

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Hewa Halpage, Samiru Gayan

Title:

Wireless Communications with Low-Resolution Quantization

Date:

2020

Persistent Link:

<https://hdl.handle.net/11343/268028>

Terms and Conditions:

Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.

Wireless Communications with Low-Resolution Quantization

Hewa Halpage Samiru Gayan

ORCID: 0000-0001-7561-1541

Submitted in total fulfilment of the requirements of the degree of
Doctor of Philosophy

Department of Electrical and Electronic Engineering
THE UNIVERSITY OF MELBOURNE

March 2021

Copyright © 2021 Hewa Halpage Samiru Gayan

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm or any other means without written permission from the author.

To Amma, Thatthha, Nangi and Hansika

Declaration

This is to certify that

1. the thesis comprises only my original work towards the PhD,
2. due acknowledgement has been made in the text to all other material used,
3. the thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Hewa Halpage Samiru Gayan, March 2021

Abstract

Wireless communication systems with low-resolution quantization are envisioned to be a major part in future wireless communication networks because of their potential to improve the energy efficiency of the network. In this thesis, we present a comprehensive and rigorous analytical investigation on the performance impact of using low-resolution phase quantization at the receiver of a wireless communication system, when compared to traditional high-resolution systems. To that end, we consider three different system setups; a point-to-point wireless communication system with coherent detection, a point-to-point wireless communication system with non-coherent detection and a multi-antenna system with coherent detection. We study the optimum detectors and draw fundamental insights on the error probability performance of low-resolution quantization systems in the presence of fading and noise.

Firstly, we focus on coherent detection with M -ary phase shift keying (M -PSK) modulation and, derive the optimum maximum likelihood (ML) detector for a single-input single-output (SISO) system. Utilizing the structure of the derived detector, a general average symbol error probability (SEP) expression for M -PSK modulation with n -bit quantization is obtained when the wireless channel is subject to Nakagami- m fading. We show that a transceiver architecture with n -bit quantization is *asymptotically* optimum in terms of communication reliability if $n \geq \log_2 M + 1$.

The coherent detection techniques discussed above require channel state information (CSI) to be available at the receiver. Due to the non-linear nature of quantization, channel estimation has been one of the major challenges associated with low-resolution quantization based systems. Taking these into account, next we focus on non-coherent detection by adopting the differential quadrature phase shift keying (DQPSK) mod-

ulation scheme to differentially encode the transmit data. At the receiver side, we employ non-coherent detection that does not require instantaneous CSI. With DQPSK modulation, the ML detector is derived using which, a general average SEP expression with n -bit quantization is obtained when the wireless channel is subject to Rayleigh fading. It is shown that a transceiver architecture with n -bit quantization is *asymptotically* optimum in terms of communication reliability if $n \geq 4$. That is, the decay exponent for the average SEP is the same and equal to 1 with infinite-bit and n -bit quantizers for $n \geq 4$. Therefore, when the channel knowledge is not available at the receiver, the quantizer has to use one additional bit to achieve optimum communication robustness.

Finally, we extend our investigation to low-resolution quantization based multi-antenna wireless communication systems equipped with one transmit antenna and N receive antennas. We derive the ML detector and then propose three sub-optimum detection rules based on selection combining which have less computational complexity compared to the ML detector. We also note that the simple sub-optimum detector that selects the path with the channel that locates the rotated constellation point furthest away from the decision boundary is *asymptotically* optimum in terms of communication reliability if $n \geq 3$. An extensive simulation study is performed to illustrate the accuracy of the derived results.

Preface

This thesis contains only original work that I conducted under the supervision of Prof. Jamie Evans, Dr Rajitha Senanayake and Dr Hazer Inaltekin during my PhD candidature in the Department of Electrical and Electronic Engineering at the University of Melbourne, Australia. This thesis has not been submitted for any other qualification. All the work towards the thesis was carried out after the enrolment in the degree. No third party editorial assistance was provided in the preparation of the thesis. The research work was supported by the Australian Government and the University of Melbourne through the Melbourne Research Scholarship (MRS). I acknowledge the student travel grants provided by the Australian Communication Theory Workshop in 2018 and 2019 and the Australian Information Theory School in 2019. I acknowledge the University of Melbourne and Prof. Jamie Evans for funding my attendance at the International Telecommunication Networks and Applications conference in Melbourne, Australia in November 2017 and IEEE International Conference on Communications in Shanghai, China in May 2019.

The main contributions of this thesis are presented in Chapters 2, 3 and 4. I conducted the majority of the research work, including the problem formulation, theoretical analysis, computer simulations and manuscript writing. Prof. Jamie Evans, Dr Rajitha Senanayake and Dr Hazer Inaltekin supervised all the work with proper guidance, fruitful discussions and constructive feedback. The summary of publications related to this thesis is given below.

- H.H.S. Gayan, R. Senanayake, H. Inaltekin and J. Evans: "Low-Resolution Quan-

tization in Phase Modulated Systems: Optimum Detectors and Error Rate Analysis," *IEEE Open Journal of the Communications Society*, pp. 1000-1021, July 2020.

- H.H.S. Gayan, H. Inaltekin, R. Senanayake and J. Evans: "Phase Modulated Communication with Low-Resolution ADCs," *IEEE International Conference on Communications (ICC)*, pp. 1-6, Shanghai, China, May 2019.
- H.H.S. Gayan, R. Senanayake, J. Evans: "On the Symbol Error Probability for QPSK with Quantized Observations", *27th International Telecommunication Networks and Applications Conference (ITNAC)*, pp. 1-6, Melbourne, VIC, Australia, Nov. 2017.

All these papers produced from this thesis are my original work. In all the papers my contributions are literature survey, problem formulation, theoretical analysis, Monte Carlo simulations and manuscript wiring. The contributions of the co-authors Prof. Jamie Evans, Dr Rajitha Senanayake and Dr Hazer Inaltekin are technical comments and guidance, discussions on result analysis and manuscript proofreading. I declare that I am the primary author and I contributed more than 50 percent for all the above publications.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my primary supervisor Prof. Jamie Evans for his invaluable support and guidance throughout my doctoral research tenure. I was constantly motivated by his wisdom, limitless enthusiasm and dedication towards excellence, and continuous encouragement. He was my main inspiration during difficult periods in my postgraduate studies and always helped me to overcome tough research challenges through his immense technical knowledge. I feel blessed and privileged to be one of his PhD students.

I would also like to express my sincere gratitude to my co-supervisors Dr Rajitha Senanayake and Dr Hazer Inaltekin for their continuous support and encouragement throughout my PhD study. The constructive feedback and relentless guidance that they offered were encouraged me to be a better researcher for which I am truly thankful. I must admit that I could attain a lot of technical skills and improve my technical writing because of them. I would also like to express my sincere gratitude to my advisory committee chair, Prof. Elaine Wong for her invaluable guidance. She supported me throughout my PhD studies to overcome the challenges.

I would also like to extend my gratitude to Prof. Dileeka Dias from the University of Moratuwa who encouraged me in studying communications theory. I also would like to be grateful to Eng. Kithsiri Samarasinghe from the University of Moratuwa for his inspirational guidance to make me a better person. I also acknowledge Prof. Peter Smith from Victoria University of Wellington for providing me useful insights for the third chapter of my PhD thesis. I would also like to thank Prof. Elaine Wong, A/Prof. Brian Krongold, A/Prof. Gavin Buskes, Prof. William Shieh and Prof. Thas Nirmalathas for offering me casual tutoring and workshop demonstration positions

during the semesters. I would also like to give special thanks to Dr Tharaka Samarasinghe, Prof. Ruwan Udayanga, Dr Chandika Wavegedara, Dr Nuwan Dayananda, Dr Duleeka Nayanasinghe and Dr Chamira Edussooriya from the University of Moratuwa who motivated me in pursuing a career in higher studies. Also, I would like to extend my heartiest gratitude to all my school teachers and university lecturers for their invaluable guidance. Further, I am also grateful to the University of Melbourne for providing me with the opportunity to pursue my PhD. I would also like to acknowledge the Australian Government and the University of Melbourne for providing me with financial assistance through the Melbourne Research Scholarship and multiple other travel scholarships during my PhD studies.

I would like to thank my friends and colleagues Sourav, Jithin, Shalanika, Bhathiya, Trang and Neil for making my life at the University of Melbourne a truly memorable one. I also would like to thank my friends Para ayya, Dinuka ayya, Hirantha ayya, Viraj ayya, Suresh, Sampath, Sadeep, Gavin, Thushara, Kanishka, Dinesh, Danesh, Upul, Manjula, Sandunil ayya, Welgama ayya, Geeth ayya, Kathika and Tharindu for being with me during this journey.

Finally, I would like to thank my wife Hansika, my parents, my parents-in-law and my sister Disney for their continuous affection and care. I would like to dedicate this work to them.

Contents

1	Introduction	1
1.1	Low-Resolution Quantization	4
1.2	Related Work	8
1.2.1	Signal Detection	9
1.2.2	Channel Estimation	11
1.2.3	Performance Analysis	12
1.2.4	Precoding	13
1.2.5	System Design	14
1.3	Focus of the Thesis	15
1.4	Main Thesis Contributions	16
1.5	List of Publications	21
1.5.1	Journal Papers	21
1.5.2	Conference Papers	21
2	Low-Resolution Quantization in Single Input Single Output Systems	23
2.1	Introduction	23
2.1.1	Background and Motivation	23
2.1.2	Main Contributions	24
2.2	System Setup	27
2.2.1	Transmitter Architecture	27
2.2.2	Channel Model	27
2.2.3	Receiver Architecture	27
2.2.4	Quantization Process	28

2.2.5	Key Assumptions	28
2.3	Optimum Signal Detection	29
2.4	Average Symbol Error Probability	33
2.4.1	Symbol Error Probability for $n \geq \log_2 M$	33
2.4.2	Centering Property: Impact of Quantization Bits on the Average SEP	36
2.5	The Decay Exponent for the Average Symbol Error Probability	39
2.6	Performance Analysis for QPSK modulation	43
2.6.1	Symbol Error Probability for QPSK modulation	43
2.6.2	Analysis of Quantization Penalty for QPSK Modulation	44
2.7	Numerical Results	45
2.7.1	Effect of Quantization Bits on System Performance	46
2.7.2	Comparison with Independent I and Q Quantization	49
2.7.3	Effect of Channel Estimation Errors on DVO	51
2.8	Conclusions and Future Generalizations	54
2.9	Appendix	55
2.9.1	Proof of Theorem 2.1	55
2.9.2	Proof of Theorem 2.2	56
2.9.3	Proof of Lemma 2.1	60
2.9.4	Proof of Lemma 2.2	63
2.9.5	Proof of Lemma 2.3	64
2.9.6	Proof of Lemma 2.4	66
2.9.7	Proof of Lemma 2.5	67
3	Low-Resolution Quantization with Non-coherent Detection	71
3.1	Introduction	71
3.1.1	Background and Motivation	71
3.1.2	Main Contributions	72
3.2	System Setup	74
3.2.1	Transmitter Architecture	74
3.2.2	Channel Model	75

3.2.3	Receiver Architecture	75
3.3	Optimum Signal Detection	77
3.4	Symbol Error Probability	79
3.5	The Decay Exponent for the Average Symbol Error Probability	80
3.6	Low-Resolution quantized systems with DQPSK modulation under co- herent detection	81
3.6.1	Receiver Architecture with Coherent Detection	81
3.6.2	Optimum Signal Detection with Coherent Detection	82
3.6.3	Symbol Error Probability with Coherent Detection	83
3.6.4	The Decay Exponent for the Average Symbol Error Probability under Coherent Detection	85
3.7	Numerical Results	86
3.8	Conclusions and Future Generalizations	91
3.9	Appendix	93
3.9.1	Proof of Theorem 3.1	93
3.9.2	Proof of Theorem 3.2	100
3.9.3	Proof of Corollary 3.1	102
3.9.4	Proof of Theorem 3.3	103
3.9.5	Proof of Theorem 3.4	108
3.9.6	Proof of Theorem 3.5	109
3.9.7	Proof of Corollary 3.2	110
3.9.8	Proof of Theorem 3.7	111
4	Low-Resolution Quantization in Multi-Antenna Systems	113
4.1	Introduction	113
4.1.1	Background and Motivation	113
4.1.2	Main Contributions	114
4.2	System Setup	116
4.2.1	Channel Model and Signal Modulation	116
4.2.2	Receiver Architecture	117
4.3	Optimum Signal Detection	118

4.3.1	ML Detector with 2-bit Quantization	118
4.3.2	ML Receiver with n -bit Quantization	119
4.4	Antenna Selection and Signal Detection	120
4.4.1	Selection Criterion 1	121
4.4.2	Selection Criterion 2	122
4.4.3	Selection Criterion 3	122
4.5	Average Symbol Error Probability	123
4.5.1	SEP with Selection Criterion 1	125
4.5.2	SEP with Selection Criterion 2	125
4.5.3	SEP with Selection Criterion 3	126
4.6	The Decay Exponent for the Average Symbol Error Probability	126
4.6.1	DVO with Selection Criterion 1	127
4.6.2	DVO with Selection Criterion 2	127
4.6.3	DVO with Selection Criterion 3	128
4.7	Numerical Results	128
4.7.1	System Performance with ML Detector	128
4.7.2	System Performance under Selection Criterion 1	131
4.7.3	System Performance under Selection Criterion 2	131
4.7.4	System Performance under Selection Criterion 3	132
4.7.5	System Performance Comparison	135
4.8	Conclusions and Future Generalizations	137
4.9	Appendix	138
4.9.1	Proof of Theorem 4.1	138
4.9.2	Proof of Theorem 4.2	140
4.9.3	Proof of Theorem 4.3	142
4.9.4	Proof of Theorem 4.4	142
4.9.5	Proof of Theorem 4.5	143
4.9.6	Proof of Theorem 4.6	148
4.9.7	Proof of Theorem 4.7	150
4.9.8	DVO bounds for $N = 2$ and $n = 2$ case with Selection Criterion 3	152

