

Low-SNR Operation of FSK Demodulators

MSc Thesis

Armin Šabanović



Delft University of Technology

Low-SNR Operation of FSK Demodulators

by

A. Šabanović

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in

ELECTRICAL ENGINEERING

at the faculty of EEMCS, Delft University of Technology,

to be defended publicly on the 28th of August 2019.

Student number: 4372441

Programme: Microelectronics

Specialization: Analog and RF/Wireless Design

Project duration: September 1, 2018 – August 28, 2019

Thesis committee: prof. dr. E. W. McCune TU Delft, ELCA

dr. M. Babaie TU Delft, ELCA

dr. ir. G. J. M. Janssen TU Delft, CAS

This thesis is confidential and cannot be made public until August 28, 2020.

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.

To my parents, my sister and my late friend Juri

Abstract

This thesis introduces the concept of phase rotation cancellation (PRC) as a previously unknown error mechanism in narrow-band FSK systems, and proposes a new FSK demodulation algorithm to resolve PRC errors. Phase rotation cancellation is defined as an event where the signal phasor rotation in the IQ plane is temporarily cancelled by the rotation of quadrature noise. The PRC errors are the dominant error mechanism in narrow-band FSK due to the small dynamic range at the output of a conventional FSK demodulator. In addition, any pre-modulation filtering further reduces the dynamic range and increases the probability for PRC errors to occur. Based on the discovered characteristics of PRC errors, a novel FSK demodulation algorithm is developed that greatly improves detection accuracy in FSK systems. Furthermore, the novel demodulation algorithm allows for sub-datarate receiver bandwidth operation, which offers improved receiver sensitivity and blocker performance. It is shown that the new demodulation algorithm eliminates nearly all PRC errors, leaving only errors due to clicks. In a GFSK system with $BT = 0.5$ and $h = 0.5$, an improvement in receiver sensitivity of up to 4.4 dB is achieved with the new demodulation algorithm.

Acknowledgements

Throughout the duration of the thesis project, I have received a substantial amount of support from many people around me. Therefore, I would like to take the time to express my gratitude to those who have been there for me.

First and foremost, I would like to thank my supervisor Earl McCune for the massive support and assistance throughout the project, even when thousands of kilometers away from each other. I cannot stress enough how many hours we have spent breaking down problems and coming up with solutions. You gave me the opportunity to work on a true research project, one with a perfect mixture of theory and practice, and I am proud that I have taken that offer. You are a role model to me, and I am sure that many others share this opinion.

I would like to thank my daily supervisor Masoud Babaie for the support and counsel on many occasions throughout the year, and always asking the right critical questions for me to answer. Together with Earl, you were a crucial factor in my progress and development throughout the thesis project. Thank you for taking part in this journey with me.

In addition, I would like to thank my friends for supporting me, helping me, occasionally offering me the necessary distraction, and listening to me talking about my project over and over again.

Finally, I would like to thank my family, in particular my parents and my sister, for their unconditional love and support and for always being there for me. We have had a tough year as a family, and I am more than happy to see that we can make it through anything as long as we stand together.

A. Šabanović
Delft, August 2019

Contents

List of Figures	xi
List of Tables	xv
List of Acronyms	xvii
1 Introduction	1
1.1 A Brief History of Frequency Modulation.	1
1.2 Problem Definition.	2
1.3 Research Objectives	5
1.4 Motivation	5
1.5 List of Contributions	6
1.6 Thesis Outline	6
2 Fundamentals of FSK	9
2.1 Binary Frequency Shift Keying	11
2.1.1 Time Domain Description	11
2.1.2 Spectrum and Bandwidth	12
2.1.3 Signal Phasor	15
2.2 Pre-modulation Filtering	17
2.3 Conventional FSK Demodulation Techniques	20
2.3.1 Theoretical Demodulators	20
2.3.2 Practical Demodulators	22
3 Error Mechanisms in FSK	25
3.1 Noise in Frequency-Modulated Systems	25
3.1.1 Band-Limited Quadrature Noise	26
3.1.2 Signal plus Noise	27
3.1.3 Clicks and Doublets	29
3.2 FM Threshold Effect	32
3.2.1 Ideal SNR Model	32
3.2.2 Rice SNR Model.	34
3.2.3 Consequences for FSK Systems	36

3.3	Phase Rotation Cancellation.	37
3.3.1	Concept	38
3.3.2	Noise Frequency	38
3.4	Simulation Results	40
4	Phase-based FSK Demodulation Algorithm	43
4.1	System Specifications	43
4.2	Phase-Based Demodulation Algorithm.	44
4.2.1	Phase Profile Reconstruction	44
4.2.2	Error Detection	46
4.3	Algorithm Non-idealities	49
4.3.1	Phase Offset.	49
4.3.2	Frequency Offset	52
4.3.3	Click Events	55
4.4	Alternative Algorithms.	56
4.4.1	Click Detection and Elimination	56
4.4.2	Data-dependent Decision Threshold	57
5	Algorithm Results	59
5.1	Simulation Results	59
5.2	Discussion	61
6	Conclusion	63
6.1	Summary	63
6.2	Recommendations for Future Work.	65
	Bibliography	67
A	GFSK Tool	69
B	MATLAB Scripts	71
C	Journal Paper	77

List of Figures

1.1	Bit error rate curves of BPSK/QPSK, 8-PSK, non-coherent FSK and coherent FSK. . . .	3
1.2	The required bit error rate to achieve packet error rate of 30.8% as a function of the packet size.	4
2.1	A basic block diagram of a digital wireless communication transmitter.	9
2.2	The time-domain signal phase of unfiltered CP-BFSK: (a) the phase tree and (b) an example of a phase profile.	12
2.3	The normalized PSD of unfiltered CP-BFSK: (a) $h = 0.1$, (b) $h = 0.3$, (c) $h = 0.5$, (d) $h = 0.7$, (e) $h = 1.0$, (f) $h = 2.6$	13
2.4	The normalized PSD of Gaussian-filtered BFSK with $BT = 0.5$: (a) $h = 0.1$, (b) $h = 0.3$, (c) $h = 0.5$, (d) $h = 0.7$, (e) $h = 1.0$, (f) $h = 2.6$	14
2.5	The occupied bandwidth of Gaussian-filtered BFSK ($BT = 0.5$) with BW6, BW20, BW60 and the adjusted Carson's rule: (a) $\alpha = \beta = 1.0$, and (b) $\alpha = 1.3$, $\beta = 0.9$	15
2.6	A vector diagram of an FSK signal (solid), including an indication of the phase rotation per symbol for $h = 0.2$ and $h = 0.6$ (dotted).	16
2.7	The phase as a function of the in-phase and quadrature components: (a) four-quadrant arctangent function (atan2), and (b) extended arctangent function.	17
2.8	PAM signals before and after Gaussian filtering, for BT -products $\{0.25, 0.5, 1.0, 2.0\}$. . .	19
2.9	The phase tree of a GFSK signal for $BT = 0.5$	19
2.10	The block diagram of the non-coherent FSK demodulator.	21
2.11	The block diagram of the coherent FSK demodulator.	21
2.12	The block diagram of the frequency discriminator demodulator.	22
2.13	The block diagram of the arctangent demodulator.	23
2.14	A digital frequency demodulation algorithm without an arctangent function.	24
3.1	The magnitude, angle and phase waveform of band-limited noise in absence of a signal, for a noise variance of $\sigma = 1$ and a relative bandwidth of $B = 0.1f_s$	27
3.2	A graphical representation of signal $r(t)$ and its components. Point P represents the momentary position of $r(t)$ in the IQ plane.	28
3.3	The Rice distribution for $\sigma = 1$ and $A = \{0, 0.5, 2, 5\}$	28
3.4	The magnitude, angle and phase waveform of an unmodulated carrier plus band-limited noise for a noise variance of $\sigma = 1$ and a relative bandwidth of $B = 0.1f_s$: (a) $\text{SNR} = 2$, and (b) $\text{SNR} = 10$	29

3.5	A simplified signal plus noise trajectory, starting from signal mean A . The trajectory consists of four sections representing different behavior: (1) Gaussian-like, (2) a click, (3) false clicks (type I, II or III) and (4) a doublet.	30
3.6	A click event representation: (a) phase waveform, (b) instantaneous frequency waveform, and (c) approximate output PSD.	31
3.7	A doublet event representation: (a) phase waveform, (b) instantaneous frequency waveform, and (c) approximate output power spectral density.	31
3.8	A block diagram of an 'ideal' FM receiver.	32
3.9	The output SNR curves as a function of the input SNR of the ideal model and the Rice model, for a deviation ratio of $B/(2f_a) = 5$	37
3.10	A vector diagram of an FSK modulated signal ($h = 0.5$) with and without AWGN channel. Highlighted in red are signal plus noise trajectories where a symbol error occurred.	39
3.11	The (normalized) instantaneous frequency distribution for $\sigma = 1$ and $A = \{0, 1, 3, 5\}$: (a) linear scale and (b) logarithmic scale.	39
3.12	A GFSK transceiver with a 4th-order Butterworth IF filter: (a) transmitter, (b) receiver.	40
3.13	The bit error rate curves as a result of click errors and PRC errors for (a) $h = 0.5$ and (b) $h = 1.0$	41
3.14	The bit error rate as a function of the modulation index h for click errors, PRC errors and the total errors, for $E_b/N_0 = 10$ dB.	41
4.1	The packet structure in BLE systems. For the specifications of Table 4.1, the PDU is 42 bytes.	44
4.2	Phase profile reconstruction: (a) transmitted and received phase profiles with an error in the third symbol, and (b) the corresponding phase error signal.	46
4.3	The instantaneous phase profile error (blue) and average phase error per symbol (red) in the presence of noise.	47
4.4	A flowchart representation of the demodulation algorithm process.	48
4.5	An example of the algorithm operation for $N = 2$ and $E_b/N_0 = 10$ dB: (a) the transmitted and received phase, and (b) the symbol-average phase profile error waveforms.	49
4.6	The transmitted (blue), received (red) and reconstructed (yellow) phase with a phase offset of $\theta_{os} = 1$ radian.	50
4.7	The bit error rate versus phase offset for $E_b/N_0 = 10$ dB.	51
4.8	The probability of failing to meet $\theta_{os} < 0.2$ rad, shown as a function of E_b/N_0 for the averaging method and non-averaging method.	52
4.9	The transmitted, received and reconstructed phase for a frequency offset of 25 kHz.	53
4.10	A frequency offset of $f_{os} = 24$ kHz: (a) phase error signal and (b) FFT magnitude of the phase error signal.	54
4.11	Click removal signal processing: (a) phase waveforms before and after CRSP, and (b) output SNR versus input CNR with and without CRSP.	57

5.1	Bit error rate curves for the demodulation algorithm with $N = \{0, 1, 2\}$, including the click error rate for a conventional demodulator.	60
5.2	Bit error rate curves for a receiver bandwidth of $B = \{0.8R_b, 1.2R_b\}$ for the conventional demodulator and the new demodulator with $N = 2$	60
5.3	The bit error rate versus receiver bandwidth for the conventional demodulator and the new demodulator with $N = 2$	61
A.1	The GUI of the developed GFSK simulation tool.	70

List of Tables

2.1	The phase rotation reduction factor c as a function of symbol sequence for Gaussian-filtered CP-BFSK with $BT = 0.5$	20
4.1	The system specifications used in the demodulation algorithm design, drawn from BLE.	44
4.2	The optimum decision threshold for GFSK with $BT = 0.5$	58
5.1	The minimum required IBEND to achieve $PER = 30.8\%$ for each performed simulation.	61

List of Acronyms

AC	Alternating Current
AM	Amplitude Modulation
ASK	Amplitude Shift Keying
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BFSK	Binary Frequency Shift Keying
BLE	Bluetooth Low Energy
BPSK	Binary Phase Shift Keying
BR	Basic Rate
CER	Click Error Rate
CNR	Carrier to Noise Ratio
CP-BFSK	Continuous Phase Binary Frequency Shift Keying
CP-FSK	Continuous Phase Frequency Shift Keying
CRSP	Click Removal Signal Processing
DC	Direct Current
DSP	Digital Signal Processing
DWC	Digital Wireless Communication
EE	Electrical Engineering
FFT	Fast Fourier Transform
FM	Frequency Modulation
FPGA	Field Programmable Gate Array
FSK	Frequency Shift Keying
GFSK	Gaussian-filtered Frequency Shift Keying
GSM	Global System for Mobile Communication
GUI	Graphical User Interface
IBEND	Individual Bit Energy to Noise Density
IEEE	Institute of Electrical and Electronic Engineers
IQ	In-phase and Quadrature phase
ISI	Inter-Symbol Interference

LO	Local Oscillator
PA	Power Amplifier
PAM	Pulse Amplitude Modulation
PDF	Probability Density Function
PER	Packet Error Rate
PRC	Phase Rotation Cancellation
PSD	Power Spectral Density
PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
RF	Radio Frequency
SNR	Signal to Noise Ratio
TCAS	Transactions on Circuits and Systems

Introduction

The focus of this thesis lies primarily on the digital modulation scheme of frequency shift keying (FSK); its characteristics, the benefits that accompany it, its drawbacks, and how to overcome these drawbacks. This introductory chapter first briefly discusses the long history of frequency modulation. The main disadvantages and challenges within the domain of FSK are discussed in the problem definition. Several research questions are formulated based on the problem definition, which comprise the research objectives of this thesis. Following the research objectives, a motivation for solving the formulated problems is given. The outline of this thesis is given at the end of this chapter, including a list of contributions to the scientific community that this thesis provides.

1.1. A Brief History of Frequency Modulation

Frequency modulation (FM), and its digital counterpart FSK, have played a significant role in the history of wireless communication. The first use of frequency modulation dates back to more than a century ago, in the early twentieth century. One of the earliest examples of FSK is found in an application of Morse code, as described in [1]; rather than distinguishing marks (dots) and spaces (dashes) by the corresponding time-length of their tone, the marks and spaces are each given a separate tone frequency. In doing so, the transmission time with respect to conventional Morse transmission is reduced by giving marks and spaces an equal time-length.

One of the most important driving forces behind the development of FM in the early twentieth century was the *static* hissing present in wireless communication systems. To solve this issue, an engineer by the name of Edwin Armstrong focused his attention on FM in 1928 [2]. His efforts were greatly rewarded; by the year 1935, he had successfully developed a system of wide-band frequency modulation for broadcasting purposes [3]. In addition to solving the problem of static, the constant-envelope characteristic of FM resulted in a transmitter efficiency unrivaled by any other broadcasting system, such as those that employ amplitude modulation (AM). However, in order to achieve this

noise suppression quality of FM, a much larger transmission bandwidth was required compared to that of AM. In addition, Armstrong discovered that frequency-modulated systems might not be very useful when considering weak signals, due to a phenomenon named the *FM threshold effect*.

Over the few decades that followed, frequency modulation had been adopted in many applications, such as in radio broadcasting, which initially employed AM. Due to its noise resilience, FM greatly outperformed its rival AM in broadcasting. In addition, with the increase of portability of wireless communication systems, the battery life – and therefore transmitter energy efficiency – became crucial, further favoring FM over the much more inefficient AM. By the year 1980, FM broadcasting had surpassed AM in popularity amongst radio audiences [2].

In recent decades, wireless communications rapidly became increasingly more accessible to the public. With the start of the information age, marked by the invention of the transistor, demand for user capacity and data rate had started increasing rapidly. Over time, the desire for transmitter energy efficiency was given up in favor of increased bandwidth efficiency. By then, digital modulation schemes such as phase shift keying (PSK) and quadrature amplitude modulation (QAM) became available, which offered a significant increase in bandwidth efficiency over FSK. It is then no surprise that nearly all applications that required high data rates adopted these more bandwidth-efficient modulation techniques rather than FSK.

As of today, QAM and PSK are amongst the most popular modulation schemes in digital wireless communication (DWC) systems, particularly due to their capability of achieving a high data rate in a narrow bandwidth. However, despite the disadvantages of FSK, its value still remains in some applications, particularly for low-power communication. Examples of existing applications of FSK are the Global System for Mobile communications (GSM) and some versions of Bluetooth, such as Bluetooth Low Energy (BLE).

1.2. Problem Definition

FSK modulation faces several drawbacks with respect to other modulation schemes. The drawbacks are demonstrated through multiple important performance metrics of DWC systems; the bandwidth efficiency, the bit error rate performance, and the receiver sensitivity.

Bandwidth Efficiency

One drawback that has already been mentioned is the relatively low bandwidth efficiency of FSK with respect to other (linear) modulation techniques such as the well-known QAM, or amplitude shift keying (ASK), the digital counterpart of AM. The bandwidth efficiency is expressed as

$$\eta_{BW} = \frac{R_b}{B}, \quad (1.1)$$

where R_b is the data rate in bit/s, and B is the occupied signal bandwidth. Transmission of an FSK signal results in an occupied signal bandwidth that is generally much larger than the data rate itself. Without any pre-modulation filtering, the use of FSK modulation would be *very* expensive in terms

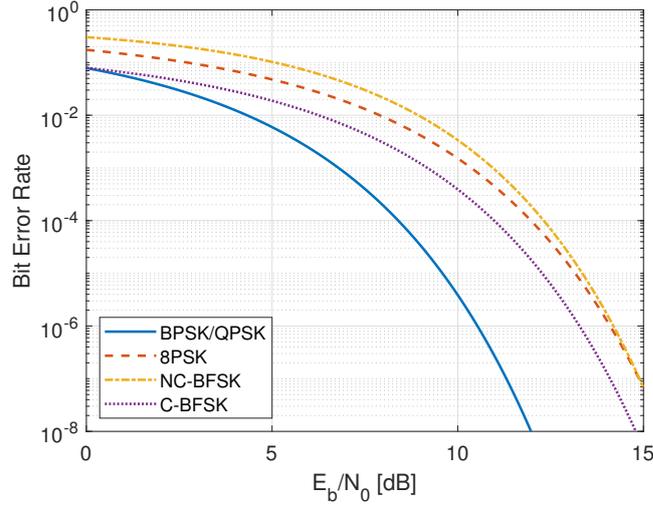


Figure 1.1: Bit error rate curves of BPSK/QPSK, 8-PSK, non-coherent FSK and coherent FSK.

of bandwidth. Furthermore, the low bandwidth efficiency does not only have consequences for the transmitted signal; at the receiver side of the DWC system, the receiver bandwidth needs to be wide enough to be able to receive the signal properly. A wider receiver bandwidth results in a larger noise power admitted to the receiver, which corrupts the received signal.

Bit Error Rate

Another important performance metric of DWC systems is the bit error rate (BER) performance. This metric is evaluated as a function of the energy per bit to noise power spectral density ratio E_b/N_0 , or equivalently the individual bit energy to noise density (IBEND) ratio. The IBEND is related to the received signal-to-noise ratio (SNR) by the bandwidth efficiency as [4]

$$SNR = \frac{E_b}{N_0} \eta_{BW}. \quad (1.2)$$

Examples of BER versus IBEND curves are shown in Figure 1.1 for binary phase shift keying (BPSK) or quadrature phase shift keying (QPSK), 8-PSK, and unfiltered (non-)coherent FSK. It is evident that the BER performance of FSK is relatively poor, particularly when taking into account that QPSK and 8-PSK have a *much* higher bandwidth efficiency than FSK.

Receiver Sensitivity

On the side of the receiver, one of the most important performance metrics is the receiver sensitivity, generally expressed as the minimum signal power required to 'successfully' demodulate the signal. The receiver sensitivity is mathematically written in logarithmic form as

$$P_{S,min}|_{dBm} = k_B T|_{dBm/Hz} + B_{RX}|_{dBHz} + NF + SNR_{min}|_{dB}. \quad (1.3)$$

Equation (1.3) separately shows all the elements that contribute to the receiver sensitivity: $k_B T$ is the power spectral density (PSD) of thermal noise; B_{RX} is the receiver bandwidth, which in turn

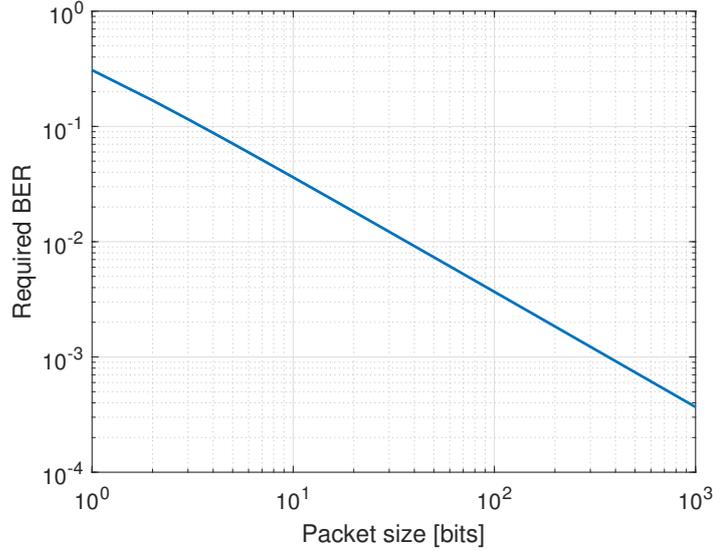


Figure 1.2: The required bit error rate to achieve packet error rate of 30.8% as a function of the packet size.

influences the total noise power at the input of the demodulator; NF is the noise figure, which is related to the contribution of the physical components in the receiver chain to the noise; and SNR_{min} is the minimum input SNR required to demodulate the signal. The latter is related through (1.2) to the minimum required IBEND. It is clear from (1.3) that both of the previously discussed performance metrics are somehow related to the receiver sensitivity. The bandwidth efficiency is related to the occupied bandwidth, which in turn determines the required receiver bandwidth, while the bit error rate performance is directly related to SNR_{min} .

Most parameters in (1.3) are clearly related to the physical characteristics of the receiver. However, this is not true for SNR_{min} ; it is an arbitrarily defined parameter within the scientific community. For example, in nearly all datasheets for FSK receivers for the application of BLE, $P_{S,min}$ is reported at a specific packet error rate (PER) of 30.8%. The PER is mathematically related to the BER through

$$PER = 1 - (1 - BER)^{N_p}, \quad (1.4)$$

where N_p is the number of bits in a single packet, or the packet size. Figure 1.2 shows the required BER to achieve a PER of 30.8% as a function of N_p . It is clear that proper documentation of the receiver sensitivity requires provision of both the used reference PER as well as the corresponding packet size.

Despite the more than a century long existence of FM and FSK, its mechanisms and intricate parameters are often not well understood, even within the electrical engineering (EE) community. This is partly due to the non-linear nature of FSK, which makes any mathematical analysis significantly more complicated. Furthermore, the response of a frequency demodulator to additive white Gaussian noise is also non-linear, which makes FSK modulation vastly different from other mod-

ulation schemes. The error mechanisms in FSK systems are not well understood nor well documented. Developing an understanding of these error mechanisms would be extremely helpful in the implementation of new FSK demodulation techniques that offer improved receiver sensitivity performance.

1.3. Research Objectives

The research objective of this thesis is to develop an understanding of the error mechanisms involved in FSK modulation, specifically when operating through an additive white Gaussian noise (AWGN) channel. Based on the developed understanding of the error mechanisms and their characteristics, a novel demodulation technique is proposed that offers an improved FSK system performance. The main research questions that this thesis answers are summarized as:

1. What are the underlying error mechanisms and their corresponding parameters in FSK systems?
2. Which demodulation method(s) can be used to solve the underlying error mechanisms in FSK systems?

1.4. Motivation

From the perspective of academic research, the investigation and documentation of the underlying error mechanisms in FSK, including their parameters and physical relationships, is in itself already a substantial contribution to the scientific community. Investment of research efforts in this topic builds a foundation for future development and improvement of FSK systems.

The value of the constant-envelope characteristic of FSK is one of the most important motivations to develop and improve FSK systems. Power efficiency is essential in this age, where a substantial part of wireless communication is portable. FSK systems are particularly valuable in applications where long battery life is important or even crucial. For example, electronic implants that are required to convey information to some place outside of the human body are preferably charged or replaced as few times as possible, particularly for implants that monitor the health of a patient. Furthermore, partly concerning the recent climate change developments, increasingly more research effort is invested in the improvement of energy efficiency in modern applications. FSK modulation is particularly useful in applications that do not require a large data rate, but could benefit greatly from a high energy efficiency. Bluetooth Low Energy is an excellent example of this. By improving FSK system performance, an incentive is given to start using FSK more in applications that are suited for it.

Finally, it is noted that most of the development of FM and FSK happened decades ago, when the attention for frequency modulation was at its largest. Problems such as the FM threshold effect had very limited solutions, primarily due to the limited resources available at that time. In the

present day, with the enormous computational power of digital signal processing (DSP), there exists a good probability that a solution to any problems in FSK systems is found in the digital domain.

1.5. List of Contributions

This thesis provides several original contributions to the scientific community. A summary of the major contributions is given below:

- The development of a MATLAB application intended for extensive simulation of GFSK modulation and demodulation, which has been used frequently throughout the thesis project. The MATLAB application description is included in Appendix A.
- A description of the effect of pre-modulation filtering in FSK systems, namely, an inevitable reduction in phase rotation per symbol, and therefore a reduction in dynamic range at the output of an FSK demodulator.
- The introduction of *phase rotation cancellation* as a dominant error mechanism in narrow-band FSK systems, which has not been described before.
- A description of the relative importance of the two dominant error mechanisms in FSK systems, clicks and phase rotation cancellation, as a function of the modulation index.
- The introduction of the concept of *noise frequency*, defined as the quadrature noise phasor rotation in the IQ plane, which has a not yet mathematically described distribution that is dependent on both noise variance and signal mean.
- A scientific paper written on the topic of 'FSK demodulation errors from phase rotation cancellation', submitted to IEEE Transactions on Circuits and Systems II (TCAS-II) on April 16, 2019. The paper is included in Appendix C.
- A novel FSK demodulation algorithm that offers a receiver sensitivity improvement of up to 4.5 dB at a packet error rate of 30.8%, and allows for sub-datarate receiver bandwidth operation, which is not feasible with conventional demodulation techniques.

1.6. Thesis Outline

The outline of this thesis is as follows. Chapter 2 describes FSK modulation and its characteristics, and provides the necessary background information for the reader to fully understand the concepts introduced in the remaining chapters. This chapter also includes a description of pre-modulation filtering and its effect on FSK signals, and provides multiple conventional FSK demodulation techniques.

Chapter 3 discusses noise in frequency-modulated systems, and describes the two fundamental error mechanisms that are dominant in FSK systems, clicks and phase rotation cancellation. The

latter error mechanism is introduced for the first time by this thesis.

Chapter 4 describes a novel FSK demodulation technique that offers a significantly improved receiver sensitivity. In addition, this chapter discusses the main non-idealities of the new demodulation algorithm and how to mitigate their effects.

Chapter 5 analyzes the receiver sensitivity improvement when using the demodulation algorithm with respect to conventional demodulation techniques. Two embodiments of the demodulation algorithm are simulated and compared to results obtained from the simulation of a conventional demodulation technique. A discussion of the obtained results is included at the end of this chapter.

Chapter 6 concludes the thesis with a summary and recommendations for future work.

2

Fundamentals of FSK

This chapter serves as a general introduction to the principles of FSK modulation. Understanding the background information that is presented in this chapter is necessary to fully understand the concepts introduced in the remainder of this thesis.

In general, the objective of a DWC system is to convey information from one place to another with the use of a wireless signal. Figure 2.1 shows the basic procedure of generating such a wireless signal.

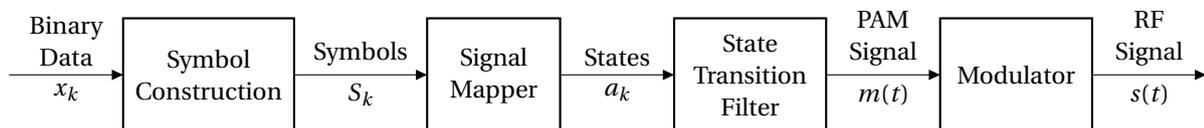


Figure 2.1: A basic block diagram of a digital wireless communication transmitter.

The input of the DWC transmitter is binary data or 'bits' x_k . It is irrelevant for the transmitter what the bits actually represent; only the polarity of the bits is of interest. In many DWC systems, multiple bits are grouped together in what is called a symbol. Each possible symbol is a unique combination of bits; a group of n bits has $M = 2^n$ possible unique combinations, and therefore M unique symbols. Generally, M is referred to as the modulation order, and a signal with M possible unique symbols is an M -ary signal. For example, in 8-PSK, there are eight unique symbols. In 2-state (binary) systems, each bit is acted upon individually rather than in groups, resulting in only two possible symbols. However, to avoid ambiguity in terminology, the term 'symbol' is still preferred regardless of the number of bits are contained in a symbol.

In order to physically transfer the desired information, the constructed symbol sequence needs to be converted to a sequence of signal states which can manipulate signal parameters. The signal mapper does this by associating each of the M unique symbols with a unique signal state a_k .

Each signal state consists of a single number (one-dimensional) or a combination of numbers (N-dimensional). For example, if each symbol is mapped to a particular value of both amplitude and phase of a signal, then each signal state a_k must be a two-dimensional set of numbers.

At this point, the binary data x_k is represented by a sequence of particular signal states. The next step in the process is to create a physical waveform that represents the information. This requires that not only the signal states themselves are defined, but also the transition between states. The state transition filter in Figure 2.1 constructs the pulse amplitude modulated (PAM) signal, or simply the *information signal*, from the signal state sequence. A generalized PAM signal is written mathematically as

$$m(t) = \sum_{k=-\infty}^{\infty} a_k g(t - kT_{sym}), \quad (2.1)$$

where $g(t)$ represents the band-limiting pulse corresponding to a particular symbol. In nearly all DWG systems, each pulse $g(t)$ is shaped with a pre-modulation or pulse-shaping filter, which is often necessary to meet occupied bandwidth requirements set by particular standards. For simplicity, a unit magnitude is chosen as the maximum for $g(t)$, such that the information signal $m(t)$ is bounded within the range ± 1 .

