Coding Optimization and Nonlinear Receiver Analysis for Wireless Communications Systems in the Presence of Interference

by

Troy Christopher Nolan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
at The University of Michigan
2001

Doctoral Committee:
Professor Wayne E. Stark, Chair
Research Scientist Jack R. East
Professor Clark T.-C. Nguyen
Professor Kim A. Winick

ABSTRACT

Coding Optimization and Nonlinear Receiver Analysis for Wireless Communications Systems in the Presence of Interference

by
Troy Christopher Nolan

Chair: Wayne E. Stark

In our work we analyze many different situations where the transmission of information is corrupted by noise. In communications systems noise can be naturally occurring, artificially generated, or self-generated. It is the job of communications engineers to combat the effects of noise on their designs. In this thesis we examine two main areas of noise mitigation: channel coding and filtering.

In the channel coding realm, we examine optimum code rates for a product code in the presence of AWGN and intelligent partial band jamming interference, and we provide asymptotic analysis of Reed-Solomon based product codes. The comparison of the optimal code rates for finite length product codes with the rates determined by asymptotic analysis illustrate the importance of considering undetected error in the code design. We develop a method of determining the probability of undetected error for the underlying Reed-Solomon codes that offers attractive advantages over known methods.

In the realm of noise mitigation by filtering, we consider MEMS filters which show great promise in achieving very high Q's at passband frequencies. We examine the performance of sub-sampled systems and systems with cosite interference which

employ filters of varying Q. We quantify the tradeoff between noise performance and required Q for these systems. Additionally we address the issues of power added efficiency (PAE) in systems that employ high Q filters and nonlinear amplifiers. We show that the application of a MEMS filter to a fixed bias power amplifier increases the PAE in the presence of strong interferers, but that there is not a clear link between lowering the bias power and increases in PAE.

For the analysis of nonlinear systems, we motivate and propose a new method called the modified instantaneous quadrature method (MIQM). With this method we can analyze, rather than simulate, the effects of a nonlinear element in weakly and strongly nonlinear systems with large or small dynamic range inputs. Once the system is known to be memoryless, or have a limited degree of memory, the MIQM will produce accurate results that match the output of more numerically complex simulation systems.

To Hyatt, whom I love with all that I am.

ACKNOWLEDGEMENTS

Where to begin. The animal that is research has good days and bad days. Some days a researcher feels like the world champion lion tamer, other days like the unlucky villager that has stumbled across a hungry lioness. Such is the nature of the beast. I have learned a great deal about myself in these past few years, some things that I liked, some that were hard to face. The lessons learned will stay with me for the remainder of my life.

First and foremost I wish to thank the Lord for his guidance and strength in my weak and difficult moments. At times, His was the only source of comfort and peace in lonely and frustrating periods. I truly believe that the Peace and Joy that is the Lord is there for the taking. Reach out for it.

On a more conventional note, I wish to thank my advisor and committee members for their guidance along the way. Discussions with these Professors has proved insightful in many situations and challenging in other situations; they always kept me honest.

The work presented in this thesis was supported in part by The United States Army under grants DOD-G-DAAH04-96-1-0001 and DOD-G-DAAH04-96-1-0377, "Low Power/Low Noise Electronics" and "Low Energy Electronics Design for Mobile Platforms" respectively. Additional support was received from from the United States Department of Defense Graduate Research Fellowship program and Raytheon Cor-

poration.

To Dr. Kurt Metzger goes a special thank you. Discussions with Kurt and resources that he was able to magically produce at just the right instant proved invaluable in the solution to one of the most difficult problems I encountered. This solution eventually paved the way to the Modified Instantaneous Quadrature Method of nonlinear analysis developed in this Thesis.

To my Parents who have always believed in me, thank you is not enough. I hope that when I become a parent I am able to convey the unconditional support that you have given me over the years. To my brother, the who will always be the first Dr. Nolan; it wasn't as much in the Ph.D. years, but healthy competition in the past and your friendship recently has meant a great deal to me. To Sr. Veronica of Jesus, my sister, whom I miss dearly, may the peace of Christ stay with you your whole life.

To my friends that I have met in Michigan, far to numerous to name here, thanks for the fun times and for the shoulder that only friends can give. Thanks for slapping me around when necessary, and building me back up when I needed it. We had some great times at the U of M, let's make sure as we spread all over the country and world that we always stay in touch. Go Tractors!

Finally, during my tenure at N.C. State University and at the University of Michigan, I courted the woman of my dreams. I was and will always be most honored that she became my wife this past year. Hyatt, I love you with all my heart. I look forward to being with you for the rest of my life, to all the crazy adventures we have, to our deep discussions, to our children that are yet a twinkle in your eye, to being your lifelong friend, and to growing old with you. Hyatt, my all forever belongs to you.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF APPENDICES

APPENDIX

CHAPTER 1
Introduction

The transmission of information from a source to a destination is the goal of communication. Amongst the many desires in communications engineering are the desires to achieve this goal with minimal power expenditure and with as small of an error rate as possible. This thesis addresses some of the issues involved in meeting these two desires.

To examine the transmission from source to destination, let us assume that a source generates a single numeric value. This value represents the state of the source at a given time instant. If the value can take on one of an infinite set of possibilities, the source is known as an analog source. If, however, the value generated is one of a finite number of possibilities the source is known as a digital source. It is the objective of the communications system to transmit this value to the destination in such a manner that the destination is able to reliably determine what the original value was.

Analog communications systems attempt to communicate an analog value to the receiver, and digital communications systems are concerned with the transmission of a digital value. Analog sources can be converted into digital sources by the use of a source coder. In one possible use, a source coder attempts to represent the infinite

set of analog values with a finite set of digital values, or a digital alphabet. Once this digital alphabet is known to the transmitter and receiver, the source coded digital value is transmitted to the destination. The error introduced by the use of a digital value to represent an analog value is a well studied field, but one which we will not concern ourselves with at this time.

We focus our efforts on digital communications systems. In a simple wireless digital transmission one of a finite set of signals is transmitted from the source antenna to the destination antenna. As the signal propagates from antenna to antenna through the communications channel, it may be interfered with by various natural and artificial effects. This interference distorts the transmitted signal so that the received signal is not equal to any of the set of potentially transmitted signals. Management and mitigation of interference is a concern of communications engineers and scientists.

At this point, we can draw a simple block diagram of the system, as shown in Figure 1.1. The flow through this system is very easy to follow: information is generated from the digital source, it is fed into the device that selects from a set of signals to transmit (called the modulator), the signal is amplified and transmitted through the channel. The signal is possibly corrupted in the channel, received by the receiving antenna, and amplified. The receiver operates on it in such a manner that a decision is made as to what signal was sent (demodulator). At this point the destination makes its best estimate as to what the original information was. An error is made if the estimate at the destination does not match the generated value at the source.

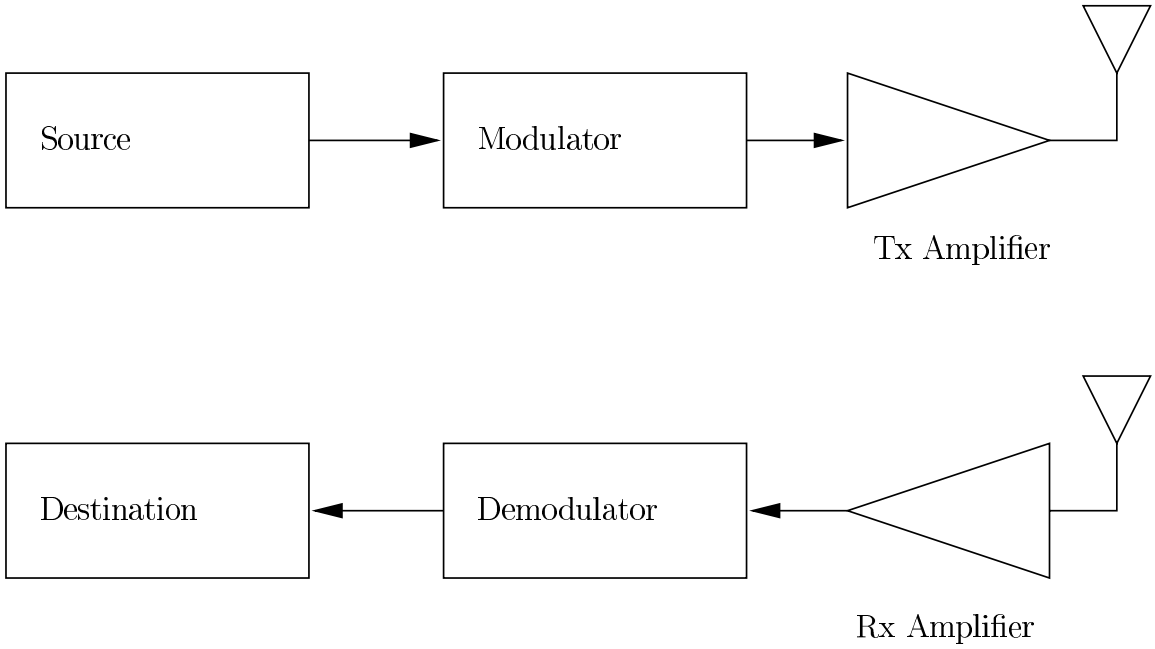


Figure 1.1: Simple communications system block diagram.

1.1 Types of Interference

1.1.1 Environmental Effects

There are many different types of interference that can corrupt information as it propagates from the source to the destination. One type of naturally occurring interference can add random noise to the transmitted signal. The most commonly analyzed natural noise is thermal noise or additive white Gaussian noise (AWGN). This noise is generally caused by the random thermal motion of molecules. The probability density function of a single observation of a band-limited AWGN process is:

$$f_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1.1}$$

where μ is the mean value and σ^2 is the variance. Assuming the mean is zero, the variance of the noise is a good measure of how corrupted a signal can become as it

effected by AWGN.

Another common variety of naturally occurring interference is caused by the propagating signal reflecting off of various objects as it travels from the transmit antenna to the receive antenna. The addition of randomly attenuated and delayed copies of the same signal can cause the envelope of the received signal to fluctuate. The random fluctuation of a single observation under certain circumstances can be shown to follow a Rayleigh density function:

$$f_y(y) = \frac{y}{\sigma^2} \mathrm{e}^{-\frac{y^2}{2\sigma^2}}, \qquad y \geq 0, \tag{1.2}$$

where $2\sigma^2$ is the variance of the random fluctuation.

1.1.2 Jamming Interference

Interference can also be caused by artificial sources, either unintended or hostile. Two types of jamming that we utilize in this thesis are continuous wave (CW) interference and partial band jamming (PBJ). Continuous wave interference, otherwise known as tone jamming, is the addition of power at a single frequency in such a manner that it interferes with the correct reception of the transmitted signal. The time domain waveform of a CW interferer is merely:

$$s(t) = A \cos 2\pi f t, \tag{1.3}$$

where the power in the tone is $\frac{A^2}{2}$ and the frequency is f. This type of interference can cause many different types of effects in receivers. Two such possible effects are: the tone can be located in-band, to directly corrupt the transmitted signal, or it can be located out of band which can cause intermodulation interference.

Partial band jamming is a technique where the jammer adds band-limited AWGN noise power to a fraction, ρ, of a given range of frequencies. If ρ is large the jammer

spreads his power thinly over a large portion of the band; if ρ is small he concentrates all his power into a much smaller band. The amount of noise power density added to the effected band is $\frac{N_J}{2\rho}$, where $\frac{N_J}{2}$ is the variance of the added Gaussian noise.

1.1.3 Interference Added by Circuit Elements

The last source of interference considered in this thesis is generated or amplified by the circuitry in the receiver. Virtually any operation on the received signal adds interference to the signal. Two types of interference will be described: thermal noise and nonlinear mixing interference. The noise figure (NF) of an amplifier is defined as the ratio of input signal power divided by thermal noise power divided by the ratio of the output signal power divided by the output thermal noise power. In the absence of signal gain it can be thought of as noise gain. The primary place of concern for the noise figure is in the very first elements that the received signal encounters i.e. the antenna, front end filters, and the low noise amplifier. This measure reflects the quality of signal that will result from the operation of the amplifier.

Nonlinear mixing noise, or more properly called intermodulation distortion, occurs when a nonlinear amplifier operates on a signal in the presence of interference. For a given single input circuit with response

$$y(t) = \mathrm{h}(x(t)), \tag{1.4}$$

given two tones, $x_1(t) = A_1 \cos(2\pi f_1 t)$ and $x_2(t) = A_2 \cos(2\pi f_2 t)$, the output due to each signal independently is:

$$y_1(t) = \mathrm{h}(x_1(t)) \tag{1.5}$$

$$y_2(t) = \mathrm{h}(x_2(t)). \tag{1.6}$$

A circuit is deemed linear if the output of the system due to the linear combination

of the two input signals is equal to the linear combination of the outputs due to each input alone, or:

$$\alpha \mathrm{h}(x_1(t)) + \beta \mathrm{h}(x_2(t)) = \mathrm{h}(\alpha x_1(t) + \beta x_2(t)), \tag{1.7}$$

where α and β are arbitrary scaling factors. However, if the input-output relationship of the circuit can be modeled by:

$$\mathrm{h}(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots, \tag{1.8}$$

with nonzero coefficients above the a_1 coefficient, then it fails the linearity test. In fact, a component of the output signal known as the third order intermodulation (IM) product contains the following term:

$$\mathrm{IM}_3(t) = A_1^2 A_2 \cos 2\pi(2f_1 - f_2)t. \tag{1.9}$$

If the desired signal has a component of frequency $2f_1 - f_2$ it will be corrupted by $\mathrm{IM}_3(t)$.

1.1.4 Noise and Interference Measures

Of the many ways to quantify the fidelity of a given signal, we focus on three commonly used definitions. The first is the well used received signal-to-noise ratio (SNR). SNR is defined by:

$$\mathrm{SNR} = \frac{P_s}{B\frac{N_0}{2}}, \tag{1.10}$$

where P_s is the power in the received signal, $\frac{N_0}{2}$ is the environmental thermal noise power density, and B is the bandwidth under consideration. Our use of the received SNR measurement accounts only for naturally occurring noise.

The second noise measure we define is the signal-to-interference ratio (SIR) at the output of an amplifier. The SIR is the ratio between the amplified received signal power and the self generated interference power. It is defined as follows:

$$\text{SIR} = \frac{\beta P_s}{I}, \tag{1.11}$$

where β is the gain of the amplifier circuit and I is the power of the intermodulation distortion.

The final measure that we will use is the signal-to-noise-plus-interference ratio (SNIR) at the output of an amplifier, which is defined as:

$$\text{SNIR} = \frac{\beta P_s}{\beta \text{NF} \left(\frac{N_0}{2} + \frac{N_J}{2\rho} \right) B + I}, \tag{1.12}$$

where $\frac{N_J}{2\rho}$ is the added noise power density of a possible PBJ jammer and NF is the noise figure of the circuit.

1.2 Methods to Combat the Effects of Interference

A communications engineer seeks to combat the effects of interference in the transmission of information. Once a proper determination of the type of interference expected in the system is made, different approaches may be used to mitigate its effects. The two methods that are focused on in this thesis are 1) channel coding and 2) filtering.

1.2.1 Channel Coding

There are various methods of reducing the probability of error in a communications system. A common technique is to add mathematical redundancy to the data, called channel encoding, before it reaches the modulator (see Figure 1.2). This redundancy

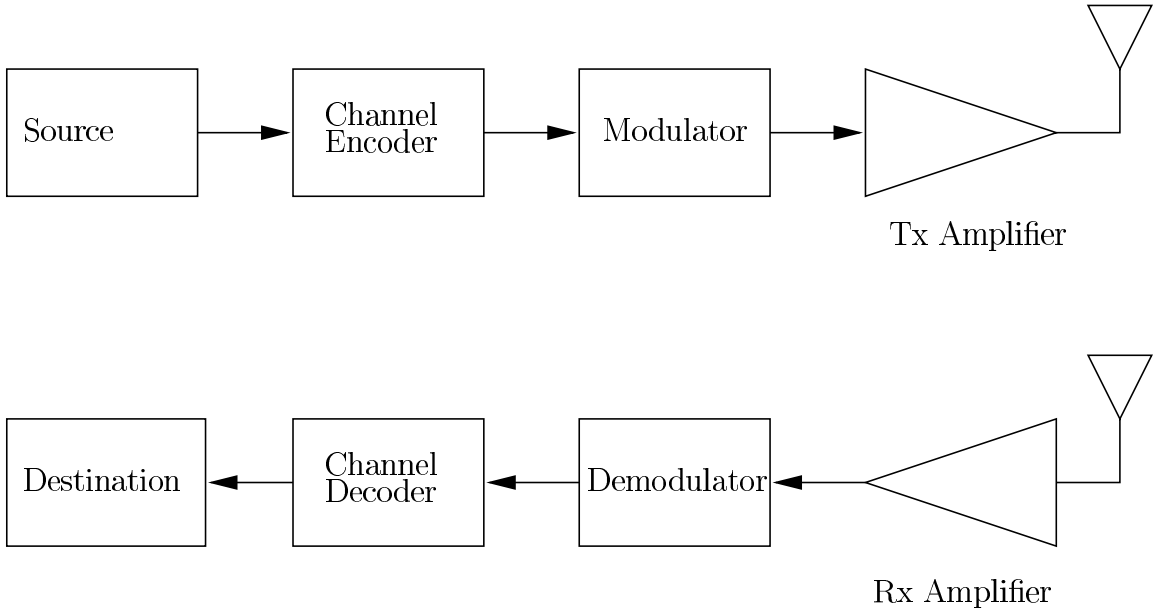


Figure 1.2: Simple communications system block diagram, with channel coder.

is extra data that must be transmitted to the destination, but it is derived from the original data in a structured fashion that is known to the receiver. Its primary purpose is to combat the effects of channel noise on the decision device at the destination. One common type of channel code is called a block code. Block codes group K data symbols together and add $N-K$ additional coding symbols. The rate of this code is defined as $\frac{K}{N}$ information symbols per channel symbol. The N transmitted symbols are known as a codeword and can be thought of as a vector in an N dimensional space. When all N, possibly corrupted symbols, are received they form a received vector. Then the decoder attempts to determine which K symbols were intended to be communicated.

Reed-Solomon (RS) codes are commonly used block codes. If the size of the alphabet is M, a power of 2, then a standard RS code can be constructed with length $N = M-1$ symbols. The minimum Hamming distance (the minimum number of positions in which two codewords differ) for RS codes is $d_{min} = N - K + 1$.

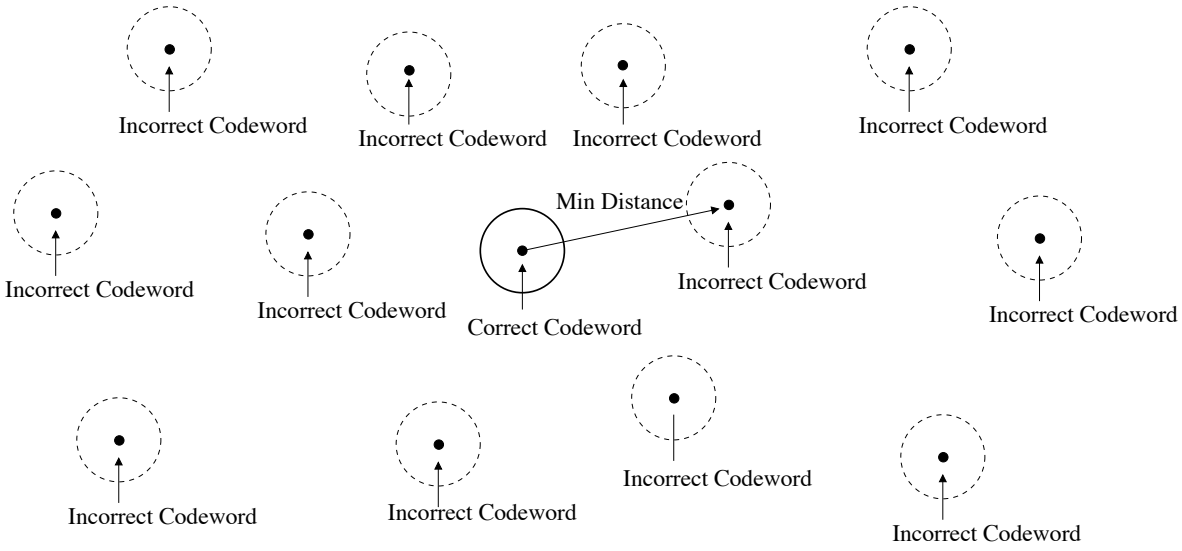


Figure 1.3: Illustration of the distances between codewords and decoding spheres for bounded distance decoding.

Decoding is typically accomplished by using a bounded distance decoder (BD) such as the Berlekamp-Massey decoder. An errors-only BD decoder maps all received vectors within a N-dimensional sphere of radius $\lfloor \frac{d_{min}}{2} \rfloor$, around each codeword, to that codeword. An illustration of this is shown in Figure 1.3. If a received vector lies outside of all codeword spheres, the decoder can determine this and a decoding failure is declared. If a transmitted codeword is decoded incorrectly, i.e. the received vector lies within the decoding radius of a codeword other than the transmitted codeword, a decoding error occurs. Without the use of an additional information, these decoding errors are *undetected*. Calculating the probability of undetected error in a RS code has been addressed by various authors, especially in the area of re-transmission schemes, but usually these algorithms are of exponential complexity in N. Even the best algorithm we encountered [1] has exponential complexity when each transmitted symbol has different noise statistics.

In Chapter 2 we motivate and develop an efficient, recursive method for calculating

the probability of undetected codeword error. Our algorithm assumes independent noise statistics from symbol to symbol but allows different noise distributions. The efficiency of the algorithm is dependent on the number of different noise distributions present, and can have exponential complexity in N as the number of noise distributions approaches N. In this situation, our result can be bounded in a fashion that is compute-limited (the bound grows monotonically tighter as the amount of CPU cycles applied increases). While the algorithm in [1] may approach our efficiency for the case with few noise distributions, it does not offer a bound in the case where there are many distributions.

1.2.2 Filtering

The second method of combating interference in a wireless system that we address, is the use of filters. Using filters to mitigate the effects of interference is useful when interfering signals have some type of frequency separation from the desired signal. A basic filter will attenuate some signals and allow other signals to pass unattenuated. These filters, when placed strategically, can effect the overall performance of the system. A common measure to describe the selectivity of filters is Q, or the quality factor. The Q of a filter is defined as:

$$Q = \frac{f_c}{B}, \tag{1.13}$$

where f_c is the center frequency of the filter and B is the bandwidth of the filter.

Filters of high Q are especially difficult to manufacture for high values of f_c. Recent advances in micro electro-mechanical systems (MEMS) have made possible filters of high center frequency, with very large Q values. In Chapter 4 we analyze three applications of MEMS filters in communications systems: sub-sampling, cosite

interference, and power efficiency. The basic receiver that we analyze in the cosite and power efficiency applications, and that we compare the sub-sampling architecture to is shown in Figure 1.4. In this front-end, the received signal is filtered by a channel

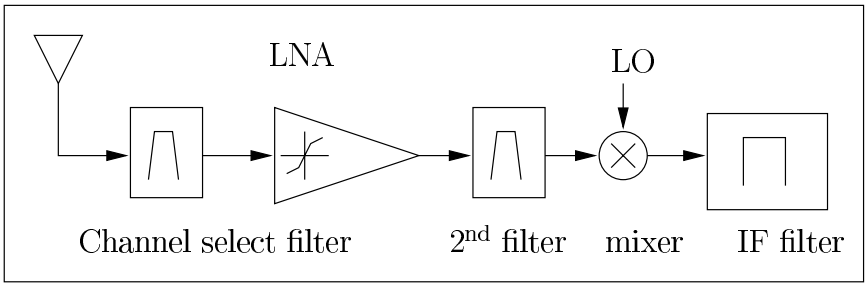


Figure 1.4: System Block Diagram.

select filter, amplified by a potentially nonlinear amplifier, filtered again, and mixed down to an intermediate frequency (IF). The goal is to perform these operations with a minimum of added noise or interference.

For sub-sampling systems, the challenge is mitigating the effects of replicated noise, while retaining as much of the desired signal as possible. A sub-sampling system seeks to eliminate the local oscillator (LO) and mixer by using a technique which creates a sampling-induced spectral replicate at the IF frequency. It is commonly known that sampling at a frequency below the Nyquist rate for low-pass signals induces spectral overlap, but bandpass sampling *below* the Nyquist rate can produce spectral replicates at advantageous frequencies. If proper system parameters are set, the mixer and LO found in the traditional radio front-end shown in Figure 1.4 can be eliminated. We give conditions on the sub-sampling system parameters to assure self-interference free communication and we show that the replicated sampling resistor noise is dominated by the thermal noise at the antenna. Additionally, we show that under another condition, the replicated wide-band noise is finite. We apply this analysis to an example system and show the performance tradeoff between sampling

rate, filter shape, and interference.

The second issue that we explore in Chapter 4 is the analysis of the suppression of cosite interference. Cosite interference occurs when one or more strong interfering signals are spectrally close to a weak desired signal. If the channel select filter is unable to adequately attenuate the interferers and the LNA has a nonlinear response, then intermodulation interference may be generated in-band. To combat this interference, MEMS filters may be employed as the channel select filter. For a system with this type of interference, we quantify the tradeoff between filter Q and SIR.

The third issue that we address is the power savings that may be achieved by using MEMS filters in nonlinear receiver front-ends. It was thought that the use of selective filters may allow the use of amplifiers that are more power efficient and at the same time less linear. By applying a MEMS channel select filter to attenuate strong interferers in a cosite scenario, intuition held that the overall system SIR could be held constant while the power used by the system decreased. The measure of power added efficiency (PAE) proposed to examine this situation is defined as:

$$\text{PAE} = \frac{P_{out} - P_{in}}{P_{sup}}, \tag{1.14}$$

where P_{out} is the power of the desired signal at the output of the amplifier, P_{in} is the input power of the desired signal, and P_{sup} is the amount of power supplied to the nonlinear amplifier. We show that the application of MEMS filters to a fixed bias power amplifier increases the PAE in the presence of strong interferers, but that there is not a clear link between lowering the amplifier bias bias power and increases in PAE. The PAE of a given set of amplifiers is both dependent on the amount of gain available and the amount of power supplied. When the power supplied decreases the gain usually decreases as well, and the rates at which these two measures change is heavily dependent on the circuit design.

1.3 Nonlinear analysis technique

In order to properly evaluate the effects of a nonlinear amplifier on the interference level in a given system, a nonlinear analysis model is generally employed. The commonly used nonlinear modeling techniques are: circuit simulation, behavioral modeling, and analytical modeling. An overview of the strengths and weaknesses of these techniques is found in Chapter 3.

To analyze, rather than simulate, the effects of a nonlinear amplifier in the systems that we examine, the Modified Instantaneous Quadrature Method (MIQM) is proposed. Once certain limitations are addressed, this method is shown to have advantages over the existing modeling techniques discussed. Most important amongst the advantages discussed in Chapter 3, is that it allows for rapid analysis of weakly and strongly nonlinear systems with large or small dynamic range inputs. Once the system is shown to be memoryless, or have a limited degree of memory, the MIQM will produce accurate results that match the output of more numerically complex simulation systems.

1.4 Organization of Thesis

In Chapter 2 the selection of optimal RS coding rates for a product code in the presence of partial band interference is addressed. Development and use of a computationally efficient method for determining the probability of undetected error is made to select code rates that achieve a given design goal. These codes are compared to the asymptoticly derived optimum codes and it is shown that undetected error is an important consideration in rate selection.

Chapter 3 details the background and design of the MIQM technique for the

analysis of nonlinear systems. The results from the MIQM analysis of a baseline amplifier circuit are found to compare favorably to the results generated by a circuit simulator. This technique is employed to analyze the use of MEMS filters in the communications systems studied in Chapter 4.

Finally, a summary of the original contributions and conclusions of our work is found in Chapter 5.

CHAPTER 2

Reed-Solomon Codes: Probability of Undetected Error and Applications

It is well known that coding is an essential element of a well designed digital communication system. In some cases, the additional structure imposed by coding can lower the required signal power by as much as 30 to 40dB [2]. Specifically, in systems where parts of the transmitted codewords experience interference or fading, coding can provide a mechanism to combat these effects. Many systems utilize Reed-Solomon (RS) codes as part of their their error control structure. While the Reed-Solomon code provides an attractive error control element due to its distance and burst error control properties, analysis of the probability of undetected error for these codes can be complex.

In this chapter we present an efficient algorithm for calculating the probabilities of correct decode, incorrect decode, and decode failure for an RS code which is decoded using a bounded distance (BD) decoder. We start with the assumption that each symbol in the codeword has identical interference statistics, and we extend the analysis and algorithm to the case where each symbol may experience different interference statistics. Using this analysis we determine the probability of undetected error for a single RS codeword transmitted through a number of common channels.

Additionally, we evaluate the correct, incorrect, and failure statistics for a frequency-

hopped spread-spectrum system where a RS based product code (PC) is transmitted in the partial band jamming (PBJ) channel. A product code consists of two codes: A row code and a column code. These codes provide large error correcting capability with low decoding complexity. We consider the case where the row code is transmitted on a single frequency of a frequency-hopped system thus each element of the column code is transmitted over a different frequency. The purpose of the row code is to provide correction against random errors and detection of bursts of errors. Thus the decoder may attempt to correct a small number of errors but detect a large number of errors. The column code corrects the erasures due to the row decoder's error detection. We determine optimal rates for the constituent row and column codes and we compare these rates with the rates determined through asymptotic analysis [3].

The remainder of this chapter is organized as follows: In Section 2.1 we provide a brief overview of the literature, in Section 2.2 we illustrate the system model that we use for the case of the single RS codeword. Sections 2.3, 2.4, and 2.5 contain the main derivation of our efficient method, and Sections 2.6 and 2.7 detail the application of our method to some channels of interest. Finally Section 2.9 details our conclusions.

2.1 Work in the Literature

In the literature, authors have generally ignored the probability of incorrectly decoding a Reed-Solomon codeword, but there are a few that tackle this problem. In [4] an expression for the probability of error is derived for a RS code with independent identically distributed (IID) noise. Direct calculation of this expression is of exponential complexity. The problem of decoding in IID noise is also addressed in [5]. While this method is less complex it does not consider the case where symbols are corrupted

by independent but differently distributed (IDD) noise. In [6] and [7] bounds are given for the probability of undetected error for the IID case but no mention is made of IDD noise. Similarly in [8], results are derived combanitorically for the IID noise but not for IDD noise. In perhaps the closest parallel to our work, Daraiseh [1], uses zero patterns to combanitorically derive results where all symbols may experience noise of different distributions, but the given expressions grow with exponential complexity in the length of the code and the framework is less flexible than our result.

In the application of RS codes to product codes, there has been some activity. In [9] the development of "good" Hamming codes for use in a product code is explored. While there are interesting problems in this area, a Hamming code cannot provide the degree of codeword separation that a constituent RS codes can. In [10], Nakamura decodes a product code array by decoding row-wise, column-wise and row-wise again in the presence of randomly distributed errors. This repetition is interesting in light of the recent turbo code iterative methods, but they do not address the probability of undetected error. In a fairly recent work, [11], Yi proposes a diagonal interleaving scheme which serves to randomize potential burst errors. In a product code the strength of the row code is used to detect and possibly erase burst errors. Since interleaving attempts to eliminate burst errors it is not clear how Yi's scheme will impact the overall correcting power of the product code. In an older work, [2], Stark evaluates the performance of concatenated codes in the partial band jamming (PBJ) channel. The constituent codes discussed are an outer RS code and an inner single parity code. This work is helpful as a performance comparison, but we will change the model slightly. In addition, Frank [12] also considers concatenated coding in the PBJ channel. He uses both soft and hard decisions, this will provide another model for comparison.

2.2 System Model

In developing an algorithm for efficient computation of undetected error probability for Reed-Solomon codes we assume a singly extended (N, K) RS codeword with minimum distance $d_{min} = N-K+1$. A standard RS code has length $N = q-1$, where the symbols are drawn from $GF(q)$, and is known as a maximum distance separable (MDS) code because its minimum codeword distance satisfies $d_{min} = N - K + 1$. A singly extended RS code is also drawn from the $GF(q)$ but it has length $N = q$, and it also satisfies the MDS criteria. This codeword is transmitted using orthogonal modulation across an M-ary symmetric channel with errors and erasures. The receiver demodulates and makes symbol-by-symbol decisions as to what symbol was transmitted. Interference in the channel results in a non-zero probability that the receiver will decide what symbol was transmitted in error. In addition, the receiver may declare a symbol erasure if some condition is satisfied (such as the ratio threshold test (RTT)). For the l'th symbol transmitted over this channel the probability that the receiver decides, in error, that *any* other symbol was transmitted is P_a^l, the probability that we error to a *specific* symbol is $P_s^l = \frac{P_a^l}{M-1}$, the probability of erasure is P_f^l, and the probability that the correct symbol is decided upon is P_c^l.

The receiver then attempts to decode the received vector into a codeword using a BD/RS/Errors and Erasures (EE) decoder. If the received vector has t errors and e erasures satisfying $2t + e < d_{min}$, the received vector decodes correctly. However, if the the received codeword lies within a d_{min} "radius" with respect to an incorrect codeword, the codeword decodes incorrectly. If neither of these conditions hold, the decoder declares a decoding failure. While it is common practice to assume the probability of undetected error is negligible, we wish to determine this probability. With out loss of generality we assume the all zeros codeword is the transmitted

codeword for all cases.

2.2.1 Zero Patterns

One further concept that is needed in our algorithm development is the concept of a "zero pattern". A zero pattern is a vector that flags the positions of zero symbols in a RS codeword. If we create a vector of length N that contains 0's where a RS codeword has zero symbols and 1's where the same RS codeword has non-zero symbols we have created a "zero pattern". In this manner, every codeword has a zero pattern, but not every zero pattern has a unique codeword. In fact, we know from the weight enumerator that there are

$$(N-1)\sum_{s=1}^{k-d_{min}}(-1)^s\binom{k-1}{s}N^{k-s-d_{min}} \tag{2.1}$$

codewords per weight k zero pattern [13]. Additionally we know that there are $\binom{N}{k}$ zero patterns for a weight k codeword [14].

2.2.2 Noise Distribution

We assume each symbol is transmitted in an M-ary symmetric channel with independent noise from symbol to symbol. For compactness of results we will assume the distribution of noise has the potential to change from symbol to symbol, and these distributions are known to us. The only case that requires a distinction is the probability of incorrect decoding; in this case if the number of different distributions is large (compared with N) the IDD noise case yields bounds, otherwise the IID case and IDD case with few distinct noise distributions yield exact results with reasonable complexity.

2.3 Correct Decoding

To determine the probability of correctly decoding a received vector, we define the event $B_{i,j}^{l}$ to be the event that i erasures and j errors have occurred in l transmitted symbols, and we build the density function for $B_{i,j}^{N}$ starting with $P\{B_{0,0}^{1}\} = P_c^1$, $P\{B_{0,1}^{1}\} = P_f^1$, and $P\{B_{1,0}^{1}\} = P_a^1$ and proceeding with $l = 2 \ldots N$. The general term then becomes

$$\begin{aligned} P\{B_{i,j}^{l}\} &= [P\{B_{i,j}^{l-1}\} \times I_{\{l \neq i+j\}}]P_c^l + \\ & \quad [P\{B_{i-1,j}^{l-1}\} \times I_{\{i \neq 0\}}]P_f^l + \\ & \quad [P\{B_{i,j-1}^{l-1}\} \times I_{\{j \neq 0\}}]P_a^l, \end{aligned} \tag{2.2}$$

where $i = 0, 1, \ldots, l$ and $j = 0, 1, \ldots, l - i$, and $I_{\{m\}} = 1$ if the event m is true and $I_{\{m\}} = 0$ otherwise. Thus, the probability of correctly decoding is

$$P_{cd} = \sum_{i=0}^{N} \sum_{j=0}^{N-i} P\{B_{i,j}^{N}\} \times I_{\{2j+i<d_{min}\}}. \tag{2.3}$$

The worst case computational complexity of evaluating (2.3) is $O(N^2)$ when there are N different noise distributions present. If the probability of erasure is zero or the BD decoder decodes errors only (EO), (2.3) reduces to complexity $O(N)$, but if there is only one noise distribution the BD/EE decoder calculation is $O(N)$ and the BD/EO decoder calculation is trivial. (2.3) can be applied when the noise distribution is any arbitrary function, the only requirement being the knowledge of the symbol transition probabilities.

2.4 Incorrect Decoding

To determine the probability of incorrectly decoding the received vector we must split the discussion into 2 sections. In the first section we determine the exact prob-

ability of incorrectly decoding in the presence of IID noise, while the second section determines the bounds on the probability of incorrectly decoding in the presence of IDD noise.

2.4.1 IID

The probability of incorrectly decoding in the presence of IID noise can be constructed in a similar fashion to that of correctly decoding. For this analysis we define the event of interest in a slightly different manner. Let $B^l_{i,j|k}$ be the event that the received vector has i erasures and is a Hamming distance j (in the non-erased positions) from a weight k codeword after l symbols have been transmitted. In addition we define transition probabilities as

$$P^l_{c^*} = \begin{cases} P^l_s & \text{if } Z^l = 1 \\ P^l_c & \text{if } Z^l = 0, \end{cases} \tag{2.4}$$

$$P^l_{e^*} = \begin{cases} P^l_s(M-2) + P^l_c & \text{if } Z^l = 1 \\ P^l_e & \text{if } Z^l = 0, \end{cases} \tag{2.5}$$

and $P^l_{f^*} = 1 - P^l_{c^*} - P^l_{e^*} = P_f$; where Z is the zero pattern for any weight k codeword.

Proceeding as above, the initial terms are $P\{B^1_{0,0|k}\} = P^1_{c^*}$, $P\{B^1_{0,1|k}\} = P^1_{f^*}$,and $P\{B^1_{1,0|k}\} = P^1_{e^*}$. The general term is

$$\begin{aligned} P\{B^l_{i,j|k}\} &= [P\{B^{l-1}_{i,j|k}\} \times I_{\{l \neq i+j\}}]P^l_{c^*} + \\ &\quad [P\{B^{l-1}_{i-1,j|k}\} \times I_{\{i \neq 0\}}]P^l_{f^*} + \\ &\quad [P\{B^{l-1}_{i,j-1|k}\} \times I_{\{j \neq 0\}}]P^l_{e^*}, \end{aligned} \tag{2.6}$$

which is independent of the weight k word decoded. Thus, the probability of error to

any weight k codeword can be found by summing

$$P_{id,k} = A_k \sum_{j=0}^{\lfloor \frac{d_{min}-1}{2} \rfloor} \sum_{i=0}^{d_{min}-2j-1} P\{B^N_{i,j|k}\}, \tag{2.7}$$

where A_k is the weight enumerator for an MDS code [14]. Summing over the potential codeword weights removes the conditioning and completes the overall probability of codeword error calculation:

$$P_{id} = \sum_{k=d_{min}}^{N} P_{id,k}. \tag{2.8}$$

2.4.2 IDD

While the analysis from the IID case provides a straightforward exact probability of incorrectly decoding the received vector, the additional dimension of differently distributed noise raises some problems. In the previous analysis we constructed the transition probability of the channel based on a zero pattern of a representative weight k codeword. Unfortunately, the addition of IDD noise prevents us from selecting a single weight k codeword to be representative of all weight k codewords. Because the channel transitions are a function of *both* the location of the zeros in a codeword and the differing distribution of noise that is possible for each symbol, we must consider each weight k codeword individually. This presents a problem in that there are N_p combinations of zero positions for codewords of weight k [14] thus the number of codewords to consider grows exponentially. For a code of non-trivial length this complexity growth is unacceptable.

To address this problem we must examine the zero pattern distribution more closely. As a result of the different noise distributions, some zero patterns are more likely to occur than others. We use this knowledge and develop a list of representative zero patterns ordered by their likelihood, with the most likely pattern first. We denote

the l-th symbol in the ζ-th representative zero pattern Z_ζ^l.

We define the correct transition probability vector, P_C, to be a vector of length N whose elements contain the probability that the received symbol type (zero or non-zero symbol) matches the type of the corresponding symbol in the zero pattern of weight k under consideration. Thus P_C is constructed of elements P_C^l where

$$P_C^l = \begin{cases} P_c^l & \text{if } Z_\zeta^l = 0 \\ P_s^l & \text{if } Z_\zeta^l = 1 \ , \end{cases} \tag{2.9}$$

and $l = 1 \ldots N$. The erasure probability vector can be determined from the channel and is denoted:

$$P_F = [P_f^1 P_f^2 \ldots P_f^N]. \tag{2.10}$$

The error probability vector, P_E is constructed from the probability space that remains,

$$P_E^l = \begin{cases} 1 - P_f^l - P_c^l & \text{if } Z_\zeta^l = 0 \\ 1 - P_f^i - P_s^l & \text{if } Z_\zeta^l = 1 \ . \end{cases} \tag{2.11}$$

The event $B_{i,j|k}^{l,\zeta}$ is defined as the event that after l symbols are transmitted the received vector has i erasures and its associated zero pattern is a Hamming distance j (in the $N - i$ non-erased positions) from the ζ'th weight k zero pattern of interest. For the specific pattern, Z_ζ, the general term for the recursion is very similar to that of (2.6) and is given by

$$\begin{aligned} P\{B_{i,j|k}^{l,\zeta}\} \quad = \quad & [P\{B_{i,j|k}^{l-1,\zeta}\} \times I_{\{l \neq i+j\}}] P_C^l + \\ & [P\{B_{i-1,j|k}^{l-1,\zeta}\} \times I_{\{i \neq 0\}}] P_F^l + \\ & [P\{B_{i,j-1|k}^{l-1,\zeta}\} \times I_{\{j \neq 0\}}] P_E^l, \end{aligned} \tag{2.12}$$

with the corresponding initial values $P\{B_{0,0|k}^{1,\zeta}\} = P_C^1$, $P\{B_{0,1|k}^{1,\zeta}\} = P_F^1$,and $P\{B_{1,0|k}^{1,\zeta}\} = P_E^1$.

Next, the probability contribution of the ζ'th zero pattern is found by summing,

$$P_{ID,k}^{\zeta} = A_{\zeta} \sum_{j=0}^{\lfloor \frac{d_{min}-1}{2} \rfloor} \sum_{i=0}^{d_{min}-2j-1} P\{B_{i,j|k}^{N,\zeta}\}, \tag{2.13}$$

where A_{ζ} is the number of codewords with zero pattern Z_{ζ}, (2.1). Continuing, we write the unconditional probability of incorrect decoding as:

$$P_{ID} = \sum_{k=d_{min}}^{N} \sum_{\zeta=0}^{N_k} P_{ID,k}^{\zeta}. \tag{2.14}$$

2.4.3 Representative Zero Patterns

The remaining question is how to identify the representative zero patterns. Utilizing Stirling's approximation to $N!$, it can be clearly seen that the number of zero patterns grows exponentially in N. Because of the exponential growth of zero patterns bounding the result may be more desirable than considering all of the patterns. On first pass we may be tempted to select the fraction L most likely zero patterns that the IDD noise would precipitate. While effective in creating a bound [15], if a small number of noise distributions are present, as in the case of a two state channel, there are much more intelligent mechanisms. To frame the problem, let us assume t indexes different noise distributions $(f_1 \dots f_t)$. Each distribution effects ψ_m symbols from a (N, K) codeword. We define $\Psi = [\psi_1 \dots \psi_t]$. For a given weight, k, and zero pattern, ζ, we define X_k^{ζ} to be the product of the associated correct transition vector from (2.9):

$$X_k^{\zeta} = \prod_{l=1}^{N} P_C^l. \tag{2.15}$$

If we fix k and plot X_k^{ζ} for all ζ we observe that some of the zero patterns share the same likelihood of occurring. Figures 2.1, 2.2, and 2.3 are generated by sorting the likelihood of *all* weight 10 zero patterns for a (16,12) code. Figures 2.1 and 2.2 have

2 and 3 distinct noise distributions each affecting 2 symbols, and Figure 2.3 has 16 different noise distributions each affecting 1 symbol.

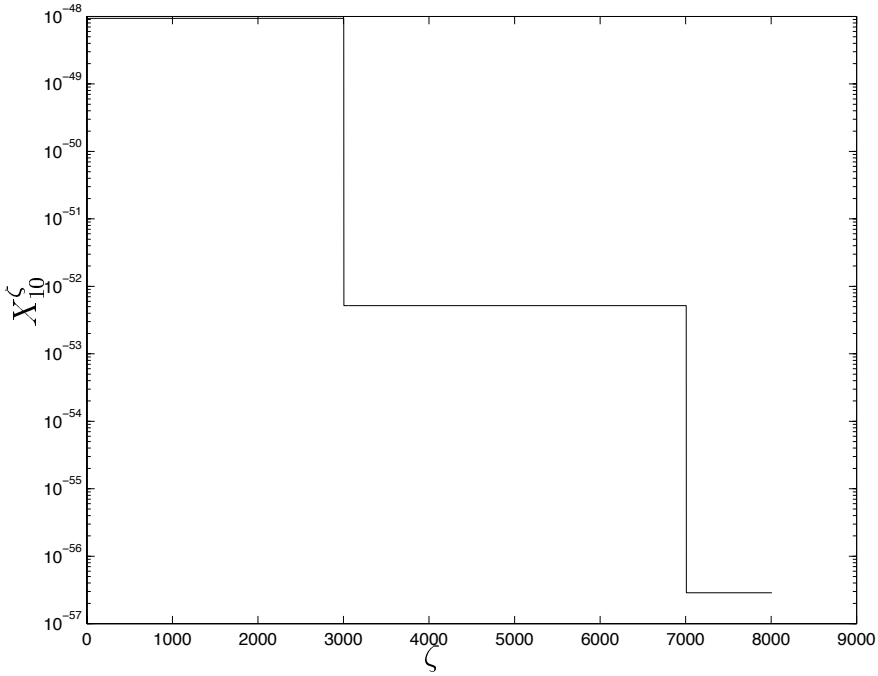


Figure 2.1: Two distinct noise distributions

When the number of distinct noise distributions is small there are zero patterns with the same likelihood of appearing, as the plateaus in Figures 2.1 and 2.2 illustrate. However, as the number of noise distributions approaches the length of the code, N, there are very few zero patterns with the same likelihood. We can see the mechanism that generates these plateaus by examining the received zero patterns more closely.