5	Conclusion and Future Work	155
5.1	Summary of Contributions	155
5.2	Future Research	156

List of Figures

1.1	Walden figure of merit (FOM_W) versus the speed of the ADCs. $FOM_W = \frac{P}{2^{ENOB} f_{snyq}}$ where P is the power dissipation, f_{snyq} is Nyquist sampling rate, and ENOB is the effective number of bits defined by the signal-to-noise and-distortion ratio (SNDR). It is evident that the energy efficiency drops rapidly beyond a sampling rate of 100 MHz.	4
1.2	MIMO or Massive MIMO receiver architecture with low-resolution ADCs. The output of each quantizer is used for baseband processing to estimate the transmitted signals.	5
1.3	The receiver architecture with independent I and Q channel quantization for a point-to-point system.	5
1.4	(a) Quantization regions when I and Q components of the received signal are quantized independently using one-bit ADCs (b) Quantization region when I and Q components of the received signal are quantized independently using b -bit quantizer (c) Quantization regions with phase quantizer (d) Quantization regions with the polar quantizer. The received signal is assumed to be landed in the coloured area.	7
1.5	Organization of the literature review	9
1.6	nML detector for massive MIMO systems with one-bit ADCs.	10
1.7	Mixed ADC architectures	15
2.1	The receiver architecture with low-resolution quantization. The signal detector observes only the n -bit quantized versions of Y to estimate the transmitted signal.	28

2.2	An illustration of average SEP calculations. If the noise does not drag the original M -PSK constellation point rotated by the channel h beyond the region $\mathcal{E}$ (shaded area), there will not be any errors in decoding. . . .	35
2.3	An illustration of the centering property for QPSK modulation with 2-bit and 3-bit quantization. Original signal points are indicated by ' $\diamond$ ', whereas the rotated ones after multiplication with $\sqrt{\text{SNR}}$ and h are indicated by ' $\bullet$ '. Quantization region boundaries and the corresponding bisectors are indicated in solid black lines and green dash lines, respectively. The shaded area represents the region of attraction of the transmitted symbol x_3	38
2.4	Average SEP curves as a function of SNR for QPSK modulation. $n = 2, 3, 4$ and $m = 1, 2$	46
2.5	Average SEP curves as a function of SNR for different modulation schemes. $n = \log_2 M, \log_2 M + 1, \log_2 M + 2$ and $m = 1$	47
2.6	Upper and lower bounds on $p(\text{SNR})$ as a function of SNR for QPSK, 8-PSK and 16-PSK modulations. $n = 2, 3, 4, 5$ and $m = 1$	48
2.7	Average SEP as a function of SNR for 8-PSK and 16-PSK modulations. $n = 2 < \log_2 M$ and $m = 0.5, 1$ and 2	49
2.8	Quantization penalty for QPSK modulation and the asymptotic average SEP curves. $n = 2, 3, 4, \infty$ and at $m = 1$	50
2.9	Average SEP as a function of SNR for QPSK modulation with phase quantization and separate I and Q quantization, $m = 1$	51
2.10	Average SEP as a function of SNR for QPSK modulation with 4-bit quantization and imperfect CSI, $m = 1$, b is the number of quantization bits used for channel estimation.	53
2.11	Average SEP as a function of SNR for QPSK modulation with n -bit quantization with and without perfect CSI, $m = 1$, $n = 2, 3, 4, 5$, $b = n, n + 1$, where b is the number of quantization bits used for channel estimation.	53

2.12	An illustration for the proof of Lemma 2.6 when μ_1 and μ_2 lie outside $\mathcal{R}^\circ$. $ \mu_1 = \mu_2 = r$, $ \mu_1 - z_{\text{mid}} \leq \mu_2 - z_{\text{mid}} $ and $\alpha_2 = -\alpha_1 = \alpha$	57
2.13	An illustration for the proof of Lemma 2.6 when $\mu_1 \in \mathcal{R}^\circ$ and $\mu_2 \notin \mathcal{R}^\circ$. $ \mu_1 = \mu_2 = r$, $ \mu_1 - z_{\text{mid}} \leq \mu_2 - z_{\text{mid}} $ and $\alpha_2 = -\alpha_1 = \alpha$	58
2.14	An illustration for the proof of Lemma 2.6 when μ_1 and μ_2 lie inside $\mathcal{R}^\circ$. $ \mu_1 = \mu_2 = r$, $ \mu_1 - z_{\text{mid}} \leq \mu_2 - z_{\text{mid}} $ and $\alpha_2 = -\alpha_1 = \alpha$	59
2.15	An illustration for the proof of Theorem 2.2 where $ hx^* - z_k \leq hx - z_k $	61
3.1	The transmitter architecture with the DQPSK modulation scheme. The modulator generates the symbol to be transmitted at time t according to (3.1).	74
3.2	The receiver architecture with low-resolution quantization with differential modulation. The signal detector observes only the n -bit quantized versions of Y_t and Y_{t-1} to estimate the transmitted message signal $\hat{M}_t$	76
3.3	The block-2 detector architecture with low-resolution quantization with differential modulation. The signal detector uses two consecutive n -bit quantized received signals $Q(Y_{t-1})$ and $Q(Y_t)$ to estimate the message $\hat{M}_t$	77
3.4	The receiver architecture with low-resolution quantization with differential modulation when perfect channel knowledge is available at the receiver. The message at time t , $\hat{M}_t$ will be decided after two steps. The signal detector first uses the n -bit quantized versions of Y_{t-1} and Y_t to estimate the transmitted signals $\hat{S}_{t-1}$ and $\hat{S}_t$ respectively. These estimated transmitted signals will be compared to estimate the message $\hat{M}_t$	81
3.5	Average SEP curves as a function of SNR for DQPSK modulation with non-coherent detection. $n = 2, 3, 4, 5, \infty$	86
3.6	Average SEP curves as a function of SNR for DQPSK modulation with coherent detection. $n = 2, 3, 4, 5, \infty$	88

3.7	Average SEP curves as a function of SNR for coherent and non-coherent detection with DQPSK modulation. $n = 2, 3, 4, 5$. For $n = 2$ -bit quantization, both coherent and non-coherent detection show the same SEP behaviour with respect to average SNR. It is also visible that the full DVO is achieved with $n = 3$ under coherent detection. On the other hand, it requires $n = 4$ for non-coherent detection to achieve full DVO.	89
3.8	Average SEP curves as a function of SNR for QPSK modulation and DQPSK modulation with non-coherent detection. $n = 2, 3, 4, 5$	90
3.9	Average SEP curves as a function of SNR for QPSK modulation and DQPSK modulation with coherent detection. $n = 2, 3, 4, 5$	91
3.10	The decoding process of the ML detector with 2-bit quantization with differential modulation.	94
3.11	An illustration for the proof of Theorem 3.1 using the detection rule 1. (a) $n = 3$ (b) $n = 4$. Given that, $Y_{t-1} \in \mathcal{R}_0$, $E[Y_t Y_{t-1}]$ should be in either one of the blue coloured regions depending on the value for ϕ . Given $Y_t \in \mathcal{E}_0$, $E[Y_t Y_{t-1}]$ become closest to the bisector $\mathcal{H}_{\mathcal{E}_0}$ of the region $\mathcal{E}_0$ when $\phi = 0$	99
3.12	An illustration for the proof of Theorem 3.1 using the detection rule 2. (a) $n = 3$ (b) $n = 4$. Given that, $Y_{t-1} \in \mathcal{R}_0$, $E[Y_t Y_{t-1}]$ should be in either one of the blue coloured regions depending on the value for ϕ . Given $Y_t \in \mathcal{E}_0$, $E[Y_t Y_{t-1}]$ become closest to the bisector $\mathcal{H}_{\mathcal{E}_0}$ of the region $\mathcal{E}_0$ when $\phi = 0$	99
3.13	An illustration of average SEP calculations for $n \geq 3$ case assuming that $M_t = m_0$. (a) $n = 3$. Given that, $Y_{t-1} \in \mathcal{R}_1$, if $Y_t \in \mathcal{E}_1$, there will be no error in the estimating the message $\hat{M}_t$. (b) $n = 4$. Given that, $Y_{t-1} \in \mathcal{R}_2$, if $Y_t \in \mathcal{E}_2$, there will be no error in the estimating the message $\hat{M}_t$. Without loss of generality, we choose k' in which $Y_{t-1} \in \mathcal{R}_{k'}$ such that $Y_t \in \mathcal{E}_{k'}$ where $\mathcal{E}_{k'} = \{z \in \mathbb{C} : 0 \leq \text{Arg}(z) < \frac{\pi}{2}\}$	101

3.14	An illustration of average SEP calculations for $n = 3$. Due to the symmetry of the problem, $\Pr \{Y_{t-1} \in \mathcal{R}_1, Y_t \in \mathcal{E}_1\} = \Pr \{Y_{t-1} \in \mathcal{R}_1, Y_t \in \mathcal{E}_0\}$. By combining (a) and (b) together, we evaluate the probability of $\Pr \{Y_{t-1} \in \mathcal{R}_0 \cup \mathcal{R}_1, Y_t \in \mathcal{E}_0\}$	103
4.1	The receiver architecture with low-resolution quantization. The signal detector observes only the n -bit quantized versions of Y_l to estimate the transmitted signal. $l = 0, 1, \dots, N-1$	117
4.2	This is an example case for $n = 2$ case. We assume $x = x_0$ (indicated by the red dot) has been transmitted and $\text{Arg}(H_l) \in [-\frac{\pi}{4}, \frac{\pi}{4}]$. The transmit signal x_0 has been moved to $\sqrt{\text{SNR}}h_l x_0$ (indicated by the blue dot) by the channel $H_l = h_l$ in path l . Since, in this case the region of attraction is $\mathcal{R}_0$, the distances to the decision boundaries are $ \text{Re}(\sqrt{\text{SNR}}h_l x_0) $ and $ \text{Im}(\sqrt{\text{SNR}}h_l x_0) $. The detector compares these distances for all l and chooses the path according to (4.12).	123
4.3	Average SEP curves as a function of SNR with QPSK modulation and ML detection. $n = 2, 3, 4, \infty$ and $N = 2, 3, 4$. We note that the non-smooth variations of the curves in the high SNR regime are a result of numerical errors in the simulation.	129
4.4	Average SEP curves as a function of SNR with QPSK modulation and selection combining with selection criterion 1. $n = 2$ and $N = 2, 3, 5, 10$	130
4.5	Average SEP curves as a function of SNR with QPSK modulation and selection combining with selection criterion 1. $n = 2, 3, 4$ and $N = 2, 3, 5$	130
4.6	Average SEP curves as a function of SNR with QPSK modulation and selection combining with selection criterion 2. $n = 2$ and $N = 2, 3, 5, 10, 50$	132
4.7	Average SEP curves as a function of SNR with QPSK modulation and selection combining with selection criterion 2. $n = 2, 3, 4$ and $N = 2, 3, 5, 10, 50$	133
4.8	Average SEP curves as a function of SNR with QPSK modulation and selection combining with selection criterion 3. $n = 2$ and $N = 2, 3, 4$	134

4.9	Average SEP curves as a function of SNR with QPSK modulation and selection combining with selection criterion 3. $n = 2, 3, 4$ and $N = 2, 3, 4$.	134
4.10	Average SEP curves as a function of SNR with QPSK modulation and selection combining with different selection criteria. $n = 2, 3, 4$ and $N = 2$.	135
4.11	Average SEP curves as a function of SNR with QPSK modulation under the ML detector and selection combining with selection criterion 3. $n = 2, 3, 4$ and $N = 2$.	136
4.12	Average SEP curves as a function of SNR with QPSK modulation under the ML detector, selection combining with selection criterion 3 and majority combining. $n = 2, 3, 4$ and $N = 3$. We note that the non-smooth variations of the curves in the high SNR regime are a result of numerical errors in the simulation.	136
4.13	This is an example case. We assume the received signal at path l has been landed in $\mathcal{R}_0$. The transmitted signal x (denoted by the red dot) has been rotated by the channel h_l to the point $\sqrt{\text{SNR}}h_l x$ (denoted by the blue dot). Unless the noise W_l drags $\sqrt{\text{SNR}}h_l x$ to a different quantization region other than $\mathcal{R}_0$, there will be no errors in estimating the transmitted signal. As such, there will be no error if $Y_l^{\text{re}} \geq 0$ and $Y_l^{\text{im}} \geq 0$.	139
4.14	An illustration of the average SEP calculation under the selection strategy 3. Without loss of generality, let us assume that the x_0 is transmitted and the region of attraction is $\mathcal{R} = \{z \in \mathbb{C} : 0 \leq \text{Arg}(z) \leq \frac{\pi}{2}\}$. Then, the distances to the decision boundaries are $Z_{L^*}^{\min} = u$ and $Z_{L^*}^{\max} = v$. Using polar coordinates, we can represent these distances as $Z_{L^*}^{\min} = \sqrt{\text{SNR}}r^* \cos \theta^*$ and $Z_{L^*}^{\max} = \sqrt{\text{SNR}}r^* \sin \theta^*$ where $r^* = h^* $ and $\theta^* = \text{Arg}(h^*x_0)$.	146

