NMOS latch devices (M_1 and M_2). The function of the reset phase is to provide a good recovery from the previous sample and to set up an initial voltage proportional to the input (but smaller) as the starting point for regeneration of the NMOS latch. At the end of the reset phase, clock ϕ_2 goes to zero. After the falling edge of ϕ_2 , there is a brief period of time during which both ϕ_2 and ϕ_1 are low. This 'non-overlap' period is typically regarded simply as a separator between phases of operation that guarantees that one phase will not interfere with the operation of another. However, at such a high frequency of operation, the brief non-overlap period represents a significant percentage of the clock period, and really can not be sacrificed as an idle period. The state of the proposed comparator during the period when both clocks are low is shown in figure 6.23. The reset switch across nodes 1 and 2 has been opened. Therefore, the differential current being provided by the input stage no longer sees two diode-connected loads. If the voltage v_{init} is small, then M_1 and M_2 are biased at about the same level, and act

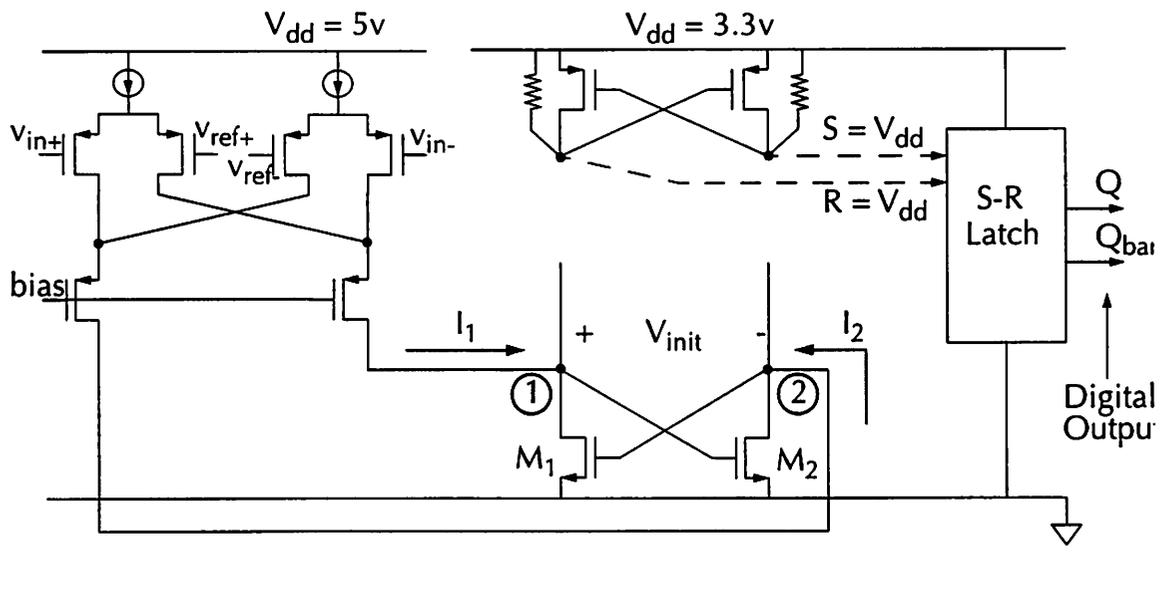


Figure 6.23: Comparator during non-overlap period ($\phi_1 = 0, \phi_2 = 0$)

as a pair of NMOS active loads. If the initial voltage, v_{init} , is larger than a few tens of millivolts, then the positive feedback of the latch will force the two nodes to split apart further towards either supply. However, if v_{init} is small (corresponding to an input voltage very close to the reference voltage) then the currents flowing from the input stage will see the output impedance, r_o , of M_1 and M_2 . Therefore, the voltage gain of the circuit changes from $A_{v1} = -\left(\frac{g_{m_input}}{g_{m_{1,2}}}\right)$ to $A_{v2} = -(g_{m_input} r_{o_{1,2}})$. While the first term, $g_{m_input} / g_{m_{1,2}}$ is actually less than one, the second term represents an amplification of the signal. The voltage across nodes 1 and 2 will begin to approach $v_{in} \cdot A_{v2}$ until either the non-overlap period ends (and the evaluation phase begins), or the voltage between the two nodes becomes large enough for the positive feedback of the latch to begin to regenerate it. It is not uncommon for comparator topologies to include a pre-amplifier (in the form of a differential pair) immediately before a latch because the preamp reduces the input referred offset of the latch by its gain.¹ The proposed topology folds the preamp and the latch together by using the gain achieved during the non-overlap period to reduce the input referred offset of the latch once the evaluation period begins. The evaluation phase of operation begins when ϕ_1 rises (figure 6.24). The two PMOS switches that shorted the output nodes to the supply are opened, and

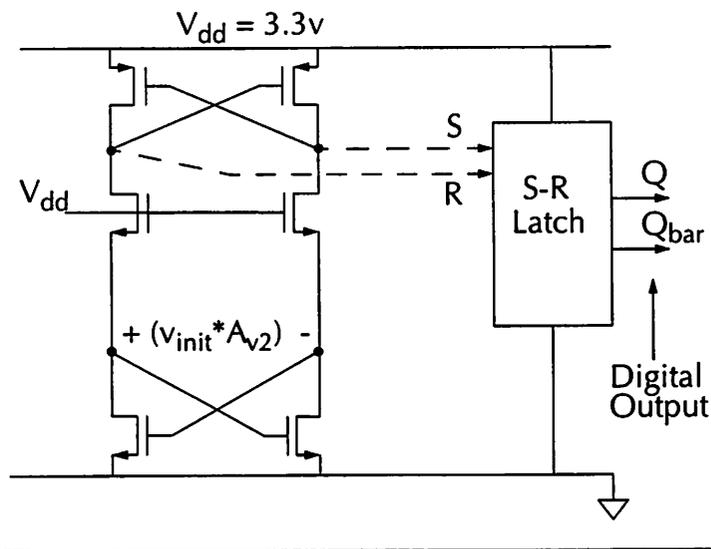


Figure 6.24: Comparator at the beginning of the evaluation phase ($\phi_1 = 1, \phi_2 = 0$)

1. A latch has a fairly bad offset characteristic because the positive feedback regeneration grows exponentially with time. Therefore, a small signal will start out with a slow rate of change.

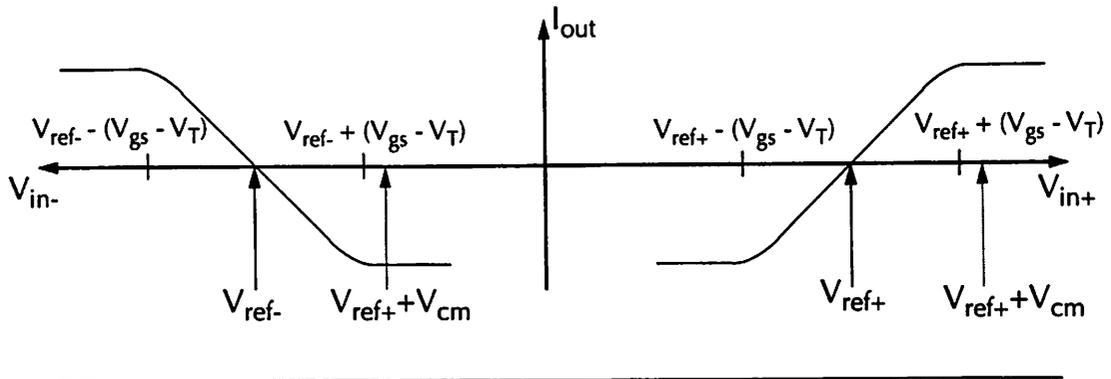


Figure 6.25: Effect of common-mode shift on the modified input stage

the two NMOS switches connecting the PMOS latch to the NMOS latch are closed. The resulting circuit is simply two inverters connected in positive feedback (a full latch), which regenerates the signal to near-digital levels. The S-R latch has been added to both bring the final output to a full digital swing, as well as to hold the comparator's last output value during the reset phase. The final combination of doubly-differential input stage, high-impedance cascode, non-overlap period preamplification, and S-R latch results in a comparator with low offset, good overload recovery, little kickback noise, low power, and high speed [Lynn95].

6.4.2. A/ D Optimizations

The previous section explained the advantages achieved from using the modified input stage shown in figure 6.19b over the standard double-differential stage of figure 6.19a. With its limited linear range, the differential pair limits the ability of the circuit in figure 6.19a to compare a large input to a large reference voltage; the improved design in figure 6.19b avoids this because it places the comparator switching point exactly in the middle of the differential pair's linear range. However, if the common mode voltage of the input is different from the common mode voltage of the reference, then the switching point no longer falls exactly in the middle of the differential pair transfer function, and results in an offset in the comparison point. This concept is illustrated in figure 6.25. The differential output current is given by the sum of the two currents given by the curves shown in the figure. However, if the difference in common mode voltage shifts the comparison point outside the linear region of the circuit (as shown by the light grey arrows in figure 6.25), then the differential output near the switching

point will be zero, regardless of the input's value. In other words, the modified topology of figure 6.19b works well even for large signal swings, but begins to fail if the common mode difference between input and reference varies significantly from zero. Therefore, the $V_{gs} - V_T$ bias of the input devices must be chosen so that the circuit will remain linear during the largest expected common mode excursion. Unfortunately, the input offset voltage of a MOS differential pair is given by:

$$V_{os} = \Delta V_t + \left(\frac{V_{gs} - V_t}{2} \right) \left[\frac{\Delta(W/L)}{W/L} + \frac{\Delta Load}{Load} \right]$$

Clearly, the offset due to geometry mismatch in the devices increases with $V_{gs} - V_t$. Furthermore, the g_m of the input stage is $\frac{2I_D}{V_{gs} - V_t}$. Not only can the g_m for a given current can be maximized by reducing $V_{gs} - V_T$, but the offset can be reduced in this manner as well. Therefore, the bias on the input devices must be carefully designed in order to have a minimal $V_{gs} - V_T$, while still ensuring some robustness against common mode excursions. To this end, an input $V_{gs} - V_T = 200\text{mV}$ has been chosen. In order to meet the stringent power budget allocated for the A/D, a total tail current of 40 mA was allowed (10 mA flowing through each input transistor). Using the drain current equation for an MOS transistor,

$I_D = \frac{\mu C_{ox} W}{2L} (V_{gs} - V_T)^2$, the W/L sizing for the input devices can be determined. The choice of $8\mu/1\mu$ geometry input PMOS's leads to an estimated worst case input offset of 25-30 mV for the comparator, just below half of an LSB ($\frac{\Delta(W/L)}{W/L} \cong 0.05$, $\Delta V_t \cong 10\text{mV}$).

Once the input device dimensions are determined, the cascode transistors can not be far behind, else the nondominant pole at the source of the cascode device becomes significant.

Therefore, the two cascode transistors have been designed to be half the size of the input devices in order to reduce the capacitance contributed to the NMOS latch. The sizes of the various transistors used in the comparator are shown in figure 6.26. The NMOS latch transistors should be made as large as possible in order to achieve a faster regeneration speed by increasing current drive. However, the parasitic capacitance at the two nodes 1 and 2 is partially determined by the width of transistors MN. Therefore, an optimum transistor width can be found from the time constant equation:

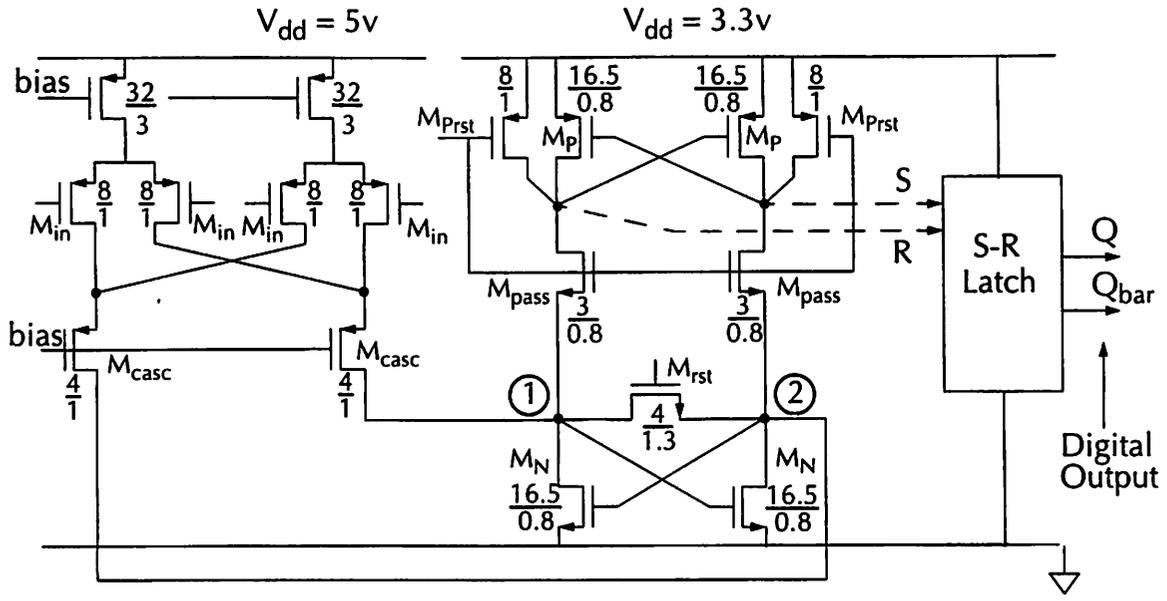


Figure 6.26: Comparator device sizing (microns)

$$\tau = \frac{C_{1\text{total}}}{g_{mN}} = \frac{C_{\text{parasitic}} + \alpha W_N}{\sqrt{2I_N \mu C_{\text{ox}} (W_N/L)}}$$

where α is the capacitance per unit width contributed to node 1 (or node 2) by the two M^N transistors. Optimization of the above yields:

$$W_N = \frac{C_{\text{parasitic}}}{\alpha} = 16.5 \mu\text{m}$$

The initial voltage from which regeneration will start is determined by:

$$V_{\text{init}} = -\left(\frac{g_{\text{min}}}{g_{mN} - (2/R_{\text{onrst}})}\right)(V_{\text{in}} - V_{\text{ref}})$$

Therefore, the R_{on} of the NMOS reset transistor must be made large enough to ensure a sizeable initial signal voltage, but small enough to provide a good overload recovery during the reset phase. Since the primary purpose of the reset switch is to drop the positive loop gain of the latch below one, the design equation becomes:

$$(-g_{mN}R_{\text{onRst}})(-g_{mN}R_{\text{onRst}}) < 1$$

The above evaluates the loop gain around the latch. By taking the square root of both sides, and substituting for g_m and R_{on} , the relative sizing of transistors M_N and M_{rst} can be determined:

$$\left(\sqrt{2\mu C_{ox} I_N (W/L)_N}\right) \left(\frac{1}{\mu C_{ox} (V_{gs} - V_t)_{rst} (W/L)_{rst}}\right) < 1$$

$$(W/L)_{rst} > \frac{\sqrt{2\mu C_{ox} I_N}}{\mu C_{ox} (V_{gs} - V_t)_{rst}} \sqrt{(W/L)_N}$$

For the 0.8 micron CMOS process available, the above becomes:

$$(W/L)_{rst} > \left(\frac{1}{6}\right) \sqrt{(W/L)_N} \cong \frac{2}{3}$$

A $W/L_{rst} = 4/1.3$ microns was chosen for the reset switch, but probably could have been more optimally designed as a smaller (i.e., lower W/L ratio) device.

Device sizes for the other transistors in the circuit have less impact on the comparator's performance. NMOS transistors M_{pass} were sized to minimize the parasitic capacitive load placed on the NMOS latch. The PMOS latch devices, M_p were sized to match the current drive of the NMOS latch transistors as is done in any digital design (once the evaluation phase arrives, the n-latch combines with the p-latch to effectively make two cross-coupled digital inverters). The final schematic and device sizings are as shown in figure 6.26.

It should be mentioned here that the sign bit comparator performs a comparison against a zero-valued reference voltage. In other words, this one comparator does not need four inputs, since it only determines if v_{in+} is greater or less than v_{in-} . The device geometries for this comparator all remain the same as for the comparator just described in this section, but the lack of the two extra inputs gives this comparator a better offset characteristic. It is the increased resolution of this design that allows it to be used in the 1-bit to 3-bit pipeline. Since the two interstage SHA's between the sign bit comparator and the 3-bit flash A/D represent a potential gain of 6dB, the sign bit comparator must have more accuracy in order to resolve a smaller signal.

SPICE simulations of the proposed comparator design (extracted from actual layout) showed the comparator performing at the speed and accuracy requirements for which it was designed. Figure 6.27 shows a transient output from a comparator simulation. The top panel of figure 6.27 shows the input samples to the comparator. The horizontal line across the panel corresponds to the reference voltage. The second panel shows the digital output of the SR latch switching every time the input crosses the reference voltage. The bottom panel shows the two non-overlapping 128 MHz clocks used to control the comparator. Simulation of the entire A/D converter, although too cumbersome to include here, was consistent with the simulation of each individual comparator.

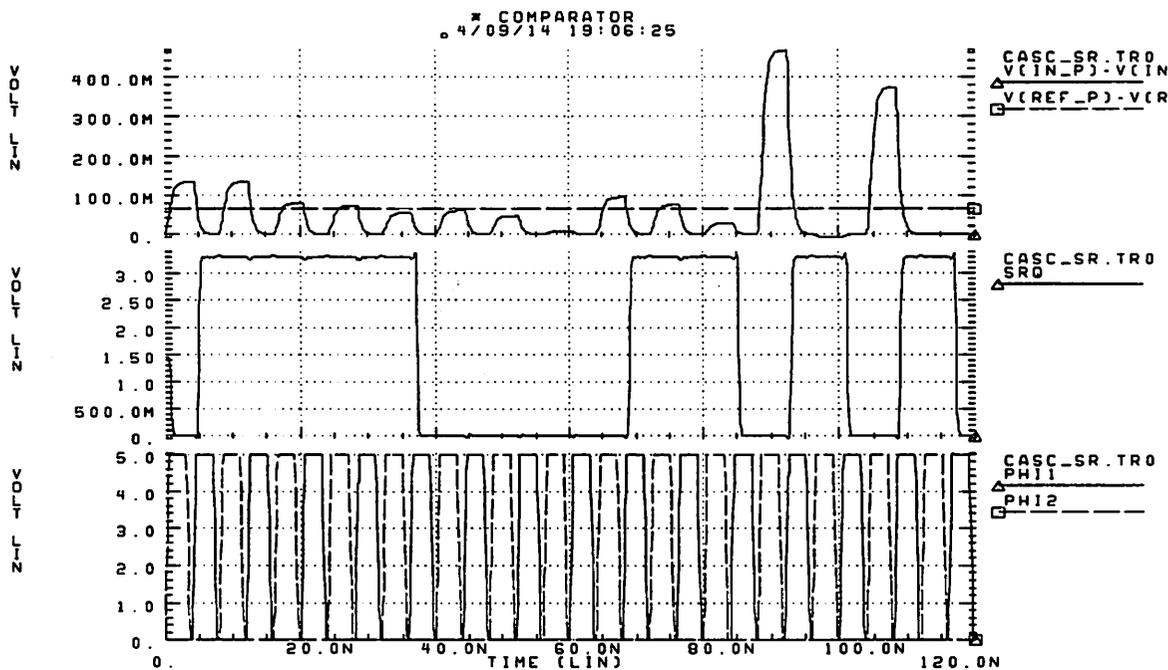


Figure 6.27: Single comparator transient analysis showing correct digital output based on comparison of input to reference.

6.5. Results

Three chips have been fabricated in a standard HP non-double-poly 0.8 micron digital CMOS process. The first die, designated “minisporf”¹, consists of the proposed 3-bit flash A/D structure. A diagram of the 3-bit flash converter layout is shown in figure 6.28, and a

close-up diagram of one of the comparators is shown in figure 6.29. It is important that the comparator layout be as symmetric as possible in order to prevent an increase in V_{os} due to geometry mismatch. Substrate and well contacts can be seen surrounding the circuits in the bottom half of the figure, forming rectangular guard rings to collect as much substrate noise current as possible. The signal flows from the bottom of figure 6.29 to the top where the SR latch can be seen (sandwiched between two bypass capacitors). It is also worth noting that the clock distribution is horizontal across the circuit (and not directly above any transistors) so that inductive and capacitive coupling between the clock and the signal path (which flows vertically through figure 6.29) is minimized.

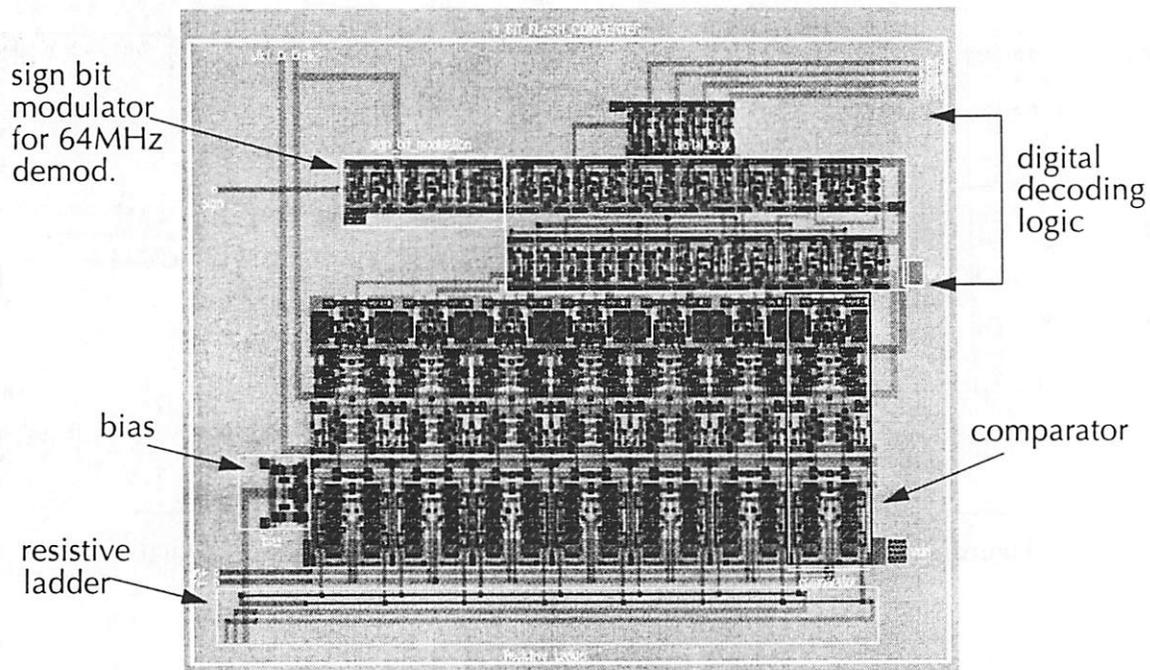


Figure 6.28: 3-bit flash comparator layout

1. The reader may question the origin of the name “sporf”: it stands for “Sam and Poe’s Outstanding RF Chip”. In actuality, it should be “sporc”, but that doesn’t sound nearly as cool.

A diagram of the 3-bit A/D test chip (which includes a sample-and-hold circuit) is shown in figure 6.30. Large on-chip bypass capacitors can be seen distributed throughout the die. The signal enters the chip from the left-hand side of figure 6.30 and the digital outputs exit the chip on the right of the figure. All pads on the right-hand side of figure 6.30 are reserved for digital signals (including digital supply and clocks) and the analog and digital supply and ground connections are completely separated on the chip (except for the fact that the grounds must be connected through the substrate of the die).

Speed tests on this die have shown that the A/D functions up to a 150 MHz clock rate (the goal of the design was 128 MHz). Beyond 150 MHz, the converter does not have sufficient time to evaluate a signal and convert it to a full digital level. Total power consumption from the analog circuitry is 2.2 mW. Digital power consumption for the converter could not be determined because the vast majority of digital power consumption on the test chip is created by the buffers driving the outputs off chip at the full 128 MHz rate (which would not exist in a fully integrated implementation).