In order to transmit the information signal, it needs to be imposed on a physical signal that can 'carry' the information at the desired frequency. This signal, typically referred to as a *carrier wave*, is written in its general form as

$$s(t) = A \cos(\omega_c t + \phi). \quad (2.2)$$

From (2.2), it is clear that three available parameters can be used to represent the information signal: the amplitude A , the angle ϕ , or the instantaneous frequency ω_c . By modulating one (or more) of these parameters with the information signal, the modulated carrier wave can be transmitted and, at a different place, received by a receiver which retrieves the information.

In FSK systems, the modulated parameter is the frequency. It is tempting to intuitively denote the modulation in (2.2) by substituting ω_c with

$$\omega_c(t) = \omega_0 + \omega_d m(t), \quad (2.3)$$

where ω_0 is a fixed frequency offset, and ω_d is the deviation frequency. This notation for FSK signals is common in many textbooks. However, the notation is only mathematically correct for the particular case in which $g(t)$ has a rectangular shape (i.e., $g(t) = \text{rect}(t/T_{sym})$). For any other pulse shapes, the product of (2.3) with t in (2.2), when differentiated, does not result in the correct relationship between modulation and instantaneous frequency. Therefore, it is preferred to denote FSK modulation as

$$\phi(t) = \omega_d \int_{-\infty}^t m(\tau) d\tau, \quad (2.4)$$

such that its time derivative $\phi'(t)$ is the modulation signal scaled with ω_d . In this notation, the instantaneous frequency of the modulated carrier has the correct relationship with $m(t)$. In addition, the notation illustrates the fact that even though the instantaneous frequency is modulated, the actual involved *physical* parameter is the phase [5].

In this chapter, a description of binary FSK (BFSK) and its characteristics is given. Then, the consequences of pre-modulation filtering on FSK signals are discussed. As a part of the literature survey, existing demodulation techniques are briefly presented and discussed at the end of the chapter. These demodulators are separated into two groups: the theoretical (non-)coherent demodulator, and demodulators that are commonly used in practice.

2.1. Binary Frequency Shift Keying

In a BFSK system, the instantaneous frequency parameter of the carrier is modulated by a PAM signal with $a_k \in \{-1, 1\}$. This means that the instantaneous frequency during transmission is either greater or less than the carrier frequency, and the carrier frequency itself is in fact never explicitly transmitted. While it is not necessary that the two transmitted instantaneous frequencies are centered around the carrier frequency, it does result in a simplified description of FSK, and there is no particular benefit to having it unbalanced.

In the following sections, the description of BFSK and its characteristics is separated in three groups, including: a time-domain description of BFSK signals; the spectral shape and occupied bandwidth of BFSK signals; and the behavior of BFSK signals in the phasor domain.

2.1.1. Time Domain Description

In the time-domain, the FSK-modulated carrier wave is written as

$$s(t) = A \cos[\omega_c t + \phi(t)] = A \cos\left[\omega_c t + \omega_d \int_{-\infty}^t m(\tau) d\tau\right], \quad (2.5)$$

where ω_c is the carrier frequency in rad/s, $\omega_d = 2\pi f_d$ is the deviation frequency in rad/s, and $m(\tau)$ is the information signal (i.e., the PAM signal). If no pre-modulation filter is applied, then the instantaneous frequency of this waveform is $\omega_c \pm \omega_d$, depending on which symbol is transmitted. From (2.5), the constant-envelope characteristic is immediately evident; while modulating the carrier, the amplitude A remains constant. As a consequence, the power amplifier (PA) used in an FSK transmitter has no circuit linearity requirement; a practical switch-mode (e.g., class E) PA can give an FSK system close to 100% power efficiency. This characteristic is very desirable in all applications.

Similarly to other modulation schemes, an FSK signal has a modulation index that indicates how strongly the modulated parameter varies. In FSK, the modulation index h is defined as

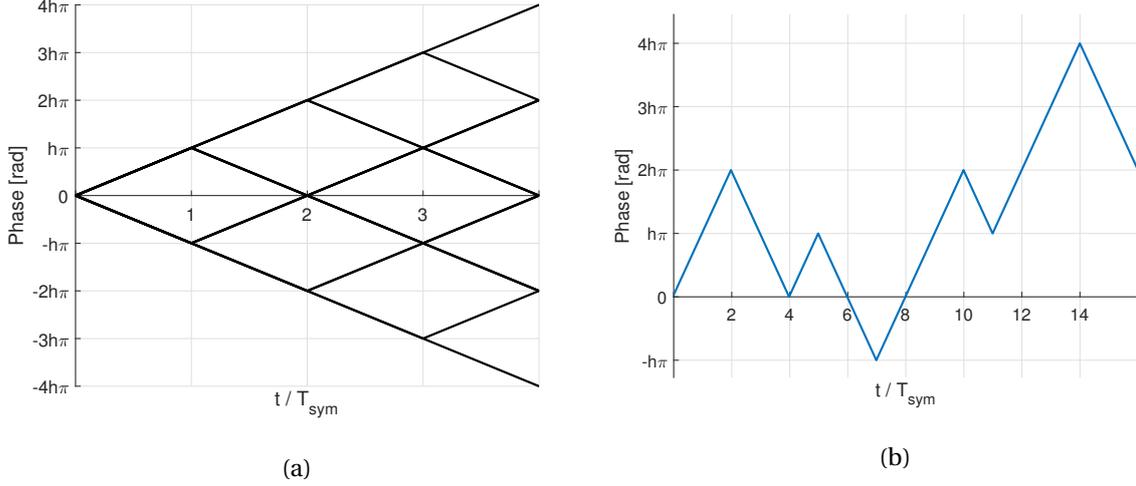


Figure 2.2: The time-domain signal phase of unfiltered CP-BFSK: (a) the phase tree and (b) an example of a phase profile.

$$h \equiv \frac{\omega_d T_{sym}}{\pi} = 2f_d T_{sym}, \quad (2.6)$$

where T_{sym} is the symbol time in seconds, and $f_d = \omega_d/(2\pi)$ is the deviation frequency in Hz. The modulation index relates the instantaneous frequency (with respect to the carrier) to the phase change per symbol time. This is easily seen by rewriting (2.6) to $\omega_d T_{sym} = h\pi$. The interpretation of this relation is as follows; for every symbol, the total phase rotation during that symbol is $\Delta\phi_{sym} = \pm h\pi$, where the polarity depends on the symbol state a_k . Under the condition that the phase is continuous even across symbol boundaries (continuous-phase FSK or CP-FSK), the total phase change over the entire FSK signal is

$$\Delta\phi = h\pi \sum_{k=-\infty}^{\infty} a_k g(t - kT_{sym}). \quad (2.7)$$

For unfiltered CP-FSK, $g(t) = \text{rect}(t/T_{sym})$ and this sum always has unit magnitude. The net phase change is determined solely by the difference between the number of times a '1' K_1 and a '0' K_0 is transmitted:

$$\Delta\phi = h\pi \left(\sum_{k=0}^{K_1} - \sum_{k=0}^{K_0} \right). \quad (2.8)$$

As a result, the phase $\phi(t)$ is, at any time, on a fixed grid of a triangular-shaped *phase tree*. The generalized phase tree for unfiltered CP-BFSK is illustrated in Figure 2.2a. A specific path on the phase tree is referred to as a phase profile. An example of a phase profile is shown in Figure 2.2b.

2.1.2. Spectrum and Bandwidth

Since FSK is a non-linear modulation scheme – as evident from its time-domain representation in (2.5) – there is no requirement that the PSD of FSK signals matches the PSD of $m(t)$. Indeed, the

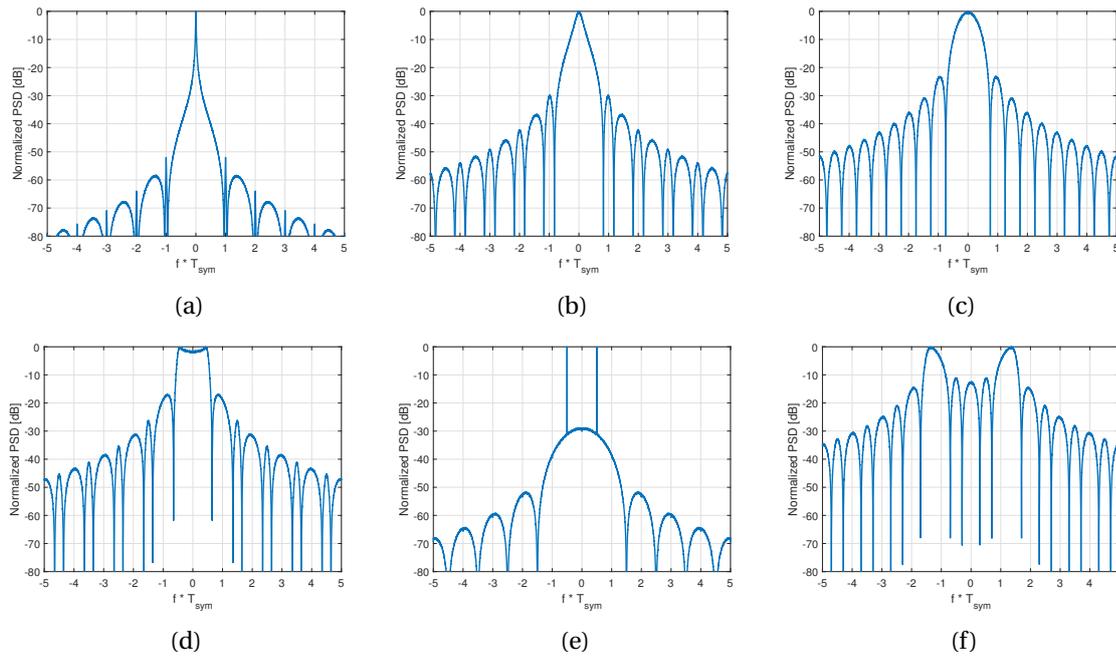


Figure 2.3: The normalized PSD of unfiltered CP-BFSK: (a) $h = 0.1$, (b) $h = 0.3$, (c) $h = 0.5$, (d) $h = 0.7$, (e) $h = 1.0$, (f) $h = 2.6$.

PSD varies widely for different values of h . In Figure 2.3, the power spectral density of an unfiltered CP-FSK signal is shown for various values of h . Aside from the varying shape of the spectrum, many side-lobes are seen in these FSK signal spectra caused by the discontinuities of $m(t)$ at the symbol transition. Typically, filtering is used to provide both waveform and derivative continuity on $m(t)$, which according to the Fourier transform directly reduces the magnitude of the side-lobes.

The spectral efficiency in any DWC system is given by

$$\eta_{BW} = \frac{R_b}{B}, \quad (2.9)$$

where R_b is the data rate of the system in bits/s, and B is the actual occupied bandwidth of the signal in Hz. In order to obtain an accurate estimate of this performance metric in the case of an FSK system, B needs to be defined somehow. One method is to define B as the frequency span where the signal exceeds a certain relative magnitude, such as -6 dB (BW6) or -60 dB (BW60) relative to the maximum of the PSD. However, as seen in Figure 2.3, the resulting occupied bandwidth can vary immensely based on which reference value is taken. This is especially true when the modulation index is close to an integer number, where the side-lobe magnitudes suddenly appear to decrease with respect to that of lower modulation indices. Consequently, the BW60 metric actually occasionally decreases with increasing modulation index. In addition, the side-lobes theoretically decay relatively slowly and extend infinitely far in the frequency domain [6], resulting in an incredibly low bandwidth efficiency if the BW60 metric is to be taken as the occupied bandwidth. This should motivate the necessity of a pre-modulation filter in FSK systems to limit these side-lobes. However, to ensure that the FSK application meets the channel separation requirements set by various stan-

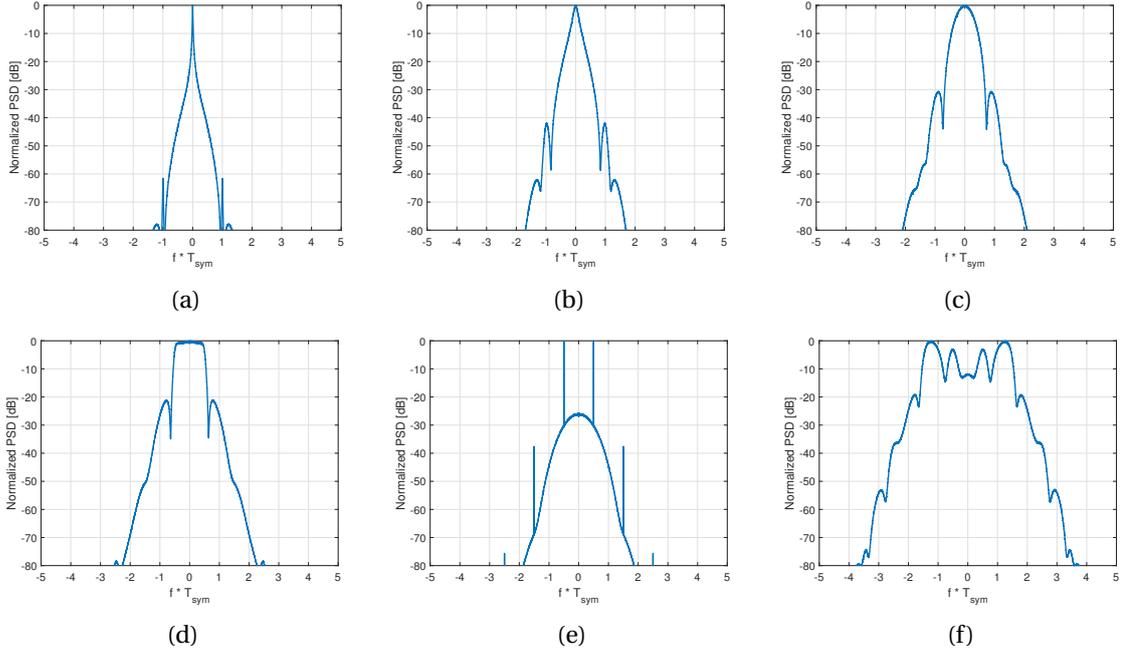


Figure 2.4: The normalized PSD of Gaussian-filtered BFSK with $BT = 0.5$: (a) $h = 0.1$, (b) $h = 0.3$, (c) $h = 0.5$, (d) $h = 0.7$, (e) $h = 1.0$, (f) $h = 2.6$.

dards around the world, it is still preferred to take a large bounded bandwidth metric such as BW_{60} rather than, for example, BW_6 . Furthermore, in order to obtain realistic and practical values of B , only filtered FSK is considered in this thesis, as unfiltered FSK is rarely used in modern practice.

Now a realistic value for the bandwidth efficiency is obtained. Similarly to Figure 2.3, the spectra of Gaussian-filtered CP-BFSK signals ($BT = 0.5$) for various values of h are shown in Figure 2.4. From these power spectra, it is evident that the BW_{60} metric for filtered CP-BFSK is much more confined than for the unfiltered case. The application of BLE, for example, uses a modulation index of $h = 0.5$ and the same Gaussian filter BT -product as in Figure 2.4. With an occupied bandwidth BW_{60} of approximately three times the symbol rate, the bandwidth efficiency of BLE is $\eta_{BW,BLE} \approx 0.33$. Though if for example BW_{20} is taken as B , η_{BW} would be vastly different. It is, therefore, vital to define B clearly before evaluating and reporting η_{BW} values.

In order to relate the occupied bandwidth to the modulation index, Carson's bandwidth rule for analog FM signals needed into an FM demodulator is taken as a starting point [5]:

$$B_{Carson} = 2(\Delta f + f_m), \quad (2.10)$$

where Δf is the peak frequency deviation, and f_m is the highest modulating frequency. For FSK, $f_m = 1/(2T_{sym})$. Carson's rule accounts for at least 98% of the energy in the power spectrum of a sine-wave modulated analog FM signal, and is a simple method to obtain an estimate on the occupied bandwidth of the signal. Equation (2.10) shows that for small values of h the occupied bandwidth is dominated by the modulation signal bandwidth, while for wide-band systems, it is the frequency

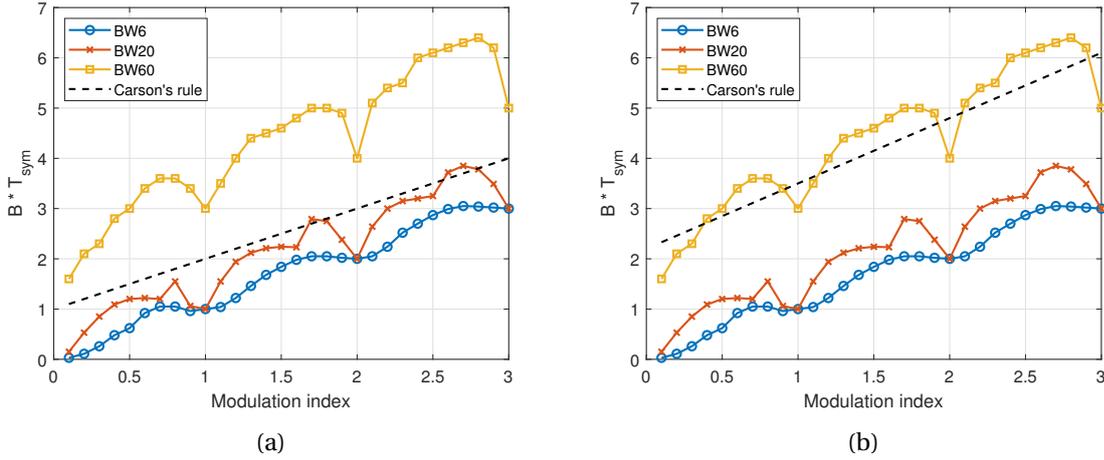


Figure 2.5: The occupied bandwidth of Gaussian-filtered BFSK ($BT = 0.5$) with BW6, BW20, BW60 and the adjusted Carson's rule: (a) $\alpha = \beta = 1.0$, and (b) $\alpha = 1.3$, $\beta = 0.9$.

deviation that dominates.

As an initial estimate for FSK signals, the direct substitution of the frequency deviation from (2.6) and half the symbol rate for the maximum modulation frequency in (2.10) results in

$$B_{BFSK} = 2 \left(\frac{h}{2T_{sym}} + \frac{1}{2T_{sym}} \right) = \frac{h+1}{T_{sym}}. \quad (2.11)$$

Equation (2.11) typically results in a bandwidth estimate of somewhere between BW20 and BW25 for Gaussian-filtered BFSK with $BT = 0.5$. A slight adjustment in slope and intercept of the line in (2.11) is sufficient for an estimate of BW60:

$$B_{BFSK} = \frac{(\alpha h + \beta)}{T_{sym}}. \quad (2.12)$$

The slope α and intercept β are dependent on the type of filter and the degree of filtering. For a Gaussian filter with $BT = 0.5$, $\alpha = 1.3$ and $\beta = 0.9$ results in a good approximation of BW60 over the provided range of values for h . Figures 2.5a and 2.5b show BW6, BW20 and BW60 versus h , including Carson's rule with $\alpha = \beta = 1.0$ and $\alpha = 1.5$, $\beta = 0.9$, respectively.

2.1.3. Signal Phasor

A signal phasor is the vector representation of a sinusoidal signal in the complex plane. Equation (2.5) shows the modulated carrier signal in polar form. In exponential notation, (2.5) becomes

$$s(t) = \Re \left\{ A e^{j\phi(t)} e^{j\omega_c t} \right\}. \quad (2.13)$$

The signal phasor in (2.13) is $Z(t) = A e^{j\phi(t)}$ is referred to as the complex amplitude of $s(t)$. The amplitude A and phase $\phi(t)$ determine the position of the signal phasor in the complex plane. Alternatively, using Euler's identity, (2.13) is written in Cartesian form as

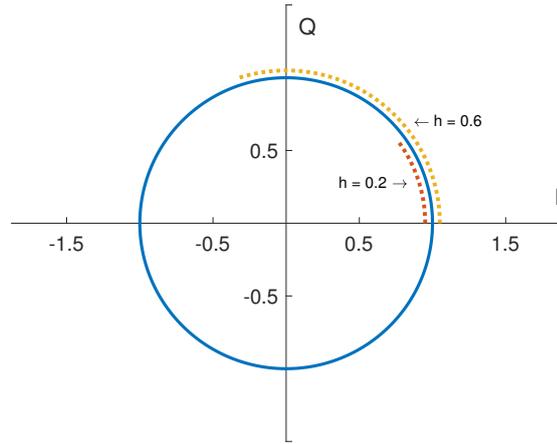


Figure 2.6: A vector diagram of an FSK signal (solid), including an indication of the phase rotation per symbol for $h = 0.2$ and $h = 0.6$ (dotted).

$$s(t) = I(t) \cos(\omega_c t) - Q(t) \sin(\omega_c t), \quad (2.14)$$

with

$$I(t) = A \cos[\phi(t)], \quad \text{and} \quad Q(t) = A \sin[\phi(t)]. \quad (2.15)$$

Signals $I(t)$ and $Q(t)$ are defined as the *in-phase* and *quadrature phase* components of signal $s(t)$, respectively. In this notation, the signal phasor is written as $Z(t) = I(t) + jQ(t)$, which shows the signal phasor position in Cartesian coordinates I and Q . In the DWC community, it is common practice to refer to the complex plane as the 'IQ plane', named after the signal components I and Q . The real axis then represents I , and the imaginary axis represents Q .

From (2.13) and (2.14), it is evident that the modulation signal determines the behavior of the signal phasor in the IQ plane. For QAM or PSK signals, each possible symbol is mapped to a certain *position* in the IQ plane through two-dimensional signal states a_k , resulting in a constellation of possible complex amplitudes. However, since the information in an FSK signal is in the *rotation* of the signal, it is impossible to define a constellation diagram in the conventional sense for FSK. Instead, a more useful diagram for FSK signals is a vector diagram, which shows the signal trajectory through the IQ plane rather than signal states. An example of the vector diagram of FSK signals is shown in Figure 2.6. In addition, the phase rotation of a single symbol, for a modulation index of $h = 0.2$ and $h = 0.6$, is indicated in the illustration. Due to the constant envelope of FSK signals, the vector diagram of FSK in the absence of noise is simply a circle with radius A and its center at the IQ plane origin.

As implied by the phase tree in Figure 2.2, there is no physical boundary on the value of the

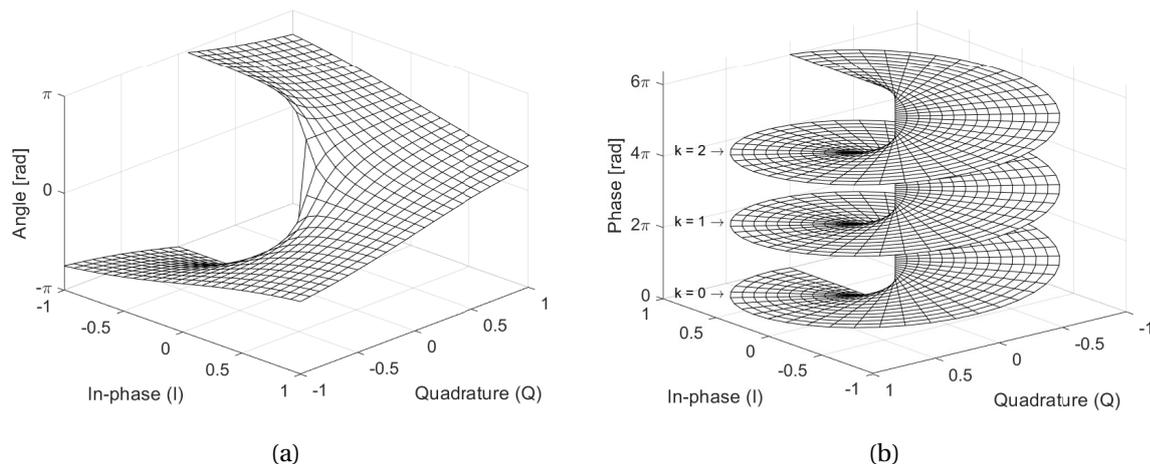


Figure 2.7: The phase as a function of the in-phase and quadrature components: (a) four-quadrant arctangent function (atan2), and (b) extended arctangent function.

phase; given enough time, the phase can increase to any real number. It is therefore not sufficient to consider the *angle* of the signal phasor, which is only defined in the range $[-\pi, \pi)$, and is calculated with the well-known four-quadrant arctangent or 'atan2' function. In order to illustrate this, the atan2 function is graphically shown in Figure 2.7a. From this surface plot, it is evident that there is a discontinuity at the radial line with angle $-\pi$. However, phase is a continuous parameter, so this description is not complete [6].

In order to correctly represent the *phase* $\phi(t)$, the four-quadrant arctangent function for angle is extended to include all real numbers for phase [6]:

$$\phi(t) = \tan_c^{-1} \left(\frac{Q(t)}{I(t)} \right) = \tan_2^{-1} \left(\frac{Q(t)}{I(t)} \right) + 2\pi k(t) + 2\pi m, \quad (2.16)$$

where $k(t)$ is a time-dependent integer-valued variable that tracks in which helicoid turn the signal phasor currently is, and m is an arbitrarily assigned phase origin. This brings $\phi(t)$ into the correct association with the physical behavior of the phase [7]. The resulting surface is shown in Figure 2.7b, where several turns of the helicoid are shown, with an indication of the value of k for each turn. For simplicity, m is chosen to be zero in this illustration. This surface meets the phase continuity criteria, as is expected from the physical phase of the signal.

2.2. Pre-modulation Filtering

As shown in subsection 2.1.2, the power spectrum of a CP-BFSK signal consists of many side-lobes that theoretically span an infinite bandwidth. The side-lobes are a consequence of the discontinuity at symbol boundaries in the PAM signal $m(t)$. This is impractical for many systems, particularly due to constraints set by adjacent channels in the frequency spectrum, since transmitting outside of the allocated bandwidth is typically not allowed through government regulation. In order to combat this, the occupied bandwidth of an FSK signal can be reduced by changing the shape of the pulse

signal $g(t)$ from a rectangular shape to a more smooth shape. The shaping of the information pulses is referred to as pre-modulation filtering or pulse-shape filtering, and is performed in nearly all modern applications for wireless communication.

While there are many filters available that are used to shape the pulses $g(t)$, this thesis focuses particularly on Gaussian filters, as they are commonly used in applications of FSK. Gaussian-filtered FSK (GFSK) is, for example, used in some of the Bluetooth standards such as Basic Rate (BR) and BLE, as well as in the GSM mobile phone standard. The Gaussian filter has a time-domain impulse response of

$$h(t) = \frac{2\pi BT}{\ln(2)} \exp\left[-\frac{2}{\ln(2)}(\pi BT t)^2\right], \quad (2.17)$$

where BT is the bandwidth-time product of the filter and is a measure for the time-domain pulse width. Equivalently, its frequency-domain transfer function is

$$H(f) = \exp\left[-\frac{\ln(2)}{2}\left(\frac{f}{BT}\right)^2\right]. \quad (2.18)$$

Both $h(t)$ and $H(f)$ are of identical shape. This is a special case of the Fourier transform. The pulse shape $g(t)$ for a single symbol is obtained through convolution of a rectangular pulse with the impulse response of the Gaussian filter $h(t)$ [5]:

$$g(t) = G \left\{ \operatorname{erf}\left[G\sqrt{2}\left(t + \frac{1}{2}\right)\right] - \operatorname{erf}\left[G\sqrt{2}\left(t - \frac{1}{2}\right)\right] \right\}, \quad \text{with} \quad G = \frac{\pi BT}{\sqrt{\ln(2)}}. \quad (2.19)$$

The impulse response in (2.17) is both non-causal and theoretically requires an infinite time-window. Because of this, a pure Gaussian filter is physically not realizable. However, due to its rapid exponential decay, the implementation of a sufficiently accurate Gaussian-shaped impulse response is feasible with a finite time-window and a delay to compensate for the non-causality.

An example of a Gaussian-filtered sequence of symbols is shown in Figure 2.8 for multiple BT values. It is evident from this figure that a small BT -product results in a partial-response filter, as the PAM signal in that case does not reach unit magnitude for every individual symbol. For small values of the BT -product, the Gaussian-shaped pulse spans multiple symbol periods, resulting in inter-symbol interference (ISI) and the reduction of response magnitude for some symbols.

While the Gaussian-filtered PAM signal can be used irrespective of the modulation scheme that is applied, there is a small difference in the way it influences FSK signals with respect to that of other modulation schemes. In other modulation schemes, such as QAM or PSK, a full-response PAM signal results in the complex baseband signal eventually reaching the constellation position in the IQ plane, even though its transitions have been filtered. For FSK signals, a full-response PAM signal means that the maximum *instantaneous frequency deviation* is (eventually) reached. However, since the actual physical parameter is the *phase*, the signal of particular interest is actually the integral of the PAM signal. Irrespective of the BT -product, the integral of the filtered PAM signal shows a reduced response per symbol with respect to the rectangular PAM signal. As a result, the to-

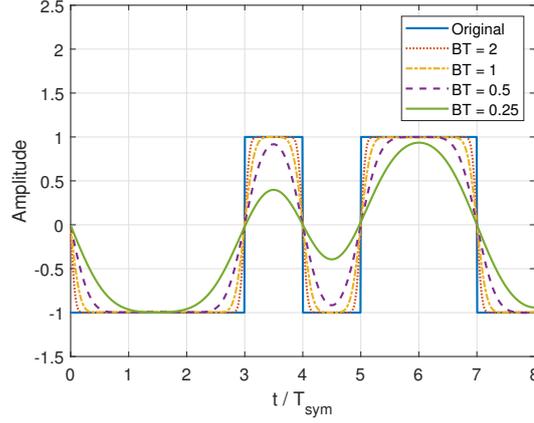


Figure 2.8: PAM signals before and after Gaussian filtering, for BT-products {0.25, 0.5, 1.0, 2.0}.