Glossary

List of Abbreviations

ADC	Analog-to-Digital Converter
BEP	Bit Error Probability
BPSK	Binary Phase Shift Keying
CDF	Cumulative Distribution Function
CSI	Channel State Information
DAC	Digital-to-Analog Converter
DQPSK	Differential Quadrature Phase Shift Keying
DVO	Diversity Order
MAP	Maximum A-Posteriori Probability
MIMO	Multiple-Input-Multiple-Output
MISO	Multiple-Input-Single-Output
ML	Maximum Likelihood
M-PSK	M-ary Phase Shift Keying
MRC	Maximal Ratio Combining
PDF	Probability Density Function
QPSK	Quadrature Phase Shift Keying
SEP	Symbol Error Probability
SIMO	Single-Input-Multiple-Output
SISO	Single-Input-Single-Output
SNR	Signal-to-Noise Ratio

Notations

$\mathbb{R}$	Real line
$\mathbb{R}^2$	2-dimensional Euclidean space
$\mathbb{N}$	Natural numbers
$\mathbb{C}$	Set of complex numbers
$[i : j]$	Discrete interval $\{i, i + 1, \dots, j\}$ for a pair of integers $i \leq j$
$\text{Arg}(Z)$	(Principle) argument of $Z \in \mathbb{C}$
$\Pr\{\cdot\}$	Probability
$E[X]$	Mean of the random variable X
$\text{Var}(X)$	Variance of the random variable X
$\log x$	Natural logarithm of x for $x > 0$
$\log_2 x$	Logarithm of x in base 2 for $x > 0$
$f_X(\cdot)$	Probability Density Function of the random variable X
$F_X(\cdot)$	Cumulative Distribution Function of the random variable X
$\mathcal{N}(\mu, \sigma^2)$	Gaussian distribution with mean μ and variance σ^2
$\mathcal{CN}(\mu, \sigma^2)$	Complex Gaussian distribution with mean μ and variance σ^2
$\mathcal{Q}(\cdot)$	Complementary distribution function of the standard normal random variable
$\text{erf}(\cdot)$	Error function
$\text{erfc}(\cdot)$	Complementary error function
$\Gamma(\cdot)$	Gamma function

Chapter 1

Introduction

Mobile wireless communication has evolved a long way since its introduction in the late seventies. From there onward, the world has experienced five successive generations of wireless standards, namely, first-generation (1G) to fifth-generation (5G), up to now and we are expecting to move to its sixth generation around 2030. When we move from one generation to another, the main focus is to improve the data-carrying capacity and decrease the latency. Let us first outline the key breakthroughs along the journey of wireless standards from 1G to where we are today, on the verge of a global 5G deployment.

All these began with the decision of the World Administrative Radio Conference (WARC) to allocate the frequency band between 890 and 960 MHz for mobile communications in 1979 [1]. Later in the same year, the first 1G mobile network was launched by Nippon Telegraph and Telephone (NTT) in Tokyo. About a decade later, in 1991 the first commercial 2G mobile network was launched with Global System for Mobile Communications (GSM) in Finland. It replaced the 1G analog systems with revolutionized digital systems where for the first time, mobile users could send text messages (SMS), picture messages and multimedia messages (MMS) from their phones. This made mobile communications very popular among the general public as well as the business world. The first commercial 3G network was launched by NTT DoCoMo in Japan in 2001 [2]. This introduced significant improvements in data transmission capabilities compared to 2G and it opened up a vast set of new services such as video conferencing, video streaming and voice over IP, in addition to the international roaming facilities. 4G was first deployed in Stockholm, Sweden and Oslo, Norway in 2009 as the Long

Term Evolution (LTE) standard. It allowed users to use high-quality video streaming by offering fast mobile internet access. Global 4G population coverage was around 80 percent at the end of 2019 and is forecast to reach over 90 percent in 2025 [3]. In April 2019, South Korea deployed the first commercial 5G network. To date, there have been more than 75 5G commercial launches across the world with a global 5G population coverage of 5 percent at the end of 2019. It has been predicted that 5G internet will cover up to 65 percent of the world population by the end of 2025 [3].

In recent years, the demand for mobile data has been increasing at an unprecedented pace, with most of that coming from video traffic. With a forecast of 31 percent annual growth of mobile traffic between 2019 to 2025 [3], mobile networks will need to provide as much as 1000 times the capacity compared to the current systems [4]. On the other hand, it is predicted that there will be more than 24.9 billion Internet of Things (IoT) connections by 2025 compared to the estimated human population of 8.1 billion with the development of massive IoT, vehicular-to-everything (V2X) technologies [3]. With over 30 percent of millennial smartphone users' primal expectation from 5G being a factor beyond better speed and coverage [5], achieving all these demands will be a formidable challenge.

To cope with the projected increase in the demand for capacity, there has been a growing interest in multiple mobile transmission technologies including massive multiple-input-multiple-output (MIMO) and millimeter wave (mmWave) communications. The main idea behind massive MIMO is to use many antennas (i.e., hundreds or even more) at the base station, which provides a large number of spatial degrees of freedom for the simultaneous serving of multiple user terminals on the same time-frequency channel [6]. On the other hand, mmWave communication aims to use the frequency band from 30 GHz to 300 GHz, where wider bandwidths for data transmission can be utilized. Thus, mmWave communication can offer more than 200 times the spectrum when compared to the current cellular allocations [4]. Despite some challenges [4], both technologies are well capable of providing the data rates required by next-generation wireless communication networks.

During the past decade, energy efficiency emerged as another important perfor-

mance measure in communication networks. This is mainly because telecommunication networks are responsible for approximately 2 percent of global energy consumption [7] and 0.2 percent of global carbon emissions [8]. Since this figure may continue to grow day by day at a rapid pace with the increase of mobile traffic, developing energy-efficient alternatives or “green” communications while keeping the spectral efficiency has become the norm of current research in wireless communications.

Analog-to-digital converters (ADCs) are known to consume a significant amount of power dissipated at a base station receiver [9]. It is shown that the power consumed by ADCs grows exponentially with their resolution level and linearly with their sampling rate [4, 10–12]. We use Fig. 1.1¹ to illustrate these variations using the data available for commercial ADCs. Thus, using high-resolution quantization with high sampling rates can significantly degrade the energy efficiency of a communication system. With the introduction of massive MIMO and millimeter wave (mmWave) technology, this is even more prominent in next-generation wireless systems. Because massive MIMO systems use hundreds of antennas where each antenna is connected to a dedicated radio frequency (RF) chain equipped with high-resolution ADCs.

MmWave systems, on the other hand, use much larger bandwidths that require higher sampling rates. Since mmWave uses larger bandwidths, the corresponding sampling rate of the ADC also goes up. As shown in Fig. 1.1, the power consumption of ADCs increases quadratically with sampling frequency beyond the sampling rates of 100 megasamples per second. In fact, the typical power consumption of a high speed (≥ 20 GSamples/s) and high-resolution (8-12 bits) ADC is around 500 mW. As an RF chain contains one ADC for the in-phase (I) component and another ADC for the quadrature (Q) component of the received signal, the total power consumption of the ADCs in a single RF chain is around 1 W. Therefore, a future mmWave massive MIMO system with 256 RF chains and 512 ADCs will require around 256 W of a significant amount of power [13].

To overcome this challenge, one can consider two possible solutions. The first solution is to use high-resolution but low-speed sub-ADCs, whereas the second one

¹Fig. 1.1 source: [10]. Accessed Sep. 2020.

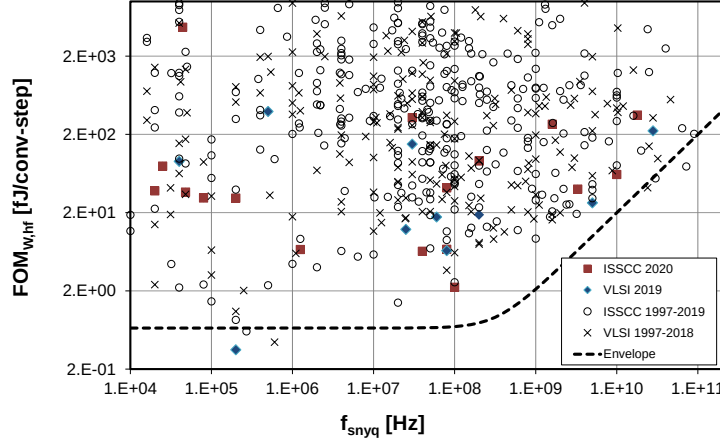


Figure 1.1: Walden figure of merit (FOM_W) versus the speed of the ADCs. $FOM_W = \frac{P}{2^{ENOB} f_{snyq}}$ where P is the power dissipation, f_{snyq} is Nyquist sampling rate, and ENOB is the effective number of bits defined by the signal-to-noise and-distortion ratio (SNDR). It is evident that the energy efficiency drops rapidly beyond a sampling rate of 100 MHz.

is to use high-speed but low-resolution ADCs. Since high-resolution but low-speed sub-ADC architecture introduces error floors due to mismatches among sub-ADCs [13], the research in the field has been more focused on the second solution. Consequently, the idea of replacing power-hungry high-resolution ADCs with low-resolution ADCs could provide a viable solution to the power consumption concerns in future wireless systems.

1.1 Low-Resolution Quantization

Fig. 1.2 illustrates a typical receiver architecture of a low-resolution quantization based MIMO or massive MIMO system. Each receive RF chain contains a quantizer (with ADCs), demodulator, down-converter, low noise amplifier (LNA), mixers, automatic gain control (AGC), variable gain amplifier (VGA), and some filters. The VGA and AGC will make sure that the power of the analog signal is within the required range for the quantizer. Similar circuitry is employed at the transmit side [13].

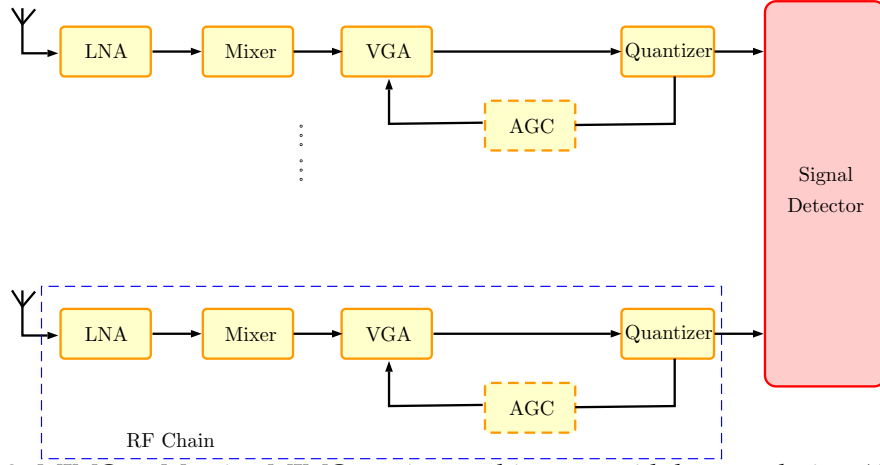


Figure 1.2: MIMO or Massive MIMO receiver architecture with low-resolution ADCs. The output of each quantizer is used for baseband processing to estimate the transmitted signals.

We now focus on the low-resolution Quantizer which is the central module relevant to our work. Each quantizer outputs the quantized observations of its received signal to the detector. The detector then decodes the transmitted signals based on those quantized observations. There are multiple ways of implementing the low-resolution quantizers. Out of those, the most common situation is quantizing the in-phase (I) and quadrature (Q) channels independently as illustrated in Fig. 1.3. To that end, most

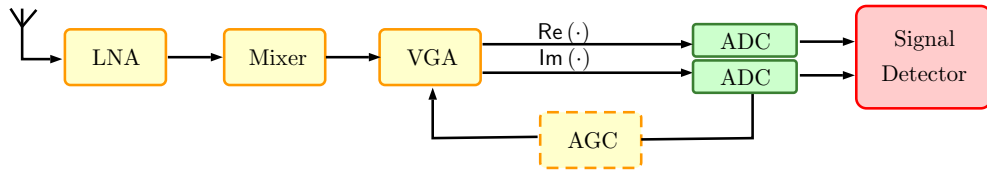


Figure 1.3: The receiver architecture with independent I and Q channel quantization for a point-to-point system.

of the systems available in the literature [14–28] use extremely low-resolution one-bit ADC, one for each I and Q components of the received signal. In this case, the quantizer only outputs the quadrant on which the received signal lands. Thus, the quantized received signal $\hat{Y}$ can be represented as

$$\hat{Y} = \text{sgn}(\text{Re}(Y)) + j \text{sgn}(\text{Im}(Y)), \quad (1.1)$$

where Y is the received signal and $\text{sgn}(a)$

$$\text{sgn}(a) = \begin{cases} 1 & \text{if } a \geq 0 \\ -1 & \text{if } a < 0. \end{cases} \quad (1.2)$$

Similarly, two low-resolution b -bit quantizers ($b \geq 2$) can be used separately to quantize the I and Q channels. In that, each quantizer maps an input to one of 2^b quantization regions which are characterized by the quantization boundaries $-\infty = q_0 < q_1 < q_2 < \dots < q_{2^b-1} < q_{2^b} = \infty$. More specifically, if b bits are used to quantize each of $\text{Re}(Y)$ and $\text{Im}(Y)$, the quantizer Q outputs the index of the region in which $\text{Re}(Y)$ and $\text{Im}(Y)$ lie as an input to the detector. As such, we declare $Q(\text{Re}(Y)) = k$ if $\text{Re}(Y) \in (q_k, q_{k+1}]$ for $k \in [0 : 2^b - 1]$ where we use $[i : j]$ to denote the discrete interval $\{i, i+1, \dots, j\}$ for a pair of integers $i \leq j$. The work related to multi-bit low-resolution quantization can be found in [29, 30].