The second die (designated minisporf2) was mentioned in chapter 5 with regard to RF layout of a sampling demodulator; it consists of a standalone I/Q demodulator, with analog outputs to measure distortion performance and RF analog track bandwidth. Settling performance of the OTA could not be verified with this die; the output capacitive loading on the OTA is that

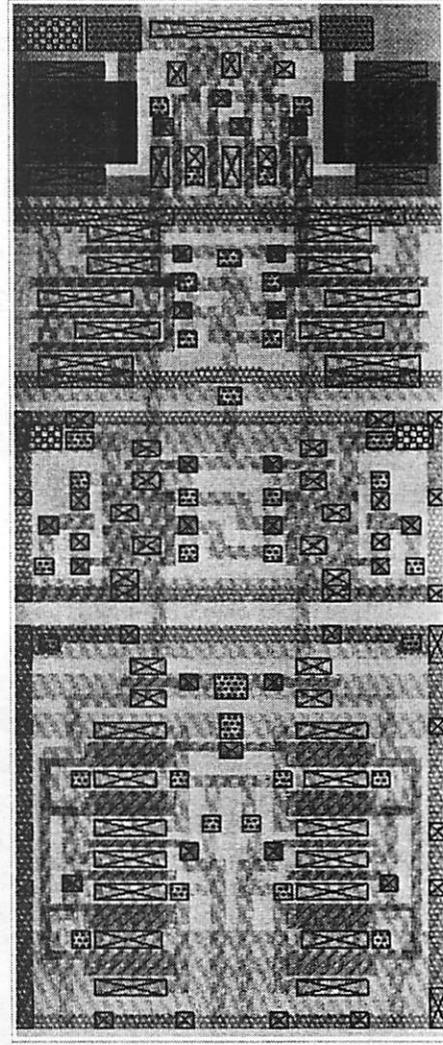


Figure 6.29: Comparator layout

of the pad, board-level parasitics, and the scope probe: far greater than the on-chip load it was designed to drive. Because of this, the die could not be clocked any faster than 20 MHz; however, this was sufficient to acquire the measurement results that it was designed for.

The die photo is shown in figure 6.31. Measured -3 dB point on the input track bandwidth is 1.78 GHz (compared to a simulated 2.1 GHz), and its IP_3 distortion intercept was +22 dBm (slightly higher than the simulated +20.4 dBm). To measure the distortion intercept, two equal-power 0 dBm sinusoidal input tones were injected, one at 1.020 GHz and the other at 1.021 MHz. Subsampling at 20 MHz, the third order intermod component at 2 MHz was

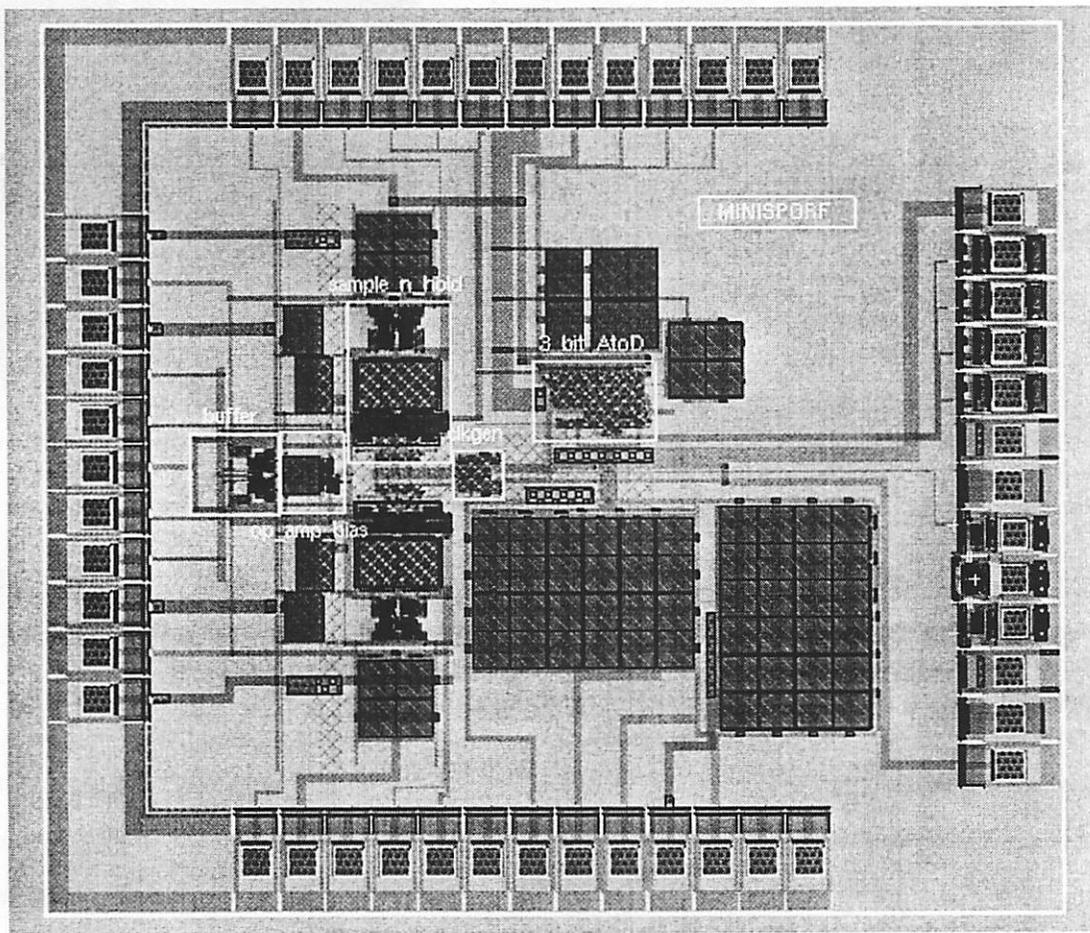


Figure 6.30: 3-bit flash A/D converter testchip (minisporf)

measured using an HP 70000-series spectrum analyzer. These results agree well with the simulated results discussed in chapter 5, and are well within the accuracy of the device simulation models provided by the foundry; a summary table is provided as part of figure 6.33.

Lastly, the final die (designated “sporf”) represents the entire analog front-end testchip: it includes the full 1 bit- 3-bit A/D converter, the complete 4-stage VGA chain, the sampling demodulator, and the RF LNA. Due to time constraints, the AGC control loop has not yet been designed/implemented; however, the gain is externally controllable, and can be set on

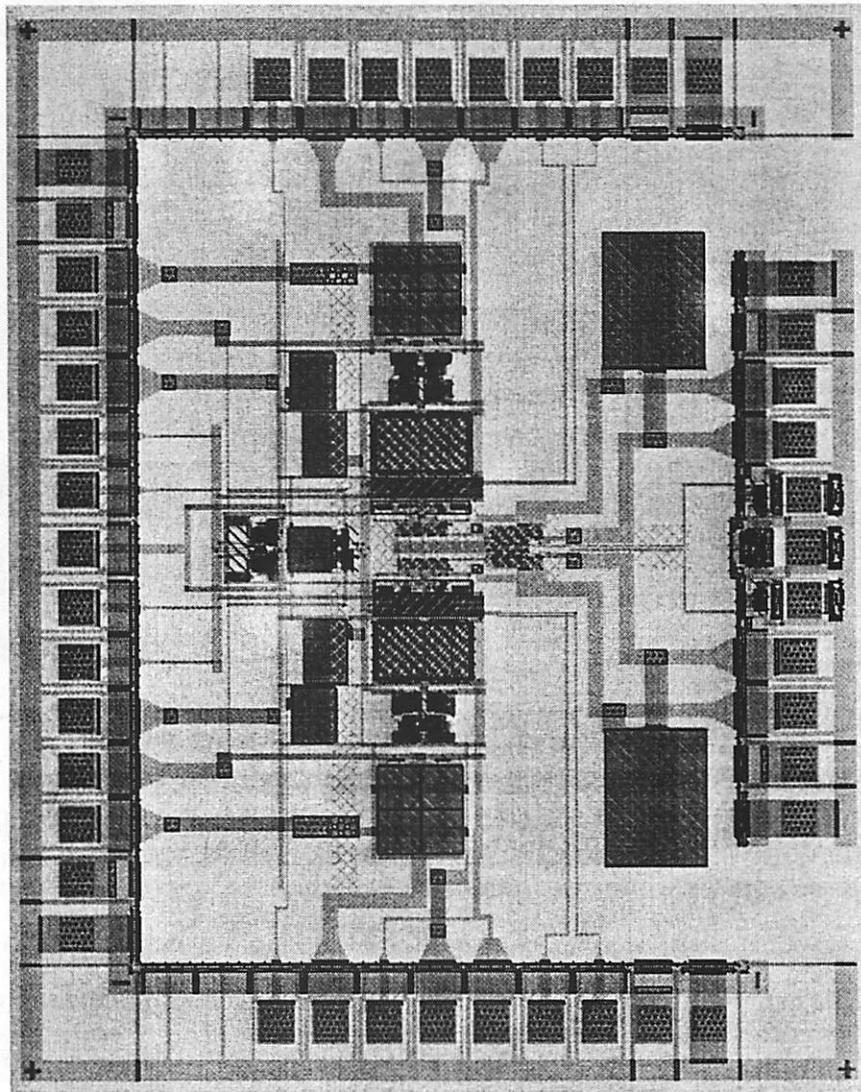


Figure 6.31: Sampling demodulator testchip (minisporf2)

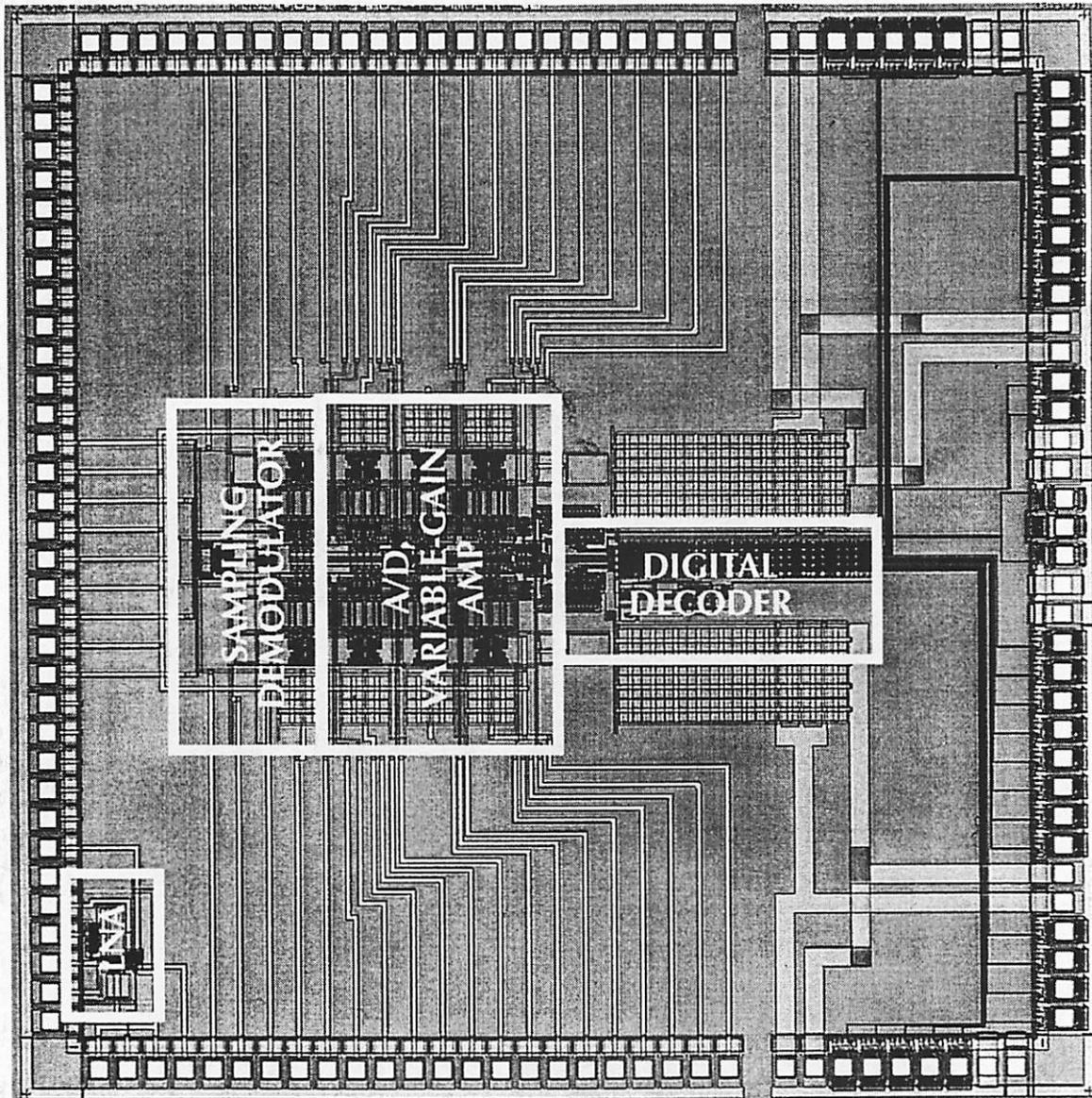


Figure 6.32: Full analog front-end receiver testchip (sportf)

the testboard. The die photo is shown in figure 6.32. Excluding the LNA, the top half and the bottom half of the core are symmetric copies of one another; the LNA (being single-ended for power considerations) is placed as far away from the high-speed switching path as possible. Again, the die is symmetric with respect to the in-phase and quadrature signals; the centerline of symmetry divides the I channel and Q channels of the receiver. The signal flows from left to right across figure 6.32, and the pad ring to the right of the chip has been broken to separate analog and digital supplies.

Parameter	Value	Parameter	Value
LNA		Baseband VGA	
Gain	11.5 dB @ 1.05 GHz	Maximum Sampling Rate	90 MHz
Noise figure	7.2 dB	Digitally Controllable Gain (4 stages)	12 dB, 3 dB/stage
Power Consumption	36 mW @ 3.3 V	Per Stage Sample & Hold Power Consumption	8.3 mW
IP ₃	-10 dBm		
Sampling Demodulator		A/D Converter	
I/Q Phase Balance	4°	Peak SNDR	25.6 dB
Track Bandwidth	1.75 GHz	4 bit A/D INL	0.42 LSB
Noise figure (computed)	18.1 dB	4 bit A/D DNL	0.36 LSB
IP ₃	22 dBm	Maximum A/D Conversion Rate	150 MHz
		A/D Power Consumption (for two converters on chip)	4.4 mW
		Total Measured Analog Power Consumption	106.8 mW

Figure 6.33: Measured results for the analog front-end receiver testchip.

LNA results are from chapter 5; sampling demod track BW and IP₃ are from the standalone sampling demodulator testchip results (minisporf2). Maximum A/D conversion rate and A/D power are from the standalone A/D testchip results (minisporf).

The RF chip-on-board high-speed test board described in the appendix to chapter 5 was used to test the integrated analog receiver chip. The final measured results for the complete receiver chain are listed in figure 6.33, including a recap of the LNA results from chapter 5. Beyond the LNA, the key test results from this die were the operating frequency and linearity of the entire baseband chain. The maximum clock frequency of the overall chain proved to be significantly lower than the 150 MHz top speed of the flash A/D converter, owing to the kickback problem described previously in section 6.3. The maximum sampling rate at which an input sinusoid can be recovered without error at the output is 90 MHz; beyond this, loss of settling time due to the signal dependent kickback noise causes the sample-and-hold amplifiers to fail. Lastly, figure 6.34 shows the FFT of the baseband sampling-demodulated output

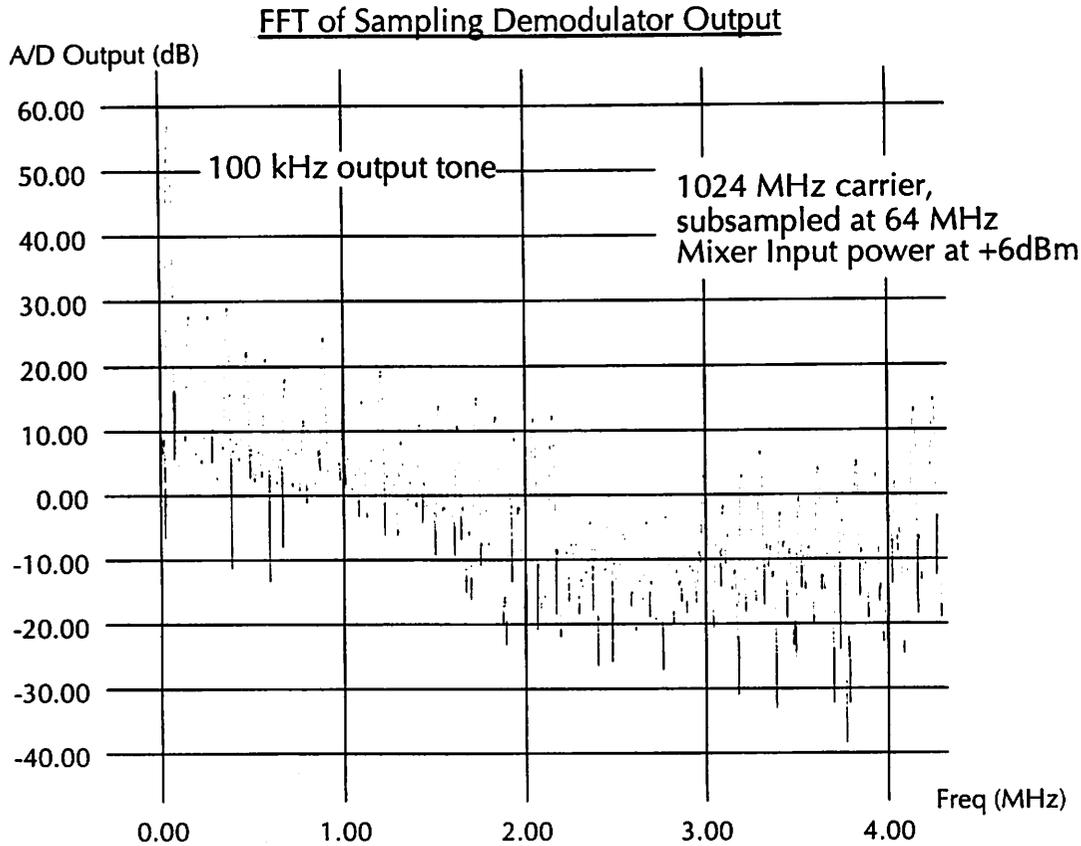


Figure 6.34: FFT of baseband sampling demodulator output. Input signal is a +6 dBm, 1.024 GHz carrier, SSB modulated by a 100 kHz sinusoid. Output is sampled at 64 MHz; distortion harmonics are consistent with that of a 4-bit A/D converter.

of an RF signal; in particular, the input was modulated 1.024 GHz carrier modulated in single-sideband by a 100 kHz sinusoidal tone, subsampled at 64 MHz. By integrating the distortion harmonics the SNDR of the entire receiver chain has been calculated to be 25.6 dB (or slightly over four bits).

Lastly, as a full demonstration of the baseband transmit modulator and analog front-end, the modulator's output was used to drive a semicustom, broadband single-sideband RF modulator [Yee96]; the SSB modulator utilizes the down-up method described in chapter 4. No off-the-shelf bandpass filter possessed sufficient bandwidth to pass the full transmit spectrum; a cascade of lowpass-highpass filters was used to implement crude bandpass filters. Due to this, the carrier frequencies were limited to 700 MHz; above this, the noise admitted into the sys-

tem due to the imprecise nature of the filters made testing above this frequency unreliable. To accommodate the digital acquisition system at the output of the A/D converters, the system spread-spectrum chipping rate was set to 20 MHz. Figure 6.35 shows the received eye diagram for a transmitted PN sequence, which is the pilot tone by itself. The eye is clearly open, with quite a bit of both noise immunity and jitter margin, and the 15 levels of the A/D converter are quite distinct at the sampling instants.¹ Furthermore, the off-time sampling point is clearly saturating the A/D converter, in comparison with the true 30% excess bandwidth eye diagram of figure 3.2. To test data integrity, a sliding correlator test was applied to the received PN sequence; a single strong peak should be readily apparent, if no data corruption was incurred. Indeed, a single peak appears at a value nearly equal to $7 \cdot 32768$ (maximum A/D output times the full PN sequence length). Thus, at least for this test, the bit-error rate is significantly lower than 1 part in $3 \cdot 10^5$ bits.

1. The sampling rate here was 20 MHz; the arcs in between the sample instants was reconstructed after digital acquisition using an ideal DAC function.

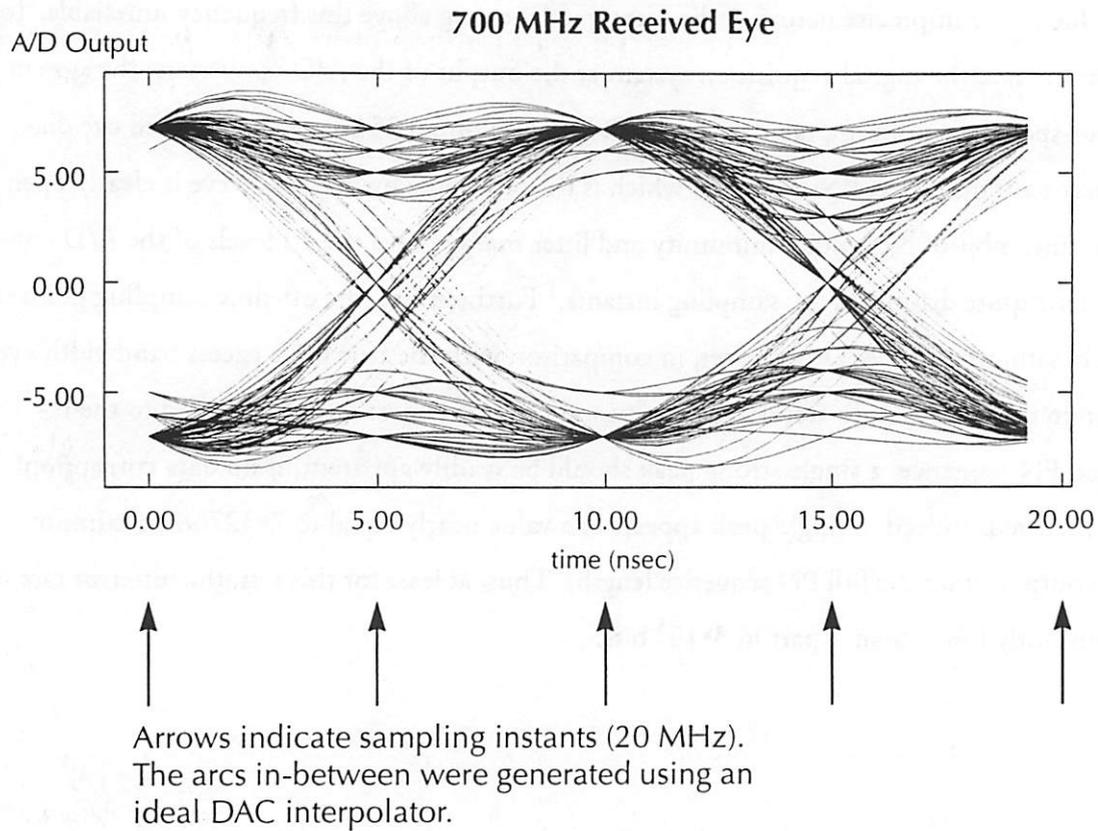


Figure 6.35: 700 MHz received eye diagram.