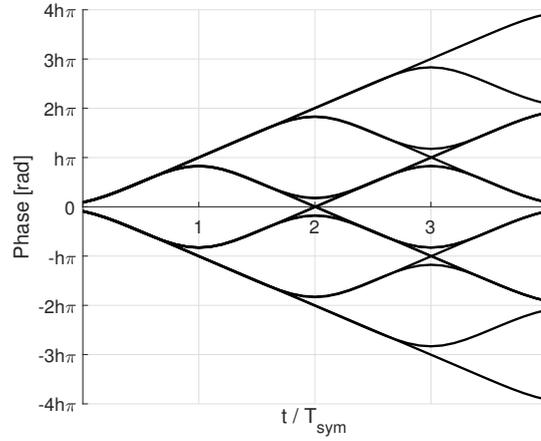


Figure 2.9: The phase tree of a GFSK signal for BT = 0.5.

tal *phase rotation* per symbol decreases when filtering is applied to the PAM signal. While this does not necessarily introduce ISI in the conventional sense, some data-dependency does occur. This is further illustrated with the Gaussian-filtered phase tree, shown for BT = 0.5 in Figure 2.9.

From Figure 2.9, it is evident that a phase rotation per symbol of $\Delta\phi_{sym} = \pm h\pi$ is no longer guaranteed when applying pre-modulation filtering. The phase rotation per symbol is now dependent on surrounding symbols:

$$\Delta\phi_{sym} = ch\pi, \quad (2.20)$$

where $c = c(BT, x_{k-N}, \dots, x_{k+N})$ is dependent on BT, the current symbol x_k and on multiple surrounding symbols x_{k-i} . Approximately for BT > 0.3, only the two directly adjacent symbols have a significant effect on the phase rotation of a particular symbol. Under this assumption, the phase reduction factor c is written as

$$c \approx (2x_k - 1) + \alpha(x_{k+1} + x_{k-1} - 2x_k), \quad (2.21)$$

Table 2.1: The phase rotation reduction factor c as a function of symbol sequence for Gaussian-filtered CP-BFSK with $BT = 0.5$.

Sequence	'000'	'001'	'010'	'011'	'100'	'101'	'110'	'111'
c	-1	$-(1 - \alpha)$	$(1 - 2\alpha)$	$(1 - \alpha)$	$-(1 - \alpha)$	$-(1 - 2\alpha)$	$(1 - \alpha)$	1

where $x_k \in \{0, 1\}$, and α depends on the filter type and the degree of filtering. The phase reduction factor c is shown in Table 2.1 for all eight possible combinations of symbols. It is evident that a reduction occurs for nearly all combinations. Furthermore, it is observed that a reduction of α is twice as likely as a 2α reduction or no reduction. It follows from simulations that $\alpha \approx 0.2$ for $BT = 0.5$, resulting in a worst-case phase rotation reduction of nearly 40%. As a consequence, the dynamic range at the output of an FSK demodulator is (in some cases dramatically) reduced when pre-modulation filtering is applied. This becomes important in Chapter 3 when noise is introduced.

2.3. Conventional FSK Demodulation Techniques

In this section, several examples of demodulation techniques for FSK signals are presented. It is important to note that the goal of this section is to be a short survey or overview of the various FSK demodulation techniques, rather than an extensive tutorial. Only short operation descriptions are included. First, an example of a coherent and non-coherent demodulator are presented, followed by three examples of widely-used demodulation techniques in modern applications.

2.3.1. Theoretical Demodulators

The coherent and non-coherent demodulator exhibited in this section are typically shown in textbooks, in order to illustrate the difference in coherent and non-coherent detection. The term 'theoretical' demodulator, in this case, does not imply that a physical realization is impossible; they are rarely used in any actual applications of FSK.

Non-coherent Demodulator

The non-coherent FSK demodulator, shown in Figure 2.10, uses two band-pass filters that each let through one of the signal states f_0 or f_1 while attenuating the other. The band-pass filter is succeeded by an envelope detector that detects the magnitude of any signal that comes through. The decision element then determines which path has the largest detected envelope, and decides on the detected symbol. The name of the demodulator comes from the fact that this demodulation technique does not require any phase synchronization. This demodulator is typically more suited for wide-band FSK signals, since the matched filter requirements are relaxed when the frequency spacing increases.

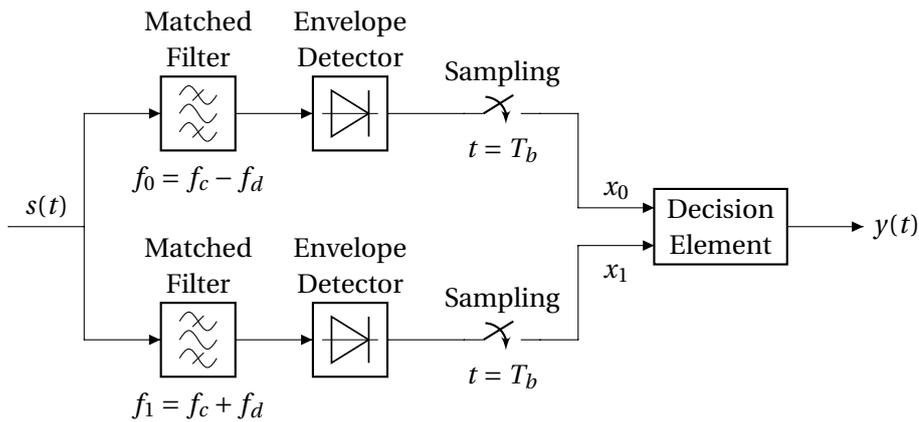


Figure 2.10: The block diagram of the non-coherent FSK demodulator.

Coherent Demodulator

The coherent demodulator in Figure 2.11 consists of two correlators, one for each signal state, and a decision element. The correlator itself includes a combination of a mixer and an integrator. The signal is down-converted using a reference signal with frequency f_0 or f_1 , after which the integrator detects the 'similarity' between the reference wave and the received signal. If the frequency of the received signal and reference signal match perfectly, then the signal is down-converted to DC, and the integrator outputs a relatively large value. The decision element determines which symbol is transmitted based on both integrator outputs. After one symbol time, the integrator 'dumps' its output value and starts over.

For the coherent demodulator to operate correctly, the reference signal needs to be phase-synchronous and phase-aligned with the received signal. If there is a phase difference between the reference signal and the received signal, the integrator output will be lower, and symbol detection will be less accurate. The need for phase synchronization makes a coherent demodulator more complex to implement than a non-coherent demodulator. However, its performance is theoretically higher.

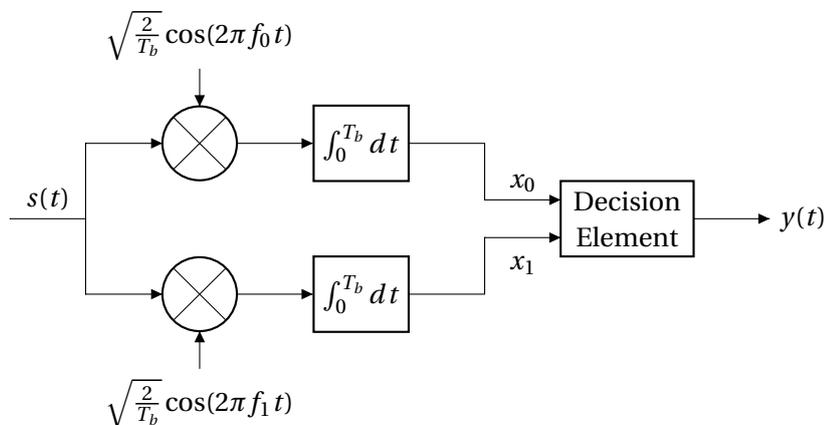


Figure 2.11: The block diagram of the coherent FSK demodulator.

2.3.2. Practical Demodulators

This section covers three examples of widely-used FSK demodulators in practice, including the frequency discriminator, the arctangent demodulator and the IQ'-I'Q demodulator (or arctangent derivative demodulator). The description of each demodulator is done similarly to the previous section.

Frequency Discriminator

In a frequency discriminator FSK demodulator, as shown in Figure 2.12, the received signal is mixed with a delayed version of itself. The delay τ_d is chosen such that a phase shift of exactly 90 degrees is achieved for the carrier frequency, or one-quarter of the carrier period. When the received carrier is modulated, this phase shift will be slightly lower or higher. Therefore, the phase shift can be related to the frequency modulation. After the multiplier, the desired signal is obtained. Additional filtering is required to remove the secondary product (at twice the carrier frequency) from the multiplication.

The sensitivity of this demodulator depends on the length of the delay that is used. As long as the quadrature condition is satisfied (i.e., 90 degrees phase shift between the two paths at the carrier frequency), the desired signal will be visible at the output. This condition is achieved for

$$\tau_d = \frac{2m-1}{4f_c}, \quad m \in \mathbb{N}, \quad (2.22)$$

where \mathbb{N} denotes the set of natural numbers (non-zero positive integers). Since longer delays are more difficult to implement accurately, this demodulator is more suited for wide-band FSK signals where a relatively small sensitivity is sufficient [5].

There are alternative methods of generating the required phase shift relationship that results in a higher sensitivity. For example, the delay line can be replaced with a resonator coupled to the input through a small capacitor. The sensitivity then depends on the quality factor of the resonator which, if large, can result in very high receiver sensitivity. In that case, the demodulator is usable for narrow-band FSK signals as well [5].

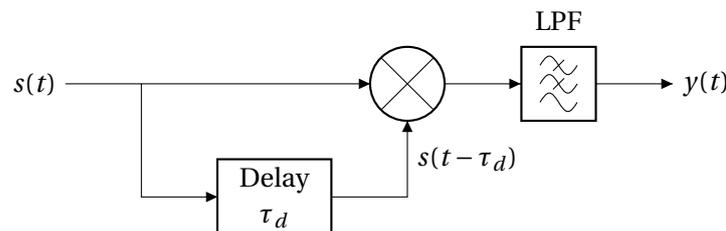


Figure 2.12: The block diagram of the frequency discriminator demodulator.

Arctangent Demodulator

Quadrature down-conversion, the conversion of a radio frequency (RF) signal to its in-phase and quadrature-phase baseband signals, is very popular in a variety of demodulator types, especially for modulation schemes such as QAM and PSK. In the demodulator shown in Figure 2.13, the quadra-

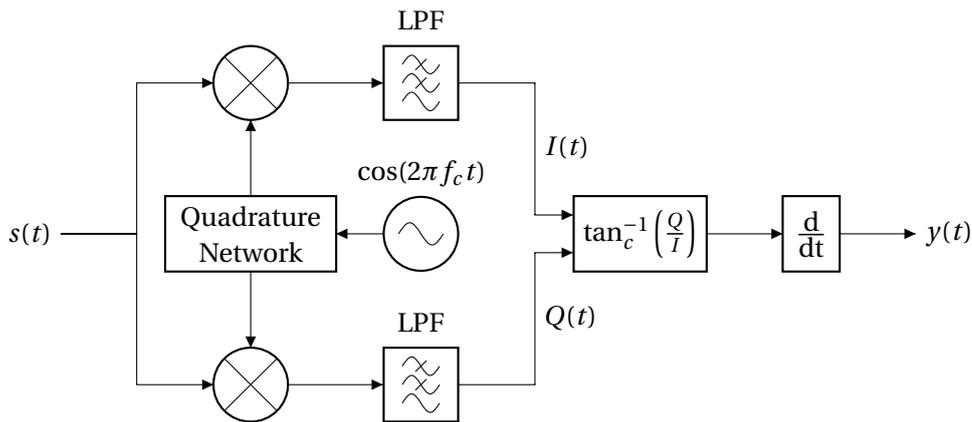


Figure 2.13: The block diagram of the arctangent demodulator.

ture down-conversion is achieved by mixing the received signal with two reference carrier signals that are in phase quadrature. After filtering out the double-frequency product, one branch then produces the in-phase baseband signal $I(t)$, while the other produces the quadrature phase baseband signal $Q(t)$. In this demodulator, the instantaneous frequency is derived from the received phase; the phase is obtained by using the continuous arctangent in (2.16), followed by (digital) differentiation to obtain the instantaneous frequency. However, this demodulator requires the continuous arctangent operation, which is non-linear and not straight-forward to implement.

IQ' - I'Q Demodulator

Instead of calculating the arctangent and taking a derivative, it is possible directly implement the mathematical derivative of the arctangent:

$$\frac{d}{dt} \left[\tan_c^{-1} \left(\frac{Q(t)}{I(t)} \right) \right] = \frac{I(t)Q'(t) - I'(t)Q(t)}{I^2(t) + Q^2(t)}. \quad (2.23)$$

This equation shows that with $I(t)$, $Q(t)$ and some operations, the modulation signal is retrieved without the use of an arctangent operation. While the time-derivatives $I'(t)$ and $Q'(t)$ are not simple to implement in the analog domain, it is relatively simple to obtain an approximate derivative through DSP techniques. The digital FSK demodulation algorithm without the arctangent function is shown in Figure 2.14 [8]. The derivative in this diagram is implemented by subtracting a delayed version of the baseband signal from the original signal. If the sampling frequency is large enough with respect to the modulation bandwidth, this approximation of the derivative is sufficiently accurate.

The scaling at the end of the receiver, and its necessary inverse operation, is not required if the received signal has a constant envelope (i.e., $I^2 + Q^2 = 1$), such as for FSK. However, any noise superimposed on $I[n]$ or $Q[n]$ disturbs the constant-envelope characteristic of FSK at the side of the receiver. Because of that, the omission of the scaling would result in distortion of the frequency modulation waveform at the output of the demodulator.

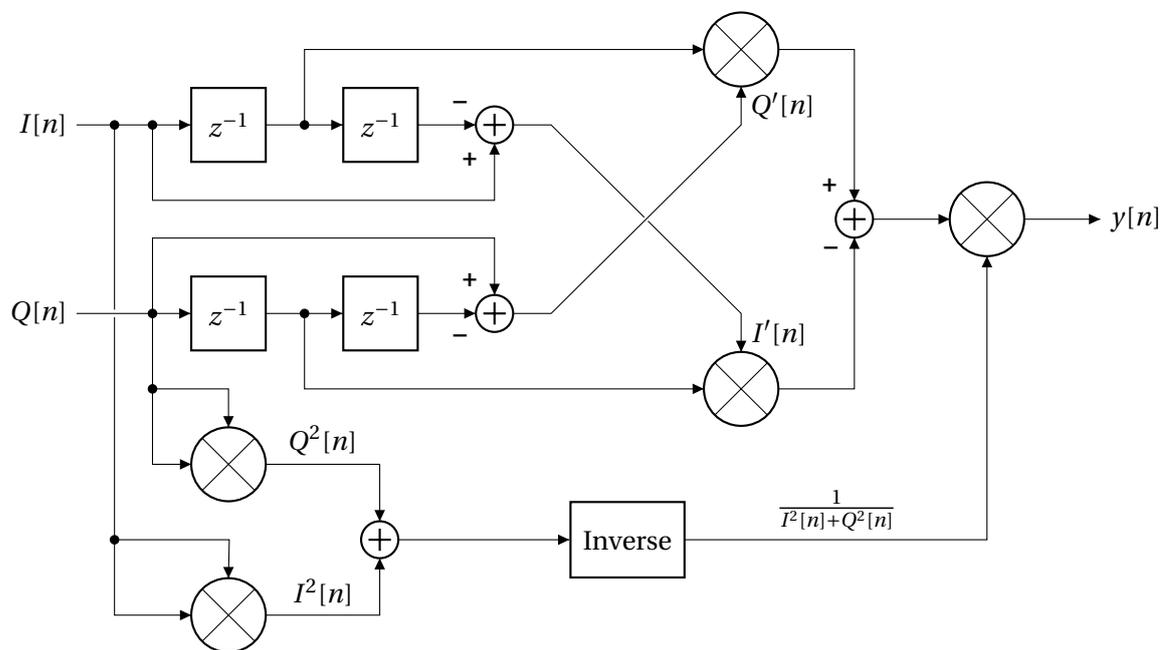


Figure 2.14: A digital frequency demodulation algorithm without an arctangent function.

3

Error Mechanisms in FSK

Any practical wireless communication system consists of three main components: a transmitter that generates the modulated signal at the desired frequency and power level; a receiver that demodulates the signal and estimates the originally transmitted information; and a channel through which the system operates. Such a channel can consist of different noise and interference sources that corrupt the transmitted signal. The most commonly considered channels in wireless communications are the Rician channel, the Rayleigh channel and an additive white Gaussian noise (AWGN) channel. The former two channels represent fading of the received signal due to multi-path reflections, with the difference between Rician or Rayleigh fading being whether there is a line-of-sight path or not, respectively. The AWGN channel originates from thermal noise in the receiver, with the noise power in an impedance-matched environment being $P_n = kTB$, or a noise power spectral density of roughly -174 dBm/Hz at room temperature. Depending on the required signal bandwidth and the modulation order, Gaussian noise can have a large impact on the receiver detection accuracy. Therefore, an AWGN channel is often applied to simulate the performance of receivers in environments dominated by thermal noise.

The focus of this chapter lies on FSK systems operating through an AWGN channel, and the error mechanisms within this boundary. First, a general introduction on noise in frequency-modulated systems is given. Then, the two dominant error mechanisms in FSK are analyzed; the FM threshold effect, which has been well-understood for over fifty years, and phase rotation cancellation, which is an error mechanism that has not been described before. Finally, the chapter concludes with a discussion on the two error mechanisms and their relationship with the modulation index.

3.1. Noise in Frequency-Modulated Systems

Before getting into the error mechanisms in FSK systems, it is important to first have an understanding of what quadrature noise is, and what its effect is on various signal parameters, such as

magnitude, phase and instantaneous frequency.

At the very input of the receiver, before any filtering takes place, the spectrum of the Gaussian noise is still flat (white). With a spectrum of theoretically infinite bandwidth, there is no limitation on how fast the noise signal $n(t)$ can change between different values. In addition, there is no limitation on the magnitude of the noise either, since the total noise power is theoretically infinite as well. Of course, this is not physically possible, which suggests that something else is happening; quantum effects limit the thermal noise spectrum at very high frequencies (in the high THz range) [9], and the bandwidth limitation imposed by any parasitic capacitance and inductance present in a receiver is typically even lower.

Despite the natural limit on the noise power spectral density, it is rarely practical to leave the received signal unfiltered. In a communication system, the transmitted signal is usually, to a certain degree, confined to a specified bandwidth. This means that anything outside of this bandwidth is considered unwanted, as it does not contribute to the energy of the desired signal. One example of an unwanted signal is thermal noise. Since it is always present, measures need to be taken to reduce the noise power and reduce the corruption of the received signal. Another important example of an unwanted signal is a blocker signal, which is a (possibly large) signal in nearby frequency bands, transmitted by other communication systems. These blockers, if left untouched, can hinder or completely prevent detection of the desired signal. While blockers are not always necessarily present at all times, they do need to be accounted for. A receiver therefore always requires a channel-select filter, passing through the desired signal (preferably untouched), and attenuating blockers and out-of-band noise. However, any noise or interference within the signal bandwidth cannot be removed; other measures need to be taken to reduce the effect of in-band noise on the receiver accuracy, such as improved detection methods or information coding [5].

3.1.1. Band-Limited Quadrature Noise

After the channel-select filter, the white Gaussian noise is band-limited. This does not only limit the noise variance, but also the rate of change of the noise waveform. In a quadrature system, the band-limited noise signal is mathematically represented as

$$n_f(t) = n_c(t) \cos(\omega_c t) - n_s(t) \sin(\omega_c t), \quad (3.1)$$

where $n_c(t)$ and $n_s(t)$ are the low-pass equivalents of two identically distributed and independent Gaussian processes. The spectrum of the filtered noise signal $n_f(t)$ is centered around the carrier frequency as a result of the multiplication of $n_c(t)$ and $n_s(t)$ with the carrier signals. Equivalently, the signal is written in polar form as

$$n_f(t) = R(t) \cos[\omega_c t + \theta(t)] = \operatorname{Re} \left\{ R(t) e^{j\theta(t)} e^{j\omega_c t} \right\}, \quad (3.2)$$

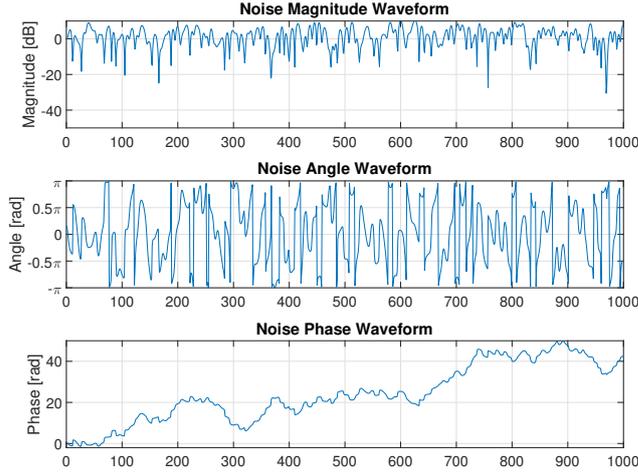


Figure 3.1: The magnitude, angle and phase waveform of band-limited noise in absence of a signal, for a noise variance of $\sigma = 1$ and a relative bandwidth of $B = 0.1f_s$.

with

$$R(t) = \sqrt{n_c^2(t) + n_s^2(t)}, \quad \text{and} \quad \theta(t) = \tan_c^{-1} \left(-\frac{n_s(t)}{n_c(t)} \right). \quad (3.3)$$

From statistics, it follows that the magnitude $R(t)$ is Rayleigh-distributed, and the angle is uniformly distributed over the range $[-\pi, \pi)$. However, the phase $\theta(t)$ – being the ‘unwrapped’ angle – has a time-dependent distribution, since its range increases over time. A larger time-frame results in an increased probability for larger phase changes with respect to the initial phase. In fact, the phase of $n_f(t)$ has some resemblance to a ‘random walk’, or Brownian motion, as it wanders up and down the phase helicoid surface erratically. An example of waveforms for magnitude, angle and phase of band-limited noise is shown in Figure 3.1.

3.1.2. Signal plus Noise

In subsection 3.1.1, band-limited noise was shown in the absence of a carrier signal. Consider now the same band-limited noise in the presence of an unmodulated carrier:

$$r(t) = s(t) + n_f(t) = [A + n_c(t)] \cos(\omega_c t) - n_s(t) \sin(\omega_c t), \quad (3.4)$$

where A is the (constant) in-phase component of the unmodulated carrier. Similarly to the previous example, the signal $r(t)$ is represented in polar form as in (3.2), with

$$R(t) = \sqrt{[A + n_c(t)]^2 + n_s^2(t)}, \quad \text{and} \quad \theta(t) = \tan_c^{-1} \left(-\frac{n_s(t)}{A + n_c(t)} \right). \quad (3.5)$$

A vector diagram of signal $r(t)$ with its Cartesian and polar components is shown in Figure 3.2.

With the addition of an unmodulated carrier, the magnitude is now Rician-distributed [10], with a

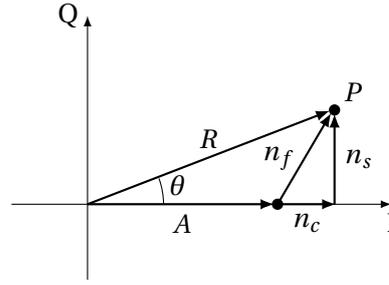


Figure 3.2: A graphical representation of signal $r(t)$ and its components. Point P represents the momentary position of $r(t)$ in the IQ plane.

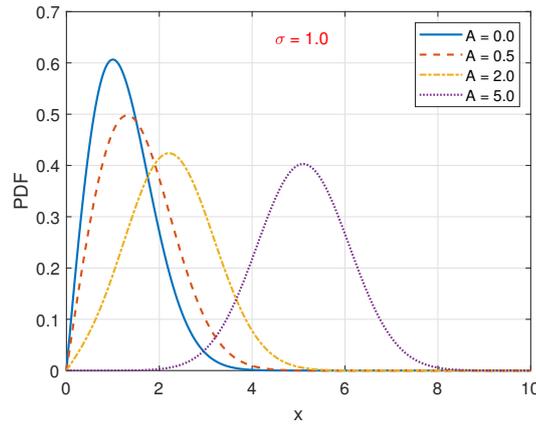


Figure 3.3: The Rice distribution for $\sigma = 1$ and $A = \{0, 0.5, 2, 5\}$.

probability density function

$$f(x|A, \sigma) = \frac{x}{\sigma^2} \exp\left[-\frac{(x^2 + A^2)}{2\sigma^2}\right] I_0\left(\frac{Ax}{\sigma^2}\right), \quad (3.6)$$

where $I_0(x)$ is the zero-order modified Bessel function of the first kind, σ is the standard deviation of the noise, and A is the signal mean. The probability density function (PDF) of the Rice distribution is shown in Figure 3.3 for multiple values of A . For $A = 0$, the Rice distribution defaults to the Rayleigh distribution. For $\sigma \ll A$ and $x \ll A$, the Rice distribution approximately takes the shape of a Gaussian distribution with mean A , which is derived using Hankel's expansion of the modified Bessel function for large arguments [11].

The angle and phase distribution are more involved; as the unmodulated carrier amplitude A increases, the distribution of both the angle and the phase start to approximate a Gaussian distribution. This conclusion is obtained by approximating the phase function by its first-order Taylor approximation, and making the assumption that A is, at all times, much larger than $n_c(t)$. However, it is when the amplitude A is smaller than the excursions caused by $n_f(t)$ that the phase waveform starts to behave interestingly. At first, the phase of $r(t)$ remains centered around the mean phase of the unmodulated carrier, which is zero in this case. However, with finite probability, the noise components $n_c(t)$ and $n_s(t)$ simultaneously have the right excursions in such a way that the origin of the

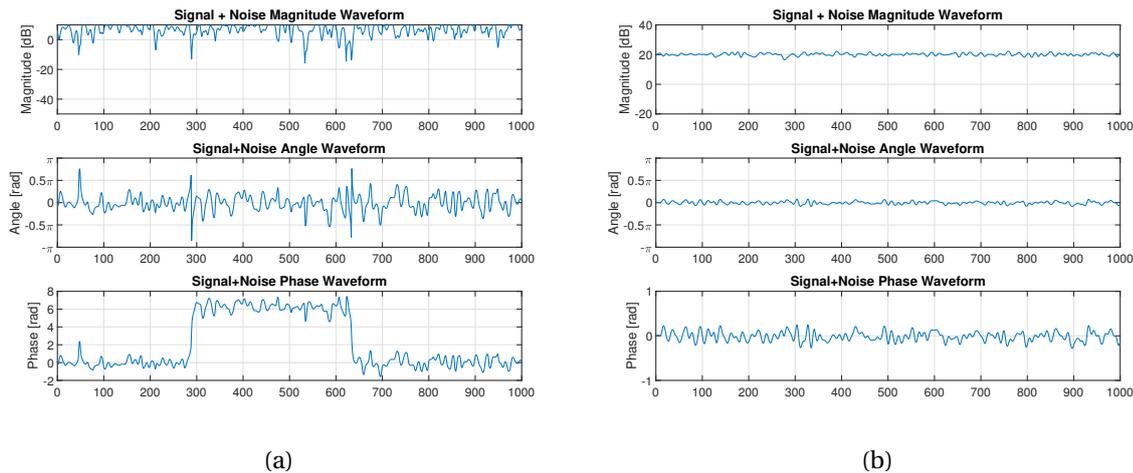


Figure 3.4: The magnitude, angle and phase waveform of an unmodulated carrier plus band-limited noise for a noise variance of $\sigma = 1$ and a relative bandwidth of $B = 0.1f_c$: (a) $SNR = 2$, and (b) $SNR = 10$.

IQ plane is encircled in a clockwise or counter-clockwise manner. In the angle waveform, such an event appears as a rapid discontinuity from $-\pi$ to π or vice versa, as expected from Figure 2.7a, after which the angle waveform returns to values centered around the original mean value. For the phase waveform, however, there is no such discontinuity, and this event has caused the signal phase mean to shift by $\pm 2\pi$. Examples of a magnitude, angle and phase waveform of an unmodulated carrier plus band-limited noise are shown in Figures 3.4a and 3.4b for $SNR = 2$ and $SNR = 10$, respectively. For $SNR = 2$, the signal mean is small enough relative to the noise variance for this phenomenon to occur, while for $SNR = 10$ it does not occur at all within this particular time-window.

In frequency-modulated systems, such a phase shift of 2π has large consequences for the demodulated signal. Typically, these 2π shifts happen very quickly, due to the proximity of the signal plus noise trajectory to the origin, and these events appear to be much faster than the input bandwidth theoretically would allow. As a consequence, at the output of an FSK receiver, large spikes or impulses of area 2π are present where such shifts in phase occur, corrupting the desired signal in an extreme manner [12]. Contrarily, if the SNR is very large, then the noise only makes small excursions in the phase waveform, and therefore also in the instantaneous frequency waveform. This is a result of the non-linear transfer from the input of the demodulator to the output, which is further discussed in Section 3.2.

3.1.3. Clicks and Doublets

The phase waveform of a signal plus AWGN drastically differs in behavior when varying the input SNR. Based on the possible trajectories that the unmodulated carrier plus noise can make through the IQ plane, four different categories are identified from the phase waveform behavior. A simplified signal plus noise trajectory that goes through all four categories is shown in Figure 3.5.