Another possible architecture for the quantizer is using phase quantizers which quantizes only the phase of the received signals as discussed in [31–34]. More specifically, if n bits are used to quantize Y , the quantizer Q divides the complex domain $\mathbb{C}$ into 2^n quantization regions and outputs the index of the region in which Y lies as an input to the detector. As such, we declare $Q(Y) = k$ if $Y \in \mathcal{R}_k$ for $k \in [0 : 2^n - 1]$, where $\mathcal{R}_k \subseteq \mathbb{C}$ is the k -th quantization region given by

$$\mathcal{R}_k = \left\{ z \in \mathbb{C} : \frac{2\pi}{2^n} k \leq \text{Arg}(z) < \frac{2\pi}{2^n} (k+1) \right\}. \quad (1.3)$$

Another quantizer architecture is using polar quantization in which the received signal is quantized based on both its phase and magnitude. First, the magnitude range is partitioned into L magnitude levels indexed by $[0 : L - 1]$ and then each magnitude level will be sub-divided to P phase levels indexed by $[0 : P - 1]$ depending on the phase of the received signal. The quantizer outputs these two indices. The work related to polar quantization can be found in [35–37].

Fig. 1.4 illustrates the different quantization regions used by the quantizers that we discussed above.

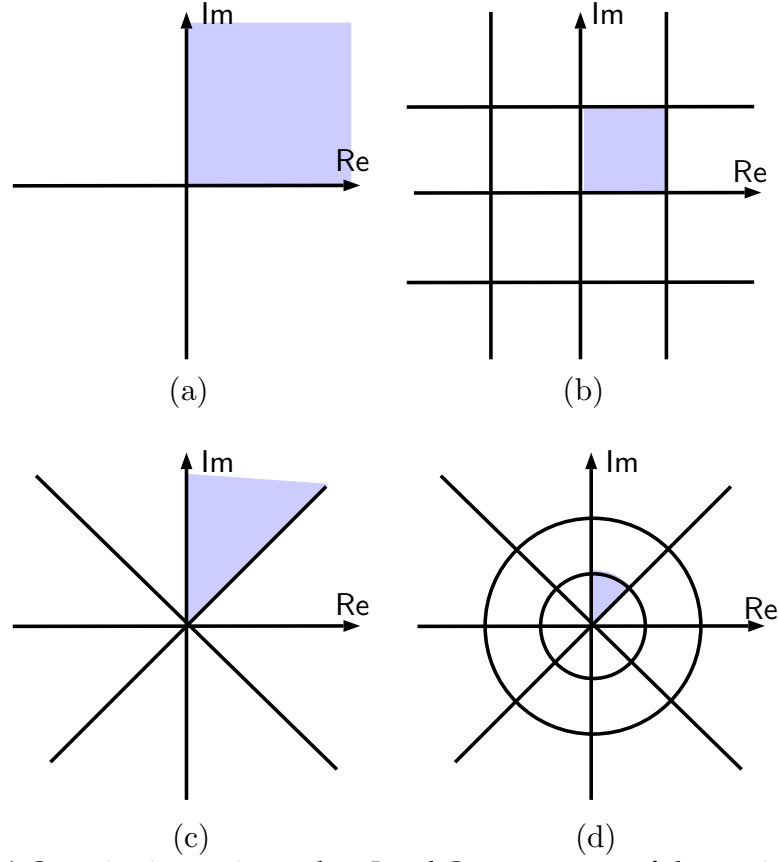


Figure 1.4: (a) Quantization regions when I and Q components of the received signal are quantized independently using one-bit ADCs (b) Quantization region when I and Q components of the received signal are quantized independently using b -bit quantizer (c) Quantization regions with phase quantizer (d) Quantization regions with the polar quantizer. The received signal is assumed to be landed in the coloured area.

Indeed, low-resolution ADCs have long been known to provide significant energy savings in digital transceiver implementations [18,38,39]. Their other benefits include simplification in design (especially with 1-bit ADCs) and reduction in transceiver form-factor [13,16,40]. On the other hand, using low-resolution ADCs will limit the amount of data that has to be transferred over the link that connects the RF components and the baseband-processing unit [29]. For example, let us consider a base station that is equipped with 256 antennas. At each antenna element, the in-phase and the quadrature components are sampled with 10-bit ADCs operating at 20 GSamples/s. Such a system produces over 10 Tbit/s of data. However, this data rate exceeds the rate supported by the common public radio interface (CPRI) used over today's fiber-optical fronthaul links [41]. This is also a severe problem in cloud radio access network (C-RAN) architectures

where the baseband processing is moved away to a central unit from the base stations [42]. Furthermore, the future long-term evolution (LTE) networks are also expected to support a wide range of Internet-of-Things (IoT) applications through protocols such as LTE-M, NB-IoT and EC-GSM, where devices are usually battery power-limited [43]. In these future application scenarios, low-resolution ADC based digital transceivers can prolong the battery lifetime of remote IoT devices as well, and thereby lessening the operating costs and the need for frequent human intervention.

Motivated by the potential of low-resolution ADCs, in this work, we aim to understand the fundamental trade-off between using low-resolution quantization and achieving higher communication reliability. More specifically, we analyse the error probability performance of a communication system, when the receiver is equipped with a phase quantizer.

Although the use of low-resolution quantization at the receiver can significantly improve the energy efficiency of a communication system, there are a certain set of challenges that still need to be answered. In the next section, we discuss these challenges under five main areas and present the related work available in the literature.

1.2 Related Work

Random House Webster's dictionary defines quantization as "the division of a quantity into a discrete number of small parts, often assumed to be integral multiples of a common quantity [44]." Although the idea of quantization appears in literature as long as 1898, the first recognition of quantization as a key role player in communication systems with modulation and analog-to-digital conversion was identified in 1948. This was from the paper published by Oliver, Pierce, and Shannon. Starting from the very early stages, the development of quantization techniques in chronological order can be found in [44].

With the aim of developing communication systems with low power consumption and low complexity, many scholars have investigated the ability to use low-resolution ADCs in various aspects of a wireless communication system. As illustrated in Fig. 1.5,

the following literature review is carried out focusing on these different aspects.

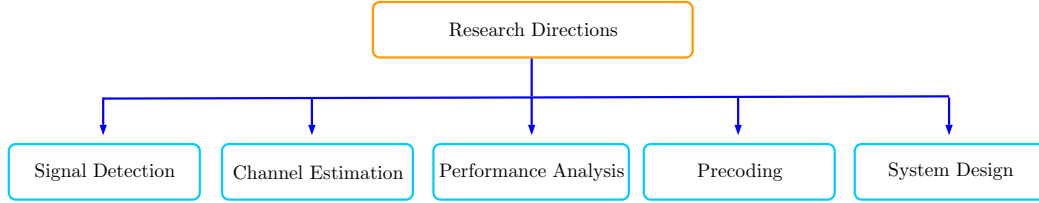


Figure 1.5: Organization of the literature review

1.2.1 Signal Detection

Under coherent detection, the signal detector aims to estimate the transmitted signals by minimizing the symbol error probability (SEP) using the knowledge of quantized observations and channel state information (CSI). In the case of multiple antenna systems with unquantized or high-resolution quantization, signal detectors such as maximal-ratio combining (MRC), Zero Forcing (ZF) [45–47], Least Squares (LS), minimum mean square error (MMSE) [48–50], maximum likelihood (ML) [51–53] and message passing detector are commonly used. However, these existing detectors may not be able to be used directly in low-quantization based systems due to the significant degradation of performances [13, 29]. Therefore, many scholars have paid their attention to develop novel signal detectors counting the effects of low-resolution quantization.

To that end, [14] presents an optimal ML detector and a low-complexity ZF receiver for a low-resolution quantization based system with one-bit ADCs. In this case, the complexity of the ML detector grows exponentially with high signal constellations, the number of transmit antennas and the number of users in the uplink. On the other hand, the sub-optimal ZF-type receiver is simple to implement and shows comparable performance with the ML receiver in the low signal-to-noise ratio (SNR) regime, however it experiences an error rate floor at high SNR. Due to the high complexity of the ML detector, [15] presents a near maximum likelihood (nML) detector which is based on convex optimization techniques. Numerical examples show that the nML detector is capable of performing well even with fewer antennas. Fig. 1.6² illustrates

²Fig. 1.6 is recreated using [13] ©2018 IEEE

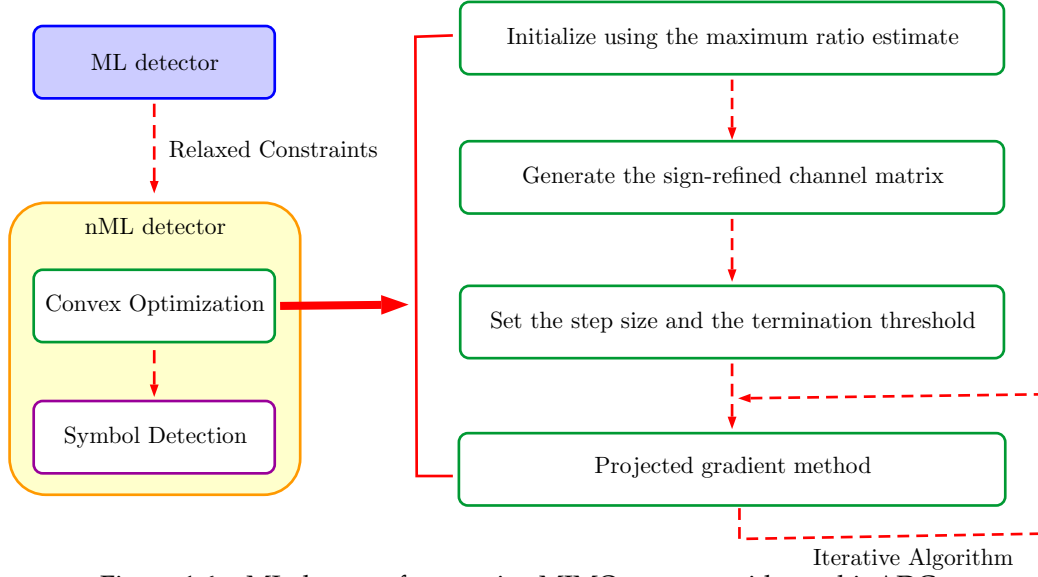


Figure 1.6: nML detector for massive MIMO systems with one-bit ADCs.

the functionality of the nML detector using [13].

In [54], the authors present another sub-optimal detector (naive-ML) for a low-resolution quantization based MIMO system. This detector is based on the distance measures to the quantization regions and it shows almost the same performance as the optimum ML detector in the high SNR regime in terms of bit error rate (BER). In [55], the authors present an MMSE multiuser detector for uplink large-scale MIMO systems with one-bit ADCs. The detector is based on the Message Passing De-Quantization algorithm. An algorithm for uplink multiuser detection in a massive MIMO system using quantized phase only measurements is presented in [56], whereas [57] presents a convex optimization based multiuser detection technique for uplink large-scale MIMO systems with low-resolution quantization. In [58], the authors present a low-complexity nML detector called *one-bit sphere decoding* for an uplink MIMO system with one-bit ADCs. The detection algorithm estimates the transmitted symbol vector sent by uplink users (a codeword vector) by searching over a sphere, which contains a collection of codeword vectors close to the received signal vector at the base station in terms of a weighted Hamming distance.

In [59], the authors propose a linear minimum mean square error (LMMSE) receiver when in-phase and quadrature components of the received signal are independently

quantized by using a low-resolution ADC. They provide an approximation for the mean squared error between the transmitted symbol and the received one, and derive an optimized linear receiver that performs better than the conventional Weiner filter. Results in [59] were further extended to an iterative decision feedback receiver with quantized observations in [60].

For a massive MIMO system with one-bit ADCs, a sphere decoding based low-complexity near-ML detection method is presented in [61]. In the proposed algorithm, which is called one-bit-sphere decoding (OSD), the complexity of the construction of the sphere is reduced by dividing the received signal vector into multiple subvectors. The authors discuss the trade-off between performance and complexity with the dimension of the sphere and the length of the subvectors. In [62], the authors focus on the development of soft-output detection methods for low-precision ADCs and propose another near-optimal detection method for a coded mmWave MIMO system. More recently, there is progress in machine learning based approaches as well [63–65]. In [63,64] a reinforcement learning approach is used to design a robust likelihood function learning method for MIMO systems with one-bit ADCs. For a similar system, a semi-supervised learning detector is proposed in [65] which is further improved to an online-learning detector.