The device we construct to illustrate this in Figure 2.4 is a zero pattern of length $N = 16$ and weight $k = 13$. We subdivide this pattern into $t = 5$ partitions each containing ψ_m symbols ($m = 1 \ldots t$). We populate the pattern with $N - k = 3$ zero symbols, "0", and $k = 13$ non-zero symbols, "X". Next, we permute the location of the zero symbols such that every successive permutation causes a zero symbol to move across partition bounds. The resulting P_C^l changes, thus the value of (2.15)

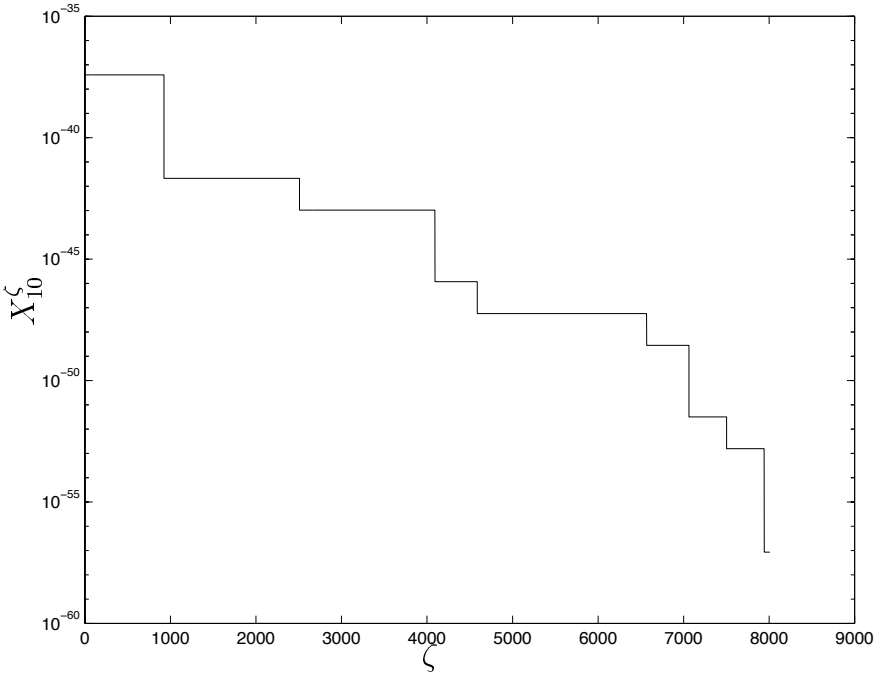


Figure 2.2: Three distinct noise distributions

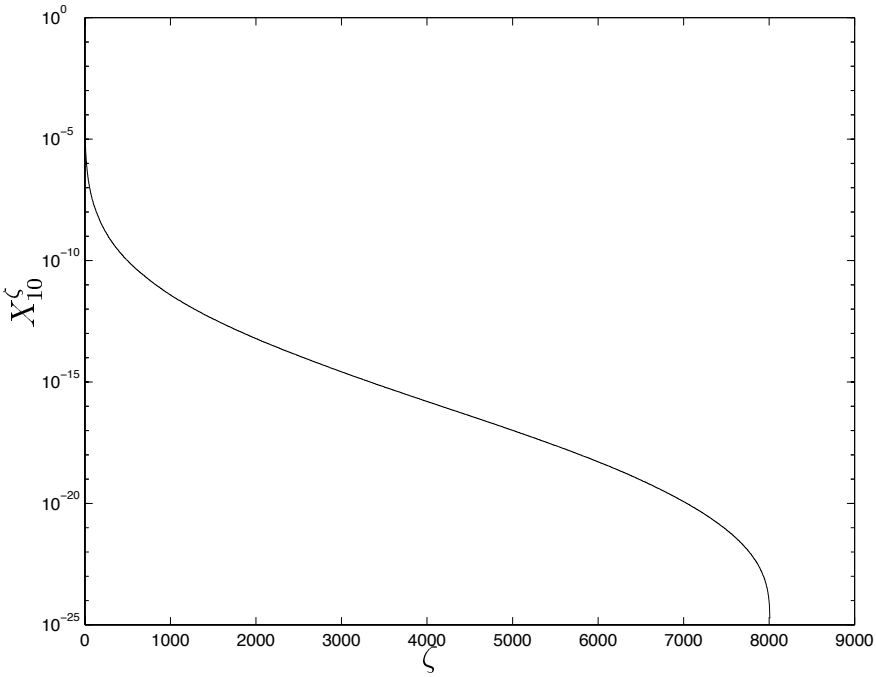


Figure 2.3: 16 distinct noise distributions

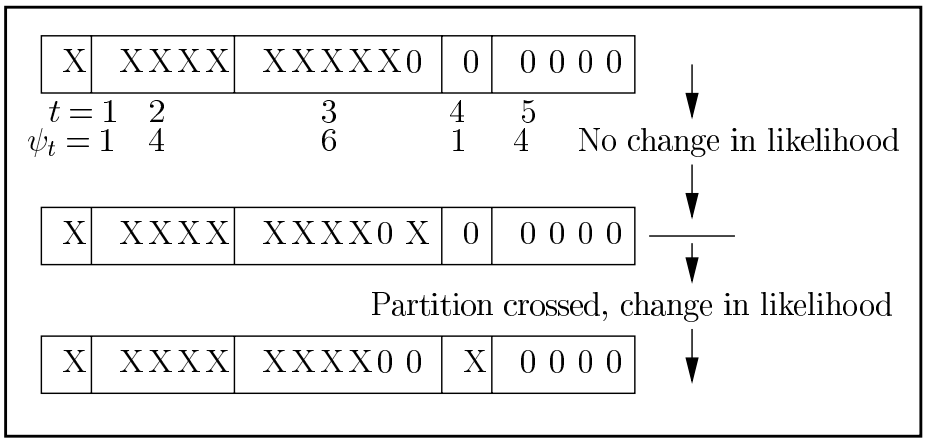


Figure 2.4: Plateau illustration device

changes.

We note that every zero pattern on a given plateau has the same likelihood of occurring. This can be realized by moving a zero symbol such that it does not cross a partition boundary. Using these zero pattern generation mechanisms, we can easily generate the set of all zero patterns corresponding to a single plateau or a set of zero patterns where each pattern is drawn from a unique plateau. Define a specific zero pattern that is drawn from a set of zero patterns with equal likelihood, a "plateau representative". We will define $Z^k = [Z_1^k \dots Z_{N_R}^k]$ to be the ordered set of plateau representatives which contain all representatives of weight k ordered by decreasing likelihood of occurrence, where N_R is the size of the set (calculated in Section 2.4.4).

The advantage of exploiting the plateaus is the ability to selectively evaluate representative zero patterns from each plateau. For instance, in a situation where 2 different noise distributions are present such as a basic PBJ channel, we would merely need to calculate statistics for a maximum of k weight k zero patterns. This provides a major decrease in complexity and accuracy as compared with our previous results in [16] which provided upper and lower bounds on the probability of incorrectly

decoding by choosing a fraction of the zero patterns without regard to the potentially equal likelihood.

2.4.4 Plateau Representative Combinatorics

To take the previous analysis a step further, we are interested in the number of plateau representatives and the number of zero patterns on a plateau for a given k and Ψ. In general there are

$$N_R = \sum_{x_1=0}^{\psi_1} \sum_{x_2=0}^{\psi_2} \cdots \sum_{x_t=0}^{\psi_t} I\left\{k + \sum_{y=1}^{t} x_y > \sum_{y=1}^{t} \psi_y\right\} \tag{2.16}$$

distinct plateaus, where $I\{x\} = 1$ if x is true and $I\{x\} = 0$ otherwise. While accurate, this calculation itself has exponential computational complexity. To reduce the complexity of this calculation, we generate a counting recursion. First we restate the problem as a counting game. Suppose we have k indistinguishable tennis balls and t buckets. Each bucket has a specific capacity ψ_m where $\sum_m \psi_m = N$ and the position of the tennis balls in a bucket is indistinguishable. The problem then becomes this: how many ways are there to put the k balls into the t buckets with the caveat that the buckets *may* be under-filled but *may not* be over-filled.

To solve this problem we define C_b^u to be the number of arrangements of b balls into u buckets, and we generate a recursion. To see the derivation of the recursion's general term consider the case where the u'th bucket is empty, there are C_b^{u-1} ways to distribute b balls into the $u-1$ buckets. If a ball is now placed in the u'th bucket, there are C_{b-1}^{u-1} ways of distributing the remaining $b-1$ balls into the $u-1$ buckets. The process of moving one ball from the $u-1$ buckets to the u'th bucket continues until the u'th bucket contains ψ_u balls or all the balls are in the u'th bucket. Thus, the general term is only dependent on the case where there are $u-1$ buckets.

This recursion can be calculated using an array, C, indexed by u and b such that b increases as we descend down the rows and u increases as we proceed to the right across the columns. We initialize the first column with 1's in the following positions: $C_0^1, C_1^1, C_2^1, ...C_{\psi_1}^1$ and zeros in the remaining positions. This is interpreted in the following manner: there is one different way to arrange $0, 1, 2 \ldots \psi_1$ balls in one bucket of size ψ_1 and there are no ways to arrange any more than ψ_1 balls. We now fill in the next column by proceeding down the rows with the general term:

$$C_b^u = \sum_{j=0}^{\min(\psi_u, b)} C_{b-j}^{u-1}. \tag{2.17}$$

Thus, the solution of the question of how many ways to arrange k balls (or non-zero symbols) in t buckets (or different noise statistics) is

$$N_R = C_k^t. \tag{2.18}$$

This solution has computational complexity $\mathrm{O}(N^2)$ and storage complexity $\mathrm{O}(N^2)$.

The number of elements on a given plateau can be determined once the plateau representative is known. Define k_i to be the number of non-zero symbols with noise distribution f_i in the plateau representative ζ. The number of different zero patterns that share this same likelihood is

$$N_p^\zeta = \prod_{i=1}^{t} \binom{\psi_i}{k_i}. \tag{2.19}$$

2.4.5 Bounding

In the case where t is much smaller than N, selecting one representative zero pattern per plateau and determining its contribution multiplied by the number of zero patterns that it represents, decreases the number of zero patterns over which we must sum. But as t approaches N, the number of plateaus and thus the number

of representatives, grows exponentially. If the complexity introduced by a growing t becomes unacceptable, bounding is an option. The tightest lower bounds using the plateau analysis comes from selecting the first fraction L of the ordered plateau representatives. Modifying (2.14) gives us

$$P_{ID} \geq P_{ID}^{LB} = \sum_{k=d_{min}}^{N} \sum_{\zeta=0}^{LN_R} N_p^{\zeta} P_{ID,k}^{\zeta}, \tag{2.20}$$

and the upper bound can be determined by summing over the first fraction L of the plateau representatives and treating the remaining fraction, $1-L$, plateau representatives as they contribute as much as the smallest evaluated plateau.

$$P_{ID} \leq P_{ID}^{UB} = \sum_{k=d_{min}}^{N} \left[\sum_{\zeta=0}^{LN_R} P_{ID,k}^{\zeta} N_p^{\zeta} + (1-L) N_R N_p^{LN_R} P_{id,k}^{LN_R} \right]. \tag{2.21}$$

In general, these bounds are compute limited, to make them tighter by increasing L more CPU time must be employed. While multiplying N_R by the fraction L certainly did not remove us from exponential complexity, it does grant the freedom to make the bounds as tight or as loose as the application demands.

2.4.6 Zero Pattern Sorting/Generation

Some mention should be made here regarding the ordering of Z^k, the set of plateau representatives. In Sections 2.4.3 and 2.4.5 we assume the zero patterns are ordered by their likelihood. For now, we generate a large number of highly likely plateau representatives, a factor of 10 times more than we require for the product LN_R and we sort these using a standard $\mathrm{O}(n \log n)$ sorting technique. It is desirable to develop an algorithm to generate these zero patterns in order, and this is an open problem described in this section.

Suppose we generate N real, not necessarily unique, numbers from the open interval $(0, 1)$. These numbers are then sorted into ascending order. We call the smallest

number of this sorted set P_1, the next larger P_2 and so forth, and we define the vector $P = [P_1, P_2, P_3, \ldots, P_N]$. In addition, we define Z_0 to be a binary vector of length N and weight k such that the rightmost k elements of Z_0 are "1"'s and the remaining elements are "0"'s. Next we define $\pi_0 = \prod_{x=1}^{N}(P_x I\{Z_x = 1\} + (1 - P_x)I\{Z_x = 0\})$, where $I\{x\} = 1$ if x is true and $I\{x\} = 0$ otherwise.

As an example we have the following for $k = 2$, $N = 5$:

$$P = [.1, .2, .3, .4, .5] \tag{2.22}$$

$$Z_0 = [0, 0, 0, 1, 1] \tag{2.23}$$

$$\pi_0 = (.9)(.8)(.7)(.4)(.5) = 0.1008 \tag{2.24}$$

The arrangement of Z_0 may now be permuted into any of $\binom{N}{k}$ possible patterns, thus creating Z_1. For instance:

$$Z_1 = [0, 0, 1, 0, 1] \tag{2.25}$$

$$\pi_1 = 0.0648 \tag{2.26}$$

The problem is this: How can the sequence of zero patterns, $Z_0, Z_1, \ldots, Z_{\binom{N}{k}}$, be reliably and efficiently generated such that the associated product sequence, $\pi_0, \pi_1, \ldots, \pi_{\binom{N}{k}}$, is monotonic non-increasing. While it is "easy" to do this with brute force searching, we are interested in developing an algorithm that does not need to enumerate all possible binary vectors before searching. While it is clear (and trivially provable) that the above selection of Z_1 is the correct choice, it is not so clear which is the correct choice for Z_2 without trying different possibilities. Some analysis leads me to believe that there must be some restrictions on the elements of P for the solution to be tractable, but further work is needed.

2.5 Decoding Failure

With the calculation of the probability of correctly decoding and incorrectly decoding complete, the probability of decoding failure occupies the remainder of the probability space. The probability of the decoder failing to decode for IID noise is $P_{df} = 1 - P_{cd} - P_{id}$ and the upper and lower bounds for IDD noise are

$$P_{DF}^{UB} = 1 - P_{cd} - P_{ID}^{LB} \tag{2.27}$$

and

$$P_{DF}^{LB} = 1 - P_{cd} - P_{ID}^{UB}, \tag{2.28}$$

which are equal when t is small and all the plateau representatives can be summed over.

2.6 Results for Single RS Codewords

The results in this section are the probability of decoding error for a single RS codeword. The transmitted codeword is modulated using M-ary orthogonal signaling and is transmitted through a channel. Symbol detection is done using the ratio threshold test (RTT), and the corresponding symbol error correct, incorrect and erasure probabilities are used to calculate the probability of decoding error. The receiver block diagram is shown and symbol correct, incorrect, and erasure probabilities are derived in Appendix A.

2.6.1 Single Codeword in AWGN

To examine the results of these calculations we, we consider a singly extended RS code generated over GF(32). Since the RS code is linear, we may transmit the all

zeros code word without loss of generality. All code symbols are transmitted as M-ary orthogonal symbols which are corrupted by AWGN noise during transmission. The receiver detects the symbols non-coherently, applies the ratio threshold test (RTT), and makes hard decisions as to which symbol was transmitted. This model can be thought of as the M-ary symmetric channel with erasures.

The first example is the IID case where all the symbols are corrupted by noise with the same distribution. In Figure 2.5, the rate and RTT parameter γ are varied and the figure shows the probability that the RS codeword is decoded in error. Note the non-monotonicity in error probability associated with a decrease in SNR for non-zero values of γ, this is a result of the RTT detecting and erasing symbols that are unreliable. As the SNR decreases, more and more symbols fail the RTT, and as a result the RS decoder is more likely to declare a decoding failure rather than make an incorrect decision, thus the probability of incorrect decoding falls. To illustrate the required receive energy for a given performance metric, Figure 2.6 gives the minimum SNR required such that the probability of incorrect decode is less than 1×10^{-20}.

2.6.2 Single Codeword in the Partial Band Jamming Channel

As a second example we transmit a codeword in the partial band jamming channel (PBJ). In this channel a hostile jammer is allocated a given amount of noise power to distribute over a fraction ρ of the channel. Thus the probability that each symbol experiences increased noise power is ρ and the noise density for that symbol is $\frac{N_0}{2} + \frac{N_J}{2\rho}$. The remaining symbols are only corrupted by background noise, with density $\frac{N_0}{2}$. As before, all code symbols are transmitted as M-ary orthogonal symbols which are corrupted during transmission. The receiver detects the symbols non-coherently,

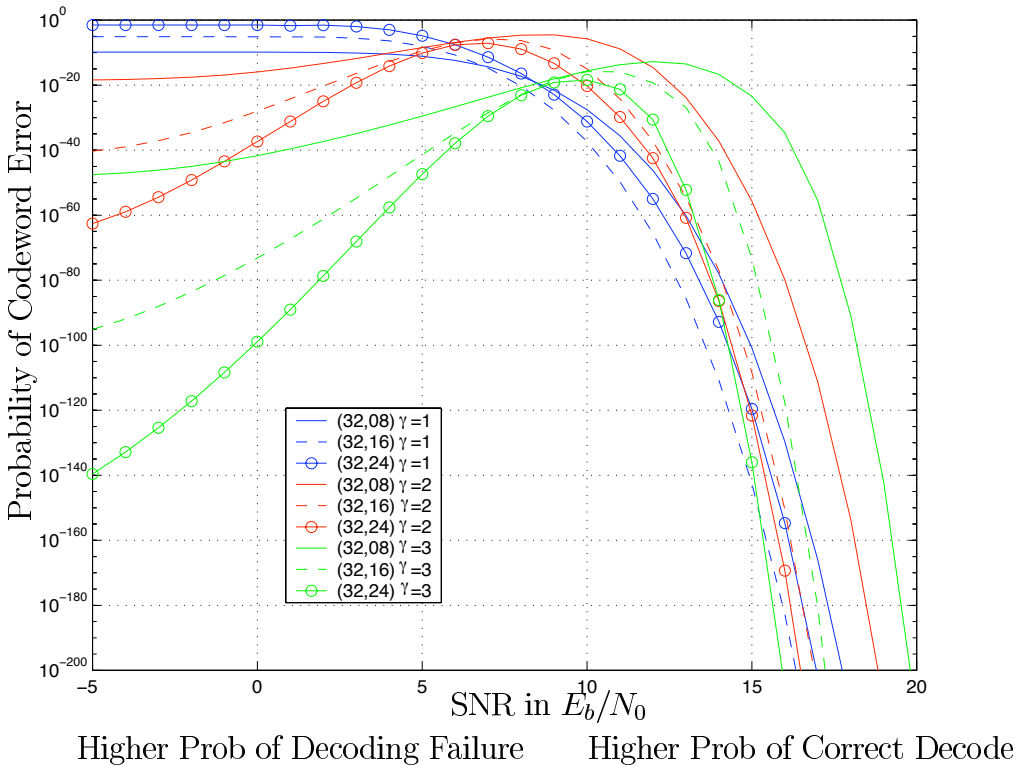


Figure 2.5: Probability of RS Codeword Decoding Error in AWGN

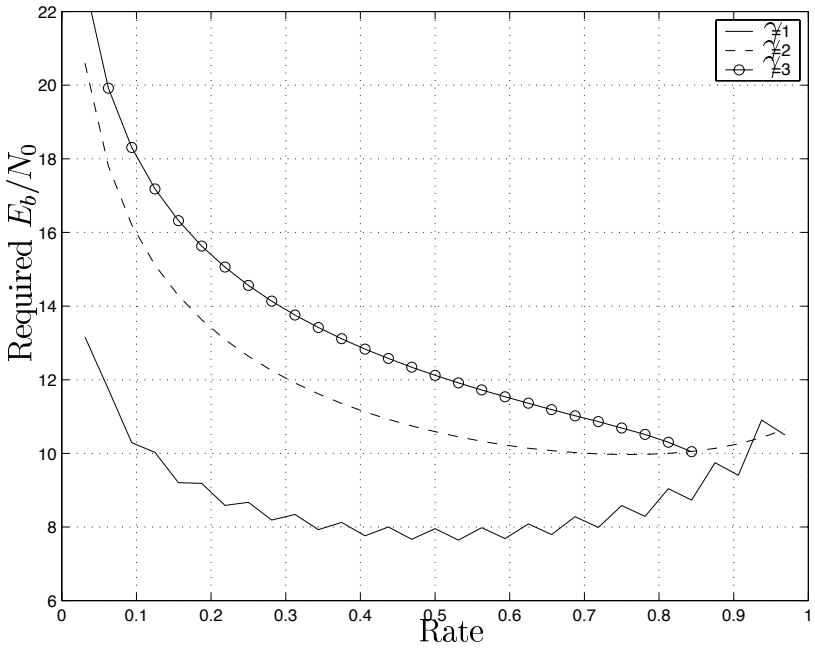


Figure 2.6: Minimum SNR required for $P_{ID} \leq 1 \times 10^{-20}$

applies the ratio threshold test (RTT), and makes hard decisions as to which symbol was transmitted. This model can be thought of as the M-ary symmetric channel with erasures with the caveat that the transition probabilities may change every symbol. Because the PBJ channel imparts one of two noise densities on each symbol transmitted through it, the symbols have one of two different transition probabilities and there are at most $K+1$ plateau's for a fixed code rate ($\frac{K}{N}$) and jamming fraction (ρ).

In Figure 2.7 the average probability that the codeword decodes in error is displayed for RTT parameter $\gamma = 1$ and varied $\frac{E_b}{N_J}$ for a (32,16) RS codeword. The average assumes the distribution of the jammer's parameter ρ is uniform over the interval $(0, 1)$. Note the error floor as the $\frac{E_b}{N_J}$ increases. This behavior is explained by the relative contribution of the jamming interference decreasing; thus, the probability of decoding error approaches the performance of the system that does not experience jamming. In Figure 2.8 we plot the performance of the system for $\gamma = 2$. The performance for low $\frac{E_b}{N_0}$ is relatively unaffected by the jammer. The probability of that symbols are erased by the RTT at a low SNR is relatively high and the addition of extra noise from the jammer will not cause significant additional degradation, thus the probability of decoding error is relatively flat across the whole range of jammer power. However, at higher $\frac{E_b}{N_0}$ the addition of a strong jammer to the system can cause a significant increase in the probability of codeword error. The point at which this transition occurs corresponds to the peak of the decoding error probability in Figure 2.5; for a $(32, 16)$ code with $\gamma = 2$ the this transition point is $7.252dB$. This leads to the conclusion that, all else being equal, if we are optimizing a system utilizing the RTT for the worst case probability of decoding error in both AWGN and PBJ channels, the worst selection of $\frac{E_b}{N_0}$ are roughly equivalent between the 2 channels.

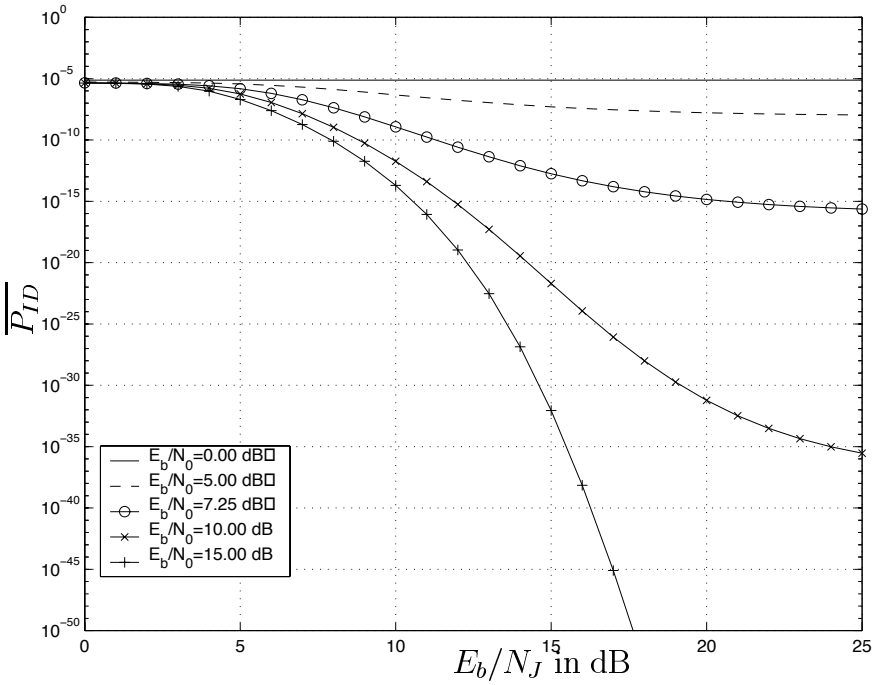


Figure 2.7: Average Decoding Error Probability in the PBJ Channel, $\gamma = 1$.

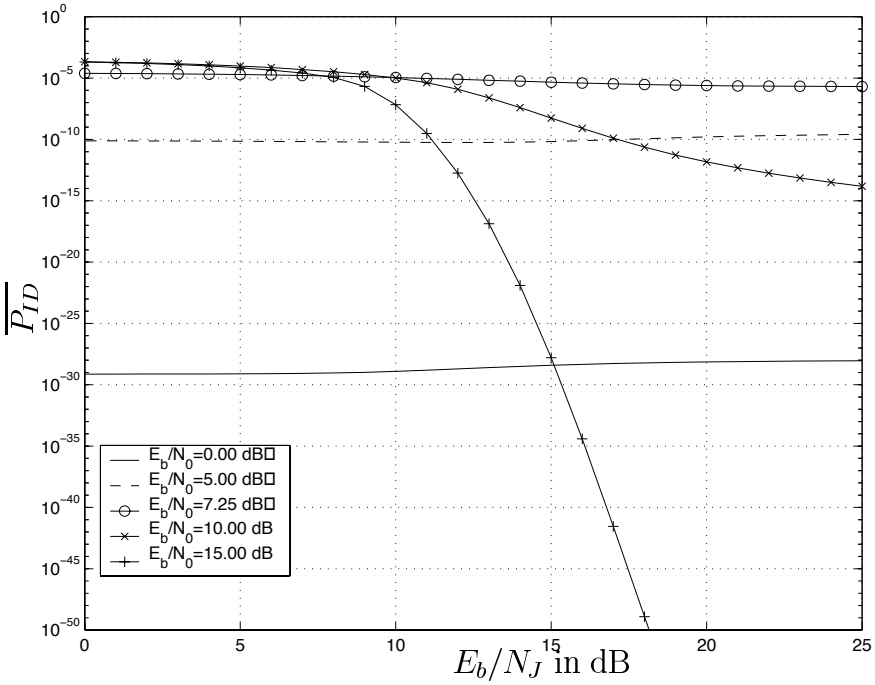


Figure 2.8: Average Decoding Error Probability in the PBJ Channel, $\gamma = 2$.

2.7 RS Codes as Constituent Codes in a Product Code

In this section we examine the performance of a Product Code (PC) in the partial band jamming channel (PBJ). We focus on three areas: optimum rate selection for the constituent codes (Sections 2.7.4 and 2.7.5), determining the probability of correct, incorrect and decoding failure (Section 2.7.6), and finally some discussion of the selection of constituent code decoding radius (Section 2.7.7).

2.7.1 System Model

In the partial band jamming channel, a fraction of the total bandwidth experiences a noise density that is higher than the normal background thermal noise. This extra noise is due to the energy imparted by a, possibly hostile, jamming source. In the frequency hopped spread spectrum (FH-SS) incarnation of this model, the overall system bandwidth is divided into channels. There are two states in this model; either a channel is jammed, or it is not. If a channel is jammed, its noise density increases. In this section we assume the background thermal noise to be zero, i.e. if a channel is not jammed its noise density is zero. We assume that the distribution of jammed channels is uniform. The fixed power jammer changes the noise density on the channels it chooses to jam by adjusting the parameter ρ, the fraction of the system bandwidth W jammed. If the jammer has power J Watts to distribute over the bandwidth jammed, then the added noise density in the jammed bandwidth is $N_J = \frac{J}{\rho W}$.

2.7.2 Product Code Basics

Product codes are arrays that are composed of two constituent codes, a row code and a column code. Data symbols are placed in an $K_r \times K_c$ symbol array and the array filled out by applying the constituent row code to the rows and the constituent column code to the columns. This increases the total packet size to $N_r \times N_c$, where N_r and N_c are the number of symbols in a complete row and column. The rate of the row code is, of course, K_r/N_r; likewise the column code rate is K_c/N_c. The overall rate of the code is:

$$R = \frac{K_r K_c}{N_r N_c} = R_r R_c, \tag{2.29}$$

the overall rate is the product of the constituent code rates. Thus, the moniker "Product" code. We use singly extended RS codes as our row and column codes. These codes are maximum distance separable (MDS) codes built from symbols in GF($M = 2^q$) of length $N_r = N_c = M$ where the minimum distance for the row code is $d_r = N_r - K_r + 1$, and likewise for the column code, $d_c = N_c - K_c + 1$. [4]

2.7.3 Modulation and Transmission

We model the transmission of data on a FH-SS system by transmitting one row per channel hop. Therefore, a row has the same noise statistics for all symbols in that row. Since a RS code whose length increases must use a growing alphabet of symbols, ($M = N_r$), we will represent each symbol as one of an M-ary orthogonal signal set. For a Rayleigh fading channel with AWGN, as in [17], the corresponding symbol crossover probability is:

$$P_a = F(\rho) = \frac{M-1}{2\sigma_0^2} \int_0^\infty \left[1 - \exp(-z/2\sigma_1^2)\right] \left[1 - \exp(-z/2\sigma_0^2)\right]^{M-2} \left[\exp(-z/2\sigma_0^2)\right] \, dz, \tag{2.30}$$

where E_b is the energy per data bit and P_a is the symbol crossover as defined in Section 2.2. Note that we drop the l superscript since all symbols in a row experience the same noise statistics. The inverse function is numerically determined and is expressed in shorthand as:

$$\rho = F^{-1}(P_a). \tag{2.31}$$

2.7.4 Worst case rate selection: the asymptotic case

We examine this problem from a game-theoretic formulation. The game is between the row decoder and a jammer. A wide-band jammer will cause random errors and thus a good strategy for the decoder is to operate to maximize the total error correction capability of the decoder. A narrow-band jammer will cause bursts of errors during a hop and thus a good strategy for the decoder will be to perform error detection and let the column code correct the induced erasures. The payoff function for the game is the rate at which reliable communication is possible. For this analysis we examine very long Reed-Solomon codes. Our results are a region of possible row and column rate pairs for which reliable communication is possible. In addition we find the maximum rate for which reliable communication is possible with Reed-Solomon codes.

For this analysis, the row and column rates are fixed and the number of symbols N_r, N_c becomes large. As the number of symbols become large, the expected number of symbol errors in any transmission scheme, with a fixed crossover probability (P_a), goes to $N_r P_a$, by the weak law of large numbers (WLLN).

For our Reed-Solomon row code, we can correct the received row from all generated

errors if

$$\begin{aligned}
\lfloor \frac{N_r - K_r}{2} \rfloor &\geq N_r P_a \\
\frac{1 - \frac{K_r}{N_r}}{2} &> P_a \\
\frac{1 - R_r}{2} &> P_a \\
1 - R_r &> 2P_a \\
R_r &< 1 - 2P_a. \qquad (2.32)
\end{aligned}$$

The conclusion we draw is that the row can be properly decoded if its rate is appropriately small.

If the row rate is too large, with high probability the row decoder will fail, which means that no codeword will be within the error correcting capability of the received vector. If this happens, we erase the entire row and the corresponding elements in the column codes are these erasures. Let e be the number of erasures passed to the column decoders. The column coders can correct e erasures if:

$$\begin{aligned}
d_c - 1 &> e \\
N_c - K_c &> e \\
1 - R_c &> \frac{e}{N_c} \\
R_c &< 1 - \frac{e}{N_c} \\
R_c &< 1 - \rho. \qquad (2.33)
\end{aligned}$$

A similar conclusion can be drawn, if the column rate is appropriately small the column decoder will correct all erasures passed to it. If the number of erasures exceeds the error correcting capability of the column codes the decoder will declare a failure. Asymptotically, the probability that an undetected error occurs goes to zero as the error correcting capability of the column code grows large.

To find the best rate we have a classic min max style optimization. Can the transmitter *minimize* the effect of the *maximum* effectiveness of the jammer. This is equivalent to finding the *maximum* rate that the transmitter can achieve in the presence of an intelligent jammer. We assume the jammer is privy to the transmitter's row and column rates and we develop the ranges of the jamming parameter, ρ.

To prevent the column codes from decoding properly, using (2.33) the jammer may select any ρ such that

$$\rho \geq 1 - R_c. \tag{2.34}$$

To prevent the row codes from decoding properly, using (2.32) the jammer may select any ρ such that

$$\rho \leq F^{-1}(\frac{1 - R_r}{2}). \tag{2.35}$$

Note the flip in the inequality, this is due to the fact that $F^{-1}(\bullet)$ is a decreasing function.

The probability that the jammer successfully prevents the receiver from decoding properly is defined by the following conditions:

$$P_{jam} = \begin{cases} 0 & \rho > F^{-1}(\frac{1-R_r}{2}) \\ 0 & \rho < 1 - R_c \\ 1 & \rho \geq 1 - R_c \text{ and } \rho \leq F^{-1}(\frac{1-R_r}{2}) \end{cases} \tag{2.36}$$

The jammer may pick any ρ such that $1 - R_c \leq \rho \leq F^{-1}(\frac{1-R_r}{2})$ and it will maintain the ability to jam the packet. Thus, the transmitter must select rates as to force the jammer's ρ choices to be the empty set. It must force

$$1 - R_c > F^{-1}(\frac{1 - R_r}{2}). \tag{2.37}$$

Because the RHS of (2.37) is a decreasing function of $(1-R_r)$, it follows that for a fixed R_c the maximum row rate, R_r, occurs when (2.37) is equality. The corresponding overall rate is

$$R = (1 - 2F(\rho))(1 - \rho). \tag{2.38}$$

To determine the maximum overall rate, we maximize (2.38) by varying ρ, which in turn determines R_c by (2.37). Since the symbol energy is a function of R_r, P_a is also a function of R_r. To determine the maximum R_r that satisfies (2.32) and (2.38) numerical solutions were necessary.

Figure 2.9 shows the region where a (R_r,R_c) pair achieves reliable communication. Outside of this region, the intelligent jammer succeeds in overcoming the error correcting capacity of the PC. Figure 2.11 illustrates the PC rates that the asymptotic analysis states are achievable for various column code rates. The SNR of the system

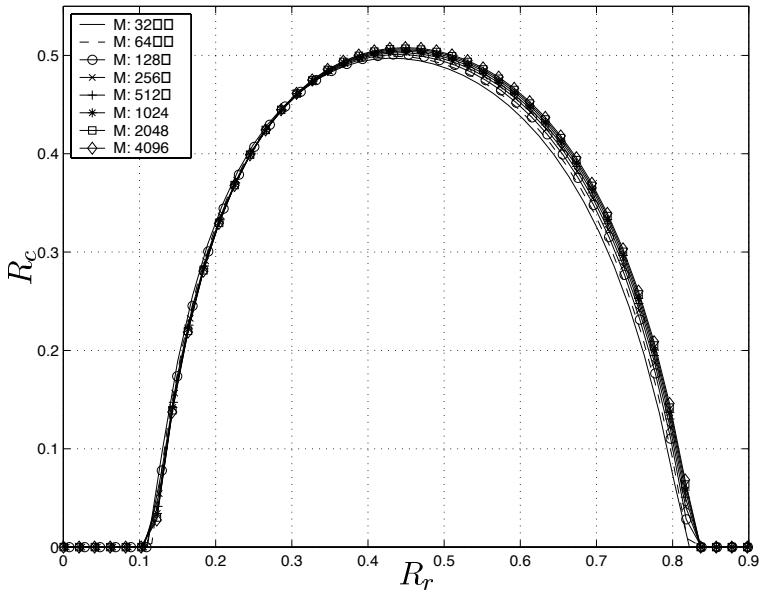


Figure 2.9: Asymptotic results: achievable constituent code rates, SJR=10dB.

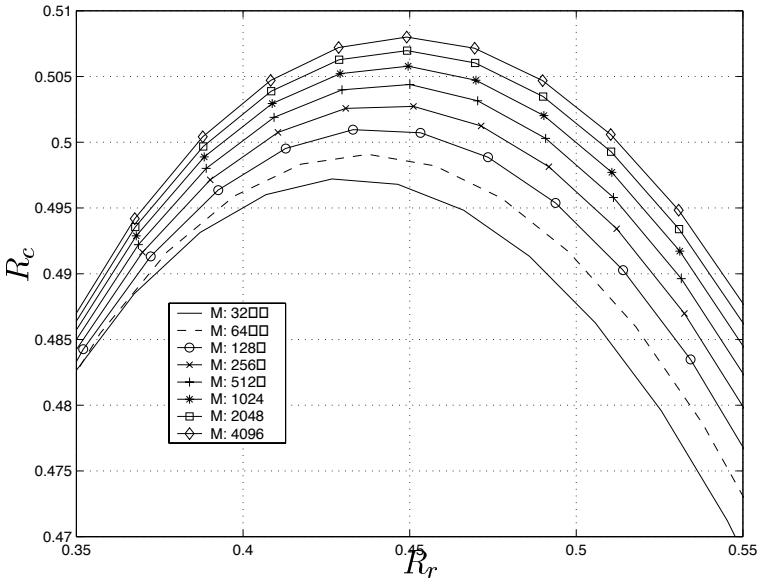


Figure 2.10: Asymptotic results: achievable constituent code rates, SJR=10dB.

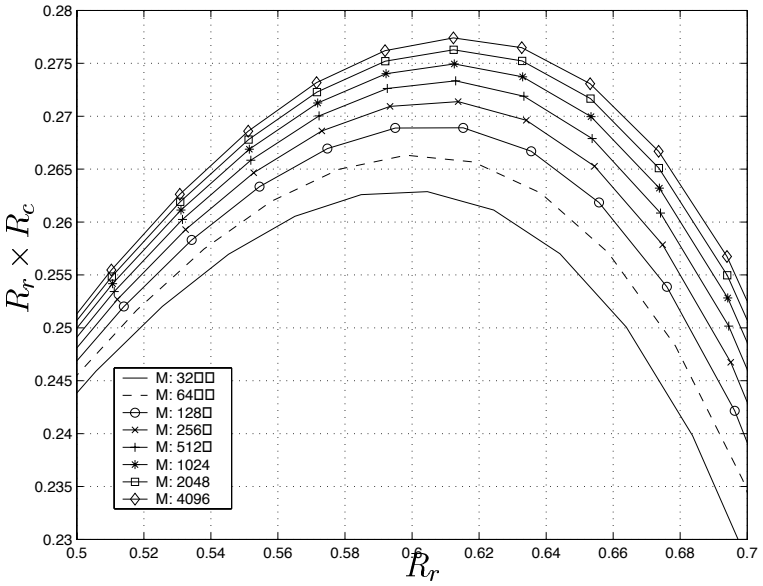


Figure 2.11: Asymptotic results: achievable PC rates, SJR=10dB.

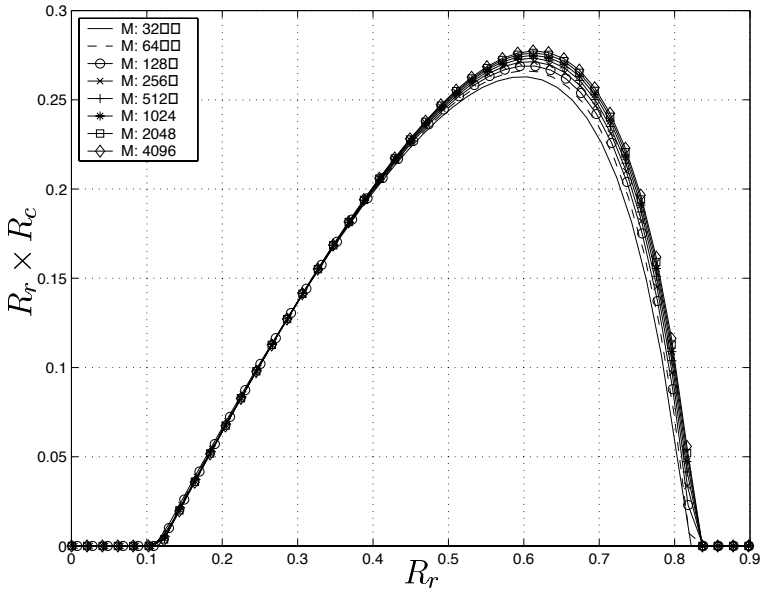


Figure 2.12: Asymptotic results: achievable PC rates, SJR=10dB.

is defined as

$$\text{SNR} = 10 \log_{10} \left(\frac{E_b}{N_J} \right). \tag{2.39}$$

Note the N_J term is the noise density if the jammer sets $\rho = 1$. By selecting the maximum rate for each SNR, we obtain the maximum achievable PC rates, as shown in Figure 2.13.

This concludes the asymptotic analysis of optimal constituent code rates and overall rate for a product code when codewords are hopped in the partial band jamming channel. We have shown the achievable rates in the presence of a jammer which has perfect knowledge of the encoding method. Additionally, (2.37) separates the rate pairs that allow communication from those whom are blocked and allows the transmitter to maximize the overall PC rate.

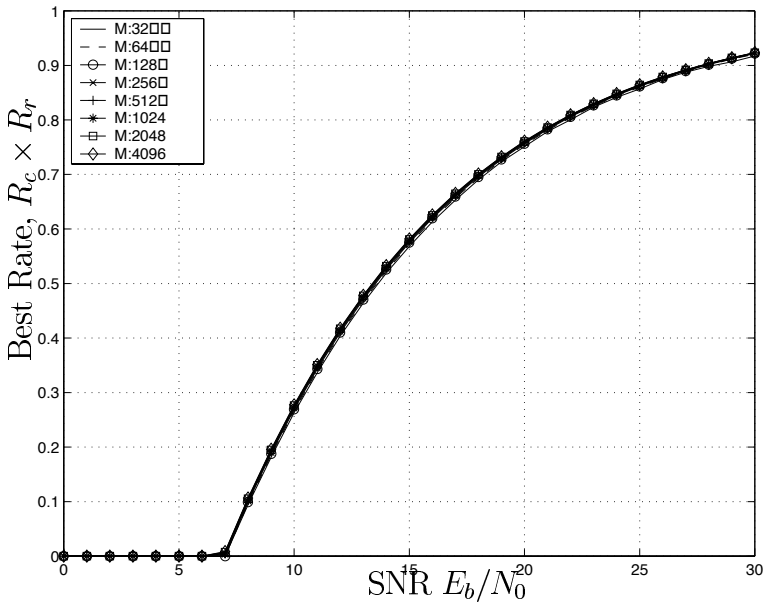


Figure 2.13: Asymptotic results: maximum achievable PC rates.

2.7.5 Worst case rate selection: the finite case

The previous analysis of large product codes is a first pass estimate of the best rates for the row and column constituent codes. The results were developed by letting the lengths of the constituent codes become large, and then applying those results to a finite code. This asymptotic/finite analysis may lead to some inaccuracy because $F(\rho)$ and $F^{-1}(P_a)$ are dependent on the the length of the codes through the growing size of the code symbols and the growing numbers of M-ary orthogonal dimensions. The degree of error becomes less and less as the code grows because the log of the length grows slowly. As the constituent codes become longer the achievable rates grow and the optimal rate shifts, as illustrated in Figure 2.14, where code length ranges from $M = N_r = N_c = 8$ to $M = N_r = N_c = 2^{12}$.

While small, this inaccuracy leads to the question of how to determine the best constituent code rates for short codes without asymptotic analysis, and how different the rates are from the asymptotic results. To answer these questions we determine

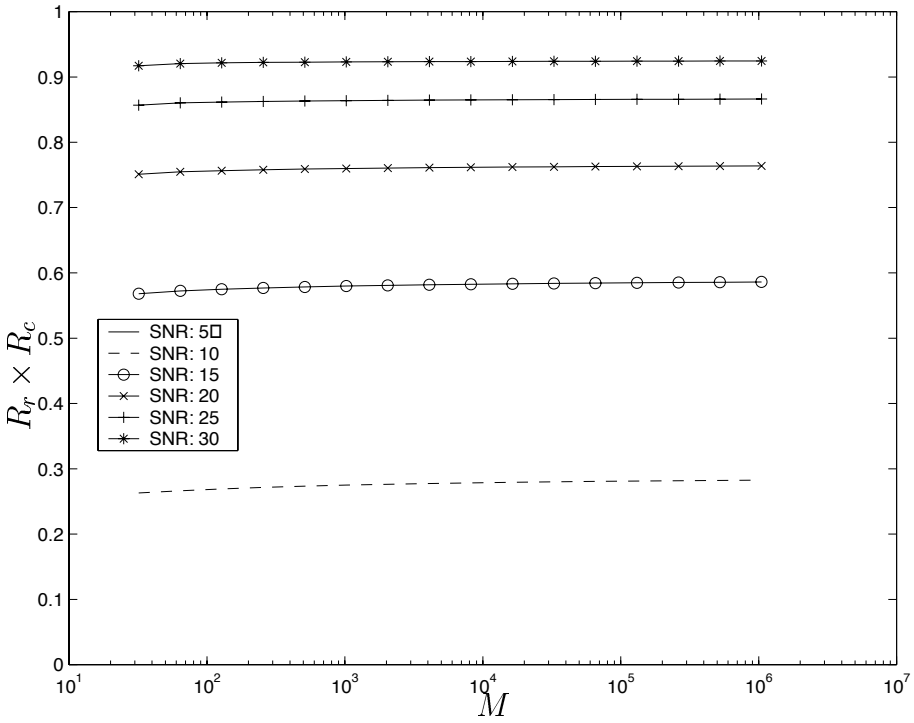


Figure 2.14: Illustration of inaccuracy

the probabilities of correct decoding, incorrect decoding, and decoding failure for the short row codes. Then we must develop expressions for the achievable rate pairs, and from these draw the maximums.

2.7.6 Error probabilities for finite length product codes

We are interested in determining the probability that a product codeword is correctly received. A product codeword consists of $N_r \times N_c$ symbols. Additionally, to analyze any of the current re-transmission schemes, knowledge of the probability of undetected error is necessary. After decoding row-wise and then column-wise, a codeword is correctly received when all columns have decoded correctly. An undetected error occurs when there are no column decoding failures and the are one or more column decoding errors. A codeword failure occurs when there are one or more column

decoding failures.

We will compute the conditional probabilities that these events occur, given r rows are jammed. To compute un-conditional probabilities we must know something about the distribution of the parameter r at the jammer, so we consider only conditional probability expressions.

The calculation of the product codeword conditional probability measures depends on a number of constituent codeword conditional probabilities. In particular we will condition on knowing the number of correct, incorrect, and erased rows for both jammed and unjammed rows. Define the following quantities:

- n_c^j = the number of jammed rows that correctly row-decode
- n_i^j = the number of jammed rows that incorrectly row-decode
- n_f^j = the number of jammed rows that fail row-decode
- n_c^o = the number of un-jammed rows that correctly row-decode
- n_i^o = the number of un-jammed rows that incorrectly row-decode
- n_f^o = the number of un-jammed rows that fail row-decode.

Let us refer to the above quantities as 2 vectors:

- $\overline{n^j} = [n_c^j, n_f^j, n_i^j]$
- $\overline{n^o} = [n_c^o, n_f^o, n_i^o]$

These quantities are subject to the constraint that the sum of the vector components must be equal to the number of elements in a column, N_c.

Row code preliminaries

Because the row symbols are either all jammed or all unjammed, we can use the results from Sections 2.3, 2.4.1, 2.5 to determine:

- $P_{cd|x}$, the probability that a row decodes correctly,
- $P_{df|x}$, the probability that a row decode fails,
- $P_{id|x}$, the probability that a row decodes incorrectly.

When a row code error is made, if we assume that errors are uniformly distributed along the whole row then we can write the probability that a symbol is in error and the row decodes in error as:

$$\overline{P_s^x} = \sum_{k=d_{min}}^{N_r} P_{id,k} \frac{k}{N_r}, \tag{2.40}$$

where $P_{id,k}$ is from (2.7). For $P_{cd|x}$, $P_{df|x}$, $P_{id|x}$ and $\overline{P_s^x}$, we replace x with o when we refer to unjammed rows and x with j when we refer to jammed rows.