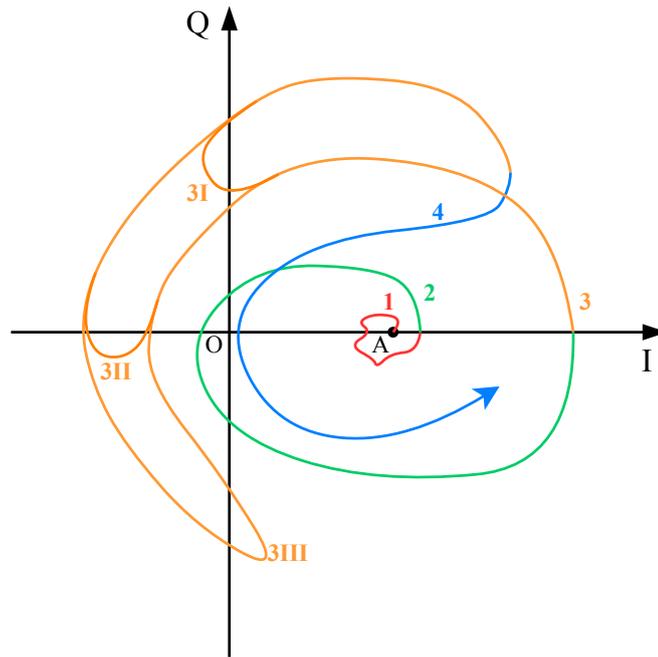


Figure 3.5: A simplified signal plus noise trajectory, starting from signal mean A . The trajectory consists of four sections representing different behavior: (1) Gaussian-like, (2) a click, (3) false clicks (type I, II or III) and (4) a doublet.

Gaussian-like The first category, indicated in red, results in Gaussian-like behavior of the phase waveform. This occurs only when the noise excursions are small relative to the signal mean A . For large SNR, this is usually the only behavior present in the phase waveform.

Click The second section of the curve, in green, shows a signal plus noise trajectory where the IQ plane origin is encircled, resulting in a (rapid) 2π shift in the phase waveform. This category is given the name 'click', originating from almost a century ago when Armstrong noticed that FM radio receivers produces a 'clicking' sound if the input SNR is too low. As discussed before, a click event results in a spike of area 2π at the demodulator output, producing the clicking sound. Because of the resemblance between a click event and the Dirac delta (or impulse) function, the PSD of a click event at the demodulator output has a low-pass shape, and the 'flatness' of the PSD depends on the click duration. In Figure 3.6, a simplified click event is shown as a phase waveform, an instantaneous frequency waveform and its PSD at the demodulator output.

False Click The third section, in orange, shows false clicks of type I, II or III. Such a trajectory initially appears as if it encircles the origin, but then returns to the proximity of the signal mean without doing so, resulting in a net phase shift of zero. Consequently, there will be little low-frequency energy in its PSD at the demodulator output. Due to the complexity of a false click trajectory relative to other categories, it is improbable for false clicks to occur.

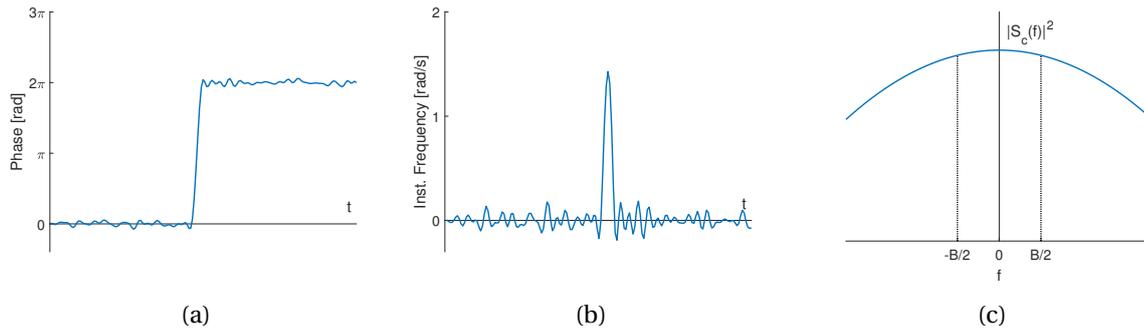


Figure 3.6: A click event representation: (a) phase waveform, (b) instantaneous frequency waveform, and (c) approximate output PSD.

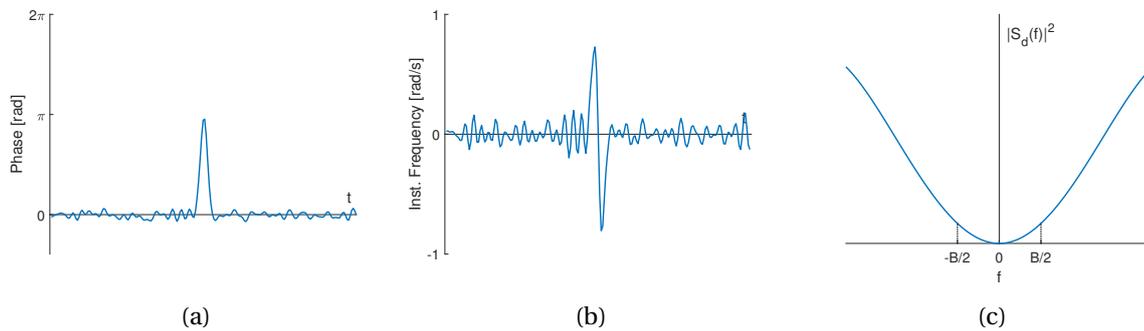


Figure 3.7: A doublet event representation: (a) phase waveform, (b) instantaneous frequency waveform, and (c) approximate output power spectral density.

Doublet The final section, marked in blue, shows a trajectory that closely approaches the IQ plane origin without encircling it, producing a rapid phase change in a short time, but with ultimately zero phase shift. Similarly to false clicks, the doublet PSD at the demodulator output has relatively little low-frequency content. In Figure 3.7, a simplified doublet event is shown as a phase waveform, an instantaneous frequency waveform and its output power spectral density. For the purpose of the phase waveform and instantaneous waveform, doublets and false clicks are considered the same category. The reason behind the distinction between the two events is explained in subsection 3.2.2.

Based on the instantaneous frequency behavior of the different categories, several conclusions are drawn. It appears that the non-linear transfer from receiver input to demodulator output dramatically enhances the noise power due to click and doublet events. This noise enhancement effect occurs when the input SNR is small enough for these events to occur, and is referred to as the *FM threshold effect*, which effectively further reduces the output SNR of an FM demodulator at these lower SNR values. From the three categories discussed in subsection 3.1.3 (excluding the Gaussian-like component), clicks appear to have the most low-frequency content in its spectral density, resulting in a significant increase of the in-band noise at the output. In the next section, the impact of clicks on the output SNR is further examined.

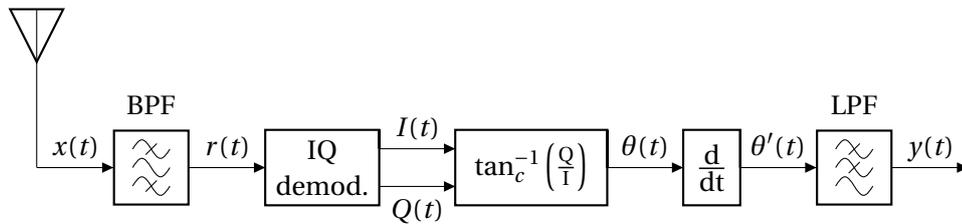


Figure 3.8: A block diagram of an 'ideal' FM receiver.

3.2. FM Threshold Effect

In this section, a model of the output SNR of an FM receiver is developed, in order to get a good understanding of how input noise affects the signal at the output of the demodulator. The block diagram of an 'ideal' FM receiver is shown in Figure 3.8. For simplicity, it is assumed that all components are ideal and do not contribute to the noise at the input of the receiver. The output SNR model is developed in two steps; first, an analysis without any click events is done, followed by an expansion of the model with the concept of click events. The model is developed under the assumption of analog FM, since it is based on the analysis done by Rice in [12]. However, some conclusions can still be drawn for FSK systems, which are included at the end of the section.

3.2.1. Ideal SNR Model

First, a model for the output SNR is developed in the absence of click events. The model is constructed in a step-by-step approach by moving through the receiver chain shown in Figure 3.8. For simplicity, the received signal is assumed to be an unmodulated carrier plus AWGN. Before filtering, the Gaussian noise has a flat, double-sided power spectral density of $w(f) = N_0/2$. Assuming the use of an ideal rectangular (brick-wall) filter, the power spectrum of the *filtered* noise is band-limited to a bandwidth B , centered around the carrier frequency f_c . After the bandpass filter, the signal of (3.4) is obtained:

$$r(t) = s(t) + n_f(t) = [A + n_c(t)] \cos(\omega_c t) - n_s(t) \sin(\omega_c t).$$

From the power spectral density of $n_f(t)$, it follows that $n_c(t)$ and $n_s(t)$ both have a power spectral density equal to $2w(f + f_c)$. With the well-defined equivalent noise bandwidth, the carrier-to-noise ratio (CNR) (or input SNR) is

$$\rho = \frac{\overline{s^2(t)}}{\overline{n_f^2(t)}} = \frac{A^2}{2BN_0}, \quad (3.7)$$

where $\overline{n_f^2(t)} = BN_0$ is the total noise power in $n_f(t)$. Next, the in-phase and quadrature components of the received signal $r(t)$ are obtained from quadrature demodulation as

$$I(t) = A + n_c(t), \quad \text{and} \quad Q(t) = -n_s(t). \quad (3.8)$$

Using the continuous arctangent function of (2.16), the phase is then

$$\theta(t) = \tan_c^{-1} \left(-\frac{n_s(t)}{A + n_c(t)} \right). \quad (3.9)$$

At this point, an assumption is made to simplify the analysis; if the input signal-to-noise ratio is large, then the carrier amplitude A will nearly always be much larger than variations caused by $n_c(t)$ and $n_s(t)$. As a result, the argument of the arctangent function will be relatively small. In that case, the phase $\theta(t)$ is reasonably approximated by the first-order Taylor expansion of the arctangent

$$\theta(t) \approx -\frac{n_s(t)}{A}. \quad (3.10)$$

Differentiation of the phase with respect to time gives the instantaneous angular frequency:

$$\theta'(t) \approx -\frac{n_s'(t)}{A}, \quad (3.11)$$

where the prime symbol denotes a derivative with respect to time. The PSD of (3.11) is of particular interest, as this is the signal plus noise PSD at the demodulator output (prior to low-pass filtering). If the power spectrum of $n_s(t)$ is $2w(f + f_c)$, then the power spectrum of $\theta(t)$ is $2w(f + f_c)/A^2$. Using the Fourier properties, the power spectrum of $\theta'(t)$, defined as $W(f)$, is $(i2\pi f)^2$ times the power spectrum of $\theta(t)$, resulting in

$$W(f) \approx \frac{2w(f_c + f)}{A^2} (2\pi f)^2 = \frac{8\pi^2 w(f_c + f)}{A^2} f^2. \quad (3.12)$$

Finally, a low-pass filter operation is performed on the signal in order to further limit the bandwidth to that of the modulation signal. Again assuming a rectangular filter, the noise power at the output is calculated using

$$N_{out} = 2 \int_0^{f_a} W(f) df, \quad (3.13)$$

where f_a is the modulation signal bandwidth. If $f_a < B$, which is the case for wide-band FM signals, then the total output noise power is

$$N_{out} \approx \frac{8\pi^2 w_0 f_a^3}{3A^2} = \frac{4\pi^2 f_a^3}{3\rho B}. \quad (3.14)$$

In order to evaluate the output signal-to-noise ratio, a modulation signal needs to be present at the output. However, so far the calculations have been done with the assumption of an unmodulated carrier, which theoretically does not contribute to any AC power at the demodulator output. Nevertheless, if a hypothetical sinusoidal modulation signal $\phi'(t) = A_m \sin(2\pi f_m t)$ is assumed with $f_m < f_a$, then this signal would pass through the receiver chain unaltered, under the condition that no noise is present. Since the calculation so far already assumes a large input CNR, this assumption is sufficient for this particular analysis. The modulation signal $\phi'(t)$ has an output signal power of

$$S_{out} = \frac{A_m^2}{2}. \quad (3.15)$$

In order to achieve the maximum possible output SNR, $A_m = 2\pi(B/2)$ is taken, such that the input signal swings across the entire input bandwidth. Substituting A_m in (3.15) results in

$$S_{out} = \frac{\pi^2 B^2}{2}, \quad (3.16)$$

and the output SNR is obtained by dividing (3.16) with (3.14):

$$\frac{S_{out}}{N_{out}} \approx 3 \left(\frac{B}{2f_a} \right)^3 \rho. \quad (3.17)$$

From (3.17), it is seen that the output SNR is linearly proportional to the input SNR. In addition, the bandwidth ratio $B/2f_a$ in (3.17) shows the noise resilience of wide-band FM receivers; for $B > 2f_a$, the output SNR improves dramatically. Such an effect is not seen in other modulation schemes. However, the output SNR of (3.17) is only valid as long as the main assumption holds; the input CNR must be large. If the input SNR is not large enough, non-linear effects such as clicks and doublets occur, which dramatically corrupt the signal at the demodulator output.

3.2.2. Rice SNR Model

The expression in (3.17) is not sufficiently accurate below the FM threshold primarily due to the occurrence of click events. A better approximation is achieved by considering the consequences of the occurrence of a click event; a 2π area impulse-like signal at the output of the demodulator, which appears as an increased in-band noise energy at the demodulator output. The output noise is then no longer Gaussian-like, but rather a superposition of a Gaussian-like component and an impulse-like component. In order to account for this, the power spectrum of the instantaneous frequency signal $\theta'(t)$ must be expanded to include the effect of click events. This derivation is done by Rice in [12], and is summarized below.

Impulse Noise

Consider the instantaneous frequency $\theta'(t)$ to consist of the summation of two independent signals $\theta'_1(t)$ and $\theta'_2(t)$, where $\theta'_1(t)$ is the Gaussian-like component and $\theta'_2(t)$ is the impulse-like component. Under the assumption that click events are independent from each other, $\theta'_2(t)$ is written as an impulse train of magnitude 2π for both positive (counter-clockwise) and negative (clockwise) clicks:

$$\theta'(t) = \theta'_1(t) + \theta'_2(t) = \frac{n'_s(t)}{Q} + \sum_{p=-\infty}^{\infty} 2\pi\delta(t-t_p) + \sum_{n=-\infty}^{\infty} (-2\pi)\delta(t-t_n), \quad (3.18)$$

where t_p and t_n denote the moments in time at which positive and negative click events occur, respectively. The DC component of $\theta'_2(t)$ is then $2\pi(N_+ - N_-)$, where N_+ and N_- denote the positive and negative click rate in Hz, respectively, which results in the power spectrum

$$W(0) = 4\pi^2(N_+ + N_-). \quad (3.19)$$

Under the assumption that the click events are sufficiently well approximated by a Dirac function, the power spectrum of (3.19) is in fact approximately flat over the entire demodulator output bandwidth. With the addition of $\theta_2'(t)$, the total power spectrum of the instantaneous frequency is now

$$W(f) \approx 4\pi^2(N_+ + N_-) + \frac{2w(f_c + f)}{Q^2}(2\pi f)^2. \quad (3.20)$$

Unmodulated Carrier

In order to calculate the noise PSD at the demodulator output from (3.20), the positive and negative click rates must be determined. Unfortunately, this task is neither straightforward nor trivial. The complex derivation has been done by Rice in [12], and its results are included in this section. In this derivation, Rice assumes that any signal passing through an odd multiple of π will result in a click event. In practice, this results in an overestimation of the click rate, since false clicks of type II and III meet this criterion without actually encircling the IQ plane origin. A more accurate analysis that takes into account false clicks is done in [13]. However, since this is only a demonstration of the FM threshold effect, the work done by Rice is sufficient for this section. For an unmodulated carrier, the click rate is calculated as

$$N_+ = N_- = \frac{r}{2} \operatorname{erfc}(\sqrt{\rho}), \quad (3.21)$$

where $\operatorname{erfc}(x)$ is the complementary error function, and

$$\begin{aligned} r &= \frac{1}{2\pi} \sqrt{\frac{b_2}{b_0}}, \\ b_0 &= \overline{n_f^2(t)} = \int_0^\infty w(f) df = \int_0^\infty 2w(f_c + f) df, \\ b_2 &= (2\pi)^2 \int_0^\infty (f - f_c)^2 w(f) df. \end{aligned} \quad (3.22)$$

If the power spectrum $w(f)$ has a rectangular shape with bandwidth B , then it follows that $r = B/\sqrt{12}$. After substitution of (3.21) in (3.20), the total output noise power N_{out} is calculated with the expression in (3.13). Similarly to subsection 3.2.1, the demodulator output SNR is calculated by dividing the S_{out} by N_{out} , which results in

$$\frac{S_{out}}{N_{out}} \approx \frac{3\rho B^3 (2f_a)^{-3}}{\rho\sqrt{3} \operatorname{erfc}(\sqrt{\rho}) (B/f_a)^2 + 1}. \quad (3.23)$$

Several observations are made from (3.23). First, if the SNR is large, then the first term in the denominator approaches zero, and the output SNR defaults back to that of (3.17). However, as the SNR decreases, the rate at which click events occur increases, resulting in a noise-enhancement

effect that degrades the output SNR. Next, if the input bandwidth is increased, then more noise is admitted to the system. For a larger input bandwidth, the input CNR degrades and the click rate increases. As a result, the threshold effect appears at a larger input CNR. Finally, it is noted that for an extremely low input SNR (below 2 dB), the model in (3.23) does not accurately predict the output SNR, since the assumption that click events appear as Dirac functions at the demodulator output no longer holds.

Modulated Carrier

A more practical and realistic model is obtained when considering the consequences of a modulated signal on the demodulator output SNR. The variation in click rates as a function of modulation is of particular interest. For a constant modulation signal $\phi'(t) = 2\pi f_0$, the positive and negative click rates are determined by Rice in [12] as

$$N_+ = \frac{1}{2} \left[\sqrt{r^2 + f_0^2} \operatorname{erfc} \left(\sqrt{\rho + \frac{\rho f_0^2}{r^2}} \right) - f_0 e^{-\rho} \operatorname{erfc} \left(\frac{\sqrt{\rho} f_0}{r} \right) \right], \quad (3.24)$$

$$N_- = N_+ + f_0 e^{-\rho}. \quad (3.25)$$

It is noted that substitution of $f_0 = 0$ in (3.24) results in the original expression for an unmodulated carrier, given in (3.21). However, when f_0 is non-zero, there is an unbalance in positive and negative click rates. Intuitively, this makes sense; if the signal mean is rotating around the IQ plane origin, a click event of opposite polarity to that of the modulation signal rotation has a simpler trajectory than one with the same polarity. Because of that, the probability for a click event to occur in opposite direction of the modulation signal rotation is larger.

The output SNR for a modulated carrier is calculated in a similar fashion as for the unmodulated carrier. Since the modulation signal is constant, the signal power is now $S_{out} = \pi^2 B^2$, resulting in an output SNR of

$$\frac{S_{out}}{N_{out}} \approx \frac{6\rho B^3 (2f_a)^{-3}}{12B f_a^{-2} \rho (N_+ + N_-) + 1}. \quad (3.26)$$

The output SNR models of (3.17), (3.23) and (3.26) are graphically shown in Figure 3.9, where a deviation ratio $B/(2f_a) = 5$ is used to obtain the curves. The 'breaking point' or 'threshold' of the FM threshold effect is visible around $SNR = 10$ dB, and appears to be worse when carrier modulation is taken into account.

3.2.3. Consequences for FSK Systems

While the focus of the previous analyses in this chapter lies primarily on analog FM, some conclusions can also be drawn for FSK systems. Ultimately, even though the information signal in DWC systems is discrete in time and amplitude, the transmitted wave must be analog. Therefore, click

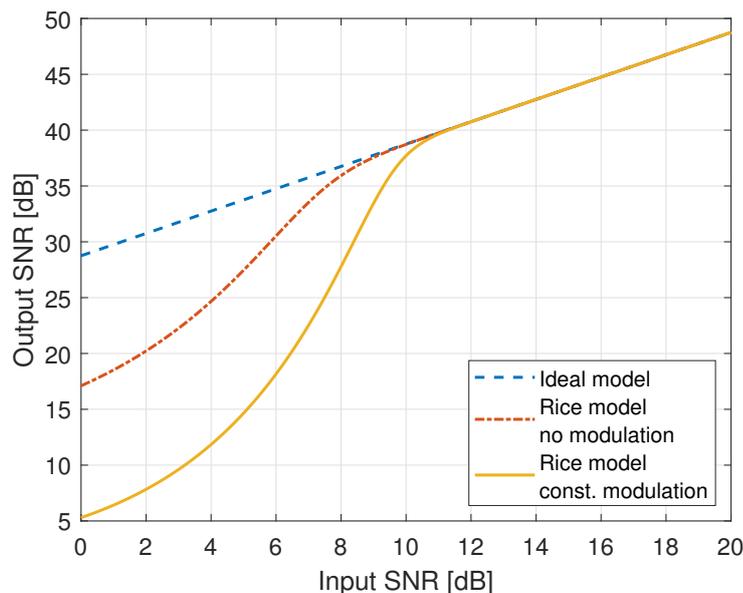


Figure 3.9: The output SNR curves as a function of the input SNR of the ideal model and the Rice model, for a deviation ratio of $B/(2f_a) = 5$.

events occur in FSK receivers in exactly the same manner as in FM receivers. The only difference is in the way that click events affect the output of the demodulator. In Chapter 2, it is shown that the modulation index determines the total phase rotation per symbol time, which is $\Delta\phi = \pm h\pi$ for unfiltered BFSK systems. Furthermore, in subsection 3.2.2, it is mentioned that click events have a larger probability of occurring in the opposite direction to that of the modulation rotation. In that case, particularly in narrow-band systems, a click event completely overwhelms the phase rotation within a single symbol period, resulting in a detection error of that particular symbol at the demodulator output. While wide-band FSK systems, in general, have a higher tolerance for click events due to their larger phase rotation per symbol, they also require a larger receiver bandwidth, which in turn increases the click rate. Finally, as the input CNR increases past the breaking point of the threshold effect, the click rates drop dramatically, leaving only other error mechanisms that affect FSK systems. The effect of click events on the BER performance as a function of the modulation index is included in the discussion at the end of this chapter.

3.3. Phase Rotation Cancellation

In this section, phase rotation cancellation is introduced, and is found to be a dominant error mechanism in narrow-band FSK systems. First, the concept of the error mechanism is explained, including parameters that play an important role in the mechanism. Following the concept of phase rotation cancellation, the idea of 'noise frequency' is introduced.

3.3.1. Concept

In narrow-band FSK systems, the total phase rotation per symbol is relatively small. As a consequence, the output instantaneous frequency signal is more easily disturbed by the presence of noise. In the case where the *rotation* of the noise completely cancels or even overwhelms the rotation of the signal over a symbol period, a detection error is made at the demodulator output. The signal plus noise trajectory within that particular symbol period is then almost entirely radial, with (typically) nearly no change in phase. This result is something that was discovered after highlighting the signal plus noise trajectories of incorrectly detected symbols in the IQ plane. For narrow-band FSK systems (e.g., $h = 0.5$), almost none of the highlighted trajectories were click events, and were instead trajectories with tiny phase rotations.

An example of this error mechanism, phase rotation cancellation (PRC), is shown in Figure 3.10. A GFSK signal with $h = 0.5$ and $BT = 0.5$ is used for the demonstration. In Figure 3.10, samples that correspond to an incorrectly detected symbol are highlighted in red. From Figure 3.10, several observations are made. PRC errors seem to most likely occur near the magnitude of the signal mean, contrarily to click events that typically pass by the origin closely. Next, while a click event results in a large phase change within a symbol period, PRC events have an almost zero net phase change due to the noise rotation canceling the signal rotation. While this is partly due to the small modulation index, it is also a consequence of pre-modulation filtering. As explained in Section 2.2, pre-modulation filtering reduces the total average phase rotation per symbol. The reduction is particularly prevalent for alternating symbol patterns ('010' or '101' in BFSK systems), where a Gaussian filter with $BT = 0.5$ results in an almost 40% reduction in phase rotation within that central symbol period. This significant decrease in dynamic range results in an increase in the probability of PRC errors to occur. It is then no surprise that nearly all PRC errors happen specifically during alternating symbol patterns.

3.3.2. Noise Frequency

In FSK systems, the demodulator extracts the rotation of the received signal (plus noise) vector in the IQ plane. In BFSK, an error occurs if the instantaneous frequency throughout a symbol period has a polarity opposite to that of the information signal. To determine the BER in FSK systems, it is important to consider the transfer from the AWGN at the input to the demodulator output. Since the transmitted information is in the *rotation* of the received signal phasor in the IQ plane, only the tangential velocity component of the noise is important. At the demodulator output, this tangential component, from now on referred to as (instantaneous) 'noise frequency', is superimposed on the instantaneous frequency deviation caused by the information signal. Since the noise is additive in the IQ plane, its phase excursions depend on the radial distance of the signal mean to the origin. This transfer from AWGN in the IQ plane to noise frequency at the demodulator output is highly non-linear, in particular for small signal mean, as expected from the analysis in Section 3.1. Therefore, the noise frequency distribution is expected to be not only a function of the noise variance, but

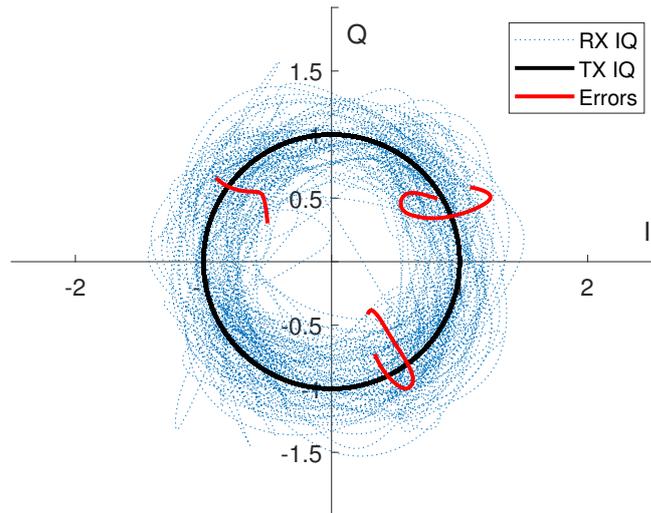


Figure 3.10: A vector diagram of an FSK modulated signal ($h = 0.5$) with and without AWGN channel. Highlighted in red are signal plus noise trajectories where a symbol error occurred.

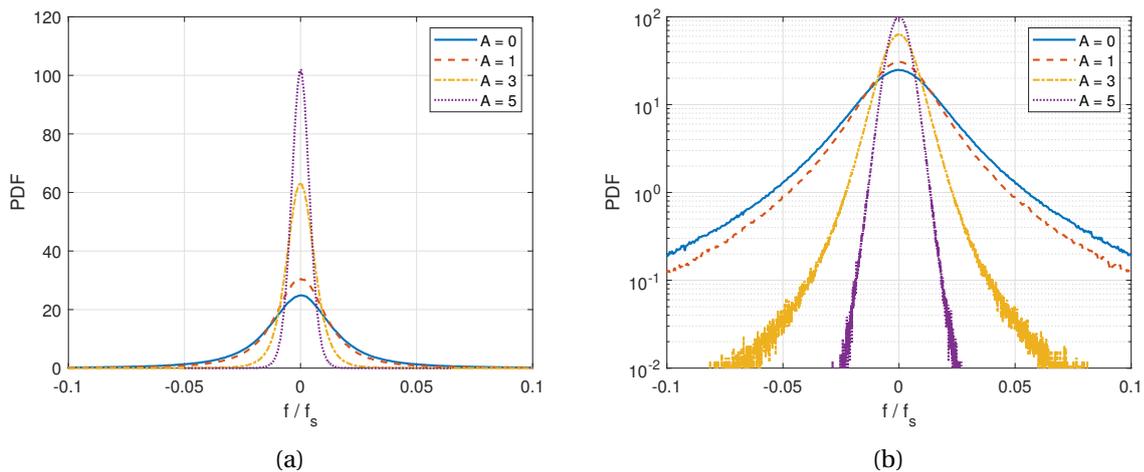


Figure 3.11: The (normalized) instantaneous frequency distribution for $\sigma = 1$ and $A = \{0, 1, 3, 5\}$: (a) linear scale and (b) logarithmic scale.

also of the received signal mean A . The instantaneous frequency distribution of an unmodulated carrier plus band-limited noise is shown in Figure 3.11 for multiple values of A .

From Figure 3.11, it is clear that the distribution is not at all Gaussian-shaped for small values of A ; the decay of the distribution is relatively slow and mostly has a positive curvature in the logarithmic scale, resulting in a wide PDF, while a Gaussian curve in the logarithmic scale has a strictly negative curvature. When the signal plus noise trajectory can get arbitrarily close to the IQ plane origin, there is theoretically no bound on the instantaneous frequency waveform. Therefore, when increasing A , the shape of the instantaneous frequency PDF becomes narrower as the signal plus noise trajectory is less likely to be in close proximity to the origin with respect to the case $A = 0$.

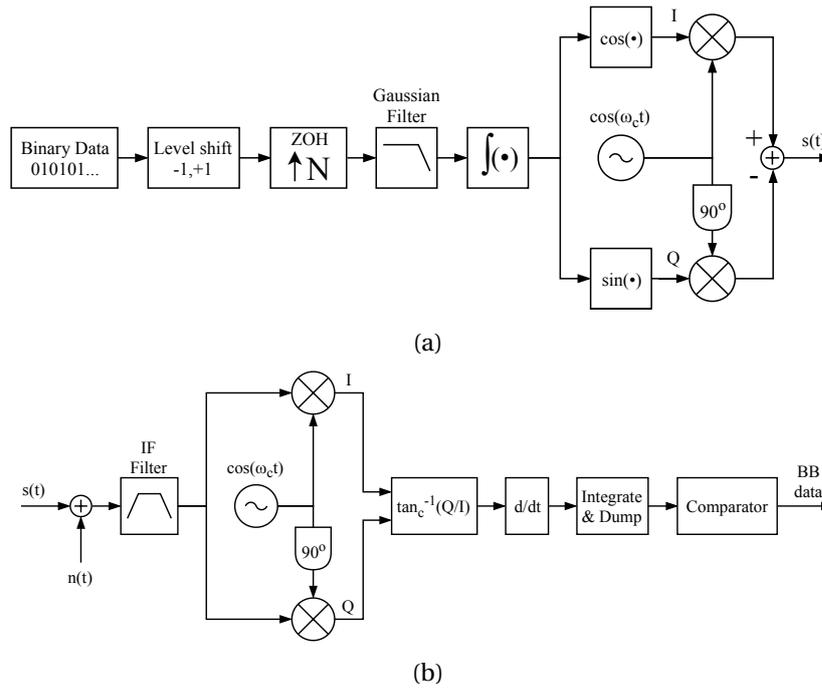


Figure 3.12: A GFSK transceiver with a 4th-order Butterworth IF filter: (a) transmitter, (b) receiver.