1.2.2 Channel Estimation

Many of the low-resolution quantization based detectors that we come across in the literature need accurate CSI at the receiver to reach their full performance. However, the nonlinearities of low-resolution ADCs degrade the performance of conventional channel estimation techniques [13]. Therefore, estimating the CSI plays an important role in low-resolution quantization based communication systems. In this sub-section, we briefly provided a discussion on some such channel estimation techniques.

An expectation maximization (EM) algorithm is presented in [66] for a MIMO system with one-bit ADCs. The EM algorithm computes the minimum mean square estimate of the quantized received signal iteratively until it converges and hence it has high computational complexity. The numerical results show that the performance

of the channel estimator compared to the unquantized case in terms of mean square error (MSE), is relatively good in low SNR regime. In [16], a modified EM algorithm is proposed for channel estimation for a mmWave MIMO system with one-bit ADCs. This estimator uses the property of sparsity in the channel to reduce the computational complexity.

In [18], the channel estimation problem for a MIMO system with one-bit ADCs is converted to a channel estimation problem for a single-input-single-output (SISO) system using orthogonal pilots to find a closed-form ML channel estimate. In [38], an iterative method is presented to find the ML channel estimate using the quantized observations. As evaluated in [15], the channel estimation problem in both [18] and [38] are not convex and the algorithm may converge to a local optimum in high SNR regime. Table 1 of [13] presents a comparison of channel estimation algorithms [15, 16, 29, 30, 66] which are commonly used in literature.

1.2.3 Performance Analysis

The performance of a low-resolution quantization based system is relatively low compared to a classical unquantized system due to unavoidable reasons. To measure the performance of low-resolution quantization based systems we can use metrics such as channel capacity, achievable rate, SEP and spectral efficiency. In this sub-section, we briefly discuss some of the key performance related findings of low-resolution quantization based systems.

There are many research papers investigating the capacity of communication systems with one-bit ADCs under the additive quantization noise model (AQNM) [9, 24, 67]. However, it is very difficult to find a general expression for capacity for a communication system with multi-bit low-resolution quantization [68]. In [23], authors present bounds on capacity in high SNR regime for a single-input-multiple-output (SIMO) and a MIMO system with one-bit ADCs under flat fading. The results show how the number of paths, number of transmit antennas, and number of receive antennas impact the capacity at high SNR regime. Further, it shows that the capacity in high SNR regime is lower bounded by the rank of the channel. The work in [23] is further extended

in [24] to obtain a closed-form expression for multiple-input-single-output (MISO) channel capacity assuming that CSI is available at both the transmitter and receiver.

Mutual information between the inputs and the quantized outputs of a one-bit ADC based MIMO system is in [25] and [26]. Using the second-order expansion of the mutual information, it is showed that the power penalty due to one-bit quantization is approximately equal to $\frac{\pi}{2}$ (1.98 dB). This confirms that low-resolution quantization performs adequately in low SNR regime while reducing power consumption. Numerical results in [69] show that for a phase-quantized block non-coherent channel with 8-bin phase quantization under quadrature phase shift keying (QPSK) modulation can recover about 80-85 percent of mutual information attained with unquantized observations when SNR is larger than 2-3 dB.

In [27], the achievable rate of a downlink of a massive MIMO system with one-bit ADCs is analyzed. The results show that it is enough to have approximately 2.5 times more antennas at the base station to overcome the loss due to one-bit ADCs.

1.2.4 Precoding

The majority of research on communication systems based on low-resolution quantization is in the direction of the uplink where the base station is equipped with low-resolution ADCs. However recently, a considerable amount of research attention has been paid to downlink communication. Most of the existing work on MIMO downlink precoding has assumed that each RF chain is equipped with high-resolution digital-to-analog converters (DACs). Similar to ADCs, the power consumption of DACs increases exponentially with their resolution and hence using low-resolution DACs in precoding has the potential to save a considerable amount of power in the communication system.

In [24], the authors show that when the channel matrix has a full row rank with a small condition number, channel inversion precoding becomes optimum. Papers [20] and [21] present low complexity ZF precoders for downlink massive MIMO multiuser system with one-bit DACs. Numerical results of [20] show that the quantized ZF precoder outperforms the ML encoder for low to moderate SNR. In [21], the authors

provide asymptotic symbol error rate expressions using the Bussgang theorem at each terminal for the above-mentioned setup. However, there is an error floor at the medium-to-high SNR regime due to the low-resolution quantization. To improve this, [22] presents a method by perturbing the transmitted signal before quantization.

1.2.5 System Design

When designing wireless communication systems, the most commonly used performance metric is spectral efficiency. However, with the emergence of the green communication concept a few years back, developing energy-efficient communication systems has also gained a considerable amount of interest [6]. Though there is a trade-off in achieving both spectral and energy efficiency at the same time [70], joint optimization of them has considered in recent years. To that end, in [70], the authors consider the joint optimization of the power amplifier and a pulse shaping filter over a point-to-point additive white Gaussian noise (AWGN) channel using one-bit ADCs and DACs. Using the product of power efficiency and spectral efficiency as the objective function, the optimization problem finds the optimal system parameters. In [6], the authors investigate the optimal design of single-cell massive MIMO systems with one-bit ADCs. By analyzing the spectral and energy efficiency trade-off, it shows that a system with one-bit ADCs needs 2-2.5 times more antennas to achieve the same ratio of spectral efficiency to energy efficiency as a classical unquantized system.

To avoid some of the adverse effects of low-resolution quantization, some of the research work has carried out on mixed-ADC architectures [71–74]. Mixed-ADC architectures contain a mixture of both high-resolution and low-resolution ADCs. This architecture is expected to inherently immune to the capacity loss in high SNR regime, to the high pilot overhead for channel estimation and to the error floors in multiuser detection [13]. In this, channel estimation can be done using the high-resolution quantizers while the data decoding phase can be carried out using the low-resolution quantizers. Further, in [71], it is presented a mixed-ADC architecture which both one-bit ADCs and high-resolution ADCs for a massive MIMO system. Results show that mixed-ADC architecture can achieve a near similar performance as conventional

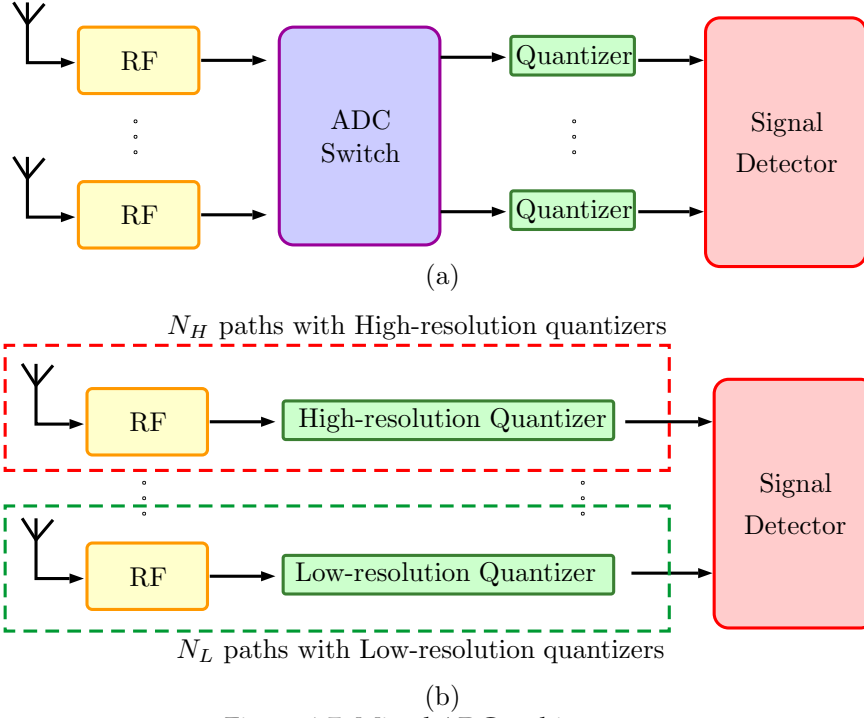


Figure 1.7: Mixed ADC architectures

architecture while reducing the power consumption considerably.

Fig. 1.7 (a) and (b) illustrate two proposed mixed ADC architectures in [71] and [74] respectively. Fig. 1.7 (a) uses an ADC switch module before the quantizers in each path. Fig. 1.7 (b) has receivers with N_H number of high-resolution quantizers and N_L number of low-resolution quantizers, where $0 \leq \frac{N_H}{N_L} \leq 1$. In each case, we can use a high-resolution section for channel estimation and a low-resolution section for data decoding [13].

1.3 Focus of the Thesis

Motivated by the potential of low-resolution ADCs, in this work, we aim to understand the fundamental trade-off between low-resolution quantization and achieving higher communication reliability. More specifically, we conduct a comprehensive analytical study on the performance impact of using low-resolution phase quantization at the receiver of a wireless communication system, when compared to traditional high-resolution systems.

In this thesis, we study the optimum and sub-optimum detectors and error rate analysis for wireless communication systems with low-resolution quantizers in the presence of fading and noise. We adopt phase shift keying modulation under Nakagami- m and Rayleigh faded wireless channels with additive white Gaussian noise (AWGN). We derive optimum ML detectors and sub-optimum detectors and based on these detectors, we conduct rigorous analysis on the average SEP performance which provides a measure of network reliability and the decay exponent of the average SEP which helps to discover the first-order effects of wireless communication systems with low-resolution ADCs. We first focus on point-to-point systems under both coherent and non-coherent detection. This work is further extended to more general multi-antenna systems later. We establish a fundamental ternary behaviour in the average SEP performance of a wireless communication system with low-resolution ADCs and phase modulation. Further, we draw fundamental insights on the optimum number of quantization bits that minimizes the performance gap between high-resolution and low-resolution quantization based wireless communication systems.

1.4 Main Thesis Contributions

The thesis is presented in three main chapters. Each chapter contains a detailed introduction, theoretical analysis and numerical results. Below, we provide a chapter-by-chapter summary of the thesis contributions. Further, a detailed version of the contributions can be found at the beginning of each chapter.

Low-resolution Quantization in Single-Input-Single-Output Systems

In Chapter 2, we consider a point-to-point wireless communication system, in which the receiver is equipped with a low-resolution ADC that quantizes the phase of the received signal. We note that the SISO system provides the basic building block of a more complex multi-antenna system. We adopt the SISO system to draw fundamental insights. It allows us to conduct a rigorous mathematical investigation to provide exact analytical expressions for the error probability performance. The performance results

obtained in this chapter hold in the general case of n -bit quantization, M -ary phase shift keying (M -PSK) modulation and Nakagami- m fading channels. We note that most of the work on low-resolution ADCs have previously focused one-bit quantization. In this chapter, our focus is on identifying the trade-off between the number of quantization bits and the error probability performance. Therefore, we take a different approach in which we allow the number of bits in the quantizer to vary until the transceiver architecture becomes asymptotically optimum in terms of communication reliability. The main contributions of this chapter are listed as follows.

- For any M -ary modulation scheme and quantizer structure, we show the existence of an error floor below which the average SEP cannot be pushed if the number of quantization bits n is less than $\log_2 M$.
- We derive the optimum ML detection rule for signal detection with low-resolution ADCs.
- We obtain exact analytical expressions for the average SEP attained by the derived ML rule when the wireless channel is subjected to Nakagami- m fading.
- We establish a fundamental ternary behaviour in the average SEP performance of a wireless communication system with low-resolution ADCs and M -PSK modulation. Based on the Nakagami- m fading model, we show that the decay exponent of the average SEP is the same as that of an infinite-bit quantization, which is equal to m , when n is larger than or equal to $\log_2 M + 1$. We also show that it is equal to $\frac{1}{2}$ and 0 for $n = \log_2 M$ and $n < \log_2 M$, respectively.
- By analysing the error probability performance in the high-SNR regime, we characterize the diversity order (DVO) for low-resolution ADC based communication systems.
- We perform numerical analysis to compare the performance of phase quantization to the more common situation where I and Q channels are independently quantized.

- We provide a numerical analysis establishing an important robustness property for our analytical results under channel state estimation errors.

Low-resolution Quantization with Non-coherent Detection

In Chapter 2, we assumed the perfect knowledge of CSI is available at the receiver and due to the non-linear nature of quantization, channel estimation has been one of the major challenges associated with low-resolution ADCs. Researchers have focused on several approaches in obtaining CSI using low-resolution ADCs. To that end, many researchers have used iterative channel estimation algorithms based on the least-squares approach and expectation maximization method, however, the complexity of these approaches is significant as it requires many iterations for convergence. As a solution to this, a mixed one-bit ADC architecture was proposed in which most antennas were equipped with one-bit ADCs, and some antennas equipped with high-resolution ADCs. These high-resolution ADCs were used to acquire precise CSI by reducing the channel estimation overhead.

In Chapter 3, we take a different approach to handle this challenge by adopting non-coherent data detection. More specifically, we adopt the differential quadrature phase shift keying (DQPSK) modulation method to differentially encode the transmit data and at the receiver side, employ non-coherent detection that does not require instantaneous CSI. we consider a point-to-point wireless communication system, in which the receiver is equipped with a low-resolution ADC that quantizes the phase of the received signal. We note that the SISO system provides the basic building block of a more complex multi-antenna system. We adopt the SISO system to draw fundamental insights into using non-coherent detection. It allows us to conduct a rigorous mathematical investigation to provide exact analytical expressions for the error probability performance. The performance results obtained in this chapter hold in the general case of n -bit quantization, DQPSK modulation and Rayleigh fading channels. We note that most of the work on low-resolution ADCs has previously focused one-bit quantization. In this chapter, our focus is on identifying the trade-off between the number of quantization bits and the error probability performance, which has not been considered previously. Therefore, we take a different approach in which

we allow the number of bits in the quantizer to vary until the transceiver architecture becomes asymptotically optimum in terms of communication reliability. The main contributions of this chapter are listed as follows.