Now we have developed the 8 probability measures for the row decoding process: $P_{cd|j}$ $P_{df|j}$ $P_{id|j}$, $\overline{P_s^j}$, $P_{cd|o}$ $P_{df|o}$ $P_{id|o}$, and $\overline{P_s^o}$. Evaluating these measures from their underlying expressions was derived in a different fashion in [4], but that derivation relied on an exhaustive counting technique which we have replaced with a recursion.

Individual column symbol decode probabilities

In order to calculate the probability of column error we must know something about the probability that individual column symbols are correct, incorrect, or were

erased. The probability that a column symbol is correct is as follows:

$$\begin{aligned}
P_{sc|x} &= P(\text{symbol is correct}|\text{row decodes correct})P(\text{row decodes correct}) \\
&\quad + P(\text{symbol is correct}|\text{row decode fails})P(\text{row decode fails}) \\
&\quad + P(\text{symbol is correct}|\text{row decodes incorrect})P(\text{row decodes incorrect}) \\
&= 1 \times P(\text{row decodes correct}) \\
&\quad + 0 \times P(\text{row decode fails}) \\
&\quad + P(\text{symbol correct and row decodes in error}) \\
&= P_{cd|x} + (1 - P_{cd|x} - P_{df|x} - \overline{P_s^x}). \qquad (2.41)
\end{aligned}$$

The probability that a column symbol is incorrect is as follows:

$$\begin{aligned}
P_{si|x} &= P(\text{symbol is incorrect}|\text{row decodes correct})P(\text{row decodes correct}) \\
&\quad + P(\text{symbol is incorrect}|\text{row decode fails})P(\text{row decode fails}) \\
&\quad + P(\text{symbol is incorrect}|\text{row decodes incorrect})P(\text{row decodes incorrect}) \\
&= 0 \times P(\text{row decodes correct}) \\
&\quad + 0 \times P(\text{row decode fails}) \\
&\quad + P(\text{symbol incorrect and row decodes in error}) \\
&= \overline{P_s^x}. \qquad (2.42)
\end{aligned}$$

Lastly, the probability that a column symbol is erased is as follows:

$$\begin{aligned} P_{sf|x} &= P(\text{symbol is erased}|\text{row decodes correct})P(\text{row decodes correct}) \\ &\quad + P(\text{symbol is erased}|\text{row decode fails})P(\text{row decode fails}) \\ &\quad + P(\text{symbol is erased}|\text{row decodes incorrect})P(\text{row decodes incorrect}) \\ &= 0 \times P(\text{row decodes correct}) \\ &\quad + 1 \times P(\text{row decode fails}) \\ &\quad + 0 \times P(\text{row decodes incorrect}) \\ &= P_{df|x}. \end{aligned} \tag{2.43}$$

This concludes development the 6 probability measures for column symbols: P_{sc}^{j}, P_{si}^{j}, P_{sf}^{j}, P_{sc}^{o}, P_{si}^{o}, and P_{sf}^{o}.

Packet decode statistics

To determine the probability of decoding the packet incorrectly, we start by conditioning on the number of rows jammed, r. Define $P_{PKTICD}(r)$ as the probability that the packet decodes incorrectly. For incorrect decoding to occur there must be at least one incorrectly decoded column with the remainder decoding correctly, no decoder failures are allowed because we are looking for undetected errors. This probability is expressed as:

$$P_{PKTICD}(r) = \sum_{i=1}^{K_r} \binom{K_r}{i} (P_{ICD1col}(r))^i (P_{CORR1col}(r))^{K_r - i}, \tag{2.44}$$

where

$$P_{ICD1col}(r) = \text{Probability that a column decodes incorrectly, given } r \tag{2.45}$$

$$P_{CD1col}(r) = \text{Probability that a column decodes correctly, given } r. \tag{2.46}$$

Note that the symbol crossover statistics for a column can be viewed as 2 different distributions. The IDD analysis from Sections 2.3 and 2.4 can easily be applied here. This is done by constructing the transition probability vectors P_c, P_f and P_a for the correctly decoding answer:

$$\begin{aligned} P_c &= [P_{sc|o}, \ldots, P_{sc|o}, P_{sc|j}, \ldots, P_{sc|j}] \\ P_a &= [P_{si|o}, \ldots, P_{si|o}, P_{si|j}, \ldots, P_{si|j}] \\ P_f &= [\underbrace{P_{sf|o}, \ldots, P_{sf|o}}_{N_c - r \text{ of these}}, \underbrace{P_{sf|j}, \ldots, P_{sf|j}}_{r \text{ of these}}], \end{aligned} \tag{2.47}$$

where the number j of elements in each vector is equal to the number of rows jammed. All that remains is to apply these vectors to (2.3). Similarly, the transition vectors for incorrect decoding can be constructed by selecting the set of plateau representatives and applying (2.9) through (2.14). If the column decoding failure is desired apply (2.27) which is tight because of the small $t = r$.

Once the column decode statistics are complete, we can determine the packet decode statistics. To decode a packet correctly we must have no column erasures and no undetected column errors,

$$P_{PKTCD}(r) = (P_{CORR1col}(r))^{K_r}. \tag{2.48}$$

The event space of the packet decoder consists solely of correct packet decode, incorrect packet decode, and packet decode failure. These 3 events are mutually exclusive, and their union is the whole probability space for the outcome of the decoder. Hence:

$$P_{PKTFAIL}(r) = 1 - P_{PKTCD}(r) - P_{PKTICD}(r). \tag{2.49}$$

In Figures 2.15, 2.16, and 2.17 we show the probabilities of packet correct decoding, incorrect decoding, and decoding failure for a 32×32 PC, with the assumption that

ρ is uniformly distributed.

While instructive, the determination of "best" coding rates is difficult to determine from Figures 2.15, 2.16, and 2.17 without performance requirements. Figures 2.18 and 2.19 illustrate the rate selection for the case where the SJR 10dB and SNR is 30dB. The asymptotic results are from the analysis in Section 2.7.4 with $M = N_r = N_c = 32$. The finite length rates are determined by selecting the row and column rates that guarantee error and erasure rates to be at most 1×10^{-5} and 1×10^{-15} as labeled in the caption. The selected row rates for the finite calculation decrease, to correct more errors in the row code, as the error performance requirement increases. For the case where SJR is 20dB and SNR is 40dB, Figures 2.20 and 2.21 illustrate a similar, albeit less severe effect in the rate selection. These results illustrate the importance of considering undetected error when determining optimum rates for the constituent codes. Ignoring an error rate requirement may lead to overly optimistic rate selection.

2.7.7 Varying the radii of the decoder

The above sections consider the effect of variable rates on the packet probability measures. Varying the encoding rate entails the adjustment of parameters on both the transmitter and the receiver. The row code of the product code serves two functions; it corrects small numbers of errors, or it erases the whole row to let the column code have a chance to correct. If we reduce the decoding radius of the row code we would intuitively expect to see the probability of correctly decoding drop, the probability of incorrectly decoding drop, and the probability of decoding failure (erasure) rise. Since the overall goal of the product code is to have no undetected column errors, any incorrect symbols in the row code that are changed to erasures by decreasing the

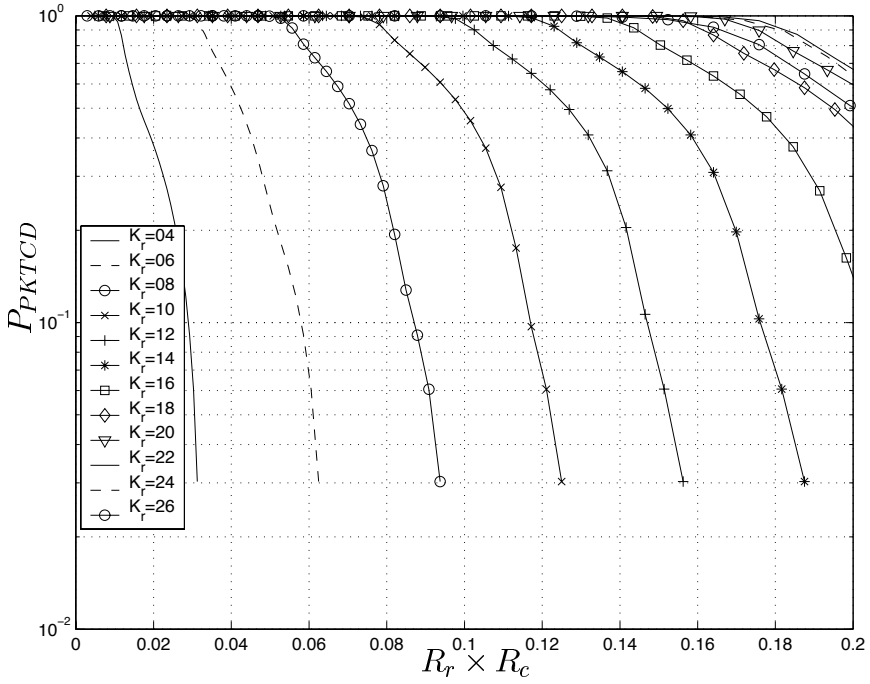


Figure 2.15: Unconditional probability of packet correct decoding, SJR=10dB.

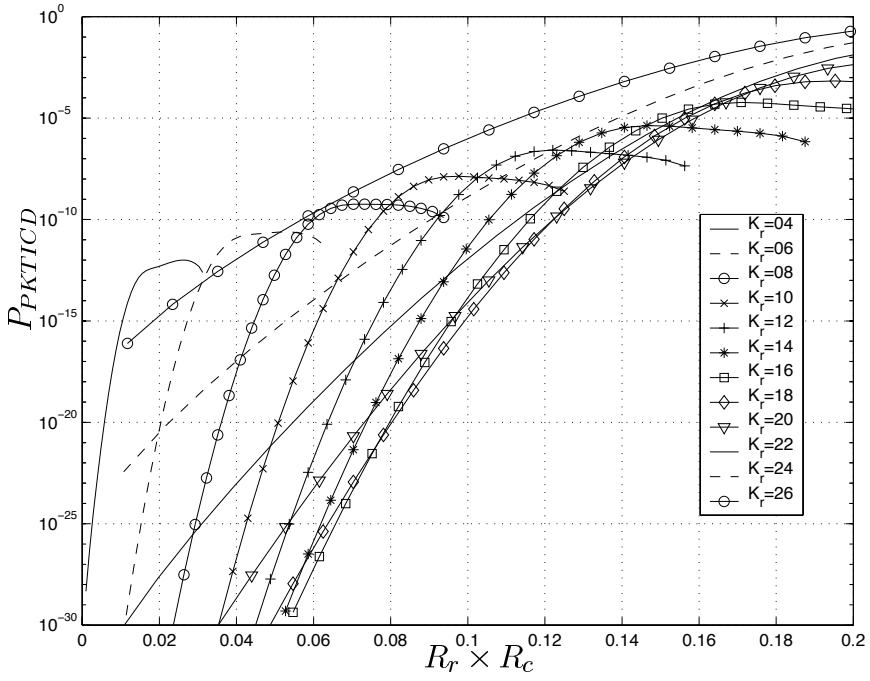


Figure 2.16: Unconditional probability of packet incorrect decoding, SJR=10dB.

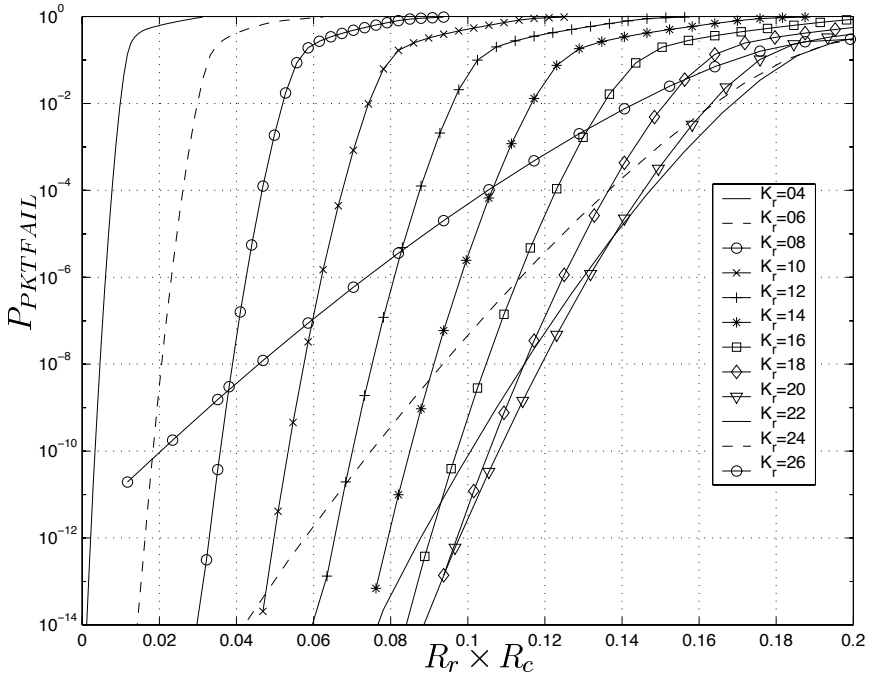


Figure 2.17: Unconditional probability of packet decoding failure, SJR=10dB.

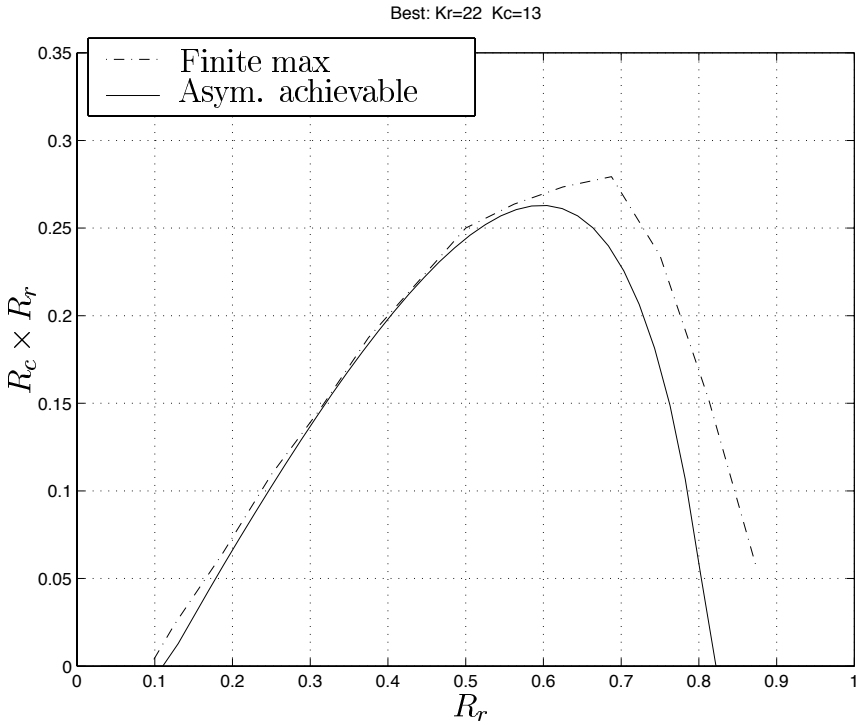


Figure 2.18: Comparison of optimum rates, $P_{PKTFAIL} < 1 \times 10^5$, $P_{PKTICD} < 1 \times 10^5$, SJR=10dB.

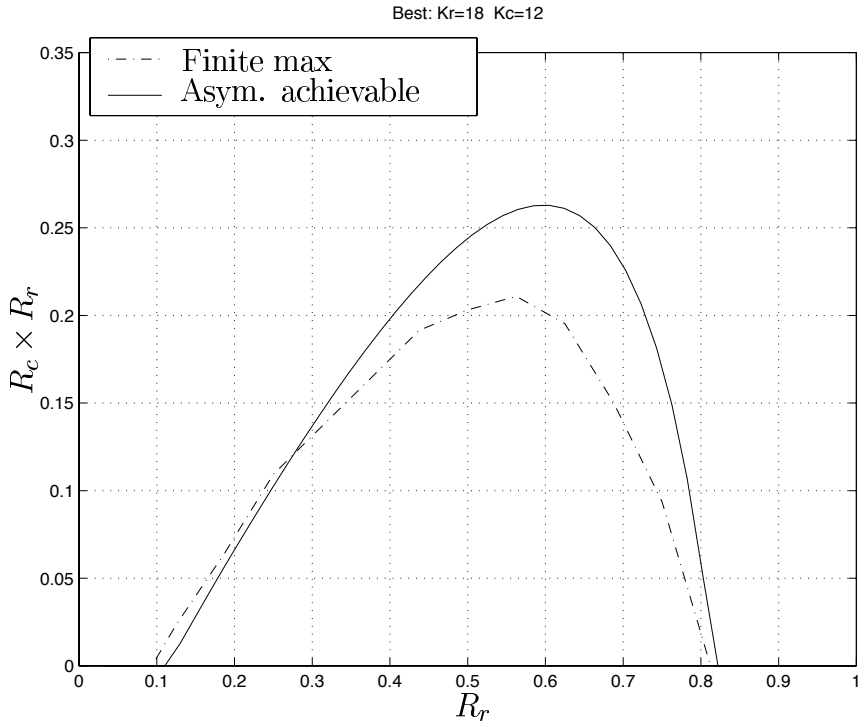


Figure 2.19: Comparison of optimum rates, $P_{PKTFAIL} < 1 \times 10^5$, $P_{PKTICD} < 1 \times 10^{15}$, SJR=10dB.

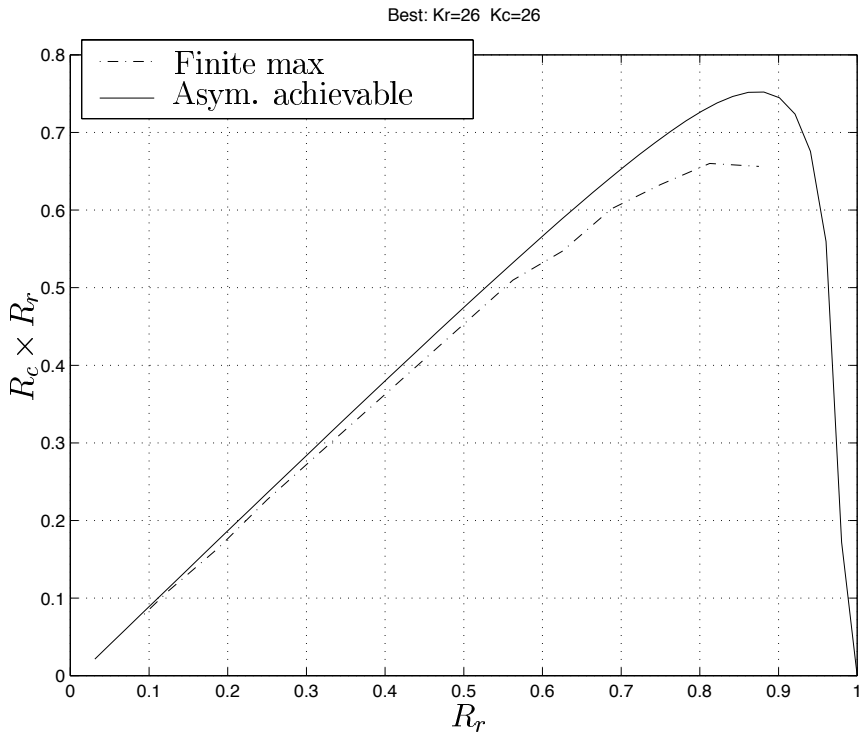


Figure 2.20: Comparison of optimum rates, $P_{PKTFAIL} < 1 \times 10^5$, $P_{PKTICD} < 1 \times 10^5$, SJR=20dB.

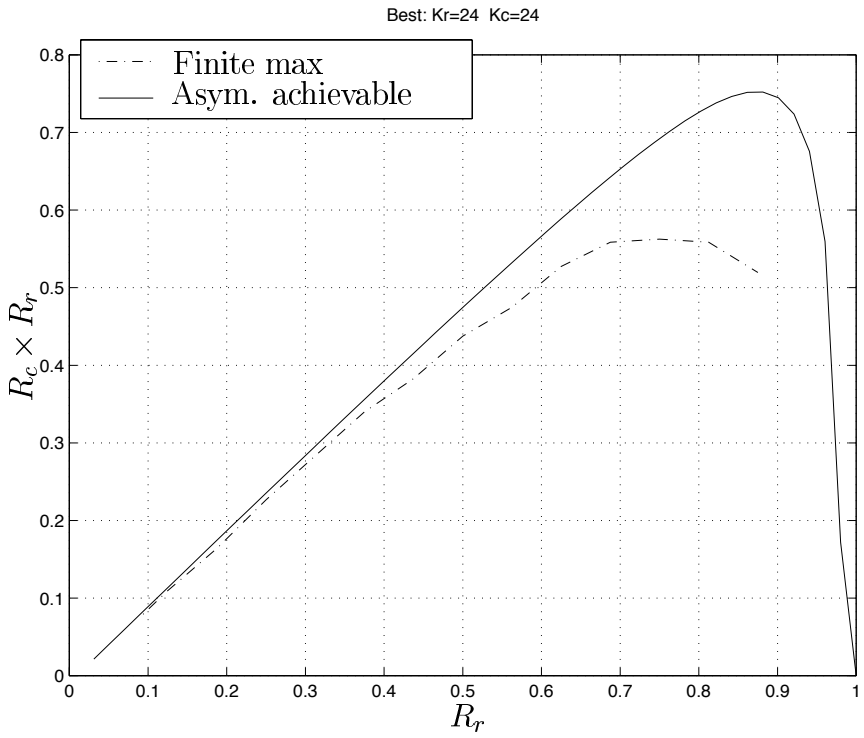


Figure 2.21: Comparison of optimum rates, $P_{PKTFAIL} < 1 \times 10^5$, $P_{PKTICD} < 1 \times 10^{15}$, SJR=20dB.

decoding radii, might advantageously impact the probability of correct packet decode.

The column code's function is to correct for the erasures flagged by the row code and the incorrect symbols generated by the row code. It is not clear if adjusting the column decoder's radii will have advantageous effects. By decreasing the radii we would expect the number of incorrectly decoded columns to decrease, but we might also expect the number of columns which fail decode to rise. Some investigation is needed here.

The advantage of this scheme lies in the fact that the receiver could rapidly adapt to changing conditions, if in fact an optimality situation exists. By reducing the recoding radii, the receiver could decrease the probability that a packet is received in error at no cost in complexity (potentially a savings) and with no adjustments in the overall packet rate.

2.8 Computational Complexity

We compare the complexity of our method with the errors only IID decoding method found in [4], and with the errors and erasures decoding IDD method found in [1].

In [4], the combination of (7.8), (7.10), and (7.16) gives the following expression:

$$P_{ICD} = \sum_{h=d}^{N} A_h \sum_{s=0}^{l} \sum_{k=h-s}^{h+s} \sum_{r=r_1}^{r_2} () () () () () \tag{2.50}$$

This expression is the probability that a bounded distance errors only decoder incorrectly decodes a received vector that has been corrupted by IID noise. The calculation complexity for this is, clearly O(N^4). As a comparison, the complexity of (2.14) specialized to IID noise and EO decoding ($i = 0$) depends on the complexity of the recursion in (2.12). The value of $P\{B_{j|k}^{N,\zeta}\}$ can be determined by populating a table

of $l = N$ columns and $j = N$ rows. In the worst case, filling this table will require populating $N^2/2$ table entries each requiring one addition, two multiplications, and two values from the previous column. The recursion's computational complexity is $\mathrm{O}(N^2)$ and the storage complexity is $\mathrm{O}(N)$. Thus, the computational complexity of (2.14) is $\mathrm{O}(N^4)$, which is equivalent to that in [4].

Next, we compare our computational complexity of EE/BD decoding with IDD noise to that found in [1]. Daraiseh's method evaluates

$$P_e\left(\overline{R}, |\overline{R}|\right) = \sum_{j=d_{min}}^{N} \hat{M}_j \sum_{s=1}^{\binom{N}{j}} P\left(\psi_j^n(s), \overline{R}, |\overline{R}|\right), \tag{2.51}$$

where $P\left(\psi_j^n(s), \overline{R}, |\overline{R}|\right)$ is an $\mathrm{O}(N^6)$ expression. Therefore the overall complexity of his method is $\mathrm{O}(N^7 N^N)$, since $\binom{N}{j}$ grows exponentially in N.

The direct comparison of Daraiseh's method and ours involves the complexity of the recursion in (2.12) and the use of that recursion in (2.14). In EE decoding the above mentioned table must have N^2 rows to account for all possible errors and erasures combinations. Thus the computational complexity for the recursion is $\mathrm{O}(N^3)$ and the storage complexity is $\mathrm{O}(N^2)$. This complexity factored into (2.14) gives a total computational complexity of $\mathrm{O}(N^6 N^N)$. The savings in of a factor of N comes from the selection of errors and erasures that are valid in the limit of our sum, whereas it is calculated as an additional step in [1].

The clear advantage of our scheme lies in the flexibility of the bounding technique from Section 2.4.5. The method developed orders the likelihood of the potential zero patterns and calculates the probability contribution of the most likely patterns first. To bound in this manner, not all zero patterns need to be considered, but the most likely will be considered first. The bound will tighten the fastest with the least compute time, thus the idea of "compute" limited bounds. The complexity of this

bounding is difficult to determine exactly since the generation of the zero patterns is still an open question. A brute force generation and sort of the potential zero patterns is an $O(N_k \log N_k)$ operation, but if some intelligent generation methods are used, significant complexity savings can result.

2.9 Conclusions

This chapter's results are in three areas. First, we developed a recursion to efficiently calculate the probability of correct decode, incorrect decode, and decoding failure for a bounded distance decoded Reed-Solomon code. These results are on the order of $O(N^4)$ for noise with a small number of distributions. As the number of distributions grows large, we give compute limited bounds. By compute limited we mean that the bounds improve as more calculations are performed and the most significant calculations are performed first. The second result is the calculation of correct decode, incorrect decode, and decoding failure for a product code with very long Reed-Solomon constituent codes in the presence of worst case partial band jamming. The third result is the selection of optimal PC rates for finite length constituent codes, in the partial band jamming channel. In developing the best rate choices, we have shown the existence of an optimal row and column rate pair where communication is possible even in the presence of an intelligent jammer. Additionally, the use of asymptotic analysis to select the best constituent code rates is matched approximately to the finite case, with variations based on the required performance goal.

CHAPTER 3

Functional Analysis of Nonlinear Amplifiers

Before beginning any nonlinear analysis endeavor one must take note of the admonition by Maas in [18]:

> ... the engineer who designs a nonlinear circuit (which the engineer may or may not have intended to be nonlinear) will get very little help from conventional sources because most textbook network theory is based on an assumption of linearity. After searching in vain for reliable design information, the engineer usually makes use of what little data he or she can find and "tweaks" the circuit empirically to make it work. The frustration, anger, and insecurity engendered by this situation is responsible for many sleepless nights, ruined marriages, and hateful children.

While this may seem a little extreme, the analysis of nonlinear systems can be very complex and the techniques used can vary dramatically from system to system.

This chapter begins with the motivation for the development of the Modified Instantaneous Quadrature Method (MIQM) and the functional analysis that it precipitates. We then describe a basic circuit to model, give an overview of the characteristics of nonlinear circuits, and present a brief survey of the advantages and disadvantages of current nonlinear analysis techniques. Following this, we show the development of the MIQM and we show that it leads to the ability to model a nonlinear circuit as the combination of two functions, under certain memory constraints. We then illustrate the accuracy of our new method and draw conclusions.

3.1 Motivation

To begin, define the input/output response of a single input circuit to be:

$$y(t) = \mathrm{h}(x(t)), \tag{3.1}$$

where $x(t)$ is the input time varying signal and $y(t)$ is the output signal. Two signals, $x_1(t)$ and $x_2(t)$, are separately input into the circuit. The output due to these inputs can be generally written as:

$$y_1(t) = \mathrm{h}(x_1(t)) \tag{3.2}$$

$$y_2(t) = \mathrm{h}(x_2(t)). \tag{3.3}$$

A circuit is deemed linear if the output of the system due to the linear combination of the two input signals is equal to the linear combination of the outputs due to each input alone, or:

$$\alpha \mathrm{h}(x_1(t)) + \beta \mathrm{h}(x_2(t)) = \mathrm{h}(\alpha x_1(t) + \beta x_2(t)), \tag{3.4}$$

where α and β are arbitrary gain factors.

When the input signal into an LNA is small, the output is generally linearly related to the input. However, when a large amount of power is driven into the input, the underlying transistor in the LNA will saturate and the output is no longer linearly related to the input. If this is the case, most of the standard circuit analysis techniques break down. The determination of the output signal and the analysis of systems when the input may be small or large is a difficult task which is usually managed by simulating the circuit's responses for specific input signals.

The overriding goal of this chapter is to develop a method of circuit analysis that provides the ability to model nonlinear circuits in such a manner that does not suffer

from the drawbacks of existing methods. The method that we introduce, the Modified Instantaneous Quadrature Method (MIQM), permits the analysis of nonlinear systems in such a way that a circuit, with certain restrictions, can be effectively modeled by in-phase and a quadrature-phase nonlinear functions. Some existing nonlinear models use relations rather than functions to model the circuit behavior. Functions by their definition are one-to-one, whereas relations can have multiple output values for a single input value. We stress the necessity of modeling the system with functions, not relations, due to the ability to accurately approximate these functions with a power series. Once this approximation is made, there are known analytical methods to calculate the nonlinear output from the circuit [19–21] (see Appendix C). The output that interests us the most is the gain effects of the fundamental and the intermodulation effects from the mixing that occurs in nonlinear devices.

The role of the MIQM is to correctly determine the functions that accurately model the circuit. This method is key to the functional analysis of nonlinear circuits under large and small signal excitations, and will be used as the analysis model in Chapter 4. Functional analysis offers key advantages over several existing techniques namely Harmonic Balance (HB) and discrete time methods. The advantages of increased calculation speed, large input dynamic range, freedom from aliasing effects and large diversity of input frequencies give this method a unique niche in the world of nonlinear analysis.

3.2 Nonlinear Amplifier Circuit

In this section we detail the design of the LNA that will be analyzed. While the design criterion may not be well motivated at this point (they will be discussed in

Chapter 4) the primary goal of the LNA in a microwave receiver is the amplification of a very low power signal without the addition of significant noise. The LNA design considered in this section will be used to illustrate the effectiveness of the Instantaneous Quadrature Method (IQM) and the new MIQM in Sections 3.4.8 and 3.5, and the effects of circuit induced memory on accuracy in Section 3.6.

The Noise Figure (NF) of a circuit is commonly defined as the ratio of the SNR at the input to the SNR at the output of the given circuit [22]. The minimization of noise figure of a circuit and the maximization of power transfer into a circuit are generally mutually exclusive goals, one can either maximize power transfer or minimize noise figure but not achieve both goals simultaneously. The design of low noise amplifiers inherently involves the tradeoff between the reduction of gain in the circuit due to the input mismatch and the reduction of resulting noise figure. These design choices are typically made in the context of a Smith chart or a CAD tool.

The basic circuit that we will analyze is a derivative of the design found in [23]. Modifications are performed to use lumped circuit elements, varied bias points, and a 900MHz center frequency. The incarnation of the circuit used is shown in Figure 3.1. This circuit is biased by a .91V DC source and matched to 900MHz operation. The input and output matching networks are designed from the Smith chart shown in Figure 3.2, where the noise figure and and available power gain circles are included. The input match is designed such that a 50Ω source is matched to circuit at the point where the circuit exhibits an available power gain of 19.45dB while maintaining a NF of 2dB. The output match of the system is designed for maximum power transfer into a 50Ω load. The S-parameter match over a frequency range of 0.5GHz to 1.5GHz is shown in Figure 3.3.

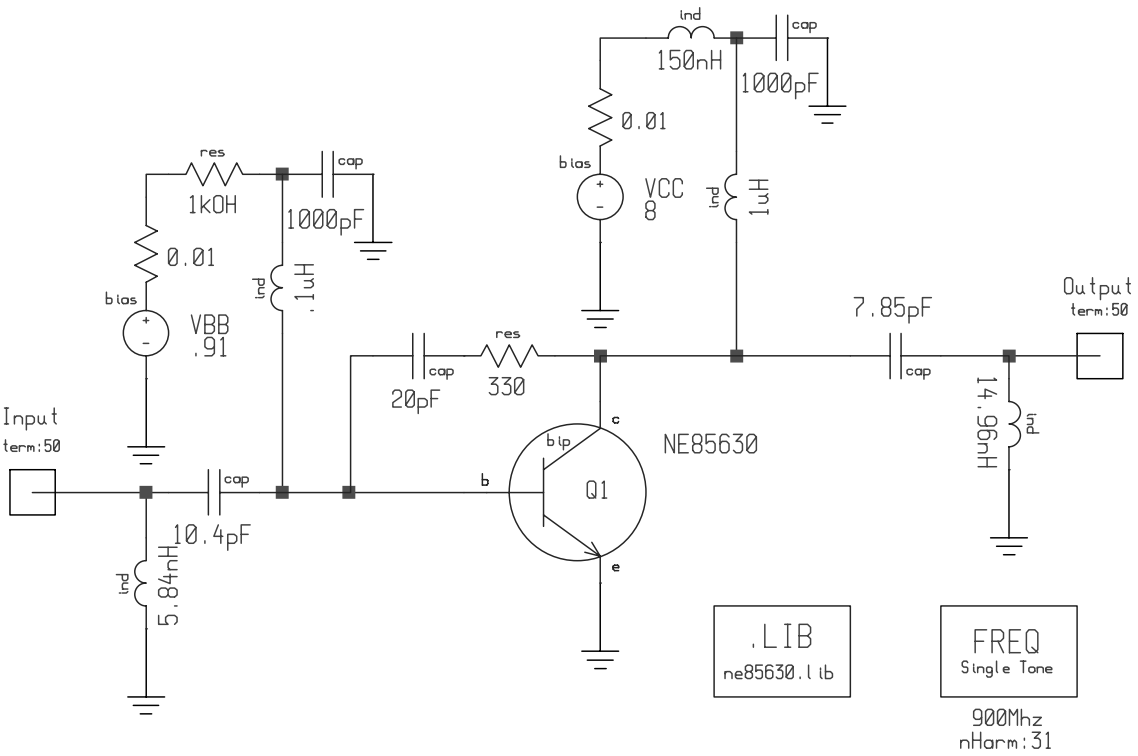


Figure 3.1: Basic 900MHz LNA, .91V bias, GA=19.45dB, NF=2dB.

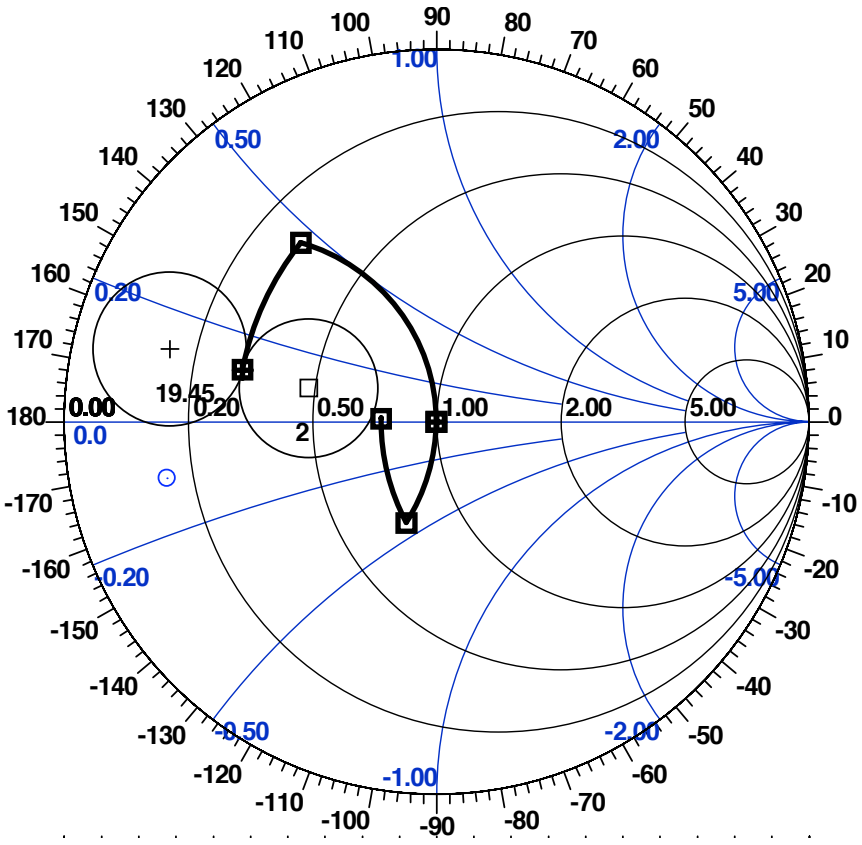


Figure 3.2: Matching network design: .91V bias LNA, 50Ω characteristic impedance.

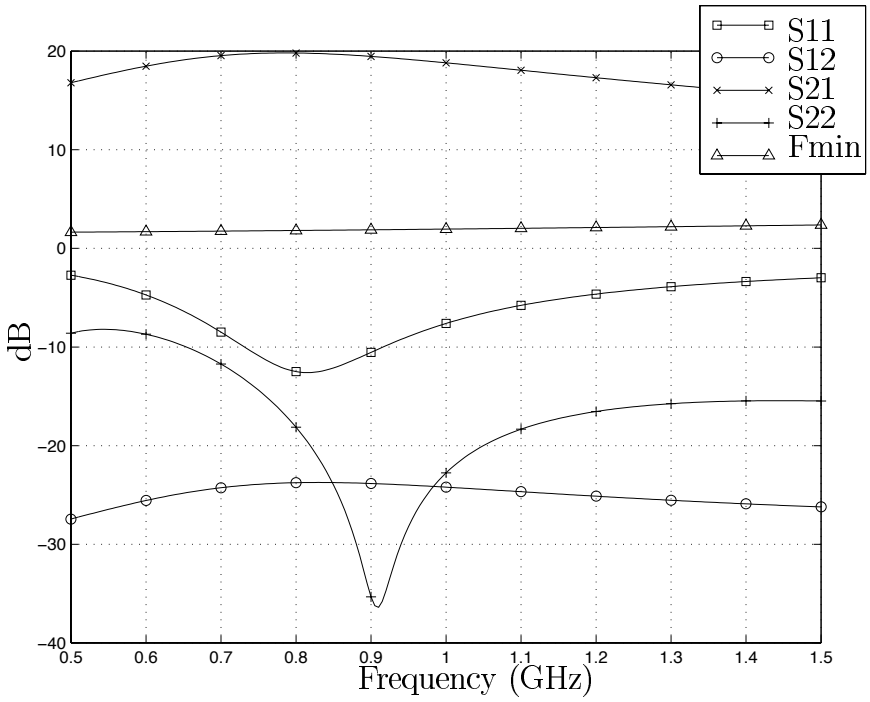


Figure 3.3: Matching network response: .91V bias LNA.

3.3 Nonlinear Circuit Characteristics

The three main elements of nonlinear circuit characterization are the amplitude of the signals involved, the degree (or strength) of nonlinearity of the system, and the memory of the circuit. For sake of discussion the nonlinear circuit we will consider is a microwave amplifier, but the nonlinear characterizations discussed can be equally well applied to mixers, multipliers, varactors, diodes and any other microwave circuit or device.

The principle that we begin with is that all circuits are nonlinear [18]. The degree of nonlinearity is a function of the nonlinearity of the component devices and the designed operating parameters of the circuit. A common assumption (and mostly a valid one) is that a microwave device is linear for "small signal" operation. While there is no precise definition of what "small signal" means, it is generally accepted that small

signal device operation is such that the input signals impart a very small perturbation to the designed operating point. In the case of a front end microwave amplifier in the absence of strong interferers the received power available from the antenna can be as small as -110dBm. This signal level represents a signal amplitude of 1.0μV across a 50Ω matched load. While we generally do not have an absolute threshold where we call one signal "small" and another signal "large", it can reasonably be assumed that the operating point for most real devices is not changed appreciably by perturbing their input by this small amount. Put another way, the transfer characteristics of devices and circuits can be assumed to be linear for inputs in a small neighborhood around their operating point.

While small-signal operation leads circuit designers to make simplifying and mostly valid assumptions about device linearity when the operating point is not moved appreciably, if the system is driven in such a manner that the bias point changes appreciably the output of the device is no longer linearly related to the input. For example, -10dBm of available power driven into a matched 50Ω amplifier generates a voltage at the input of the amplifier of approximately 0.1V. This significant swing in input voltage can no longer be termed "small" in relation to the input bias point of the device, and is thus deemed "large".

The second descriptive quality of a nonlinear device is the "strength" or degree of the nonlinearity which depends on the input conditions. Microwave circuits are generally described as linear, weakly-nonlinear, or strongly nonlinear. Due to small signal assumptions, the definition of linear can generally be lumped into weakly-nonlinear. The characteristic quality of the circuit/device is (again) subject to interpretation. As examples of extremes of each case consider an ideal limiter and a small-signal LNA. The small signal amplifier is generally biased in the active region of the underlying

nonlinear device. At this operating point the assumptions of small signal linearity are usually made, but if the input signal is such that the input/output relation has a gentle curve rather than a straight line relationship, the amplifier is said to be weakly nonlinear. This is most often observed in front-end amplifiers with pseudo-large signal interferers or in power amplifiers operating near saturation. In contrast to these weakly-nonlinear examples a hard limiter is very clearly a strongly nonlinear device. It's input/output relation approaches that of a step function. In a less extreme case an amplifier driven with a very strong input signal is likely to have its device operating point driven into the saturation or cutoff regions where the output signal appears "clipped". This type of operation is deemed "strongly" nonlinear.

The remaining descriptive characteristic of a nonlinear microwave circuit is its memory. In practice, any change in input signal shifts the operating point of a nonlinear device. The nonlinear element of a device is generally characterized by its so-called V/I curve; the output current is nonlinearly related to some control voltage (whether it be a voltage present at the terminals of the device or the voltage across an internal element). The operating point of the device is set by the load line of the linear portions of the circuit intercepting the V/I characteristic. The rate at which the operating point of the device changes in relation to the control voltage is crucial to the analysis of the system. Unfortunately, in most microwave circuits the matching, bias and feedback networks introduce reactive elements that retain charge and can cause the voltages and currents in the circuit to become dependent on past inputs. This behavior is evidenced by a circuit exhibiting amplitude dependent phase shifts (AM/PM). If it is assumed that the operating point of the nonlinear device reacts very quickly (if not instantaneously) to changes in its control voltage, the circuit is called *quasi-static* or *quasi-memoryless* [18]. Note this does not imply that the amalgam of

elements in the circuit is memoryless, merely that the change of the operating point of the nonlinear device is not dependent on some long time constant.

The three characteristics discussed: signal size, strength of nonlinearity, and circuit memory are commonly used to classify circuits and to select the type of analysis or model used.

3.4 Nonlinear Model Characterization

In this section we provide an overview of the existing methods of nonlinear microwave circuit analysis. This section is not meant to be an exhaustive and detailed discussion, rather an overview of the strengths, weaknesses, and references for each technique. Existing techniques are classified into three main types based on their underlying model: analytical, circuit, and behavioral. Analytical modeling is done by selecting a function that is fairly easy to analyze and determining "best-fit" parameters to match some measured circuit or device response; it is the least accurate model. Circuit modeling is performed by analyzing the node currents and voltages in an iterative fashion; while precise, this model can involve very long computer run times and deterministic outputs. Behavioral modeling attempts to relate the output of a system to its input. The behavioral model is usually optimized for analyzing a specific portion of the system output. The remainder of this section details some of the common models available in the literature and texts.

3.4.1 Analytical

Analytical models are chosen as a first pass approximation to the AM/AM performance (and sometimes the IIO performance)[1] of a circuit when more detailed models are unavailable. While these models can provide a gross match to the performance of a given amplifier if the parameters of the model are fit to measured data, they usually introduce another level of approximation due to their parametric nature and thus are inaccurate. Five common analytical models [26] used in modeling different devices are given below:

$$\text{Cann:} \qquad y(t) = G\frac{L}{\sqrt[s]{1+(\frac{L}{|x(t)|})^s}}, \tag{3.5}$$

$$\text{sine:} \qquad y(t) = \frac{1}{32}(38\sin(x(t)) + 7\sin(3x(t)) + 5\sin(5x(t))), \tag{3.6}$$

$$\text{arctan:} \qquad y(t) = \frac{2}{\pi}\arctan(\frac{\pi x(t)}{2}), \tag{3.7}$$

$$\text{tanh:} \qquad y(x) = \tanh(x(t)), \tag{3.8}$$

$$\text{erf:} \qquad y(x) = \frac{2}{\pi}\int_0^x \exp(-x^2(t))dx. \tag{3.9}$$

Recent work [27,28] suggests that while these analytical models may provide a workable approximation to AM/AM curves, the results obtained for intermodulation product circuit-based analysis are not accurate due to their derivatives not matching the derivatives of the underlying device characteristic or their derivatives not existing. As discussed in [29], to determine the n'th order intermodulation products in a circuit based model, the device model must accurately match the first n derivatives of the

[1] It is common in the literature and texts to discuss the AM/AM performance of an amplifier. This characteristic is defined as the fundamental output amplitude (envelope) for a given fundamental input amplitude. Determining this relationship can be done using simulation, measurement, or analysis of a model. This curve should not be confused with the instantaneous input/output (IIO) relationship. The IIO relates the input voltage or current to output voltage or current at a given instant in time. Note that in the literature the AM/AM is sometimes referred to as the instantaneous envelope transfer function [24] [25], not to be confused with the IIO.

devices V/I characteristic[2].

3.4.2 Circuit

Perhaps one of the most flexible and realistic models of the performance of the microwave circuit is done by computer simulation of circuit currents and voltages. Computer aided design (CAD) for microwave circuits is generally is performed in one of two ways: time domain (TD) or harmonic balance (HB). In arguably the most ubiquitous TD simulator, SPICE [32–34] characterizes the node voltages and currents of the linear parts of the circuit and solves for the operating point of the nonlinear devices at given time intervals (or so called time-steps). The solution is typically an iterative Newton-Raphson descent and can experience numerical convergence problems, but these can be minimized by appropriate choice of time step. Long run times are also known for multiple closely spaced input frequencies.

The other well known CAD method is the HB method. In this method the linear sections of the circuit are modeled in the frequency domain and the nonlinear sections are modeled in the time domain; details of this method can be found in [18]. The accuracy and speed of this method is dependent on the number of harmonics tracked by the program and the numerical accuracy of the underlying FFT. In the case of a large number of harmonics, the computational complexity can grow exponentially [27], so judicious selection of analysis parameters is a must for efficient design. Common CAD software that use this method for analysis include LIBRA and HARMONICA. There are many different flavors of HB, a brief survey can be found in [35]. The primary difference between the various methods are in the multi-dimensional opti-

[2]However, this is said with some caution in that authors have accurately modeled n'th order intermodulation with higher order polynomials in behavioral modeling [18, 24, 25, 27, 30, 31] of the circuit response rather than the device response.

mization method used to reach a stable solution. The different HB methods converge at different rates for different systems; it is the job of the particular CAD program to manage the task of method selection.