Furthermore, the instantaneous frequency PDF is approximately Gaussian for large signal mean, similar to the Rice distribution for the magnitude. This approximation is intuitively obtained without any distribution function by assuming a large signal magnitude, such that the phase deviation due to noise perturbations in the IQ plane is small. When the phase deviation is reasonably approximated by a tangent line of a circle (with a radius equal to the signal magnitude), then the bi-variate Gaussian distribution obtained from noise components $n_c(t)$ and $n_s(t)$ maps approximately linearly to the phase. As a result, the distribution of phase and instantaneous frequency both become approximately Gaussian.

At this time, there is no statistical model for the distribution of noise frequency at moderate or low SNR. Such a distribution could help determine the bit error rate accurately for FSK systems with small modulation indices. However, the derivation of this distribution is outside of the scope of this thesis.

3.4. Simulation Results

This chapter has demonstrated and discussed the two dominant error mechanisms in FSK systems. Click events, which result from the non-linear transfer from the input of the demodulator to the output, are mostly prevalent in wide-band FSK, while PRC errors are dominant in narrow-band FSK systems. In order to demonstrate the relation between the two error mechanisms and the modulation index, a GFSK transceiver is simulated while distinguishing errors caused by the two error mechanisms. The simulated GFSK transceiver is shown in Figure 3.12. The simulation results for $h = 0.5$ and $h = 1.0$ are shown in Figures 3.13 and 3.14, respectively.

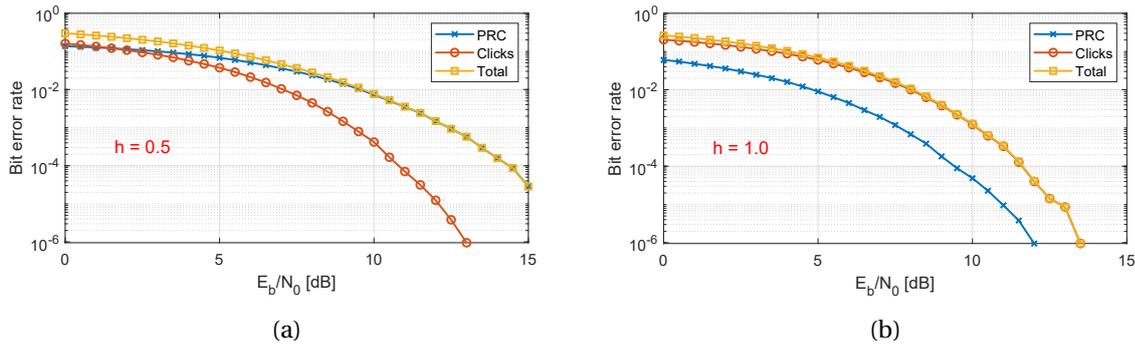


Figure 3.13: The bit error rate curves as a result of click errors and PRC errors for (a) $h = 0.5$ and (b) $h = 1.0$.

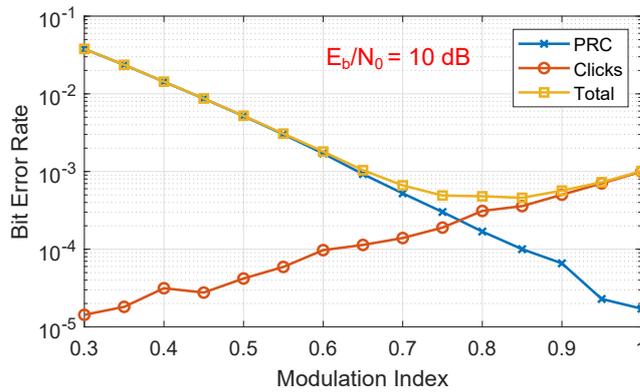


Figure 3.14: The bit error rate as a function of the modulation index h for click errors, PRC errors and the total errors, for $E_b/N_0 = 10$ dB.

It is observed that changing the modulation index drastically changes the relative importance of click errors and PRC errors. For $h = 0.5$, PRC errors are dominant for approximately $E_b/N_0 > 3$ dB. For even lower IBEND, the noise excursions with respect to the modulation signal are so large that click events frequently disturb the detection of the signal, resulting in many errors. Contrarily, for $h = 1.0$, barely any PRC errors occur with respect to the total number of errors. The intended signal rotation is large enough such that it is considerably more likely that click events result in errors rather than PRC events. These two results suggest that there is some modulation index which, for a fixed IBEND, results in an even distribution between PRC errors and click errors. Figure 3.14 shows a simulation result of simulating the same transceiver at a fixed IBEND value of $E_b/N_0 = 10$ dB for multiple values of h . According to Carson’s rule, the bandwidth of narrow-band FSK signals is dominated by the data rate. However, as the modulation index increases, so does the influence of the frequency deviation in the bandwidth of the FSK signal. This is taken into account by taking the IF bandwidth that minimizes the number of errors at the output for each value of h . The optimal receiver bandwidth turned out to be approximately increasing linearly from 1 times the data rate at $h = 0.3$ to 1.3 times the data rate at $h = 1.0$.

From Figures 3.13 and 3.14, it is evident that the modulation index plays a significant role in the trade-off between error mechanisms. For small values of h , the dynamic range at the demodulator

output is small, and most errors occur due to cancellation of the modulation signal rotation by the noise rotation. This is particularly true when pre-modulation filtering is applied, since doing so further reduces the dynamic range for specific local symbol sequences. With large modulation indices, modulation signal rotation is too large for most PRC errors occur. Instead, the noise variance allows the signal to encircle the IQ plane origin with a higher probability than actually canceling the signal rotation within the same helicoid turn. The probability of click errors is further increased due to the increase in required receiver bandwidth for larger values of h .

It is important to note here that phase rotation cancellation and click events are difficult to distinguish once the modulation index exceeds $h = 2$, which is when a full helicoid turn is traversed during a single symbol period. However, there is no application in practice that utilizes such wide-band FSK modulation. Nearly all modern applications of FSK use a small modulation index, for which the dominant error mechanism is phase rotation cancellation, and not click events. Any improvement in the demodulation of FSK signals with small h should therefore first be sought in the elimination of PRC errors.

A scientific paper is written on the topic of 'FSK demodulation errors from phase rotation cancellation', which is included in Appendix C.

4

Phase-based FSK Demodulation Algorithm

In Chapter 3, it has been concluded that there are two dominant error mechanisms in FSK systems. The modulation index h controls the relative importance of the error mechanisms. Since most modern FSK applications use narrow-band FSK signals (i.e., $h < 1$), such as GSM and Bluetooth, these systems are dominated by phase rotation cancellation errors when an AWGN channel is considered. In this chapter, a new FSK demodulation algorithm is developed that takes into account the characteristics of PRC errors in an effort to improve FSK system performance.

First, an overview is given of the system specifications used for the algorithm and simulation purposes. Next, the new demodulation algorithm is introduced and explained. After that, some non-idealities are introduced, and methods to solve these issues are proposed. Finally, at the end of this chapter, some alternative algorithms are described that have been considered throughout the project, but have not been thoroughly explored nor implemented in the final system.

4.1. System Specifications

In order to obtain practical and realistic results of the demodulation algorithm performance, the specifications that are used in simulation and testing are taken from a widely used modern application. Parameters that closely resemble the operation of Bluetooth Low Energy, one of the major present applications of FSK, are listed in Table 4.1. With a modulation index of $h = 0.5$ and a Gaussian pre-modulation filter with $BT = 0.5$, the maximum phase rotation per symbol is $\Delta\phi_{sym} = \pi/2$. The carrier frequency does not necessarily play a direct role in the baseband FSK signal, but becomes important in subsection 4.3.2 when considering the consequence of frequency drift. For the on-air data rate, $R_b = 1$ Mbit/s is taken, which results in a symbol period of $1 \mu\text{s}$. A sampling frequency of $f_s = 10$ MHz then means that each symbol is up-sampled by a factor ten. Finally, the packet size is set to 400 symbols. While the packet size is not necessarily a fixed value, as illustrated

Table 4.1: The system specifications used in the demodulation algorithm design, drawn from BLE.

Parameter	Value
Modulation index, h	0.5
TX filtering	Gaussian, BT = 0.5
Carrier frequency, f_c	2402 - 2480 MHz
Data rate, R_b	1 Mbit/s
Sampling frequency, f_s	10 MHz
Packet size, N_{sym}	400 bits

BLE Packet

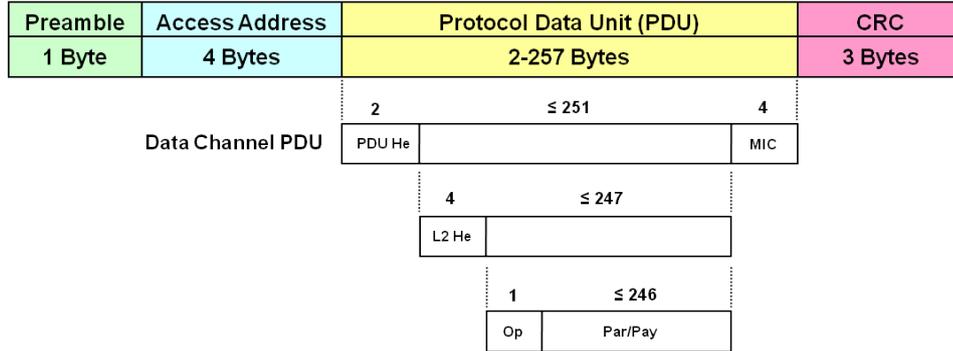


Figure 4.1: The packet structure in BLE systems. For the specifications of Table 4.1, the PDU is 42 bytes.

by Figure 4.1 [14], it is used as a reference value in the simulation. The total time required to transmit one packet is 400 μ s.

4.2. Phase-Based Demodulation Algorithm

As explained in Chapter 2, the actual physical parameter in FSK is the total phase. Even though the information signal modulates the instantaneous frequency of the carrier, some information is lost when differentiating the phase to obtain the instantaneous frequency. The demodulation techniques discussed in Chapter 2 all have one thing in common; the symbol state is estimated only using the information within a single symbol, which for a binary modulation equals a single bit. Since the phase of the received signal contains not only information about the present symbol, but also that of all past symbols, a better algorithm may be obtained by taking advantage of this memory effect in FSK. This section details a possible method to use this information through a new phase-based demodulation algorithm for FSK signals.

4.2.1. Phase Profile Reconstruction

As a starting point for this new demodulation algorithm, the phase trees in Figures 2.2 and 2.9 are remembered. Ultimately, in the absence of noise, the received FSK signal is expected to be somewhere on one of these curves in the phase tree. While the maximum range of the phase value grows over time, the *phase profile* of an FSK signal is limited to only one particular sequence of branches within the phase tree. When noise is added into the picture, it is no longer guaranteed that the sig-

nal plus noise phase is located exactly on the phase tree due to noise excursions in the IQ plane. However, because the additive noise is a zero-mean process, it is expected that the signal plus noise phase follows roughly the same phase trajectory as the modulation signal phase. Deviations from the actual phase tree over multiple symbol periods are unlikely to happen, and become increasingly less likely the more symbol periods are considered. However, this is strictly under the assumption that no click events occur. Click events are not a zero-mean process; if a click does occur, the phase tree is shifted up or down by 2π , and the phase profile does not necessarily coincide with the original phase tree anymore. For the purpose of explaining the demodulation algorithm, it is assumed that no clicks occur at all. The effect of clicks on the algorithm is discussed in subsection 4.3.3.

At the output of any FSK demodulator, a decision is made on which symbol state the signal is estimated to have occupied during a particular symbol period. A set of consecutive decisions then leads to an estimate of the originally transmitted symbol sequence. From the estimated symbol sequence, an estimation of the phase profile of the transmitted signal is obtained through the process of *phase tree reconstruction*:

$$\phi_{gen}[n] = f(y[m]), \quad (4.1)$$

where $y[m]$ is the estimated output symbol sequence, and $f(y)$ is a reconstruction function applied to y such that the phase profile $\phi_{gen}[n]$ is obtained. Since the phase profile and the symbol sequence have a different sampling rate, m is used to denote symbols, while n denotes samples. In GFSK, the reconstruction function is ideally a Gaussian filter (after up-sampling), identical to that in the transmitter.

When the demodulator makes no detection errors, then the phase profile of the transmitted signal *exactly* overlaps with the reconstructed phase profile. However, when a detection error does occur, then the reconstructed phase profile diverges from the transmitted phase profile at the point of the incorrectly detected symbol, as shown in Figure 4.2a. The difference between the estimated phase profile ϕ_{gen} and the originally transmitted phase profile ϕ_{TX} is defined as the phase profile error:

$$\phi_e[n] = \phi_{gen}[n] - \phi_{TX}[n]. \quad (4.2)$$

When a detection error occurs, ϕ_{gen} will diverge from ϕ_{TX} , resulting in a mean shift of $2h\pi$ in $\phi_{err}[n]$. This is illustrated in Figure 4.2b, where an error is made in the detection of the third symbol. Any error in detection clearly results in a dramatic shift in the mean of the phase profile error.

Unfortunately, at the side of the receiver, the transmitted phase profile is not available. However, due to the zero-mean characteristic of additive noise, the received signal plus noise phase ϕ_{RX} approximately follows the transmitted phase profile, again under the assumption that no click events occur. As a result, by substituting the received signal plus noise phase for the transmitted phase,

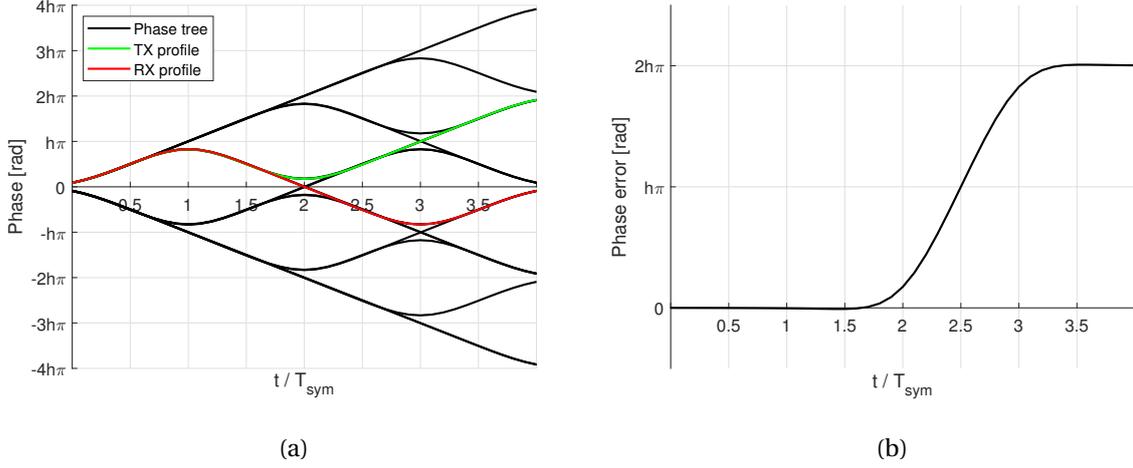


Figure 4.2: Phase profile reconstruction: (a) transmitted and received phase profiles with an error in the third symbol, and (b) the corresponding phase error signal.

$$\phi_e[n] = \phi_{gen}[n] - \phi_{RX}[n], \quad (4.3)$$

a rough approximation of the phase profile error is obtained. Both ϕ_{gen} and ϕ_{RX} are available at the receiver, which makes the phase profile error signal incredibly useful for finding symbols which have been incorrectly detected. By tracking the phase of the received signal and comparing it to what is expected from the FSK phase tree, PRC errors are potentially resolved in situations where other demodulators cannot do so. The next section focuses on the identification and elimination of these errors.

4.2.2. Error Detection

It is evident from Figure 4.2 that every individual error due to phase rotation cancellation is immediately visible in ϕ_e . Through the use of digital signal processing, it is possible to detect the typical behavior of this signal around errors. An example of ϕ_e with a noisy ϕ_{RX} and with multiple detection errors is shown in Figure 4.3. The blue curve indicates the phase profile error per sample or 'instantaneous' phase profile error ϕ_e as shown in (4.3), and the red curve represents the average phase profile error per symbol period $\phi_{e,avg}$:

$$\phi_{e,avg}[m] = \frac{1}{K_s} \sum_{k=0}^{K_s-1} \phi_e[n+k-mK_s], \quad (4.4)$$

where $K_s = f_s T_{sym}$ is the number of samples per symbol, such that $\phi_{e,avg}$ is the average phase profile error of the m -th symbol.

As a starting point, a simple threshold detection method can be applied. As seen from Figure 4.3, the variation in phase error is relatively small compared to the massive shift in ϕ_e when a detection error is made. By setting a threshold value ϕ_{th} such that symbols are considered as

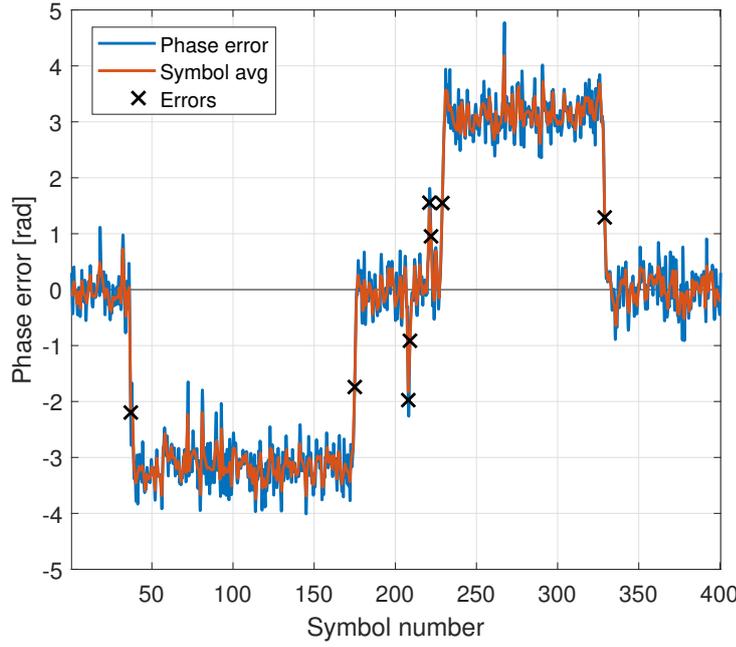


Figure 4.3: The instantaneous phase profile error (blue) and average phase error per symbol (red) in the presence of noise.

incorrectly detected when $\phi_e > \phi_{th}$, many singular errors are potentially removed. However, a threshold-based error detection algorithm is not very robust because of several reasons. First, the exact threshold value is critical in determining which symbol is actually considered incorrect; an additional error is introduced if the phase error threshold is exceeded in the symbol period prior to or after the incorrectly detected symbol. Therefore, a threshold that is too low or too high might introduce additional errors, deteriorating the performance of the algorithm. Furthermore, Figure 4.3 illustrates that the occurrence of two consecutive errors almost cancels the short mean shift of ϕ_e , which might be missed by a threshold-based detection algorithm. The value of ϕ_{th} is a trade-off between the number of missed errors and the number of false alarms, a classic situation from radar systems [15], which is most likely not the optimal error detection method. Therefore, a more accurate and robust algorithm is proposed.

Instead of relying on the reconstruction of the phase profile from an estimated symbol sequence, a better approach is to compare the received phase to multiple possible *local* phase profiles, and determine the symbol state based on the most probable option (i.e., the option with the smallest $\phi_{e,avg}$). This process is illustrated in Figure 4.4 with a simplified flowchart. The demodulation algorithm process is explained as follows: in each iteration, one symbol state is estimated by generating all combinations of phase profiles for a length of N symbols, where N is chosen by the designer. As a result, a total of 2^N reference phase profiles, one for each unique combination of reference symbols \hat{y}_m through \hat{y}_{m+N-1} , are generated and compared to the received phase. The symbol state of the m -th symbol then results from whichever phase profile has the smallest absolute value of $\phi_{e,avg}$. At the end of each iteration, the initial phase for the next iteration is determined from the reference

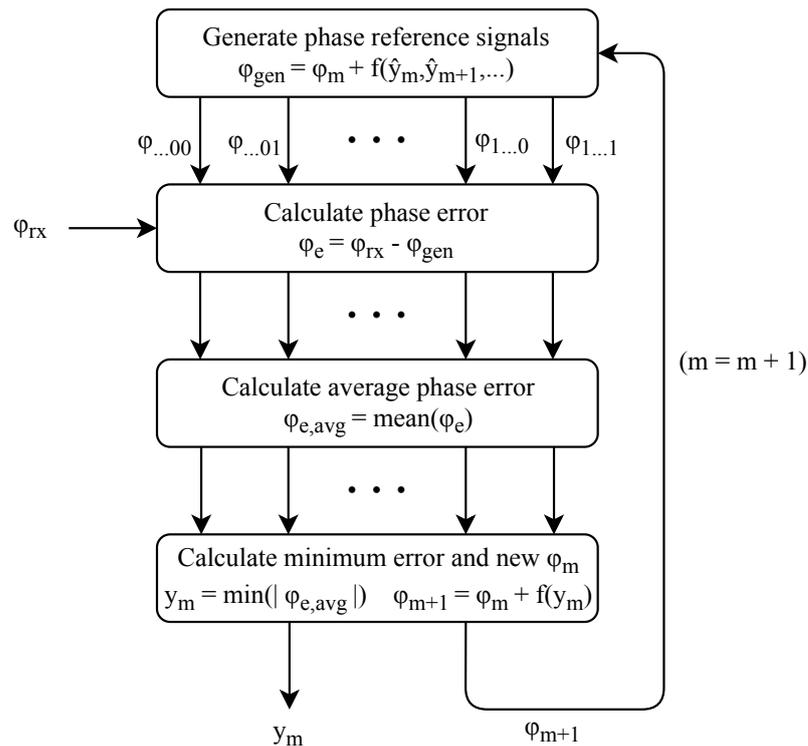


Figure 4.4: A flowchart representation of the demodulation algorithm process.

phase profile with the smallest phase profile error. This process is repeated until the entire symbol sequence is verified by the demodulation algorithm.

The main design parameter for this algorithm is the number of symbols N that are considered in each iteration. This is essentially a trade-off between algorithm complexity and functionality; an incremental increase in N results in an exponential increase in the number of required reference phase profiles. This is a massive increase in required computation time, and therefore, N should not be chosen too large. Contrarily, if N is too small (e.g., $N = 1$), then the deviations from the originally transmitted phase profile due to noise excursions that last for more than one symbol period – which is not uncommon in FSK demodulation – are potentially incorrectly interpreted by the algorithm. In order to guarantee an approximate zero-mean characteristic of the noise, N must be made large enough.

An example of the present algorithm operation is shown in Figure 4.5 for $N = 2$ and $E_b/N_0 = 10$ dB. Figure 4.5a shows the transmitted and received phase waveform, and Figure 4.5b shows the corresponding $\varphi_{e,avg}[m]$ of each reference phase profile. During each iteration, the minimum average phase profile error is used to decide the symbol state of the m -th symbol. This is of particular importance when the signal rotation is almost canceled, such as in the last few symbols of the waveform in Figure 4.5a where the demodulator output dynamic range is smallest. A conventional demodulator struggles with this, while the algorithm clearly has no problem detecting the symbol states.

It is important to note the differences between the present demodulation algorithm and a Viterbi-based demodulation algorithm, such as in [16]. A Viterbi algorithm is designed to undo con-

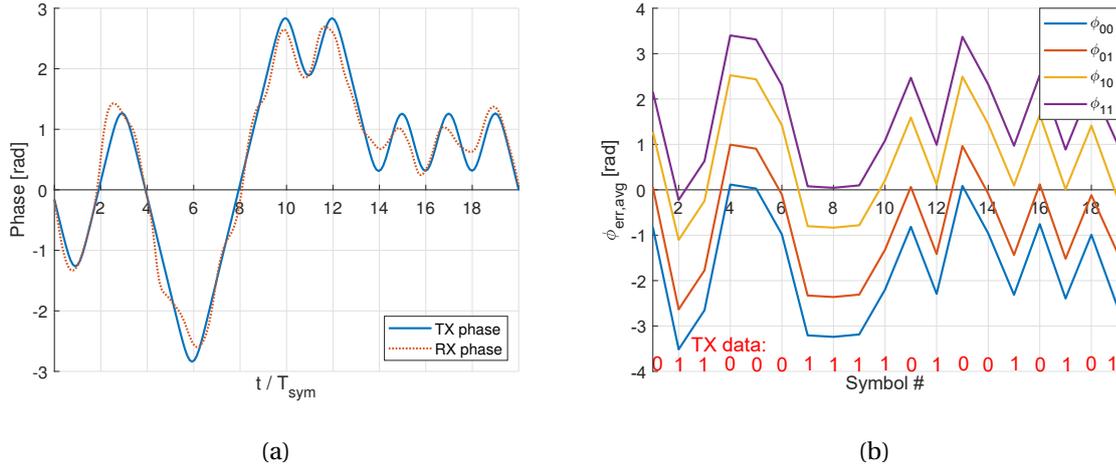


Figure 4.5: An example of the algorithm operation for $N = 2$ and $E_b/N_0 = 10$ dB: (a) the transmitted and received phase, and (b) the symbol-average phase profile error waveforms.

volutional coding, which means that the communicated data stream must be modified, resulting in a reduced data rate for the same symbol rate. The present algorithm does not require any modifications to the communicated information; the original data rate is maintained. Furthermore, a Viterbi-based demodulator employs a soft-decision phase detector, while the proposed algorithm uses hard decisions to determine the symbol sequence.

4.3. Algorithm Non-idealities

When applicable, the operation of the algorithm has so far been assumed to be under 'perfect' conditions. In this section, some non-ideal concepts are introduced that disturb the functionality and performance of the new demodulation algorithm. If applicable, a method is proposed to resolve or diminish the effects of each of these concepts.

4.3.1. Phase Offset

With the physical parameter in the demodulation algorithm being the phase, it is no surprise that any non-ideal error sources are also somehow related to phase. For FSK signals, there is no requirement on the initial position in the IQ plane; the information in FSK modulation is only in the rotation of the signal. However, in the present algorithm description, it is assumed that the received signal phase and the multiple reference phase signals have the same phase origin. In practice, the initial phase of the received signal is not defined, meaning that the received signal phasor starts at an arbitrary position in the IQ plane. If the phase origin of the received phase and the reference phases are not aligned, the algorithm will have trouble detecting the information properly, even in the absence of noise. This is because the algorithm is essentially trying to fit a phase profile procedurally on top of the received phase. Depending on the degree of misalignment of the phase origins, this fit may or may not be possible to do accurately. This is demonstrated in Figure 4.6, where a static

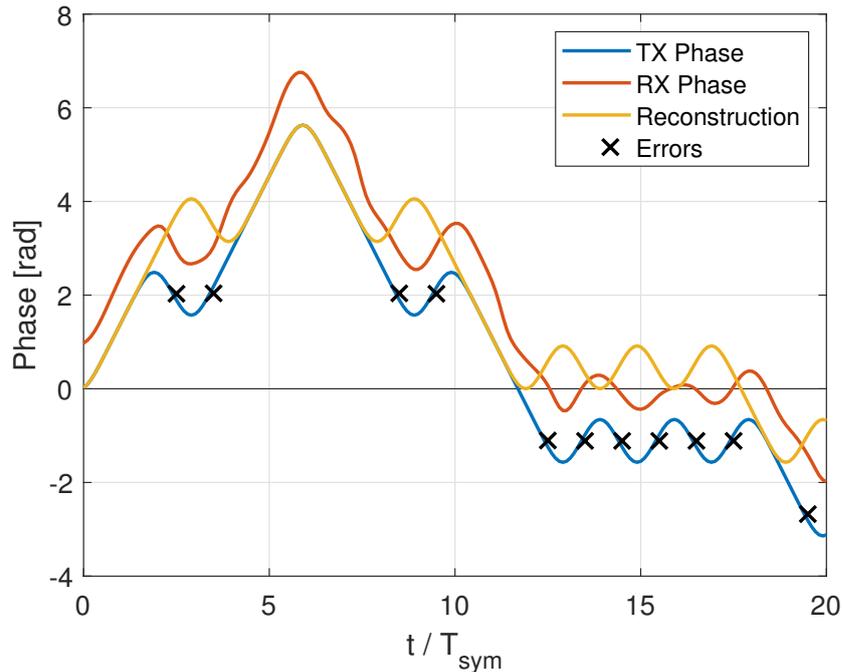


Figure 4.6: The transmitted (blue), received (red) and reconstructed (yellow) phase with a phase offset of $\theta_{os} = 1$ radian.

phase offset of $\theta_{os} = 1$ radian is introduced for the received phase with respect to the reference phase profiles. It is evident that the algorithm itself introduces some errors, despite the 'relative' shape of the received phase closely resembling that of the transmitted phase. Therefore, some form of phase calibration is necessary for the algorithm to function.