- We derive the optimum ML detection rule when the detection is performed by using two consecutive received symbols.
- We derive the exact analytical expressions for the average SEP when the channel is subjected to Rayleigh fading.
- By analysing the asymptotic behaviour of the SEP performance, we establish that under non-coherent detection with DQPSK modulation, we can achieve the same DVO as with that of an infinite-bit quantization, when the number of quantization bits $n \geq 4$. Comparing this result with coherent detection considered in Chapter 2, we highlight that, for DQPSK modulation with non-coherent detection we need one additional quantization bit to achieve the full diversity when compared to the number of quantization bits required for QPSK modulation with coherent detection.
- We also analyse a system with DQPSK modulation under coherent detection when low-resolution phase quantization is employed. We obtain the optimum ML detection rule and analyse the average SEP results under Rayleigh fading. We also analyse the asymptotic behaviour of the SEP performance by focusing on the high SNR regime.

Low-resolution Quantization in Multi-Antenna Systems

In Chapter 4, we focus on optimum and sub-optimum detectors and error rate analysis for wireless systems with low-resolution quantizers in the presence of fading and noise for a wireless communication network consisting of one transmit antenna and N receive antennas. This is an extension of our work in Chapter 2 to a more general system with multiple antennas. We adopt the SIMO system to draw fundamental insights. It allows us to conduct a rigorous mathematical investigation to provide exact analytical expressions for the error probability performance. The performance results

obtained in this chapter hold in the general case of n -bit quantization, QPSK modulation and Rayleigh fading channels. We note that most of the work on low-resolution quantization based systems have previously focused on one-bit quantization. In this chapter, our focus is on identifying the trade-off between the number of quantization bits and the error probability performance, which has not been considered previously. The main contributions of this chapter are listed as follows.

- We derive the optimum ML detection rule for signal detection with low-resolution ADCs.
- We establish a fundamental ternary behaviour in the average SEP performance of a wireless communication system with low-resolution ADCs and QPSK modulation by using Monte Carlo simulations. Based on the Rayleigh fading model, we show that the decay exponent of the average SEP is the same as that of an infinite-bit quantization, which is equal to N when n is larger than or equal to 3 with N antennas at the receiver. We also show that it is equal to $\frac{N}{2}$ and 0 for $n = 2$ and $n < 2$, respectively.
- Due to the complexity of the derived ML detector, We propose three sub-optimum detection rules based on selection combining for signal detection with low-resolution ADCs in a SIMO wireless communication system.
- We obtain exact analytical expressions for the average SEP attained by the proposed sub-optimum detection rules when the wireless channel is subjected to Rayleigh fading.
- By analysing the error probability performance in the high-SNR regime, we characterize the DVO for low-resolution ADC based communication systems. We note that the simple sub-optimum detector that selects the path with the channel that locates the rotated constellation point furthest away from the decision boundary is asymptotically optimum in terms of communication reliability if $n \geq 3$. An extensive simulation study is performed to illustrate the accuracy of the derived results.

1.5 List of Publications

The following papers are based on the materials presented in this thesis. In some cases, materials in the journal papers are presented in part in the conferences papers.

1.5.1 Journal Papers

1. H.H.S. Gayan, R. Senanayake, H. Inaltekin and J. Evans: "Low-Resolution Quantization in Phase Modulated Systems: Optimum Detectors and Error Rate Analysis," *IEEE Open Journal of the Communications Society*, pp. 1000-1021, July 2020.
2. H.H.S. Gayan, R. Senanayake, H. Inaltekin and J. Evans: "Differentially Modulated Communications with Low-Resolution ADCs: Optimum Detectors and Error Rate Analysis," under preparation and will be submitted.
3. H.H.S. Gayan, R. Senanayake, H. Inaltekin and J. Evans: "Multi-Antenna Communications with Low-Resolution ADCs: Optimum Detectors and Error Rate Analysis," under preparation and will be submitted.

1.5.2 Conference Papers

1. H.H.S. Gayan, H. Inaltekin, R. Senanayake and J. Evans: "Phase Modulated Communication with Low-Resolution ADCs," *IEEE International Conference on Communications (ICC)*, pp. 1-6, Shanghai, China, May 2019.
2. H.H.S. Gayan, R. Senanayake, J. Evans: "On the Symbol Error Probability for QPSK with Quantized Observations", *27th International Telecommunication Networks and Applications Conference (ITNAC)*, pp. 1-6, Melbourne, VIC, Australia, Nov. 2017.
3. H.H.S. Gayan, R. Senanayake, H. Inaltekin and J. Evans: "Selection Combining for Multi-Antenna Communication with Low-Resolution ADCs," *IEEE International Symposium on Information Theory*, Melbourne, VIC, Australia, submitted and under review.

4. H.H.S. Gayan, R. Senanayake, H. Inaltekin and J. Evans: "Differentially Modulated Communication with Low-Resolution ADCs with Block-2 Detectors," under preparation and will be submitted.

Chapter 2

Low-Resolution Quantization in Single Input Single Output Systems

This chapter studies optimum detectors and error rate analysis for wireless systems with low-resolution quantizers in the presence of fading and noise. A universal lower bound on the average SEP, correct for all M -ary modulation schemes, is obtained when the number of quantization bits is not enough to resolve M signal points. In the special case of M -PSK, the ML detector is derived. Utilizing the structure of the derived detector, a general average SEP expression for M -PSK modulation with n -bit quantization is obtained when the wireless channel is subject to fading with a circularly-symmetric distribution. For the Nakagami- m fading, it is shown that a transceiver architecture with n -bit quantization is asymptotically optimum in terms of communication reliability if $n \geq \log_2 M + 1$. That is, the decay exponent for the average SEP is the same and equal to m with infinite-bit and n -bit quantizers for $n \geq \log_2 M + 1$. On the other hand, it is only equal to $\frac{1}{2}$ and 0 for $n = \log_2 M$ and $n < \log_2 M$, respectively. An extensive simulation study is performed to illustrate the accuracy of the derived results, energy efficiency gains obtained by means of low-resolution quantizers, performance comparison of phase modulated systems with independent in-phase and quadrature channel quantization and robustness of the derived results under channel estimation errors.

2.1 Introduction

2.1.1 Background and Motivation

Motivated by the potential of low-resolution ADCs, in this chapter, we aim to understand the fundamental trade-off between using low-resolution quantization and achieving higher communication reliability. More specifically, we analyse the error

probability performance of a communication system when the receiver is equipped with a phase quantizer that only outputs the knowledge of the phase quantization region on which the received signal lands. Phase quantization is a generalization of the common approach of using independent 1-bit quantization on I and Q arms, considered in several previous works [14–28]. In this case, the quantizer only outputs the quadrant on which the received signal lands. Using the derived structure of the ML detector, we fully characterize the SEP performance of a SISO system when the receiver is equipped with an n -bit phase quantizer.

We note that the SISO system provides the basic building block of a more complex MIMO system. We adopt the SISO system to draw fundamental insights. It allows us to conduct a rigorous mathematical investigation to provide exact analytical expressions for the error probability performance. The performance results obtained in this chapter hold in the general case of n -bit quantization, M -PSK modulation and Nakagami- m fading channels. We note that the Nakagami- m fading model characterizes a broad range of fading phenomena ranging from severe to moderate and no fading conditions as m varies over $[0.5, \infty)$ [75,76] and it reduces to Rayleigh fading for $m = 1$.

We note that some work on low-resolution ADCs has previously focused on MIMO systems [14–27] but under the case of one-bit quantization. In this chapter, our focus is on identifying the trade-off between the number of quantization bits and the error probability performance, which has not been considered previously. Therefore, we take a different approach in which we allow the number of bits in the quantizer to vary until the transceiver architecture becomes asymptotically optimum in terms of communication reliability.

2.1.2 Main Contributions

In this work, we consider a point-to-point wireless communication system, in which the receiver is equipped with a low-resolution ADC that quantizes the phase of the received signal. Our main contributions are summarized as follows.

- For any M -ary modulation scheme and quantizer structure, we show the existence of an error floor below which the average SEP cannot be pushed if the

number of quantization bits n is less than $\log_2 M$. This provides a fundamental performance limit for the SEP of a communication system that does not have enough quantization bits to resolve the M possible transmitted signals. This result is presented in Theorem 2.1.

- Motivated by the capacity achieving property of circularly symmetric input distributions for low-resolution ADCs [77], we adopt M -PSK modulation and derive the optimum ML detection rule for signal detection with low-resolution ADCs. For the general n -bit quantization, we obtain exact analytical expressions for the average SEP attained by the derived ML rule when the wireless channel is subjected to Nakagami- m fading. We use extensive numerical examples to compare the analytical expressions against the Monte-Carlo simulations. These results are presented in Theorems 2.2 and 2.3 respectively.
- We establish a fundamental ternary behaviour in the average SEP performance of a wireless communication system with low-resolution ADCs and M -PSK modulation. Based on the Nakagami- m fading model, we show that the decay exponent of the average SEP is the same as that of an infinite-bit quantization, which is equal to m when n is larger than or equal to $\log_2 M + 1$. We also show that it is equal to $\frac{1}{2}$ and 0 for $n = \log_2 M$ and $n < \log_2 M$, respectively. This result is presented in Theorem 2.4.
- By analysing the error probability performance in the high-SNR regime, we characterize the DVO for low-resolution ADC based communication systems. From a system design point of view, our results show that using one additional bit on top of $\log_2 M$ of them can achieve optimum communication robustness in the high-SNR regime. In particular, for fading environments with a large value of m , using an extra quantization bit improves communication reliability significantly. On the other hand, for fading environments with a low value of m , it may be more beneficial to use $\log_2 M$ bits, without sacrificing communication robustness too much but doubling system energy efficiency.
- We perform numerical analysis to compare the performance of phase quantiza-

tion to the more common situation where I and Q channels are independently quantized. For this analysis, we focus on Rayleigh fading ($m = 1$) with QPSK ($M = 4$) modulation to obtain some initial key insights into the potential bit savings that can be provided by the phase quantization approach employed in this work. While clearly 2-bit phase quantization is identical to using 1-bit quantization on each of the I and Q arms, we show that 3-bit phase quantization performs very similarly to the optimized independent I and Q quantization using 4-bits. This indicates that phase quantization provides a 1-bit saving in this particular case. Further, our analysis also demonstrates that the optimized independent quantization with 6-bits in total performs the same as 4-bit phase quantization, i.e., a saving of 2 bits. An important conclusion from this analysis is that not only does phase quantization have the potential to save bits, but the independent case needs to be optimized and adjusted as a function of the channel statistics, something that is not required for phase quantization.

- We provide numerical analysis establishing an important robustness property for our analytical results under channel state estimation errors. A salient feature arising from this numerical analysis is that the derived fundamental ternary behaviour of the decay exponent of $p(\text{SNR})$ stays the same if $n + 1$ bits are used during the channel estimation stage, while n being the number of bits used for data decoding. Moreover, the performance difference between using n and $n + 1$ bits for channel estimation becomes negligibly small as n increases. We note that the use of a different number of bits for data decoding and channel estimation is possible by employing mixed-ADC receiver architectures commonly used in the literature. Furthermore, the same effect to improve channel estimation accuracy can be achieved by using carefully designed training sequences to perform a bisection search during the channel estimate phase.

2.2 System Setup

2.2.1 Transmitter Architecture

In order to formalize the receiver architecture and the optimum signal detection problem below, we will assume that the transmitted signals are taken from the constellation set of information signals $\mathcal{C} = \left\{ e^{j\pi(\frac{2k+1}{M}-1)} \right\}_{k=0}^{M-1} \in \mathbb{C}$, which is the classical M -PSK signal constellation¹ and for ease of exposition, we only consider the case in which M is an integer power of 2².

2.2.2 Channel Model

We consider the classical point-to-point wireless channel model with flat-fading. For this channel, the received discrete-time baseband equivalent signal Y can be expressed by

$$Y = \sqrt{\text{SNR}}HX + W, \quad (2.1)$$

where $X \in \mathcal{C} \subset \mathbb{C}$ is the transmitted signal, $\mathcal{C}$ is the constellation set of information signals in $\mathbb{C}$, SNR is the ratio of the transmitted signal energy to the additive white Gaussian noise (AWGN) spectral density, $H \in \mathbb{C}$ is the unit power channel gain between the transmitter and the receiver which is random with zero-mean and generated according to a continuous distribution, and W is the circularly-symmetric zero-mean unit-variance AWGN, i.e., $W \sim \mathcal{CN}(0, 1)$. Here, W is independently generated at each channel use, independent of its past realizations and the realization of H .

2.2.3 Receiver Architecture

The receiver architecture is based on a low-resolution ADC. As illustrated in Fig. 2.1, the received signal Y is first sent through a low-resolution quantizer, and then the resulting quantized signal information is used to determine the transmitted symbol X .

¹This choice of $\mathcal{C}$ ensures that the phase of X always lies in $[-\pi, \pi)$

²Extensions of our results to the more general case of M being any positive integer is straightforward, albeit with more complicated notation and separate analyses in some special cases.