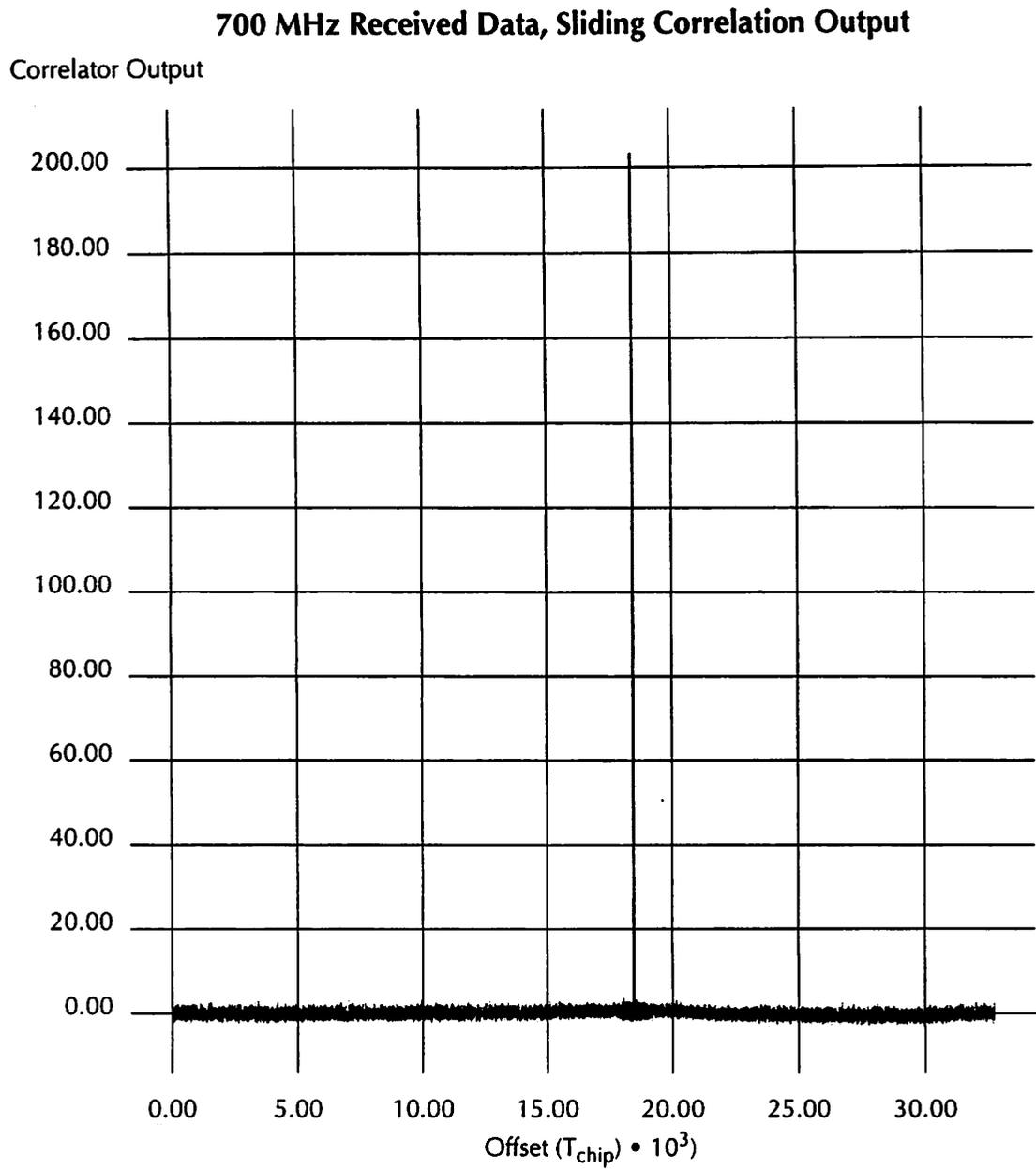


Figure 6.36: 700 MHz received data, sliding correlation test

6.6. Appendix: Electrical Interface Specifications

This appendix contains the pad pinout information for the three analog testchips described in this chapter. Since these die are intended for use in a chip-on-board type package, only the pad-level descriptions are provided. For the most part, high pin count packages have lead inductances that are too large to be tolerated (especially by the full analog receiver testchip). If, for some reason, packaging these die is desirable, the following is a recommended list:

Full analog receiver testchip (sporff): 132-pin CQFP

Sampling demodulator testchip (minisporff2): 44-pin PLCC

Standalone A/D testchip (minisporff): 68-pin PLCC

The full analog receiver will be limited to 600 MHz by the CQFP package, so the package is only truly useful for “low” frequency testing. Obviously, if the die are refabricated in a better CMOS technology, their input capacitances will be commensurately scaled, and can tolerate more series input inductance.

Pad	Function	Description
1	analog GND	
2	analog GND	
3	analog GND	
4	analog GND	
5	analog Vdd	Analog supply, nominal 5V
6	V_{icm}	OTA Common-mode Input Bias Voltage, nominal 2.5V Due to headroom concerns, only +/- 0.2V is tolerable.
7	analog GND	
8	analog Vdd	Analog supply, nominal 5V
9	i_agcctrl1	in-phase path, VGA stage one control (5V digital input)
10	analog Vdd	Analog supply, nominal 5V
11	analog GND	
12	analog Vdd	Analog supply, nominal 5V
13	i_agcctrl2	in-phase path, VGA stage two control (5V digital input)
14	analog Vdd	Analog supply, nominal 5V
15	analog GND	
16	i_agcctrl3	in-phase path, VGA stage three control (5V digital input)
17	analog Vdd	Analog supply, nominal 5V
18	analog GND	
19	i_agcctrl4	in-phase path, VGA stage four control (5V digital input)
20	refn	External voltage bias input, negative end of resistor ladder ($V_{icm} - 0.25V$)
21	refp	External voltage bias input, positive end of resistor ladder ($V_{icm} + 0.25V$)
22	adcbias	External current input, 20 uA source to ground
23	analog Vdd	Analog supply, nominal 5V
24	digital GND	
25	digital Vdd	Digital supply, nominal 5V
26	loffH3	Digital output, in-phase sample, bit #3, time position 3
27	lonH3	Digital output, in-phase sample, bit #3, time position 1
28	lonL3	Digital output, in-phase sample, bit #3, time position 2
29	QonL3	Digital output, quadrature sample, bit #3, time position 2
30	loffL3	Digital output, in-phase sample, bit #3, time position 4
31	digpaddingvdd	Isolated Vdd for digital pad ring buffers, nominal 5V
32	digpaddinggnd	
33	QoffL3	Digital output, quadrature sample, bit #3, time position 4
34	QonH3	Digital output, quadrature sample, bit #3, time position 1
35	QoffH3	Digital output, quadrature sample, bit #3, time position 3
36	loffH2	Digital output, in-phase sample, bit #2, time position 3
37	lonH2	Digital output, in-phase sample, bit #2, time position 1

Figure 6.37: Analog front-end receiver pinout (sporf). Pad numbers are with respect to the upper-left hand corner of the die photo (figure 6.32).

Pad	Function	Description
38	IonL2	Digital output, in-phase sample, bit #2, time position 2
39	QonL2	Digital output, quadrature sample, bit #2, time position 2
40	IoffL2	Digital output, in-phase sample, bit #2, time position 4
41	digital Vdd	Digital supply, nominal 5V
42	digital GND	
43	QoffL2	Digital output, quadrature sample, bit #2, time position 4
44	QonH2	Digital output, quadrature sample, bit #2, time position 1
45	digpadvdd	Isolated Vdd for digital pad ring buffers, nominal 5V
46	digpadgnd	
47	clkIn-	Differential clock input, 128 MHz sinusoidal, common mode voltage between 1V and 3V.
48	clkbias	Current bias for clock input buffer, 1 mA source to ground
49	clkIn+	Differential clock input, 128 MHz sinusoidal, common mode voltage between 1V and 3V.
50	digpadvdd	Isolated Vdd for digital pad ring buffers, nominal 5V
51	digpadgnd	
52	QoffH2	Digital output, quadrature sample, bit #2, time position 3
53	IoffH1	Digital output, in-phase sample, bit #1, time position 3
54	IonH1	Digital output, in-phase sample, bit #1, time position 1
55	IonL1	Digital output, in-phase sample, bit #1, time position 2
56	QonL1	Digital output, quadrature sample, bit #1, time position 2
57	IoffL1	Digital output, in-phase sample, bit #1, time position 4
58	QoffL1	Digital output, quadrature sample, bit #1, time position 4
59	QonH1	Digital output, quadrature sample, bit #1, time position 1
60	digital GND	
61	digital Vdd	Digital supply, nominal 5V
62	QoffH2	Digital output, quadrature sample, bit #1, time position 3
63	IoffH0	Digital output, in-phase sample, bit #0, time position 3
64	IonH0	Digital output, in-phase sample, bit #0, time position 1
65	IonL0	Digital output, in-phase sample, bit #0, time position 2
66	digpadgnd	
67	digpadvdd	Isolated Vdd for digital pad ring buffers, nominal 5V
68	QonL0	Digital output, quadrature sample, bit #0, time position 2
69	IoffL0	Digital output, in-phase sample, bit #0, time position 4
70	QoffL0	Digital output, quadrature sample, bit #0, time position 4
71	QonH0	Digital output, quadrature sample, bit #0, time position 1
72	QoffH0	Digital output, quadrature sample, bit #0, time position 3
73	digital GND	
74	digital Vdd	Digital supply, nominal 5V

Figure 6.37: Analog front-end receiver pinout (sporfi). Pad numbers are with respect to the upper-left hand corner of the die photo (figure 6.32).

Pad	Function	Description
75	analog Vdd	Analog supply, nominal 5V
76	ADC current bias	External current input, 20 uA source to ground
77	refp	External voltage bias input, positive end of resistor ladder ($V_{icm} + 0.25V$)
78	refn	External voltage bias input, negative end of resistor ladder ($V_{icm} - 0.25V$)
79	q_agcctrl4	quadrature path, VGA stage four control (5V digital input)
80	analog GND	
81	analog Vdd	Analog supply, nominal 5V
82	q_agcctrl3	quadrature path, VGA stage three control (5V digital input)
83	analog GND	
84	analog Vdd	Analog supply, nominal 5V
85	q_agcctrl2	quadrature path, VGA stage two control (5V digital input)
86	analog Vdd	Analog supply, nominal 5V
87	analog GND	
88	analog Vdd	Analog supply, nominal 5V
89	q_agc_ctrl1	quadrature path, VGA stage one control (5V digital input)
90	analog Vdd	Analog supply, nominal 5V
91	analog GND	
92	V_{icm}	OTA Common-mode Input Bias Voltage, nominal 2.5V Due to headroom concerns, only +/- 0.2V is tolerable.
93	analog Vdd	Analog supply, nominal 5V
94	lnagnd	Isolated RF ground for the LNA
95	lnagnd	Isolated RF ground for the LNA
96	lnarfout	RF LNA output; series inductance < 22 nH for 1 GHz operation. Requires shunt RF choke for operation (DC bias through choke to V_{icm}).
97	lnagnd	Isolated RF ground for the LNA
98	lnagnd	Isolated RF ground for the LNA
99	lnagnd	Isolated RF ground for the LNA
100	lnavdd	Isolated supply for the LNA, 3.3V - 5V
101	ampnbias	External current source bias for the LNA, 0.1 mA sink from Vdd
102	lnagnd	Isolated RF ground for the LNA
103	lnavin	RF LNA input; series inductance < 6 nH for 1 GHz operation
104	lnagnd	Isolated RF ground for the LNA
105	analog GND	
106	analog GND	
107	analog GND	
108	analog GND	
109	analog GND	
110	bufout-	Negative output of pre-demodulator buffer; also allows direct input to the sampling demodulator itself. Requires DC common-mode choke bias to V_{icm} .

Figure 6.37: Analog front-end receiver pinout (sporf). Pad numbers are with respect to the upper-left hand corner of the die photo (figure 6.32).

Pad	Function	Description
111	analog GND	
112	bufin-	Negative input to the pre-demodulator buffer; common-mode input voltage at V_{icm} .
113	analog GND	
114	otabias	External current bias for the OTA chain; 0.1 mA current sink from Vdd
115	analog GND	
116	bufin+	Positive input to the pre-demodulator buffer; common-mode input voltage at V_{icm} .
117	analog GND	
118	bufout+	Positive output of pre-demodulator buffer; also allows direct input to the sampling demodulator itself. Requires DC common-mode choke bias to V_{icm} .
119	analog GND	
120	analog GND	
121	analog GND	
122	analog GND	
123	analog GND	
124	analog GND	
125	analog GND	
126	analog GND	
127	analog GND	
128	analog GND	
129	analog GND	
130	analog GND	

Figure 6.37: Analog front-end receiver pinout (sporf). Pad numbers are with respect to the upper-left hand corner of the die photo (figure 6.32).

Pad	Function	Description
1	V _{icm}	OTA Common-mode Input Bias Voltage, nominal 2.5V Due to headroom concerns, only +/- 0.2V is tolerable.
2	analog Vdd	Analog supply, nominal 5V
3	i_agcctrl1	in-phase path, VGA control (5V digital input)
4	analog GND	
5	analog Vdd	Analog supply, nominal 5V
6	iout-	in-phase demodulator, negative output. Analog common mode at Vicm
7	iout+	in-phase demodulator, positive output. Analog common mode at Vicm
8	anapadgnd	
9	anapadvdd	Isolated supply for the analog pad ring, nominal 5V
10	digpadvdd	Isolated supply for the digital pad ring, nominal 5V
11	digpadgnd	
12	digital GND	
13	digital VDD	Digital supply, nominal 5V
14	clkin-	Differential clock input, 128 MHz sinusoidal, common mode voltage between 1V and 3V.
15	clkbias	Current bias for clock input buffer, 1 mA source to ground
16	clkin+	Differential clock input, 128 MHz sinusoidal, common mode voltage between 1V and 3V.
17	digital VDD	Digital supply, nominal 5V
18	digital GND	
19	digpadgnd	
20	digpadvdd	Isolated supply for the digital pad ring, nominal 5V
21	anapadvdd	Isolated supply for the analog pad ring, nominal 5V
22	anapadgnd	
23	iout+	quadrature demodulator, positive output. Analog common mode at Vicm
24	iout-	quadrature demodulator, negative output. Analog common mode at Vicm
25	analog Vdd	Analog supply, nominal 5V
26	analog GND	
27	q_agc_ctrl1	quadrature path, VGA control (5V digital input)
28	analog Vdd	Analog supply, nominal 5V
29	V _{icm}	OTA Common-mode Input Bias Voltage, nominal 2.5V Due to headroom concerns, only +/- 0.2V is tolerable.
30	anapadvdd	Isolated supply for the analog pad ring, nominal 5V
31	anapadgnd	
32	analog GND	
33	analog Vdd	Analog supply, nominal 5V
34	analog GND	
35	bufin+	Positive input to the pre-demodulator buffer; common-mode input voltage at V _{icm} .

Figure 6.38: Sampling demodulator testchip pinout (minisporf2). Pad numbers are with respect to the upper-left hand corner of the die photo (figure 6.31).

Pad	Function	Description
36	bufout-	Negative output of pre-demodulator buffer; also allows direct input to the sampling demodulator itself. Requires DC common-mode choke bias to V_{icm} .
37	otabias	External current bias for the OTA chain; 0.1 mA current sink from Vdd
38	bufout+	Positive output of pre-demodulator buffer; also allows direct input to the sampling demodulator itself. Requires DC common-mode choke bias to V_{icm} .
39	bufin-	Negative input to the pre-demodulator buffer; common-mode input voltage at V_{icm} .
40	analog GND	
41	analog Vdd	Analog supply, nominal 5V
42	analog GND	
43	anapadgnd	
44	anapadvdd	Isolated supply for the analog pad ring, nominal 5V

Figure 6.38: Sampling demodulator testchip pinout (minisporf2). Pad numbers are with respect to the upper-left hand corner of the die photo (figure 6.31).

Pad	Function	Description
1	V_{icm}	OTA Common-mode Input Bias Voltage, nominal 2.5V Due to headroom concerns, only +/- 0.2V is tolerable.
2	analog Vdd	Analog supply, nominal 5V
3	i_agcctrl1	in-phase path, VGA control (5V digital input)
4	analog GND	
5	analog Vdd	Analog supply, nominal 5V
6	i_signbitinput	MSB of the in-phase A/D (Externally applied for this testchip); digital 5V input
7	analog GND	
8	analog Vdd	Analog supply, nominal 5V
9	refn	External voltage bias input, negative end of resistor ladder ($V_{icm} - 0.25V$)
10	refp	External voltage bias input, positive end of resistor ladder ($V_{icm} + 0.25V$)
11	adcbias	External current input, 20 uA source to ground
12	anapadgnd	
13	digpadvdd	Isolated supply for the digital pad ring, nominal 5V
14	digpadgnd	
15	lout0	In-phase digital output, bit 0 (LSB)
16	lout1	In-phase digital output, bit 1
17	lout2	In-phase digital output, bit 2
18	lout3	In-phase digital output, bit 3 (MSB)
19	digital GND	
20	digital VDD	Digital supply, nominal 5V
21	clk-	Differential clock input, 128 MHz sinusoidal, common mode voltage between 1V and 3V.
22	clkbias	Current bias for clock input buffer, 1 mA source to ground
23	clk+	Differential clock input, 128 MHz sinusoidal, common mode voltage between 1V and 3V.
24	digital VDD	Digital supply, nominal 5V
25	digital GND	
26	Qout3	Quadrature digital output, bit 3 (MSB)
27	Qout2	Quadrature digital output, bit 2
28	Qout1	Quadrature digital output, bit 1
29	Qout0	Quadrature digital output, bit 0 (LSB)
30	digpadgnd	
31	digpadvdd	Isolated supply for the digital pad ring, nominal 5V
32	anapadgnd	
33	adcbias	External current input, 20 uA source to ground
34	refp	External voltage bias input, positive end of resistor ladder ($V_{icm} + 0.25V$)
35	refn	External voltage bias input, negative end of resistor ladder ($V_{icm} - 0.25V$)
36	analog Vdd	Analog supply, nominal 5V
37	analog GND	
38	q_signbitinput	MSB of the in-phase A/D (Externally applied for this testchip); digital 5V input
39	analog Vdd	Analog supply, nominal 5V
40	analog GND	

Pad	Function	Description
41	q_agc_ctrl1	quadrature path, VGA control (5V digital input)
42	analog Vdd	Analog supply, nominal 5V
43	V _{icm}	OTA Common-mode Input Bias Voltage, nominal 2.5V Due to headroom concerns, only +/- 0.2V is tolerable.
44	anapadvdd	Isolated supply for the analog pad ring, nominal 5V
45	anapadgnd	
46	analog GND	
47	analog Vdd	Analog supply, nominal 5V
48	analog GND	
49	bufin+	Positive input to the pre-demodulator buffer; common-mode input voltage at V _{icm} .
50	bufout-	Negative output of pre-demodulator buffer; also allows direct input to the sampling demodulator itself. Requires DC common-mode choke bias to V _{icm} .
51	otabias	External current bias for the OTA chain; 0.1 mA current sink from Vdd
52	bufout+	Positive output of pre-demodulator buffer; also allows direct input to the sampling demodulator itself. Requires DC common-mode choke bias to V _{icm} .
53	bufin-	Negative input to the pre-demodulator buffer; common-mode input voltage at V _{icm} .
54	analog GND	
55	analog Vdd	Analog supply, nominal 5V
56	analog GND	
57	anapadgnd	
58	anapadvdd	Isolated supply for the analog pad ring, nominal 5V

7 The Receiver: Baseband Spread-Spectrum DSP

Last, but certainly not least, the digital signal processing required to recover the original transmitted data needs to be addressed. From the output of the receiver's analog front-end, two 4-bit, 128 MHz interleaved streams are emitted: the only processing that has been done so far is to bring the signal to baseband, and A/D converting it into a digital stream. Even beyond sheer data recovery, the issues of timing recovery and carrier frequency recovery have not been mentioned yet, nor has the implementation of the benefits of a spread-spectrum receiver – channel estimation, RAKE reception, and adjacent cell detection. None of these issues are touched in the analog domain; instead, they are relegated to the baseband signal processor that will be discussed in this chapter. A more detailed description of the circuit implementation can be found in [Stone95]; the discussion below is intended to focus on the specific system and circuit design tradeoffs that were made.

As mentioned before, this partitioning of the receiver architecture – an “A/D front-end” coupled with a highly complex DSP – has resulted in a low-power solution for the analog aspects of the receiver. However, this may place the power burden in the digital domain: the 128 MHz programmable DSP's of today consume literally watts of power, and it might be surmised that a custom ASIC solution would be comparable. Fortunately, this is not true; from the classic equation for digital switching power, $P_{\text{digital}} = CV^2f$, reduction of the supply voltage V is the key to low power consumption, owing to the quadratic dependency. With an ASIC, algorithm-specific architectural techniques can be applied to preserve the operating frequency f , potentially with some expense in the switching capacitance C , and yield tremendous net gains in power efficiency [Chand92]. In particular, the total power consumption of the baseband receiver DSP is a mere 46.5 mW: far lower than if a programmable DSP was

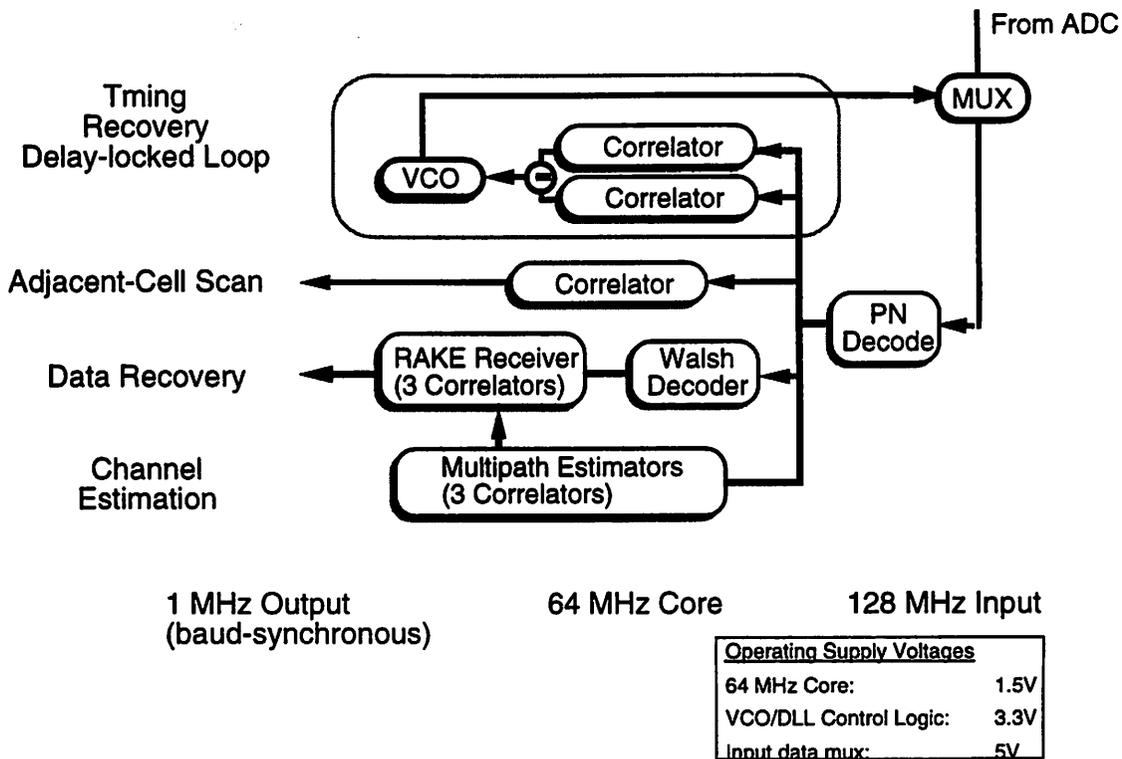


Figure 7.1: Receiver DSP Block Diagram

employed to provide the same functionality. In general, digital techniques are better suited than analog in creating circuits that can operate at high speed on a low supply voltage – a critical aspect of low-power design.

7.1. Receiver DSP Architecture

In figure 7.1,¹ the block diagram of the receiver DSP is shown. As expected, it implements all of the functionality delineated in chapter 3 for the spread-spectrum receiver: baseband timing recovery, adjacent-cell detection, channel estimation, and data recovery. Interestingly, this architecture is particularly amenable to low-power implementation: all of the functional blocks run in parallel, with little interaction between them. In particular, this parallel nature allows the input data (clocked from the analog receiver front-end at 128 MHz = $T_{\text{chip}}/2$) to be immediately decimated to 64 MHz (at T_{chip}), since the sampling instant for the delay-locked pilot tone tracking loop (chapter 3) is decided via an input multiplexer. The “on-time” samples

1. The implementation prototype for the DSP implements a subset of this block diagram; see section 7.6.

are fed to the data recovery, channel estimation, and adjacent-cell scan blocks; the “off-time” interleaved samples are fed to the delay-locked loop for tracking purposes, as will be discussed below.

However, one crucial block is missing: frequency recovery. Due to the need for minimal phase noise in the sampling demodulator architecture, frequency “pulling” of the baseband clock oscillator driving the demodulator is highly undesirable; it would render the entire demodulating front-end unusable. Digital methods of frequency compensation are thus called for. The most straightforward – an all-digital phase-locked loop – is also the costliest in power and complexity, given the 128 MHz input rate and the required ancillary hardware: ROM lookup tables, interpolators, and digital multipliers. Instead, the high frequency accuracy and stability of crystals are leveraged: the worst-case frequency offset for the crystal oscillators employed by this system is 25 kHz. Compared to the 64 MHz spread-spectrum chipping time, the phase is varying extremely slowly, and can be compensated in the recovery process. As each of the functional blocks is discussed below, the impact of incoherent recovery will be made apparent.