The advantage of circuit level modeling is the fact that the three aforementioned characteristics of nonlinear circuits: signal size, nonlinearity strength, and circuit memory are generally built into the circuit device models. The results are then dependent on the accuracy of the underlying transistor device models used. For instance, at very high frequencies the Ebers-Moll model exhibits more realistic response than the standard Gummel-Poon model used in most CAD programs [36] due to non-negligible signal wave propagation times. Thus, one must be cautious as to the underlying models in a TD simulator used at high frequencies. However, at the frequency we are considering the inter-device and intra-device transit times are arguably negligible due to the relatively long wavelength as compared to the physical circuit dimension.

While circuit modeling offers some very positive advantages, its drawbacks include the inability to give results for inputs with statistically random distributions, e.g. a sum of sinusoids with random phases, without exhaustive variation of input parameters. HB simulators have exponentially complexity in the number of harmonics tracked and thus the intermodulation products desired. TD simulators generally have large run time especially with systems with closely spaced input frequencies. For some applications, these disadvantages can lead to prohibitively long simulation times.

3.4.3 Behavioral

The complexity and iterative disadvantages of the above analytical and circuit models illustrate the need for a flexible and fast modeling scheme. The many types of behavioral modeling fulfill this goal, with a few caveats. The distinction between the major behavioral models lie in their ability retain accuracy in the face of the difference in the strength of nonlinearities, the signal size and the circuit memory.

3.4.4 Behavioral: Power Series Method

The most basic, and possibly well explored, method of modeling a nonlinearity is by representing its IIO and sometimes its AM/AM characteristic as a power series. Depending on the size of the input signal and the gain of the model, this series may be a Taylor Series in the case of small signal/small perturbation or some other polynomial fit in the case of a larger signal dynamic range. It is important to note that the selected series must only be tight in the signal range of interest [29]. The Taylor series polynomial has the quality of fitting a curve very well in the neighborhood around a given operating point, with the error of the fit being pushed out to the extremes. However, a Taylor polynomial has problems fitting a curve where there are sharp corners, as in the case in a strongly nonlinear circuit. Other polynomial fits are found to have better properties when applied to a larger interval; specifically the Chebychev polynomials are found to give an adequate fit to larger signal swings and sharper nonlinearities [24, 27, 30, 37].

As mentioned in section 3.4.1, there is some question as to the accuracy of the intermodulation information that is derived from a power series fit to either the IIO or the AM/AM curves. Some authors, including Staudinger in [24] and Loyka in [30],

show that the tightness of a higher order power series fit determines the accuracy of the result, while other authors including [29] claim that the n'th order IM products are a function of the n'th order derivatives of the curve fit and that a power series has a poor higher order derivative fit. Different, non-power series, fits to the IIO and AM/AM characteristic functions have been proposed in [38] and [24]. In [38] a model based on a combination of sinusoids is proposed to fit the IIO. A fit of the form:

$$y(A,B) = \left[a_0 x + \sum_{i=1}^{n} a_i \sin(ix)\right] \tanh(\alpha B), \tag{3.10}$$

where x, A, and B functions of transistor parameters and voltages, is said to match the derivatives of the IIO more closely than the standard power series fits. In [24] a Bessel series fit is proposed that does not suffer from the classical ill-conditioning found in higher order polynomial fits.

The most basic power series fit to an arbitrary curve is probably the method of least squares. Unfortunately, the system of equations that needs to be solved in this method involves inverting a matrix that resembles a Hilbert matrix, and is therefore an ill-posed problem for numeric evaluation. This behavior is primarily as a result of the basis functions being the monomials (x,x^2,x^3 ...) which are relatively linearly *dependent* in the range $(0,1)$. The solution to this problem is to use a Chebychev polynomial fit, which has orthogonal Chebychev polynomials as basis functions. It is noted that the polynomial model in [24] suffers from the ill conditioning problems, a problem that Chebychev polynomials do not suffer from [39]. Since the polynomial model chosen does provide comparable performance to the Bessel fit before the ill-conditioning error dominates, we assume that the Chebychev polynomials will continue to track the performance of the Bessel fit.

The debate as to the accuracy of power series fit stems from the use of the fit. If the actual device characteristics are modeled with a power series then the higher

order derivatives are important. If, however, the transfer response of the entire circuit is being studied the higher order derivatives are not as critical [40] and the circuit response can be accurately modeled by a power series.

The power series can be used to model both large and small signal circuits under strong or weakly nonlinear operation. Their strength is in the rapidity that they can be incorporated into a simulation and in their dynamic range. The AM/AM and IIO functions that these series fit can be determined experimentally or from an appropriate simulation. Since the power series (and likewise any of the other fitting methods) are functional, they are not able to fit a nonlinear system with a large amount of memory. The characteristic IIO function in this case is not one-to-one and the definition of a series "fit" can not be meaningfully defined.

3.4.5 Behavioral: Volterra Series Method

The Volterra Series (VS) is a useful tool for the analysis of small-signal, weakly nonlinear systems with memory. The interested reader can find a detailed discussion of this technique in [18]. The basic principal of the VS can be summed up by Wiener's suggestion in [41] that a weakly nonlinear system under appropriately small-signal conditions has the following response:

$$w(t) = \int \mathrm{h}_1(\tau_1)\,\mathrm{s}(t-\tau_1)d\tau_1 \tag{3.11}$$

$$+ \int\int \mathrm{h}_2(\tau_1,\tau_2)\,\mathrm{s}(t-\tau_1)\,\mathrm{s}(t-\tau_2)d\tau_1 d\tau_2 \tag{3.12}$$

$$+ \int\int \mathrm{h}_3(\tau_1,\tau_2,\tau_3)\,\mathrm{s}(t-\tau_1)\,\mathrm{s}(t-\tau_2)\,\mathrm{s}(t-\tau_3)d\tau_1 d\tau_2 d\tau_3 + \ldots \tag{3.13}$$

The functions $\mathrm{h}_n(\tau_1,\tau_2,\ldots,\tau_n)$ are called the n'th order nonlinear impulse response or the n'th order kernel. Determining these functions and the appropriate frequency domain representations are the crux of the VS analysis technique. While this technique

is deemed a powerful one for representing nonlinearities with memory, the complexities of determining higher-order kernels usually keep this analysis from being used for large-signal and strongly nonlinear systems [42].

3.4.6 Behavioral: Quadrature Modeling Method

The quadrature modeling (QM) technique breaks the nonlinear response of a circuit into an in-phase and a quadrature-phase nonlinearity, and thus is able to account for a greater range of nonlinear effects than a single nonlinear function. This modeling technique's specific advantage over a single nonlinear function is that is able to account for both AM/AM and AM/PM nonlinear effects in quasi-memoryless systems.

Understanding of the QM is necessary in understanding the so called "Instantaneous" methods found in Sections 3.4.8 and 3.5. The block diagram of this method is illustrated in Figure 3.4. The development of the theory of quadrature modeling follows from [40, 43, 44].

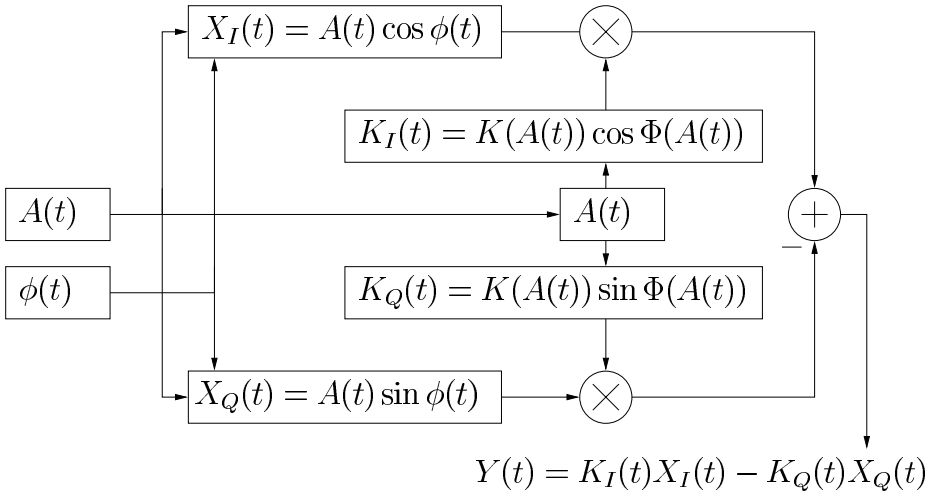


Figure 3.4: Quadrature Model Block Diagram.

Necessary in understanding of the following derivation is the idea of so called

output zones. The first zone output is the component of the output signal that has the carrier as a component. The n'th zone output are the components of the output signal that have the $n-1$'th harmonic of the carrier as a component. For example, if the carrier is defined to be ω_0, and component of the output that possesses $\cos\omega_0 t$ as its carrier frequency is a first zone output while any component of the output that possesses $\cos 2\omega_0 t$ as its carrier frequency is a second zone output.

Consider a real bandpass signal:

$$\mathrm{x}(t) = \mathrm{Re}\left\{\mathrm{A}(t)\exp(j\omega_0 t + j\phi(t))\right\}, \tag{3.14}$$

$$= \mathrm{A}(t)\cos(\omega_0 t + \phi(t)), \tag{3.15}$$

where $\mathrm{A}(t)$ and $\phi(t)$ are the amplitude and phase modulation and ω_0 is the carrier frequency. The complex envelope of this signal can be expressed as:

$$\mathrm{Env}\left\{\mathrm{x}(t)\right\} = \mathrm{A}(t)\exp\left[j\phi(t)\right], \tag{3.16}$$

$$= \mathrm{A}(t)\cos\phi(t) + j\,\mathrm{A}(t)\sin\phi(t). \tag{3.17}$$

The corresponding low-pass in-phase and quadrature signals are denoted:

$$X_I(t) = A(t)\cos\phi(t) \tag{3.18}$$

$$X_Q(t) = A(t)\sin\phi(t). \tag{3.19}$$

Thus

$$x(t) = X_I(t)\cos\omega_0 t - X_Q(t)\sin\omega_0 t. \tag{3.20}$$

The nonlinear effects, or so-called transfer factors, are dependent on the envelope amplitude and can be expressed as:

$$K_I(A(t)) = K(A(t))\cos\Phi(A(t)), \tag{3.21}$$

$$K_Q(A(t)) = K(A(t))\sin\Phi(A(t)), \tag{3.22}$$

where $K(A(t))$ and $\Phi(A(t))$ are the AM/AM and AM/PM responses of the circuit. These functions are typically measured or simulated using a CAD program and are restricted to the first zone output; second zone and higher outputs are not considered. The output due to each transfer factor is multiplicative:

$$\begin{aligned} Y_I(t) &= K_I(t)X_I(t), \\ &= A(t)K(A(t))\cos(\Phi(A(t))\cos(\phi(t)), \end{aligned} \tag{3.23}$$

and

$$\begin{aligned} Y_Q(t) &= K_Q(t)X_Q(t) \\ &= A(t)K(A(t))\sin(\Phi(A(t))\sin(\phi(t)). \end{aligned} \tag{3.24}$$

The combined output is:

$$\begin{aligned} Y(t) &= Y_I(t) - Y_Q(t), \\ &= A(t)K(A(t))\cos(\Phi(A(t))\cos(\phi(t)) - A(t)K(A(t))\sin(\Phi(A(t))\sin(\phi(t)), \\ &= A(t)K(A(t))\cos(\phi(t)) + \Phi(A(t)). \end{aligned} \tag{3.25}$$

Similarly, we note that the system can also be expressed as:

$$\begin{aligned} Y(t) &= \mathrm{Re}\left\{A(t)\,\mathrm{e}^{j\phi}\right\}\mathrm{Re}\left\{K(A(t))\,\mathrm{e}^{j\Phi(A(t))}\right\} - \mathrm{Im}\left\{A(t)\,\mathrm{e}^{j\phi}\right\}\mathrm{Im}\left\{K(A(t))\,\mathrm{e}^{j\Phi(A(t))}\right\}, \\ &= \mathrm{Re}\left\{A(t)\,\mathrm{e}^{j\phi}\,K(A(t))\,\mathrm{e}^{j\Phi(A(t))}\right\}, \\ &= A(t)K(A(t))\cos(\phi(t)) + \Phi(A(t))). \end{aligned} \tag{3.26}$$

To determine the actual first zone output, $y_1(t)$, the complex carrier is multiplied by

the complex envelope of (3.26):

$$\begin{aligned} y_1(t) &= \mathrm{Re}\left\{A(t)K(A(t))\,\mathrm{e}^{j(\phi+\Phi(A(t)))}\,\mathrm{e}^{j\omega_0 t}\right\}, \\ &= \mathrm{Re}\left\{A(t)K(A(t))\ [\cos(\Phi(A(t))) + j\sin(\Phi(A(t)))]\right. \\ &\qquad \left.\times\,[\cos(\omega_0 t+\phi) + j\sin(\omega_0 t+\phi)]\right\}, \\ &= A(t)\ K(A(t))\cos(\Phi(A(t)))\ \cos(\omega_0 t+\phi) \\ &\qquad - A(t)\ K(A(t))\sin(\Phi(A(t)))\ \sin(\omega_0 t+\phi). \end{aligned} \tag{3.27}$$

Equation (3.27) shows that the first zone output of the system is the equivalent of an envelope times the $\cos\omega_0 t$ portion of the carrier coupled with an envelope times the $\sin\omega_0 t$ portion of the carrier.

Zone n nonlinearities may be determined by measuring/simulating the appropriate zone one input vs. zone n output envelope transfer factors. For example, if the zone two output envelope vs. zone one input envelope relationship is measured as $K_2(A)$ and $\Phi_2(A)$ the envelope nonlinearities can be expressed as:

$$\begin{aligned} y_2(t) &= \mathrm{Re}\left\{A(t)K_2(A(t))\,\mathrm{e}^{j(2\phi+\Phi_2(A(t)))}\,\mathrm{e}^{j2\omega_0 t}\right\}, \\ &= \mathrm{Re}\left\{A(t)K_2(A(t))\ [\cos(\Phi_2(A(t))) + j\sin(\Phi_2(A(t)))]\right. \\ &\qquad \left.\times\,[\cos(2\omega_0 t+2\phi) + j\sin(2\omega_0 t+2\phi)]\right\}, \\ &= A(t)K_2(A(t))\ \cos(\Phi_2(A(t)))\ \cos(2\omega_0 t+2\phi) \\ &\qquad - A(t)K_2(A(t))\ \sin(\Phi_2(A(t)))\ \sin(2\omega_0 t+2\phi). \end{aligned} \tag{3.28}$$

Use will be made of (3.27) and (3.28) in Section 3.4.8 to develop the Instantaneous Quadrature Method.

The combined system where the in-phase signal envelope is acted on by the in-phase nonlinearity and the quadrature phase signal envelope is acted on by the quadrature nonlinearity is depicted in Figure 3.4. The main drawback of the QM is

that the output is only determined in the measured zones, thus it is unable to model higher order harmonics without additional circuit measurements or simulations. The outputs of this system can be measured or simulated for different harmonic zones to increase its utility, but this represents a large amount of measurement or computing power to properly characterize the full spectrum of system output. Additionally, the modulation signal must be narrow-band since the AM/AM and AM/PM response is assumed constant over the frequency range of interest so it is unable to model matching networks or bias-coupling effects [40].

3.4.7 Behavioral: Discrete Method

In the Discrete Method, systems are typically represented in the time domain for nonlinear elements and the frequency domain for the linear elements. Use of the FFT and IFFT is made to change between these domains. Because this type of analysis is heavily dependent on the FFT it suffers from the following drawbacks: the numerical resolution of the FFT can introduce errors in systems where both large and small signals are present, the selection of the sampling rate of the signals can lead to aliasing of intermodulation products and harmonics into the band of interest and can significantly slow the simulation, and the output of the system is determined by the exact input which in turn requires many simulations to determine effects of varied parameters. Other problems arise with system causality, especially in systems or operations that require an integration over all time (such as the Hilbert transform). The truncation of these operations coupled with the dependence on the FFT introduce error and force long run times for accurate results. Additionally, if the system analyzed consists of a large number of closely spaced input tones or the nonlinear effects are significant, equivalent baseband analysis may may not be possible

and simulation at passband may require an inordinately large number of cycles for an accurate answer.

However, Discrete Method modeling of a nonlinear circuit under the conditions of large sampling rate and fast processing is able to treat wide-band systems and determine even and odd order intermodulation and harmonic effects. This is, of course contingent upon the numerical accuracy of the processing and the other disadvantages listed.

3.4.8 Behavioral: Instantaneous Quadrature Method

A method is needed that preserves the advantages of the power series and discrete analysis, namely the ability to analyze wide-band, strongly nonlinear, large signal circuits while not prejudicing the analysis of weakly nonlinear, small signal, narrow band circuits. These advantages can be coupled with the ability of the QM to analyze circuits that exhibit a nontrivial AM/PM response. In [27] and [40] the Instantaneous Quadrature Method (IQM) of Figure 3.5 is developed. The principal difference

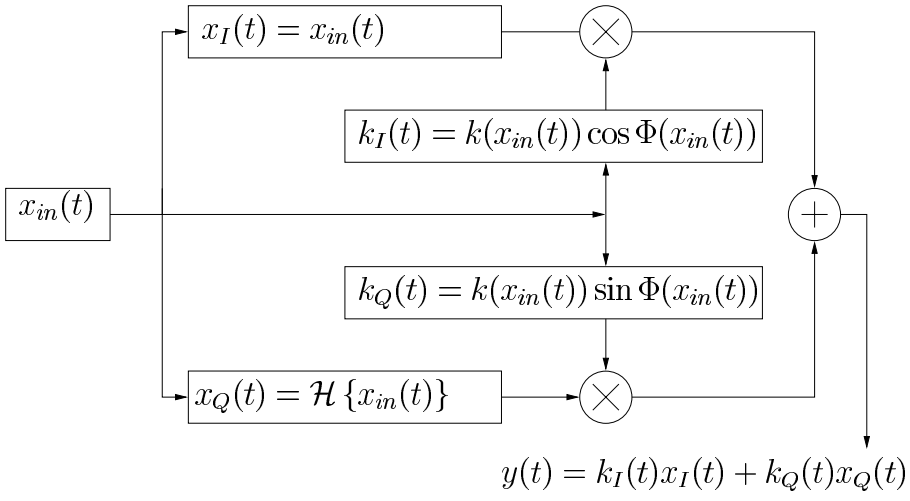


Figure 3.5: Instantaneous Quadrature Model Block Diagram.

between the IQM and the QM is the nonlinearities in the IQM are dependent on the

instantaneous value of the the input signal rather than the envelope of the signal. Following Loyka's development in [40] and [45] the real and quadrature parts of the input signal are determined as follows:

$$x_{in}(t) = \mathrm{Re}\left\{A\,\mathrm{e}^{j\phi}\,\mathrm{e}^{j\omega t}\right\}, \tag{3.29}$$

$$x_I(t) = x_{in}(t) = A\cos(\omega t + \phi), \tag{3.30}$$

$$x_Q(t) = \mathcal{H}\left\{x_{in}(t)\right\} = -A\sin(\omega t + \phi), \tag{3.31}$$

where $\mathcal{H}$ is the Hilbert transform operator. The block diagram of this method/model can be seen in Figure 3.5. The goal of this method is to determine the instantaneous transfer factors, k_I and k_Q from the corresponding envelope transfer factors K_I and K_Q where the input is a tone of constant amplitude and phase: $x_{in}(t) = A\cos\omega t$. The transfer factors can be separated into even and odd parts:

$$k_I(t) = k_{I_e}(t) + k_{I_o}(t), \tag{3.32}$$

$$k_Q(t) = k_{Q_e}(t) + k_{Q_o}(t), \tag{3.33}$$

and the determination of these portions of the nonlinearities can proceed separately. The output, $y(t)$, can be written as:

$$\begin{aligned}
y(t) &= k_I(x_{in}(t))x_{in}(t) + k_Q(x_{in}(t))\mathcal{H}\left\{x_{in}(t)\right\}, \\
&= Ak_I(A\cos\omega t)\cos\omega t - Ak_Q(A\cos\omega t)\sin\omega t, \\
&= A\left[k_{I_e}(A\cos\omega t) + k_{I_o}(A\cos\omega t)\right]\cos\omega t \\
&\quad - A\left[k_{Q_e}(A\cos\omega t) + k_{Q_o}(A\cos\omega t)\right]\sin\omega t, \\
&= A\left[k_{I_e}(A\cos\omega t)\cos\omega t + k_{I_o}(A\cos\omega t)\cos\omega t\right. \\
&\quad \left. -k_{Q_e}(A\cos\omega t)\sin\omega t - k_{Q_o}(A\cos\omega t)\sin\omega t\right].
\end{aligned} \tag{3.34}$$

Since $k_{I_e}(A\cos\omega t)\cos\omega t$ and $k_{I_o}(A\cos\omega t)\cos\omega t$ are even periodic functions of t and $k_{Q_e}(A\cos\omega t)\sin\omega t$ and $k_{Q_o}(A\cos\omega t)\sin\omega t$ are odd periodic functions of t they can

be fully represented by their equivalent Fourier cosine and sine series. Thus,

$$\begin{aligned} y(t) = A \Bigg[& \frac{Y_{I_e 0}}{2} + \sum_{n=1}^{\infty} Y_{I_e n} \cos n\omega t + \frac{Y_{I_o 0}}{2} + \sum_{n=1}^{\infty} Y_{I_o n} \cos n\omega t \\ & - \frac{Y_{Q_e 0}}{2} - \sum_{n=1}^{\infty} Y_{Q_e n} \sin n\omega t - \frac{Y_{Q_o 0}}{2} - \sum_{n=1}^{\infty} Y_{Q_o n} \sin n\omega t \Bigg], \\ = A \Bigg[& \frac{Y_{I_e 0}}{2} + \frac{Y_{I_o 0}}{2} - \frac{Y_{Q_e 0}}{2} - \frac{Y_{Q_o 0}}{2} \\ & + \sum_{n=1}^{\infty} [Y_{I_e n} + Y_{I_o n}] \cos n\omega t - \sum_{n=1}^{\infty} [Y_{Q_e n} + Y_{Q_o n}] \sin n\omega t \Bigg], \end{aligned} \tag{3.35}$$

where, after the substitution $\theta = \omega t$:

$$Y_{I_e n} = \frac{1}{\pi} \int_{\frac{-\pi}{2}}^{\frac{3\pi}{2}} k_{I_e}(A \cos\theta) \cos\theta \cos n\theta \; d\theta, \tag{3.36}$$

$$Y_{I_o n} = \frac{1}{\pi} \int_{\frac{-\pi}{2}}^{\frac{3\pi}{2}} k_{I_o}(A \cos\theta) \cos\theta \cos n\theta \; d\theta, \tag{3.37}$$

$$Y_{Q_e n} = \frac{1}{\pi} \int_{\frac{-\pi}{2}}^{\frac{3\pi}{2}} k_{Q_e}(A \cos\theta) \sin\theta \sin n\theta \; d\theta, \tag{3.38}$$

$$Y_{Q_o n} = \frac{1}{\pi} \int_{\frac{-\pi}{2}}^{\frac{3\pi}{2}} k_{Q_o}(A \cos\theta) \sin\theta \sin n\theta \; d\theta. \tag{3.39}$$

Combining cosine terms from (3.27), (3.28) and (3.35) yield:

$$[Y_{I_e 1} + Y_{I_o 1}] = K(A) \cos\Phi(A), \tag{3.40}$$

$$[Y_{I_e 2} + Y_{I_o 2}] = K_2(A) \cos\Phi_2(A). \tag{3.41}$$

Similarly, combining sine terms from (3.27), (3.28) and (3.35) yield:

$$[Y_{Q_e 1} + Y_{Q_o 1}] = K(A) \sin\Phi(A), \tag{3.42}$$

$$[Y_{Q_e 2} + Y_{Q_o 2}] = K_2(A) \sin\Phi_2(A). \tag{3.43}$$

Solving for the Y terms:

$$Y_{I_e 1} = \frac{4}{\pi} \int_{0}^{\frac{\pi}{2}} k_{I_e}(A \cos\theta) \cos^2\theta \; d\theta, \tag{3.44}$$

$$Y_{I_e 2} = 0, \tag{3.45}$$

$$Y_{I_o 1} = 0, \tag{3.46}$$

$$Y_{I_o 2} = \frac{4}{\pi} \int_0^{\frac{\pi}{2}} k_{I_o}(A\cos\theta)\cos\theta\cos 2\theta \, d\theta, \tag{3.47}$$

$$Y_{Q_e 1} = \frac{4}{\pi} \int_0^{\frac{\pi}{2}} k_{Q_e}(A\cos\theta)\sin^2\theta \, d\theta, \tag{3.48}$$

$$Y_{Q_e 2} = 0, \tag{3.49}$$

$$Y_{Q_o 1} = 0, \tag{3.50}$$

$$Y_{Q_o 2} = \frac{4}{\pi} \int_0^{\frac{\pi}{2}} k_{Q_o}(A\cos\theta)\sin\theta\sin 2\theta \, d\theta. \tag{3.51}$$

Thus,

$$K_I(A) = K(A)\cos\Phi(A) = Y_{I_e 1}, \tag{3.52}$$

$$K_{I2}(A) = K_2(A)\cos\Phi_2(A) = Y_{I_o 2}, \tag{3.53}$$

$$K_Q(A) = K(A)\sin\Phi(A) = Y_{Q_e 1}, \tag{3.54}$$

$$K_{Q2}(A) = K_2(A)\sin\Phi_2(A) = Y_{Q_o 2}. \tag{3.55}$$

(3.52), (3.53), (3.54), and (3.55) are known as integral equations of the first kind (IFK) and can be re-written in their Fredholm and Volterra integral equation forms via two variable changes: $t = \cos\theta$ or $t = \sin\theta$ and $x = At$. These transformations follow:

$$K_I(A) = \frac{4}{\pi} \int_0^{\frac{\pi}{2}} k_{I_e}(A\cos\theta)\cos^2\theta \, d\theta, \tag{3.56}$$

$$= \frac{4}{\pi} \int_0^1 k_{I_e}(At) \frac{t^2}{\sqrt{1-t^2}} \, dt, \tag{3.57}$$

$$= \frac{4}{\pi} \int_0^A k_{I_e}(x) \frac{x^2}{A^2\sqrt{A^2-x^2}} \, dx, \tag{3.58}$$

$$K_Q(A) = \frac{4}{\pi}\int_0^{\frac{\pi}{2}} k_{Q_e}(A\cos\theta)\sin^2\theta\, d\theta, \tag{3.59}$$

$$= \frac{4}{\pi}\int_0^1 k_{Q_e}(At)\sqrt{1-t^2}\, dt, \tag{3.60}$$

$$= \frac{4}{\pi}\int_0^A k_{Q_e}(x)\frac{\sqrt{1-\left(\frac{x}{A}\right)^2}}{A}\, dx,$$

$$= \frac{4}{\pi}\int_0^A k_{Q_e}(x)\frac{\sqrt{A^2-x^2}}{A^2}\, dx, \tag{3.61}$$

$$K_{2I}(A) = \frac{4}{\pi}\int_0^{\frac{\pi}{2}} k_{I_o}(A\cos\theta)\cos\theta\cos 2\theta\, d\theta, \tag{3.62}$$

$$= \frac{4}{\pi}\int_0^1 k_{I_o}(At)\frac{t(2t^2-1)}{\sqrt{1-t^2}}\, dt, \tag{3.63}$$

$$= \frac{4}{\pi}\int_0^A k_{I_o}(x)\frac{2x^3-xA^2}{A^4\sqrt{1-\left(\frac{x}{A}\right)^2}}\, dx,$$

$$= \frac{4}{\pi}\int_0^A k_{I_o}(x)\frac{2x^3-A^2x}{A^3\sqrt{A^2-x^2}}\, dx, \tag{3.64}$$

$$K_{2Q}(A) = \frac{4}{\pi}\int_0^{\frac{\pi}{2}} k_{Q_o}(A\cos\theta)\sin\theta\sin 2\theta\, d\theta, \tag{3.65}$$

$$= \frac{4}{\pi}\int_0^1 k_{Q_o}(At)2t\sqrt{1-t^2}\, dt, \tag{3.66}$$

$$= \frac{4}{\pi}\int_0^A k_{Q_o}(x)\frac{2x}{A^2}\sqrt{1-\left(\frac{x}{A}\right)^2}\, dx,$$

$$= \frac{4}{\pi}\int_0^A k_{Q_o}(x)\frac{2x}{A^3}\sqrt{A^2-x^2}\, dx. \tag{3.67}$$

IQM: Resulting nonlinear factors and transfer functions

The Fredholm forms of (3.57), (3.60), (3.63), and (3.66) are equal to (23a), (23b), (36a) and (36b) of [40] respectively. The method of moments, described in Appendix B, is used to solve the corresponding Volterra equations (3.58), (3.61), (3.64), and

(3.67) for the unknowns k_{I_e},k_{I_o},k_{Q_e}, and k_{Q_o}. Using this method, k_I and k_Q can be determined from the measured quantities $K(A)$, $K_2(A)$, $\Phi(A)$ and $\Phi_2(A)$. Figures 3.6 and 3.7 show the even, odd, and combined transfer factors of the circuit described in Section 3.2. The corresponding I and Q transfer functions are shown in Figure 3.8.

Note that the k_Q curve in Figure 3.8 is a relation, not a function, thus fitting

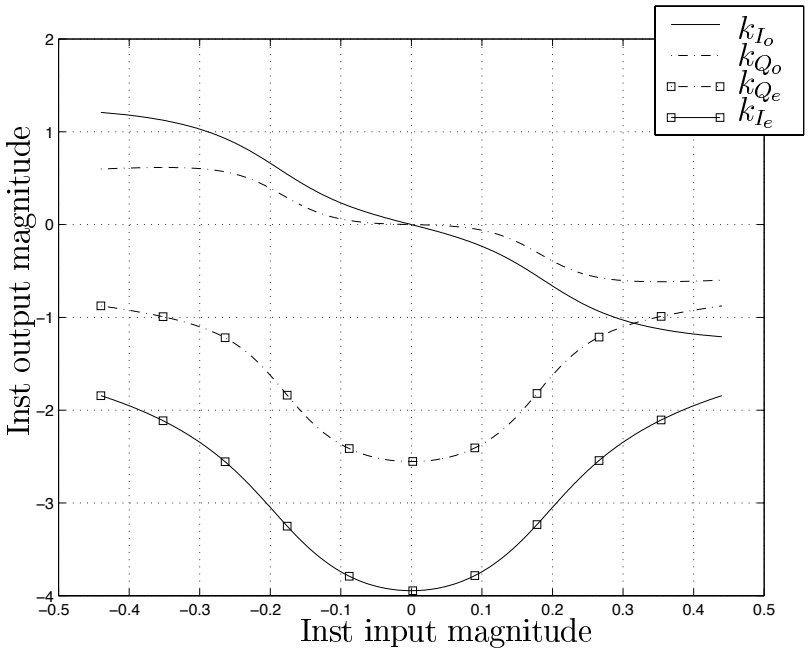


Figure 3.6: IQM: Even and odd parts of k_I and k_Q nonlinear transfer factors.

with a polynomial to determine output intermodulation products is not possible. The open "eye" in k_Q stems from the use of the Hilbert transform of the input signal with the actual input signal in the quadrature side of the nonlinear model:

$$y_{Q_e}(t) = \mathcal{H}\{x_{in}(t)\}\, k_{Q_e}(x_{in}(t)), \tag{3.68}$$

$$= -A \sin \omega t k_{Q_e}(A \cos \omega t). \tag{3.69}$$

This is seen by selecting two values of t, $t = t_1$ and $t = t_2$, such that $\sin \omega t_1 = \sin \omega t_2$, then the corresponding input is equal, but the cosine portion of the output is such

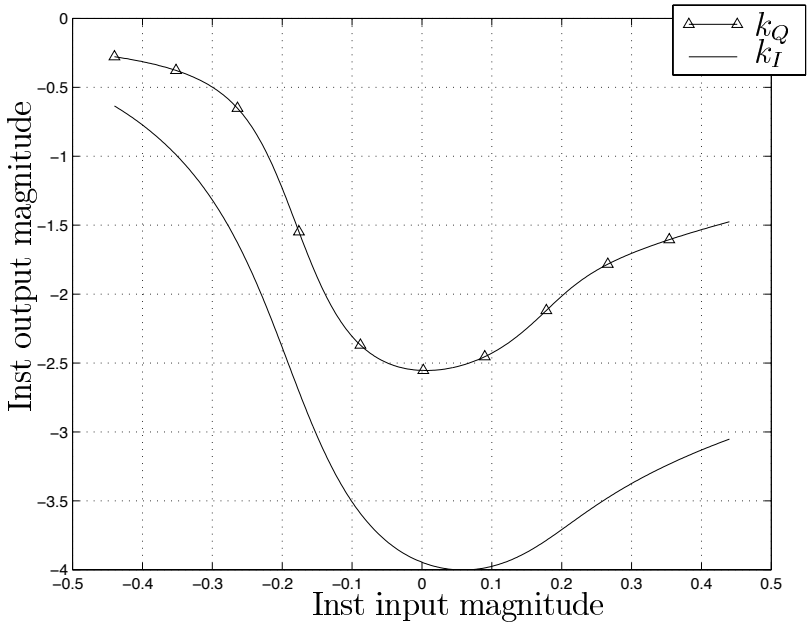


Figure 3.7: IQM: k_I and k_Q nonlinear transfer factors.

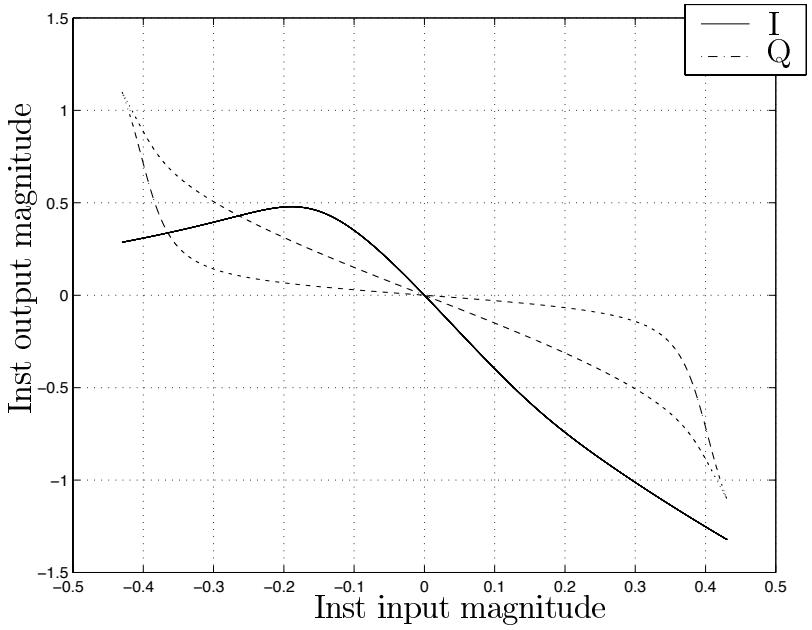


Figure 3.8: IQM: k_I and k_Q nonlinear transfer characteristic.

that $k_{Q_o}(A\cos\omega t_1) \neq k_{Q_o}(A\cos\omega t_2)$ and $k_{Q_e}(A\cos\omega t_1) \neq k_{Q_e}(A\cos\omega t_2)$. Thus, the input/output transfer relationship is not one-to-one, and therefore not a function.

The advantages of the IQM are that it can model large and small signal inputs into strongly nonlinear or weakly nonlinear circuits, using a small number of measurements from the actual circuit. Once the in-phase and quadrature-phase nonlinearities are determined, time domain simulation will provide accurate results. The main disadvantages to this system are twofold. First, the quadrature nonlinear transfer factor may be a function, but coupled with the input signal, the input/output transfer relationship is not a function. Because the quadrature channel IIO relationship is non-functional fitting with a power series is impractical; thus the same drawbacks as discrete modeling apply.

3.5 Modified Instantaneous Quadrature Method

The non-functional relationship between the input and output of the Q-channel of Figure 3.8 presents a significant impediment to modeling the amplifier nonlinearities as polynomial and thereby easily determining the intermodulation effects. While the IQM follows naturally from the QM, a modification to the system can be made to make the nonlinearities more amenable to functional analysis. The proposed Modified Instantaneous Quadrature Method (MIQM) is shown in Figure 3.9. By re-arranging the location of the Hilbert transformer we will show that the analysis and functional relationship of the nonlinearity is preserved.

The problem of determining the even and odd portions of the transfer factors

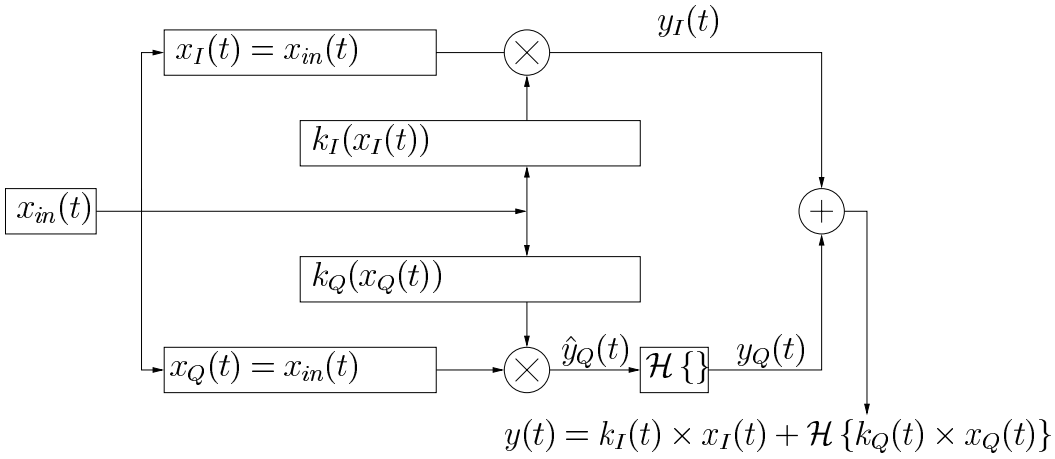


Figure 3.9: Modified Instantaneous Quadrature Model Block Diagram.

proceeds, as before, by examining the output of the system:

$$\begin{aligned}
y(t) &= k_I(x_{in}(t))x_{in}(t) + \mathcal{H}\left\{k_Q(x_{in}(t))x_{in}(t)\right\}, \\
&= Ak_{I_e}(A\cos\omega t)\cos\omega t + Ak_{I_o}(A\cos\omega t)\cos\omega t \\
&\quad + \mathcal{H}\left\{Ak_{Q_e}(A\cos\omega t)\cos\omega t\right\} + \mathcal{H}\left\{Ak_{Q_o}(A\cos\omega t)\cos\omega t\right\}, \\
&= Ak_{I_e}(A\cos\theta)\cos\theta + Ak_{I_o}(A\cos\theta)\cos\theta \\
&\quad + \mathcal{H}\left\{Ak_{Q_e}(A\cos\theta)\cos\theta\right\} + \mathcal{H}\left\{Ak_{Q_o}(A\cos\theta)\cos\theta\right\}, \qquad (3.70)
\end{aligned}$$

with $\theta = \omega t$. The $\mathcal{H}\left\{Ak_{Q_e}(A\cos\theta)\cos\theta\right\}$ and $\mathcal{H}\left\{Ak_{Q_o}(A\cos\theta)\cos\theta\right\}$ terms are odd functions of θ, and can be represented by a Fourier sine series; the $Ak_{I_e}(A\cos\theta)\cos\theta$ and $Ak_{I_o}(A\cos\theta)\cos\theta$ are even functions of θ and can be represented by a Fourier cosine series as the following:

$$\begin{aligned}
y(t) &= A\left[\frac{Y_{I_e0}}{2} + \sum_{n=1}^{\infty} Y_{I_en}\cos n\omega t + \frac{Y_{I_o0}}{2} + \sum_{n=1}^{\infty} Y_{I_on}\cos n\omega t\right. \\
&\quad \left. + \frac{Y_{Q_e0}}{2} + \sum_{n=1}^{\infty} Y_{Q_en}\sin n\omega t + \frac{Y_{Q_o0}}{2} + \sum_{n=1}^{\infty} Y_{Q_on}\sin n\omega t\right], \\
&= A\left[\frac{Y_{I_e0}}{2} + \frac{Y_{I_o0}}{2} + \frac{Y_{Q_e0}}{2}\right. \\
&\quad \left. + \frac{Y_{Q_o0}}{2} + \sum_{n=1}^{\infty}\left[Y_{I_en} + Y_{I_on}\right]\cos n\omega t + \sum_{n=1}^{\infty}\left[Y_{Q_en} + Y_{Q_on}\right]\sin n\omega t\right], \qquad (3.71)
\end{aligned}$$

where

$$Y_{I_e n} = \frac{1}{\pi} \int_{\frac{-\pi}{2}}^{\frac{3\pi}{2}} k_{I_e}(A\cos\theta)\cos\theta\cos n\theta \; d\theta \tag{3.72}$$

$$Y_{I_o n} = \frac{1}{\pi} \int_{\frac{-\pi}{2}}^{\frac{3\pi}{2}} k_{I_o}(A\cos\theta)\cos\theta\cos n\theta \; d\theta \tag{3.73}$$

$$Y_{Q_e n} = \frac{1}{\pi} \int_{\frac{-\pi}{2}}^{\frac{3\pi}{2}} \mathcal{H}\left\{k_{Q_e}(A\cos\theta)\cos\theta\right\}\sin n\theta \; d\theta \tag{3.74}$$

$$Y_{Q_o n} = \frac{1}{\pi} \int_{\frac{-\pi}{2}}^{\frac{3\pi}{2}} \mathcal{H}\left\{k_{Q_o}(A\cos\theta)\cos\theta\right\}\sin n\theta \; d\theta. \tag{3.75}$$

k_{I_e} and k_{I_o} can be determined from (3.72) and (3.73) by the method of moments (see Appendix B), but (3.74) and (3.75) are nested integral equations with unknown kernels. The solution to these nested equations may be obtainable by iterative means, but for our purposes there is an easier way to determine k_{Q_e} and k_{Q_o}.

To determine k_{Q_e}, we compared the output of the MIQM system with the output of the first zone of the QM system output and equated the non-vanishing terms; this is the procedure that leads to the nesting of integral equations. If, however, we equate the product of $x_Q(t)$ and $k_Q(x_Q(t))$ with the *inverse* Hilbert transform of the sine portion of the output from the QM system, the resulting integral equations are solvable. Because the output of the system from the QM is sinusoidal, the inverse Hilbert transform is easy to determine. Mathematically, in the quadrature branch before the Hilbert transformer, we have:

$$\begin{aligned} \hat{y}_Q(t) &= A\cos\theta k_{Q_e}(A\cos\theta) + A\cos\theta k_{Q_o}(A\cos\theta) \\ &= A\left[\frac{Y_{Q_e 0}}{2} + \frac{Y_{Q_o 0}}{2} + \sum_{n=1}^{\infty}\left[Y_{Q_e n} + Y_{Q_o n}\right]\cos n\theta\right]. \end{aligned} \tag{3.76}$$

The $Y_{Q_o 1}$ and $Y_{Q_e 2}$ terms vanish and the remaining terms are equated to the QM

output:

$$\mathcal{H}^{-1}\left\{K_Q(A)\sin\theta\right\} = -K_Q(A)\cos\theta = Y_{Qe1}\cos\theta, \tag{3.77}$$

$$\mathcal{H}^{-1}\left\{K_{Q2}(A)\sin\theta\right\} = -K_{Q2}(A)\cos\theta = Y_{Qo2}\cos\theta, \tag{3.78}$$

and the following quadrature-phase Volterra IFK's are determined:

$$K_{Q1}(A) = -\frac{4}{\pi}\int_0^A k_{Q_e}(x)\frac{x^2}{A^2\sqrt{A^2-x^2}}\,dx, \tag{3.79}$$

$$K_{Q2}(A) = -\frac{4}{\pi}\int_0^A k_{Q_o}(x)\frac{2x^3-A^2x}{A^3\sqrt{A^2-x^2}}\,dx,. \tag{3.80}$$

The in-phase IFK's:

$$K_I(A) = \frac{4}{\pi}\int_0^A k_{I_e}(x)\frac{x^2}{A^2\sqrt{A^2-x^2}}\,dx, \tag{3.81}$$

$$K_{2I}(A) = \frac{4}{\pi}\int_0^A k_{I_o}(x)\frac{2x^3-A^2x}{A^3\sqrt{A^2-x^2}}\,dx, \tag{3.82}$$

are the Volterra IFK form of (3.72) and (3.73).

MIQM: Resulting nonlinear factors and transfer functions

The resulting nonlinear transfer factors and functions determined by the MIQM are shown in Figures 3.10, 3.11, and 3.12. As a result of the re-ordering of the Hilbert transform and the multiplication operators, the transfer function between $x_Q(t)$ and $\hat{y}(t)$ is a functional relationship which submits to polynomial curve fitting. After noting that the signal $\hat{y}(t)$ is the sum of sinusoids and that the Hilbert transform of that sum is trivially performed, our result is achieved.

3.6 Circuits with Memory

Memory in a circuit refers to the retention of energy in any of the components [18]. The assumption of quasi-memoryless circuits is crucial to the development of

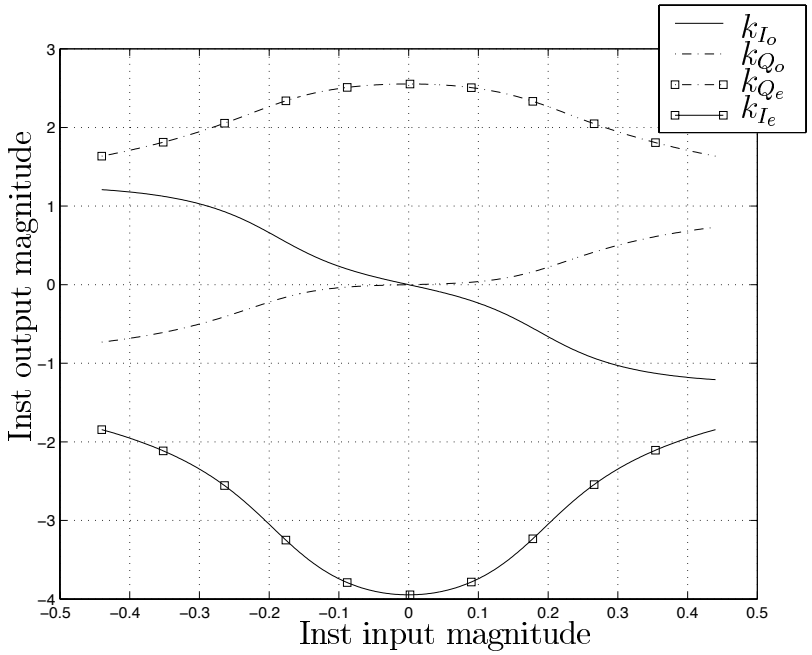


Figure 3.10: MIQM: Even and odd parts of k_I and k_Q transfer factors.