A BER simulation as a function of θ_{os} is done to obtain an idea of the maximum phase offset that the new demodulation algorithm can tolerate. The result is shown in Figure 4.7. Clearly, the phase alignment must be within approximately $\theta_{os} \approx 0.2$ radians for the algorithm to maintain close to its maximum performance. It should be noted that the curve in Figure 4.7 is symmetric around $\theta_{os} = 0$ rad and roughly repeats after every multiple of $2h\pi$. The symmetry around zero phase offset is self-explanatory. However, the cyclic behavior is more complex and requires an explanation. If $\theta_{os} > h\pi$, then a series of one or more detection errors introduced by the algorithm re-aligns the reconstructed phase profile with the received phase to within $\pm h\pi$. As a result, the shape of the curve in Figure 4.7 roughly repeats after every multiple of $2h\pi$, with a reduced local minimum of the BER for values of θ_{os} that are further away from zero. However, it is noted that in practice it is not possible for θ_{os} to exceed an absolute value of π . This is because even though the physical parameter is phase, the phase origin is automatically chosen such that the initially measured received phase is in the principal turn of the phase helicoid. In other words, the initial phase is always measured as an 'angle', since no additional information about the phase is available. Therefore, from the perspective of static (initial) phase offset, the cyclic behavior of the BER performance as a function of θ_{os} is meaningless in practice. However, the periodicity does become important in subsection 4.3.2 when considering a frequency offset.

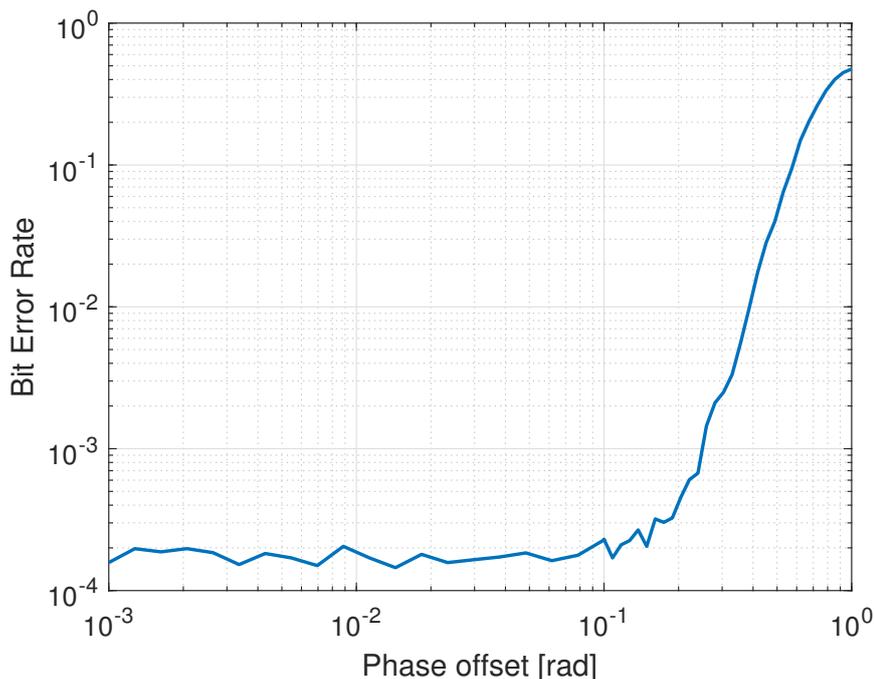


Figure 4.7: The bit error rate versus phase offset for $E_b/N_0 = 10$ dB.

As a starting point for eliminating this problem, it is relatively simple to take the initial received phase as the phase origin. However, this does not offer a guaranteed solution to the problem; if there is any deviation from the transmitted phase at the moment of sampling the initial phase, whether it is due to noise or interference, then the phase origins of the received signal phase and reconstructed phase remain misaligned throughout the algorithm process for that particular packet. Instead, a better approach is to use averaging, such that the influence of noise is diminished. Fortunately, looking at the packet structure of BLE in Figure 4.1, it is observed that the packet always begins with a preamble, which consists of eight symbol periods. The preamble consists of an alternating symbol pattern, and is predominantly used for time synchronization and clock recovery. Since this segment of the packet does not contain any actual information, it can be used to synchronize the phase origins before applying the demodulation algorithm to the actual information segment of the packet. By taking the average of the received signal over eight symbol periods, the noise influence on the initial phase estimate is reduced significantly, offering a much better phase alignment with respect to simply taking the initial phase. By taking the average *angle* instead of the *phase*, the effect of any click events within the preamble is diminished. However, the accuracy and consistency of this method do depend on the SNR at the input of the demodulator. Figure 4.8 shows the probability of failing to meet the criterion $\theta_{os} < 0.2$ radians as a function of the IBEND. Assuming that $\theta_{os} > 0.2$ radians *always* results in a packet error – which is a gross overestimation based on Figure 4.7 – then the averaging method works excellent for this application even for extremely low IBEND, and performs far better than the non-averaging method.

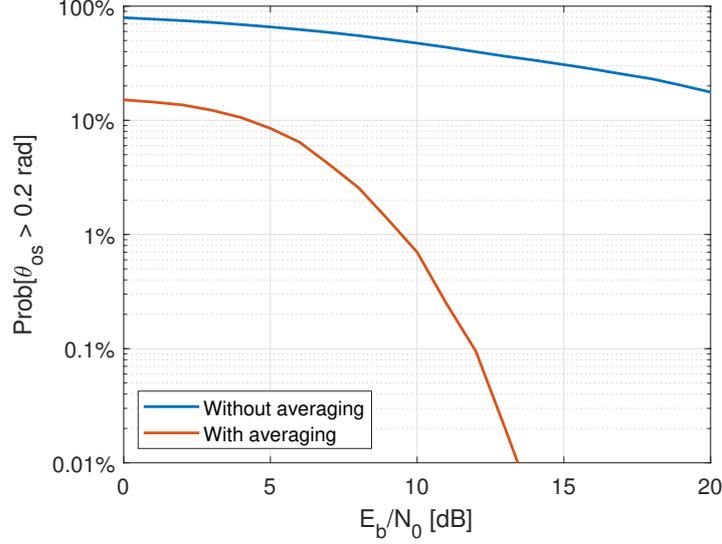


Figure 4.8: The probability of failing to meet $\theta_{os} < 0.2$ rad, shown as a function of E_b/N_0 for the averaging method and non-averaging method.

4.3.2. Frequency Offset

Another non-ideal condition for the algorithm is a frequency offset, which manifests itself as a dynamic phase offset, as demonstrated in Figure 4.9 for a frequency offset of 25 kHz in the absence of noise. The frequency offset appears as an increasing difference between transmitted and received phase. There are several causes active for frequency errors. For example, a frequency offset predominantly originates from a mismatch between the transmitter's and receiver's local oscillator (LO) frequency. The LO signals in a transmitter and receiver are typically derived from a crystal oscillator with finite frequency accuracy, for example, due to temperature sensitivity. Typical crystals used in Bluetooth have an accuracy of about ± 20 ppm [17], resulting in a maximum frequency offset between transmitter and receiver of roughly $f_{os} = 100$ kHz in Bluetooth applications, where the carrier is between 2.4 GHz and 2.5 GHz. Since the new demodulation algorithm uses the received signal phase in the demodulation process, any frequency offset is integrated and added directly to the received signal phase. This manifests itself as an increase in phase rotation per symbol for one particular signal state and a decrease for the other, depending on the polarity of the frequency offset. Therefore, if not taken care of, a frequency offset can give rise to problems in the demodulation algorithm. The impact of frequency drift on the demodulation algorithm is demonstrated through the example below.

Assuming that the phase offset is zero at the start of the packet, and that the frequency offset does not change through the course of a single packet, then the phase offset at the end of a packet is

$$\theta_{os} = 2\pi f_{os} N_{sym} T_{sym}, \quad (4.5)$$

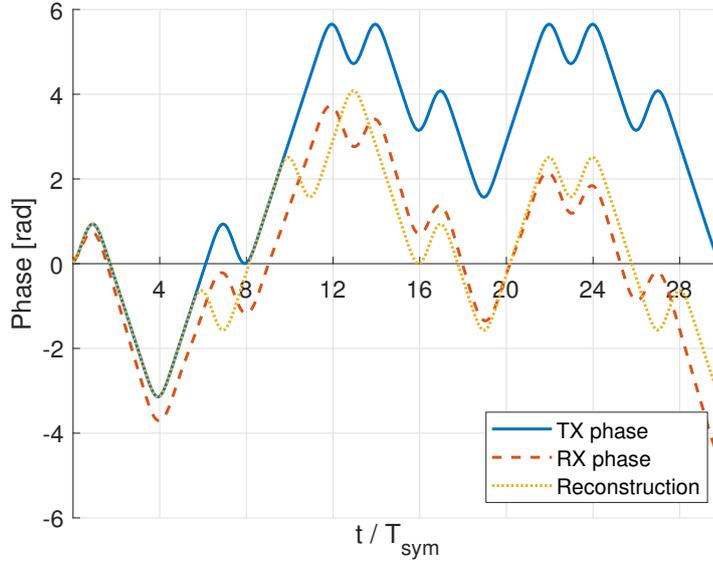


Figure 4.9: The transmitted, received and reconstructed phase for a frequency offset of 25 kHz.

where f_{os} is the frequency offset in Hz, N_{sym} is the number of symbols in the packet and T_{sym} is the symbol time in seconds. In subsection 4.3.1, it has already been established that the maximum tolerable phase offset is $\theta_{os,max} = 0.2$ radians. By rewriting (4.5), the maximum tolerable frequency offset within the BLE specifications of Table 4.1 is

$$f_{os,max} = \frac{\theta_{os,max}}{2\pi N_{sym} T_{sym}} \approx 80 \text{ Hz.} \quad (4.6)$$

It is evident that $f_{os,max}$ is orders of magnitude lower than the frequency drift of the LO carrier. Without any calibration, a small frequency offset already results in many errors due to the gradual misalignment of the transmitted and received phase profiles throughout the packet.

Furthermore, $f_{os,max}$ is only a small fraction of the modulation deviation frequency of BLE, which is $f_d = 250$ kHz, as obtained from (2.6). It is therefore difficult to obtain an accurate estimate on the frequency offset solely through the change in the slope of the phase profile. Fortunately, the phase profile error waveform used in the new demodulation algorithm does provide a solution to this problem. As the phase offset increases or decreases, the algorithm goes through cycles of the phase offset curve in Figure 4.7, as shown in Figure 4.9. In the phase profile error waveform, there is then an alternation of regions where the algorithm operates normally (small phase offset modulo π), and regions where many errors occur, as shown in Figure 4.10a for a frequency offset of 24 kHz at $E_b/N_0 = 8$ dB. It is then evident that the periodicity of ϕ_{err} is an indication of the frequency offset. Since the phase offset curve repeats every multiple of $h\pi$, the fundamental frequency component of ϕ_{err} is double the frequency offset for BLE. Using a fast-Fourier transform (FFT), the frequency offset is then estimated with a resolution of

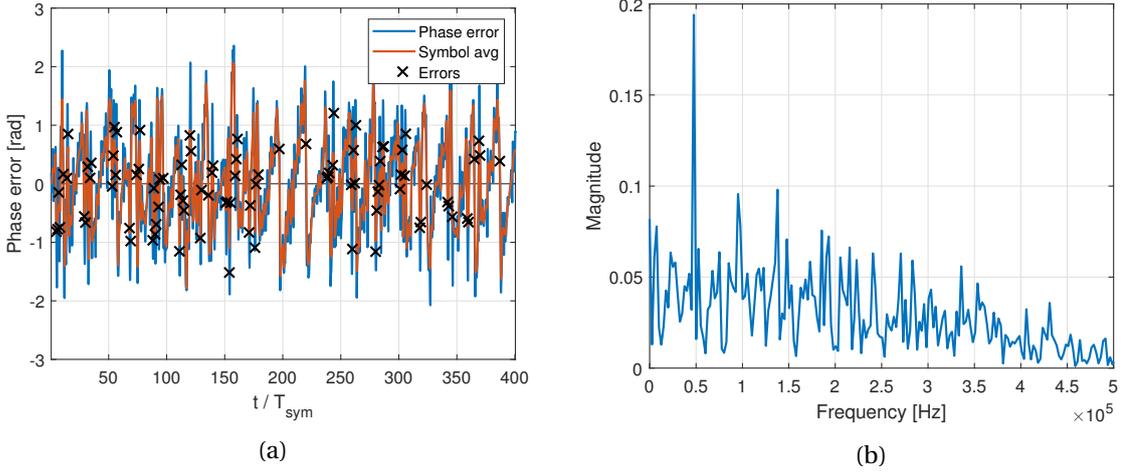


Figure 4.10: A frequency offset of $f_{os} = 24$ kHz: (a) phase error signal and (b) FFT magnitude of the phase error signal.

$$f_{res} = \frac{1}{2} \frac{f_s}{K_s N_{sym}} = \frac{R_b}{2N_{sym}}, \quad (4.7)$$

where K_s is the number of samples per symbol. If $\phi_{err,avg}$ is used to determine the frequency offset instead of ϕ_{err} , an identical frequency resolution is obtained; both the sampling rate and the number of samples of $\phi_{err,avg}$ decrease by a factor f_s/R_b . The only difference is the maximum unambiguously measurable frequency in the FFT, half the sampling rate, which remains well above the BLE specification of $f_{os} = 100$ kHz for both ϕ_{err} as well as $\phi_{err,avg}$.

With the specifications of Table 4.1, the frequency resolution is $f_{res} = 1250$ Hz, which is a coarse estimate of the frequency offset. The FFT of the phase error signal of Figure 4.10a is shown in Figure 4.10b. The peak magnitude corresponding to the frequency offset extends by roughly a factor 2 with respect to other peaks, and is therefore easily detectable, even for small IBEND values. The frequency offset estimate is then used in the generation of the phase reference signals in the algorithm to account for the additional increase or decrease in phase rotation.

With a frequency resolution of 1250 Hz, the frequency offset is estimated to within the range ± 625 Hz. Looking at the phase offset curve in Figure 4.7, this still results in significant performance degradation of the algorithm; even if the initial phase is perfectly aligned, errors will most likely occur near the end of the packet. In order to avoid this, the estimation must be accurate within ± 80 Hz, as obtained from (4.6). However, under the assumption that the frequency drift does not vary much in between packets, it is possible to first do a coarse frequency offset estimate with an FFT, followed by a finer estimate of the remaining frequency offset. For example, the remaining frequency offset can be determined by calculating the average slope of ϕ_e over the entire packet length. Alternatively, the DC component of the FFT and its polarity can be used in a step-by-step (incremental) approach to estimate the remaining frequency offset.

Unfortunately, as of now, there exists no frequency offset estimation technique that does not

require initial packet errors. If there is any frequency offset present, then this estimation technique requires *at least* one initial packet for the demodulation algorithm to function. Other FSK demodulators that do not rely on the phase signal estimate the originally transmitted information do not suffer as much from the frequency offset, since the offset only results in a small shift in the optimum decision threshold. For example, for a frequency offset of 25 kHz in a BLE system, the phase slope for a '1' symbol increases with 10%, while decreasing by 10% for a '0' symbol. The optimum threshold (in terms of phase) shifts from $\phi_{th,opt} = 0$ radians to $\phi_{th,opt} = \pm 0.2h\pi$ radians. As a temporary solution, a conventional demodulator such as the arctangent demodulator of Figure 2.3.2 can be used until the frequency offset estimate is accurate enough. Most parts of this demodulator are already required in order to support the present algorithm; the only necessary addition is an integrate and dump operation to make a decision on the occupied symbol states. Once the frequency offset is sufficiently well estimated, the algorithm is used again.

4.3.3. Click Events

So far, the algorithm process is discussed under the assumption that no click events occur. While it is true that click events are less prevalent in narrow-band FSK, they do still occur at lower IBEND values. Therefore, the consequences of click events for the algorithm must be evaluated. It is described in Section 4.2 that a click event moves the position of the received signal phase in the phase tree by a static 2π radians. Depending on the structure of the phase tree, the new position of the received signal phase after the click event may or may not be located on the original phase tree 'grid'. Consequently, a click event is observed as a sudden introduction of a static phase offset, which produces the same results as a regular phase offset, as detailed in subsection 4.3.1. However, without any method of detecting click events, this phase offset is not easily resolved without post-processing the phase error signal. Currently, there are no reliable methods to detect click events, as discussed in subsection 4.4.1.

Fortunately, there are special cases where a click event does not introduce a phase offset with severe consequences. For the special instances where

$$\gamma h = 1, \quad \gamma \in \mathbb{N}, \quad (4.8)$$

the new demodulation algorithm recovers from a click event by introducing a total of γ errors such that the received phase and the reconstructed phase align. Since a click event generally results in a detection error by itself in conventional demodulators, only $\gamma - 1$ additional errors are introduced by the new algorithm for every click event. While this in itself is not optimal, it is only a minor consequence for the algorithm performance, since click events are not the dominant error mechanism in narrow-band FSK. For BLE, the new demodulation algorithm introduces two errors for each click event in order to compensate for the mean shift in the phase tree. From Figure 3.14, it is evident that the consequences of additional errors for click events are small due to the relatively low click error rate with respect to PRC error rate for narrow-band FSK systems.

4.4. Alternative Algorithms

In this section, some alternative algorithms are discussed that either compliment the developed algorithm or act as a replacement. These are algorithms that were initially considered during the project duration, but were abandoned in pursuit of the algorithm described in the previous sections.

4.4.1. Click Detection and Elimination

As discussed in Chapter 3, click events are mainly dominant for wide-band FSK signals. However, as the signal-to-noise ratio at the input decreases, click events become more and more prevalent in FSK signals. In addition, the removal of PRC errors by the developed algorithm leaves mostly click events as the main cause for the remaining detection errors. In fact, as described in subsection 4.3.3, click events always cause two errors in the case of the developed algorithm when a modulation index of $h = 0.5$ is used, and many more for $h \neq 0.5$ due to the introduced phase offset with respect to the original phase tree. Therefore, some interest remains in the development of an algorithm that can successfully identify and eliminate clicks from the FSK signal.

After the characterization of click events as an error mechanism in frequency-modulated systems, many attempts were made to diminish the effect of clicks on the output signal-to-noise ratio. For instance, in [18] and [19], FM threshold extension is attempted through a click elimination process, such as the addition of an 'inverse' impulse at the position of a click event. Other methods for threshold extension are the phase-locked loop and frequency feedback [4]. However, the attempts at threshold extension only showed small improvements of up to 2-3 dB, and only in wide-band systems. The effort to extend the FM threshold has been long abandoned since the 1970s, as is evident by the decrease in scientific papers on this topic.

In order to reliably detect clicks, it is useful to keep in mind the typical characteristics of click events [6]:

1. A click event results in a static phase shift of $\pm 2\pi$ in the received signal phase.
2. For larger input SNR, the apparent bandwidth of a (fast) click event greatly exceeds the input bandwidth of the receiver.
3. For larger input SNR, the magnitude of the signal plus noise is typically small during a (fast) click event.
4. For narrow-band FSK signals, a click event results in a much larger phase change than that caused by modulation.

Using these characteristics to identify and correct symbol errors due to click events is key to further improving the developed algorithm. An example of the application of a simple click removal signal processing (CRSP) algorithm on an unmodulated carrier is shown in Figure 4.11a. This figure shows multiple phase waveforms before and after removing clicks. The unmodulated carrier in this simulation has a mean phase of zero, which is used to identify clicks; if the absolute value of the

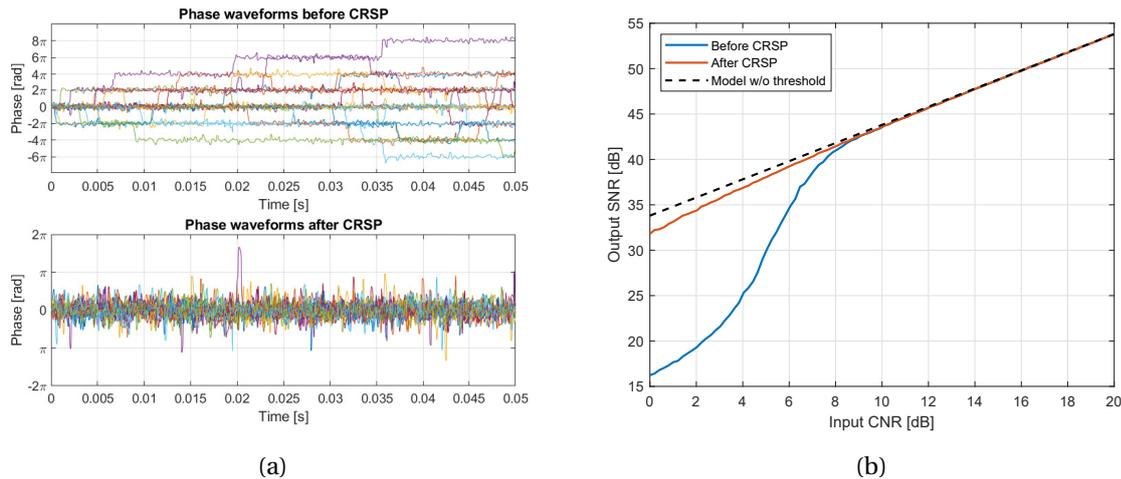


Figure 4.11: Click removal signal processing: (a) phase waveforms before and after CRSP, and (b) output SNR versus input CNR with and without CRSP.

phase waveform surpasses 2π , a click has occurred. In this CRSP algorithm, 2π is subtracted from the phase waveform whenever a click occurs, followed by a linear interpolation between the starting point and endpoint of the click. The resulting SNR plot is shown in Figure 4.11b. It is observed that the removal of click events has eliminated a significant part of the noise enhancement, which shows that most of the non-linear noise enhancement is, in fact, due to click events. Other non-linear effects, such as doublets, form the remaining part of the noise enhancement.

The problem of click detection and elimination is more complex when a modulated carrier is considered. Because of the signal phasor rotation due to modulation, a click event is no longer guaranteed to be exactly 2π . In addition, the mean phase is not fixed, contrarily to the unmodulated case. Currently, there is no reliable CRSP algorithm available for modulated carrier signals.

4.4.2. Data-dependent Decision Threshold

By using a pre-modulation filter, the phase rotation per symbol reduces for alternating symbol patterns, as discussed in Chapter 2. As a consequence, the optimum decision threshold for some symbols changes depending on the surrounding symbols. For example, for a GFSK signal with the specifications of Table 4.1, if a symbol sequence '101' is transmitted, then the phase rotation during the middle symbol is approximately $\Delta\phi \approx -0.6h\pi$. However, if the middle symbol were to be a '1', then the phase change would be $\Delta\phi \approx h\pi$. Clearly, the optimum decision threshold is no longer at exactly zero phase change, as it would be for unfiltered FSK. Instead, the optimum threshold is $\phi_{th,opt} \approx 0.3h\pi$. The optimum decision threshold is derived in a similar fashion for other symbol sequences, and is summarized in Table 4.2. From this table, it is seen that the decision threshold for patterns '010' and '101' are no longer zero. These particular patterns are cases where the dynamic range is minimal and where most PRC errors occur, as discussed in Chapter 2. Therefore, it would be very useful in those cases to take the optimum decision threshold rather than the standard one.

The main disadvantage of this concept is that an estimation of a future symbol is necessary to

Table 4.2: The optimum decision threshold for GFSK with BT = 0.5.

Sequence	'000'	'001'	'010'	'011'	'100'	'101'	'110'	'111'
$\phi_{th,opt}$ [rad]	$-0.3h\pi$	0	$-0.3h\pi$	0	0	$0.3h\pi$	0	$0.3h\pi$

determine the optimal decision threshold. This estimation can be obtained by merely putting the decision threshold at 0 temporarily for that symbol. It is not known in advance whether this estimation is correct, making this algorithm inconsistent. If the future symbol estimation is incorrect, then the optimum decision threshold is missed by $0.3h\pi$, and the advantage of the concept is lost. Fortunately, this is not a problem in the developed algorithm of Section 4.2; no assumption or estimate is made on future values, since all possible phase profiles are tested simultaneously. The developed demodulation algorithm will give more consistent results and is, therefore, more reliable in practice.

5

Algorithm Results

In this chapter, the performance of the demodulation algorithm developed in Chapter 4 is tested using an AWGN channel simulation. The algorithm performance is measured as the BER over a range of values of E_b/N_0 . The required BER to achieve a PER of 30.8% is of particular interest, since the receiver sensitivity for BLE is generally determined based on that reference.

5.1. Simulation Results

The specifications provided in Table 4.1 are used for all simulations. As a result, using (1.4), the receiver sensitivity is determined at $\text{BER} = 9.2 \cdot 10^{-4}$. In order to obtain simulation results with sufficient resolution within the IBEND range of 0-15 dB, the simulation length is set to 10^7 symbols. The arctangent demodulator from Figure 2.3.2 with an 'integrate and dump' decision element is used as a reference to show the relative change in performance.

A total of three different groups of simulations are performed:

1. BER versus E_b/N_0 for the conventional demodulator (referred to as $N = 0$) and the new demodulator for $N = 1$ and $N = 2$, with a receiver bandwidth of $1.2R_b$.
2. BER versus E_b/N_0 for $N = 0$ and $N = 2$, with a receiver bandwidth of $B = 0.8R_B$ and $B = 1.2R_b$.
3. BER versus receiver bandwidth at $E_b/N_0 = 8$ dB, for $N = 0$ and $N = 2$.

The simulation results are shown in Figures 5.1, 5.2 and 5.3, respectively. It is noted that the simulation results in this chapter are obtained under the condition that no phase or frequency offset is present. Therefore, the simulation results demonstrate the ideal peak performance of the new demodulation algorithm. For each performed simulation, the resulting minimum required

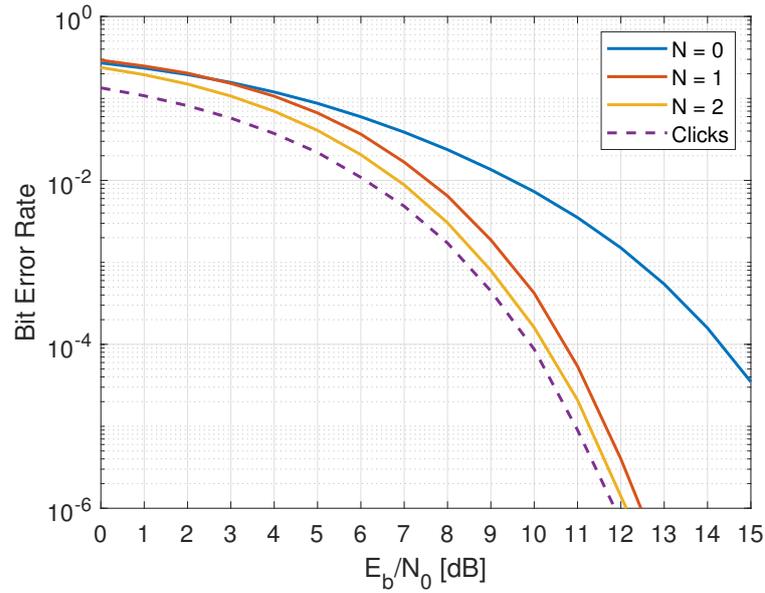


Figure 5.1: Bit error rate curves for the demodulation algorithm with $N = \{0, 1, 2\}$, including the click error rate for a conventional demodulator.

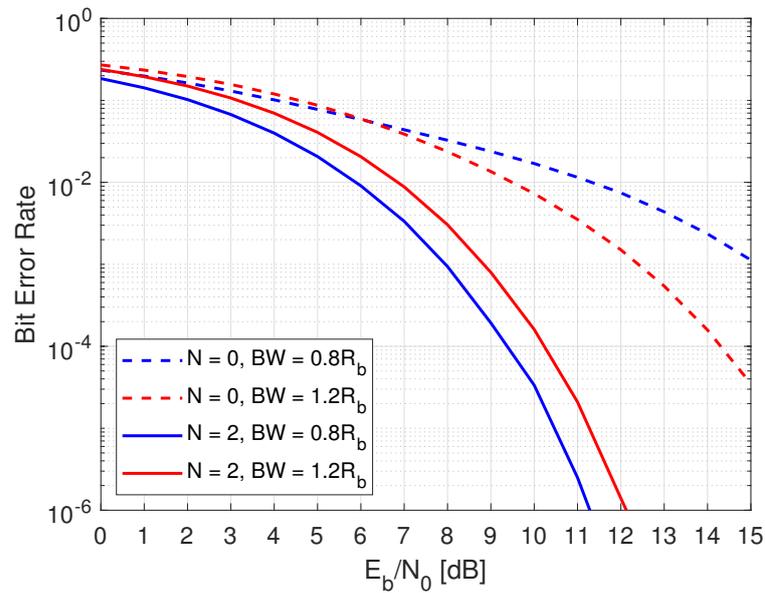


Figure 5.2: Bit error rate curves for a receiver bandwidth of $B = \{0.8R_b, 1.2R_b\}$ for the conventional demodulator and the new demodulator with $N = 2$.

IBEND to achieve a PER of 30.8% is shown in Table 5.1. The MATLAB script for the simulation of the modulation and demodulation process including the new demodulation algorithm is included in Appendix B.

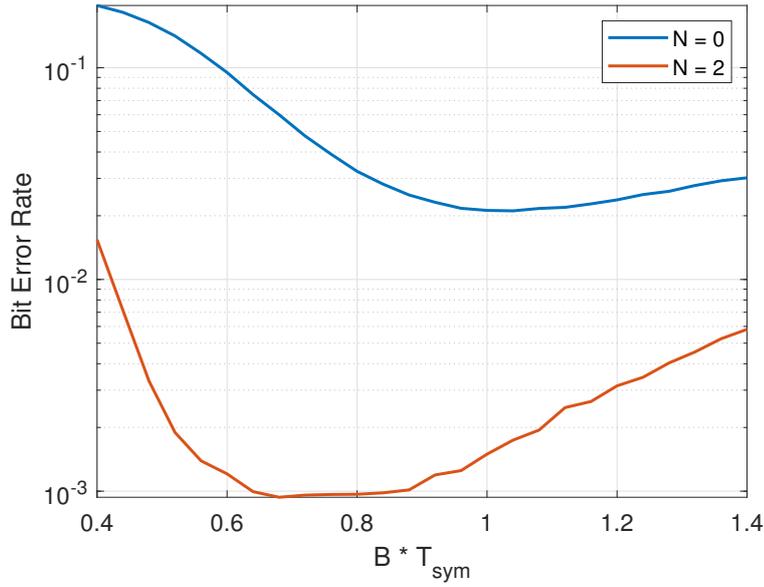


Figure 5.3: The bit error rate versus receiver bandwidth for the conventional demodulator and the new demodulator with $N = 2$.