Lastly, it is also apparent from figure 7.1 that the key block required in the receiver DSP is the matched-filter correlator. If a low-power implementation can be developed – operating at 64 MHz on as low of a supply voltage as possible – a low-power spread-spectrum receiver is thus produced. As the viability of the entire “A-to-D” front-end architecture hinges upon power efficiency in the DSP back-end, implementation of the correlator is the first concern.

7.2. The Matched-Filter Correlator

In essence, the correlator is simply an accumulator, with a PN and/or Walsh sequence modulator preceding it. Its role is to take a 64 MHz, 4-bit sign magnitude input, accumulate it into a 10-bit word, and then dump its sum and reset at the conclusion of the correlation cycle; effectively, the output is decimated back to 1 MHz from the spread 64 MHz rate. The 10-bit word length is sufficient for a length 64 correlation, corresponding to the processing gain of the system; longer correlations, such as those needed for the delay-locked loop, will further accumulate the output of these length-64 correlators. As described in chapter 3, 1024 sample

correlations are used by several of the receiver functions; the output data, CORROUT, can be processed and fed into a second 14-bit accumulator which would be updated 16 times (once every microsecond) to produce the final 1024-length correlation output.

It should be mentioned that the first design parameter that was decided upon was the supply voltage, given the extreme need for low power. The mobile terminal itself uses 3 supply voltages, at 5V, 3.3V, and 1.5V, and the lowest was utilized to minimize power consumption. In fact, the DSP uses all three supply voltages, as indicated in figure 7.1; higher supply rails than 1.5V are necessary in the control blocks to meet critical path requirements. However, for the correlators, 1.5V was chosen with power as the overriding constraint. Given the limitations of a standard HP 0.8 micron CMOS technology, an aggressively low supply rail, and a critical cycle time of 16 nsec, implementation of even a "simple 14-bit accumulator" is hardly trivial.

The basic correlator block can be seen in Figure 7.2; this is one of the two blocks that are needed to implement a full in-phase/quadrature correlator, and the in-phase and quadrature blocks are identical. The incoming 4-bit sign-magnitude datum is latched by a register. Depending on its sign bit, it will either then be latched from the first register into the POSACC

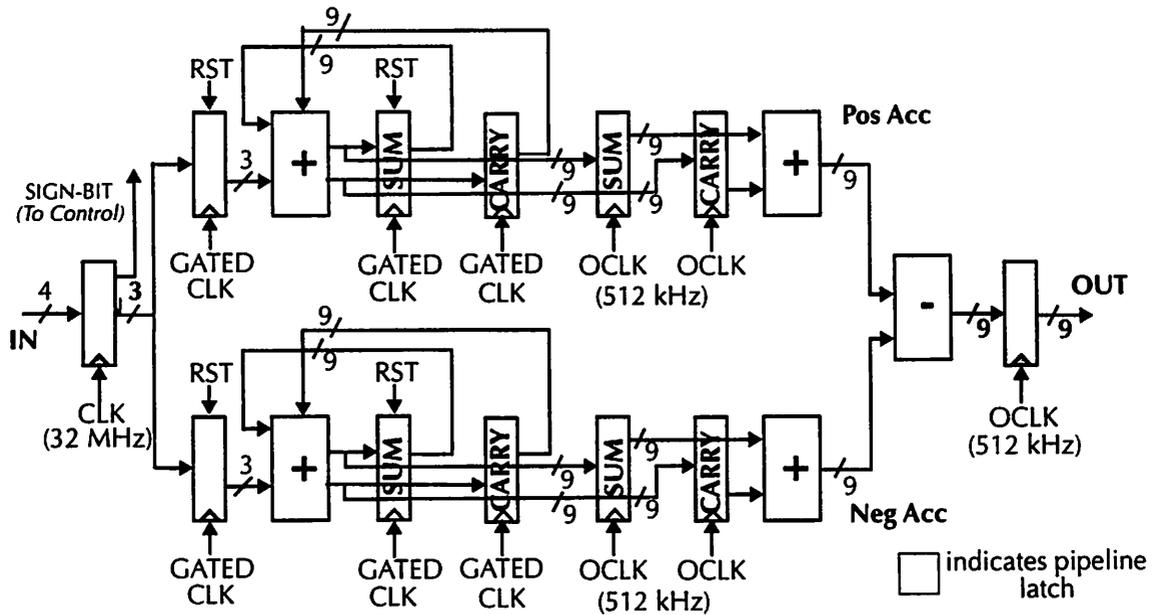


Figure 7.2: Basic 64 MHz, length-64 correlator datapath
(one-half of complex I/Q correlator; feeds second 14-bit,
1 MHz accumulator for length-1024 accumulate)

datapath or the NEGACC datapath. A simple subtractor cannot be implemented to meet the 16 nsec critical path; instead, the positive data values and the negative data values are accumulated separately, with the subtraction being performed at the 1 MHz decimated rate. As the polarity of the new input data is being determined, the previously accumulated POSACC /NEGACC SUM and CARRY vectors are latched as well and fed back into their respective adders, and the addition is performed between the previous sum and the new data. Since the addition is happening on only one of the POSACC or NEGACC datapaths, the unused datapath is not clocked, as to conserve switching activity and hence power.

Clearly, the critical path resides in the implementation of the actual accumulator: to meet the 16 nsec critical path requirement with a supply voltage of 1.5V, a highly pipelined design was necessary. In effect, pipelining was performed at the bit-level, by employing a carry-save architecture instead of a more traditional ripple-carry approach: the critical path was reduced to that of a single half-adder and a latch. To contrast, had this been implemented using a standard ripple implementation, a 3.3V supply voltage would have been required to meet the critical path: nearly a fourfold increase in power¹. In a carry-save architecture, a final addition is required between the carry-save path and the sum path; like the subtraction between the POSACC and NEGACC datapaths, this final addition can be performed at the 1 MHz output rate. It should be mentioned that both the low-speed 10-bit adders and subtractor at the output use a two's-complement ripple-carry architecture; even at 1.5V, 1000 nsec allows for plenty of margin, even with such simple implementations.

Lastly, the choice of signed magnitude versus two's-complement in the 64 MHz accumulator block has serious ramifications in terms of power consumption. From simulations done in [Chand94], the power consumption of a two's complement versus a sign-magnitude adder is strongly dependent on the power spectral density of the input data (figure 7.3). In particular, for spectrally white input data, like that of spread-spectrum data, a sign-magnitude implementation empirically consumes approximately 30% less power than a two's-complement adder. Intuitively, this is due to reduction in the number of bits toggled from input to input:

1. Due to the presence of extra registers required for the carry-save implementation, the power savings is less than the 4X predicted by the formula CV^2f .

with white data, the input data tends to go from positive-to-negative extremes on a sample-by-sample basis. In a two's-complement representation, all of the bits will toggle in such a transition; in sign-magnitude, only the sign bit toggles, resulting in a reduction in power.

Input Pattern (1024 cycles)	Two's Complement Power (3V)	Sign-Magnitude Power (3V)
constant (IN=7)	1.97 mW	2.25 mW
ramp (-7, -6, ..., 6, 7)	2.13 mW	2.43 mW
random	3.42 mW	2.51 mW
min→max→min (-7, 7, -7, 7, ...)	5.28 mW	2.46 mW

Figure 7.3: Comparison of the power dissipation of two's-complement vs. sign-magnitude representations in adder implementations.

7.3. Timing Recovery and Adjacent-Cell Detection

Arguably, the most difficult and important problem that faces the receiver DSP is that of baseband timing recovery: without accurate and consistent timing lock, all of the complex channel estimation and RAKE reception strategies are meaningless. As discussed in chapter 3, the PN pilot tone, with its strong autocorrelation properties, is used to synchronize the receiver with the base station. The implementation of the acquisition-and-tracking loop used to track the pilot tone in the receiver is the subject of this section.

7.3.1. Coarse Acquisition

First, when the mobile unit is turned on, detection of the base station and coarse timing acquisition must be performed in the receiver DSP. By "coarse", it is implied that the receiver's internal PN sequence is synchronized to within $\pm 1 T_{\text{chip}}$ of the base station's transmitted PN sequence. Effectively, coarse acquisition consists of stepping through each possible phase of the PN sequence, correlating at each phase, until a significant correlation output is detected. The full 32768-length correlation is not necessary at each PN phase; from simulations, 1024-length correlations are sufficient to minimize the probability of false lock, as well as minimize the time to acquisition (from chapter 3, the worst-case synchronization-lock time is

128 milliseconds). To achieve this lock time, four parallel correlators are used, searching four separate PN code phases simultaneously. A diagram of the coarse acquisition block is shown in figure 7.4; algorithmically, it can be described as the following:

1. Initialize PN generator and reset all correlators. Each correlator is time-offset by $1 T_{\text{chip}}$ with respect to one another, i.e., correlator₀ will correlate over phase_n, correlator₁ over phase_{n-1}, correlator₂ over phase_{n-2}, and correlator₃ over phase_{n-3}.
2. Correlate at this phase for 1024 cycles.
3. Latch output of each correlator.
4. Check each the magnitude of each output to see if it exceeds the lock threshold. If any do so, a positive lock has been achieved, and go to stage 6. Otherwise, continue to stage 5.
5. Reset the correlators to search for the next group of 4 PN phases. Stall the PN generator 4 four clock cycles. This will shift correlator₀ from phase_n to phase_{n-4}, correlator₁ from phase_{n-1} to phase_{n-5}, etc. Note that phase₋₁ is phase₃₂₇₆₇ (phases are modulo 32768). Go to stage 2 and repeat.
6. A positive lock has been found. Stall the PN generator a number of cycles equal to the index of the correlator that found lock; this forces the PN phase on correlator₀ to be the positive-lock phase for simplicity in control. Thus, if correlator₀ found lock, do not stall; if correlator₁ found lock, stall 1 cycle, etc.
7. PN phasing is set within $\pm 1 T_{\text{chip}}$ of the correct value. Switch to fine tracking mode, delay-locked loop.

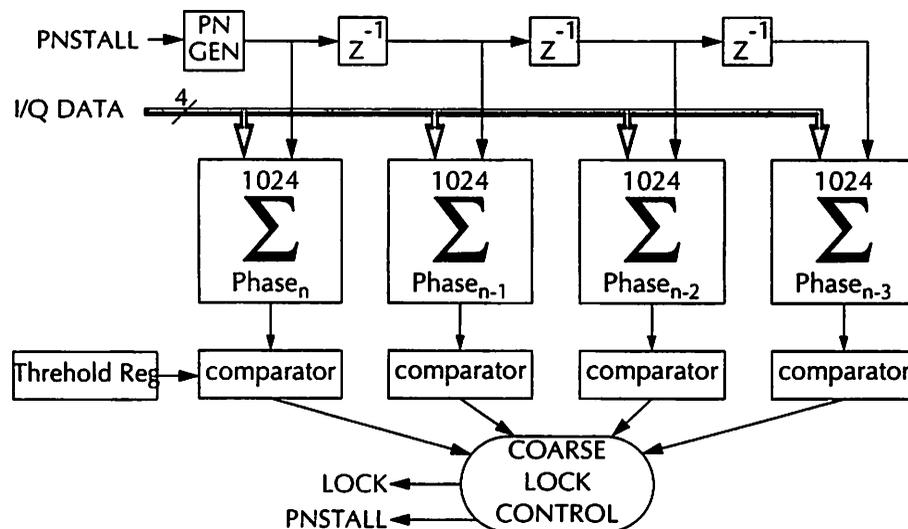


Figure 7.4: Block diagram from coarse lock acquisition.

Since silicon die area is at a premium, it would be a waste to expend the four correlators used in coarse-lock only for this purpose. In particular, these four correlators are well suited to perform the functions of channel estimation and adjacent-cell detection, which will be discussed later. However, some hardware multiplexing is necessary to support switching functionality; a more detailed diagram of the long correlators can be seen in figure 7.5. The signal `CMPTH_L` is fed from each of the long correlators to the lock circuitry, whereas the other signals, `IPQORI` and `Q` are used after lock for the remaining functions.

7.3.2. Fine Timing Acquisition and the Digital Phase-Locked Loop

After coarse lock has been achieved, the sampling point must then be moved as close to the maximum eye opening as possible, and this point must be tracked with respect to receiver movement, channel fading, and baseband oscillator offset between transmitter and receiver. Shown in figure 7.6a is the 1024-length correlator output as a function of time offset, for an ideal received transmit pulse (with zero offset being perfect synchronization). As expected, it is simply proportional to the 30% excess bandwidth raised-cosine transmit pulse. By locking

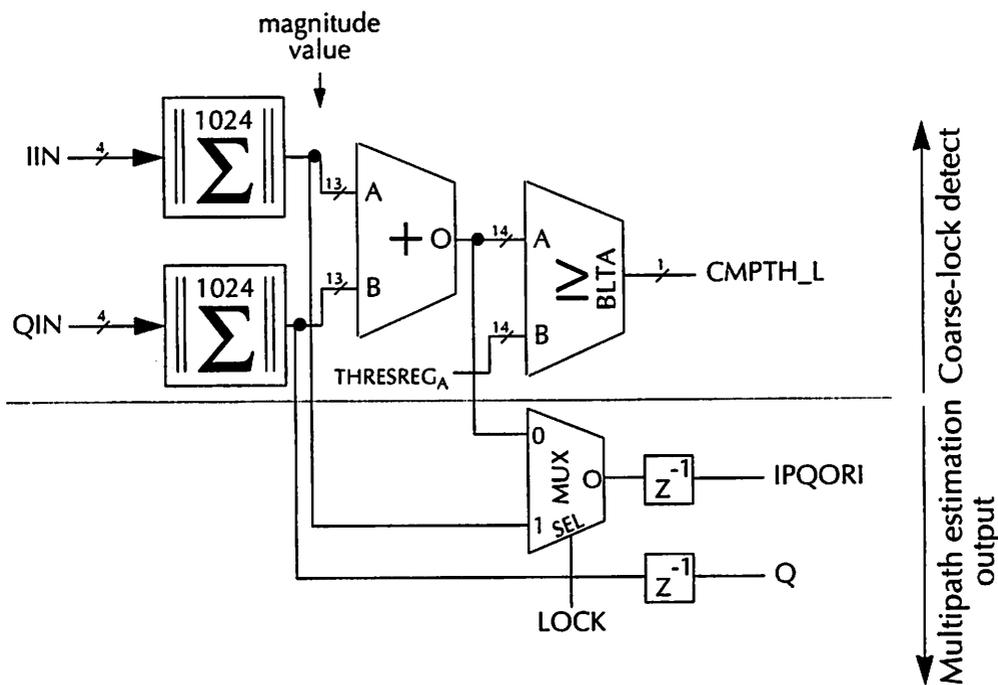


Figure 7.5: Block diagram of a multipath correlator.

two separate correlators E_{early} and E_{late} on either side of the peak (at $\pm T_{\text{chip}}/2$ relative to the center), an estimate of the peak location can be determined by the difference between the two; the symmetric nature of the pulse shape is critical. If the sampling point is early, then the difference in magnitude output between E_{early} and E_{late} will be negative, driving the sampling point forward. If the magnitude difference is positive, then the sampling point will be driven in the opposite direction (figure 7.6b).

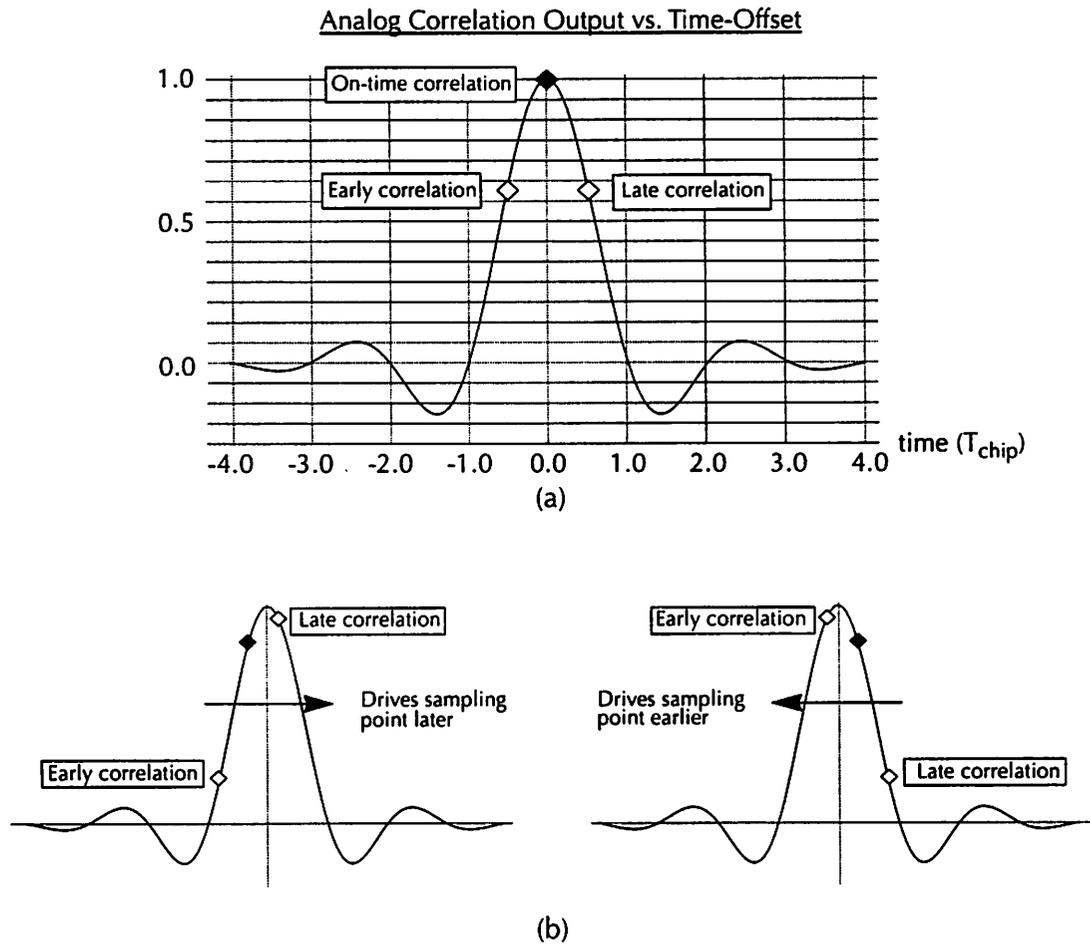


Figure 7.6: (a) Correlator outputs as a function of time offset ($t=0$ is synchronized).
 (b) Early/Late phase sensing in the delay-locked loop.

This forms the basis of the fine-tracking delay-locked loop; however, its implementation, especially with regard to the analog front-end, needs to be carefully considered. In the analog front-end, the 128 MHz baseband oscillator is used to drive the demodulator, subsampling the 1 GHz received signal. However, the same oscillator must also be used to drive the baseband DSP: “shifting” the sampling point in the analog domain for timing recovery becomes extremely difficult, since it will have dramatic effect on the RF demodulation characteristics. Fortunately, another solution is available. As discussed in chapter 5, recovery of both in-phase and quadrature signals requires two separate 128 MHz sampling switches, interleaved in time to effectively form a 256 MHz converter. As far as the analog front-end is concerned, designation of one stream as “in-phase” and the other as “quadrature” is purely notational; at its output, the two streams are effectively equivalent. The receiver DSP, on the other hand, can take advantage of this symmetry: with respect to the 256 MHz data stream, the only needed fact is that any two consecutive samples will be 90° out-of-phase with respect to one another, regardless of which demodulator actually generated which sample (figure 7.7).

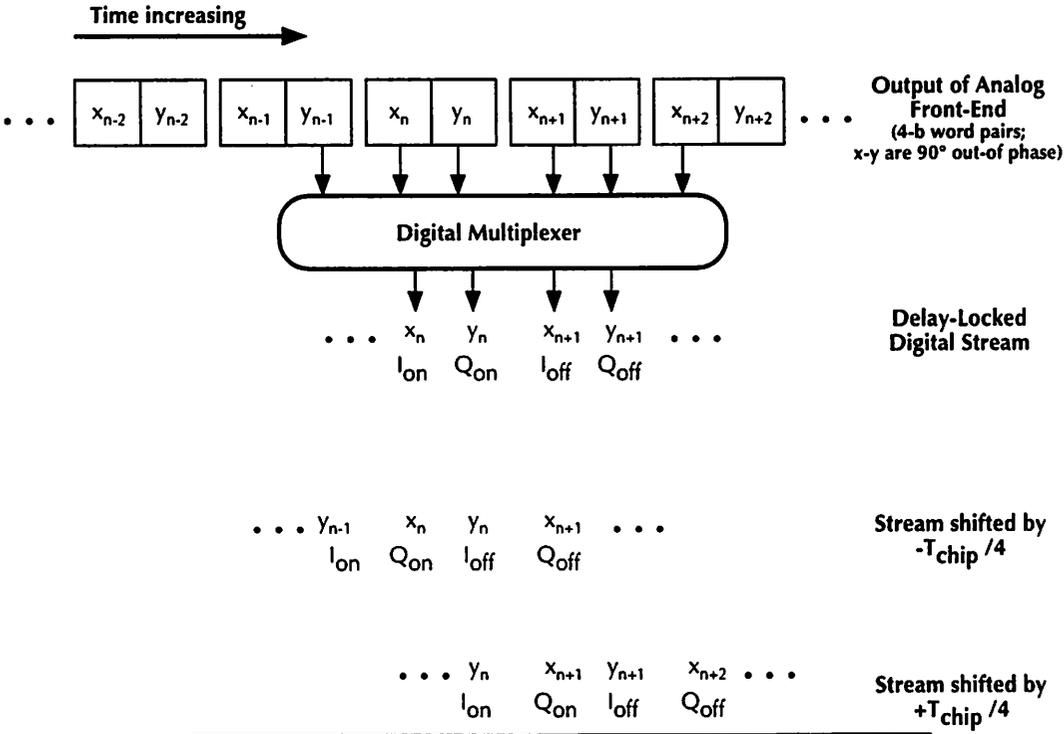


Figure 7.7: Mapping of analog front-end outputs into the four delay-locked digital data streams. The multiplexer “shifts” the samples appropriately.

Thus, for the purposes of timing recovery, the data stream is effectively oversampled at 4X with respect to the chipping rate; no real need exists to shift the sampling phase in the analog domain. Instead of driving an analog VCO, the delay-locked loop instead drives a 256 MHz multiplexer at the input of the receiver DSP; it is this multiplexer that decides which sample is “in-phase” or “quadrature”, and which sample is “on-time” or “off-time”, as shown in figure 7.7. The resulting granularity in sampling is 2 nsec ($T_{\text{chip}}/8$) in the worst-case, which is more than sufficient from system-level simulations.¹ One comment needs to be made on notation: for the discussion below, the multiplexer outputs are designated I_{on} , Q_{on} , I_{off} , and Q_{off} , being the I/Q on-time and I/Q off-time samples respectively. It is the off-time samples that are used to drive the DPLL, since these represent the early/late samples of figure 7.6.

The two basic building blocks of the DPLL need to be addressed: the phase detector, which consists of the $E_{\text{early}}/E_{\text{late}}$ correlators, and the input multiplexer, which serves the role of VCO. First, a functional block diagram of the phase detector can be seen in 7.8. One of the 1024-length correlators receives an $I_{\text{off}}/Q_{\text{off}}$ pair, while the other receives an $I_{\text{off}}/Q_{\text{off}}$ pair which has been delayed by T_{chip} (the late-correlator input is simply the early-correlator input delayed by 1 T_{chip}). Both correlators extract the required pilot tone magnitude information from the off-time data, over a period of 1024 samples (16 μsec), and then latched. This determines the phase detector’s update cycle, which is more than sufficient to track both the transmit-receiver oscillator offsets (25 kHz worst-case), as well as the slowly-varying indoor channel variations. From the recovered early and late pilot tone energies, a comparison is made between the absolute value of the difference between E_{early} and E_{late} ($|E_{\text{early}}-E_{\text{late}}|$) to a threshold register. If the comparator output CMPTH3 is greater than the value programmed in the threshold register, then the timing error is sufficient to warrant a shift in the sampling phase; the difference ($E_{\text{early}}-E_{\text{late}}$) is then used to determine the direction of the shift (figure 7.6). By setting an “acceptable” threshold on the difference through CMPTH3, needless ping-ponging in the sampling phase is prevented when E_{early} and E_{late} have similar values.