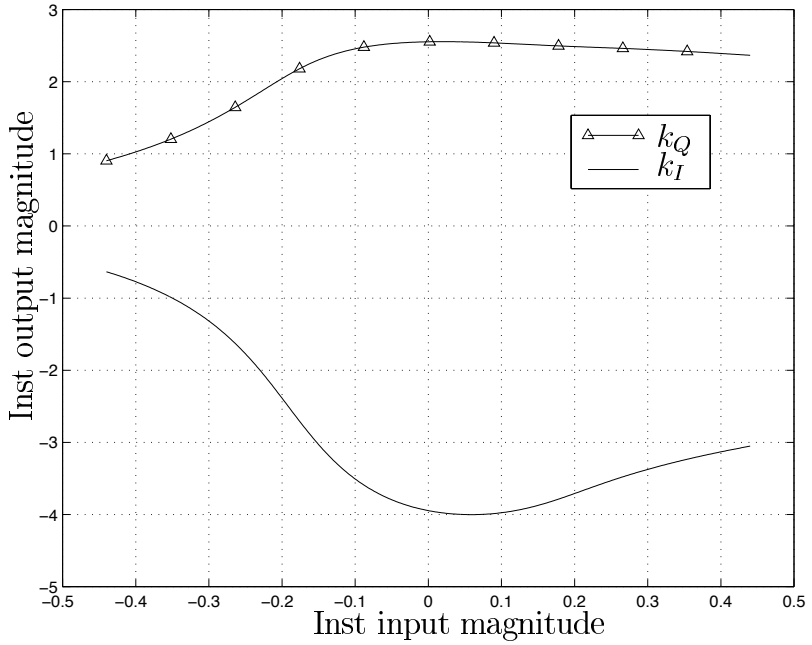


Figure 3.11: MIQM: k_I and k_Q nonlinear transfer factors.

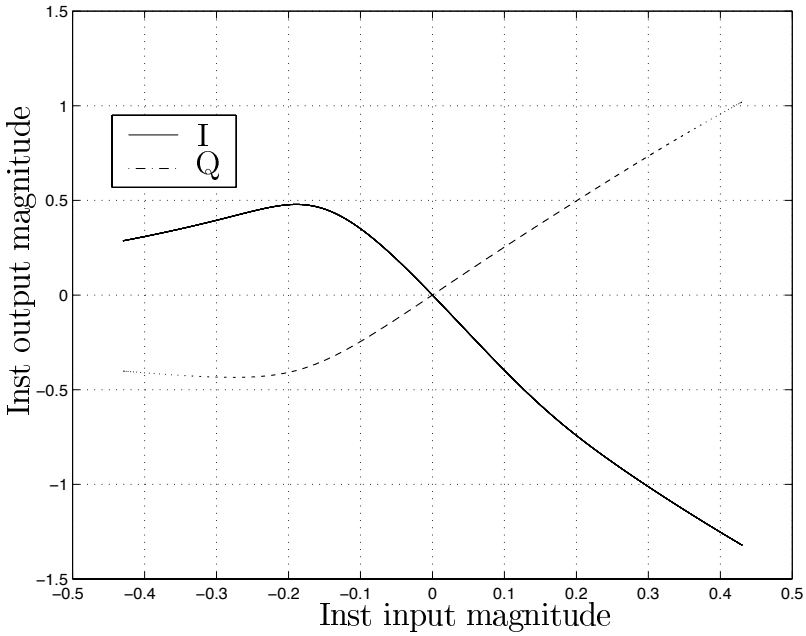


Figure 3.12: MIQM: k_I and k_Q nonlinear transfer functions.

the IQM and MIQM analysis models. There is inherent memory in a circuit any time nonlinear elements are coupled with energy storage elements such as inductors and capacitors. Recall that in quasi-memoryless systems the operating point of the underlying nonlinear device is assumed to be instantaneously related to the voltage across or current into the device. When there exist elements in the circuit, such as resonators, with large time constants the quasi-memoryless assumption no longer holds and the accuracy of the QM, IQM, and MIQM methods is reduced. Care must be taken in the circuit design to remove the presence of long time constant paths [46]. One such path in a BJT based LNA path is the feedback path. The adjustment of the components to mitigate the long time constant in this path mainly effects the S22 parameter, and care must be taken so that the design goals are still satisfied when making these adjustments.

In [46] a method is given for observing the amount of memory present in a circuit.

A two tone frequency sweep under heavy drive conditions is used to analyze the effect of a modulated signal on the input/output gain of the system. The test fixes the input frequency of one tone while it sweeps the frequency of the second tone past that of the first tone. This test simulates the varying envelope of a modulated signal. If the gain of the fixed tone varies then there exist paths in the circuit that are inhibiting the instantaneous change of the bias point of the nonlinear element which causes memory effects. A memoryless system will exhibit perfectly flat response, the gain will be the same across all frequencies of interest. A system with a large amount of memory will exhibit a swift change in gain during part of the sweep, and a system with a small amount of memory will exhibit a gradual change in gain across the frequency band of interest. Examples of the two tone gain test for the circuit with memory shown in Figure 3.1 and a quasi-memoryless circuit shown in Figure 3.15 may be seen in Figures 3.13 and 3.17. The corresponding small signal S-parameters are shown in Figures 3.3 and 3.16, where it can be seen that the small signal match for S22 flattens as the memory of the system is reduced by adjusting the feedback path components. Note the gain of the two tone test is magnitude gain of an input ideal source in relationship to the output across a 50Ω load. The S21 parameter is the ratio of the power waves into and out of the ports. Under perfect matching the ideal source will sink half its power into the input port of the amplifier, thus the actual port to port transducer gain is a factor of two larger than the magnitude gain plotted. Some gain loss is also due to the gain compression of the amplifier under hard driving conditions. The important characteristic to note is not the gain mismatch between the two plots but the severity of the uncovered memory response.

The effect of mitigating the frequency dependence of the circuit can be observed by observing the sum of the odd order $2f_1 - f_2$ intermodulation products' magnitude

as two input tones are swept in magnitude. Illustrated in Figures 3.14 and 3.18, the match between the harmonic balance method of the CAD program HARMONICA and the predicted $2f_1 - f_2$ magnitude from the MIQM improves markedly as the frequency dependence (and hence the memory) of the circuit is reduced. As a side note, Figures 3.14 and 3.18 were generated with a hybrid MIQM/discrete technique (time domain signals and FFT analysis) and exhibit inaccuracy in the lower input magnitudes due to numeric noise and the Hilbert transform of a finite sequence of points.

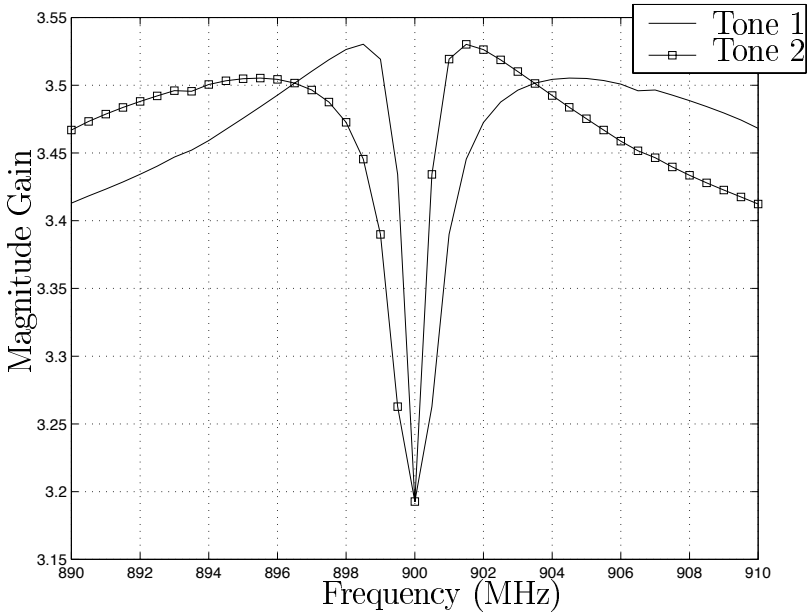


Figure 3.13: Two tone frequency sweep, resulting tone gains, circuit with memory.

Since the narrow-band S22 rejection is generally a function of a long time constant circuit path, the effect on the circuit match by modification of this path may give a circuit designer pause. However, one must remember that the S-parameters are small signal measurements and are not strictly applicable in a large signal system. In a system where the input signal is large enough to shift the bias point of the nonlinear

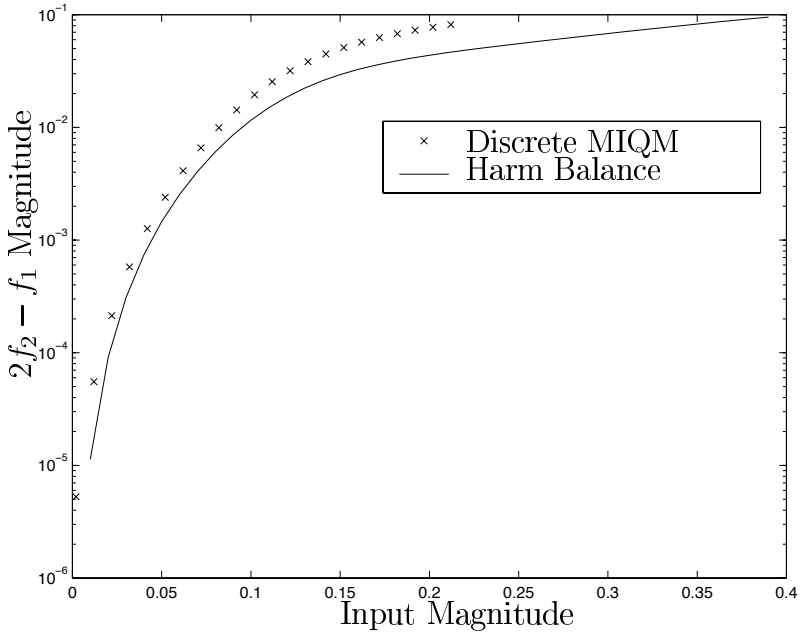


Figure 3.14: Two tone magnitude sweep, resulting $2f_2 - f_1$ Magnitude, circuit with memory.

device the small circuit S-parameters can be used as guidelines, not gospel.

In [46] suggestions are made as to where long time constant paths may be found. For BJT amplifiers, they are generally located in the feedback path and the biasing networks; if the long time constant paths cannot be effectively removed from the circuit design, the "Quadrature" based models (QM, IQM, MIQM) will break down. Additionally, no effort is made to quantify to what degree the system may exhibit memory. The MIQM shares this weakness with the QM and the IQM, so if the degree of memory is quantified for one it is quantified for all.

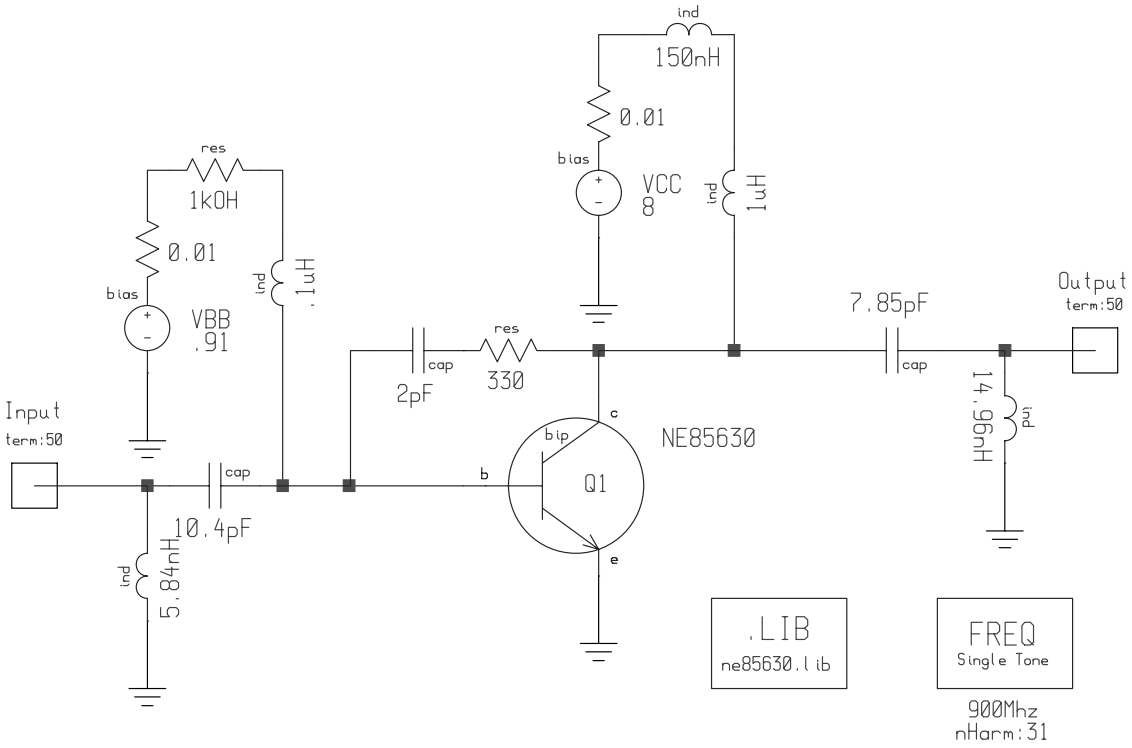


Figure 3.15: 900MHz LNA, .91V bias, quasi-memoryless.

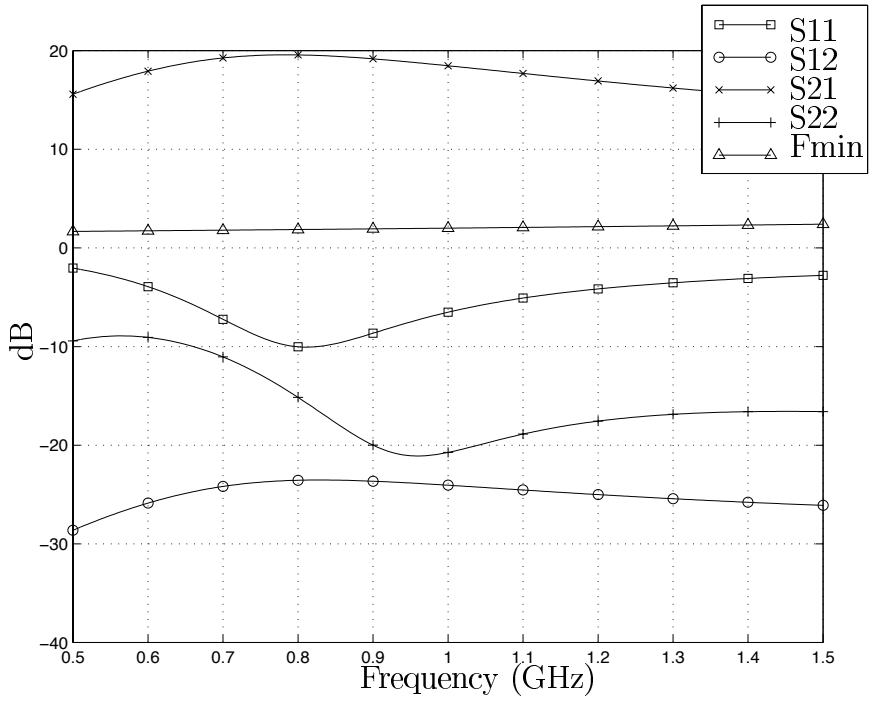


Figure 3.16: Matching network response: .91V bias LNA, quasi-memoryless circuit.

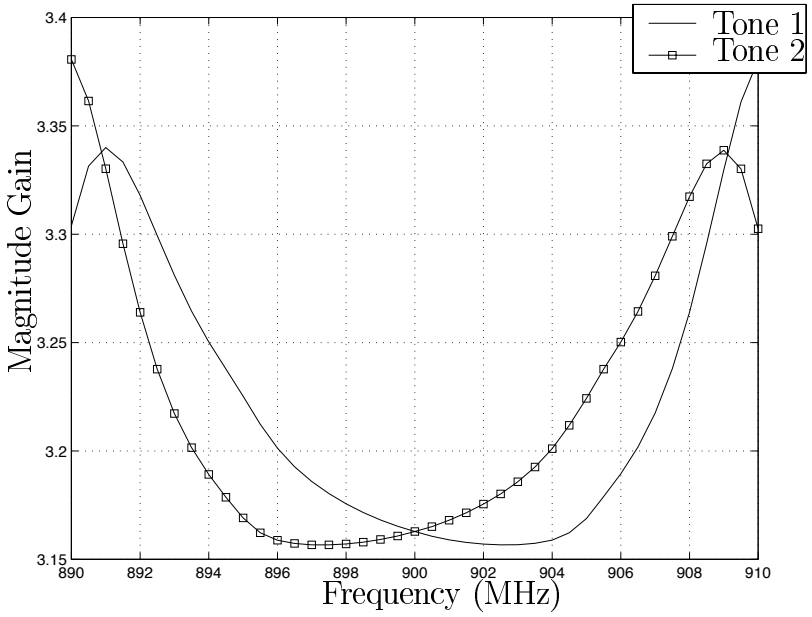


Figure 3.17: Two tone frequency sweep, resulting tone gains, quasi-memoryless circuit.

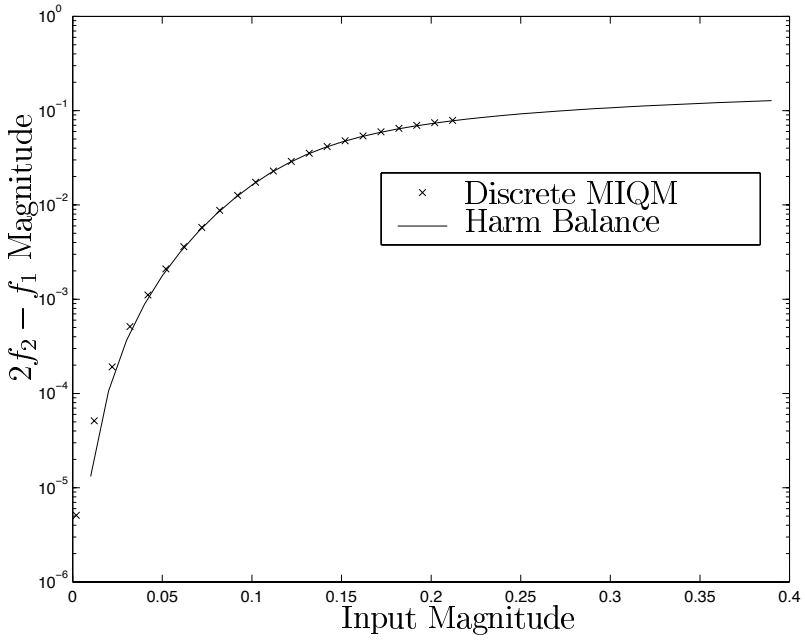


Figure 3.18: Two tone magnitude sweep, resulting $2f_2 - f_1$ Magnitude, quasi-memoryless circuit.

3.7 Power Series Analysis Coupled with MIQM

With the nonlinear functional representation of the MIQM comes the ability to model the system nonlinearities with a power series fit. The power series fit gives the ability to directly determine the magnitude and phases of intermodulation and harmonic tones without the drawbacks of the FFT based discrete methods. As discussed in Section 3.4.3, a Chebychev polynomial is fit to the I and Q channel MIQM nonlinear transfer functions using the least mean square criteria found in [39].

In [19], Wass developed a table method to determine the coefficients, phases, and frequencies of N'th order M tone intermodulation products from a polynomial nonlinearity. The polynomial fit to the quasi-memoryless system of Figure 3.15 is shown in Figure 3.19, the error of the fit is shown in Figure 3.20, and the resulting $2f_2 - f_1$ magnitude from an amplitude swept two tone test is shown in Figure 3.21. This test is comparable to the two tone tests generated via the discrete method in Figures 3.14 and 3.18, with the benefit of the knowledge of the magnitude, phase, and frequency of each IM tone and without the drawbacks of the discrete methods.

3.8 Conclusions

The MIQM offers unique advantages to the analysis of real nonlinear circuits. Under the constraint that the circuit is at least quasi-memoryless, we are able to determine the in-phase and quadrature-phase nonlinear transfer functions which accurately model the intermodulation performance of the circuit. In fact, the circuit output is modeled correctly across the whole band where the memory test response is flat. These functions, coupled with the ability to determine the magnitude, phase, and

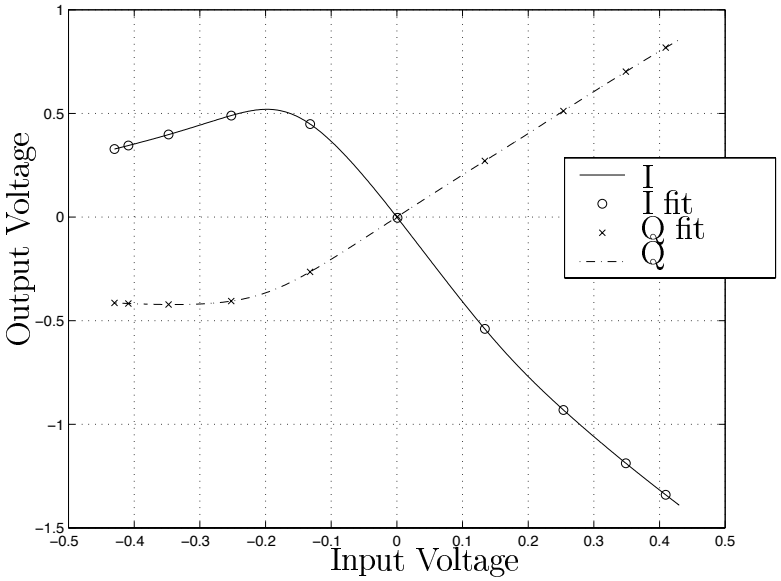


Figure 3.19: 24'th order Chebychev polynomial fit to I and Q nonlinearities.

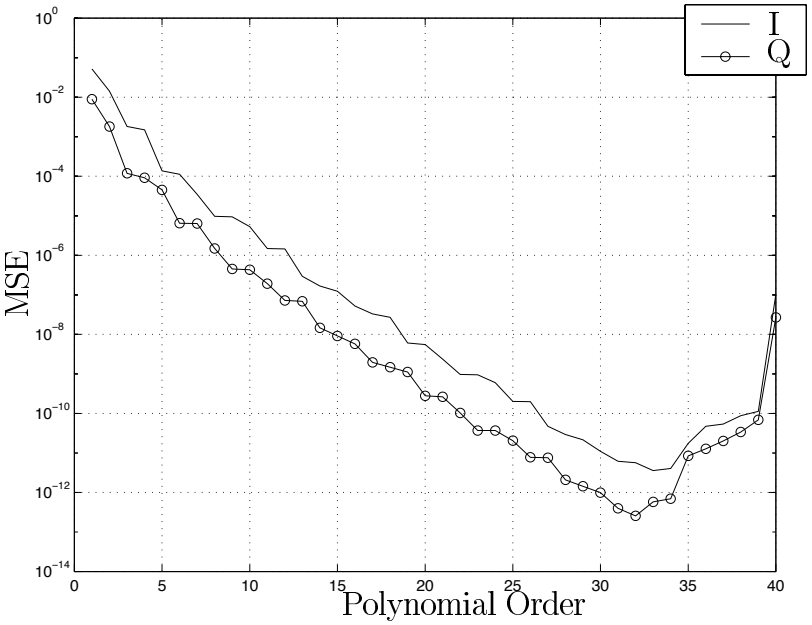


Figure 3.20: MSE of Chebychev polynomial fit to I and Q nonlinearities.

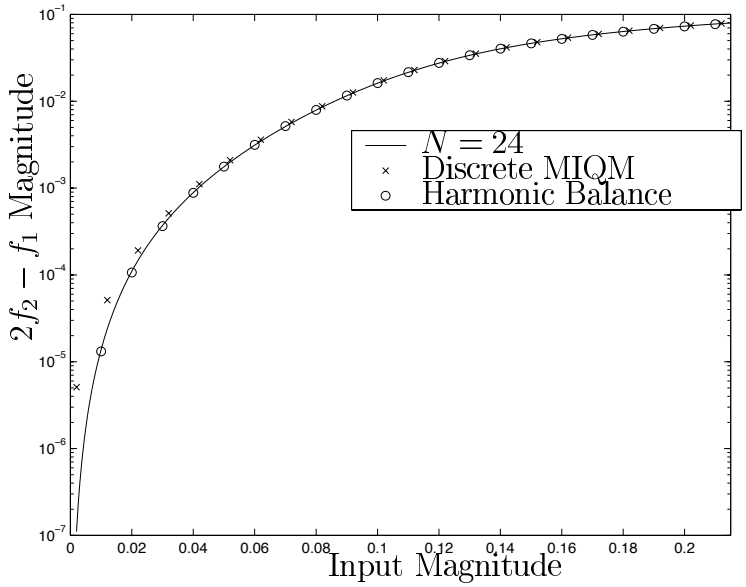


Figure 3.21: $2f_2 - f_1$ performance of 24'th order polynomial fit to the MIQM.

frequency of all intermodulation products form a new nonlinear analysis framework.

The primary advantages of this framework are the ability to model both large and small input signal levels accurately, the lack of aliasing problems and numerical noise introduced by the FFT, and the speed at which results can be achieved for large numbers of input tones spaced near to each other in frequency. One further advantage that is not as obvious is that the methods used to generate the IM products give a phase relationship between the input frequencies and the output products. This relationship can be used to determine variations in results due to random phase inputs without exhaustive simulation. Meeting these goals simultaneously is a strong advantage that MIQM has over other behavioral models, HB, time domain analysis, and analytical models.

In all fairness, there are weaknesses in this framework. Since the quadrature methods do not accurately model systems with memory, care must be taken in the design of narrow-band amplifiers so the frequency band of interest exhibits a flat two

tone response. The second weakness is the generation and storage of the IM products of interest. Since the number of intermodulation products grows exponentially with the order of the polynomial fit, the MIQM can exhibit similar complexity problems as HB simulation. The difference, however, is the ability to calculate the magnitude, phase, and frequency of these tones in advance. Hence, the complexity can be shifted from run-time complexity as in HB to storage complexity, which is largely cheaper.

This model will be used in Chapter 4 to illustrate the effects of very high Q filters on LNA performance and power efficiency. Implementation specifics not covered in this chapter will be introduced in Chapter 4 along with the results of the calculations.

CHAPTER 4

Analysis of MEMS Filters in Communications Systems

The micro electro–mechanical (MEMS) filter is very small (on the order of 420 μm^2), it can be constructed to be very narrow at passband. The ration of center frequency to filter bandwidth is called Q; MEMS filters can achieve Q's on the order of 80,000 [47]. A scanning electron microscope image from [47] is illustrated in Figure 4.1. This chapter explores the use of MEMS filters in communications systems.

We start this chapter by analyzing the addition of a narrow filter into a sub-sampling architecture, and we show the reduction of system signal to interference ratio (SIR). The second issue we explore is the filter bandwidth requirements in a system that receives a weak signal in the presence of two strong interfering signals. We quantify the increase in interference rejection as a result of decreasing front end filter bandwidth. The third issue we explore is the increase in amplifier power added efficiency (PAE) that is achievable by using narrow front end filters.

This chapter is structured in a manner where each of the three topics is separately motivated and analyzed with a common conclusion following all sections. The common thread is the application of narrow MEMS filters to communications systems.

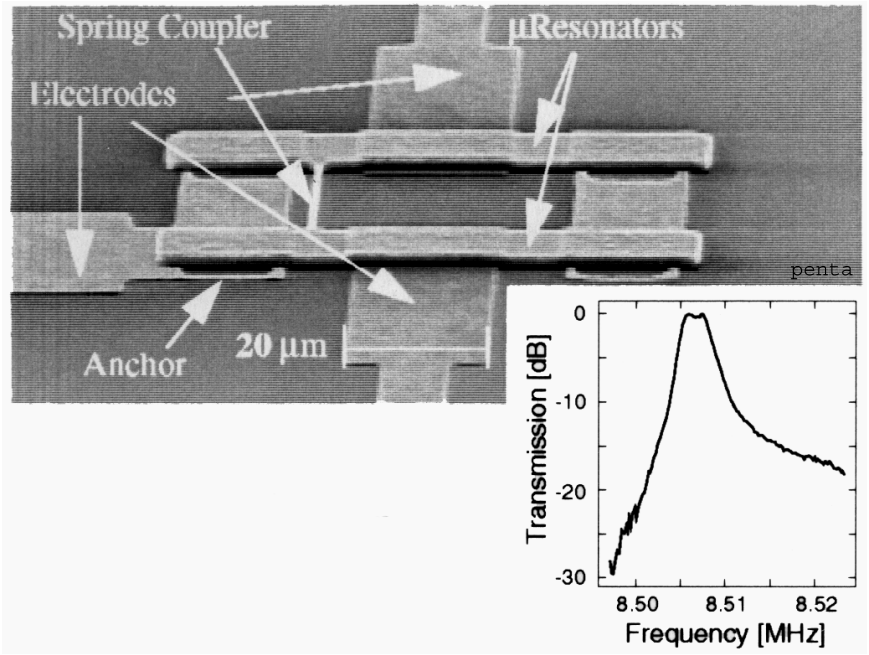


Figure 4.1: SEM of a spring-coupled bandpass MEMS filter and its measured frequency response.

4.1 MEMS filter use in sub-sampling systems

4.1.1 Sub-sampling motivation

In a traditional super heterodyne receiver the received passband signal is filtered by a wide filter, mixed down to an intermediate frequency (IF), filtered by a narrow IF filter, and then mixed down to baseband where processing occurs. As receivers begin to shrink in size and power requirements become more stringent this heterodyning architecture begins to exhibit three main problems: the physical filtering components are too large to integrate on a chip, the ability to manufacture narrow filters at passband becomes more difficult, and the power requirements for two high frequency oscillators can be large. A micro electro-mechanical (MEMS) filter potentially offers a very attractive solution to these problems.

When a passband signal is sampled at a rate less than the Nyquist rate, or sub-sampled, spectral replicates of this signal appear at all integer multiples of the sampling frequency. These replicates generate interference in the desired signal image. In [48] conditions are offered for self-interference free sub-sampled signal reception, but it can be shown that these conditions do not prevent spectral overlap. In this section we address the system restrictions necessary for self-interference free communication in a sub-sampled system which utilizes MEMS filtering. One common drawback to sub-sampling is the spectral replication of wide-band sampler thermal noise. We show that under a further restriction the spectrally replicated wide-band noise is bounded. Additionally, we give results quantifying the SNR loss in the system due to this replicated noise.

In Section 4.1.2 we model the MEMS filter and we propose a simple architecture for a digital radio that uses sub-sampling and MEMS filters. In Section 4.1.3 we develop conditions for which sub-sampling may be employed to achieve an interference free direct down conversion. We prove, in Section 4.1.4, that the additive effect of the wide-band thermal noise and sampler resistor noise replication as a result of the sub-sampling is bounded. Finally, in Sections 4.1.5 and 4.4 we show the calculated SNR loss and present conclusions.

4.1.2 Proposed system architecture

Our proposed architecture, shown in Figure 4.2, consists of using a very narrow bandwidth MEMS filter coupled with a front-end low noise amplifier (LNA), with gain α, and a sub-sampling mechanism to down-convert the desired passband information signal to a carrier frequency that is small enough such that relatively inexpensive real-time digital signal processors can process it. In general the sampling low pass

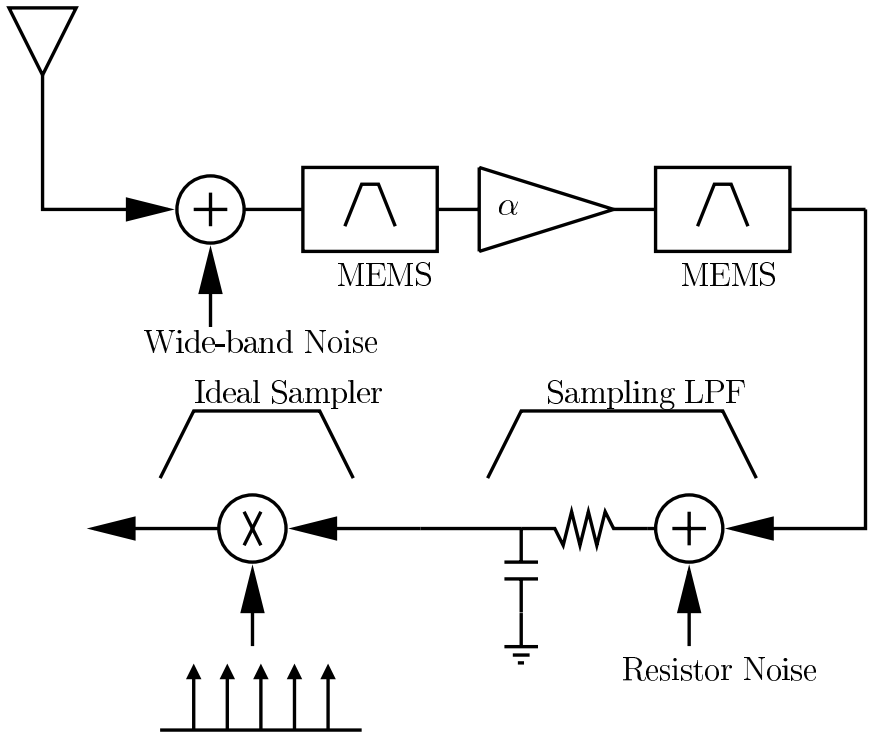


Figure 4.2: Block diagram of MEMS/sub-sampling receiver.

filter (LPF) introduces noise after the front end MEMS filtering. In reality, the noise can be modeled as a noisy resistor in the sample and hold circuit of the analog to digital converter that is performing the sampling operation. We use a colored noise source followed by an ideal sampler to model a noisy sample and hold device. Since the MEMS filters are small in size and sub-sampling is an operation that can be accomplished in silicon, our architecture can be completely integrated.

To simplify the analysis, we model the MEMS filter shape as a mathematical function with the following assumptions: the filter response is piecewise linear on the dB scale and the passband is flat with no attenuation, as shown in Figure 4.3, where W is the bandwidth of the filter and f_c is the center frequency. While this is a rudimentary model, it is sufficient to promote understanding of the fundamental problems associated with direct down conversion by sub-sampling. We write this

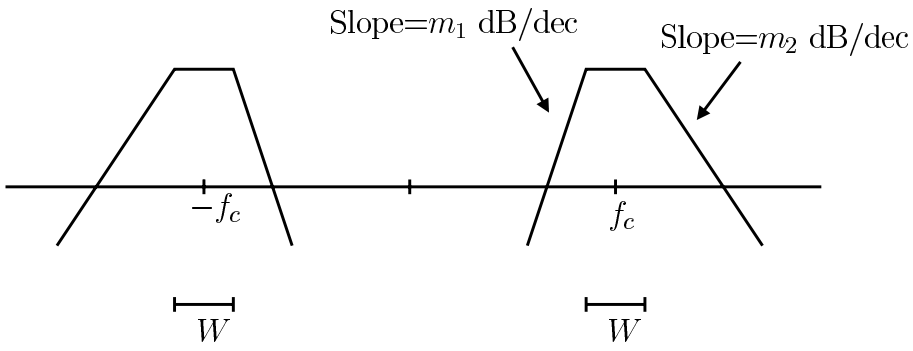


Figure 4.3: Piecewise linear MEMS filter model.

model on a linear scale as:

$$|H_m(f)|^2 = \begin{cases} -f^{\frac{m_2}{10}}(f_c + \frac{W}{2})^{\frac{-m_2}{10}} & f < -f_c - \frac{W}{2} \\ 1 & -f_c - \frac{W}{2} < f < -f_c + \frac{W}{2} \\ |f|^{\frac{m_1}{10}}(f_c - \frac{W}{2})^{\frac{-m_1}{10}} & -f_c + \frac{W}{2} < f < f_c - \frac{W}{2} \\ 1 & f_c - \frac{W}{2} < f < f_c + \frac{W}{2} \\ f^{\frac{m_2}{10}}(f_c + \frac{W}{2})^{\frac{-m_2}{10}} & f > f_c + \frac{W}{2}. \end{cases} \tag{4.1}$$

The received signal is passed through a narrow MEMS/amplifier/MEMS chain. We assume the combination of the 2 MEMS filters provide the response given by $|H_m(f)|^2$.

While it would be simplest to assume the amplifier provides a gain α for the whole spectrum, this is not a very practical assumption. Additionally, in practice, the amplifier may have a non-linear response, but we assume the surrounding MEMS filters remove all generated harmonics. At this point, we can assume the amplifier to be linear with a single pole low-pass response with breakpoint, f_{ha}. We model the

amplifier's response with the following transfer function:

$$|H_a(f)|^2 = \begin{cases} \alpha(-f^{\frac{-20}{10}}(f_{ha})^{\frac{20}{10}}) & f < -f_{ha} \\ \alpha & -f_{ha} < f < f_{ha} \\ \alpha(f^{\frac{-20}{10}}(f_{ha})^{\frac{20}{10}}) & f > f_{ha}, \end{cases} \tag{4.2}$$

where f_{ha} is significantly beyond the passband we are considering.

One further quantity in need of definition is the response of the sampler. The sampling LPF is a consequence of the charge gathering capacitor and the small resistance in the sample capture switch. The bandwidth of this sampling action is typically designed to be larger than the frequency of the signal that is being sampled. In this case we assume a piecewise linear model with a single pole breakpoint, f_{hm}, significantly beyond the passband we are considering. We model the sampler's response with the following transfer function:

$$|H_r(f)|^2 = \begin{cases} -f^{\frac{-20}{10}}(f_{hm})^{\frac{20}{10}} & f < -f_{hm} \\ 1 & -f_{hm} < f < f_{hm} \\ f^{\frac{-20}{10}}(f_{hm})^{\frac{20}{10}} & f > f_{hm}. \end{cases} \tag{4.3}$$

We note that (4.3) is very similar to (4.2). For the simple case of the amplifier having a LPF response we could have merely combined the poles, but to provide flexibility in amplifier use we choose not to combine the poles at this point.

4.1.3 Interference Free Conditions

To avoid self-interference, the standard sampling procedure is to sample at the Nyquist rate, or twice the largest frequency in the source. For low-power bandpass systems the Nyquist rate would be prohibitively large, so a smaller, sub-sampling

rate is employed. The standard sub-sampling procedure is to sample at a rate that is twice B, the *bandwidth* of the passband signal. However, we will show that this is not the only concern.

Ignoring the effects of the LPF, it is well known that spectral replicates are generated whenever any analog system is sampled [49] (Figure 4.4). These spectral

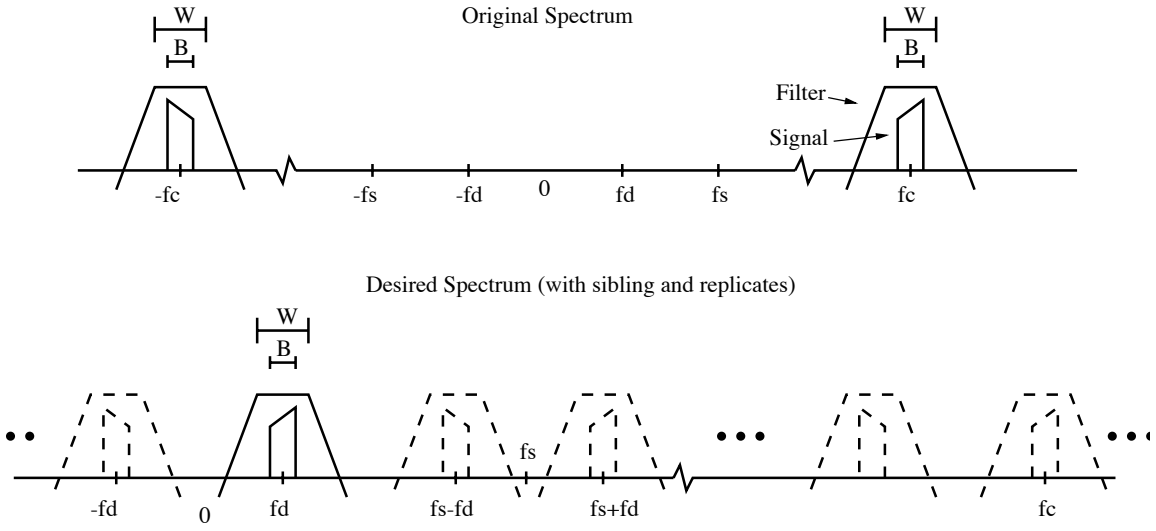


Figure 4.4: Spectral replication.

replicates occur at every multiple of the sampling frequency, and it is the replicated centered at f_d that we will use as our down-converted passband signal. However, these replicates may introduce additional interference, so we impose additional restrictions on the parameters of our system.

Restriction I: Main Shift

The first restriction specifies the relationship between the sampling rate (f_s), the desired image frequency (f_d), and the carrier frequency (f_c). If we wish to mix the signal present at f_c down to f_d we must choose f_s such that

$$f_c = kf_s + f_d, \tag{4.4}$$

where k is an integer and $0 < f_d < \frac{f_s}{2}$. This restriction assures placement of a spectral image centered at f_d.

Restriction II: Nearest Neighbor

The next restriction specifies the minimum separation between the desired image and the nearest spectral image to its right, or the nearest spectral "neighbor". We require all energy from non-desired spectral images to be attenuated, i.e. the flat region of the nearest neighbor's filter response must not overlap the the desired image's signal bandwidth. This leads to the following restriction for the sampling rate and the desired frequency:

$$\begin{aligned} f_s - f_d - \frac{W}{2} &> f_d + \frac{B}{2}, \\ f_s &> 2f_d + \frac{1}{2}(B+W). \end{aligned} \tag{4.5}$$

Restriction III: Sibling

The "sibling" image is the image that is centered at $-f_d$ which bears special attention because it may be closer in frequency than the nearest neighbor. The final restriction specifies the minimum distance between the desired image and its sibling. We require the non-attenuated region of the sibling image's filter response to not overlap the desired signal image, i.e.

$$\begin{aligned} f_d - \frac{B}{2} &> -f_d + \frac{W}{2}, \\ f_d &> \frac{1}{4}(B+W). \end{aligned} \tag{4.6}$$

Together, (4.4) (4.5) and (4.6) will prevent spectral overlap and the corresponding corruption of the demodulated signal due to sub-sampling.

4.1.4 Noise Replication

The previously mentioned restrictions prevent self interference, but they do not account for wide-band thermal noise replication. With sub-sampling, we introduce spectral images that are fairly close together but we also replicate the thermal noise in each image as well. Since real filters never completely block energy, some finite amount of energy from *every* spectral image will be present in the bandwidth of the desired image. This noise will be additive and must be examined to assure the sum of the noise from the infinite number of spectral images is bounded. In addition, we have thermal noise introduced from the noisy resistor in the sampling mechanism. This "resistor noise" is added after the MEMS filter and therefore it is only filtered by the low-pass structure of the sampler.

Noise Multiplier

We assume the noise density of the sampling resistor noise N_r is proportional to the wide-band thermal noise N_0 in the following manner: $\frac{N_r}{2} = \phi\frac{N_0}{2}$. Using (4.1), (4.2), and (4.3) we write the noise contribution of the k'th spectral replicate as:

$$\int_{f_d-\frac{W}{2}}^{f_d+\frac{W}{2}} \frac{N_0}{2}|H_m(f-kf_s)|^2|H_a(f-kf_s)|^2|H_r(f-kf_s)|^2 + \frac{N_r}{2}|H_r(f-kf_s)|^2 df. \tag{4.7}$$

To calculate the noise contribution of all replicates we write:

$$N_T = \frac{1}{W}\sum_{k=-\infty}^{\infty}\int_{f_d-\frac{W}{2}}^{f_d+\frac{W}{2}} \frac{N_0}{2}|H_m(f-kf_s)|^2|H_a(f-kf_s)|^2|H_r(f-kf_s)|^2 + \frac{N_r}{2}|H_r(f-kf_s)|^2 df,$$

$$= \frac{N_0}{2}\times\frac{1}{W}\sum_{k=-\infty}^{\infty}\int_{f_d-\frac{W}{2}}^{f_d+\frac{W}{2}} |H_m(f-kf_s)|^2|H_a(f-kf_s)|^2|H_r(f-kf_s)|^2 + \phi|H_r(f-kf_s)|^2 df. \tag{4.8}$$

The total noise expression, N_T can be split into the thermal noise and a noise multiplier, N_m. We assume the thermal noise power density is flat across all frequencies so the convergence of N_m is necessary for the additive noise to be bounded. To prove convergence we upper bound the integral as follows:

$$\begin{aligned} N_m = \frac{1}{W}\sum_{k=-\infty}^{\infty}\int_{f_d-\frac{W}{2}}^{f_d+\frac{W}{2}} & |H_m(f-kf_s)|^2\times \\ & |H_a(f-kf_s)|^2\times \\ & |H_r(f-kf_s)|^2+ \\ & \phi|H_r(f-kf_s)|^2 df, \\ \leq A + B & \end{aligned} \tag{4.9}$$

where

$$A = \sum_{k=-\infty}^{\infty}\max(|H_m(f_d+\frac{W}{2}-kf_s)|^2\times|H_a(f_d+\frac{W}{2}-kf_s)|^2\times|H_r(f_d+\frac{W}{2}-kf_s)|^2, |H_m(f_d-\frac{W}{2}-kf_s)|^2\times|H_a(f_d-\frac{W}{2}-kf_s)|^2\times|H_r(f_d-\frac{W}{2}-kf_s)|^2), \tag{4.10}$$

and

$$B = \sum_{k=-\infty}^{\infty}\max(\phi|H_r(f_d+\frac{W}{2}-kf_s)|^2, \phi|H_r(f_d-\frac{W}{2}-kf_s)|^2). \tag{4.11}$$

Examining the expression for A, we note $H_a(f) \leq \alpha$ and $H_r(f) \leq 1$, therefore $\sum_{k=-\infty}^{\infty} |H_m(f-kf_s)|^2 |H_a(f-kf_s)|^2 |H_r(f-kf_s)|^2 \leq \sum_{k=-\infty}^{\infty} \alpha |H_m(f_d + \frac{W}{2} - kf_s)|^2$. The range of the central regions of $|H_m(f)|^2$, where $f \in (-f_c - \frac{W}{2}, f_c + \frac{W}{2})$, is bounded by 0 and 1 and only intersects with $(f_d - \frac{W}{2} - kf_s, f_d + \frac{W}{2} - kf_s)$ in a finite range of k, therefore the sum in this region must converge. Thus, we are concerned only with the convergence of the sum of the tail regions. We concentrate on the left shifts of the right tail, with slope m_2 (where m_2 is a negative number). Let a be the point where $|H_m(f)|^2$ begins its downward slope, so $f_c + \frac{W}{2} + kf_s < a$ implies $k < \frac{a - \frac{W}{2} - f_c}{f_s}$. Now we show $E = \sum_{k=-\infty}^{\lfloor \frac{a-\frac{W}{2}-f_c}{f_s} \rfloor} |H_m(f_d - \frac{W}{2} - kf_s)|^2$ converges:

$$E \leq A \sum_{k=-\lfloor \frac{a-\frac{W}{2}-f_c}{f_s} \rfloor}^{\infty} \frac{1}{(f_d - \frac{W}{2} + kf_s)^{\frac{-m_2}{10}}} \tag{4.12}$$

$$\leq A \sum_{k=-\lfloor \frac{a-\frac{W}{2}-f_c}{f_s} \rfloor}^{\infty} \frac{1}{(-\frac{W}{2} + kf_s)^{\frac{-m_2}{10}}}, \tag{4.13}$$

$$\leq A \sum_{k=-\lfloor \frac{a-\frac{W}{2}-f_c}{f_s} \rfloor}^{\infty} \frac{1}{(-\frac{kW}{2} + kf_s)^{\frac{-m_2}{10}}}, \tag{4.14}$$

$$= A \sum_{k=-\lfloor \frac{a-\frac{W}{2}-f_c}{f_s} \rfloor}^{\infty} \frac{1}{(k(-\frac{W}{2} + f_s))^{\frac{-m_2}{10}}}, \tag{4.15}$$

$$\leq \frac{A}{(-\frac{W}{2} + f_s)^{\frac{-m_2}{10}}} \sum_{k=-\lfloor \frac{a-\frac{W}{2}-f_c}{f_s} \rfloor}^{\infty} \frac{1}{k^{\frac{-m_2}{10}}}, \tag{4.16}$$

$$\tag{4.17}$$

where $A = (f_c + \frac{W}{2})^{\frac{-m_2}{10}}$. The final expression converges via the p-series test [50] when $m_2 < -10$. Thus, the sum of the right tail images converge and the same holds true for the left tail images which implies the convergence of A. To show convergence of B we note the center region of $|H_r(f)|^2$ is bounded and intersects the desired range for finite k. A convergence argument similar to (4.17) holds for the tails of $|H_r(f)|^2$,

and hence the sums in (4.8) and (4.11) converge, and the noise multiplier converges under the condition that the filter slope is sufficiently steep. Next we examine the relative contribution of the resistor noise and the replicated wide-band noise.