Table 5.1: The minimum required IBEND to achieve PER = 30.8% for each performed simulation.

Algorithm	RX Bandwidth	$E_b/N_0 _{\min}$
$N = 0$	$0.8 R_b$	15.3 dB
	$1.2 R_b$	12.5 dB
$N = 1$	$1.2 R_b$	9.5 dB
$N = 2$	$1.2 R_b$	8.9 dB
	$0.8 R_b$	8.0 dB

5.2. Discussion

The results in Figure 5.1 show the relative performance improvement with respect to conventional demodulation when using the new algorithm. In addition, the click error rate (CER) (i.e., the BER exclusively caused by clicks), is also indicated in Figure 5.1. It is noted that the CER is obtained from simulating the conventional demodulator (denoted as $N = 0$). Therefore, the CER in Figure 5.1 does not account for the additional errors due to the recovery of the new demodulation algorithm in response to a click event. The performance improvement obtained with the $N = 1$ algorithm at the minimum required BER is almost 3 dB with respect to $N = 0$, which demonstrates that many PRC errors have been successfully eliminated. Furthermore, by taking $N = 2$ instead of $N = 1$, an additional improvement of 0.6 dB is obtained. It is clear that phase deviations due to noise are not limited to a single symbol period, and as a result, $N = 1$ is not sufficient to eliminate all PRC errors. Finally, it is noted that the BER curve for $N = 2$ and the CER curve differ approximately by a factor of two. For $h = 0.5$, it is established in subsection 4.3.3 that the demodulation algorithm responds to a click event with two detection errors. Therefore, nearly all remaining errors are actually caused by click events, and that negligibly few PRC errors remain.

Figure 5.2 shows the BER curves for $N = 0$ and $N = 2$ and for a receiver bandwidth of both $B = 0.8R_b$ and $B = 1.2R_b$. With a bandwidth below the data rate, a significant part of the transmitted information is filtered, which increases signal waveform distortion. A conventional receiver, such as $N = 0$, has a massively reduced performance when the received bandwidth is this narrow. However, the new demodulation algorithm actually performs better with a narrow channel bandwidth. In spite of the increased signal waveform distortion, the phase profile pattern is still properly recognized, while the reduction in receiver bandwidth reduces the noise power admitted to the receiver, and correspondingly reduces the click rate. By reducing the receiver bandwidth from $B = 1.2R_b$ to $B = 0.8R_b$, an additional performance improvement of 0.9 dB is obtained.

The trade-off between receiver bandwidth and BER performance for $N = 0$ and $N = 2$ is shown in Figure 5.3 for a specific IBEND value of $E_b/N_0 = 8$ dB. The conventional demodulator requires a receiver bandwidth of more than one times the data rate for optimal performance, while the new demodulation algorithm has its optimum performance at a bandwidth below the data rate. Furthermore, it is evident that the receiver bandwidth cannot be arbitrarily narrowed to reduce the noise; a sufficient amount of the signal energy must be received in order to demodulate the received signal correctly. From Figure 5.3, it appears that the optimal bandwidth for $N = 2$ is approximately $B = 0.7R_b$. Furthermore, with a fixed channel spacing, such as in Bluetooth, reducing the receiver bandwidth also results in a stronger attenuation of adjacent channels with the same filter order. Therefore, this particular characteristic of the proposed demodulation algorithm not only improves the noise resilience of the receiver, but also the blocker resilience.

The maximum performance improvement relative to conventional demodulation is obtained when using the $N = 2$ algorithm with a receiver bandwidth below the data rate (e.g. $B = 0.8R_b$), and the largest simulated improvement is 4.5 dB. Using the receiver sensitivity equation of (1.3), the new demodulation algorithm results in a receiver sensitivity (at $T = 298K$) of

$$P_{S,min} = -106 \text{ dBm} + NF|_{\text{dB}}. \quad (5.1)$$

The receiver sensitivity is still dependent on the physical circuit implementation of the receiver chain. Unfortunately, no reference noise figure is provided for this thesis project. However, based on typical provided noise figures in journal papers, such as in [20–25], $NF = 6$ dB appears to be a high-end and modern value. Assuming a noise figure of $NF = 6$ dB, a receiver sensitivity of $P_{S,min} = -100$ dB is obtained with the new demodulation algorithm. The receiver sensitivity performance with the new demodulation algorithm is better than any receiver sensitivity given in [20–25], which is typically reported to be around $P_{S,min} = -95$ dB.

6

Conclusion

6.1. Summary

This thesis discusses the two dominant error mechanisms present in FSK systems operating through an AWGN channel. The prevalence of both error mechanisms, clicks and phase rotation cancellation, is shown to have a relationship with the modulation index. It is demonstrated that errors in wide-band FSK systems are predominantly due to click events, while phase rotation cancellation errors are dominant in narrow-band FSK systems. Pre-modulation filtering, which is generally required in FSK systems to improve bandwidth efficiency, can significantly decrease the dynamic range at the output of an FSK demodulator due to the reduction of the phase rotation per symbol for particular symbol sequences. The reduction in phase rotation is a function of the type of pre-modulation filter and how strong the filtering action is. For symbol periods undergoing a reduced phase rotation, the probability of errors occurring due to phase rotation cancellation is (possibly) dramatically increased. Since nearly all modern applications of FSK modulation use narrow-band FSK, the emphasis here is put on resolving phase rotation cancellation errors to improve FSK system performance. The most important characteristic of phase rotation cancellation is that the noise distortion process remains zero-mean; the signal cancellation is only temporary, and every signal plus noise trajectory is required to follow the transmitted signal phase eventually, as long as no click events occur.

Based on the concept of phase rotation cancellation as an error mechanism, a novel FSK demodulation algorithm is proposed, with Bluetooth Low Energy as the target application. The demodulation algorithm uses the phase of the received FSK signal to detect symbols, rather than the instantaneous frequency. Based on the concept of phase trees, the phase profile of the transmitted FSK signal is estimated by comparing the received phase with multiple possible reference phase profiles. The number of reference phase profiles depends on the number of consecutive symbols

the algorithm considers simultaneously. The decision element determines the signal state based on the output signal with the smallest average phase error between the received phase and the reference phase. This process is applied to every symbol in the sequence. The initial phase for the next iteration is determined from the symbol state estimate of the current iteration. Through this process, the new demodulation algorithm procedurally estimates the originally transmitted phase profile while making use of the zero-mean characteristic of additive noise to improve detection accuracy significantly.

Calibration methods for phase and frequency offset are proposed, which are both necessary for the proper operation of the demodulation algorithm. The initial phase is calibrated by averaging the received signal angle during the preamble of the packet, where no actual information is transmitted. The frequency offset, often caused by an LO frequency mismatch between transmitter and receiver is coarsely estimated using the N -point FFT of the phase error signal. After that, the remaining frequency offset is determined using the total average of the phase profile error in a step-by-step approach. Since frequency drift is a relatively slow process, it is not necessary to compute the FFT for every packet.

It is discussed that click events, after removal of most phase rotation cancellation errors, are the main remaining source of demodulator output errors. An additional problem in the algorithm is that a click event results in a 2π shift of the received phase, which potentially results in the misalignment of the phase profile of the received signal with the phase tree. The misalignment is equivalent to a sudden phase offset, which can degrade the performance of the algorithm. Exceptions to this are FSK signals with $\gamma h = 1$, $\gamma \in \mathbb{Z}$, where the phase profile after a 2π shift remains on the 'grid' of the original phase tree. For a particular value of γ , the phase tracked by the algorithm is re-aligned with the received phase after introducing a total of γ errors. As of now, there is no reliable method to combat the consequences of click events.

Two algorithm embodiments, '1-symbol' ($N = 1$) and '2-symbol' ($N = 2$), are simulated in MATLAB using an AWGN channel simulation. The results from the new algorithm are compared to the performance of an arctangent demodulator with an 'integrate and dump' decision element. The simulation results show that the algorithm with $N = 1$ offers a 2.9 dB receiver sensitivity improvement at a reference packet error rate of 30.8%. The $N = 2$ algorithm improves the sensitivity by an additional 0.6 dB over the $N = 1$ algorithm. Furthermore, narrowing the input bandwidth from $B = 1.2R_b$ to $B = 0.8R_b$ results in a receiver sensitivity improvement of 0.9 dB for the $N = 2$ algorithm at a packet error rate of 30.8%. Narrowing the input bandwidth is not practical for the conventional arctangent demodulator; filtering out part of the received energy has a larger impact on system performance than the due to less noise. It is concluded that a phase-based FSK demodulation algorithm has a much higher performance than conventional demodulation techniques that rely on the instantaneous frequency and only use the available information in individual symbol periods.

6.2. Recommendations for Future Work

Due to time restrictions, not all aspects of the algorithm have been explored to the full extent during this project. The list below summarizes ideas and recommendations for future work.

Click removal It is shown in the results that, after the removal of most PRC errors by the algorithm, most of the remaining errors are caused by click events. In order to further improve the performance in FSK systems, the algorithm needs to be expanded to include click removal techniques as described in subsection 4.4.1. It is discussed that attempts made so far have not been successful in a significant way, and that there is still a lack of a robust and reliable algorithm. Therefore, one of the next steps could be the development and implementation of a click removal algorithm. If successful, the algorithm would also be useful in FSK systems with an arbitrary modulation index, rather than only for a select few modulation indices.

Fading channels In practice, wireless communication systems are not only corrupted by thermal noise, but also by multi-path propagation resulting in a phenomenon called *fading*. A more realistic approximation of the algorithm performance in practice is obtained if both thermal noise and multi-path propagation are taken into account. Multi-path propagation is simulated with the use of a fading channel that consists of several parameters: the number of additional paths, dependent on the environment of the system; the magnitude or 'signal strength' of each path, dependent on the travelled path and (possibly) reflection coefficients of materials; and the delay of each path, dependent on the travelled path. Furthermore, it is possible to have a fading channel with or without a line-of-sight signal component, referred to as Rician fading and Rayleigh fading, respectively. An extensive analysis of the algorithm performance when operating through fading channels is complex, but can give a good additional indication about the practicality of the algorithm.

Hardware implementation and testing So far, the algorithm has been shown to work in a simulated environment (MATLAB). The next step in the testing of the algorithm would be a hardware implementation that can be used in a practical context. The ultimate method to verify the algorithm performance is by physical testing in a practical environment, where both thermal noise and multi-path propagation affect the accuracy of detection at the side of the receiver. In addition, the effect of a phase offset and frequency offset can be tested. A hardware implementation would require the conversion of the algorithm from a floating-point representation to fixed-point. The algorithm could, for example, then be implemented in a hardware design language, and synthesized for use on a field programmable gate array (FPGA).

Improved frequency offset estimation At this time, estimation of the frequency offset is done using the FFT of the phase error signal. A packet with N symbols requires an N -point FFT, which might not be feasible in real-time if the packet size is too large. For example, the maximum packet

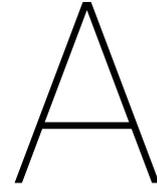
size in BLE is 2120 symbols, requiring a 2120-point FFT calculation, or a 4096-point FFT including zero-padding for a decreased computation time. In addition, the packet size increases not only FFT frequency resolution, but also the required accuracy of the frequency offset estimate; the same maximum phase offset tolerance applies over more symbol periods, resulting in a lower frequency offset tolerance. While the FFT method of frequency offset estimation theoretically works for arbitrary packet size, there might be other methods to explore.

Modulation index variations Depending on the transmitter architecture, it is possible that the modulation index is not always exactly equal to what is expected within the application. It is therefore important to test the algorithm performance under the condition that the modulation index is not static, but rather has some variations in its value. Testing for modulation index variations would result in an upper bound on the algorithm tolerance for this phenomenon. If necessary, measures need to be taken to eliminate the consequences of modulation index variations.

Bibliography

- [1] A. H. Morse, *Radio: Beam and Broadcast*. London, United Kingdom: E. Benn Ltd., 1925.
- [2] “FM radio,” Available at https://ethw.org/FM_Radio.
- [3] E. H. Armstrong, “A Method of Reducing Disturbances in Radio Signalling by a System of Frequency Modulation,” *Proceedings of the IRE*, vol. 24, no. 5, pp. 689–740, 1936.
- [4] H. Taub and D. L. Schilling, *Principles of Communication Systems*. New York, NY: McGraw-Hill, 1986.
- [5] E. McCune, *Practical Digital Wireless Signals*. New York, NY: Cambridge University Press, 2013.
- [6] —, “Extended phase shift keying,” Ph.D. dissertation, U. C. Davis, 1998.
- [7] M. J. E. Golay, “The Application of Radio Interferometry to Extraterrestrial Metrology,” *IRE Transactions on Space Electronics and Telemetry*, vol. 5, no. 4, pp. 186–193, 1959.
- [8] R. Lyons, “DSP tricks: Frequency modulation algorithms,” Available at <https://www.embedded.com/design/configurable-systems/4212086/DSP-Tricks--Frequency-demodulation-algorithms->.
- [9] H. Nyquist, “Thermal agitation of electric charge in conductors,” *Phys. Rev.*, vol. 32, pp. 110–113, 1928.
- [10] M. K. Simon, S. M. Hinedi, and W. C. Lindsey, *Digital Communication Techniques*. Englewood Cliffs, NJ: PTR Prentice Hall, 1995.
- [11] “Bessel functions: Asymptotic expansions for large argument,” Available at <https://dlmf.nist.gov/10.41>.
- [12] S. O. Rice, *Noise in FM Receivers*, ser. Time Series Analysis. New York, NY: Wiley, 1963.
- [13] D. Yavuz and D. T. Hess, “False Clicks in FM Detection,” *IEEE Transactions on Communication Technology*, vol. COM-18, no. 6, pp. 751–756, 1970.
- [14] “Bluetooth Low Energy packet types,” Available at <https://microchipdeveloper.com/wireless:ble-link-layer-packet-types>.
- [15] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 2002.

- [16] J. P. Foneska, "Soft-Decision Phase Detection with Viterbi Decoding for CPM Signals," *IEEE Transactions on Communications*, vol. 47, no. 12, pp. 1802–1810, 1999.
- [17] "BGM121/BGM123 Blue Gecko Bluetooth SiP module data sheet," Available at <https://www.silabs.com/documents/public/data-sheets/bgm12x-datasheet.pdf>.
- [18] M. J. Malone, "FM Threshold Extension Without Feedback," *Proceedings of the IEEE*, vol. 56, no. 2, pp. 200–201, 1968.
- [19] M. Polacek, "On FM Threshold Extension by Click Noise Elimination," *IEEE Transactions on Communications*, vol. 36, no. 3, pp. 375–378, 1988.
- [20] F. Kuo *et al.*, "A Bluetooth Low-Energy Transceiver With 3.7-mW All-Digital Transmitter, 2.75-mW High-IF Discrete-Time Receiver, and TX/RX Switchable On-Chip Matching Network," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 4, pp. 1144–1162, 2017.
- [21] H. Yi *et al.*, "A 0.18-V 382- μ W Bluetooth Low-Energy Receiver Front-end with 1.33-nW Sleep Power for Energy-harvesting Applications in 28-nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 6, pp. 1618–1627, 2018.
- [22] Y. Liu *et al.*, "A 1.9nJ/b 2.4GHz multistandard (Bluetooth Low Energy/Zigbee/IEEE802.15.6) transceiver for personal/body-area networks," in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 2013, pp. 446–447.
- [23] T. Sano *et al.*, "A 6.3mW BLE transceiver embedded RX image-rejection filter and TX harmonic-suppression filter reusing on-chip matching network," in *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, 2015, pp. 1–3.
- [24] Y. Liu *et al.*, "A 3.7mW-RX 4.4mW-TX fully integrated Bluetooth Low-Energy/IEEE802.15.4/proprietary SoC with an ADPLL-based fast frequency offset compensation in 40nm CMOS," in *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, 2015, pp. 1–3.
- [25] J. Prummel *et al.*, "A 10mW Bluetooth Low-Energy transceiver with on-chip matching," in *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, 2015, pp. 1–3.



GFSK Tool

In order to greatly reduce simulation effort, a tool is developed for GFSK signals in MATLAB App Designer. The graphical user interface (GUI) is shown in Figure A.1. From this GUI, several simulation parameters are easily controlled; signal parameters such as data rate, the number of samples per symbol, the carrier frequency and the modulation index, and filter parameters of both the TX filter as well as the RX filter. The noise channel can be set to noiseless or AWGN with variable E_b/N_0 . The simulation type can be set such that either a single point simulation is done, or a (multi-)variable sweep. Additionally, for two-variable sweep, an option is offered to create an image plot '2D Surface' instead of a surface plot. The length of the simulation is set through the input of a packet size and a number of packets. The output options allow the user to select which figure to display. The selection of figures depends on the simulation type. For variable sweeps, the typical output options are 'bit error rate' and 'packet error rate'. For a single-point simulation, the output options include a vector diagram, filter transfer functions, signals at the transmitter or receiver side (e.g. information signal, phase, instantaneous frequency, etc.), and the PSD of multiple signals. For a PSD output, averaging can be used to create a smooth PSD plot. For a BER or PER output, the decision element can be applied to either a 'sample and hold' or 'integrate and dump' operation. Optionally, the user has the option to separate detection errors into click errors and PRC errors. Doing so increases simulation time while scanning through signals for clicks, but does offer insight on the relative importance of the two error mechanisms. For some figures, an option is offered to control the domain of the x-axis.

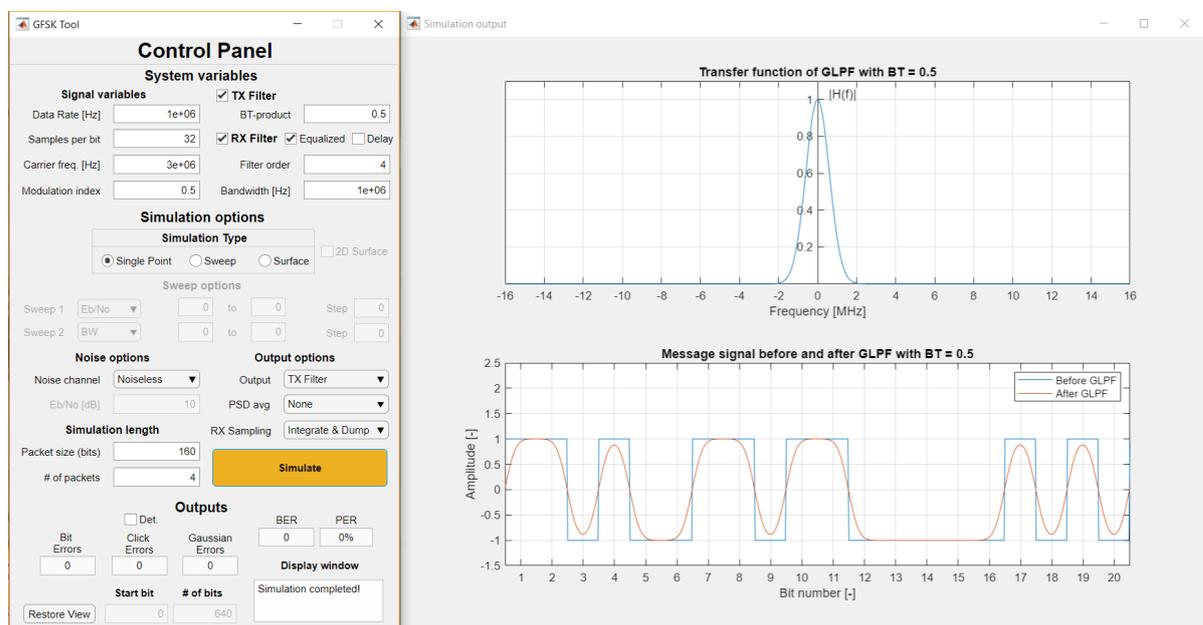


Figure A.1: The GUI of the developed GFSK simulation tool.

B

MATLAB Scripts

```
1 % This script is used to simulate the modulation and demodulation of a GFSK
2 % transmitter, including conventional demodulation and a newly proposed
3 % demodulation algorithm. The user inputs are the number of packets nPacks
4 % and the desired IBEND (Eb/NO) range. If the input is nPacks = 1, no IBEND
5 % range is requested, and the algorithm functionality plots are shown. For
6 % nPacks > 1, the BER curves are simulated.
7 % Filename: phase_reconstruct_N2.m
8 % Author: Armin Sabanovic
9 % Last edited: 14/08/2019
10 clear variables; close all; clc;
11
12 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% GFSK System Variables %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
13 dataRate = 1e6; % Data rate [Hz] %
14 sPerBit = 10; % Samples per bit [--] %
15 nBits = 400; % Number of bits (2^n) [--] %
16 Fc = 3e6; % Carrier frequency [Hz] %
17 h = 0.5; % Modulation index (BLE: h = 0.5) [--] %
18 BT = 0.5; % TX filter bandwidth-time product [--] %
19 n = 4; % RX filter order (Butterworth) [--] %
20 BW = 1.2e6; % RX filter bandwidth [Hz] %
21 EbNo = 10; % Bit energy to noise density ratio [dB] %
22 Fo = 0; % Frequency offset %
23 p0 = 0; % Phase offset %
24
25 % User inputs
26 nPacks = round(input('Please enter the number of packets to simulate:\n'));
27 if nPacks < 1
28     error('Negative number of packets. Terminating script.');
```

```

34 % Timing information
35 Fs = dataRate*sPerBit; % Sampling rate [Hz]
36 Ts = 1/Fs; % Sampling interval [s]
37 Tsym = 1/dataRate; % Symbol time [s]
38 T0 = nBits*Tsym; % Message signal duration [s]
39 len = round(Fs*T0); % Sample length of signal [-]
40 t = (0:len-1)*T0/len; % Time vector [s]
41 f = (0:len-1)*Fs/len; % Frequency vector [Hz]
42
43 % Modulation information
44 kf = pi*h/Tsym; % Peak angular frequency deviation [rad/s]
45
46 % TX filter: Gaussian lowpass filter
47 tmp = exp(-log(2)/2*(f(1:end/2)*Tsym/BT).^2); % Generate half spectrum
48 GLPF = [tmp conj(tmp(end:-1:1))]; % Generate full spectrum
49
50 % RX filter: Butterworth IF filter
51 s = 1i*f/BW; % Laplace variable (normalized)
52 Q = Fc/BW; % Filter quality factor
53 st = s + Q^2./s; % LPF to BPF transformation
54 p = exp(1j*pi*(2*(1:n)+n-1)/(2*n)); % Pole positions for n-th order
55 tmp = 1./prod(st(1:end/2)-p',1); % Generate half spectrum
56 BBPF = [tmp conj(tmp(end:-1:1))]; % Generate full spectrum
57
58 for bb = 1:length(EbNo)
59 bErrors1 = 0;
60 bErrors2 = 0;
61 for aa = 1:nPacks
62
63 %%% 2. Signal generation (TX) %%%
64 % Message signal generation
65 dataTX = randi([0 1],1,nBits); % Random bit sequence
66 % Ensure equal 1's and 0's
67 while sum(dataTX) ~= nBits/2 % Make equal number of 1s and 0s
68 dataTX(randi(nBits)) = 1*(sum(dataTX) < nBits/2);
69 end
70 msg = 2*repelem(dataTX,1,sPerBit)-1; % Message signal
71
72 % TX filter: Gaussian pulse-shaping
73 msg_fil = real(ifft(fft(msg).*GLPF)); % Apply GLPF to message signal
74
75 % Integration (frequency --> phase)
76 msg_int = Ts*cumsum(msg_fil); % Integrated message signal
77
78 % TX signal generation
79 TX_phase = kf*msg_int; % Baseband signal phase
80 TX_IQ = cos(TX_phase) - 1i*sin(TX_phase); % Baseband IQ signal
81 TX_RF = cos(2*pi*Fc*t + TX_phase + p0); % Modulated RF signal
82
83 %%% 3. Channel propagation (CH) %%%
84 nvar = sqrt(0.5*sPerBit/2*10^(-EbNo(bb)/10)); % Noise variance
85 noise = nvar*randn(1,len); % Noise vector
86 RX_RF = TX_RF + noise; % Received RF signal
87

```

```

88 %%%% 4. Conventional signal demodulation (RX) %%%%
89 % RX Filter: Noise/blocker filtering
90 RX_RF_fil = real(iff(fft(RX_RF).*abs(BBPF))); % Filtered RF signal
91
92 % Hilbert transformation demodulation
93 RX_hil = hilbert(RX_RF_fil); % Analytical signal f+
94 RX_IQ = RX_hil.*exp(-1j*2*pi*(Fc+Fo)*t); % Received baseband signal
95
96 % Demodulation
97 RX_phase = unwrap(angle(RX_IQ)); % Received signal phase
98 RX_demod = [(1/kf)*(diff(RX_phase)/Ts) 0]; % Demodulated signal
99
100 % Integrate & Dump filter
101 dataRX_up = reshape(RX_demod,sPerBit,nBits); % Reshape for trapz()
102 dataRX_int = trapz(dataRX_up)/sPerBit; % Averaging
103 dataRX1 = (sign(dataRX_int)+1)/2; % Decision
104
105 % Bit errors calculation
106 errorVec1 = 1*(dataTX ~= dataRX1);
107 bErrors1 = bErrors1 + sum(errorVec1);
108
109 %%%% 5. New Demodulation algorithm (RX) %%%%
110 % Algorithm
111 dataRX2 = dataRX1; % Initialization
112 x_data = repmat(dataRX2,4,1); % ...
113 phase_bavg = zeros(length(dataRX2),4);
114 comp = [0 0; 0 1; 1 0; 1 1]; % Reference matrix
115 for ii = 1:nBits-1
116     % Generate reference phases
117     x_data(:,ii:ii+1) = comp; % Insert reference
118     x_msg = 2*repelem(x_data,1,sPerBit)-1; % Message signal
119     x_msg_fil = real(iff(fft(x_msg,[],2).*GLPF,[],2)); % Filtering
120     phase_rct = kf*Ts*cumsum(x_msg_fil,2); % Reference phases
121
122     % Compare phase error signals
123     phase_diff = phase_rct(:,(1+(ii-1)*sPerBit):(ii+1)*sPerBit) ...
124                 - RX_phase(1+(ii-1)*sPerBit:(ii+1)*sPerBit));
125     phase_bavg(ii,:) = mean(phase_diff,2); % Symbol-average phase error
126     [M,I] = min(abs(phase_bavg(ii,:))); % Take index of minimum error
127     dataRX2(ii) = x_data(I,ii); % Output data vector
128     x_data(:,ii) = x_data(I,ii); % Adjust loop variable
129 end
130
131 % Bit errors calculation
132 errorVec2 = 1*(dataTX ~= dataRX2); % Error vector
133 bErrors2 = bErrors2 + sum(errorVec2); % Total bit errors
134
135 if nPacks == 1
136     % Phase error calculation
137     RX_phase_r1 = kf*Ts*cumsum(real(iff( ... % Reconstruction
138                               fft(2*repelem(dataRX1,1,sPerBit) - 1).*GLPF)));
139     RX_phase_diff1 = RX_phase_r1-RX_phase; %
140     RX_phase_bavg1 = mean(reshape(RX_phase_diff1,sPerBit,[]));
141

```

```

142 RX_phase_r2 = kf*Ts*cumsum(real(iff( ...
143         fft(2*repelem(dataRX2,1,sPerBit) - 1).*GLPF))); % Filtering
144 RX_phase_diff2 = RX_phase_r2-RX_phase ; % Phase error
145 RX_phase_bavg2 = mean(reshape(RX_phase_diff2,sPerBit,[])); % " Average
146
147 % Error vector preparation for plot
148 errorVec1(errorVec1 == 0) = NaN;
149 errorVec2(errorVec2 == 0) = NaN; % Needed for figure
150 end
151
152
153 end
154 BER1(bb) = bErrors1/(nBits*nPacks); % Bit error rate (conventional)
155 BER2(bb) = bErrors2/(nBits*nPacks); % Bit error rate (proposed)
156 fprintf('Starting iteration %g\n',bb)
157 end
158
159 %%% Figures %%%
160 if nPacks == 1
161     % TX Phase, RX phase and reconstructed phase
162     figure(1);
163     plot(linspace(0,nBits,length(TX_phase)),TX_phase,'LineWidth',1.5);
164     hold on;
165     plot(linspace(0,nBits,length(RX_phase)),RX_phase,'LineWidth',1.5);
166     plot(linspace(0,nBits,length(RX_phase_r1)),RX_phase_r1,'LineWidth',1.5);
167     stem(0.4:nBits-0.6,errorVec1.*TX_phase((sPerBit/2):sPerBit:end),'xk', ...
168         'LineStyle','none','MarkerSize',10,'LineWidth',1.5);
169     hold off;
170     grid on;
171     set(gca,'FontSize',12);
172     xlabel('t / T_{sym}');
173     ylabel('Phase [rad]');
174     title('Phase signal at TX and RX, and reconstructed phase before correction');
175     legend('TX Phase','RX Phase','Reconstruction','Errors');
176
177     % Phase error difference
178     figure(2);
179     plot(linspace(0,nBits,length(RX_phase)),RX_phase_diff1,'LineWidth',1.5);
180     hold on;
181     plot(0.5:nBits-0.5,RX_phase_bavg1,'LineWidth',1.5);
182     stem(0.4:nBits-0.6,errorVec1.*RX_phase_diff1((sPerBit/2):sPerBit:end),...
183         'xk','LineStyle','none','MarkerSize',10,'LineWidth',1.5);
184     hold off;
185     grid on;
186     xlabel('t / T_{sym}');
187     ylabel('Phase error [rad]');
188     title('Phase error signal before correction');
189     set(gca,'FontSize',12);
190     legend('Phase error','Bit avg','Errors');
191
192     % TX Phase, RX phase and reconstructed phase
193     figure(3);
194     plot(linspace(0,nBits,length(TX_phase)),TX_phase,'LineWidth',1.5); hold on;
195     plot(linspace(0,nBits,length(RX_phase)),RX_phase,'LineWidth',1.5);

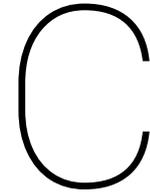
```

```

196     plot(linspace(0,nBits,length(RX_phase_r2)),RX_phase_r2,'LineWidth',1.5);
197     stem(0.4:nBits-0.6,errorVec2.*TX_phase((sPerBit/2):sPerBit:end),'xk', ...
198         'LineStyle','none', 'MarkerSize',10,'LineWidth',1.5);
199     hold off;
200     grid on;
201     set(gca,'FontSize',12);
202     xlabel('t / T_{sym}');
203     ylabel('Phase [rad]');
204     title('Phase signal at TX and RX, and reconstructed phase after correction');
205     legend('TX Phase','RX Phase','Reconstruction','Errors');
206
207     % Phase error difference
208     figure(4);
209     plot(linspace(0,nBits,length(RX_phase)),RX_phase_diff2,'LineWidth',1.5);
210     hold on;
211     plot(0.5:(nBits-0.5),RX_phase_bavg2,'LineWidth',1.5);
212     stem(0.4:nBits-0.6,errorVec2.*RX_phase_diff2((sPerBit/2):sPerBit:end),...
213         'xk', 'LineStyle','none', 'MarkerSize',10,'LineWidth',1.5);
214     hold off;
215     grid on;
216     xlabel('t / T_{sym}');
217     ylabel('Phase error [rad]');
218     title('Phase error signal after correction');
219     set(gca,'FontSize',12);
220     legend('Phase error','Symbol avg','Errors');
221
222     % Symbol-average phase error for each reference signal
223     figure(5);
224     plot(phase_bavg,'LineWidth',1.5); grid on;
225     set(gca,'FontSize',12);
226     set(gca,'XAxisLocation','origin');
227     ylabel('\phi_{ferr,avg} [rad]');
228     Xlb=mean(xlim);
229     Y = ylim;
230     Ylb=0.99*Y(1);
231     hXLbl=xlabel('Symbol #','Position',[Xlb Ylb],'VerticalAlignment','top',...
232                 'HorizontalAlignment','center');
233     legend('\phi_{00}','\phi_{01}','\phi_{10}','\phi_{11}');
234     box off;
235
236     % Phase error FFT
237     figure(6);
238     plot(linspace(0,1e6,nBits),abs(fft(RX_phase_bavg2))/nBits,'LineWidth',1.5);
239     hold off;
240     grid on;
241     xlim([0 Fs/20]);
242     xlabel('Frequency [Hz]'); ylabel('Magnitude');
243     title('FFT of the phase error signal');
244     set(gca,'FontSize',12);
245 else
246     % Plot BER curves
247     figure(1);
248     semilogy(EbNo,BER1,'LineWidth',1.5);
249     hold on;

```

```
250     semilogy(EbNo, BER2, 'LineWidth', 1.5);
251     hold off;
252     grid on;
253     xticks(EbNo);
254     set(gca, 'FontSize', 12);
255     xlabel('E_b/N_0 [dB]');
256     ylabel('Bit Error Rate');
257     legend('Before correction', 'After correction');
258 end
```



Journal Paper

This appendix chapter consists of a journal paper written on FSK demodulation errors from phase rotation cancellation, initially submitted to IEEE journal TCAS-II. The feedback of the reviewers is shown below.