1. It is this $T_{\text{chip}}/8$ sampling error that gives rise to the timing jitter specification mentioned in chapter 3. Also, since the sampling-point adjustment is thus fixed, the DPLL tracking is akin to step adaptation in adaptive equalizers, with the error signal being the difference between the E_{early} and E_{late} correlators.

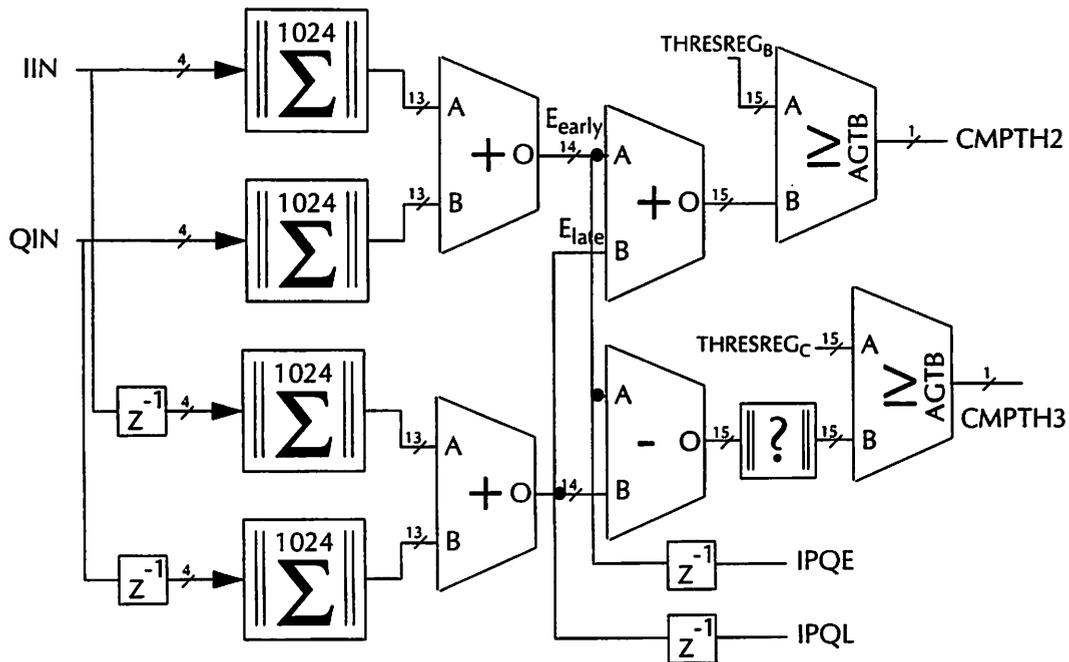


Figure 7.8: Block diagram of DLL phase detector. CMPH3 indicates whether the system clock's phase should be adjusted. CMPH2 indicates whether the received tone energy is large enough to stay in lock.

Second, the “VCO” for the DLL – the front-end I/Q data multiplexing – is shown in figure 7.9. By selecting which sample is in-phase or quadrature, as well as establishing which sample is designated as being “on-time”, it can shift the effective sampling point by the desired resolution of $T_{\text{chip}}/4$. From an architectural standpoint, there are four registers at the input to the DSP, designated I_{on} , Q_{on} , I_{off} , and Q_{off} . The two 4-bit samples from the analog front-end are first passed through a 2×2 digital switching network (implemented with true digital multiplexers), designating which sample is “I” or “Q”. The “on-time” or “off-time” designation is then made by controlling which register is clocked, latching the appropriate sample into the appropriate register. By multiplexing the samples with this gated clocking technique, shifting on-time sampling point by $\pm T_{\text{chip}}/4$ becomes a matter of extending or shortening the clock period for one cycle, instead of inserting or deleting samples from the in-

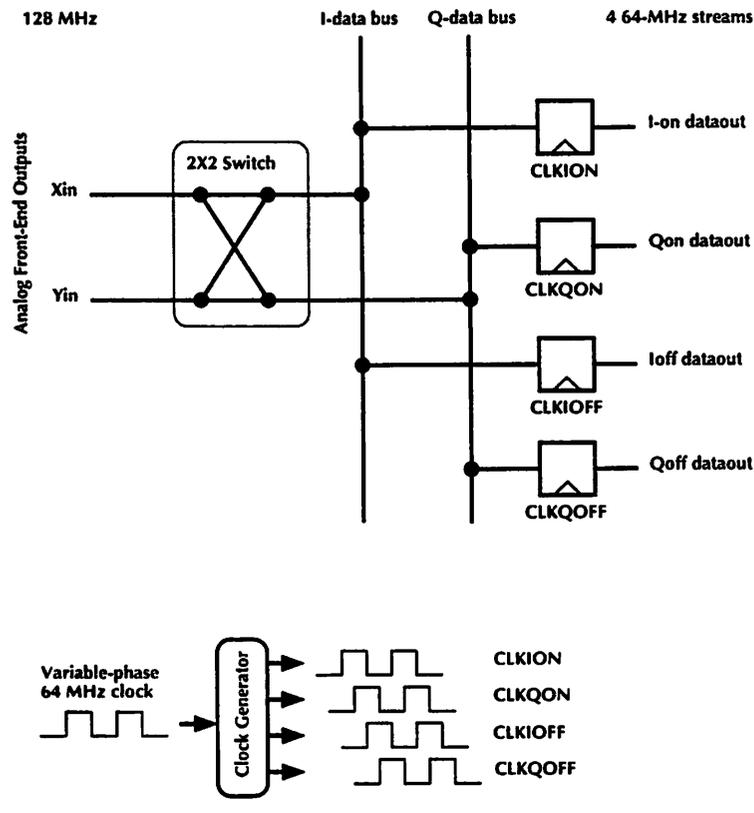


Figure 7.9: DLL VCO (data multiplexer) block diagram. All datapaths above are 4-bits wide.

put stream. This is desirable from both a complexity and power standpoint, since the control only needs to act on the clock (a 1-bit stream), instead of attempting to shift 4-bit sample words around from register to register.

As always, an architectural description is far simpler than its implementation; the $T_{\text{chip}}/4$ adjustment of 4 nsec implies that the clock generators and their control logic must run at 256 MHz, resulting in a key problem with critical path delays and necessitating a supply voltage of 5V. Likewise, when the clock phase is shortened by $T_{\text{chip}}/4$ to perform the timing adjustment, care must be taken that this shorter cycle time (12 nsec; figure 7.10) does not impact the correlators, which are designed to operate with a 16 nsec cycle time. Thus, shortening the clock phase by $T_{\text{chip}}/4$ in actuality is implemented as an extension in time by $3T_{\text{chip}}/4$, with the obvious loss of a single sample. Since the time-phase adjustment can happen only once every 1024 samples, a potential loss of 1 sample every 1024 is acceptable, re-

Variable-phase 64 MHz Clock

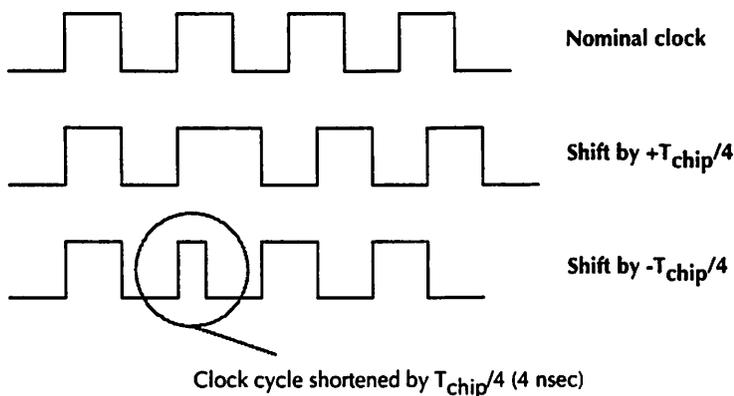
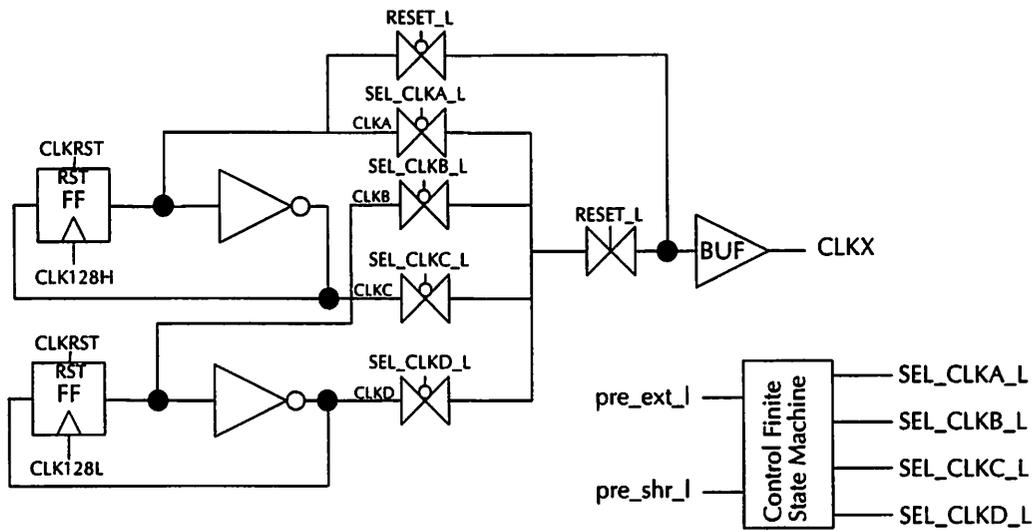


Figure 7.10: Variable-phase 64 MHz clock
(adjustable by $\pm T_{\text{chip}}/4$)

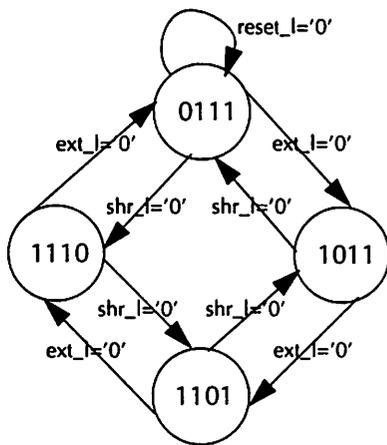
sulting in SNR degradations of 0.01 dB in the timing recovery block and 0.13 dB in the data recovery block. However, the PN and Walsh generators cannot lose a clock edge, otherwise the entire DLL scheme would be useless; separate clocks are necessary for the correlators and the PN/Walsh generators, with these generators required to meet the shorter 12 nsec cycle time.¹

The implementation of the variable-phase clock generator is shown in figure 7.11, along with its timing diagram. It consists of two stages, with the first stage generating a variable-phase 64 MHz clock (designated CLKX), and the second stage generating the four required clock phases (CLKION, CLKIOFF, CLKQON, CLKQOFF) relative to CLKX. It is important to note that when the clock phase is being reduced by $T_{\text{chip}}/4$, CLKX actually does shorten by that amount; the second stage's KILLPULSE gate control line guarantees that this shortened pulse is not fed to the datapaths on the chip or to the sample latching. As described above, the clock for the PN/Walsh generators cannot lose this pulse, and thus their clock is derived from CLKX directly, without any gating. The four DETFF blocks in the second stage are true dual

1. This requirement is not severe; the PN/Walsh generators are fairly straightforward, with only a few gate delays in the critical path; their implementation is the same as that described in chapter 4.



(a)

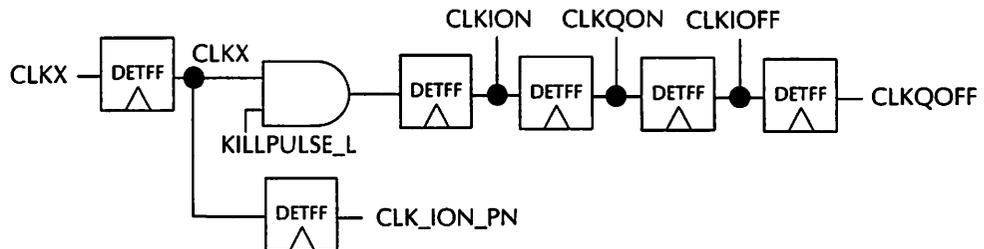


State transition diagram for the four SEL_CLK signals. Each state bit corresponds to a control bit for the clock synthesizer's pass-gates. Thus:
 bit3(msb)=SEL_CLKA_L
 bit2=SEL_CLKB_L
 bit1=SEL_CLKC_L
 bit0(lsb)=SEL_CLKD_L.

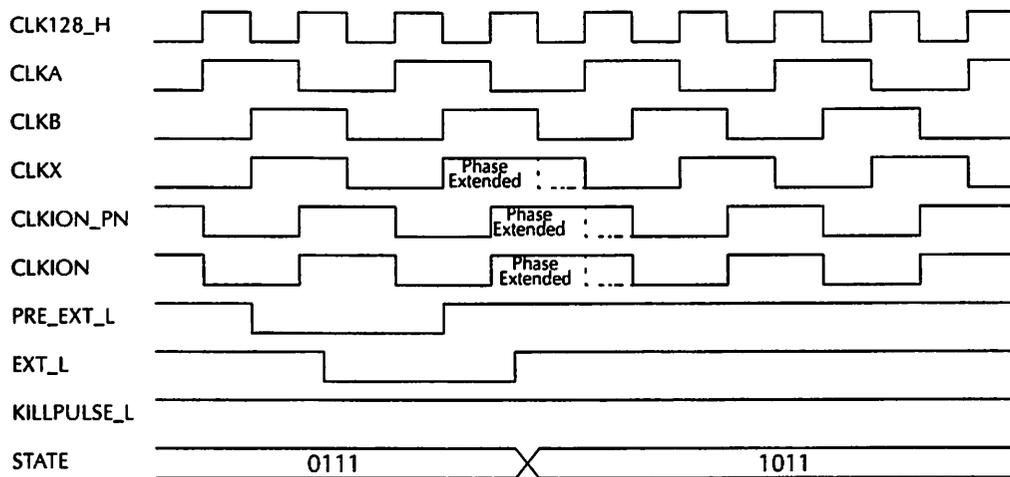
pre_ext_l, if asserted, extends the phase by $T_{chip}/4$.
 pre_shr_l, if asserted, shortens the phase by $T_{chip}/4$

ext_l = latch (pre_ext_l)
 shr_l = latch (pre_shr_l)
 reset_l = 0 forces return to 0111 state.

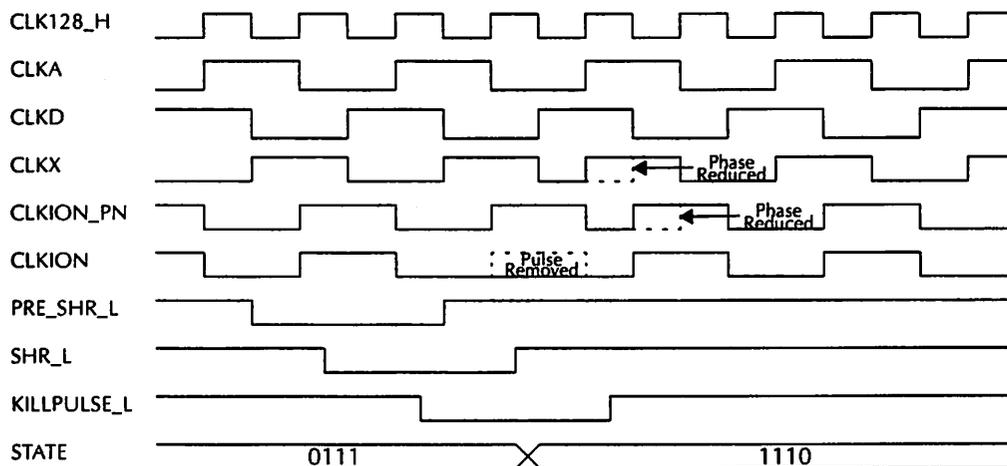
(b)



(c)



Timing diagram for extending the clock's phase.



Timing diagram for reducing the clock's phase.

(d)

Figure 7.11: (a) 64 MHz variable-phase clock synthesizer.
 (b) State machine to control clock phase.
 (c) 4-phase clock generator (derives the four required clocks from CLKX)
 (d) Timing diagrams

edge-triggered flip-flops [Afgh91] being clocked by the $T_{\text{chip}}/2 = 128$ MHz input clock; by latching the gated version of CLKX on both edges of the 128 MHz clock, a $T_{\text{chip}}/4$ delay line is thus formed, generating the four required clock phases.

Lastly, the implementation of the input data multiplexer is shown in figure 7.12; it differs significantly from figure 7.9 in two ways. First, the “on-time” and “off-time” designation registers have been moved before the I/Q designation mux; this is to avoid critical path problems with data changing at the mux input and a clock edge attempting to latch the output data 4 nsec later. Second, to simplify clocking in the remainder of the chip, all of the samples are then time-aligned to the I_{on} -clock phase, allowing the remainder of the chip to run at $1/T_{\text{chip}}$ (64 MHz) synchronous with CLKION. This resynchronization is performed by a sequence of three latches; the first latch R_1 synchronizes I_{on} with Q_{on} and I_{off} with Q_{off} , the second latch R_2 synchronizes all four together, and the third latch R_3 resynchronizes the data with the correlator’s clock (CLKION1_5X). One might question the need for the third latch; since no clock skew can be allowed between CLKION, CLKQON, CLKIOFF, CLKQOFF, CLKION itself cannot drive the huge capacitive load presented by the remainder of the chip, since the other three clocks will not see a similar loading. Thus, CLKION1_5X is derived from CLKION through a series of buffers; the third latch minimizes the impact of the skew between CLKION and CLKION1_5X.

7.3.3. Carrier-Offset Compensation

As discussed in chapter 5, for the crystals employed in this system, there is a worst-case frequency offset of 25 kHz between the transmit and receive RF local oscillators. This frequency shift results in a continuous, undesirable phase change of the received baseband signal, i.e., the baseband received constellation will be slowly “spinning,” in addition to the phase change seen by channel variations and receiver motion. With a 25 kHz carrier offset, a 64-long correlation spanning 1 μsec will see a shift of approximately 9° during the correlation length, which is acceptable for DQPSK data recovery purposes: the spin is strongly correlated from sample-to-sample; so long as its impact on the actual correlation output is small, it will tend to differentially cancel. However, for timing recovery, all of the blocks described above rely on a 1024-long correlation: 144° worst-case across its 16 μsec span, and intolerable for purposes

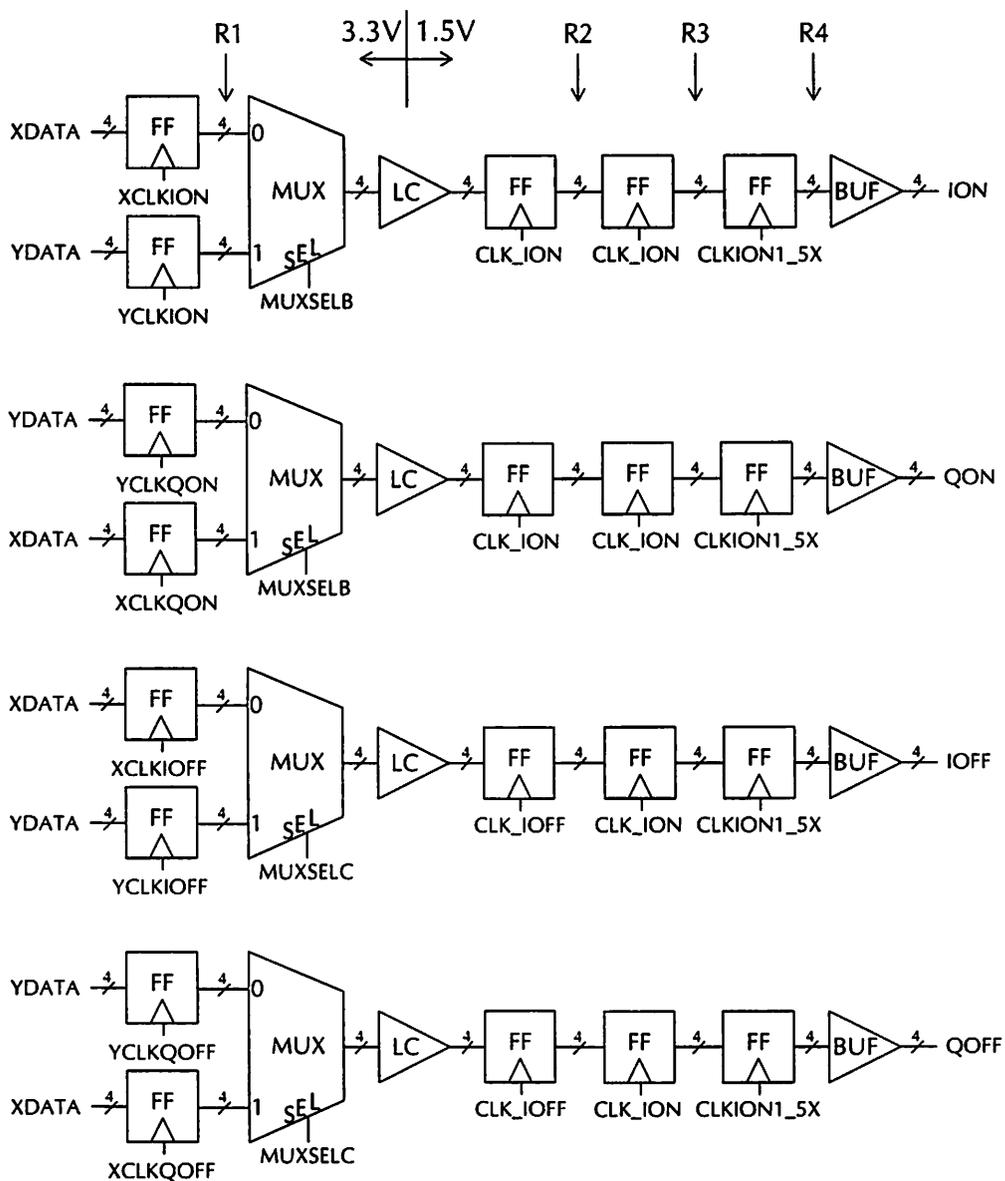


Figure 7.12: Block diagram of the input multiplexer. (LC= level converting buffer; XDATA, YDATA are the 4-bit words from the analog front-end).

of accurately estimating the channel peak. The impact of such spin can be seen in figure 7.13¹; instead of a ramp function, the peak becomes degraded, since the phase change results in a cancellation of the desired component of the correlation. Likewise, the phase is constantly rotating in the correlator output; instead of a flat phase output, the constellation is effectively “spinning”. Shorter correlations are not possible, due to subcorrelation noise effects and reduced processing gain – extremely detrimental to the critical function of timing recovery. A method of compensating this phase change must be developed.

For the magnitude estimates needed in both timing recovery and adjacent cell detection, the compensation algorithm is based on the fact that (after PN multiplication), the pilot-tone sequence component of the received data is constant, with additive high-frequency noise from other users, multipath, and neighboring cells. Effectively, the actual hardware correlator functions as a high-order lowpass filter. Thus, provided that the high-frequency noise component can be damped away, the following Cauchy-Schwartz (in)equality holds:

$$\left| \sum_{i=1}^{1024} a_i \right| \approx \sum_{j=1}^m \left| \sum_{i=1}^n a_i \right|$$

where $n \cdot m$ is equal to 1024, the correlation length. In essence, since the desired component is constant, the magnitude of the sum is equal to the magnitude of the partial sums without phase rotation. Under phase rotation, the partial (shorter) sums will have a far smaller error, and thus yield an accurate, long magnitude correlation output without being impacted by the carrier offset. Empirically, it was found that there was no difference between $n=64, m=16$ and $n=128, m=8$; either of these values was optimal. The case of $n=64, m=16$ was chosen for convenience, since $n=64$ corresponds to the system processing gain. A plot of the modified correlator output is shown in figure 7.14, and a block diagram of the magnitude correlator structure is given in figure 7.15.

1. Instead of simply showing the correlator output at the dump instant, the entire integration process is shown. Ideally, it should be a ramp; instead, there is significant degradation due to the phase rotation caused by the frequency offsets. It is important to note that for a length-64 correlation, there is no significant deviation; for a length-1024 correlation, the error becomes severe.