A major problem with resistor noise in a sub-sampling architecture is that we are unable to use a MEMS (narrow) filter to mitigate the effects of neighboring resistor noise images in the desired image because the noise is added after the filtering has taken place. As a result, a noise density of approximately $m = \frac{f_c}{f_s}$ [22] times $\frac{N_r}{2}$ added to the desired image after the MEMS filters. The solution to this problem is to dominate the resistor noise by amplifying the front end signal and wide-band noise by a factor α, where α can be potentially very large. This will force the SNR of the desired image to have a small dependence on the resistor noise. To this end, we define the resistor to wide-band noise ratio (RWNR) as the ratio between the resistor noise contribution and the contribution of the amplified wide-band noise. In general this quantity can be expressed as follows:

$$RWNR = \frac{\sum_{k=-\infty}^{\infty} \int_{f_d-\frac{W}{2}}^{f_d+\frac{W}{2}} \phi |H_r(f-kf_s)|^2 df}{\sum_{k=-\infty}^{\infty} \int_{f_d-\frac{W}{2}}^{f_d+\frac{W}{2}} |H_m(f-kf_s)|^2 |\alpha Z_a(f-kf_s)|^2 |H_r(f-kf_s)|^2 df}, \quad (4.18)$$

where $\alpha Z_a(f-kf_s) = H_a(f-kf_s)$. We note that the denominator of (4.18) is a function of α and the numerator converges, therefore

$$\lim_{\alpha \to \infty} RWNR = 0. \quad (4.19)$$

This tells us that for sufficiently high gain at the front end, the impact of the resistor noise is negligible and the SNR loss of the system can be accounted for solely by the replicated wide-band noise.

4.1.5 Numerical Results

In Section 4.1.4 we established that the sum of the noise contribution of the infinite spectral replicates with large front end LNA gain converges under the condition that the filter slope is steeper than 10 dB/decade. We now present an example system where one user transmits data in a W=B=25 kHz bandwidth centered at f_c=50 MHz and the desired sub-sampled mix down point is f_d=50 kHz. We compare the loss due to wide-band noise replication for sub-sampling rate of about 100 kHz, 200 kHz, 400 kHz, 1 MHz, and 4 MHz. To do this we must select values f_s that satisfy the the restrictions in Section 4.1.3. Figure 4.5 illustrates the SNR loss due to the noise replication as the filter slope becomes steeper and it quantifies this result.

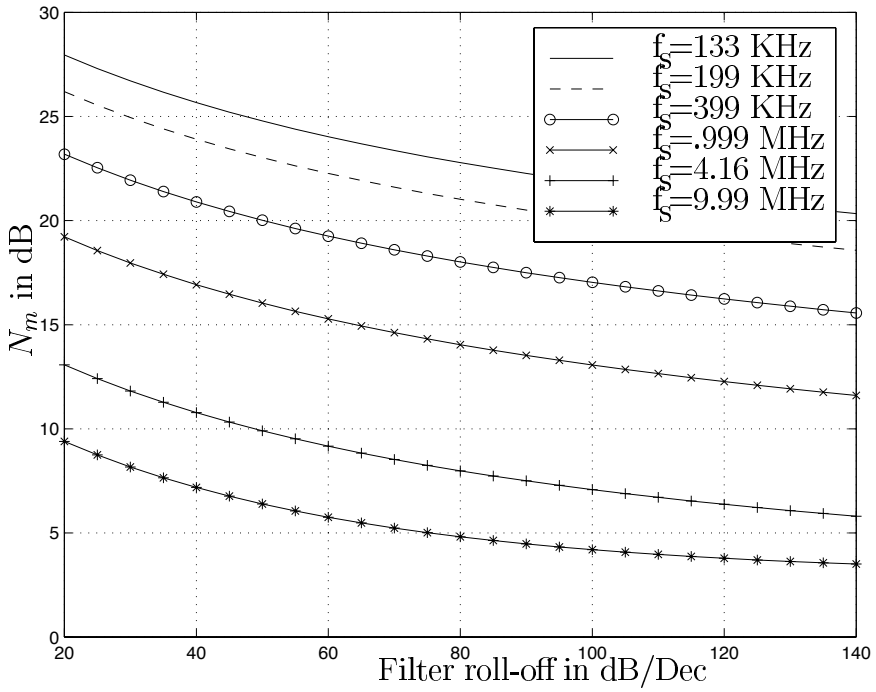


Figure 4.5: SNR loss for a system with dominated resistor noise.

In Figure 4.5 we show the noise multiplier when the resistor noise is completely dominated, essentially for an amplifier with infinite gain. However, for an amplifier

with finite gain, it is of interest to determine the effect of resistor noise on the total noise. In Figure 4.6 the filter slope is fixed at 80dB/dec, $\phi = 1$, and the LNA gain is varied. We can see that the effect of resistor noise is negligible as the amplifier gain exceeds 30dB. Figure 4.7 shows the RWNR versus the gain for a system with fixed

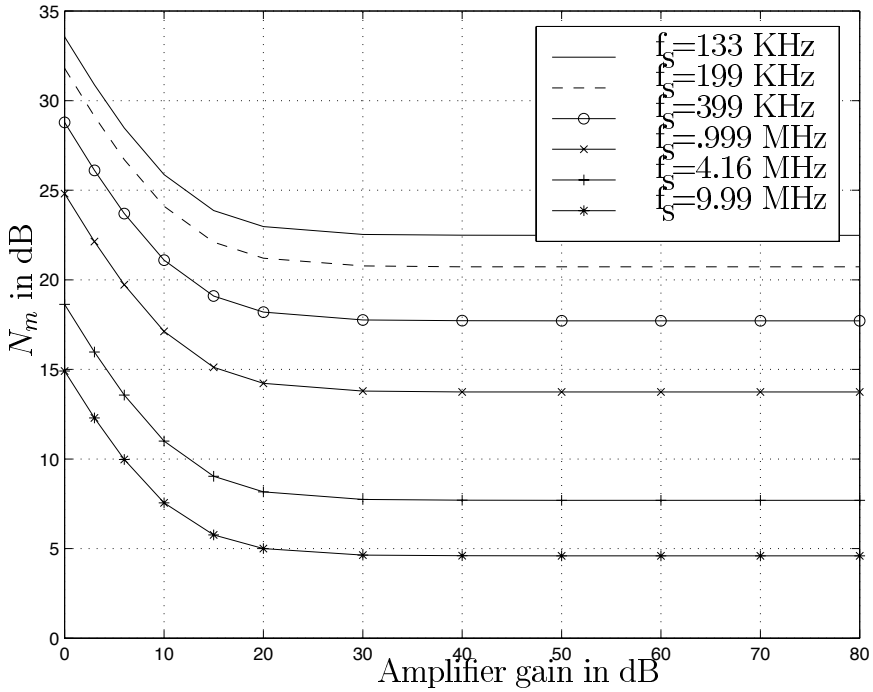


Figure 4.6: Loss vs. amplifier gain for fixed filter slope.

sampling rate of about 1MHz. It is interesting to note that the gain required for a fixed RWNR can vary by about 5dB depending on the filter slope, with the shallower filter slope requiring less amplifier gain even though the filter slope limits the overall loss of the system. For a given sampling rate a poor choice of filter slope causes a maximum of about 5-7dB loss, while a poor choice of amplifier gain can cause an additional 10dB loss. Stated another way, given a reasonable overall loss requirement for a fixed sampling rate, there are different slope/gain choices to achieve the design goal.

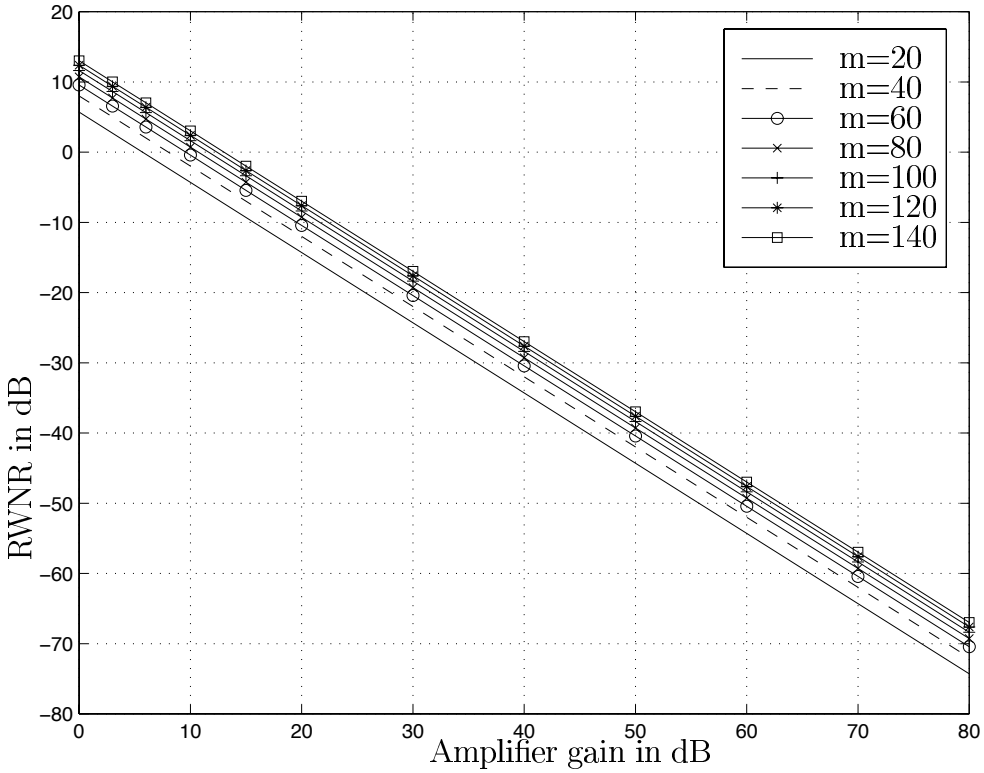


Figure 4.7: RWNR for fixed sampling rate.

4.2 MEMS Use in Cosite Problems

4.2.1 Cosite problem motivation

The main topic in this section is the problem of cosite interference in communications systems. In various commercial and military communication systems the problem of detecting and effectively demodulating a weak signal in the presence of strong interferers arises. Whether it be an IS-95 digital device or a command radio co-located with other transmitting antennae, the design documentation for these systems usually specify a minimum performance metric for the reception of the weak signal. For example, in an IS95 system the receiver must be capable of receiving a -101dBm signal in the presence of two -30dBm interferers. In [51, 52], others have considered different techniques to combat the corruption of the weak signal from spillover power from the large signals and corruption due to intermodulation interference. These works have generally simulated the appropriate system and provided interference reduction by various filters and frequency hop patterns. In this section, we apply the MIQM technique to a wireless receiver that attempts to receive a small signal in the presence of two strong interferers. The MIQM analysis framework developed in Chapter 3 accounts for the introduction of a nonlinear amplifier into the receiver front end. In addition, we incorporate a variable filter model as to quantify the tradeoff between filter quality and interference in the demodulated small signal. Where this work differs from others, is in its use of MIQM analysis rather than simulation to illustrate nonlinear effects in co-located receivers.

In a typical RF receive chain, an LNA is located as one of the first elements. If this amplifier is nonlinear, amplifying the desired and nearby interferers can cause interference at the frequency of the desired signal and at nearby surrounding frequen-

cies. This interfering power effects the overall quality of the received signal as early as the IF stage. In the following analysis, we determine the SIR at the IF point, and give requirements for the front-end filtering to achieve the desired SIR.

With the advances in MEMS filtering technologies, very high Q filters at passband can be fabricated. These filters, when compared to their predecessors, can provide the advantage of attenuating a large portion of the strong interfering signal before the nonlinear effect "mixes" the interferers into the desired band. Additionally, MEMS filters can be employed to further attenuate the interfering power after the LNA to reduce the distortion into a potentially non-ideal mixer element.

In Section 4.2.2 we give our proposed system model, and in Section 4.2.3 we develop the analysis framework to determine the effect of IM product interference in the nonlinear front end. In Sections 4.2.3 and 4.2.4 we use the analysis to determine the system signal-to-interference ratio and we determine the necessary filter requirements to guarantee a given SIR design goal. In Section 4.2.5 we discuss the approximation of our system with a third order intermodulation assumption. Finally, we draw conclusions in Section 4.4.

4.2.2 System Model

Consider a system in which the desired signal is $s_d(t) = A_d \cos(2\pi f_d t)$ and there are two nearby continuous wave (CW) tone interferers at distinct frequencies that are ΔP dB stronger in power than the desired signal.

The system we will analyze has the form of Figure 4.8. In a standard radio receiver, the received spectrum may be filtered, amplified, and filtered again. The conventional filters are low Q (wide) filters which pass the desired signal and any neighboring signals to the LNA, while in the system we consider the Q of the filters

grows large and the neighboring signals are attenuated. In our proposed radio receiver the filters of Figure 4.8 are MEMS filters. As in the standard system, the received spectrum is filtered by a MEMS filter, amplified by a non-linear LNA, and filtered again by a MEMS filter. The use of MEMS filters' high Q properties will allow us to achieve better rejection of the interfering signals and thereby reduce the SIR.

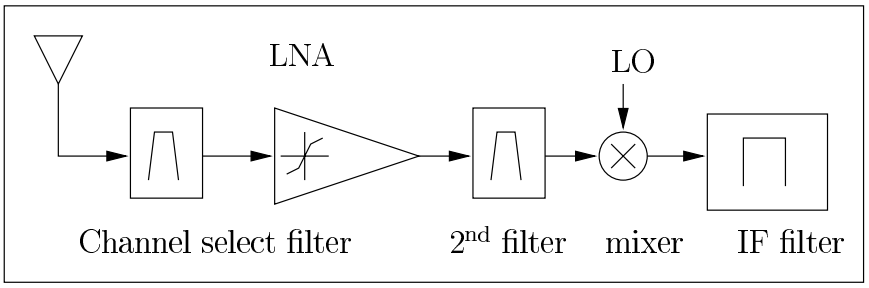


Figure 4.8: System Block Diagram.

Our models have a relatively tight set of restrictions at this point. First we assume there is no phase noise in any of the system oscillators. This allows us to concentrate solely on determining which IM products interfere with the desired signal by simply comparing frequency and phase. Second, for simplicity, we assume the MEMS filter has a Chebychev response. This is a refinement over the model in section 4.1.2 in order to achieve a better representation of the actual performance of the system. Third, we assume that the local oscillator (LO) mixer is ideal and adds no distortion. And finally, we assume that the IF filter has an ideal band-pass response of bandwidth W_{IF}. These assumptions allow the SIR to be calculated at the output of the second filter.

Nonlinear Amplifier Model

The circuit model that we use in this section is a slight modification from the model used in Chapter 3. We select this circuit design because it shares some parameters

that we will use later in this chapter. The designed available gain is 21.85dB and the designed noise figure is 2dB at 900MHz. This amplifier exhibits a quasi-memoryless two tone frequency sweep response and therefore readily submits to modeling by the MIQM from Chapter 3.

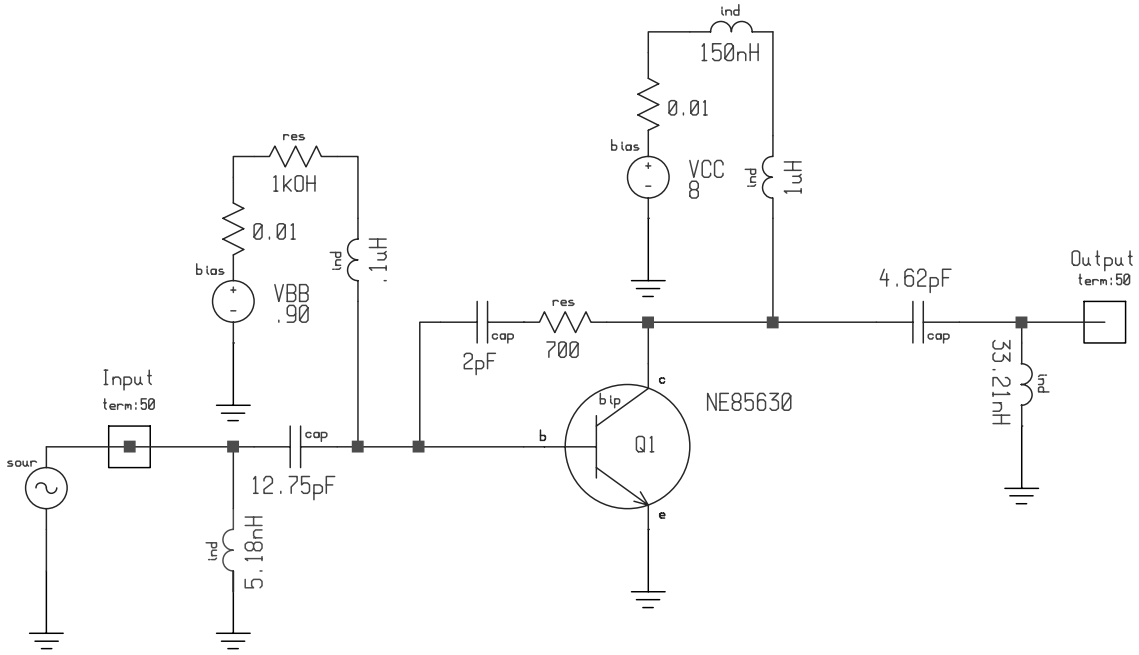


Figure 4.9: Basic LNA.

Following the method outlined in Appendix C the Chebychev polynomial order is selected partially based on the MSE plot in Figure 4.10 and partially based on the work done by Loyka in [45]. It is suggested that a balance be struck between the tightness of the curve fit and the introduction of error due to inaccuracy or noise in the AM/AM and AM/PM measurements. A 24'th degree order is suggested as an optimum order to minimize the effect of measurement error and minimize the errors due to a poor fit. The MSE traces found in Figure 4.10 tend to support this conclusion, with the added comment that for orders higher than the low 30's the fit tends to become very sensitive to the numerical precision of the computer. The I

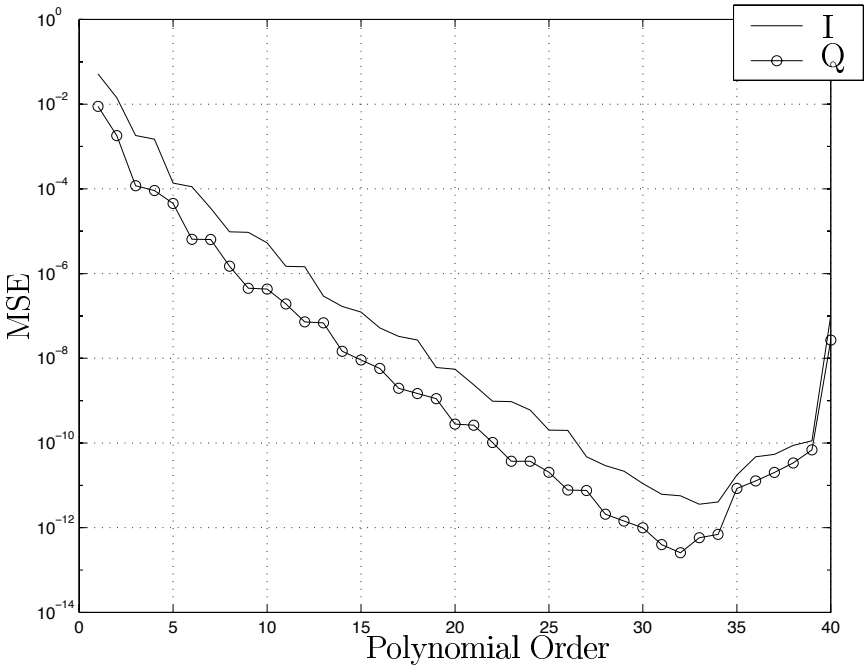


Figure 4.10: MSE of increasing order Chebychev polynomial fit.

and Q transfer functions and the 24'th order Chebychev polynomial fits are shown in Figure 4.11. The polynomials can be written as:

$$p_I(x) = \sum_{i=0}^{N} c_{I_i} x^i, \tag{4.20}$$

and

$$p_Q(x) = \sum_{i=0}^{N} c_{Q_i} x^i, \tag{4.21}$$

where the c_{I_i} and c_{Q_i} terms are the Chebychev coefficients, generated by the method outlined in [39].

In [19], Wass developed a table method to determine the coefficients and frequencies of N'th order M tone intermodulation products. Using the MIQM nonlinearities we can represent our nonlinear function as a polynomial, and are therefore able to determine the intermodulation product magnitudes, frequencies, and phases. Once

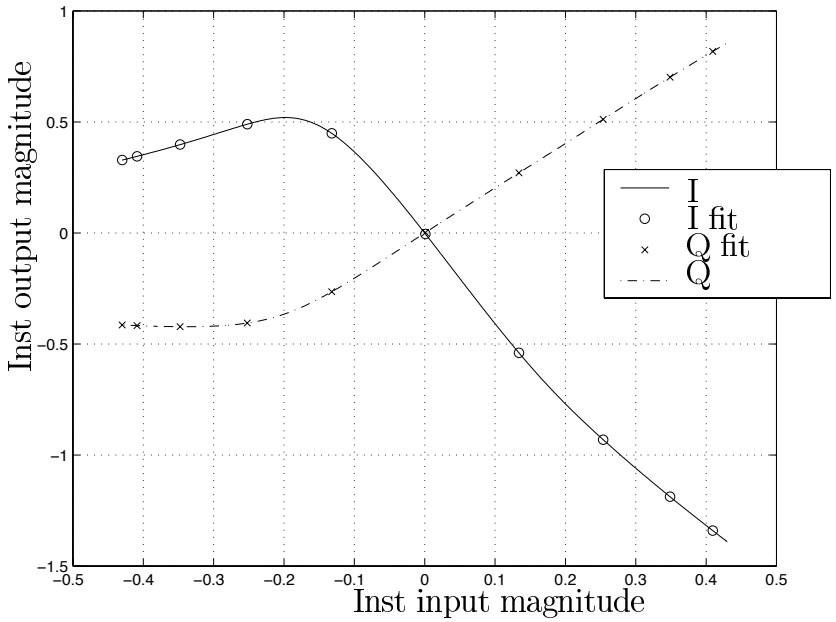


Figure 4.11: MIQM nonlinear transfer functions and 24'th order Chebychev polynomial fit.

the input signal and the group of interfering IM products are determined, we are able to calculate the signal and interference powers at the output of the amplifier. The details of this procedure are found in Appendix C.

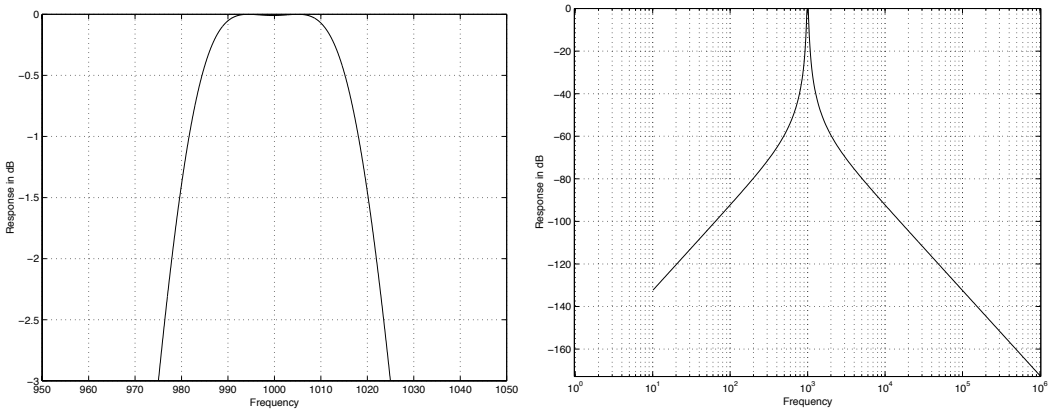
Filter Model Detail

To model the MEMS filter we select the well known Chebychev filter model. While a MEMS filter response may not exactly match the Chebychev filter response, the goal at this point is to introduce the effects of a very narrow filter to the system. We can adjust the Q of the Chebychev model to suit either a narrow (MEMS) filter or a wide (non-MEMS) filter. The Chebychev filter takes the following parameters:

- n - the order of the filter,
- m - the asymptotic slope of the filter,
- f_c - the center frequency of the filter,

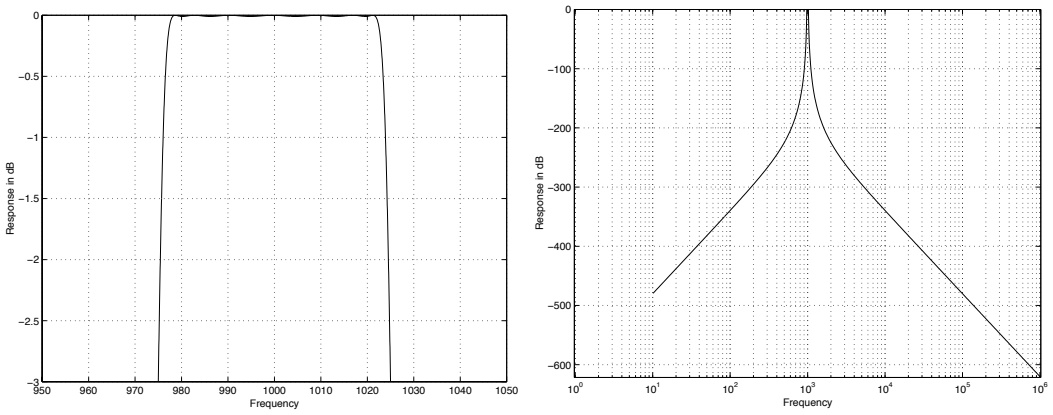
- W - the filter 3dB bandwidth,
- Q - the Q of the filter where $Q = \frac{f_c}{W}$,
- P_{br} - the acceptable passband ripple.

The order of the filter is determined by the asymptotic slope m. Additionally, the filter design is such that Q and m are independent of each other, but the sharpness of the response in the transition region between f_c and $f_c + \frac{W}{2}$ is dependent on both. To illustrate the difference, we examine the passband region of a filter with $Q = 20$ designed with low order ($m = 40$, $n = 2$) in Figure 4.12 and high order ($m = 140$, $n = 7$) in Figure 4.13. Both of these filters have the same Q, but their associated asymptotic performance is markedly different. For our use, we will fix the order of the filter to $n = 3$ and adjust Q. This adjustment will assure similar transition band characteristics as to not prejudice the performance of a lower order filter while allowing us to vary the filter bandwidth. We will denote the filter response as $H_f(f)$.



(a) 3dB Bandwidth = 50Hz (b) Asymptotic Response,$m = -40$

Figure 4.12: Order 2 Chebychev Filter Response illustrating a smooth transition band.



(a) 3dB Bandwidth = 50Hz (b) Asymptotic Response,$m = -140$

Figure 4.13: Order 7 Chebychev Filter Response illustrating a sharp transition band.

4.2.3 Analysis Framework

In this section we develop the analysis framework to determine the system SIR. We define the SIR as the ratio of the output power of the fundamental to the output power of the interfering intermodulation products, at the output to the second filter. The analysis follows the following steps: First compute the output of the first filter due to the incoming three tones. Second, calculate all the magnitude, frequency, and phase of the intermodulation products generated with order N Chebychev polynomial fits to the MIQM amplifier model that are within $\pm\frac{W_{IF}}{2}$ of the carrier (the interfering IM products) using the method detailed in Appendix C. Third, calculate the output of the second filter resulting from the IM products. Fourth, determine the maximum amount of power that the interferers can provide, ΔP, such that the desired SIR is achieved. ΔP is how much stronger the interferers are than the desired tone; e.g. $P_I = P_D + \Delta P$, where P_I is the received power in the interfering tone and P_D is the received power in the desired tone.

Define:

- $s_d(t) = A_d \cos(2\pi f_d t)$ - the desired tone,
- $s_a(t) = A_a \cos(2\pi f_a t)$ - the first interfering tone,
- $s_b(t) = A_b \cos(2\pi f_b t)$ - the second interfering tone.

The input to the amplifier is a filtered version of the sum of the desired and the two interfering signals. Formally, we define:

$$x_{amp}(t) = H_f(f_a)A_a \cos(2\pi f_a t) + H_f(f_b)A_b \cos(2\pi f_b t) + H_f(f_d)A_d \cos(2\pi f_d t). \tag{4.22}$$

For our three tone model we implement Wass' method on a computer with $M = 3$, $N = 24$. c_I and c_Q are taken from the polynomial fit in Section 4.2.2 and the the only the tones that are within a bandwidth W_{IF} around the carrier frequency are considered. This is a valid rule based on our assumption that the mixer is linear, the LO has no phase noise, and the IF filter is an ideal band-pass filter.

When the SIR is determined, or more appropriately the fundamental and interfering tones are determined, we employ a numerical technique to determine the maximum ΔP that will still permit the SIR to remain above a given design goal.

4.2.4 Numerical Results

The parameters of the filter used model are: $n = 3$, $m = 60$, $f_c = 900$MHz, $W_{IF} =$ 100KHz, $P_{br} = 0.001$, and Q which is varied. The frequency plan is: $f_d = 900$MHz, $f_a = 900.9$MHz, $f_b = 901.8$MHz, $P_D = -101$dBm, and $\Delta P = +71$dB. Note that the placement of the interferers is worst case: the third order $2f_a - f_b$ IM product lands exactly at the desired frequency. The amplifier model is shown in Figure 4.9, it provides available power gain of about 20dB with a 2dB noise figure.

For IS-95 systems, the receiver must be capable of demodulating the desired signal in the presence of two $\Delta P = +71$dB interferers placed at f_a and f_b. The desired SIR is determined from the following:

$$\mathrm{SIR} = \mathrm{SNIR} + \mathrm{SNR} - 10\log_{10}\left(10^{\frac{\mathrm{SNR}}{10}} + 10^{\frac{\mathrm{NF}+\mathrm{SNIR}}{10}}\right), \tag{4.23}$$

where SNR is the received signal to noise ratio (no interfering tones are included), SNIR is the signal to noise plus interference ratio at the output of the second filter, and NF is the noise figure of the amplifier. If NF=2dB, SNR=6.3dB, and SNIR=0dB [53] then the required SIR is 2dB. For a weaker received signal of SNR=2.01dB, the required SIR is 25dB. For a received signal of SNR very close to 2dB the required SIR is 50dB. The system can never achieve SNIR of 0dB with a NF=2dB without the received SNR being at least 2dB.

To determine the interferer level at which our target SIR is obtained, we numerically solve for the maximum ΔP such that the target SIR is achieved, while varying Q. Figure 4.14 shows the maximum permitted strong interferer ΔP for a given Q. As we would expect, this value increases as the Q of the filters increase. For SIR=2dB, the minimum filter Q to adequately suppress the intermodulation interference is about 500. If the frequency scheme were such that the strong interferers were spaced 100KHz apart, even more stringent filtering requirements would be required, as shown in Figure 4.15.

4.2.5 Third Order Approximation

For simplicity, most authors select the third order $2f_a - f_b$ IM product's interference as an approximation to the interference of all IM products combined. For a large number of cases involving nonlinearities that are not too severe, this approximation

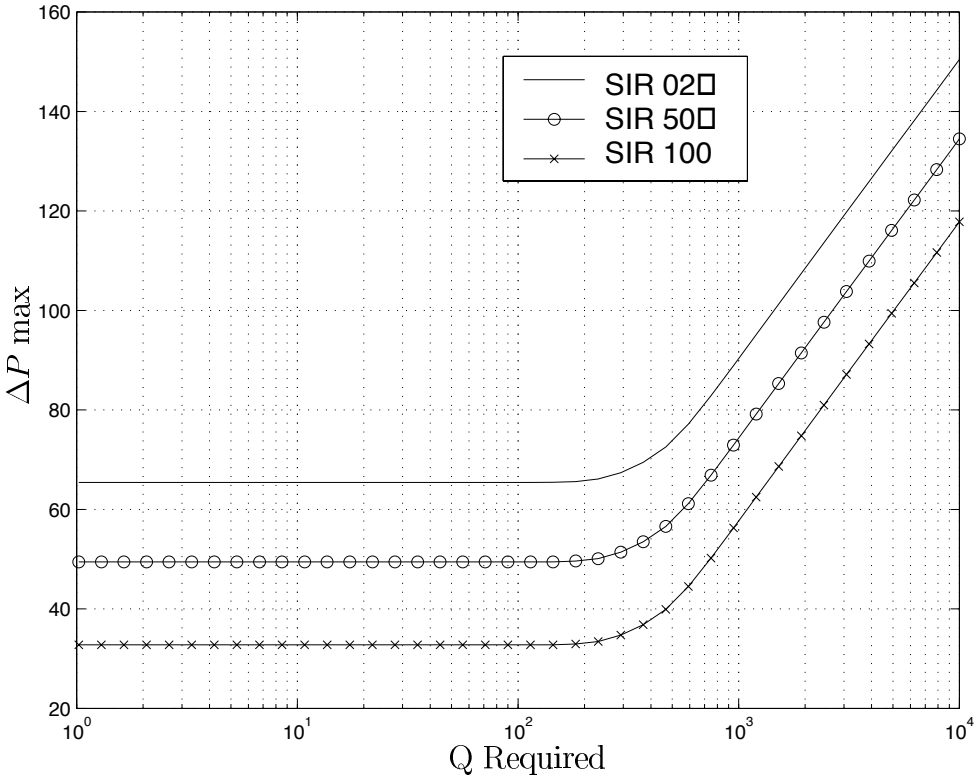


Figure 4.14: Filter requirements for system.

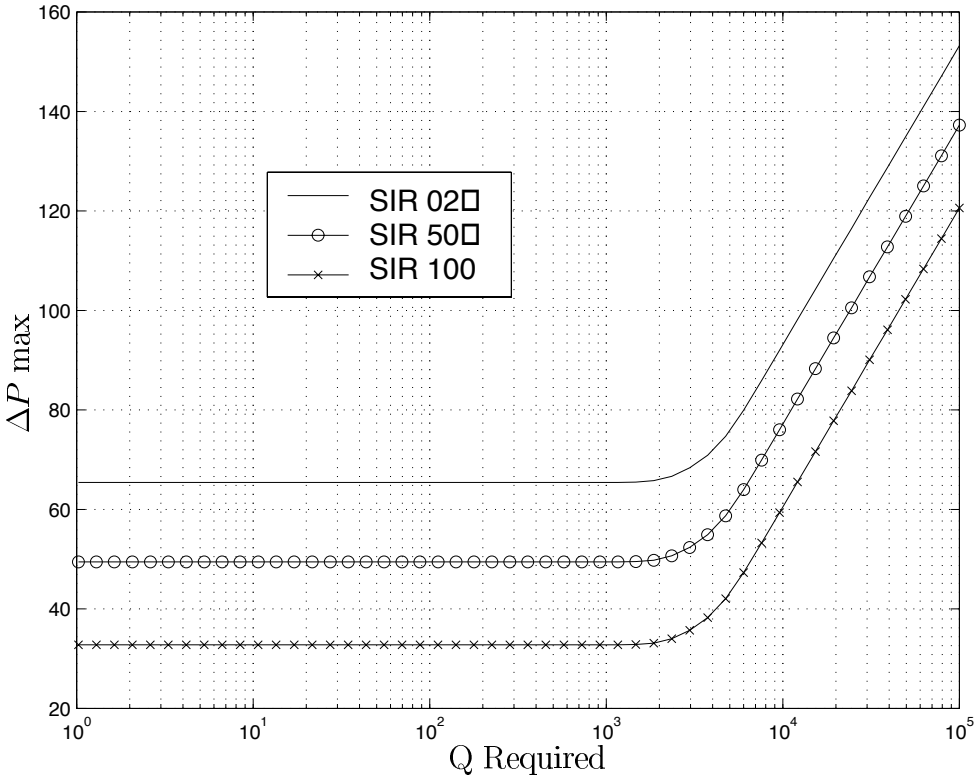


Figure 4.15: Filter requirements for system with spectrally closer interference.

is a good one. After observing the results in Figures 4.14 and 4.15, note the uniform slope and distance between the SIR lines. We approximate our system by using only the third order IM product, and we attempt to characterize the regular slope and distances.

The first approximation fixes the SIR and determines Q's effect on the line slope. We write:

$$SIR = 10\log_{10}(\frac{P_S}{P_N}) \tag{4.24}$$

where

$$P_N \approx \frac{c_3\frac{3}{4}A_1^2A_2}{2R_d}H_f(f_a)^4H_f(f_b)^2, \tag{4.25}$$

and $c_3\frac{3}{4}A_1^2A_2$ is the multiplier for the $i = 3$, $j = 1$, $2f_a - f_b$ intermodulation tone from Appendix C. We note that the filter power response falls off as $\frac{1}{Q^{2n}}$. If $H_f(f_a) = H_f(f_b)$ then the product $H_f^2(f_a)H_f(f_b)$ rolls off as $\frac{1}{Q^{6n}}$. To maintain the same SIR, the interferer power must *increase* as Q^{6n}. This rate increase on a log scale equates to:

$$10\log_{10}(Q^{2n}) = (2n)10\log_{10}(Q), \tag{4.26}$$

or $20n$ dB/decade, which matches the slope in Figures 4.14 and 4.15. In fact the slope of the lines of constant SIR match the asymptotic slope of the Chebychev filter model. This matches intuition since constant SIR is achieved when the attenuated interfering input power matches the unattenuated input power required for a given SIR.

The second approximation fixes the filter Q and determines the ΔP vs. SIR

interaction. We start with (4.25) and write:

$$P_N \approx \frac{\zeta}{2R_d} M_a^2 H_f(f_a)^4 M_b H_f(f_b)^2, \tag{4.27}$$

$$\approx \gamma^6 M_a^4 M_b^2, \tag{4.28}$$

$$\approx \gamma^6 M_a^6, \tag{4.29}$$

where γ represents the product of the fixed filter attenuations. Thus,

$$\text{SIR} = 10\log_{10}\left(\frac{P_S}{\gamma^6 M_a^6}\right), \tag{4.30}$$

$$= 10\log_{10}(P_s) - 10\log_{10}(\gamma^6 M_a^6). \tag{4.31}$$

Reordering the terms in (4.31) we obtain:

$$10\log_{10}(\gamma^6 M_a^6) = 10\log_{10}(P_S) - \text{SIR}, \tag{4.32}$$

$$10\log_{10}(M_a^2) = \frac{10}{3}\log_{10}(P_S) - 10\log_{10}(\gamma^6) - \frac{\text{SIR}}{3}, \tag{4.33}$$

$$= \phi - \frac{\text{SIR}}{3}, \tag{4.34}$$

where ϕ is fixed. This matches the distance between the lines of constant SIR which are approximately $\frac{\text{SIR}}{3}$dB apart.

The result of these approximations is the conclusion that the system is such that a 3rd order approximation gives accurate results for the relative relationship between SIR design goals, and it accurately predicts the change in intermodulation interference as Q changes. However, these relative relationships are give no insight into the absolute magnitude of the interference for which the full analysis is needed since all odd order intermodulation products have a $2f_a - f_b$ component.

4.3 Power efficiency applications

4.3.1 Power efficiency motivation

While most RF engineers are familiar with the nonlinear effects generated by a power transistor being driven into saturation, the goal of this analysis is to use the MIQM to analyze the nonlinearities that arise as a result of a single transistor LNA being driven into cutoff as a result of a low bias point. The tradeoff between power and linearity has been known since the fundamental equations of transistor (and for that matter, vacuum tube) electronics were discovered. The solution to this "problem" has historically been solved by biasing the LNA sufficiently deep into the linear region as to assure the peak input level will not remove the transistor from linear operation. While this may be a solution to the problem of linearity it may not be not power efficient in the sense of power added efficiency (PAE). Power added efficiency is defined as:

$$\text{PAE} = \frac{\text{Power imparted to the desired signal}}{\text{Power supplied from the power supplies.}} = \frac{P_{\text{desired}_{\text{out}}} - P_{\text{desired}_{\text{in}}}}{P_{DC}}. \tag{4.35}$$

In the presence of strong CW interfering tones, the nonlinear effects in an amplifier biased in the strongly linear region of operation will be minimal, but in the amplifier that is biased with much less bias power the nonlinear effects may be significant. The question we will ask is this: with the addition of MEMS filters into the receive chain of weakly biased systems in such a way mitigate the nonlinear effects to match those in a heavily biased system, what will be the resulting gain in PAE? We will find that there is not a clear link between the bias point of an amplifier and the corresponding PAE. As we will see, the PAE of the circuit is largely a function of the circuit design rather than the mitigation of strong interferers.

The second issue addressed in this section is the PAE gains in a a fixed bias system with the addition of MEMS filters to the receive chain. We know that strong interfering signals cause the output gain of the desired signal to compress. Since the bias power will be fixed, the attenuation of the interfering signals will increase the gain of the desired signal, and thus increase the PAE of the system. While this link may seem intuitive, our goal is to quantify the power savings using the MIQM.

The remainder of this section will detail the design of the various amplifiers that will be analyzed, the method of determining the power supplied to the circuit, the analysis of the varied bias PAE, and the analysis of the fixed bias PAE.

4.3.2 Amplifiers used in the analysis

The amount of bias power to a circuit is largely dominated by the bias point selected in the amplifier design. We will use a similar circuit design to that found in Figure 4.9 but the V_{BB} bias voltage will be varied. When designing microwave amplifiers the combination of the transistor, the feedback network and the bias network is characterized by S-parameters. Using the measured S-parameters, the design of the input and output matching networks are done on the Smith chart to match some desired performance criteria. For low noise amplifiers, gain of the circuit is typically sacrificed for a low noise figure. For our circuits, the noise figure of 2dB is used and the amplifiers are designed to provide as much gain as stability criteria allow. Each of the following amplifiers has been checked for input and output stability, quasi-memoryless operation, and the maximum possible resulting gain. The designs are illustrated in Figures 4.16, 4.17, 4.18, 4.19 and 4.20. Their one tone compression performance is shown in Figure 4.21. Note the gain *expansion* of the 0.70V LNA, this is due to the transistor turning on as the input signal becomes larger; with a small

input signal the transistor is mostly turned off.

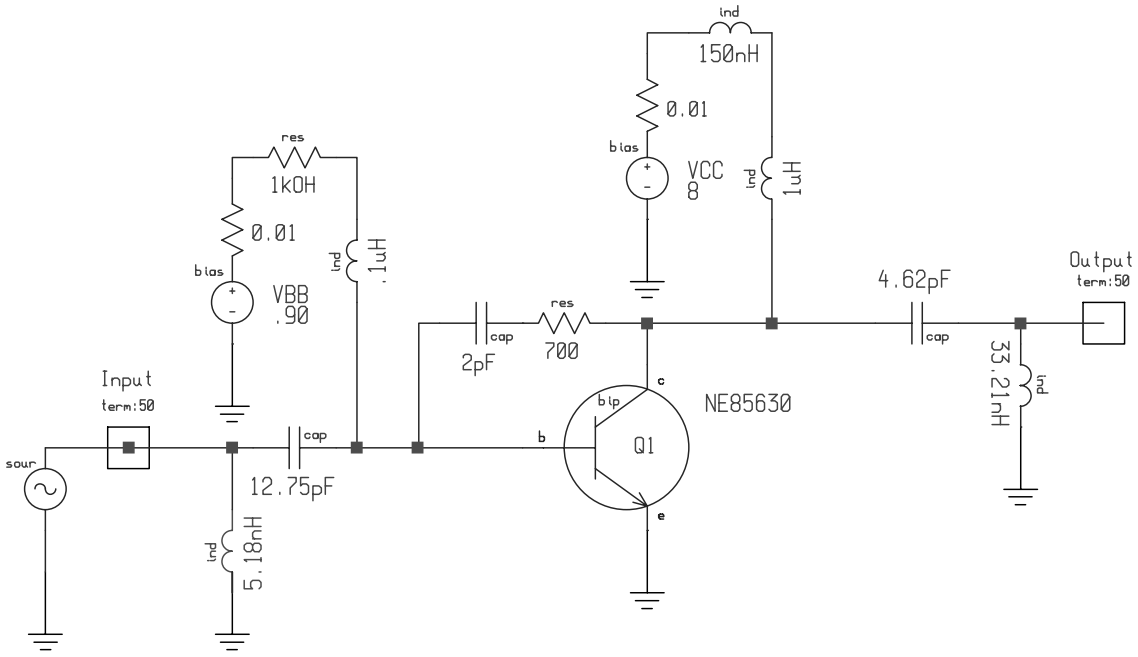


Figure 4.16: Quasi-memoryless LNA with 0.90V bias, NF=2dB, Gain=20.70dB.

4.3.3 Power supplied

To determine the power supplied to each of the different circuits we measure the supply power consumed under different input conditions using a circuit simulator. In practice, this information can be measured from a circuit and incorporated into the behavioral model. When the relationship is known, the supply power can be determined by performing an interpolated table lookup. Figure 4.22 illustrates the measured relationship between input power and supply power for each of the circuits under consideration.

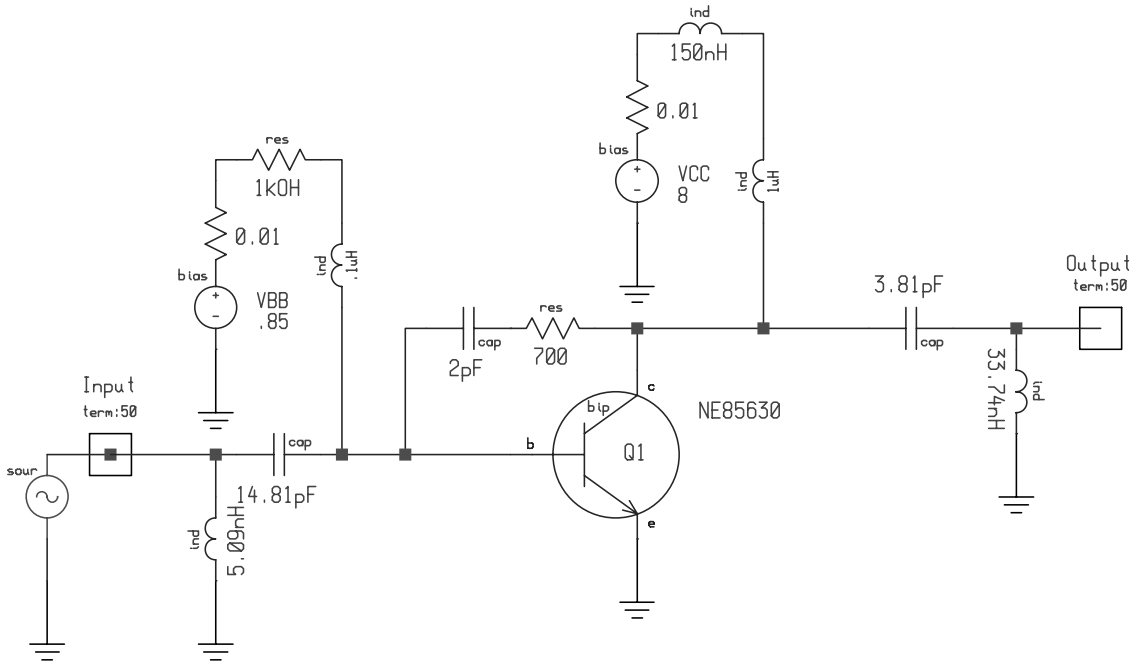


Figure 4.17: Quasi-memoryless LNA with 0.85V bias, NF=2dB, Gain=19.41dB.