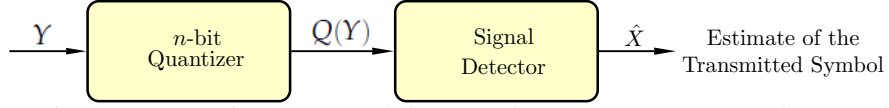


Figure 2.1: The receiver architecture with low-resolution quantization. The signal detector observes only the n -bit quantized versions of Y to estimate the transmitted signal.

In the following, we describe the quantization process in detail.

2.2.4 Quantization Process

If n bits are used to quantize Y , the quantizer Q divides the complex domain $\mathbb{C}$ into 2^n quantization regions and outputs the index of the region in which Y lies as an input to the detector. As such, we declare $Q(Y) = k$ if $Y \in \mathcal{R}_k$ for $k \in [0 : 2^n - 1]$, where $\mathcal{R}_k \subseteq \mathbb{C}$ is the k -th quantization region. Since information is encoded in the phase of X with the above choice of constellation points, we choose $\mathcal{R}_k$ as the convex cone given by

$$\mathcal{R}_k = \left\{ z \in \mathbb{C} : \frac{2\pi}{2^n}k \leq \text{Arg}(z) + \pi < \frac{2\pi}{2^n}(k+1) \right\}. \quad (2.2)$$

The use of phase quantization in our receiver architecture is further motivated by the following two factors. First, considering channel impairments as phase rotations in transmitted signals, quantization and decision regions for M -PSK modulation are conveniently modelled as convex cones in the complex plane [78] as described above, and without requiring the use of automatic gain control. Second, phase quantizers can be implemented using one-bit ADCs that consist of simple comparators, and they consume negligible power (in the order of mW).

2.2.5 Key Assumptions

We assume that the n -bit quantizer is fixed and independent of the wireless channel.

We also assume that full CSI is available at the receiver. The motivation behind our model with full-precision CSI in the paper is two-fold. First, in previous work [40], it was shown that it is possible to attain a high channel estimation precision (for $\text{SNR} = 10$ dB, the estimator based on 3-bit ADC gives a mean square error of -20

dB with 11 training symbols) with the use of low-resolution ADCs by increasing the number of training symbols in the closed-loop estimation process.

Second, the mixed-ADC architectures are also commonly investigated in the literature, and they can be employed to achieve high channel estimation accuracy [71]. In a mixed-ADC architecture, high-resolution ADCs (structured through either serial or parallel connections) are used during the channel estimation stage [18,40,71] and the data transmission phase, the receiver switches to low-resolution operation by using less number of quantization bits. Although the energy consumption is increased in this approach, this is not a restrictive degradation for our purposes. Each fading state will span a large group of information bits at the target multiple Gbits per second data rates in next-generation wireless systems. Hence, the energy saving during data transmission is more significant than the increased energy consumption during channel estimation. Based on the existing results for channel estimation accuracy with low-resolution quantizers, the assumption on the availability of full CSI at the receiver is also commonly used by the pioneering papers in the field [30,40,66,79]. The effect of channel estimation errors on the SEP performance is further investigated in Section 2.7.3.

We also assume that there is no feedback path from the receiver to the transmitter to get the CSI that is available at the receiver. Therefore, there is no CSI available at the transmitter and hence it is not possible to pre-equalize the channel phase at the transmitter. Further, in this work, we are not employing any error correction codes.

2.3 Optimum Signal Detection

The detector aims to minimize the SEP by using the knowledge of $Q(Y)$ and CSI, which can be represented as selecting a signal point $\hat{x}(k, h)$ satisfying

$$\hat{x}(k, h) \in \arg \max_{x \in \mathcal{C}} \Pr \{X = x | Q(Y) = k, H = h\}, \quad (2.3)$$

for $h \in \mathbb{C}$ and $k \in [0 : 2^n - 1]$. The main performance figure of merit for the optimum detector is the average SEP given by

$$p(\text{SNR}) = \Pr \{X \neq \hat{x}(Q(Y), H)\}. \quad (2.4)$$

It is important to note that $p(\text{SNR})$ depends on SNR as well as the number of quantization bits.

We will analyze the communication robustness that can be achieved with low-resolution ADCs by focusing on the decay exponent for $p(\text{SNR})$, which is given by³

$$\text{DVO} = - \lim_{\text{SNR} \rightarrow \infty} \frac{\log p(\text{SNR})}{\log \text{SNR}}. \quad (2.5)$$

Following the convention in the field, we will call DVO the *diversity order*, although there is only a single diversity branch in our system.

We mainly focus on the error probability perspective in this analytical study on the performance impact of using low-resolution quantization in an uncoded system. Therefore, we select SEP and DVO as the two main performance metrics. Especially, the DVO can be used to quantify the error probability behaviour in the high SNR regime. Though it is not considered in this thesis, throughput is another important performance metric that can be considered when analyzing the performance impact of using low-resolution quantization, especially when coding is entertained. In the literature, most of the works on analyzing the information-theoretic throughput focus on one-bit ADC based systems with independent I and Q quantization [29, 80, 81]. The SEP results that we generate in this thesis can be used to find information-theoretic bounds for channel capacity and throughput with suitable mathematical analysis.

Our first result indicates that there is an SNR-independent error floor such that the average SEP values below which cannot be attained for $n < \log_2 M$. The following theorem establishes this result formally. We note that this theorem establishes a more general result that holds for any choice of modulation scheme and quantizer structure.

Theorem 2.1. *Let $p_{\min}$ be the probability of the least probable transmitted symbol. If $n <$*

³We will show that the limit in (2.5) exists, and hence there is no ambiguity in the definition of DVO.

$\log_2 M$, then for any choice of modulation scheme and quantizer structure

$$p(\text{SNR}) \geq \frac{M - 2^n}{2^n} p_{\min} \quad (2.6)$$

for all $\text{SNR} \geq 0$.

Proof: See Appendix 2.9.1. □

Firstly, we note that the error floor in (2.6) is always a valid lower bound since $p_{\min} \leq \frac{1}{M}$. Secondly, it does not depend on the fading model. The average SEP values below $\frac{M-2^n}{2^n} p_{\min}$ cannot be achieved due to the inherent inability of low-resolution ADC receivers to resolve different signal points when $n < \log_2 M$. We also note that the Fano's inequality can also be used to obtain similar, perhaps tighter, lower bounds on $p(\text{SNR})$ [82]. However, this will require the calculation of equivocation between X and $Q(Y)$ for each choice of modulation scheme and quantizer structure. Hence, it is not clear how the minimization can be carried out over the modulation and quantizer selections in this approach.

Next, we will assume that all signal points in $\mathcal{C}$ are equiprobable, with probability $\frac{1}{M}$, and hence the optimum detector in (2.3) is equivalent to the ML detector given by

$$\hat{x}(k, h) \in \arg \max_{x \in \mathcal{C}} \Pr \{Q(Y) = k | X = x, H = h\} \quad (2.7)$$

for $h \in \mathbb{C}$ and $k \in [0 : 2^n - 1]$. Since Y is a proper complex Gaussian random variable with mean $E[Y] = \sqrt{\text{SNR}}hx$ and variance $\text{Var}(Y) = 1$, we can write the probability in (2.7) as

$$\Pr \{Q(Y) = k | X = x, H = h\} = \int_{\mathcal{R}_k} \frac{1}{\pi} \exp \left(-|y - \sqrt{\text{SNR}}hx|^2 \right) dy, \quad (2.8)$$

where the integral in (2.8) is with respect to the standard Borel measure in $\mathbb{C}$ [83].

It is also important to note that the error floor in Theorem 2.1 is independent of the fading model. The unachievability of the average SEP values below $\frac{M-2^n}{2^n} p_{\min}$ arises from the inherent inability of low-resolution ADC receivers to resolve different signal points when $n < \log_2 M$. Therefore, in the rest of this work we only focus on

$n \geq \log_2 M$ case.

The next theorem describes the operation of the ML detector for the above signal detection problem.

Theorem 2.2. *Assume H has a continuous probability density function (pdf). Then, $\hat{x}(k, h)$ is unique with probability one, i.e., the set of h values for which*

$$\arg \max_{x \in \mathcal{C}} \Pr \{Q(Y) = k | X = x, H = h\}$$

is singleton has probability one, and the ML detection rule for the low-resolution ADC based receiver architecture can be given as

$$\hat{x}(k, h) = \arg \min_{x \in \mathcal{C}} \text{dist}(hx, \mathcal{H}_k), \quad (2.9)$$

where $h \in \mathbb{C}$, $k \in [0 : 2^n - 1]$, $\text{dist}(z, \mathcal{A})$ is the distance between a point $z \in \mathbb{C}$ and a set $\mathcal{A} \subseteq \mathbb{C}$, which is defined as $\text{dist}(z, \mathcal{A}) \triangleq \inf_{s \in \mathcal{A}} |z - s|$, and $\mathcal{H}_k = \{z \in \mathbb{C} : \text{Arg}(z) + \pi = \frac{\pi}{2^n} (2k + 1)\}$.

Proof: See Appendix 2.9.2. □

We first note that although the ML detection rule given in Theorem 2.2 is written as dependent on the full channel knowledge, we only need to know the phase of H in order to determine $\hat{x}(k, h)$. This can be seen by means of simple scaling arguments as the set $\mathcal{H}_k$ is invariant under any amplitude scaling. The half-hyperplane $\mathcal{H}_k$ in Theorem 2.2 bisects the k th quantization region $\mathcal{R}_k$ into two symmetric regions. This is an important observation since it indicates that we are only required to estimate the channel phase to implement the ML detector. For phase estimation with circularly symmetric fading distributions, the estimation errors due to low-resolution quantization can be modelled as bounded uniform random variables, which lead to a key robustness property for our analytical results under channel estimation errors as explained in Section 2.7.3.

Secondly, given $X = x$ and $H = h$, the conditional probability of the event $Q(Y) = k$ corresponds to the integral of unit-variance and mean $\sqrt{\text{SNR}}hx$ proper complex Gaussian distribution over $\mathcal{R}_k$. As we vary x in $\mathcal{C}$, we change the mean value $\sqrt{\text{SNR}}hx$. Hence, Theorem 2.2 indicates that the most probability mass is accumulated in the region $\mathcal{R}_k$ when the unit-variance proper complex Gaussian distribution with mean

$\sqrt{\text{SNR}}hx$, $x \in \mathcal{C}$, closest to $\mathcal{H}_k$ is integrated over $\mathcal{R}_k$, which coincides with the intuition. We will use the structure of the ML detection rule to derive integral expressions for $p(\text{SNR})$ for $M \geq 2$ in Section 2.4. Further, in order to characterize the communication robustness with low-resolution ADCs in the high SNR regime, we will also provide a detailed analysis of the asymptotic decay exponent of $p(\text{SNR})$ in Section 2.5.

2.4 Average Symbol Error Probability

2.4.1 Symbol Error Probability for $n \geq \log_2 M$

We first obtain a key lemma that simplifies the calculations for deriving $p(\text{SNR})$ when the number of quantization bits is at least $\log_2 M$. We note that this lemma holds for general circularly-symmetric fading processes without assuming any specific functional form.

Lemma 2.1. *Let $H = Re^{j\Lambda}$ be a circularly-symmetric fading coefficient with R and Λ denoting the magnitude and the phase of H , respectively. Let the joint pdf of R and Λ be given by $f_{R,\Lambda}(r, \lambda) = \frac{1}{2\pi} f_R(r)$ for $\lambda \in [-\pi, \pi)$ and $r \geq 0$. Then, $p(\text{SNR})$ is equal to*

$$p(\text{SNR}) = \frac{2^{n-1}}{\pi} \int_{\frac{\pi}{M} - \frac{\pi}{2^n}}^{\frac{\pi}{M} + \frac{\pi}{2^n}} \int_0^\infty \Pr \left\{ \sqrt{\text{SNR}} Re^{j\theta} + W \notin \mathcal{E} \right\} f_R(r) dr d\theta, \quad (2.10)$$

where $\mathcal{E} = \{z \in \mathbb{C} : 0 \leq \text{Arg}(z) < \frac{2\pi}{M}\}$.

Proof: See Appendix 2.9.3. □

Using Lemma 2.1, next we obtain integral expressions for $p(\text{SNR})$ when H is circularly-symmetric with the generalized Nakagami- m fading magnitude.

Considering these advantages, we will focus on the Nakagami- m fading model for H to derive integral expressions for $p(\text{SNR})$ in the remainder of the paper. This will be done so for all parameter combinations of $M \geq 2$ (as an integer power of 2), $n \geq \log_2 M$ and $m \geq 0.5$. It will be seen that the derived integral expressions are easy to calculate numerically and they reduce to simple closed-form expressions in some special cases. Further, we will also show that using $\log_2 M + 1$ bits is enough to achieve the maximum

communication robustness achieved by using an infinite number of quantization bits.