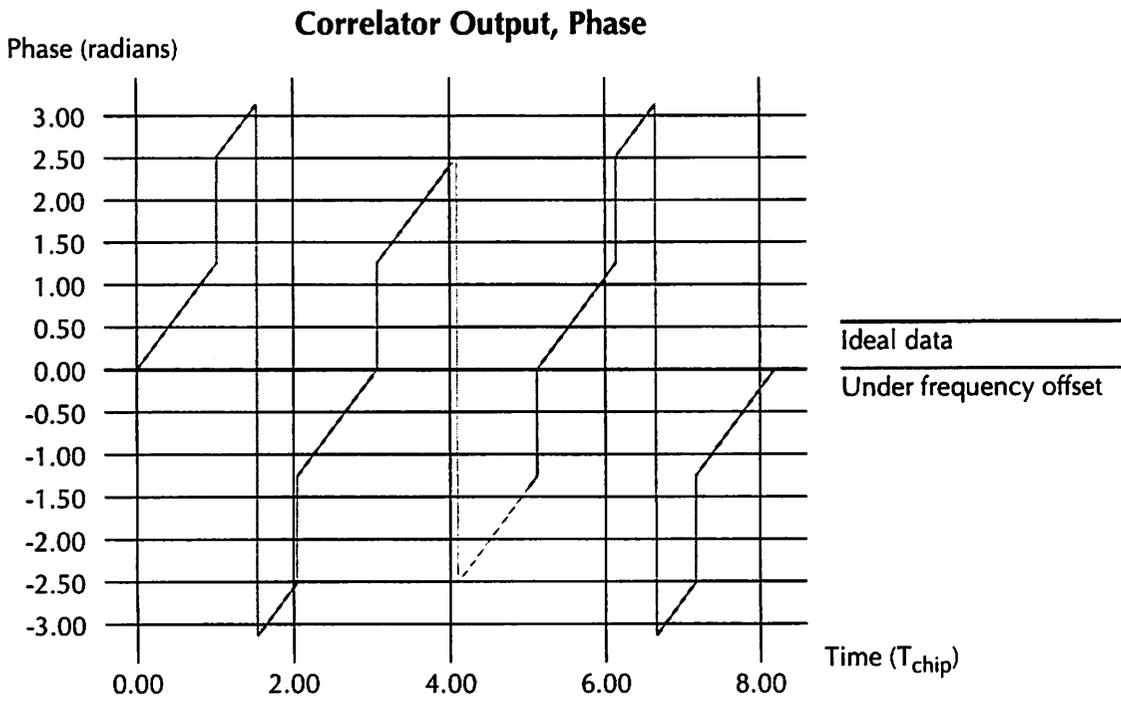
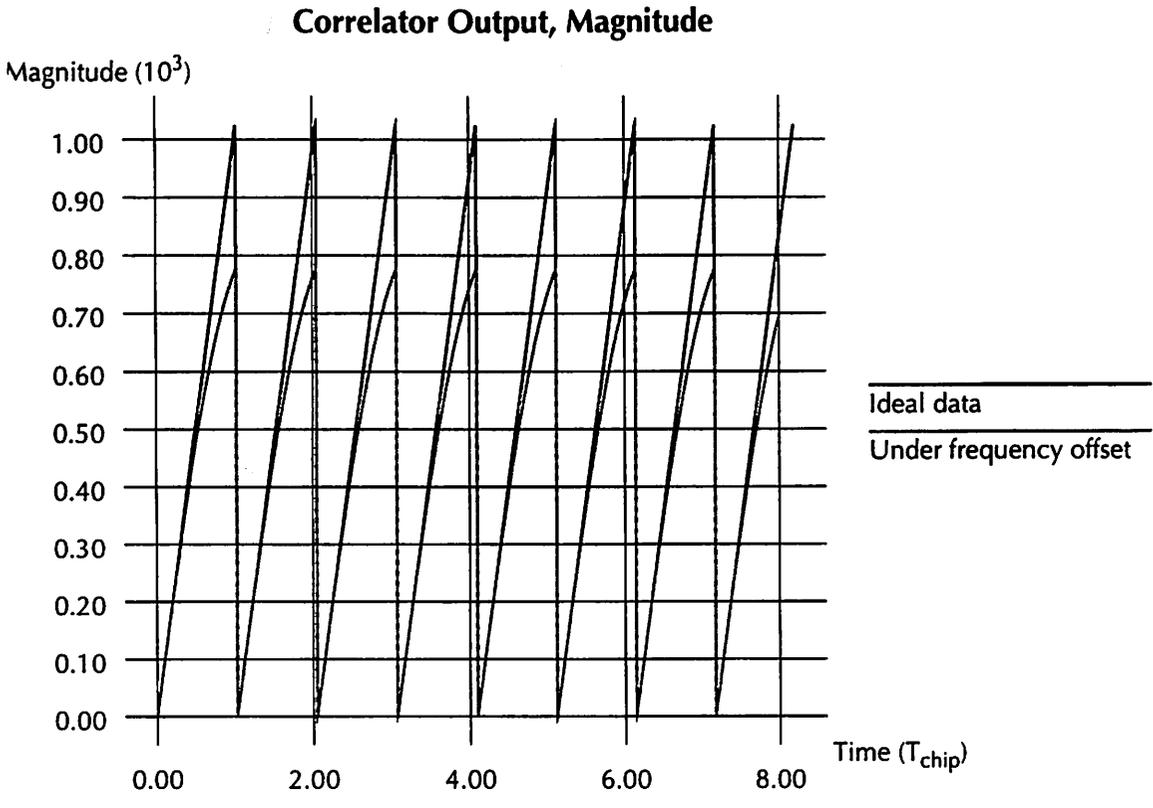


Figure 7.13: 1024-long correlation output, ideal and under 25 kHz carrier frequency offset.

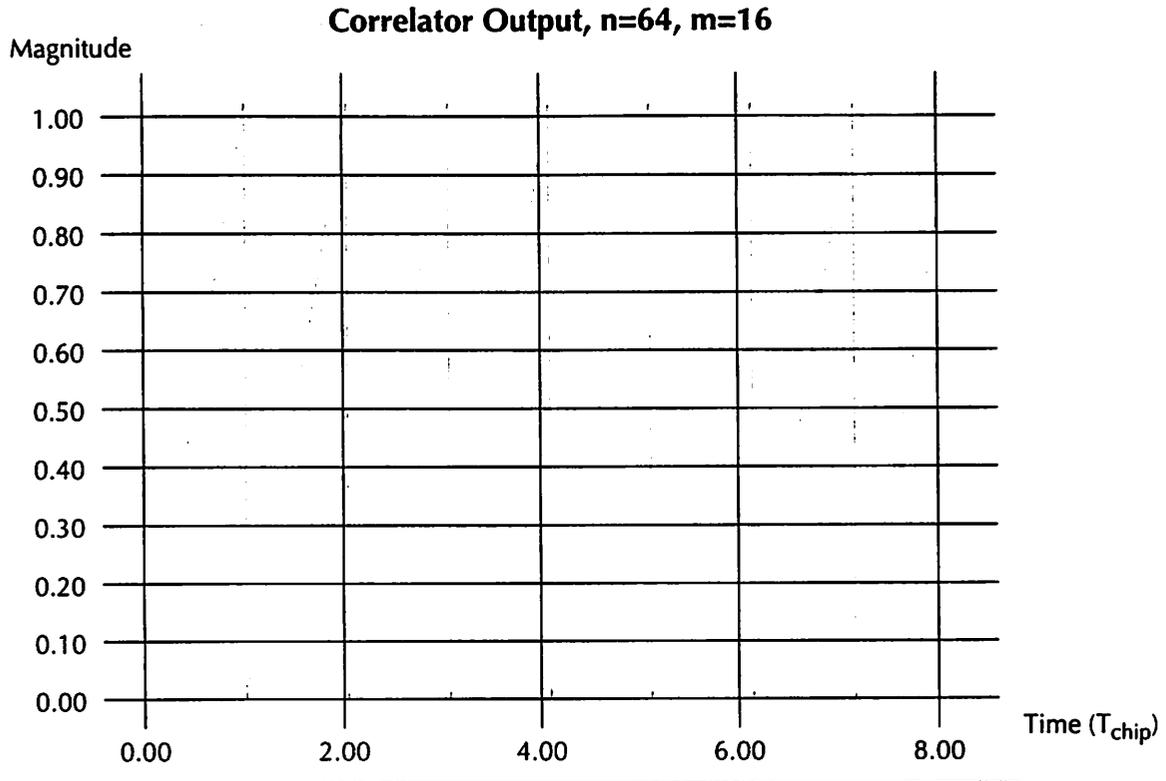


Figure 7.14: Compensated magnitude correlator output, $n=64$, $m=16$. Peak value is 1022, compared to a nominal output of 1024.

It should be mentioned that the magnitude estimators are implemented with an L_1 norm (given as $(|I| + |Q|)$) in the complex-valued I/Q space, instead of a conventional Euclidean L_2 norm $\sqrt{I^2 + Q^2}$. In figure 7.16, the impact of this choice is shown – the unit circle in the complex plane has been deformed into a cycloid, with a worst-case magnitude error of +3dB. The reason for this was complexity minimization; a true Euclidean L_2 norm would have been extremely difficult to implement in a low-power fashion without consuming excessive die area; from simulation, the performance degradation due to this was found to be negligible.

7.3.4. Adjacent-Cell Detection

Lastly, the problem of adjacent-cell detection for mobile-assisted soft-handoff must be addressed. It is quite similar to the problem of coarse timing acquisition: the pilot-tone energy from adjacent cells must be recovered. By maintaining a single sliding correlator, correlating

over a 1024-long interval and changing its phase by $1 T_{\text{chip}}$ every correlation interval, a search can be made for PN-sequence correlation peaks due to nearby cells (figure 7.17). Peaks within $\pm 4 T_{\text{chip}}$ of the current cell's PN correlation timing peak must be discounted, since they are most likely caused by multipath effects. As described in chapter 3, with the long PN sequence, there exists a sufficient number of code-phase offsets to accommodate a $K=7$ cellular reuse pattern, even accounting for PN code synchronization errors between cells. Since only the magnitude of the correlation is necessary to determine if an adjacent cell is present, the carrier-offset compensation technique described in the previous section can be employed.

However, once this sliding correlator has found an adjacent cell (decided when its 1024-long correlation output has exceeded a certain threshold CMP_{TH4}), the issue remains of how to store where these peaks occurred, so a soft-handoff can be performed quickly by the mobile. In essence, some kind of "register file" would be needed to store the PN code phase; three related methods have been devised to perform this.

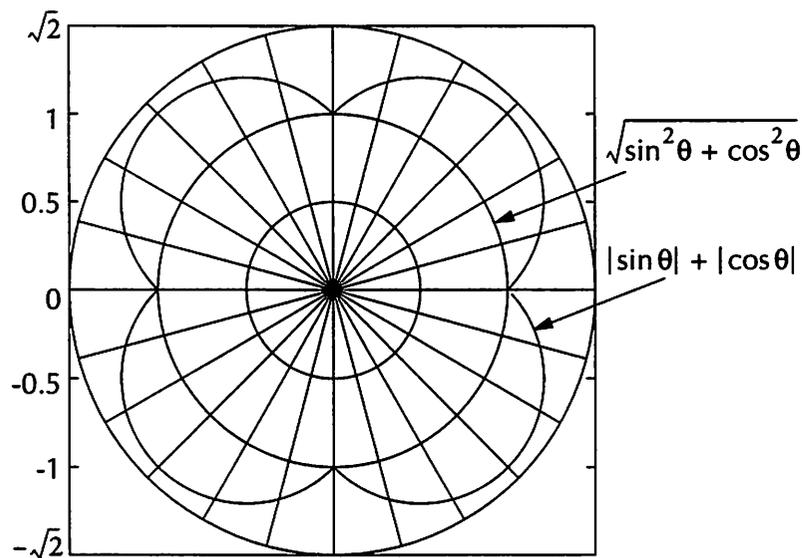


Figure 7.15: Graph comparing the absolute value L_1 norm against the true L_2 norm.

- Seed storage. After detection of an adjacent cell, the PN code phase of the sliding correlator is no longer adjusted. When the code phase of the primary PN generator (locked to the current cell's pilot tone) reaches the all-one's state in its shift register, the state of the sliding correlator's shift register is parallel-stored into a register file, along with the measured strength of the adjacent-cell pilot and the Walsh code counter value at that instant¹. The advantage of this is clear; very little digital hardware is necessary to store the phase relationship of the adjacent-cell pilot. Shifting the primary PN generator over to the new phase is simple: when it reaches the all-one's state in its shift register, the adjacent-cell's seed stored in the register file is parallel-loaded into its shift register, thus forcing it into coarse syn-

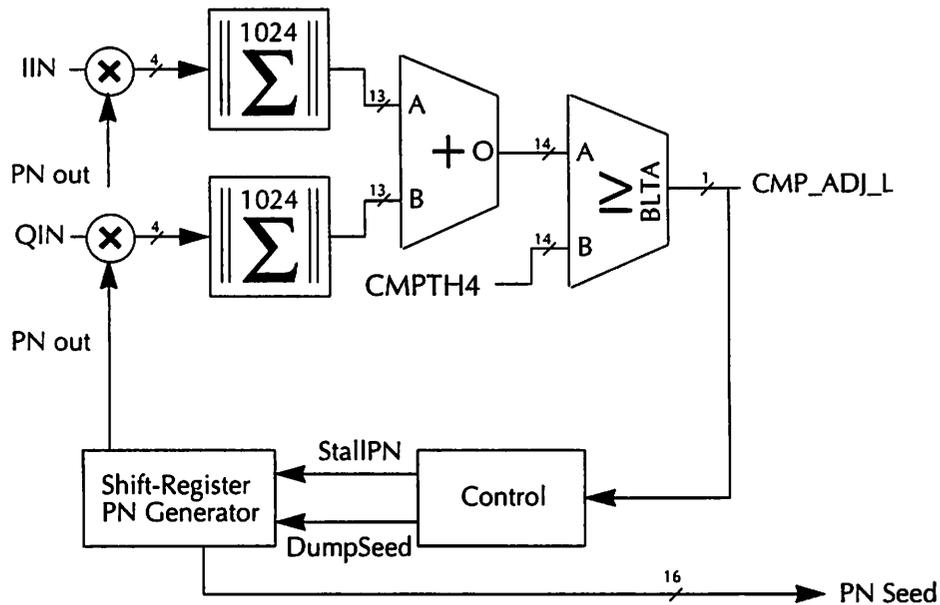


Figure 7.16: Block diagram of the adjacent-cell detector (sliding correlator)

1. The storage of the Walsh code is required to guarantee that both PN and Walsh are synchronized when hand-off occurs. Otherwise, a large time penalty may be incurred to recover the Walsh start point (all one's state in the shift-register PN generator).

chronization with the adjacent cell. The disadvantages are that fine-lock acquisition is still necessary, and moving “back” to the code phase of the original cell is difficult, as its code phasing is lost.

- Multiple PN/Walsh generators. Instead of simply storing the seed, a second strategy would be the implementation of multiple PN/Walsh generators, all running independently. After detection by the sliding correlator, the PN and Walsh code phases are retained by storing them into another PN/Walsh generator. These secondary PN/Walsh generators can be switched-in to replace the primary generators when a handoff is necessary. This solves the problem of moving “back” to the original cell’s code phase after handoff, since its code phasing has been retained. However, the issue of fine-lock acquisition latency still remains, and there is a larger penalty in hardware overhead.
- Replication of the entire receiver delay-locked loop. Lastly, the solution that overcomes all of these issues is to replicate the entire delay-locked loop to track the adjacent cells [QualC92]. Fine acquisition is accomplished by default; the tracking loop is ready to recover data the instant a handoff is called. Obviously, the hardware penalty is enormous - all of the correlators, clock generation, and control circuitry must be replicated¹.

Clearly, there is a tradeoff between the speed at which the soft-handoff can be performed, and the hardware required to implement such a “phase register”. However, in light of the size of the prototype receiver DSP (which does not implement adjacent cell-scan), a compromise solution between simple seed storage and having multiple PN/Walsh generators would seem to be the clear solution: seed storage of the detected adjacent cells, with two parallel PN/Walsh generators to allow quickly jumping back to the original cell’s PN phase. Although replication of the entire receiver DLL provides the best performance, it simply is not practical unless a more advanced CMOS fabrication technology was to become available.

1. 0.9 cm x 0.9 cm, fabricated in a 0.8 micron HP CMOS process. The DLL and data multiplexer blocks comprise slightly over 30% of this.

7.4. Channel Estimation

Like the problem of timing recovery, channel estimation relies on the receiver being able to accurately recover long correlation streams of the pilot tone. As discussed in chapter 3, by utilizing correlators with PN sequences synchronized to on-time, $+1 T_{\text{chip}}$, and $+2 T_{\text{chip}}$ time-sampling instants, the multipath channel profile may be estimated. Since partial correlations are used, a moving-average lowpass filter can be used to filter out the subcorrelation noise (figure 3.9). However, the issue of frequency compensation remains: as before, even with subcorrelations as short as 1024 in length, the worst-case phase change is still intolerable. Unlike the timing recovery case, phase information is critical to the operation of a coherent-combining RAKE receiver. A simple magnitude estimate will not suffice; a method to eliminate the frequency offset of the oscillators must be developed.

The channel estimation algorithm is shown schematically in figure 7.18; it is akin to the magnitude estimator of figure 7.17, in that correlations are taken over length 64, 1 microsecond chunks with a small phase error due to the frequency offset, and the result of the correlations then averaged over 16 microseconds to statistically minimize the noise in the shorter correlations. However, instead of averaging the magnitude of the length-64 chunks, the complex division between the first and second correlations, and the first and third are averaged. By taking the complex division, this provides not only the relative amplitude of the multipath arrival to the primary arrival, but also provides the phase difference between the two. Thus, frequency-offset compensation can be achieved, while preserving the slowly-varying multipath phasing information that is desired.

Figure 7.19 shows a simulation of this estimation technique, and a comparison to the case where a simple 1024-long correlator is used. A three-path channel is assumed, with tap weights $(1.0, 0.8e^{j\pi/5}, 0.2e^{-j4\pi/5})$. Likewise, a 25 kHz oscillator offset between the transmitter and the receiver is intentionally introduced as a sinusoidal modulation at the input to the receiver block, and a three-tap MA filter is applied after the 1024-long output to minimize the noise effects due to subcorrelations (chapter 3). The magnitude and phase estimates are shown for this estimation technique, as well as the case where the simple correlator is as-

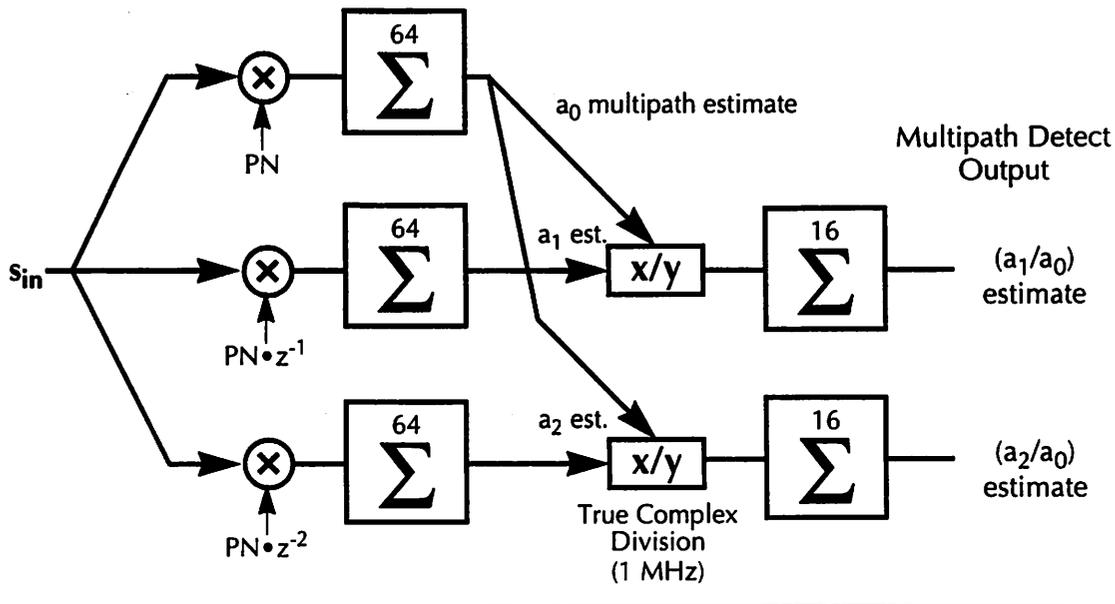


Figure 7.17: Block diagram of multipath estimator.
 At the output of the length-64 correlators, carrier offset is still present. The complex division removes the effect of the carrier, while preserving the relative phasing and amplitude between (a_0, a_1) and (a_0, a_2) . s_{in} is fully complex-valued, being equal to (I_{in}, Q_{in}) .
 Likewise, all thick arcs in the block diagram are complex-valued.

sumed. The results for this technique are quite good; even for an oscillator offset of 25 kHz, the worst-case estimation error (on the weakest multipath arrival) is only 2 dB in amplitude, and 12° in phase.

7.5. Data Recovery

Lastly, data recovery must be discussed. Given that the sampling instant is being tracked by the timing recovery delay-locked loop, and that reliable multipath profiles can be garnered by the channel estimation loop, what remains is the RAKE receiver itself, along with the DQPSK data slicer block. As described in chapter 2, a RAKE coherently combines data recovered: so long as the relative phasing between the first, second, and third arrivals is known, it is sufficient to perform coherent combining. In some sense, the scheme described below is a differential version of the conventional RAKE algorithm [Dixon84].

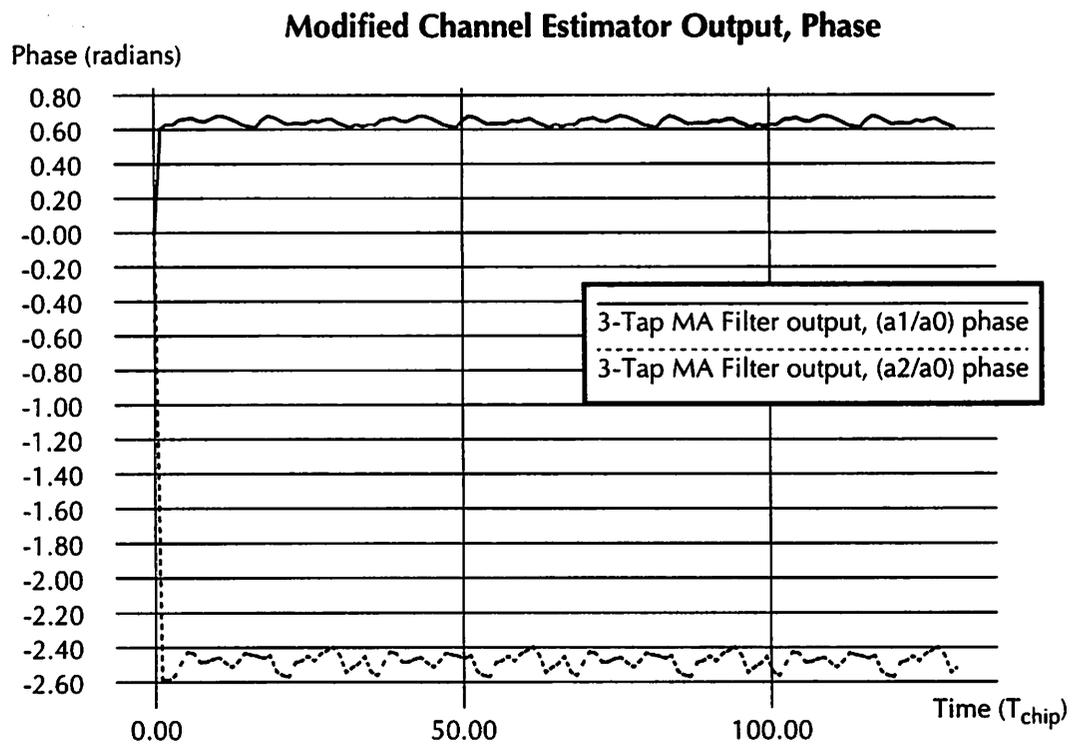
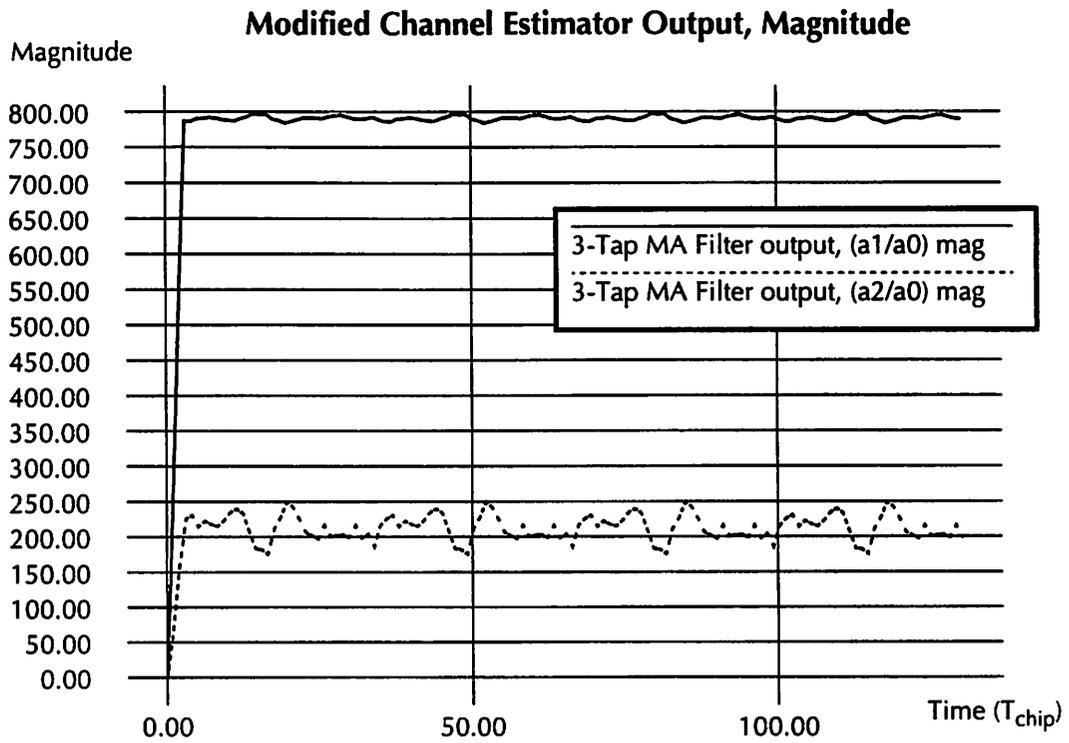


Figure 7.18: Modified channel estimator output (25 kHz frequency offset between transmit and receive oscillators).