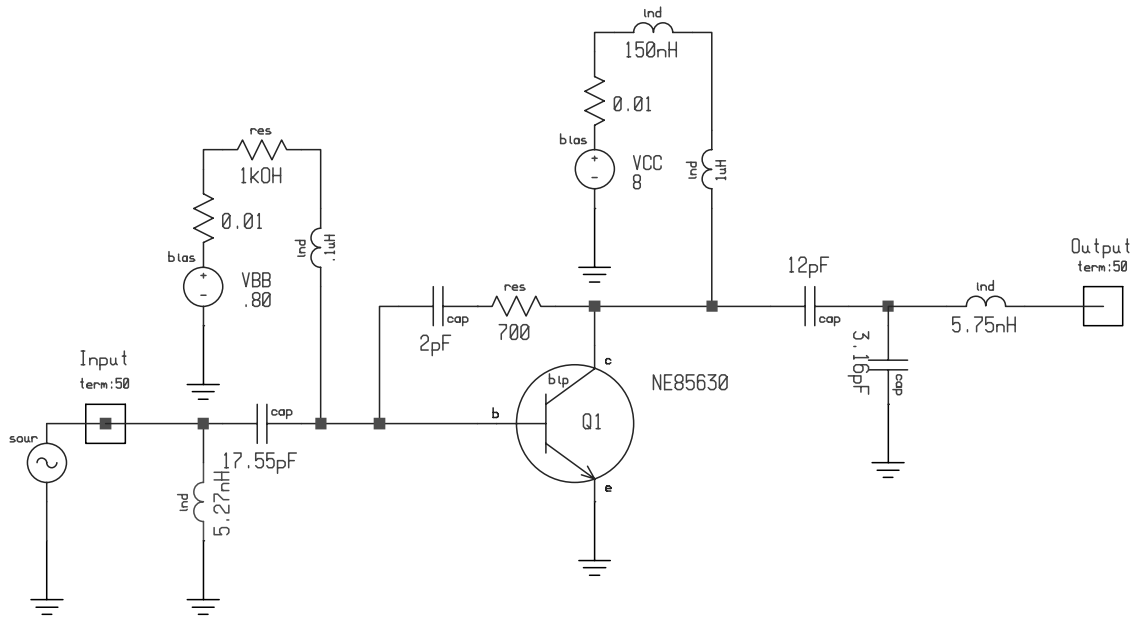


Figure 4.18: Quasi-memoryless LNA with 0.80V bias, NF=2dB, Gain=17.28dB.

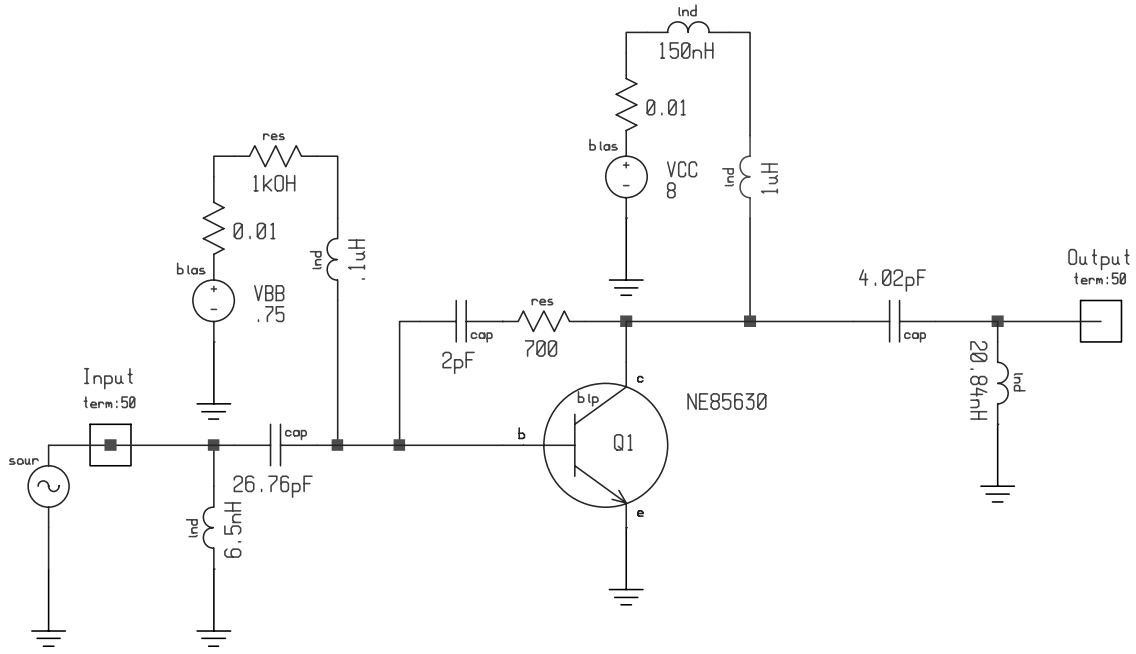


Figure 4.19: Quasi-memoryless LNA with 0.75V bias, NF=2dB, Gain=14.04dB.

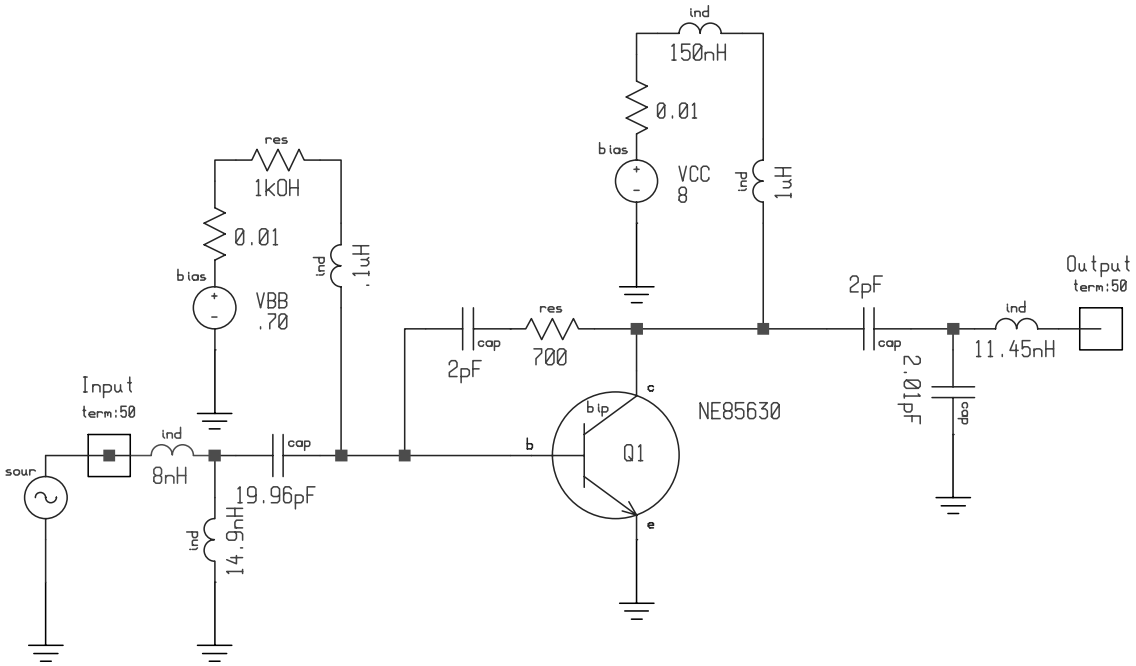


Figure 4.20: Quasi-memoryless LNA with 0.70V bias, NF=2dB, Gain=6.99dB.

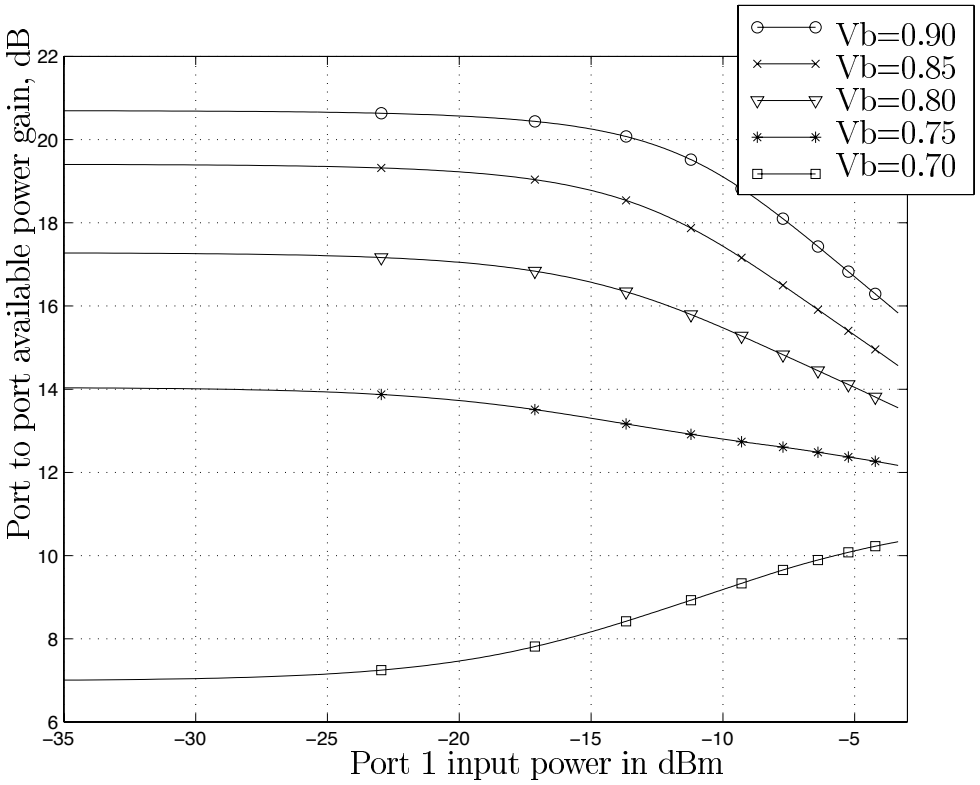


Figure 4.21: One tone compression illustration for Figures 4.16 - 4.20.

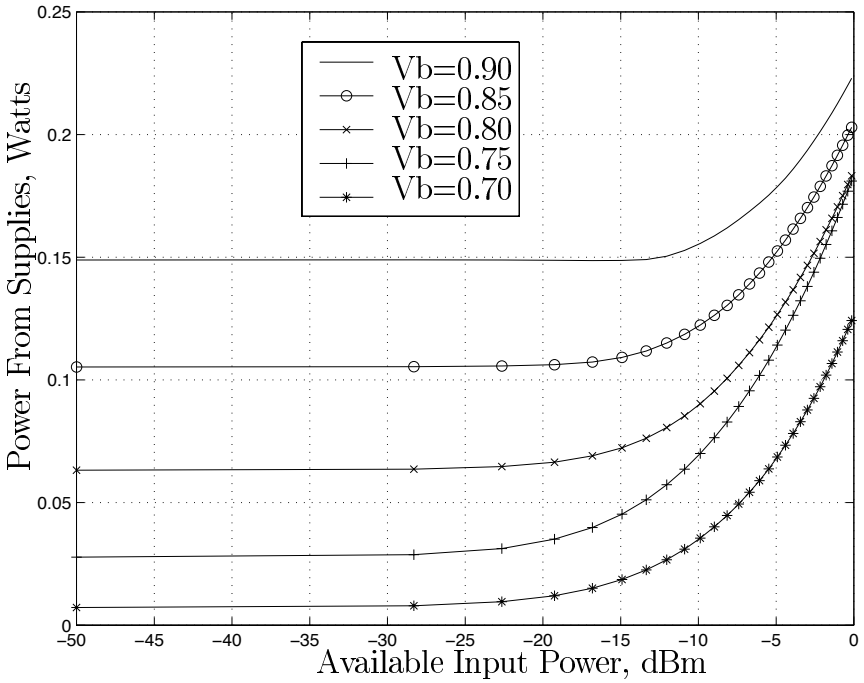


Figure 4.22: Relationship between input power and supplied power.

4.3.4 Varied bias PAE

One of the main questions to be addressed in this Chapter is the power savings that MEMS filters can foster. Fix the SIR requirement for the front end system, if the amplifier experiences the effects of strong interfering tones while attempting to receive a weaker desired tone, the nonlinear products that we have been considering come into play. If the front end filters are wide the amplifier must maintain a degree of linearity to meet the design goal. However, if the front-end filters are very selective and attenuate the interferers adequately the linearity requirement on the amplifier can be relaxed. The bias point of the amplifier typically determines the amount of power consumed. Compared to amplifiers with high biasing power, amplifiers with lower biasing power are typically less linear because the transistor is biased closer to cutoff. Given the appropriate MEMS filter, nearby interferers can be attenuated to such a degree that the SIR requirement may still be achieved. The task, then, is to quantify the power added efficiency gains that can be achieved with the amplifiers whose bias power is lower, and what the corresponding filtering requirements are.

To determine the answer to this question, we select the amplifier models illustrated in Figures 4.16-4.19. These amplifiers have been designed to achieve their maximum gain for which the noise figure does not exceed 2dB. The amplifier with bias voltage 0.70 has been dropped from consideration because of the overall poor gain response. The remaining amplifiers are modeled using the MIQM and the method described in Appendix C with $N = 24$. The frequency plan is the same as we considered in Section 4.2.4, $f_d = 900$MHz, $f_a = 900.9$MHz, and $f_b = 901.8$MHz. The interfering IM products are selected if they are within a $W = 100$kHz wide region centered on f_d. The filtering requirements for each amplifier is given in Figure 4.23 when the input desired tone is -101dBm and the two strong interferers are each received at -30dBm.

While not extraordinarily steep, there is a trend toward relaxing filter requirements

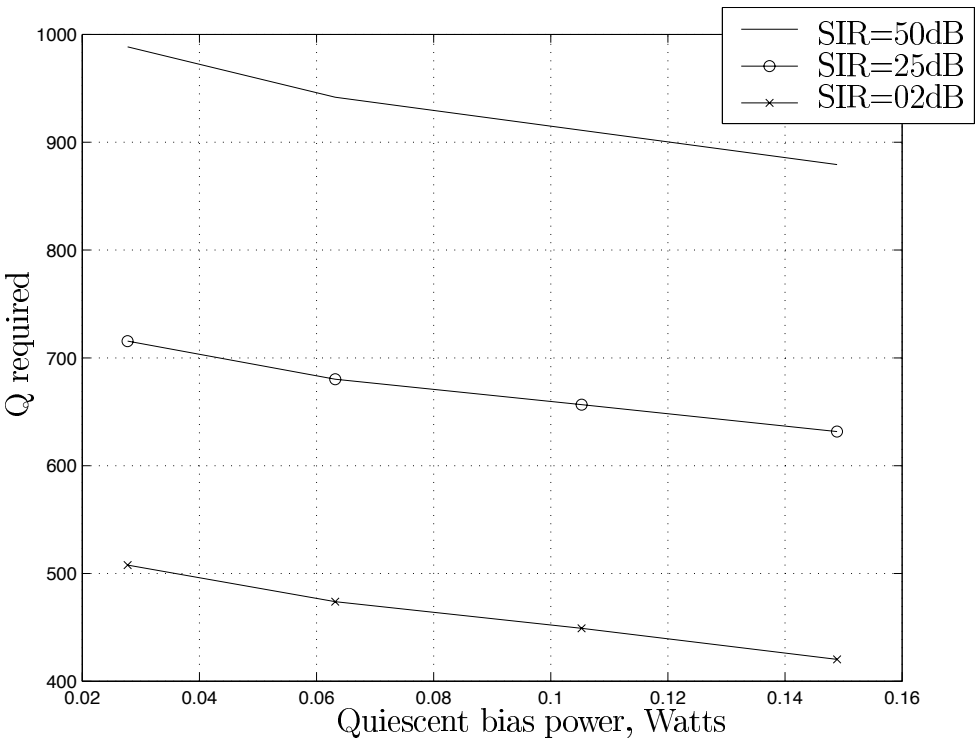


Figure 4.23: Q required for SIR design goal and different amplifiers, input is -101dBm.

as the bias power increases.

If the amplifier is driven harder by the desired signal (-41dBm and the interferers at -10dBm), as in Figure 4.24, an interesting phenomenon appears. The amplifiers with smaller bias power requires less filtering to achieve SIR=2dB than the amplifiers with greater bias power. This is explained by noting the one tone compression plot in Figure 4.21. While the $V_b = 0.75$ amplifier may have a smaller gain than the rest, its 1dB compression point is is at a larger input signal level than the remainder of the circuits. This implies that the $V_b = 0.75$ may, in fact, be more linear than the $V_b = .90$ amplifier. Also note that the rate at which the gain decreases is not as fast as the remaining amplifiers, which serves to strengthen this conjecture. The conclusion to draw here is that the linearity of an amplifier is not necessarily a function of the

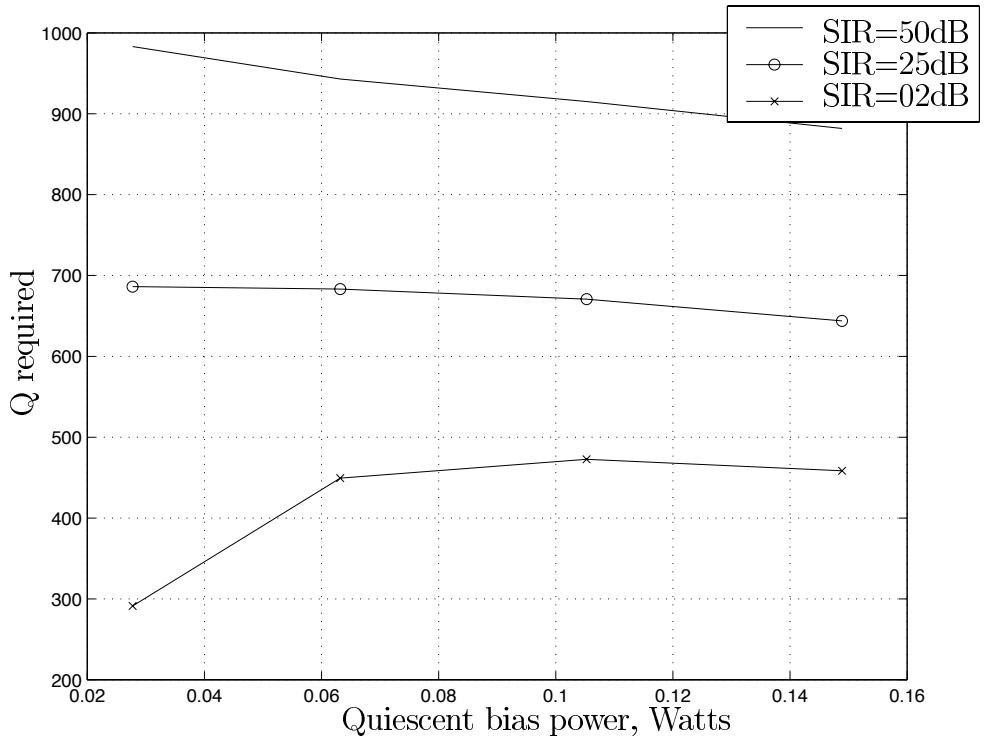


Figure 4.24: Q required for SIR design goal and different amplifiers, input is -41dBm.

bias power supplied, even for circuits with similar designs.

The next result to examine is the PAE for the amplifiers when utilizing the calculated Q that will give the desired SIR performance. The results of the PAE determination for the system driven with a -101dBm desired input and two -30dBm CW interferers. Under these conditions, the amplifiers with the lower bias power achieve a 10-14% increase in PAE as compared with the most heavily biased amplifier.
If the input of the amplifier is driven by larger signals, -41dBm desired power and -10dBm interferer power the PAE results are shown in Figure 4.26. Just the opposite effect is seen here; the circuits with smaller bias power have smaller PAE than the circuits that receive more bias power. At first glance, this result seems counter-intuitive. But, upon examination of the PAE calculation from (4.35) this effect can be explained. We know, from the power traces in Figure 4.22, that the bias power

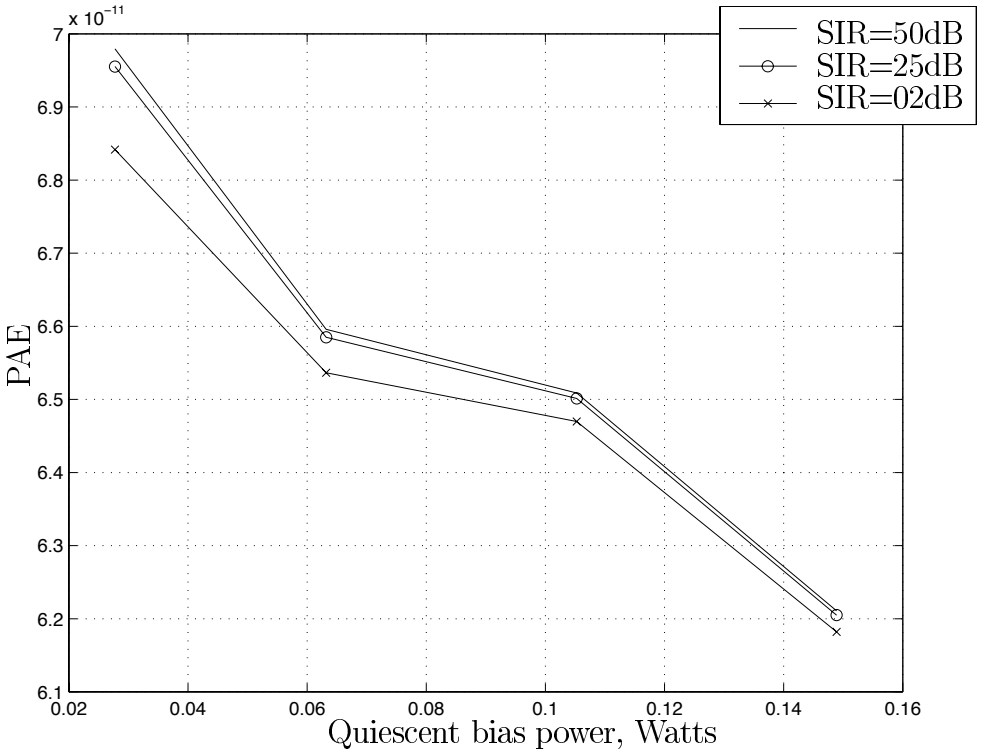


Figure 4.25: PAE for SIR design goal and different amplifiers, input is -101dBm, interferers are -30dBm.

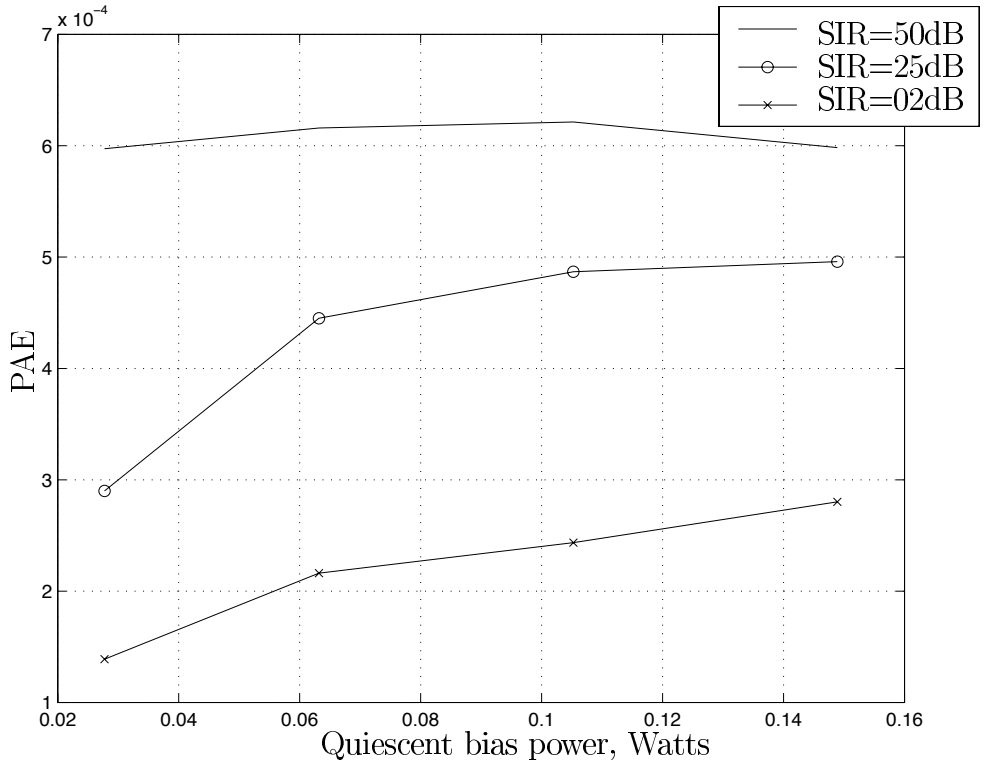


Figure 4.26: PAE for SIR design goal and different amplifiers, desired signal power is -31dBm, interferers are -10dBm.

decreases as the overall input signal power is attenuated. We also know, from Figure 4.21 that the fundamental gain decreases with a decrease in bias level. However, what we do not know is how the decreases react in relationship with each other. If the gain of the system shrinks with the bias power, then the only way to show PAE gains is to make sure the bias power decreases at a faster rate, or else the PAE may actually grow worse.

The conclusion to draw is that the PAE of an amplifier is not clearly related to the bias power applied. The PAE is heavily dependent on the actual design of the microwave circuit. A prudent design for fixed noise figure at a small bias power may be more linear than a design at higher bias powers. This coupled with the fact that the maximum gain of a system with shrinking bias power also shrinks. Thus, changes in both the numerator and denominator of (4.35) lead to no clear intuition about PAE, but the result can be quantified by using the MIQM.

The results of this section must be viewed in light of the assumptions made. In selecting the amplifier designs, the noise figure of the circuits was fixed to 2dB. This restriction was made in an attempt to make fair comparisons between amplifiers driving 50Ω loads. Some examination of the basic assumptions may need to be made in the future when circuit designers examine this work. For instance, it may be advantageous to design the load for a different impedance and match the input of the next element with an ideal transformer in order to achieve better gain or noise performance. Additionally, we do note that matching small circuit gain across amplifiers for comparison purposes was considered. If matching gain is done under the restriction of matched noise figure the comparison of response is not fair because the more heavily biased circuits must be intentionally mismatched in order to limit their gain to be equal to the least biased circuit. On the other hand, relaxing the

equal noise figure restriction may lead to an interesting tradeoff between linearity and noise and is an open question.

4.3.5 Fixed bias PAE

Since the relationship between bias power and PAE seems to be heavily design dependent, if we fix the design and analyze the effects of a MEMS filter on PAE we may be able to gain some insight. The main problem with the PAE determination in Section 4.3.4 is that both the numerator and denominator of (4.35) varied with the bias power. To avoid this we will examine the PAE with a fixed amplifier design. Thus the supply power will remain fixed and the denominator will as well. The gain variation is the only effect that will change the PAE.

In Figure 4.27 the gain for the desired tone is shown for all SIR$\geq$2dB. Since the bias power to the system is fixed, the gain compression seen contributes directly to the reduction in PAE, seen in Figure 4.28. For a given strength of interferer, increasing filter Q leads to greater power efficiency up to the limit of the small signal gain of the amplifier.

4.4 Conclusions

In this chapter's sub-sampling sections we present a model for direct down conversion systems that use MEMS filters and sub-sampling. This architecture allows the elimination of several high power oscillators. We develop 3 restrictions on the choices of parameters to eliminate interference, and we show that the replicated additive thermal noise is bounded under a fourth restriction on the MEMS filter slope. We also show that the additive resistor noise can be dominated by selecting intelligent

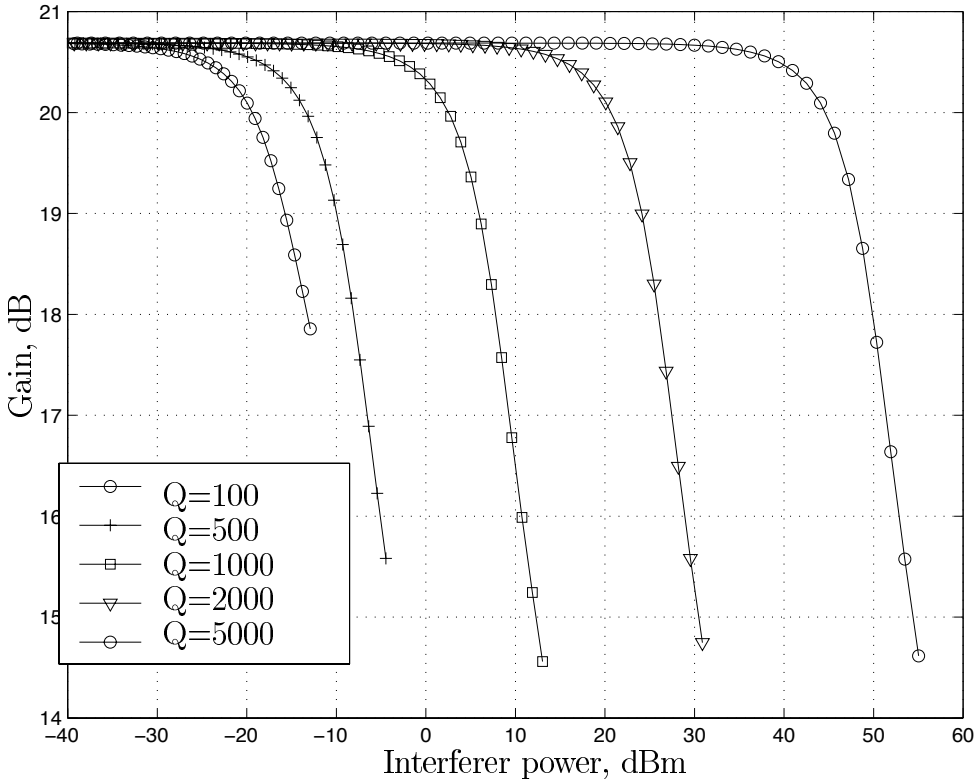


Figure 4.27: Gain for SIR$\geq$2, $V_b = 0.90$V, desired signal level is -31dBm.

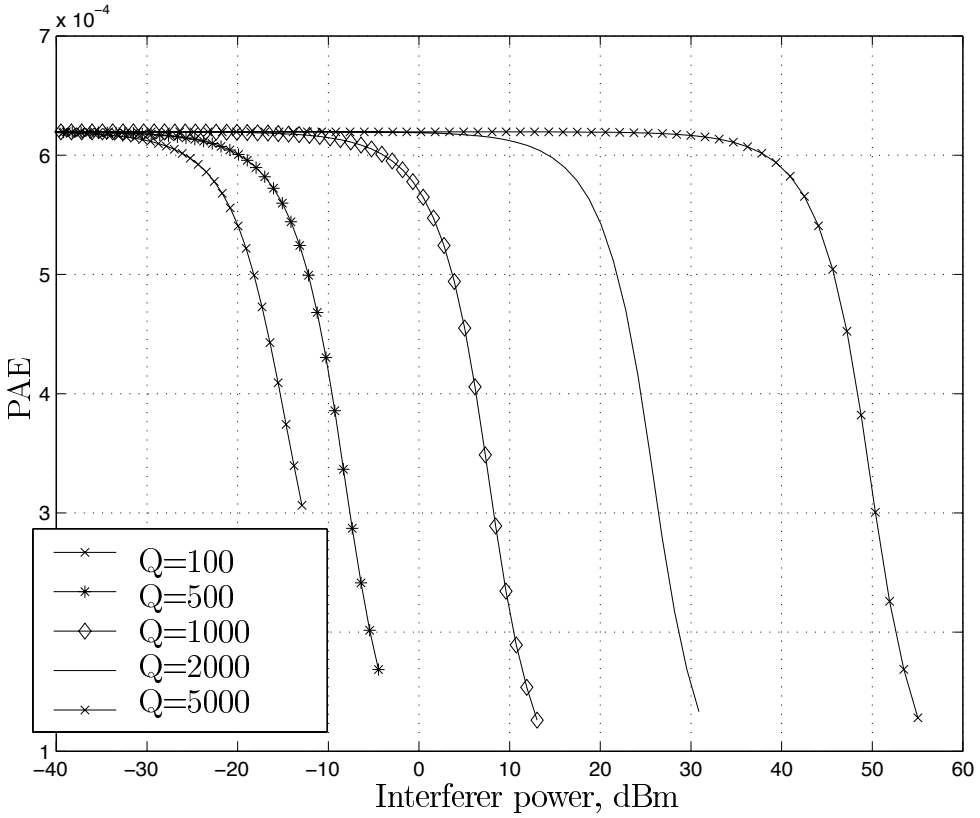


Figure 4.28: PAE for SIR$\geq$2, $V_b = 0.90$V, desired signal level is -31dBm.

gain values in the front end LNA, which is contrary to [48] and [22].

Our numerical results consist of 1) the tradeoff between the sampling rate (f_s), the filter slope (m_2), and the SNR loss as a result of wide-band noise and completely dominated resistor noise 2) the tradeoff between f_s, the LNA gain, and the SNR loss due to wide-band noise and resistor noise 3) tradeoff between a given RWNR requirement and LNA gain for a fixed f_s. These results will allow direct down-converted MEMS based radios to be designed around given system requirements.

The cosite interference sections provide a framework for evaluating the effects of nonlinearities in amplifier front-ends. It can be used to accurately predict the required Q for a given design, and because Wass' model is not dependent on either the filter shape or the amplifier characteristic, modifying the basic system and generating new results costs little compute time, resulting in a repeatable design methodology. The key contributions of this work are to quantify our intuition that narrower bandpass filters give better interferer rejection and to determine the maximum nearby interference level for a desired system performance. This type of analysis will become advantageous as narrower MEMS filters are designed.

The power added efficiency section raised some question as to the efficacy of filters on PAE. The question as to how much efficiency can be gained by the use of MEMS filters is answered in two ways. First, the link between PAE and bias power is highly dependent on the circuit design. The reduction in bias power can be completely offset by the reduction in amplifier gain. Additionally, the required filtering to reduce nonlinear effects is dependent upon the circuit design as well. The design of the amplifier is a question better suited for optimal circuit design research, while the system performance analysis is effectively handled by behavioral modeling. The second manner in which this question can be answered is by fixing the design

of the amplifier and determining the PAE increase by reducing the gain compression brought on by strong interferers. These effects are quantified for a given circuit model using the MIQM.

CHAPTER 5
Summary of Contributions and Conclusions

In this thesis we have analyzed many different situations where the transmission of information is corrupted by noise. This chapter offers a review of the original contributions that can be found in the preceding pages.

In Chapter 2 we motivated and developed an efficient, recursive method for calculating the probability of undetected codeword error. Our algorithm assumes independent noise statistics from symbol to symbol but allows different noise distributions. The efficiency of the algorithm is dependent on the number of different noise distributions present, and can have exponential complexity in N as the number of noise distributions approaches N. In this situation, our algorithm offers bounds in a fashion that is compute-limited (the bound grows monotonically tighter as the amount of CPU cycles applied increases). While the best algorithm found [1] may approach our efficiency for the case with few noise distributions, it does not offer a bound in the case where there are many distributions.

We applied the recursive method to the case of product code transmission in the presence of an omniscient fixed power partial band jammer. Code rates were selected using both asymptotic and finite length analysis. It was shown that the probability of undetected error is an important consideration in accurately determining the optimum

constituent code rates.

In Chapter 3 we motivated and proposed the MIQM, a new nonlinear modeling technique which is used to analyze, rather than simulate, the effects of a nonlinear element in receiver front-ends. This method has advantages over many of the existing modeling techniques discussed. Primarily, it allows for rapid analysis of weakly and strongly nonlinear systems with large or small dynamic range inputs. Additional advantages over existing methods include numerical precision over FFT based discrete methods, functional input-output relationships over the IQM, and accuracy of results over analytical methods. Once the system is known to be memoryless, or have a limited degree of memory, the MIQM will produce accurate results that match the output of more numerically complex simulation systems.

In Chapter 4 we analyze the effect that MEMS filters have on combating interference in various systems. We show that if proper system parameters are set, the mixer and local oscillator (LO) found in a traditional radio front-end can be eliminated. We give conditions on the sub-sampling system parameters to assure self-interference free communication and we show that the replicated sampling resistor noise is dominated by the thermal noise at the antenna. Additionally, we show that under another condition, the replicated wide-band noise is finite. We apply this analysis to an example system and show the performance tradeoff between sampling rate, filter shape, and interference.

The second issue that we explored in Chapter 4 is the analysis of the suppression of cosite interference with MEMS filters. We develop a system to quantify the interference and to show the tradeoff between the SIR design goal and the filter Q requirements. The results show that the SIR achievable with a given Q is strongly related to the order of the filter used. This is not only reinforces intuition, but we

quantify that intuition with the direct calculation and illustration of the tradeoff.

The final issue addressed is the power savings that may be achieved by using MEMS filters in receiver front-ends that employ nonlinear amplifiers. It was thought that the use of selective filters may allow the use of amplifiers that are more power efficient and at the same time less linear. By applying a MEMS filter to attenuate strong interferers in a cosite scenario, intuition held that the overall system SIR could be held constant while the power used by the amplifier decreased and thus show an increase in PAE. We show that the application of a MEMS filter to a fixed bias power amplifier increases the PAE in the presence of strong interferers, but that there is not a clear link between decreasing the amplifier bias power and increases in PAE under fixed noise figure constraints.

The primary contributions of this thesis are the methods for the analysis of communication systems in the presence of interference. The structure of the algorithm proposed in Chapter 2 leads to the ability to determine probability statistics that are important in both direct calculation of undetected error and in analyzing retransmission systems. Further work in this area can be done in determining the zero pattern generation algorithm, such that a global search and sort is not needed. The MIQM analysis technique can, most likely, be further refined and used for analysis of adjacent power interference and other nonlinear system behaviors. Finally, the PAE results in Chapter 4, need further exploration by researchers versed in optimum circuit design. While we show that the relationship between amplifier bias power and PAE is not clear in general, some design optimization research may shed light on this relationship.

APPENDICES

APPENDIX A

Derivation of probability of correct, incorrect, and failure for M-ary orthogonal non-coherent symbol detection using the ratio threshold test (RTT)

A.1 AWGN Preliminaries

In a transmission of one of many possible signals the ratio threshold test provides side information as to the reliability of the decision of which signal is detected. The output of the energy detectors for an M-ary receiver are denoted as Z_i where $i = 1 \ldots M$. The general rule is to select symbol j as the detected symbol if $\max_i Z_i = Z_j$. The RTT flags the detected symbol as un-reliable if $Z_j < \gamma Z_i$ for any $i \neq j$, with $\gamma > 1$. The special case of $\gamma = 1$ degenerates to the general decision rule.

The block diagram of the decision device is shown in Figure A.1.

The transmitted signals are

$$s_i(t) = \sqrt{\frac{2E_s}{T_s}} \cos\left(2\pi\left(f_c + \frac{i}{T_s}\right)t\right), \qquad i = 1 \ldots M, \qquad 0 < t < T_s, \qquad \text{(A.1)}$$

where the offset frequency is selected to be a multiple of $\frac{1}{T_s}$ so the signals are orthogonal. The low-pass equivalent of these signals are:

$$s_{li}(t) = \sqrt{\frac{2E_s}{T_s}} \cos\left(2\pi\frac{i}{T_s}t\right), \qquad i = 1 \ldots M, \qquad 0 < t < T_s. \qquad \text{(A.2)}$$

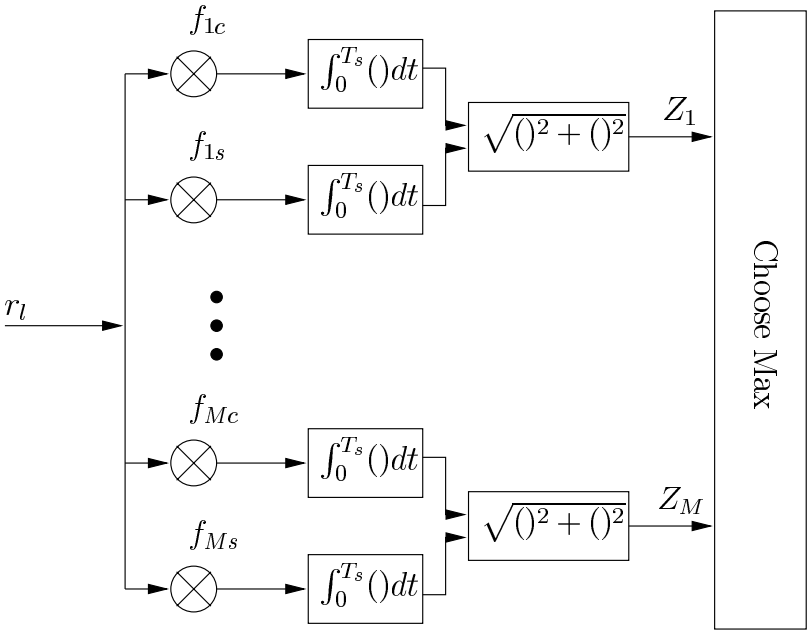


Figure A.1: M-ary orthogonal non-coherent detector

The received signals are:

$$r_(t) = \sqrt{\frac{2E_s}{T_s}} \cos\left(2\pi\left(f_c + \frac{i}{T_s}\right)t + \phi\right) + n(t), \qquad i = 1 \ldots M, \qquad 0 < t < T_s, \tag{A.3}$$

where $n(t)$ is an AWGN process, mean 0 variance $\frac{N_0}{2}$, and ϕ is a uniformly distributed phase error. The low-pass equivalent of the received signal is:

$$r_l(t) = \frac{1}{2}\sqrt{\frac{2E_s}{T_s}} \cos\left(2\pi\frac{i}{T_s}t + \phi\right) + n(t), \qquad i = 1 \ldots M, \qquad 0 < t < T_s. \tag{A.4}$$

The correlating functions for $i'th$ signal are:

$$\begin{aligned} f_{ic} &= \sqrt{\frac{2}{T_s}} \cos\left(2\pi\frac{i}{T_s}t\right), \\ f_{is} &= \sqrt{\frac{2}{T_s}} \sin\left(2\pi\frac{i}{T_s}t\right). \end{aligned} \tag{A.5}$$

The optimum receiver in AWGN selects the signal that maximizes:

$$Z_i = \sqrt{Z_{ic}^2 + Z_{is}^2}. \tag{A.6}$$

A.1.1 Outputs with signal present

The output of the cosine correlator corresponding to the transmitted signal is:

$$\begin{aligned} Z_{1c} &= \int_0^{T_s} r_l(t) f_{1c}(t) dt, \qquad i = 1 \ldots M, \\ &= \sqrt{E_s} \cos\phi + \sqrt{\frac{2}{T_s}} \int_0^{T_s} n(t) dt. \end{aligned} \tag{A.7}$$

Z_{1c} is Gaussian distributed with mean

$$\begin{aligned} \overline{Z_{1c}} &= \mathrm{E}\left[Z_{1c}\right] \\ &= \sqrt{E_s} \cos\phi, \end{aligned} \tag{A.8}$$

and variance

$$\begin{aligned} \mathrm{Var}(Z_{1c}) &= \mathrm{E}\left[Z_{1c}^2\right] - \mathrm{E}^2\left[Z_{1c}\right], \\ &= \frac{N_0}{2}. \end{aligned} \tag{A.9}$$

The output of the sine correlator is derived similarly:

$$Z_{1s} = \sqrt{E_s} \sin\phi + \sqrt{\frac{2}{T_s}} \int_0^{T_s} n(t) dt, \tag{A.10}$$

with similar mean

$$\begin{aligned} \overline{Z_{1s}} &= \mathrm{E}\left[Z_{1s}\right] \\ &= \sqrt{E_s} \sin\phi, \end{aligned} \tag{A.11}$$

and variance

$$\begin{aligned} \mathrm{Var}(Z_{1s}) &= \mathrm{E}\left[Z_{1s}^2\right] - \mathrm{E}^2\left[Z_{1s}\right], \\ &= \frac{N_0}{2}. \end{aligned} \tag{A.12}$$

For envelope detection, the input to the decision device is:

$$Z_1 = \sqrt{Z_{1c}^2 + Z_{1s}^2}. \tag{A.13}$$

Since Z_{1c} and Z_{1s} are Gaussian distributed random variables, with non-zero mean, Z_1 is a Rician random variable with non-centrality parameter $s^2 = \overline{Z_{1c}}^2 + \overline{Z_{1s}}^2 = E_s$. The distribution of Z_1 is written as:

$$\begin{aligned} f_{Z_1}(z) &= \frac{z}{\sigma^2} \mathrm{e}^{-\frac{z^2+s^2}{2\sigma^2}} \mathrm{I}_0\left(\frac{sz}{\sigma^2}\right), \\ &= \frac{2z}{N_0} \mathrm{e}^{-\frac{z^2+E_s}{N_0}} \mathrm{I}_0\left(\frac{2zE_s}{N_0}\right), \end{aligned} \tag{A.14}$$

where I_0 is the modified Bessel function of the 0'th kind. As a side note, when the parameter to I_0 exceeds 700 the approximation:

$$\mathrm{I}_0(x) \approx \frac{1}{\sqrt{2\pi x}} \mathrm{e}^x \tag{A.15}$$

is necessary on a computer. The corresponding cumulative distribution function is written as:

$$\begin{aligned} F_{Z_1}(z) &= 1 - \mathrm{Q}_1\left(\frac{s}{\sigma}, \frac{z}{\sigma}\right), \\ &= 1 - \mathrm{Q}_1\left(\sqrt{\frac{2E_s}{N_0}}, z\sqrt{\frac{2}{N_0}}\right), \end{aligned} \tag{A.16}$$

where Q_1 is the Marcum-Q function with parameter 1. From [54] we can evaluate $\mathrm{Q}_1(\alpha, \beta)$ as follows:

$$0 \le \alpha < \beta$$

$$Q_1(\alpha, \beta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1 + \frac{\alpha}{\beta}\sin\theta}{1 + 2\frac{\alpha}{\beta}\sin\theta + \left(\frac{\alpha}{\beta}\right)^2} \mathrm{e}^{\frac{-\beta^2}{2}\left(1+2\frac{\alpha}{\beta}\sin\theta+\left(\frac{\alpha}{\beta}\right)^2\right)} d\theta, \tag{A.17}$$

$$0 \le \beta < \alpha$$

$$Q_1(\alpha, \beta) = 1 + \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\left(\frac{\beta}{\alpha}\right)^2 + \frac{\beta}{\alpha}\sin\theta}{1 + 2\frac{\beta}{\alpha}\sin\theta + \left(\frac{\beta}{\alpha}\right)^2} \mathrm{e}^{\frac{-\alpha^2}{2}\left(1+2\frac{\beta}{\alpha}\sin\theta+\left(\frac{\beta}{\alpha}\right)^2\right)} d\theta, \tag{A.18}$$

and

$$0 \le \alpha = \beta$$

$$Q_1(\alpha, \alpha) = \frac{1}{2}\left[1 + \mathrm{e}^{-\alpha^2} \mathrm{I}_0(\alpha^2)\right]. \tag{A.19}$$

A.1.2 Outputs with no signal present

The output of a cosine correlator corresponding to any other signal than the transmitted signal is:

$$\begin{aligned} Z_{2c} &= \int_0^{T_s} r_l(t) f_{2c}(t) dt, \qquad i = 1 \ldots M, \\ &= 0 + \sqrt{\frac{2}{T_s}} \int_0^{T_s} \cos\left(2\pi \frac{1}{T_s} t\right) n(t) dt. \end{aligned} \tag{A.20}$$

Z_{2c} is Gaussian distributed with mean

$$\begin{aligned} \overline{Z_{2c}} &= \mathrm{E}\left[Z_{2c}\right] \\ &= 0, \end{aligned} \tag{A.21}$$

and variance

$$\begin{aligned} \mathrm{Var}(Z_{2c}) &= \mathrm{E}\left[Z_{2c}^2\right] - \mathrm{E}^2\left[Z_{2c}\right], \\ &= \frac{N_0}{2}. \end{aligned} \tag{A.22}$$

The output of the sine correlator is derived similarly:

$$Z_{2s} = 0 + \sqrt{\frac{2}{T_s}} \int_0^{T_s} n(t) dt, \tag{A.23}$$

with similar mean

$$\begin{aligned} \overline{Z_{2s}} &= \mathrm{E}\left[Z_{2s}\right] \\ &= 0, \end{aligned} \tag{A.24}$$

and variance

$$\begin{aligned} \mathrm{Var}(Z_{2s}) &= \mathrm{E}\left[Z_{2s}^2\right] - \mathrm{E}^2\left[Z_{2s}\right], \\ &= \frac{N_0}{2}. \end{aligned} \tag{A.25}$$

For envelope detection, the input to the decision device is:

$$Z_2 = \sqrt{Z_{2c}^2 + Z_{2s}^2}. \tag{A.26}$$

Since Z_{2c} and Z_{2s} are Gaussian distributed random variables, with zero mean, Z_2 is a Rayleigh random variable. The distribution of Z_2 is:

$$\begin{aligned} f_{Z_2}(z) &= \frac{z}{\sigma^2}\,\mathrm{e}^{-\frac{z^2}{2\sigma^2}}, \\ &= \frac{2z}{N_0}\,\mathrm{e}^{-\frac{z^2}{N_0}}\,. \end{aligned} \tag{A.27}$$

The cumulative distribution function is:

$$\begin{aligned} F_{Z_2}(z) &= 1 - \mathrm{e}^{-\frac{z^2}{2\sigma^2}}, \\ &= 1 - \mathrm{e}^{-\frac{z^2}{N_0}}\,. \end{aligned} \tag{A.28}$$

A.2 Rayleigh Preliminaries

The special functions that appear in (A.19) and (A.16) cause some numerical precision problems during evaluation on a computer. However analyzing a system in an environment where the signal experiences AWGN and fading avoids the use of special functions. The expressions for f_{Z_1} and f_{Z_2} come from Kim [17]. The expression for the pdf of the output of the energy correlator with signal present is:

$$\begin{aligned} f_{Z_1}(z) &= \frac{1}{2\sigma_1^2}\,\mathrm{e}^{-\frac{z}{2\sigma_1^2}} \\ &= \frac{1}{E_s + N_0}\,\mathrm{e}^{-\frac{z}{E_s+N_0}}, \end{aligned} \tag{A.29}$$

and the pdf of the output of the correlators with no signal present is:

$$\begin{aligned} f_{Z_2}(z) &= \frac{1}{2\sigma_2^2}\,\mathrm{e}^{-\frac{z}{2\sigma_2^2}} \\ &= \frac{1}{N_0}\,\mathrm{e}^{-\frac{z}{N_0}}\,. \end{aligned} \tag{A.30}$$

Integrating to find F_{Z_1} and F_{Z_2} gives:

$$F_{Z_1}(z) = 1 - \mathrm{e}^{-\frac{z}{E_s+N_0}} \tag{A.31}$$

and

$$F_{Z_1}(z) = 1 - \mathrm{e}^{-\frac{z}{N_0}}, \tag{A.32}$$

which do not contain numerically sensitive special functions.