Reviewer 1 This paper introduced phase rotation cancellation as a new error mechanism in FSK systems. The authors showed that this error mechanism is dominant in FSK systems with small modulation index. The paper is well-written and clear. However, I have the following concerns:

1. The paper is light on new research. The paper only showed a new error mechanism without proposing any schemes or demodulation algorithms that try to counter such errors.
2. All cited references (except one) are very old. Please cite more of the state of the art research.

Reviewer 2 This paper needs to be revised to correct the grammatical and formatting issues. I prefer to see equation numbers prefixed by "Equation x" rather than just being referred to as "(x)". Referring to an equation by just a number enclosed with parentheses are easily confused with paper reference numbers enclosed with brackets. Make it easier for the user to consume your information. Figure references should use the full word "Figure 7" rather than the abbreviation "Fig. 7". It seems like these shortcuts were used to make the article shorter. The wordiness of some of the sentences can be tightened up if the length of the article exceeds the maximum for this journal.

FSK Demodulation Errors from Phase Rotation Cancellation

Armin Šabanović, *Student Member, IEEE*, Masoud Babaie, *Member, IEEE*, and Earl W. McCune, *Fellow, IEEE*

Abstract—The fundamental difference between FSK and other commonly used modulation schemes is in the way the information is mapped onto the carrier signal; in FSK, the instantaneous frequency of the carrier is modulated, mapping the information to rotation of the signal phasor. Therefore, the error mechanisms in FSK operating through an AWGN channel have a different nature than in other common modulations. The two dominant error mechanisms in FSK, clicks and phase rotation cancellation, are established. The former is a well-known error mechanism, while the latter is described for the first time by this paper. Phase rotation cancellation errors occur due to signal rotation being canceled by noise rotation, i.e., ‘noise frequency’. This error mechanism is dominant in FSK systems with a small modulation index, particularly when pre-modulation filtering is applied to band-limit the FSK signal. The trade-off between the two error mechanisms as a function of the modulation index and the input SNR is demonstrated in this paper.

Index Terms—AWGN, demodulation, FSK, noise frequency

I. INTRODUCTION

FREQUENCY SHIFT KEYING (FSK) modulation has several advantages over other commonly used modulation schemes, such as quadrature amplitude modulation (QAM) and phase shift keying (PSK). One of the most important advantages of FSK is its constant-envelope characteristic, allowing for power-efficient transmitters that are suitable for low-power communication systems, since there is no linearity restriction on the power amplifier. However, in terms of bandwidth efficiency, FSK typically performs worse than other modulation schemes [1]. For most of the commonly used linear modulation schemes, such as QAM and PSK, the signal information is mapped to different positions in the in-phase and quadrature (IQ) plane, referred to as a constellation diagram. After demodulation, a decision is made as to which constellation point the signal vector is most likely representing. However, this is not the case for FSK; the information signal modulates the instantaneous frequency of the carrier (i.e., the information is mapped to *rotation* of the signal vector), generally leaving no fixed constellation diagram. In addition, as opposed to QAM and PSK, FSK is a non-linear modulation technique. Consequently, the error mechanisms in FSK differ from those of other commonly used modulation schemes.

First, a brief introduction on binary FSK (BFSK) and its properties is given. This paper then describes the two dominant error mechanisms active in FSK systems operating through an additive white Gaussian noise (AWGN) channel.

This work is partially supported by InnoPhase Inc.

A. Šabanović, M. Babaie and E. W. McCune are with Delft University of Technology, 2628 CT Delft, The Netherlands (e-mail: a.sabanovic@student.tudelft.nl; m.babaie@tudelft.nl; E.W.McCuneJr-1@tudelft.nl).

Next, the concept of ‘noise frequency’ is introduced and discussed. Finally, simulation results are shown to demonstrate the relationship between the modulation index and the two error mechanisms. In turn, this analysis provides an entrance to the development of new demodulation algorithms and circuit implementations based on these error mechanisms to improve FSK system performance.

II. BINARY FREQUENCY SHIFT KEYING

In FSK systems, the information signal modulates the instantaneous frequency of the carrier signal. An FSK modulated carrier is mathematically represented as

$$s(t) = A \cos(\omega_c t + \phi(t)) = A \cos\left(\omega_c t + \omega_d \int_{-\infty}^t m(\tau) d\tau\right), \quad (1)$$

where ω_c is the carrier frequency in rad/s, $\omega_d = 2\pi f_d$ is the deviation frequency in rad/s and $m(\tau)$ is the information signal, bounded within the range ± 1 .

Since FSK typically does not have a conventional constellation diagram in the IQ plane, the received signal vector position itself is not important, but rather its angular velocity or instantaneous frequency is. A vector diagram is therefore a more appropriate representation of an FSK signal, as illustrated in Fig. 1. From this figure, the constant-envelope characteristic of FSK modulated signals is evident; the signal magnitude remains constant while rotating clockwise (‘0’ transmitted) or counter-clockwise (‘1’ transmitted) around the IQ plane origin. The total signal vector rotation for each symbol period is determined by the modulation index

$$h = 2f_d T_s, \quad (2)$$

where f_d is the deviation frequency in Hz, and T_s is the symbol time in seconds. While a symbol in BFSK consists of only one bit, the single bits are referred to as symbols for completeness. Equation (2) states that within every symbol period, a phase rotation of $\Delta\phi = \pm h\pi$ radians is made. This is indicated in Fig. 1 for modulation indices $h = 0.2$ and $h = 0.6$.

Most modulation schemes do not exhibit a memory effect; the position of the signal vector during any symbol does not depend on previous symbols. However, this is not true for FSK. From (1), it is seen that the phase of the carrier is determined by integration of the information signal. Therefore, all of the previous symbols need to be taken into account to determine the present position of the signal vector. Since in equation (1) the phase $\phi(t)$ can grow indefinitely, it is not sufficient to consider only the angle of the carrier. The *angle* of a phasor is bound within the range $[-\pi, \pi)$, which is calculated using

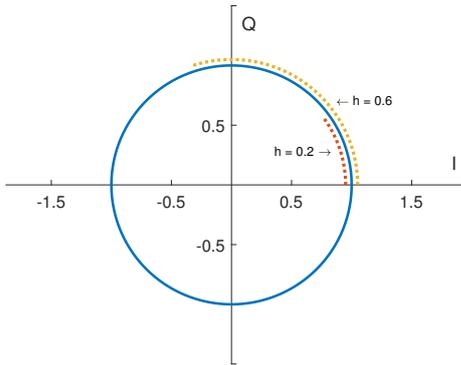


Fig. 1. Vector diagram of an FSK signal (solid), and the phase rotation per symbol for two different modulation indices (dashed).

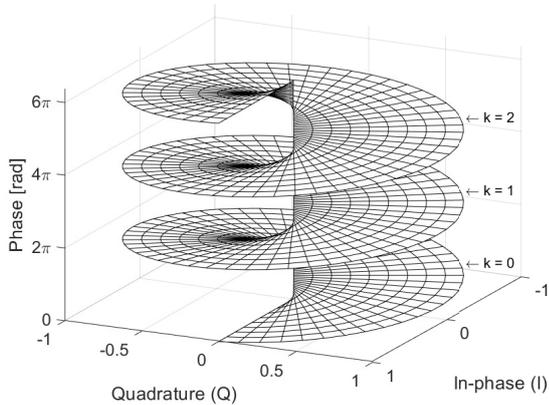


Fig. 2. A segment of three turns of the helicoid phase function (3), as a function of the in-phase and quadrature signal components.

the well-known four-quadrant arc-tangent function. The signal *phase*, however, is valid over all real numbers [2]. In order to correctly represent $\phi(t)$, the four-quadrant arc-tangent function is extended to cover all possible phase values [2]:

$$\phi(I, Q) = \tan_c^{-1} \left(\frac{Q}{I} \right) = \tan_2^{-1} \left(\frac{Q}{I} \right) + 2\pi k, \quad (3)$$

where $\tan_c^{-1}(x)$ is the new, continuous arc-tangent, $\tan_2^{-1}(x)$ is the four-quadrant arc-tangent, and k is an integer-valued number representing the memory effect. Here, $k = 0$ is an arbitrarily assigned origin. This brings the phase function into correct association with the physical behavior of the phase [3]. A segment of the phase function is graphically shown in Fig. 2. It is seen that the phase function is now multi-valued with I and Q , and the value of the phase is on one of the turns of this infinitely extending helicoid surface.

When filtering is used to band-limit the FSK signal, the total phase rotation per symbol becomes dependent on the symbol pattern, depending on the degree of filtering. As a result, the phase rotation per symbol can be significantly

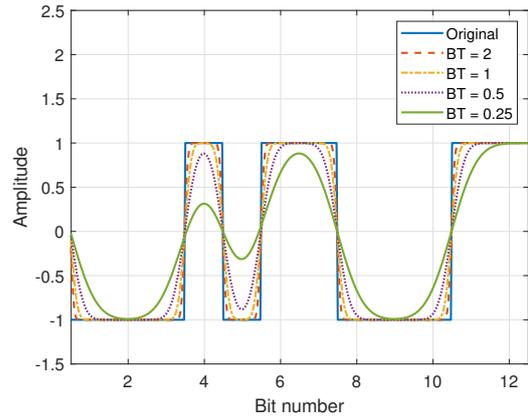


Fig. 3. Information signal before and after applying a Gaussian lowpass filter, for various BT-products.

reduced for alternating symbol patterns. Filtering is, however, generally necessary in modern applications, as unfiltered FSK signals have a poor bandwidth-efficiency due to many side-lobes occupying a large bandwidth. In order to minimize the bandwidth, and to avoid blocking adjacent channels, a pre-modulation filter is applied to the information signal. The filter smooths the transitions between symbols, reducing the occupied bandwidth. Data dependence is introduced if the shape of the information signal during a symbol depends on its adjacent symbols. Furthermore, inter-symbol interference (ISI) may be introduced if the filtering is too narrow, which means that the information signal does not have a full response for every symbol. These phenomena are shown in Fig. 3 for multiple Gaussian filters with different bandwidth-time (BT) products. Such filters are, for example, used in Gaussian FSK (GFSK) systems. Since the information signal modulates the instantaneous frequency of the carrier, the phase rotation is determined by the integral of the information signal. Consequently, it is clear from Fig. 3 that data dependence is always introduced in FSK systems with a pre-modulation filter, while ISI occurs for $BT < 0.7$. The reduced phase rotation due to pre-modulation filtering has consequences for the error rate, as is discussed in the next section.

III. ERROR MECHANISMS

In this section, the two dominant error mechanisms in FSK systems operating through an AWGN channel are introduced and discussed. First, the concept of clicks is explained, which is a well-known error mechanism in frequency modulation systems described by Stephen Rice in 1963 [4]. The concept of the second error mechanism, phase rotation cancellation, is introduced by this paper. In both cases, the received signal is perturbed by additive noise which, after the intermediate frequency (IF) filter, is mathematically represented in quadrature form as

$$n(t) = n_c(t) \cos(\omega_c t) - n_s(t) \sin(\omega_c t), \quad (4)$$

where $n_c(t)$ and $n_s(t)$ are both low-pass equivalent, independent and equally distributed Gaussian processes. The noise

magnitude is then Rayleigh-distributed, while the noise angle is uniformly distributed over the range $[-\pi, \pi]$ [5]. The noise phase, however, is not bounded within this range.

A. Clicks

In order to simplify the demonstration of clicks, an unmodulated carrier perturbed by AWGN is considered

$$r(t) = [A + n_c(t)] \cos(\omega_c t) - n_s(t) \sin(\omega_c t), \quad (5)$$

where the carrier only has an in-phase component A . The magnitude distribution, when adding this signal component to the AWGN, becomes Rician [5]:

$$f(x | A, \sigma) = \frac{x}{\sigma^2} e^{-\frac{(x^2 + A^2)}{2\sigma^2}} I_0\left(\frac{x A}{\sigma^2}\right), \quad (6)$$

where $I_0(x)$ is the zero-order modified Bessel function of the first kind, σ is the standard deviation of the noise, and A is the signal mean. For $A = 0$, the Rice distribution defaults to the Rayleigh distribution. For large A and $x \ll A$, an approximate Gaussian distribution is obtained, which is derived using Hankel's expansion of the modified Bessel function for large arguments [6].

Since the signal in (5) only has an in-phase component, the mean angle is expected to be zero. However, as the signal-to-noise ratio (SNR) decreases, perturbations caused by the noise become larger, and a finite probability exists for the signal plus noise trajectory to encircle the IQ plane origin. An example is shown in Fig. 4, where the origin is encircled twice. As a consequence, while the mean angle is still approximately zero, the phase value has moved to different turns of the helicoid surface of Fig. 2. This has consequences for modulation schemes like FSK, as this 2π step in the phase results in an impulse-like signal of area 2π at the demodulator output, due to the deletion or insertion of carrier cycles. This effect is referred to as a 'click', and is considered a noise enhancement effect since the noise power at the demodulator output appears larger than that initially admitted to the system. The demodulator SNR reduction due to this noise enhancement effect has been referred to as the 'FM threshold effect' by many communications engineers, including Rice [4].

In addition to clicks, it is possible for the signal plus noise trajectory to pass by the origin closely without encircling it, causing a rapid, zero-net phase variation which is referred to as a 'doublet'. As a result, the demodulator output during a doublet event can have large impulses that are similar to those of clicks. But, since there is no net phase change, i.e., the phase remains on the same helicoid turn, the total area underneath these impulses is also zero. This implies that the Fourier spectrum of doublets mostly consists of high-frequency content. Therefore, with a post-demodulation filter, doublets generally have a negligible contribution to the noise enhancement effect.

Clicks, click rates and the FM threshold effect, both for unmodulated and modulated carriers, have all been extensively modeled by Rice [4]. It is shown by Rice how the click rate

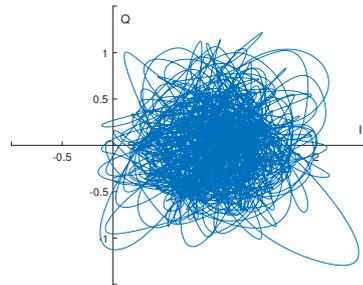


Fig. 4. Example signal plus noise trajectory that encircles the origin, with a unit signal amplitude and a noise variance of $\sigma^2 = 0.2$.

is affected by the modulation index, IF bandwidth and input SNR. Another important result that Rice found, particularly interesting for FSK, is that the polarity of clicks generally is opposite to that of the modulation [4]. As a result, click events are likely to cause errors in FSK demodulation.

B. Phase rotation cancellation

To improve bandwidth efficiency, FSK systems generally use a small modulation index in practice, which is referred to as narrow-band FSK. For small modulation indices and low SNR, the phase rotation per symbol period can be small relative to the noise variance. In that case, there is a finite probability that the independent noise components $n_c(t)$ and $n_s(t)$ have the right magnitude and polarity such that the transmitted signal rotation during a symbol period is completely overwhelmed or canceled. This error mechanism, phase rotation cancellation (PRC) or *phroX*, is shown graphically in Fig. 5, where the signal plus noise trajectories of several erroneous symbols are highlighted in red. The modulation index used is $h = 0.5$, and Gaussian pre-modulation filtering with a BT-product of 0.5 is applied to band-limit the FSK signal. It is clear from Fig. 5 that these errors, in this particular simulation, have a higher probability of occurring than clicks, making it the dominating error mechanism for small modulation indices in this SNR region. In addition, the PRC errors typically occur around the signal magnitude, where the radial velocity of the noise is minimal and the tangential velocity is maximal. Because of this, these errors would be missed by impulse-searching FSK threshold extension methods that focus on click elimination [7], [8], [9].

IV. NOISE FREQUENCY

In FSK systems, the demodulator extracts the received rotation of the signal vector in the IQ plane. In BFSK, an error occurs if the instantaneous frequency during a symbol period has a polarity opposite to that of the information signal. To determine the error rate, it is important to consider the transfer from the AWGN channel to the demodulator output. Since the transmitted information is in the rotation of the received signal vector in the IQ plane, only the tangential velocity component of the noise is important. At the demodulator output, this tangential component, from now on referred to

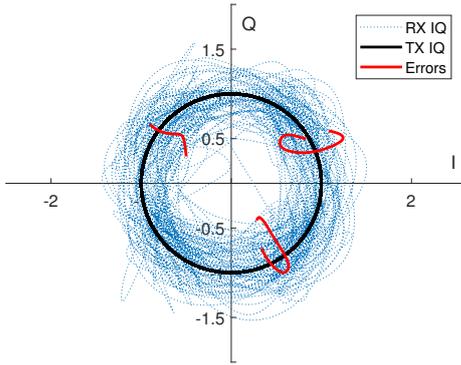


Fig. 5. Vector diagram of an FSK modulated signal with and without AWGN channel, with highlighted symbol error trajectories.

as (instantaneous) 'noise frequency', is superimposed on the instantaneous frequency deviation caused by the information signal. Since the noise is additive in the IQ plane, its phase excursions depend on the radial distance from the origin (i.e., the signal mean). This transfer from AWGN in the IQ plane to noise frequency at the demodulator output is highly non-linear, in particular for small signal mean, as expected from the phase function in (3).

The distribution of the instantaneous frequency of an unmodulated carrier plus AWGN becomes approximately Gaussian for large signal mean, similar to the Rice distribution given in (6). This approximation is intuitively explained by assuming a signal mean large enough such that the noise perturbations in the IQ plane are only on a small fraction of a circle with radius equal to the signal mean. When this circle segment is reasonably approximated by its tangent line, the bi-variate Gaussian distribution (from $n_c(t)$ and $n_s(t)$) in the IQ plane is linearly mapped to a straight line, and both the phase and instantaneous frequency distribution become approximately Gaussian.

At this time, there is no statistical model for the distribution of noise frequency at moderate or low SNR. Such a distribution could help determine the bit error rate accurately for FSK systems with small modulation indices. However, the derivation of this distribution is outside of the scope of this paper.

V. SIMULATION RESULTS

Simulations are performed in order to demonstrate the impact of the two previously mentioned error mechanisms and the relationship between the modulation index and the error mechanisms. A GFSK transceiver is simulated which closely resembles the operation of Bluetooth Low Energy (BLE), one of the main practical applications of FSK. The BLE application uses a modulation index of $h = 0.5$ and a BT-product of 0.5 for the Gaussian pre-modulation filter. A system-level overview of the simulated transceiver is given in Fig. 6.

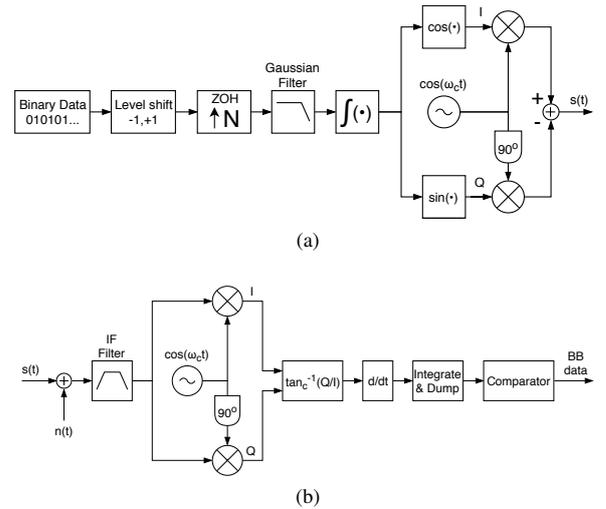


Fig. 6. System-level diagram of the simulated GFSK system: (a) transmitter and (b) receiver.

The transceiver is simulated for various levels of E_b/N_0 , and determines for every error whether it is caused by a click event or by phase rotation cancellation. Click events are detected by determining in which turn of the phase helicoid of (3) the transmitted and received signal phase are located, (i.e., the value of k). Whenever the difference between transmitted and received phase exceeds 2π , the origins are re-aligned and any error within the transition between the helicoid turns is assumed to be caused by a click event. Errors within the same helicoid turn (no change in k) are considered PRC errors.

Fig. 7 shows the bit error rate versus E_b/N_0 of both error mechanisms for modulation indices $h = 0.5$ and $h = 1$, and Fig. 8 shows the bit error rate versus modulation index for $E_b/N_0 = 10$ dB. It should be noted that as the modulation index increases, Carson's rule for frequency modulation bandwidth requires that the IF bandwidth is increased to ensure that enough signal power is captured by the receiver [1]. For narrow-band FSK, data rate dominates the signal bandwidth, while frequency deviation dominates for wide-band FSK. For the simulation results of Fig. 8, Carson's rule is taken into account by increasing the bandwidth linearly from 1 MHz at $h = 0.3$ to 1.3 MHz at $h = 1.0$, which is close to the optimum.

From Fig. 7 and Fig. 8, it is evident that the modulation index plays a significant role in the trade-off between various mechanisms. This is intuitively explained by the fact that if the modulation index is small, the noise frequency distribution is more likely to have the ability to cancel the signal rotation rather than the ability to encircle the IQ plane origin. This is particularly true when pre-modulation filtering is applied, which is generally the case in modern applications. With larger modulation indices, the exact opposite happens; the noise variance allows the signal to encircle the origin with a higher probability than actually canceling the signal rotation within the same helicoid turn. It is important to note here that phase rotation cancellation and click events become difficult to distinguish once the modulation index exceeds $h = 2$,

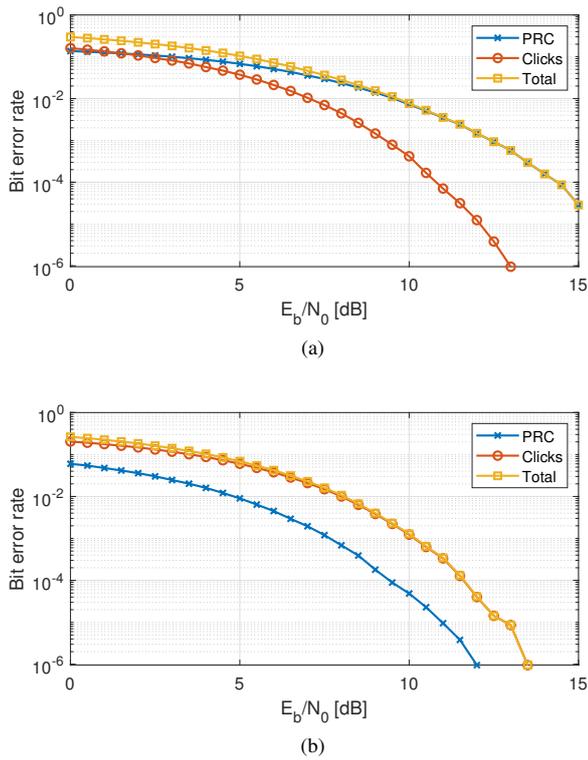


Fig. 7. Bit error rate curves of different error mechanisms for (a) $h = 0.5$ and (b) $h = 1.0$, for a data rate of 1 Mbps and a fixed IF bandwidth of 1.5 MHz.

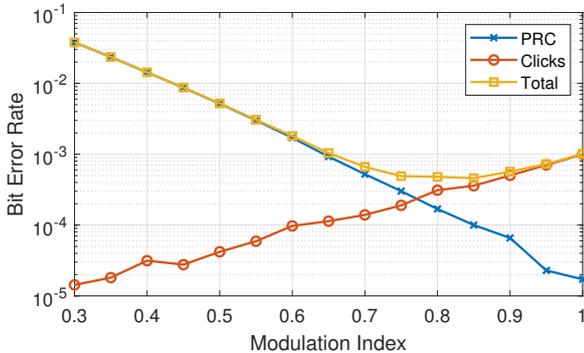


Fig. 8. Bit error rate versus modulation index for $E_b/N_0 = 10$ dB, a data rate of 1 Mbps and an IF bandwidth of 1 MHz ($h = 0.5$) to 1.3 MHz ($h = 1.0$).

which is when a full helicoid turn is traversed during a single symbol period. However, there is no application in practice that utilizes such wide-band FSK systems. For narrow-band FSK applications such as BLE, the dominant error source is phase rotation cancellation, and not click events.

VI. CONCLUSION

This paper discusses the dominant error mechanisms present in FSK systems operating through an AWGN channel. The concept of noise frequency is introduced by this paper, for which a potential distribution as a function of signal mean and noise variance can be derived in the future. This paper shows the relationship between the two error mechanisms and

the modulation index. It is demonstrated that wide-band FSK is dominated by click errors, while narrow-band FSK is dominated by phase rotation cancellation. Pre-modulation filtering, which is generally necessary in FSK systems to improve bandwidth efficiency, further reduces the dynamic range of the demodulated output. Using the knowledge from this paper, new demodulation algorithms and circuit implementations can be developed that improve FSK system performance.

REFERENCES

- [1] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. New York, N.Y.: McGraw-Hill, 2008.
- [2] E. W. McCune, *Extended Phase-Shift Keying*, Ph. D. dissertation U. C. Davis, Aug. 1998.
- [3] M. J. E. Golay, "The Application of Radio Interferometry to Extraterrestrial Metrology", *IRE Transactions on Space Electronics and Telemetry*, vol. 5, no. 4, pp. 186-193, 1959.
- [4] S. O. Rice, *Noise in FM Receivers*, Time Series Analysis New York, N.Y.: Wiley, pp. 395-422, 1963.
- [5] M. K. Simon, S. M. Hinedi and W. C. Lindsey, *Digital Communication Techniques*, 1st ed. Englewood Cliffs, N.J.: PTR Prentice Hall, 1995.
- [6] dlmf.nist.gov, *Bessel Functions: Asymptotic Expansions for Large Argument*, 2019. [Online]. Available: <https://dlmf.nist.gov/10.4>. [Accessed: 28-Mar-2019].
- [7] M. Polacek, S. Shitz and I. Bar-David, "On FM threshold extension by click noise elimination", *IEEE Transactions on Communications*, vol. 36, no. 3, pp. 375-380, 1988.
- [8] M. Malone, "FM threshold extension without feedback", *Proceedings of the IEEE*, vol. 56, no. 2, pp. 200-201, 1968.
- [9] H. Taub and D. L. Schilling, *Principles of Communication Systems*, pp. 371-440. New York, N.Y.: McGraw-Hill, 1986.



Delft University of Technology