The proposed modification is shown schematically in figure 7.20; it is akin to figure 2.14, with a division by a_0^* on all three RAKE fingers followed by a DQPSK slicer. The multipliers on each finger then become the complex conjugate of the differential channel estimates described in the previous section. In the original RAKE scheme, the multiplication by a_0^* , a_1^* , and a_2^* on each finger eliminates any channel phase rotation associated with each multipath arrival: the phase is brought back to zero by the complex conjugate, and the output could be coherently decoded with a simple QPSK slicer instead. Since absolute phase information is not available, dividing all three paths by a_0^* removes the need for absolute phase. Instead of referencing each multipath finger back to zero phase offset, each multipath finger is referenced to the primary arrival a_0 . Thus, a true DQPSK decoder is necessary, since the combined data is phase incoherent, and dependent on the previous symbol for data recovery.

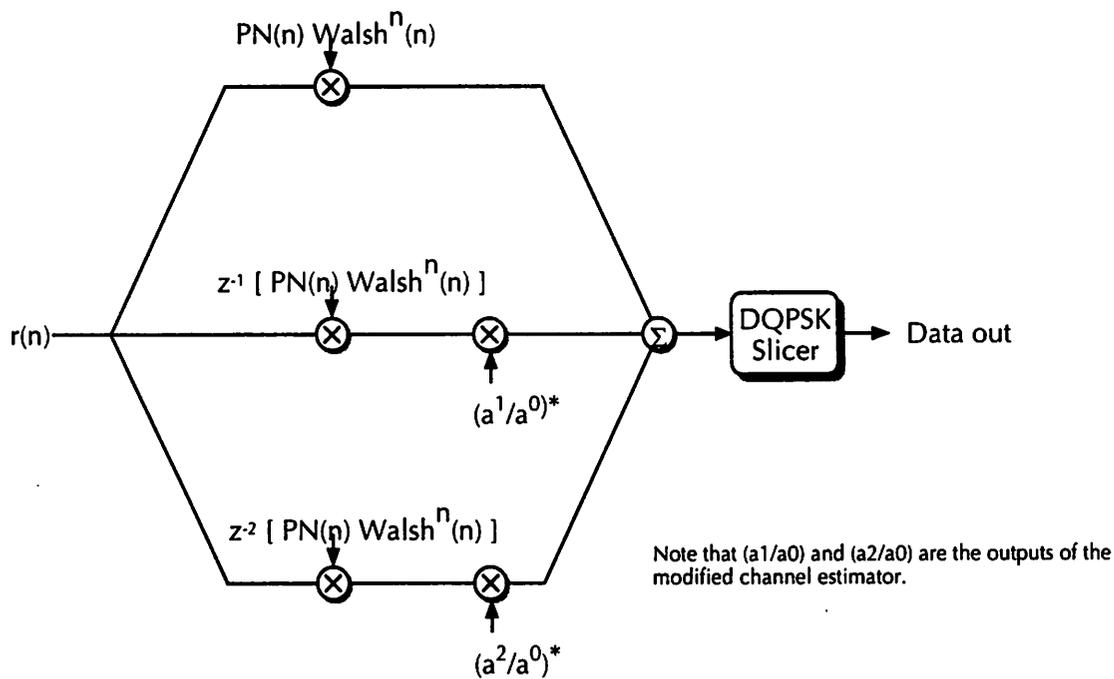


Figure 7.19: Modified RAKE receiver

Lastly, the DQPSK decoder itself needs to be discussed. In essence, the phase difference between two consecutive received symbols must be determined, and the data decoding scheme is fairly straightforward, with the caveat that a sampling demodulator is being used in the an-

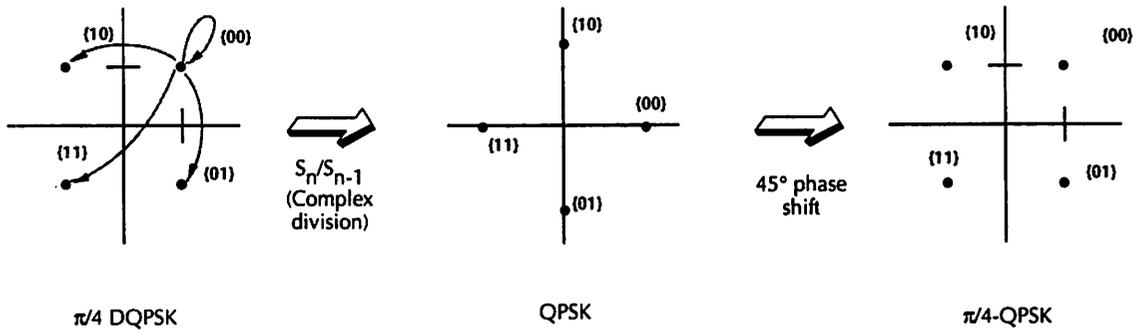


Figure 7.20: Conceptual overview of the DQPSK slicer.

alog front-end. If the delay-locked loop slips the sampling instant, it in effect exchanges the “in-phase” designation with the “quadrature” designation; i.e., a change of $T_{\text{chip}}/4$ at base-band results in a 90° phase shift at RF. Thus, the DQPSK decoder needs to implement a 90° correction in the phase if a change in sampling instant was made by the delay-locked loop. If S_n is the current output of the RAKE, and S_{n-1} is the previous output, then the slicing algorithm is as follows:

$$\text{databits} = \text{sgn}\left(\text{Re}\left(\frac{S_n}{S_{n-1}} e^{j\omega\phi}\right)\right), \text{sgn}\left(\text{Im}\left(\frac{S_n}{S_{n-1}} e^{j\omega\phi}\right)\right)$$

where $\text{sgn}()$ is the sign function, and ϕ is a phase correction. The division of S_n by S_{n-1} results in a complex number with a phase equal to the phase difference between the two symbols. If there is no shift in the DLL sampling instant, a ϕ of 45° translates the DQPSK constellation into an absolute phase QPSK, as shown in figure 7.21. The slicing operation is then particularly simple, with the data bits corresponding to the sign bits of the real and imaginary parts. Although it would seem that the phase correction would require a full complex-valued multiplier, the amplitude-independence of the sgn function implies that the phase correction can be implemented with two adders, since the multiplication will be against $(+1.0, +1.0)$, $(-1.0, 1.0)$, or $(1.0, -1.0)$, for $\phi = 45^\circ$, 135° , and -45° respectively. However, the key question that remains is the implementation of the complex division of S_n by S_{n-1} ; well-known iterative methods such as CORDIC are well suited to this application.

7.6. Results

In figure 7.22, the die photo of the receiver DSP testchip is shown. It implements a subset of the functionality described in section 7.1; the block diagram for this die is shown in figure 7.23. In particular, the complex-valued computation (required for channel estimation and data recovery) is not present; these functions were not necessary to test the full functionality

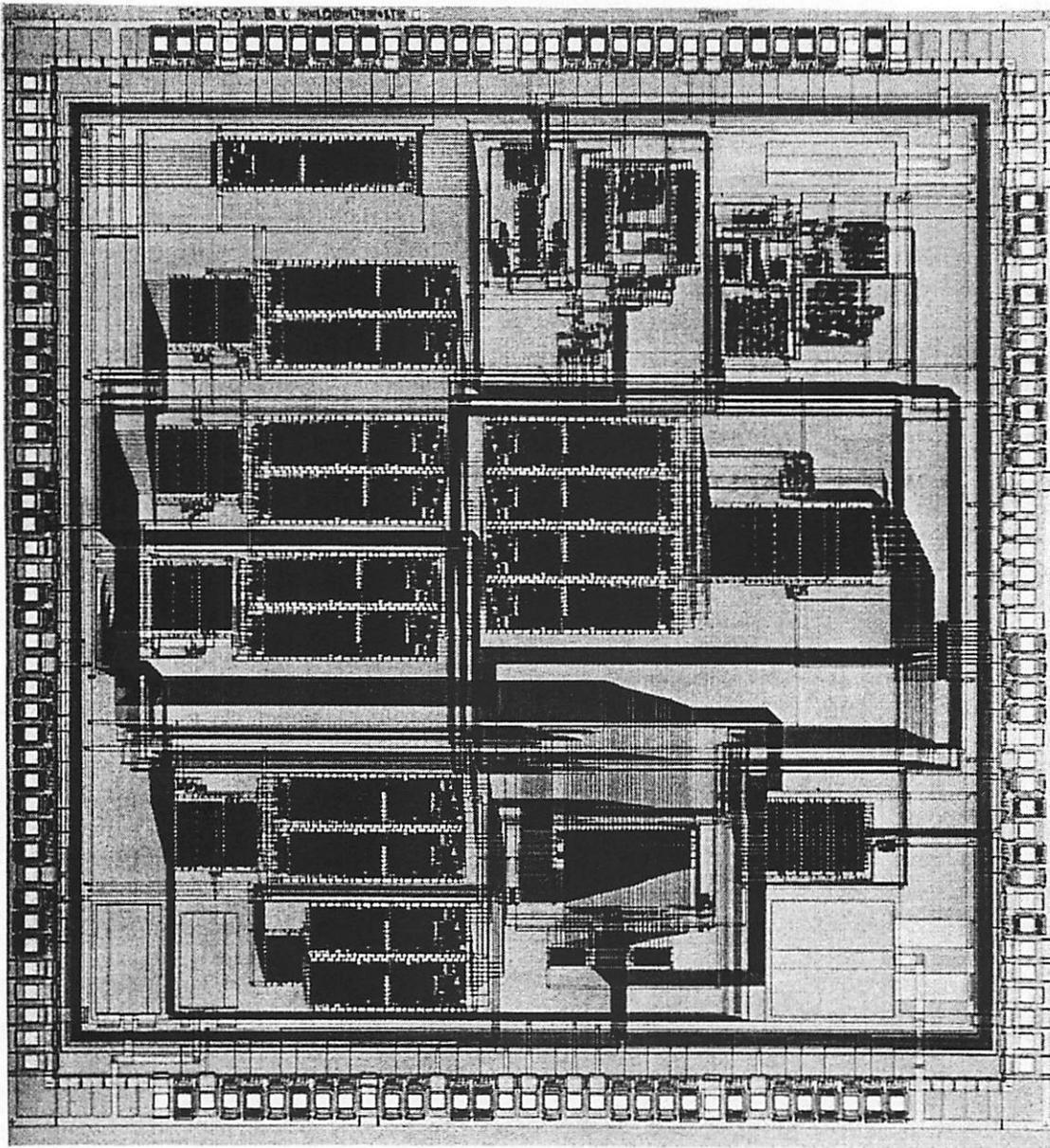


Figure 7.21: Die photo of the receiver DSP prototype testchip.

the standalone correlator and the full receiver DSP testchip bear testament that high-speed digital circuitry, coupled with careful architectural design, can operate with extremely low power consumption.

7.7. Appendix: Receiver DSP Electrical Interface

This appendix documents the electrical interface for the DSP receiver prototype described in section 7.6. More detailed circuit-level documentation can be found in [Stone95].

7.7.1. Input/Output Description

PIN	I/O/ SUPPLY	NAME	PIN	I/O/ SUPPLY	NAME	PIN	I/O/ SUPPLY	NAME
8	S	GND_1	9	S	VDD_11	10	I	DATAIN5
11	O	ODATA27	12	I	OSCH	13	O	ODATA10
14	S	VDD_1	15	I	CURRBIAS	16	S	VDD_6
17	I	ECRSTDUMP	18	I	OSCL	19	O	ODATA9
20	O	ODATA26	21	S	GND_5	22	S	PWR1_5
23	I	RESETL	24	S	VDD_4	25	I	QINX2
26	O	ODATA25	27	O	DUMP1024H	28	O	ODATA8
29	I	IINX3	30	I	DATAIN14	31	O	DUMP64H
32	S	PWR3_1	33	S	PWR5_4	34	I	ECDATA2
35	S	PWR5_1	36	S	PWR3_2	37	O	ODATA7
38	I	IIN3	39	I	IIN0	40	O	CMP3THL
41	I	ADDR1	42	I	DATAIN13	43	S	GND_8
44	I	ADDR0	45	O	ODATA17	46	O	ODATA6
47	O	ICLK	48	I	DATAIN12	49	S	VDD_7
50	O	QCLK	51	O	ODATA16	52	I	ECDATA1
53	S	PWR1_1	54	I	DATAIN11	55	O	ODATA5
56	S	PWR5_2	57	I	IINX0	58	S	PWR1_6
59	S	VDD_2	60	O	ODATA15	61	O	ODATA4
62	I	CSL	63	S	PWR1_8	64	I	QINX1
65	I	WRL	66	I	ECW	67	S	PWR3_5
68	I	EC64CLK	69	O	ODATA14	70	S	VDD_8
71	I	IIN2	72	O	CMP2THH	73	I	QIN1
74	S	PWR1_2	75	S	VDD_5	76	I	DATAIN4
77	O	ODATA24	78	O	ODATA13	79	I	DATAIN3

Table 7.1: Pinout of demodulator chip listed in order of pin number.

PIN	I/O/ SUPPLY	NAME	PIN	I/O/ SUPPLY	NAME	PIN	I/O/ SUPPLY	NAME
80	I	TSTMODE0	81	S	PWR5_5	82	I	CLKRST
83	O	ODATA23	84	I	ECDATA3	85	S	VDD_9
86	I	TSTMODE1	87	O	ODATA12	88	S	PWR5_6
89	I	IINX2	90	I	DATAIN10	91	S	PWR3_6
92	O	ODATA22	93	O	ODATA11	94	S	PWR1_7
95	S	GND_2	96	I	QINX3	97	S	PWR5_7
98	I	ECDUMP	99	I	DATAIN9	100	S	PWR3_7
101	O	ODATA21	102	S	GND_6	103	S	GND_9
104	O	LOCK	105	I	QIN3	106	I	QIN0
107	S	PWR5_2	108	I	DATAIN8	109	I	DATAIN2
110	O	ODATA20	111	S	PWR1_3	112	O	ODATA3
113	I	OMODE2	114	O	DUMPRST	115	I	DATAIN1
116	I	ECPN	117	O	STALL	118	O	ODATA2
119	O	ODATA19	120	S	PWR1_4	121	I	DATAIN0
122	S	VDD_3	123	S	PWR3_3	124	I	QINX0
125	O	ODATA18	126	I	DATAIN7	127	O	ODATA1
128	I	IINX1	129	S	PWR3_4	130	S	VDD_10
131	I	OMODE1	132	S	GND_7	133	I	ECDATA0
134	I	OMODE0	135	I	DATAIN6	136	O	ODATA0
137	I	IIN1	138	I	QIN2	139	O	STRETCHSAMP

Table 7.1: Pinout of demodulator chip listed in order of pin number.

NAME	PIN COUNT	DESCRIPTION	PIN(S)
CLKRST	1	Reset signal for clock generator	113
RESETL	2	Chip reset signal	6
OSCH	3	Positive phase of 128MHz sinusoidal oscillator	46
OSCL	4	Negative phase of 128MHz sinusoidal oscillator	48
CURRBIAS	5	Bias voltage for clock pad	47
DATAIN[14:0]	20	Bus used to load values into registers	52,56,58,60,72,75,78,84,87,89,111,112,122,124,126
ADDR[1:0]	22	Register address lines	12,13
CSL	23	Chip select	19
WRL	24	Write signal for register	20
OMODE[2:0]	27	Select which signals to output	36,42,43
TSTMODE[1:0]	29	Determines the type of input to the chip	27,25
IIN[3:0]	33	Input data from ADC or transmitter chip	11,22,44,55

Table 7.2: Pinout for demodulator chip grouped by signal name.

NAME	PIN COUNT	DESCRIPTION	PIN(S)
QIN[3:0]	37	Input data from ADC or transmitter chip	77,88,110,121
IINX[3:0]	41	Input data from transmitter chip (for testing only)	8,28,41,61
QINX[3:0]	45	Input data from transmitter chip (for testing only)	74,94,107,127
ODATA[27:0]	73	Output data bus	2,5,7,24,26,29,32,35,38,40,57,59,62,65,68,71,73,90,92,95,98,101,104,106,123,125,128,131
ICLK	74	unused-connected to ground	14
QCLK	75	internal system clock	15
LOCK	76	high when mobile has acquired lock	33
STALL	77	low when PN generator is being stalled	81
STRETCH-SAMP	78	0: Adjusted phase by $-T_{chip}/4$ 1: Adjusted phase by $+T_{chip}/4$	132
CMP2THH	79	0: No adjustment made to phase of clock 1: Adjusted phase of clock	66
CMP3THL	80	0: Not enough energy to be in lock 1: Still have energy to be in lock	99
DUMPRST	81	Indicates that the MP1 correlator was cleared	80
DUMP64H	82	Indicates when output of data recovery correlator is valid	96
DUMP1024H	83	Indicates when output of long correlators are valid	51
ECRSTDUMP	84	Extra correlator reset latches while dumping	4
EC64CLK	85	Extra correlator clock	21
ECDUMP	86	Extra correlator dump output	31
ECPN	87	Extra correlator PN input	37
ECW	88	Extra correlator walsh input	64
ECDATA[3:0]	92	Extra correlator data	70,97,103,130
PWR1[100	1.5V internal power supply	16,23,63,79,82,93,105,117
PWR3	107	3.3V internal power supply	9,54,83,85,108,116,119
PWR5	114	5V internal power supply	10,17,34,53,69,115,118
VDD	125	5V pad power supply	3,18,39,45,50,67,91,102,109,114,129
GND	132	ground	1,30,49,76,86,100,120

Table 7.2: Pinout for demodulator chip grouped by signal name.

7.7.2. Programmable Registers

The registers in the demodulator are two-level latches. This allows the front half of the register to be externally written into without affecting the internal operation of the chip which depends on the registers' values. A block diagram of a register is shown in figure 7.23, along with its control signals. The signal CLKX (X signifies the register number) is externally controlled by the CS_L, WR_L and ADDR signals. Since there is only one set of the ADDR, CS_L and WR_L lines, only one register can be updated at a time. The backend of the register is updated when CLK_BACK goes high. This occurs every rising edge of CLK64 during reset and when the PN generator's output is all ones during normal operation. There are four registers in the demodulator whose functions and bit-widths are described in Table 7.3.

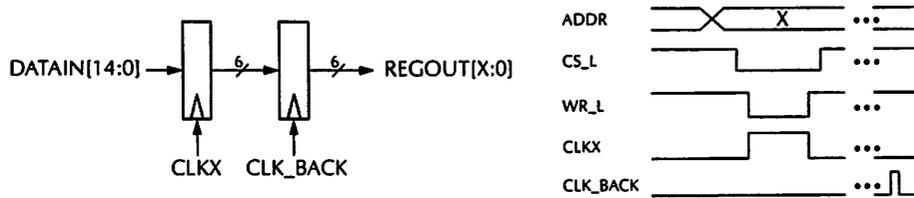


Figure 7.23: Block diagram of two level registers along with timing diagram.

ADDR	NAME	BIT WIDTH	FUNCTION
00	WALSH	6	Walsh number
01	THRESA	14	Used to determine whether coarse lock has been acquired
10	THRESB	15	Used to determine whether the Infopad is still in lock
11	THRESC	15	Used to determine whether clock phase should be adjusted by $T_{chip}/4$

Table 7.3: Describes the demodulator's four writable registers.

7.7.3. Observation Modes

The OMODE signal is a 3-bit signal that controls which signals will be driven onto the 28-bit output bus. A description of all eight possibilities along with an explanation of the signals outputted can be seen in figure 7.24.

7.7.4. Test Modes

The demodulator has three functional modes (see table 7.4) which are controlled by the 2-bit TSTMODE signal. One mode is used for regular operation and two are used for testing purposes. Independent of the test mode, after the data has passed through the testmode block, it will consist of two 4-bit sign-magnitude interleaved streams at a frequency equal to one-half OSCH's frequency.

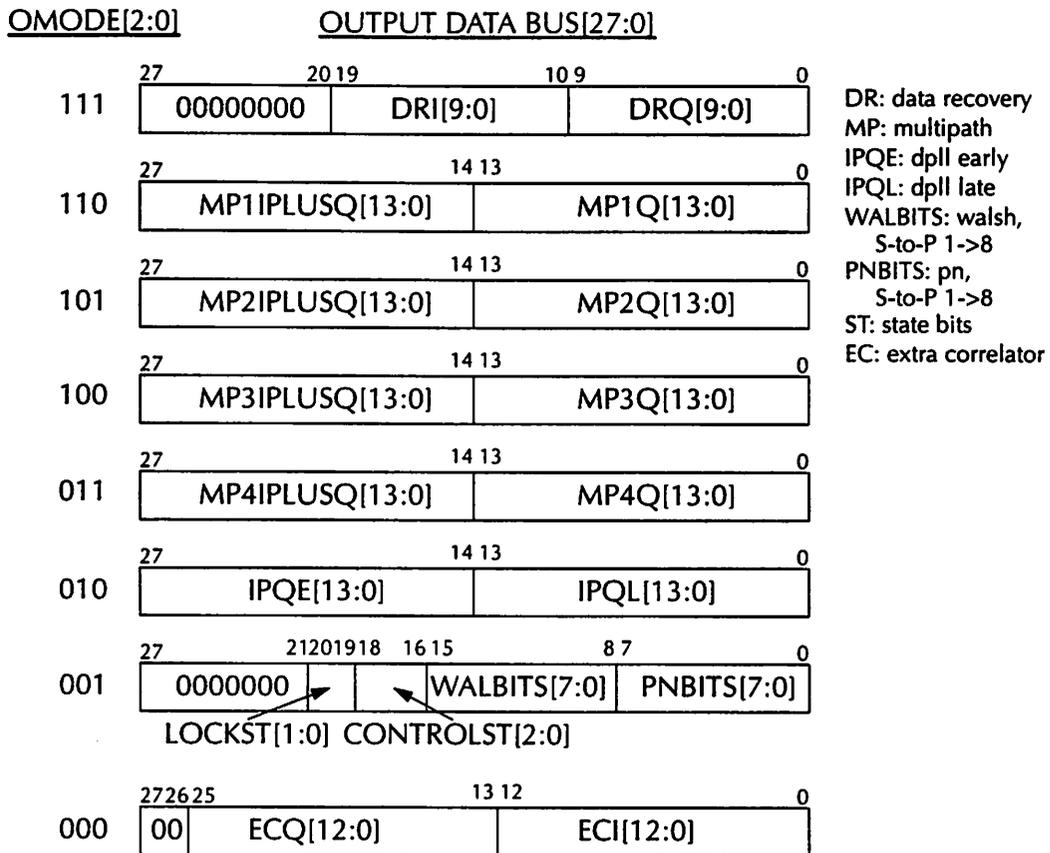


Figure 7.24: Receiver DSP output signals, as a function of observation mode select.