A.3 Putting it all together

A.3.1 The probability of correct detection

The decision device selects the correct energy correlator when the following occurs: $Z_1 \geq \gamma \max_{m \neq i} Z_m$, given signal number 1 is transmitted. The probability of this event is:

$$P_c\left(\gamma|1\right) = \mathrm{P}\left(Z_1 \geq \gamma \max_{m \neq i} Z_m | 1\right), \tag{A.33}$$

$$= \int_0^\infty \underbrace{F_{Z_2}\left(\frac{z}{\gamma}|1\right)^{M-1}}_{\substack{\text{all others are} \\ \text{smaller than} \\ z}} \underbrace{f_{Z_1}\left(z|1\right)}_{\substack{\text{prob} \\ Z_1 = z}} dz. \tag{A.34}$$

$$\tag{A.35}$$

A.3.2 The probability of incorrect detection

The decision device makes an incorrect decision when there exists one energy correlator output that is larger than γ times the remaining, and this output is not as-

sociated with the transmitted signal. Thus the probability of this event is determined as follows:

$$P_e(\gamma|1) = \mathrm{P}\left[\bigcup_{m=2}^{M} Z_m \geq \gamma \max_{j \neq m} Z_j | 1\right] \tag{A.36}$$

$$= (M-1)\int_0^\infty \underbrace{F_{Z_1}\left(\frac{z}{\gamma}|1\right)}_{\substack{Z_1 \text{ is smaller} \\ \text{than } z}} \underbrace{F_{Z_2}\left(\frac{z}{\gamma}|1\right)^{M-2}}_{\substack{\text{all other} \\ \text{incorrect} \\ \text{outputs are} \\ \text{smaller than} \\ z}} \underbrace{f_{Z_2}(z|1)}_{\substack{\text{prob} \\ Z_2 = z}} dz. \tag{A.37}$$

A.3.3 The probability of RTT failure

The last remaining event of interest occurs when the outputs of the energy detectors fail the RTT. In this situation, no output is larger than γ times any of its peers. This represents a situation where there is some strong interference or the low probability chance that the noise in one detector is large at the sample instant. We shall determine the probability that one, or more, outputs has a large noise component at the sample time.

Typically, the probability of RTT failure can be expressed as $P_f(\gamma) = 1 - P_c(\gamma) - P_e(\gamma)$, but on a computer this calculation can cause numerical problems. In this case it is desirable to determine the integral (or direct) expression for $P_f(\gamma)$. If Z_1 is the maximum output, the event that will fail the RTT is the event that one or more outputs satisfy $\frac{Z_1}{\gamma} < Z_j < Z_1$. If v outputs satisfy this condition then:

$$P_f(\gamma,\, Z_1 \text{ is max}, |1) = \int_0^\infty \sum_{v=1}^{M-1} \binom{M-1}{v} \underbrace{\left[F_{Z_2}(z|1) - F_{Z_2}(\frac{z}{\gamma}|1)\right]^v}_{\substack{v\ Z_2\text{'s outputs are between} \\ Z_1 \text{ and } \frac{Z_1}{\gamma}}} \times$$

$$\underbrace{\left[F_{Z_2}(\frac{z}{\gamma}|1)\right]^{M-1-v}}_{\substack{M-1-v\ Z_2\text{'s are} \\ \text{less than } \frac{Z_1}{\gamma}}} \times [f_{Z_1}(z|1)]\, dz. \tag{A.38}$$

The probability of Z_1 being the largest output is:

$$P(Z_1 \text{ is max}) = P_c(1). \tag{A.39}$$

If Z_1 is not the maximum output then one of the Z_2 must be. The probability of this event occurring is:

$$\mathrm{P}(Z_2 \text{ is max}) = \mathrm{P}(\max_k Z_k = Z_2) \tag{A.40}$$

$$= \int_0^\infty F_{Z_1}(z|1) F_{Z_2}(z|1)^{M-2} f_{Z_2}(z|1) dz. \tag{A.41}$$

Given Z_2 is the maximum output the failure event is the event that $\frac{Z_2}{\gamma} < Z_j < Z_2$ for some j. We split this into two separate joint probabilities:

$$P_f(\gamma,\, Z_2 \text{ is max}, |1) =$$

$$P_f(\gamma, Z_1 \text{ is below the range}, Z_2 \text{ is max}, |1) +$$

$$P_f(\gamma, Z_1 \text{ is in the range}, Z_2 \text{ is max}, |1), \tag{A.42}$$

$$P_f(\gamma,\, Z_2 \text{ is max},\, |1) = \int_0^\infty \sum_{v=1}^{M-2} \binom{M-2}{v} \underbrace{\left[F_{Z_2}(z|1) - F_{Z_2}(\frac{z}{\gamma}|1)\right]^v}_{\substack{v\ Z_2\text{'s are in the} \\ \text{failure range}}} \underbrace{\left[F_{Z_2}(\frac{z}{\gamma}|1)\right]^{M-2-v}}_{\substack{M-2-v \\ Z_2\text{'s are} \\ \text{below the} \\ \text{failure range}}}$$

$$\underbrace{\left[F_{Z_1}(\frac{z}{\gamma}|1)\right]}_{\substack{Z_1 \text{ below} \\ \text{the failure} \\ \text{range}}} f_{Z_2}(z|1)dz \tag{A.43}$$

$$+ \int_0^\infty \sum_{v=0}^{M-2} \binom{M-2}{v} \underbrace{\left[F_{Z_2}(z|1) - F_{Z_2}(\frac{z}{\gamma}|1)\right]^v}_{\substack{v\ Z_2\text{'s in the failure} \\ \text{range}}} \underbrace{\left[F_{Z_2}(\frac{z}{\gamma}|1)\right]^{M-2-v}}_{\substack{M-2-v \\ Z_2\text{'s are} \\ \text{under the} \\ \text{failure range}}}$$

$$\underbrace{\left[F_{Z_1}(z|1) - F_{Z_1}(\frac{z}{\gamma}|1)\right]}_{\substack{Z_1 \text{ is in the failure} \\ \text{range}}} f_{Z_2}(z|1)dz. \tag{A.44}$$

Combining (A.38), (A.39), (A.41), and (A.44):

$$\begin{aligned} P_f(\gamma|1) &= P_f(\gamma,\, Z_1 \text{ is max}|1) + \sum_{i=2}^{M} P_f(\gamma,\, Z_i \text{ is max}|1), \\ &= P_f(\gamma,\, Z_1 \text{ is max}|1) + (M-1)P_f(\gamma,\, Z_2 \text{ is max}|1). \end{aligned} \tag{A.45}$$

APPENDIX B

The method of moments

Given an equation of the form:

$$\int_0^1 K_f(A,t)u(t)\,dt = g(A), \tag{B.1}$$

where $u(t)$ is known and $g(A)$ is known at discrete values $g(A_m)$, determine the discrete function $K_f(A,t)$. Equation (B.1) is known as an integral equation of the first kind (IFK), in particular it is known as a Fredholm IFK[1]. $g(A)$ is generally referred to as the excitation, $u(t)$ the unknown, and $K_f(A,t)$ as the transform kernel. Comparing to (3.56), (3.59), (3.64), and (3.67) we note $u(t)$ is actually a known quantity and $K_f(A,t)$ is unknown. Thus, the problem is cast as a IFK with unknown kernel. The solution this problem is generally not known, but in the case where $K_f(A,t)$ is of the form $K_f(At)$ we may use the change of variable $x = At$ to transform (B.1) as follows:

$$\begin{aligned} g(A) &= \int_0^1 K_f(A,t)u(t)\,dt, \\ &= \int_0^1 K_f(At)u(t)\,dt, \\ &= \frac{1}{A}\int_0^A K_f(t)u(\frac{t}{A})\,dt, \end{aligned} \tag{B.2}$$

[1]The interested reader is referred to [55–60] for thorough discussion of closed form and numerical solutions to integral equations.

which can be re-written as:

$$g(A) = \int_0^A f(t) K_v(\frac{t}{A})\, dt, \tag{B.3}$$

where $\frac{1}{A}$ is subsumed into K_v. Equation (B.3) is known as a Volterra IFK, and there is a body of work that deals with its solution.

To solve, we select the method of moments [61] (otherwise known as Galerkin's method). This method involves projecting both sides of (B.3) onto a common orthonormal basis and projecting the unknown function, $f(t)$, onto a (potentially) different orthonormal basis. We select the common RHS and LHS basis functions to be Dirac delta functions to enforce equality at a finite set of points A_m; this method is known as point matching or co-location.

Starting with the Volterra IFK:

$$g(A) = \int_0^A f(t) K_v(A,t) dt, \tag{B.4}$$

$$= \int_0^A \left[\sum_n \alpha_n B_n(t) \right] K_v(A,t) dt, \tag{B.5}$$

where the set $\{B_n(t)\}$ is an orthonormal basis of domain t for $f(t)$. If $\{T_m(A)\}$ is an orthonormal basis for $g(A)$ then

$$\sum_m \beta_m T_m(A) = \int_0^A \left[\sum_n \alpha_n B_n(t) \right] K_v(A,t) dt, \tag{B.6}$$

$$= \sum_n \alpha_n \int_0^A B_n(t) K_v(A,t) dt. \tag{B.7}$$

If we now project the integral onto the basis functions $T_m(A)$ we can re-write (B.7) as:

$$\sum_m \beta_m T_m(A) = \sum_n \alpha_n \sum_m l_{m,n} T_m(A). \tag{B.8}$$

We note that

$$\beta_m = \sum_n \alpha_n l_{m,n}, \tag{B.9}$$

which in matrix form gives:

$$\alpha = l^{-1}\beta. \tag{B.10}$$

While generally the orthonormal bases are complete, truncation of the basis set to a subspace is equivalent to approximating α. Determining β and l is the remaining challenge.

To determine the basis functions for $g(A)$ we recall that the measured or simulated representation is evaluated at discrete points $g(A_m)$. Using the point matching method to enforce equality in (B.3) at $\{A_m\}$ we select the orthonormal basis functions to be:

$$T_m(A) = \delta(A - mT), \tag{B.11}$$

with $T = A_2 - A_1$ and $\{A_m\}$ uniformly spaced. For the orthonormal basis $\{B_n\}$ we choose the pulse functions:

$$B_n(t) = \begin{cases} \frac{1}{\sqrt{\tau}} & n\tau < t < (n+1)\tau \\ 0 & \text{otherwise} \end{cases}, \tag{B.12}$$

where $\tau = \max(A_m)/N$ and N is the size of the subspace $\{B_1(t) \ldots B_N(t)\}$. Since t is bounded from above by A or (A_m in the discrete case) the selection of subspace is made under the restriction $(n+1)\tau \leq A_m$.

To determine β_m we project $g(A)$ onto its basis, $\{T_m(A)\}$,:

$$\begin{aligned} \beta_m &= \int T_m(A) g(A) dA, \\ &= \int \delta(A - mT) g(A) dA, \\ &= g(A - mT), \\ &= g(A_m). \end{aligned} \tag{B.13}$$

Determination of $l_{m,n}$ proceeds in a similar fashion:

$$\begin{aligned} l_{m,n} &= \int \int_0^A B_n(t) K_v(A,t)\, dt\, T_m(A)\, dA, \\ &= \int \frac{1}{\sqrt{\tau}} \int_{n\tau}^{(n+1)\tau} K_v(A,t)\, dt\, T_m(A)\, dA, \\ &= \frac{1}{\sqrt{\tau}} \int_{n\tau}^{(n+1)\tau} K_v(A - mT, t)\, dt, \\ &= \frac{1}{\sqrt{\tau}} \int_{n\tau}^{(n+1)\tau} K_v(A_m, t)\, dt. \end{aligned} \tag{B.14}$$

Since $f(t)$ is unknown, solving for α_n is done in the manner of (B.10). Once α_n is known we note that:

$$f(t) \cong \sum_{n=0}^{N-1} \alpha_n B_n(t), \tag{B.15}$$

and discretely:

$$f_n(t) = f(t - n\tau) \cong \frac{1}{\sqrt{\tau}} \alpha_n. \tag{B.16}$$

Note that care must be taken for the matrix inversion in (B.10) that l is well conditioned. If l is ill-conditioned some regularization techniques may need to be employed.

APPENDIX C

Generation and verification of nonlinear system output

Once a nonlinear circuit has been characterized by a polynomial fit to the MIQM I and Q curves, determination of the intermodulation products becomes the focus of the analysis. The mechanics of determining the IM magnitude, phase and frequency given the input signal and the polynomial fits will be detailed in this appendix. The result of this appendix will be twofold. First, a list of all IM products from which fundamental and interfering IM products can be culled. Second, the total fundamental and interfering signal power will be determined. The output will be verified against known good one and two tone gain compression plots from the CAD tool HARMONICA.

C.1 Inputs

We define the input signal as:

$$x(t) = A_1 \cos(2\pi f_1 t + \phi_1) + \sum_{i=2}^{M} A_i \cos(2\pi f_i t + \phi_i), \tag{C.1}$$

where the $A_1 \cos(2\pi f_1 t + \phi_1)$ tone is considered the desired tone and the remainder of the $M-1$ tones are interfering tones. If the amplifier that is under consideration is

memoryless or quasi-memoryless in a fashion such that the two tone frequency sweep is fairly flat at the frequencies f_i, then the MIQM is able to accurately model the amplifier with the following two polynomials:

$$p_I(x) = \sum_{i=0}^{N} c_{I_i} x^i, \tag{C.2}$$

and

$$p_Q(x) = \sum_{i=0}^{N} c_{Q_i} x^i. \tag{C.3}$$

Once these polynomials and the input signals are known the intermodulation products can be determined.

C.2 IM product determination

C.2.1 Method of Wass

In [19], Wass developed a table method to determine the coefficients and frequencies of N'th order M tone intermodulation products. The results of Wass' method can be represented by the following "typical product":

$$P_{i,j}(t) = c_i q_{i,j} \left[\prod_{k=1}^{M} A_k^{p_{i,j,k}} \right] \cos\left(\sum_{k=1}^{M} n_{j,k} (f_k(t) + \phi_k) \right), \tag{C.4}$$

where the j index the order i intermodulation products, the c_i term is the i'th order polynomial coefficient, the $q_{i,j}$ is the leading fraction of the j'th order i'th product, the $p_{i,j,k}$ are the number of times input tone k appears in i'th order product j, and the $n_{i,j,k}$ are the frequency contribution of the k'th input tone in the j'th order i product. As an example, the $M = 2$ $N = 1 \ldots 3$ order products are enumerated in Table C.1. The main advantage of this tabular method is that large tables can be calculated off-line and stored. The individual polynomial coefficients and input frequencies can be combined with the stored products once they are determined.

Table C.1: Two tone, order 1...3 intermodulation products

i	j	c_i	$q_{i,j}$	$p_{i,j,}$	$n_{j,}$	product
1	1	c_1	1	1,0	1,0	$c_1 A_1 \cos(2\pi(f_1+\phi_1)t)$
1	2	c_1	1	0,1	0,1	$c_1 A_2 \cos(2\pi((f_2+\phi_2)$
2	1	c_2	$\frac{1}{2}$	0,2	0,0	$c_2 \frac{1}{2} A_2$
2	2	c_2	$\frac{1}{2}$	2,0	0,0	$c_2 \frac{1}{2} A_2^2$
2	3	c_2	1	1,1	1,-1	$c_2 A_1 A_2 \cos(2\pi((f_1+\phi_1)-(f_2+\phi_2)))$
2	4	c_2	$\frac{1}{2}$	2,0	2,0	$c_2 \frac{1}{2} A_1^2 \cos(2\pi 2(f_1+\phi_1))$
2	5	c_2	1	1,1	1,1	$c_2 A_1 A_2 \cos(2\pi((f_1+\phi_1)+(f_2+\phi_2)))$
2	6	c_2	$\frac{1}{2}$	0,2	0,2	$c_2 \frac{1}{2} A_2^2 \cos(2\pi 2(f_2+\phi_2))$
3	1	c_3	$\frac{3}{4}$	2,1	2,-1	$c_3 \frac{3}{4} A_1^2 A_2 \cos(2\pi(2(f_1+\phi_1)-(f_2+\phi_2)))$
3	2	c_3	$\frac{3}{4}$	3,0	1,0	$c_3 \frac{3}{4} A_1^3 \cos(2\pi(f_1+\phi_1))$
3	3	c_3	$\frac{3}{2}$	1,2	1,0	$c_3 \frac{3}{2} A_1 A_2^2 \cos(2\pi(f_1+\phi_1))$
3	4	c_3	$\frac{3}{4}$	0,3	0,1	$c_3 \frac{3}{4} A_2^3 \cos(2\pi(f_2+\phi_2))$
3	5	c_3	$\frac{3}{2}$	2,1	0,1	$c_3 \frac{3}{2} A_1^2 A_2 \cos(2\pi(f_2+\phi_2))$
3	6	c_3	$\frac{3}{4}$	1,2	1,-2	$c_3 \frac{3}{4} A_1 A_2^2 \cos(2\pi((f_1+\phi_1)-2(f_2+\phi_2)))$
3	7	c_3	$\frac{1}{4}$	3,0	3,0	$c_3 \frac{1}{4} A_1^3 \cos(2\pi 3(f_1+\phi_1))$
3	8	c_3	$\frac{3}{4}$	2,1	2,1	$c_3 \frac{3}{4} A_1^2 A_2 \cos(2\pi(2(f_1+\phi_1)+(f_2+\phi_2)))$
3	9	c_3	$\frac{3}{4}$	1,2	1,2	$c_3 \frac{3}{4} A_1 A_2^2 \cos(2\pi((f_1+\phi_1)+2(f_2+\phi_2)))$
3	10	c_3	$\frac{1}{4}$	0,3	0,3	$c_3 \frac{1}{4} A_2^3 \cos(2\pi 3(f_2+\phi_2))$

C.2.2 Method of Sea

In [20, 21] Sea proposed two methods to directly calculate the intermodulation magnitude of a given frequency combination, at run time. The method in [20] has a high degree of computational complexity which is not feasible for run-time operation unless M and N are small. However, in [21] a recursion is proposed and the complexity is reduced to a roughly $O(MN)$ for each frequency desired. Unfortunately, the numbers of frequencies present grows exponentially in N and makes his calculation difficult for run time use in determining all magnitudes, phases and frequencies of large order systems.

C.3 Combining Wass with MIQM

Once the table of IM products is generated the task at hand becomes deciding which products are interference and which products are signal. In the case of input interfering signals with no modulation all products with table element $n_j = 1, 0, \ldots, 0$ are deemed fundamental power. These products are found in odd orders due to the mixing of even order DC components with the fundamental. For example, the fundamental tones found in Table C.1 are $(i, j) = (1, 1), (3, 2), (3, 3)$ and the remainder of the tones are considered to be interferers. The output tones may be restricted to a bandwidth of interest around the fundamental tone, so all interfering tones may not be present in any given listing of tones.

Recalling the MIQM model block diagram found in Figure 3.9 and repeated here in Figure C.1, the fundamental and interfering tones can be determined by matching the I and Q Chebychev polynomials to the Wass table. The signal that results on

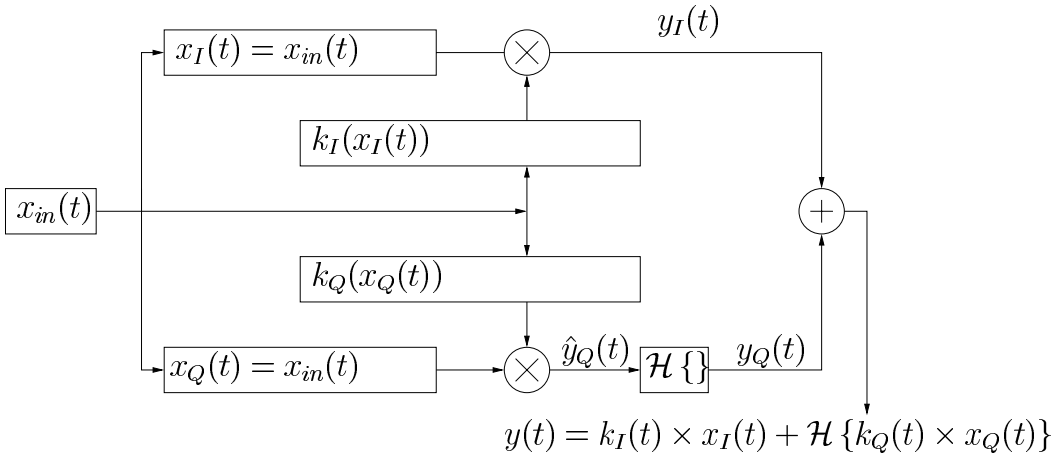


Figure C.1: Modified Instantaneous Quadrature Model Block Diagram.

the I side is

$$y_I(t) = \sum_{i,j} P^I_{i,j}(t), \tag{C.5}$$

with the c_i replaced by c_{I_i}, and the Q side

$$\hat{y}_Q(t) = \sum_{i,j} P^Q_{i,j}(t), \tag{C.6}$$

with the c_i replaced by c_{Q_i}.

To determine the magnitude of the signal at each frequency, all IM products of like frequency and phase from $y_I(t)$ are added coherently, as are all the products from $\hat{y}_Q(t)$. The resulting signals are written as:

$$y_{I,m}(t) = \underbrace{\sum_i \sum_j}_{n_{j,k}=n_{j,l}\ \forall i,k,l} P^I_{i,j}(t), \tag{C.7}$$

and

$$y_{Q,m}(t) = \underbrace{\sum_i \sum_j}_{n_{j,k}=n_{j,l}\ \forall i,k,l} P^Q_{i,j}(t), \tag{C.8}$$

where m indexes the resulting tones (by convention let $m = 1$ for the fundamental output).

Next, the tones on the I side and Q side are compared to each other. Tones sharing the same frequency and phase will be technically 90 degrees out of phase when $\hat{y}_Q(t)$ is transformed to $y_Q(t)$ via the Hilbert transformer. The magnitude of the resulting tone will be:

$$|y_m(t)| = \sqrt{|y_{I,m}(t)|^2 + |y_{Q,m}(t)|^2}. \tag{C.9}$$

The phase of the m'th tone is:

$$\phi(y_m(t)) = \tan^{-1} \frac{|y_{I,m}(t)|}{|y_{Q,m}(t)|}. \tag{C.10}$$

Thus, the resulting output is

$$y(t) = \sum_m |y_m(t)| \cos\left(2\pi f_m + \phi(y_m(t))\right). \tag{C.11}$$

C.3.1 SIR of the output signal

Once the IM products are separated into fundamental tones and interfering tones and the magnitudes of like phase and frequency terms are determined for $y(t)$, the SIR calculation is rather trivial. The total normalized power due to the fundamental signal is

$$P_S = |y_1(t)|^2/2, \tag{C.12}$$

the total power due to the interfering signals is

$$P_N = \sum_{m,m\neq 1} |y_m(t)|^2/2, \tag{C.13}$$

and the resulting SIR is

$$SIR = 10\log_{10}\frac{P_S}{P_N}. \tag{C.14}$$

C.4 Method Verification

In this section we will compare the one and two tone compression test results of the amplifier depicted in Figure 4.9 using HARMONICA and the MIQM/Wass method. The one tone compression test is a good indicator that our model is correctly identifying all the fundamental power in the output signal. The two tone test is a good indicator that out model is correctly identifying all the fundamental signal power due to the mixing of the even orders of interferers and the odd fundamental orders. When these two tests match, we will have reasonable assurance that the algorithms used to determine the IM products are correct.

The one tone compression test measures the gain of the fundamental as the input fundamental is swept in magnitude (or power). There are no interfering tones in this test and Figure C.2 shows the gain compression of our model matches the simulated gain compression.

The two tone gain compression test measures the gain of a weak input fundamental as the strength of a nearby interfering tone grows large. As Figure C.3 shows our model tracks the simulated output very closely. The divergence in the -7dBm range most likely has to do with the inversion of the matrix in the calculation of the MIQM transfer functions starting to lose precision.

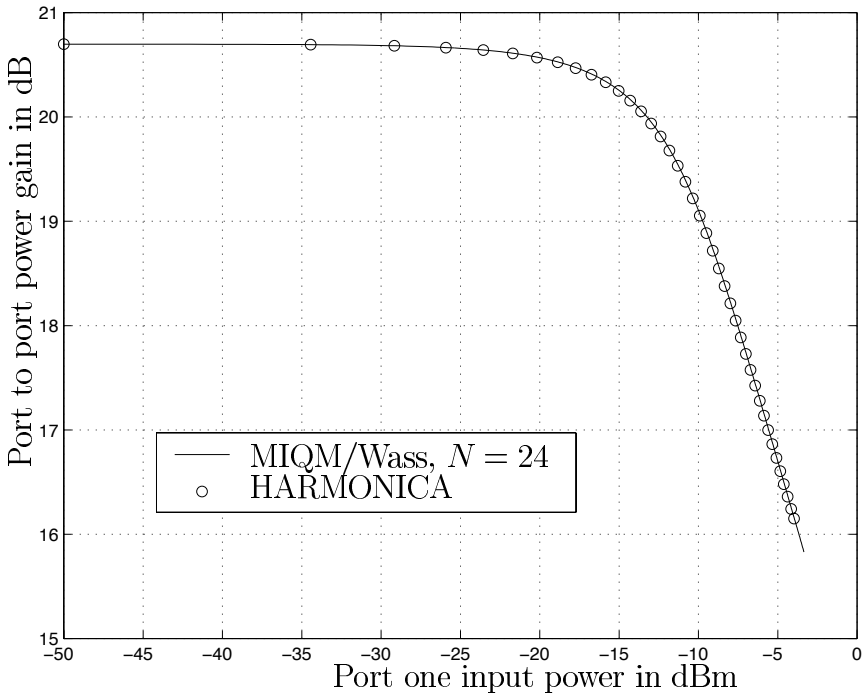


Figure C.2: One tone gain compression comparison of MIQM/Wass to HARMONICA.

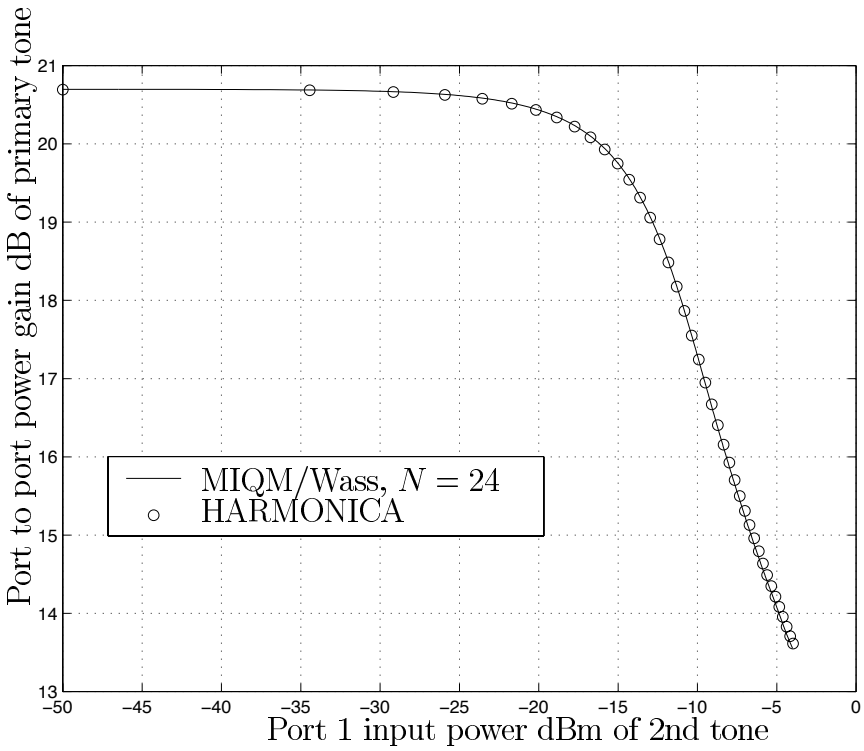


Figure C.3: Two tone gain compression comparison of MIQM/Wass to HARMONICA.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] A.-G. A. Daraiseh, "Advances in the performance of Reed-Solomon codes," *IEEE Transactions on Communications*, vol. 47, no. 1, , January 1999.

[2] W. E. Stark, "Performance of concatenated codes on channels with jamming," *IEEE International Conference on Communication*, June 1982.

[3] T. C. Nolan and W. E. Stark, "Analysis of product codes for frequency-hopped spread-spectrum with partial band interference," *Proceedings of the IEEE Military Communications Conference*, 1997.

[4] A. Michelson and A. Levesque, *Error Control Techniques for Digital Communication*, John Wiley and Sons, Inc., New York, 1985.

[5] S. B. Wicker, "Reed-Solomon error control coding for Rayleigh fading channels with feedback," *IEEE Transactions on Vehicular Technology*, vol. 41, no. 2, , May 1992.

[6] R. J. McEliece and L. Swanson, "On the decoder error probability for Reed-Solomon codewords," *IEEE Transactions on Information Theory*, vol. 32, no. 5, pp. 701–703, September 1986.

[7] K.-M. Cheung and R. J. McEliece, "The undetected error probability for Reed-Solomon codes," *IEEE Military Communications Conference*, vol. 3, pp. 163–167, October 1988.

[8] A.-G. A. Daraiseh and C. W. Baum, "Packet combining in frequency-hop spread-spectrum communication systems," *IEEE Transactions on Communications*, vol. 46, no. 1, pp. 23–33, January 1998.

[9] P. S. Guinand and J. H. Lodge, "Design of generalized product codes suitable for iterative decoding," *IEEE International Symposium on Information Theory and its Applications*, September 1996.

[10] T. Nakamura and H. Yoshida, "Error rate calculation of iterated decoding for product codes,".

[11] C. Yi and J. H. Lee, "Interleaving and decoding scheme for a product code for a mobile data communication," *IEEE Transactions on Communications*, vol. 45, no. 2, pp. 144–147, February 1997.

[12] C. D. Frank and M. B. Pursley, "Concatenated coding for frequency-hop spread-spectrum with partial-band jamming," *IEEE Transactions on Communications*, vol. 44, no. 3, pp. 377–387, March 1996.

[13] F. J. Macwilliams and N. J. A. Sloane, *The Theory of Error Correcting Codes*, North-Holland Pub. Co., New York, 1977.

[14] R. E. Blahut, *Theory and Practice of Error Control Codes*, Addison-Wesley, Reading, Mass., 1983.

[15] T. C. Nolan and W. E. Stark, "A recursive method for calculating error probabilities for a Reed-Solomon codeword," *Proceedings of the IEEE Vehicular Technology Conference*, 1998.

[16] T. C. Nolan and W. E. Stark, "A recursive method for calculating error probabilities for a Reed-Solomon codeword with bounded distance errors and erasures decoding," *Proceedings of the IEEE Military Communications Conference*, 1998.

[17] S. W. Kim and W. E. Stark, "Performance limits of Reed-Solomon coded CDMA with orthogonal signaling in a Rayleigh-fading channel," *IEEE Transactions on Communications*, vol. 46, no. 9, pp. 1125–1134, September 1998.

[18] S. A. Maas, *Nonlinear Microwave Circuits*, New York: IEEE Press, 1997.

[19] C. A. A. Wass, "A table of intermodulation products," *Journal of the IEE*, vol. 95, no. 3, pp. 31, 1948.

[20] R. G. Sea, "An algebraic formula for amplitudes of intermodulation products involving an arbitrary number of frequencies," *Proceedings of the IEEE*, vol. 56, pp. 1388–1389, August 1968.

[21] R. G. Sea and A. G. Vacroux, "On the computations of intermodulation products for a power series nonlinearity," *Proceedings of the IEEE*, vol. 57, pp. 337–338, March 1969.

[22] B. Razavi, *RF Microelectronics*, New Jersey: Prentice-Hall, 1998.

[23] D. Newkirk, "Simulating circuits and systems with serenade SV," *QST*, pp. 37–43, January 2001.

[24] J. Staudinger, "The importance of sub-harmonic frequency terminations in modeling spectral re-growth from CW AM/AM and AM/PM derived non-linearities," *Wireless Communications Conference*, 1997.

[25] J. Staudinger, "Applying the quadrature modeling technique to wireless power amplifiers," *Microwave Journal*, November 1997.

[26] A. J. Cann, "Nonlinearity model with variable knee sharpness," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-16, no. 6, pp. 874, November 1980.

[27] S. L. Loyka and I. D. Cheremisinov, "Validation of the high-order polynomial models used in behavioral-level simulation," *TELSKIS*, 1999.

[28] S. L. Loyka, "On the use of Cann's model for nonlinear behavioral-level simulation," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 5, , September 2000.

[29] S. A. Mass, "How to model intermodulation distortion," *Microwave Symposium Digest*, 1999.

[30] S. L. Loyka, "Applying genetic algorithm to modeling nonlinear transfer functions," *TELSKIS*, 1999.

[31] G. L. Heiter, "Characterization of nonlinearities in microwave devices and systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-21, no. 12, , December 1973.

[32] R. Kielkowski, *SPICE Practical Device Modeling*, New York: McGraw-Hill, 1995.

[33] L. Pillage, R. Rohrer, and C. Visweseariah, *Electronic Circuit and System Simulation Methods*, New York: McGraw-Hill, 1995.

[34] J. A. Connelly and P. Choi, *Macromodeling with SPICE*, New Jersey: Prentice Hall, 1992.

[35] E. V. D. Eijnde and J. Schoukens, "Steady-state analysis of a periodically excited nonlinear system," *IEEE Transactions of Circuits and Systems*, vol. 37, no. 2, , February 1990.

[36] D. A. Teeter, J. R. East, R. K. Mains, and G. I. Haddad, "Large-signal numerical and analytical HBT models," *IEEE Transactions on Electron Devices*, vol. 40, no. 5, , May 1993.

[37] N. B. deCarvalho and J. C. Pedro, "Multitone frequency-domain simulation of nonlinear circuits in large- and small-signal regimes," *IEEE Transactions on Microwave Theory and Techniques*, vol. 46, no. 12, , December 1998.

[38] S. A. Maas and D. Neilson, "Modeling MESFET's for intermodulation analysis of mixers and amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 38, no. 12, , December 1990.

[39] K. E. Atkinson, *An Introduction to Numerical Methods*, New York: John Wiley and Sons, 1989.

[40] S. L. Loyka and J. R. Mosig, "New behavioral-level simulation technique for RF/microwave applications. part i: Basic concepts," *International journal of RF and microwave computer-aided engineering*, vol. 10, no. 4, , June 2000.

[41] N. Wiener, "Response of a nonlinear device to noise," *MIT Radiation Lab Report no. 129*, April 1942.

[42] S. Narayanan, "Transistor distortion analysis using Volterra series representation," *The Bell System Technical Journal*, pp. 991–1024, May 1967.

[43] J. S. Kenney and A. Leke, "Power amplifier spectral re-growth for digital cellular and PCS applications," *Microwave Journal*, October 1995.

[44] A. R. Kaye, D. A. George, and M. J. Eric, "Analysis and compensation of bandpass nonlinearities for communication," *IEEE Transactions on Communications*, vol. 20, no. 11, , October 1972.

[45] S. L. Loyka and J. R. Mosig, "New behavioral-level simulation technique for RF/microwave applications. part ii: Approximation of nonlinear transfer functions," *International journal of RF and microwave computer-aided engineering*, vol. 10, no. 4, , June 2000.

[46] W. Bosch and G. Gatti, "Measurement and simulation of memory effects in predistortion linearizers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 37, no. 12, pp. 1885–1890, December 1989.

[47] C. T.-C. Nguyen, "High Q micromechanical oscillators and filters for communications," *Proceedings of the 1997 IEEE International Symposium on Circuits and Systems*, pp. 2825–2828, June 1997.

[48] D. H. Shen, C. M. Hwang, B. B. Lusignan, and B. A. Wooley, "A 900 MHz RF front-end with integrated discrete-time filtering," *IEEE Journal of Solid State Circuits*, vol. 31, no. 12, , December 1996.

[49] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing Principles, Algorithms, and Applications*, New York: Macmillan Publishing Company, 1992.

[50] S. Lay, *Analysis with Introduction to Proof*, New Jersey: Prentice-Hall, 1990.

[51] O. Carlsson, "Predicting CNR frequency hopping interference by simulation,", 1986.

[52] G. Williman and T. Rivingston, "A performance analysis of cosite filters in a collocated SINCGARS environment," *Proceedings of the Tactical Communications Conference*, vol. 1, pp. 594–604, 1990.

[53] S.-W. Chen, "Linearity requirements for digital wireless communications," *Gallium Arsenide Integrated Circuit (GaAs IC) Symposium*, pp. 29–32, 1997.

[54] M. K. Simon, "A new twist on the Marcum Q-function and its application," *IEEE Communications Letters*, vol. 2, no. 2, pp. 39–41, February 1998.

[55] G. M. Wing, *A Primer on Integral Equations of the First Kind*, Port City Press: Baltimore, 1991.

[56] R. P. Kanwal, *Linear Integral Equations Theory and Technique*, Birkhauser: Boston, 1997.

[57] J. Kondo, *Integral Equations*, Clarendon Press: Oxford, 1991.

[58] M. A. Goldberg and C. S. Chen, *Discrete Projections Methods for Integral Equations*, Computations Mechanics Publications: Boston, 1997.

[59] K. E. Atkinson, *The Numerical Solution of Integral Equations of the Second Kind*, Cambridge University Press: New York, 1997.

[60] A. D. Polyanin and A. V. Manzhirov, *Handbook of Integral Equations*, CRC Press: New York, 1998.

[61] A. F. Peterson, S. L. Ray, and R. Mittra, *Computational Methods for Electromagnetics*, IEEE Press: New York, 1998.

[62] T. C. Nolan, W. E. Stark, , and C. T.-C. Nguyen, "Direct down conversion of passband signals using MEMS filters and sub-sampling," *Proceedings of the IEEE Vehicular Technology Conference*, 1999.

[63] J. Lim, D. L. Neuhoff, and T. C. Nolan, "Allocating complexity between source and channel coding," *Workshop on Data Compression Processing Techniques for Missile Guidance Data Links*, vol. US Missile Command MG-99-5, pp. 669–679, 1999.

[64] T. C. Nolan and W. E. Stark, "Mitigation of cosite interference in nonlinear receivers with MEMS filters," *Proceedings of the IEEE Military Communications Conference*, 2000.

[65] M. N. Horenstein, *Microeletronic Circuits and Devices*, New Jersey: Prentice-Hall, 1990.

[66] A. V. Oppenheim and A. S. Willsky, *Signals and Systems*, New Jersey: Prentice-Hall, 1983.

[67] J. R. McChesney, "System applications for SINCGARS VHF tactical radio," *Proceedings of the Workshop on Research Trends in Spread Spectrum Systems*, August 1985.

[68] S. B. Wicker, *Reed Solomon Codes and their Applications*, New York: IEEE Press, 1994.

[69] D. G. Finck and H. W. Beaty, *Standard Handbook for Electrical Engineers, 13th Edition*, McGraw-Hill, Inc., 1993.

[70] S. Wilson, *Digital Modulation and Coding*, Prentice-Hall, Inc., New Jersey, 1996.

[71] A.-G. A. Daraiseh and C. W. Baum, "Decoder error and failure probabilities for Reed-Solomon codes: Decodable vectors method," *IEEE Transactions on Communications*, vol. 46, no. 7, , July 1998.

[72] C. P. Liang, J. H. Jong, W. E. Stark, and J. R. East, "Nonlinear amplifier effects in communications systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, no. 8, , August 1999.

[73] A. N. Karanicolas, "A 2.7-V 900-MHz CMOS LNA and mixer," *IEEE Journal of Solid State Circuits*, vol. 31, no. 12, , December 1996.

[74] C. T.-C. Nguyen, "Micromechanical resonators for oscillators and filters," *Proceedings of the 1995 IEEE International Ultrasonics Symposium*, pp. 489–499, November 1995.

[75] J. Jong, K. Yang, G. I. Haddad, and W. E. Stark, "Power optimization criteria for wireless communication systems with nonlinear amplifiers," *Submitted to IEEE Transactions on Vehicular Technology*.

[76] D. B. Leeson, "A simple model of feedback oscillator noise spectrum," *IEE Proceedings*, pp. 329–330, February 1966.

[77] J. Rutman, "Characterization of phase and frequency instabilities in precision frequency sources: Fifteen years of progress," *IEE Proceedings*, vol. 66, no. 9, pp. 1048–1075, September 1978.

[78] B. Razavi, "A study of phase noise in CMOS oscillators," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 3, pp. 331–343, March 1996.

[79] R. L. Beurle, "A comparison of the noise and random frequency and amplitude fluctuations in different types of oscillator," *IEE Proceedings*, vol. 103b, pp. 182–189, March 1956.

[80] W. P. Robins, *Phase Noise in Signal Sources*, England: Short Run Press Ltd., 1998.

[81] C. N. Wong, W. H. Chan, K. F. Tsang, and L. P. Wong, "Time domain modeling of nonlinear BJT amplifiers using frequency domain analysis," *Proceedings of the 1999 IEEE Canadian Conference on Electrical and Computer Engineering*, May 1999.

[82] K. F. Tsang, C. W. Tang, P. C. L. Yip, G. B. Morgan, and W. S. Chan, "Large signal transistor device measurement," *Proceedings of the 1991 International Conference on Circuits and Systems*, June 1991.

[83] D. Barataud, A. Mallet, M. Campovecchio, J. M. Nebus, J. P. Villotte, and J. Verspecht, "Measurement of time domain voltage/current waveforms at RF and microwave frequencies for the characterization of nonlinear devices," *Proceedings of the IEEE Instrumentation and Measurement Technology Conference*, May 1998.

[84] M. Lazaro, I. Santamaria, and C. Pantaleon, "A smooth and derivable large-signal model for microwave HEMT transistors," *IEEE International Symposium on Circuits and Systems*, May 2000.

[85] G. W. Rhyne, M. B. Steer, and B. D. Bates, "Frequency-domain nonlinear circuit analysis using generalized power series," *IEEE Transactions on Microwave Theory and Techniques*, vol. 36, no. 2, , February 1988.

[86] J. Ibanez-Diaz, C. Pantaleon, I. Santamaria, T. Fernandez, and D. Martinez, "Nonlinearity estimation in power amplifiers based on undersampled temporal data," *Proceedings of the 17th IEEE Instrumentation and Measurement Technology Conference*, vol. 3, pp. 1266–71, May 2000.

[87] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems: Modeling, Methodology, and Techniques*, New York: Kluwer Academic/Plenum, 2000.