TSTMODE	MODE	Inputs	Comments
00	normal	IIN,QIN	Data is unmodified
01	test	IIN,QIN	Data converted from 4-bit binary (0 to 15) to 4-bit sign magnitude (-7 to 7)
10	test	IIN,QIN, IINX,QINX	Data converted from 4 parallel 2's complement (-8 to 7) streams at 1/2 OSCH's frequency to two interleaved sign-magnitude (-7 to 7) streams at OSCH's frequency
11	UNDEF.	UNDEF.	UNDEF.

Table 7.4: Description of the three operating modes for the demodulator.

8 Conclusions and Future Work

In providing future personal communication services, broadband multimedia data services - the most demanding of which being real-time, on-demand video - will be delivered over wireless transmission systems to high densities of users. Even with the best video and data compression schemes available today, this would require a per-user sustained data rate in excess of 2 megabit/second - a sustained rate far higher than what existing cellular phones can provide, and an aggregate total rate over all users far higher than what existing wireless LAN's can provide. In postulating such a scenario, the two obvious questions that then arise are how to provide the necessary data rate to each individual while maintaining overall system capacity, and the feasibility of implementing such a transceiver in a low-power, portable fashion.

In this thesis, one possible answer to these questions has been explored. Restricting the system to an indoor environment only, and assuming that overall capacity can be maintained by picocellular deployment, the design described here simultaneously supplies a 2 Mbps/user bandwidth for up to 50 users in a single cell. This data rate is assumed to be asymmetric; the high data rate is only available in the downlink, with a lower-rate uplink to support simple control and user input; developing a downlink capable of an aggregate system bandwidth of 100 Mbps is thus the goal of this project. The use of spread spectrum techniques, in particular direct sequence code-division multiple access (CDMA), provides the multiple-access strategy to maintain parallel, separate streams of real-time data to all the users. Since the base station can provide synchronization and serves as a single point of transmission, the near-far problems and capacity issues with traditional direct-sequence systems can be avoided. Likewise, the intrinsic noise rejection inherent in spread-spectrum is particularly desirable, as it can be used to reduce sensitivity to interference and narrow band multipath fades in the radio environment, as well as relax hardware performance constraints.

Since the overall goal is to provide high-speed wireless communications in a low-cost, portable device, a standard digital single-poly 0.8 micron CMOS technology is used in the implementation of both the analog and digital circuitry. Three chips have been designed which implement most of the transceiver path: the digital baseband modulator; the analog receiver front-end and A/D; and the digital receiver baseband processor. Since it is assumed that this set of chips will be used in a downlink (base station to portable), particular focus has been placed on the power consumption of the analog and digital receiver circuits.

The transmit architecture of the system is based loosely on that of the IS-95 digital cellular standard, scaled to a spread-spectrum chipping rate of 64 MHz; it uses a hybrid Walsh-PN code to achieve the desired cross-correlation and autocorrelation properties. To aid receiver synchronization, the zeroth Walsh code is expended as a pilot tone; this minimizes the probability of false lock, and enables the receiver to perform continuous multipath profile monitoring as well as adjacent cell detection. A monolithic prototype modulator has been developed for the base station, implementing all of the baseband processing required before single-sideband upconversion to the RF carrier; this includes spreading each data bit with a unique 64-long Walsh code, and then modulating the combined user data streams with the 32768-long PN code. To provide better control of capacity and interference within each cell, the separate data streams are combined with provision for individual user power control, as well as an overall power scaling factor. Lastly, to bandlimit the transmit signal in a controlled, zero-intersymbol interference fashion, a digital raised-cosine filter has also been included on the modulator prototype; this filter effectively runs at 256 MHz, implemented as four interleaved 64 MHz subfilters. Currently, the die accepts 15 parallel user inputs, and several die can be synchronously connected to achieve up to 50 users operating from a single base station. Lastly, for analog upconversion purposes, a semicustom single-sideband RF modulator has been implemented, providing D/A conversion of the prototype modulator's output and frequency translation to a carrier of 1.088 GHz.

Clearly, the most critical design issues of the hardware lie in the development of the receiver radio-frequency components. In addition to the requirement of low power consumption, the analog RF circuitry is complicated by the need to operate at frequencies near or above 1 GHz.

Such circuitry has been traditionally dominated by designs using discrete GaAs, silicon bipolar transistors and stripline filters, which consume significant amounts of area on a circuit board and excessive amounts of power, especially when matching to standard 50Ω impedance levels. However, by the use of proper tuning networks, CMOS operation in the 1 GHz band has been demonstrated to be quite feasible.

At the receiver, the incoming RF signal is amplified by a continuous-time low-noise amplifier, and the mixed to baseband (actually 64 MHz) via direct sampling of the modulated signal. This eliminates the need for a conventional superheterodyne mixing and IF processing, and makes it possible for the variable gain function of the AGC to be implemented in sampled-data amplifiers. The signal is then A/D converted using dual interleaved 4-bit flash converters at 256 MHz (128 MHz I and Q). The remainder of the processing - timing and data recovery - is executed by the digital baseband receiver circuit, which performs coarse and fine timing recovery at 64 MHz using scanning correlators and a delay-locked loop with a time resolution of 4 nsec. After timing acquisition is obtained, the I and Q data are then despread at 64 MHz using length 64 matched-filter correlators.

A two-chip prototype receiver - consisting of a monolithic analog front-end and a high-speed baseband DSP - has been built and tested running at a carrier of 1.056 GHz; due to technology limitations, the baseband data rate is limited to 1 Mbps/user instead of the desired 2 Mbps. Thus, functionality has been demonstrated in the receiver with a spread-spectrum rate of 32 MHz instead of 64 MHz, and an analog sampling rate of 128 MHz (64 MHz I and Q). However, total peak power consumption for the receiver prototype (both analog and digital integrated circuits) is 154 mW, far lower than any receiver reported to date, and especially so in light of the supported data rate.

The large power savings that has been demonstrated arise from the architectural design of the receiver: the chosen architecture is quite unlike the superheterodyne architectures used in radio systems to date. Channel selection, timing recovery, data recovery, and frequency recovery are all performed in the digital domain; the analog front-end consists only of amplification, filtering, and a fast, low-resolution A/D conversion. The reason for this is clear: low-power, high-speed digital signal processing can be achieved through architectural voltage scaling

techniques in the digital domain – techniques that do not translate easily into the analog domain. In spite of the tremendous amount of digital processing required, and the minimization of the amount of analog hardware required by the system, the power consumed by the prototype receiver DSP is still less than one half of that consumed by the analog front-end – testament to the effectiveness of these techniques. Such an “A-to-D front-end” receiver architecture hinges on this: by pushing the analog/digital boundary as close to the antenna as possible, these digital techniques can be brought to bear as early in the signal chain, resulting in a receiver that can process a broadband 64 Mbps data stream while consuming a total of 154 mW of power.

With this in mind, several directions can be taken for future research. First, the limits of tuned-CMOS RF can be explored further: there is unlicensed spectrum and new services emerging in the 2 GHz and 5 GHz bands. Given recent research in on-chip inductor tuning techniques [Nguyen91, Ou95], one potential avenue of research is the adaptation of sampling demodulation techniques to these higher frequencies. While this project has demonstrated frequency limits due to jitter of approximately 1 GHz for a 0.8 micron process, 0.5 micron and 0.35 micron processes are becoming readily available, with concomitant improvements in sampling-edge jitter performance.

Furthermore, it is not clear that the RAKE strategy is optimal in the case of correlated interference, as what is seen in an indoor environment. Recent research [Teus94] has shown that there may be better multipath combining strategies than a straight RAKE; likewise, application of trellis codes and Viterbi detection may also yield gains at the system level, especially with regard to relaxing analog performance requirements. Along these lines, with the high levels of integration and low power consumption demonstrated, multiple receivers with phase-coherent combining become quite viable in an integrated implementation. Effectively, a monolithic, electronically steerable antenna array could be formed, further improving multipath rejection and improving SNR at the receiver.

Lastly, there is a great deal of current interest with regard to discrete multitone (DMT), also known as orthogonal frequency-division multiplexing (OFDM) [Ho96, Zogak95, Wu95]. In some sense, OFDM provides spreading redundancy in the frequency domain: the time do-

main and the frequency domain are interchanged through the use of a Fast Fourier Transform in both the transmitter and receiver. DMT-based systems have become popular in wired communications, having been employed in the CCITT v.34 modem standard and emerging broadband copper services like the asymmetric digital subscriber loop (ADSL). DMT promises the benefits of spread-spectrum, while improving immunity to multipath and narrow-band interference. However, it requires precise knowledge of the transmission channel profile: adaptation into a wireless environment, where fading is a serious issue, is an open research question. In any case, a wireless DMT system can leverage off the design techniques explored in this project – compensating for the performance limitations of integrated CMOS RF analog by careful design tradeoffs at the system level, and by exploiting low-power, high-performance digital techniques as much as possible.

Bibliography

- [Afgh91] M. Afghahi and J. Yuan. "Double Edge-Triggered D-Flip-Flops for High-Speed CMOS Circuits." *IEEE Journal of Solid-State Circuits*, Vol. 26, No. 8. August 1991.
- [Bakog90] H.B. Bakoglu. *Circuits, Interconnections, and Packaging for VLSI*. New York: Addison-Wesley Publishing Co., 1990.
- [Beau84] K. G. Beauchamp. *Applications of Walsh and Related Functions*. Orlando: Academic Press, 1984.
- [Behzad95] A. Behzad. *The Implementation of A High Speed Experimental Transceiver Module with an Emphasis on CDMA Applications*. M.S. Thesis, U.C. Berkeley, Berkeley, CA, 1995.
- [Berle68] E.R. Berlekamp. *Algebraic coding theory*. New York: McGraw-Hill Book Co., 1968.
- [Black91] K.L. Blackard et al. "Radio Frequency Noise Measurements and Models for Indoor Wireless Communications at 918 MHz, 2.44 GHz, and 4.0 GHz." Submitted to *IEEE Transactions on Antennas and Propagation*, June 1991.
- [Buck93] J. Buck. "The Ptolemy Kernel: A Programmer's Companion for Ptolemy 0.4," UCB/ERL Memorandum No. M93/8, January 19, 1993.
- [Chand92] A. Chandrakasan, S. Sheng, and R.W. Brodersen. "Low-Power CMOS Digital Design." *IEEE Journal of Solid-State Circuits*, Vol. 27, No. 4, pp. 473-484. April 1992.
- [Chand94] A. Chandrakasan. *Low Power Digital CMOS Design*. Ph.D. Thesis, U.C. Berkeley, Berkeley, CA, 1994.
- [Conroy94] C. Conroy. *A High-Speed Parallel Pipeline A/D Converter Technique in CMOS*. Ph.D. Thesis, U.C. Berkeley, February 1994.
- [Cook83] C.E. Cook and H.S. Marsh. "An Introduction to Spread Spectrum," *IEEE Communications Magazine*, pp. 8-16. March 1983.
- [Daden91] Daden Associates. *Standard-Series Micro-Miniature Bandpass Filters*, Data Sheet. Fountain Valley, CA, 1991.
- [Dixon84] R.C. Dixon. *Spread Spectrum Systems, 2nd ed.*, New York: J. Wiley and Sons, 1984.

- [Ekel87] B. Ekelund and M. Torkelson. "Waveform Generation for CPM Transmitters." Memorandum of the Department of Applied Electronics, University of Lund, Lund, Sweden, January 1987.
- [Feher87] K. Feher. *Advanced Digital Communications*. New Jersey: Prentice-Hall Inc. 1987.
- [Fields82] D. Fields. *Mrs. Fields Cookie Book*. New York: Time-Life Books. 1992.
- [Gold67] R. Gold. "Optimal Binary Sequences for Spread-Spectrum Multiplexing." *IEEE Transactions on Information Theory*, Vol. IT-13, pp. 619-621. October 1967.
- [Gold68] R. Gold. "Maximal Recursive Sequences with 3-Valued Recursive Cross Correlation Functions." *IEEE Transactions on Information Theory*, Vol. IT-14, pp. 154-156. January 1968.
- [Gonz84] G. Gonzalez. *Microwave Transistor Amplifiers: Analysis and Design*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
- [Gray84] P.R. Gray and R.G. Meyer. *Analysis and Design of Analog Integrated Circuits, 2nd Edition*. New York: John Wiley and Sons, 1984.
- [Gray89] P.R. Gray. Course Notes for EECS 290Y, Analog-Digital Converter Technologies. Department of EECS, U.C. Berkeley, Berkeley, CA. Spring 1989.
- [Ho96] M. Ho M, J.M. Cioffi, and J.A.C. Bingham. "Discrete Multitone Echo Cancellation." *IEEE Transactions on Communications*, Vol. 44, No. 7, pp. 817-825. July 1996.
- [Kasami66] T. Kasami. "Weight Distribution Formula for Some Class of Cyclic Codes," Technical Report No. R-285. Coordinated Science Laboratory, Univ. of Illinois, Urbana, IL, April 1966.
- [Koul91] S. Koul and B. Bhat. *Microwave and Millimeter Wave Phase Shifters*. Boston: Artech House, 1991
- [LNA81] H. Fukui, ed. *Low-Noise Microwave Transistors and Amplifiers*. New York: IEEE Press, 1981.
- [Lau94] J. Lau. *Chip On Board: Technologies for Multichip Modules*. New York: Van Nostrand-Reinhold, 1994.
- [Lee88] E.A. Lee and D.G. Messerschmitt. *Digital Communication*. New York: Kluwer Academic Publishers, 1988.
- [Lee89] W. C-Y Lee. *Mobile Cellular Telecommunications Systems*. New York: McGraw-Hill Book Co., 1989
- [Lin90] Y-M. Lin. *Performance Limitations of High-Resolution Video-Rate Analog-Digital Interfaces*. Ph.D. Dissertation, Memorandum No. UCB/ERL M90/55, U.C. Berkeley, Berkeley, CA, June 19,1990.
- [Linn95] J.P. Linnartz. "Wireless communication networks: an intensive short course." U.C. Berkeley, Berkeley, Berkeley, CA, 1995.

- [Lynn95] L. Lynn. *Low-Power Analog Circuits for an AII-CMOS integrated CDMA receiver*. M.S. Thesis, U.C. Berkeley, Berkeley, CA, 1995.
- [MCM91] R. Johnson, R. Teng, and J. Balde, eds. *Multichip Modules: System Advantages, Major Constructions, and Materials Technologies*. New York: IEEE Press, 1991.
- [MPEG90] *Coding of Moving Pictures and Associated Audio, ISO-IEC/JTC1/SC2/WG11/MPEG, ISO 11172, MPEG Video Committee*. International Organization for Standardization, December 18, 1990.
- [Murata93] Murata-ERiE Corporation. *Microwave Ceramic Filter Product Guide, Catalog No. M-10-B*. Smyrna, GA, 1993.
- [Murota81] K. Murota and K. Hirade. "Transmission Performance of GMSK Modulation." *Transactions of the IECE of Japan*, Vol. 64-B, p. 1123. October 1981.
- [Nguyen91] N. Nguyen. "A 1.8 GHz Monolithic LC Voltage-Controlled Oscillator." Integrated Circuits Seminar, Department of EECS, U.C. Berkeley, Berkeley, CA, October 7, 1991.
- [Nichol91] H.T. Nicholas, III and H. Samueli. "A 150-MHz Direct Digital Frequency Synthesizer in 1.25-um CMOS with -90-dBc Spurious Performance." *IEEE Journal of Solid-State Circuits*, Vol. 26, No. 12, pp. 1959-1969. December 1991.
- [Ohrvi89] S.O. Ohrvick et al. "Microcell Radio Channel: Preliminary Report on Indoor Field-Strength Measurements at 900 and 1700 MHz." Memorandum of the Department of Applied Electronics, University of Lund, Lund, Sweden, January 1989.
- [Ou95] J.J. Ou. "CMOS Low-Noise Amplifier Design for Monolithic RF Modems." Ph.D. Qualifying Exam Presentation, U.C. Berkeley, Berkeley, CA, November 1995.
- [Peroul94] J. Peroulas. *Design and implementation of a High-Speed CDMA Modulator for the INFOPAD Basestation*. M.S. Thesis, U.C. Berkeley, Berkeley, CA, 1996.
- [Phil95] W.S. Djen, Philips Semiconductors. *Using the SA7025A and SA8025A for Narrowband Systems, Application Note AN-1890*. Eindhoven, The Netherlands, Jun 18, 1995.
- [Pick82] R.L. Pickholtz and D.L. Schilling. "Theory of Spread-Spectrum Communications - A Tutorial." *IEEE Transactions on Communications*, Vol. COM-30, No. 5, pp. 855-883. May 1982.
- [Pozar90] D.M. Pozar. *Microwave Engineering*. New York: Addison-Wesley Publishing Co., 1990.
- [Proak83] J.G. Proakis. "Adaptive Equalization for TDMA Digital Mobile Radio." *IEEE Transactions on Vehicular Technology*, Vol. VT-40, No. 2, pp. 323-332. May 1991.
- [ProakB83] J.G. Proakis. *Digital Communications*. New York: McGraw-Hill Book Co., 1983
- [Prox94] RangeLAN2 Wireless Modem Product, Press Release. Proxim, Inc., Mountain View, CA, March 30, 1994.

- [Qual92] *ETSI/TIA IS-95 Interim Standard, Wideband Spread-Spectrum Digital Cellular Dual-Mode Mobile Station-Base station Compatibility Standard.* Telecommunications Industry Association, April 1992
- [QualC92] *An Overview of the Application of Code Division Multiple Access (CDMA) to Digital Cellular Systems and Personal Cellular Networks.* Qualcomm Inc., May 21, 1992.
- [RFM90] F. Perkins, R.F. Monolithics. *High-Speed Video Dot Clocks for High-Resolution Computer Display Applications, Application Note.* Dallas, TX, 1990.
- [Raith91] K. Raith and J. Uddenfelt. "Capacity of Digital Cellular TDMA Systems." *IEEE Transactions on Vehicular Technology*, Vol. VT-40, No. 2, pp. 323-332. May 1991.
- [Saleh87] A.M. Saleh and R. A. Valenzuela. "A Statistical Model for Indoor Multipath Propagation." *IEEE Journal of Selected Areas in Communications*, Vol. SAC-5, No. 2, pp. 128-137. February 1987.
- [Sarwa80] D.V. Sarwate and M.B. Pursley. "Crosscorrelation Properties of Pseudorandom and Related Sequences." *Proceedings of the IEEE*, Vol. 68, pp. 593-619. May 1980.
- [Schill90] D.L. Schilling, R.L. Pickholtz, and L.H. Milstein, eds. "Spread-Spectrum Communications I and II." Special Issues of the *IEEE Journal of Selected Areas in Communications*, Vol. JSAC-8, No. 4-5. May-June 1990.
- [Scholtz79] R.A. Scholtz. "Optimal CDMA Codes." *1979 National Telecommunications Conf. Record*, Washington, D.C., pp. 54.2.1-54.2.4. November 1979.
- [Seidel89] S.Y. Seidel and T.S. Rappaport. "Radio Channel Models in Manufacturing Environments." *Wireless Information Networks Workshop*, Rutgers University, Rutgers, New Jersey, June 15, 1989.
- [Seidel91] S.Y. Seidel and T.S. Rappaport. "914 MHz Path-Loss Prediction Models for Indoor Wireless Communications in Multi-floored Buildings." Submitted to *IEEE Transactions on Antennas and Propagation*, May 17, 1991.
- [Seven91] J. Sevenhans et al. "An Integrated Si-Bipolar RF Transceiver for a Zero-IF 900 MHz GSM Digital Mobile Radio Front-end of a Hand Portable Phone." *IEEE 1991 Custom Integrated Circuits Conference*, San Diego, CA, May 12-15, 1991.
- [Sheng91] S. Sheng. *Wideband Digital Portable Communications: A System Design.* M.S. Thesis, Memorandum No. UCB/ERL M91/108, U.C. Berkeley, Berkeley, CA, December 3, 1991.
- [Sheng92] S. Sheng, A. P. Chandrakasan, and R. W. Brodersen, "A Portable Multimedia Terminal." *IEEE Communications Magazine*, pp. 64-75. December 1992.
- [Sheng94] S. Sheng, et al. "A Monolithic CMOS Radio System for Wideband CDMA Communications." *Proceedings of the Wireless' 94 Conference*, June 1994.
- [Simon85] M.K. Simon, et al. *Spread-Spectrum Communications, Vol. 1-3.* New York: Computer Science Press Inc., 1985.
- [Stir87] R.C. Stirling. *Microwave Frequency Synthesizers.* New Jersey: Prentice-Hall Inc., 1987.

- [Stone95] K. Stone. *Low Power Spread Spectrum Demodulator for Wideband Wireless Communications*. M.S. Thesis, U.C. Berkeley, Berkeley, CA, 1995.
- [Sven90] L. Svennson. *Implementation Aspects of Decision-Feedback Equalizers for Mobile Telephones*. Ph.D. Dissertation, Department of Applied Electronics, University of Lund, Lund, Sweden, May 1990.
- [Symb95] Symbol Technologies, Inc. *Spectrum 24 Wireless Modem Product Press Release*. Holtsville, NY, June 12, 1995.
- [Teus94] C. Teuscher. *Software Simulation for the Infopad Downlink*. M.S. Thesis, U.C. Berkeley, Berkeley, CA, December 1994.
- [Trew90] G. Trewitt. *Local-Area Internetworks: Measurements and Analysis*. NSL Research Report RR-1, Digital Equipment Corp. Network Systems Laboratory, Palo Alto, CA. March 1990.
- [Uehara93] G. Uehara. *Circuit Techniques and Considerations for Implementation of High-Speed CMOS Analog-to-Digital Interfaces for DSP Based PRML Magnetic Disk Read Channels*. Ph.D. Thesis, U.C. Berkeley, Berkeley, CA, 1993.
- [Unger87] G. Ungerboeck. "Trellis-Coded Modulation with Redundant Signal Sets, Parts I and II." *IEEE Communications Magazine*, Vol. 25, No. 2, February 1987.
- [Vendel90] G.D. Vendelin, A.M. Pavio, and U.L. Rohde. *Microwave Circuit Design Using Linear and Nonlinear Techniques*. New York: John Wiley and Sons, 1990.
- [Viter95] A.J. Viterbi. *CDMA: Principles of Spread Spectrum Communications*. Reading, MA: Addison-Wesley Pub. Co., 1995.
- [Walsh73] *Theory and applications of Walsh functions and other non-sinusoidal functions*. Proceedings of a colloquium held on June 28th and 29th at the Hatfield Polytechnic.
- [Watts89] R.K. Watts. *Submicron Integrated Circuits*. New York: John Wiley and Sons, 1989.
- [Wong91] B. Wong and H. Samueli. "A 200 MHz All-Digital QAM Modulator and Demodulator in 1.2um CMOS for Digital Radio Applications." *IEEE Journal of Solid-State Circuits*, Vol. 26, No. 12, pp. 1970-1980. December 1991.
- [Wu95] Y.Y. Wu and W.Y. Zou. "Orthogonal Frequency Division Multiplexing- A Multi-Carrier Modulation Scheme." *IEEE Transactions on Consumer Electronics*, Vol. 41, No. 3, pp. 392-399. August 1995.
- [Yee96] D. Yee. *The Design and Implementation of a Semi-Custom Transmitter for a CDMA Direct Sequence Spread-Spectrum Transceiver*, M.S. Thesis, U.C. Berkeley, Berkeley, CA, December 1996.
- [Yin92] G. Yin, F. Op't Eynde, and W. Sansen. "A High-Speed CMOS Comparator with 8-b resolution." *IEEE Journal of Solid-State Circuits*, Vol. 27, No. 2, pp. 208-211. February 1992.

- [Yun94] L. Yun and D. Messerschmitt. "Power Control for Variable QOS on a CDMA Channel." *Proceedings of IEEE MILCOM*, Fort Monmouth, NJ, pp. 178-182. October 2-5, 1994.
- [ZComm95] Z-Communications, Inc. *Application Note AN-106: Using Z-COMM Sub-Miniature VCO's for WLAN Systems*. San Diego, CA. 1995.
- [Zog95] T. Zogakis, J. Aslanis, and J.M. Cioffi. "A Coded and Shaped Discrete Multitone System." *IEEE Transactions on Communications*, Vol. 43, No. 12, pp. 2941-2949. Dec. 1995.