Theorem 2.3. Assume H is a unit-power fading coefficient distributed according to a circularly-symmetric distribution with Nakagami- m fading magnitude. Let $\mathcal{Q}(\cdot)$ be the complementary distribution function of the standard normal random variable and $\Gamma(\cdot)$ be the gamma function [84]. Then, for $n \geq \log_2 M$ and $M \geq 2$, $p(\text{SNR})$ is given according to (2.11).

$$p(\text{SNR}) = \begin{cases} p_1(\text{SNR}) + p_2(\text{SNR}) - p_3(\text{SNR}) + p_4(\text{SNR}) & M \geq 4 \\ p_2(\text{SNR}) & M = 2 \end{cases}, \text{where} \quad (2.11)$$

$$p_1(\text{SNR}) = \frac{2^{n-1}m^m}{\pi^2} \int_0^{\frac{\pi}{2}} \int_{\frac{\pi}{M}-\frac{\pi}{2^n}}^{\frac{\pi}{M}+\frac{\pi}{2^n}} \left(\frac{\text{SNR}}{\sin^2 \beta} \cos^2 \theta + m \right)^{-m} d\theta d\beta, \quad (2.12)$$

$$p_2(\text{SNR}) = \frac{2^{n-1}m^m}{\pi^2} \int_0^{\frac{\pi}{2}} \int_{\frac{\pi}{M}-\frac{\pi}{2^n}}^{\frac{\pi}{M}+\frac{\pi}{2^n}} \left(\frac{\text{SNR}}{\sin^2 \beta} \sin^2 \theta + m \right)^{-m} d\theta d\beta, \quad (2.13)$$

$$p_3(\text{SNR}) = \frac{2^{n-1}m^m}{\pi^3} \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} \int_{\frac{\pi}{M}-\frac{\pi}{2^n}}^{\frac{\pi}{M}+\frac{\pi}{2^n}} \left(\frac{\text{SNR} \cos^2 \theta}{\sin^2 \beta} + \frac{\text{SNR} \sin^2 \theta}{\sin^2 \gamma} + m \right)^{-m} d\theta d\beta d\gamma, \quad (2.14)$$

$$p_4(\text{SNR}) = \frac{2^n m^m}{\pi \sqrt{\pi} \Gamma(m)} \int_{\frac{\pi}{M}-\frac{\pi}{2^n}}^{\frac{\pi}{M}+\frac{\pi}{2^n}} \int_0^\infty \int_{-\sqrt{\text{SNR}r} \cos \theta}^\infty \exp(-(w^2 + mr^2)) \cdot \mathcal{Q}\left(\sqrt{2\text{SNR}r} \sec\left(\frac{2\pi}{M}\right) \sin\left(\frac{2\pi}{M} - \theta\right) + \sqrt{2}w \tan\left(\frac{2\pi}{M}\right)\right) dw dr d\theta. \quad (2.15)$$

Proof: In the following we provide the proof for $M \geq 4$. Please note that the proof for $M = 2$ is similar and simpler. With a slight abuse of notation, we define

$$p(\text{SNR}, h) = \Pr \left\{ \sqrt{\text{SNR}r} e^{j\theta} + W \notin \mathcal{E} \right\}, \quad (2.16)$$

where the set $\mathcal{E}$ is defined as in Lemma 2.1. The probability in (2.16) can be calculated by conditioning on the real part of W , which is denoted by W_{re} . By using Fig. 2.2 as a visual guide, we can write $p(\text{SNR}, h)$ after conditioning on W_{re} as (2.17) for $w \geq$

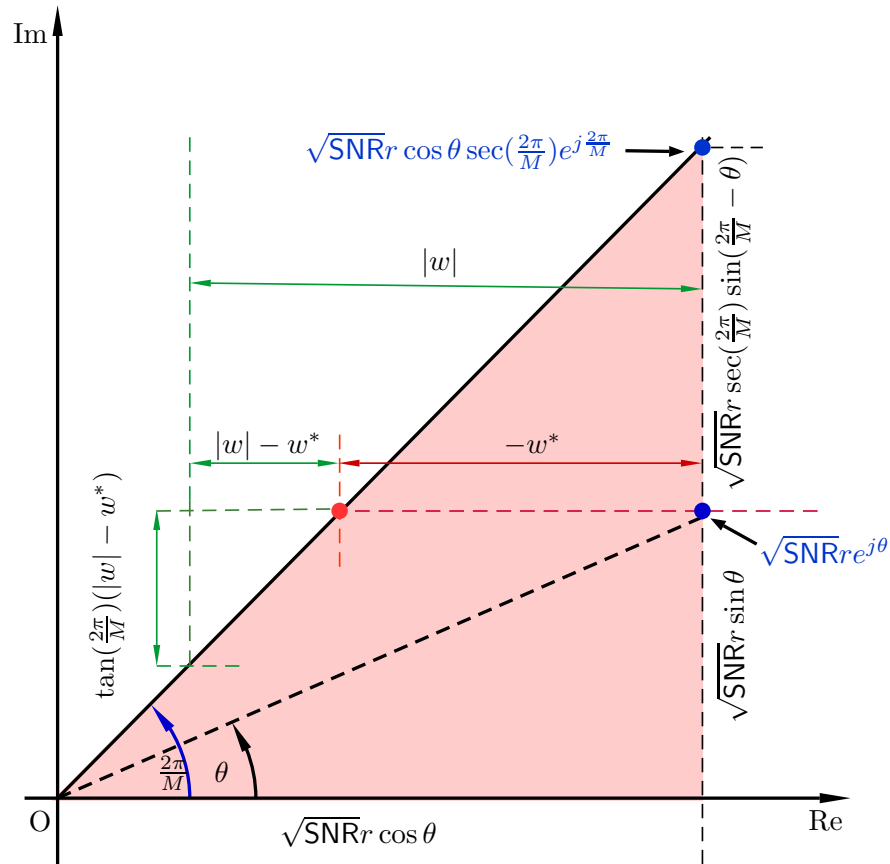


Figure 2.2: An illustration of average SEP calculations. If the noise does not drag the original M -PSK constellation point rotated by the channel h beyond the region $\mathcal{E}$ (shaded area), there will not be any errors in decoding.

$$-\sqrt{\text{SNR}r} \cos \theta.$$

$$\Pr \left\{ \sqrt{\text{SNR}} r e^{j\theta} + W \notin \mathcal{E} \mid W_{\text{re}} = w \right\} =$$

$$\mathcal{Q} \left(\sqrt{2\text{SNR}} r \sin \theta \right) + \mathcal{Q} \left(\sqrt{2\text{SNR}} r \sec \left(\frac{2\pi}{M} \right) \sin \left(\frac{2\pi}{M} - \theta \right) + \sqrt{2} w \tan \left(\frac{2\pi}{M} \right) \right)$$

$$(2.17)$$

Similarly, for $w < -\sqrt{\text{SNR}r} \cos \theta$, we get

$$\Pr\left\{\sqrt{\text{SNR}}e^{j\theta} + W \notin \mathcal{E} \mid W_{\text{re}} = w\right\} = 1. \quad (2.18)$$

Integrating (2.17) and (2.18) with respect to the pdf of W_{re} , which is given by

$f_{W_{\text{re}}}(w) = \frac{1}{\sqrt{\pi}}e^{-w^2}$, we obtain $p(\text{SNR}, h)$ as (2.19).

$$\begin{aligned}
p(\text{SNR}, h) = & \mathcal{Q}\left(\sqrt{2\text{SNR}}r \cos \theta\right) + \mathcal{Q}\left(\sqrt{2\text{SNR}}r \sin \theta\right) - \mathcal{Q}\left(\sqrt{2\text{SNR}}r \cos \theta\right) \mathcal{Q}\left(\sqrt{2\text{SNR}}r \sin \theta\right) \\
& + \frac{1}{\sqrt{\pi}} \int_{-\sqrt{\text{SNR}}r \cos \theta}^{\infty} \mathcal{Q}\left(\sqrt{2\text{SNR}}r \sec\left(\frac{2\pi}{M}\right) \sin\left(\frac{2\pi}{M} - \theta\right) + \sqrt{2}w \tan\left(\frac{2\pi}{M}\right)\right) e^{-w^2} dw.
\end{aligned} \tag{2.19}$$

For Nakagami- m fading distribution with shape parameter $m \geq 0.5$ and spread parameter $\Omega > 0$ [85], we can write the pdf of the fading magnitude as $f_R(r) = \frac{2m^m}{\Gamma(m)\Omega^m} r^{2m-1} e^{-\frac{m}{\Omega}r^2}$ for $r \geq 0$. We set $\Omega = 1$ in our calculations to make sure that H has unit-power. We average $p(\text{SNR}, h)$ over the fading distribution and solve the resulting integral based on Lemma 2.1, and the fact that θ lies between 0 and $\frac{2\pi}{M}$, to obtain $p(\text{SNR})$ in Theorem 2.3. $\square$

2.4.2 Centering Property: Impact of Quantization Bits on the Average SEP

In this subsection, we will present an intuitive explanation as to why $p(\text{SNR})$ improves with the increasing number of quantization bits. In particular, we will observe that one extra bit, on top of $\log_2 M$ of them, provides a desirable centering property that steers the received signal away from the error-prone decision boundaries. This intuition will help to understand the underlying dynamics leading to the ternary behaviour for the decay exponent of $p(\text{SNR})$ that we establish in the high SNR regime in Section 2.5.

For $i \in [0 : M-1]$, let $x_i = e^{j\pi(\frac{2i+1}{M}-1)}$ be the i th signal point in the constellation set $\mathcal{C}$ and $\mathcal{E}_i = \{z \in \mathbb{C} : \text{Arg}(x_i) - \frac{\pi}{M} \leq \text{Arg}(z) < \text{Arg}(x_i) + \frac{\pi}{M}\}$. It can be shown (i.e., see Appendix 2.9.3) that the regions defined by $\mathcal{E}_{i,k} \triangleq \exp(j(k-2^{n-1})\frac{2\pi}{2^n})\mathcal{E}_i$ for $i \in [0 : M-1]$ and $k \in [0 : 2^n-1]$ contains all $\mathcal{H}_k$'s to which $\sqrt{\text{SNR}}hx_i$ is the closest for $h \in \mathcal{D}_k$, where

$$\mathcal{D}_0 = \left\{z \in \mathbb{C} : \pi - \frac{\pi}{2^n} \leq \text{Arg}(z) < \pi\right\} \cup \left\{z \in \mathbb{C} : -\pi \leq \text{Arg}(z) < \frac{\pi}{2^n} - \pi\right\} \tag{2.20}$$

and

$$\mathcal{D}_k = \left\{ z \in \mathbb{C} : (2k-1) \frac{\pi}{2^n} \leq \text{Arg}(z) + \pi < (2k+1) \frac{\pi}{2^n} \right\}. \quad (2.21)$$

This means that all the received signal points in $\mathcal{E}_{i,k}$ will be detected as x_i , and hence $\mathcal{E}_{i,k}$ can be considered as the *region of attraction* of x_i . This also means that if the received signal lands in $\mathcal{E}_{i,k}$ when x_i is transmitted, then there will not be any detection errors.

Let us consider an example of QPSK modulation with 2-bit and 3-bit quantization. Without loss of generality, we will assume that $x_3 = e^{j\frac{\pi}{M}}$ is the transmitted signal. Our analysis will be for two cases of $\lambda = \frac{\pi}{18}$ and $\lambda = \frac{4\pi}{18}$, where $\lambda = \text{Arg}(h)$. Table 2.1 summarizes these two cases, and Fig. 2.3 illustrates them. In this figure, we show both the original signal points (indicated by ‘ $\diamond$ ’) and the rotated ones (indicated by ‘ $\bullet$ ’) after multiplying with $\sqrt{\text{SNR}}$ and h .

	$\lambda = \frac{\pi}{18}$	$\lambda = \frac{4\pi}{18}$
$n = 2$	$h \in \mathcal{D}_2$ $i = 3, k = 2$ $\mathcal{E}_{i,k} = \mathcal{E}_3$	$h \in \mathcal{D}_4$ $i = 3, k = 4$ $\mathcal{E}_{i,k} = \mathcal{E}_3$
$n = 3$	$h \in \mathcal{D}_2$ $i = 3, k = 2$ $\mathcal{E}_{i,k} = \mathcal{E}_3$	$h \in \mathcal{D}_4$ $i = 3, k = 5$ $\mathcal{E}_{i,k} = e^{j\frac{\pi}{4}} \mathcal{E}_3$

Table 2.1: Centering property for QPSK modulation with 2-bit and 3-bit quantization. $\mathcal{E}_{i,k}$ is the region of attraction of the symbol x_i when the quantizer output $Q(Y) = k$.

For both 2-bit and 3-bit quantization, we observe that $h \in \mathcal{D}_2$ and $h \in \mathcal{D}_4$ for $\lambda = \frac{\pi}{18}$ and $\lambda = \frac{4\pi}{18}$, respectively. Therefore, for 2-bit quantization, the region of attraction for x_3 will be $\mathcal{E}_3$ for both cases. Here, we can see that the rotated constellation point $\sqrt{\text{SNR}}hx_3$ is very close to the decision boundary when $\lambda = \frac{4\pi}{18}$. Hence, there is a high probability that the received signal $\sqrt{\text{SNR}}hx_3 + w$ lands in the adjacent quantization region for $\lambda = \frac{4\pi}{18}$. In this instance, we will have a detection error. However, with the addition of one bit to the quantizer (i.e. with 3-bit quantization), the region of attraction of x_3 will be $e^{j\frac{\pi}{4}}\mathcal{E}_3$, and hence the ML detector can correctly decode the transmitted signal even if the received one lands in the adjacent quantization region. This is illustrated in Fig. 2.3(d). Therefore, the addition of one extra bit to the quantizer steers the received signal

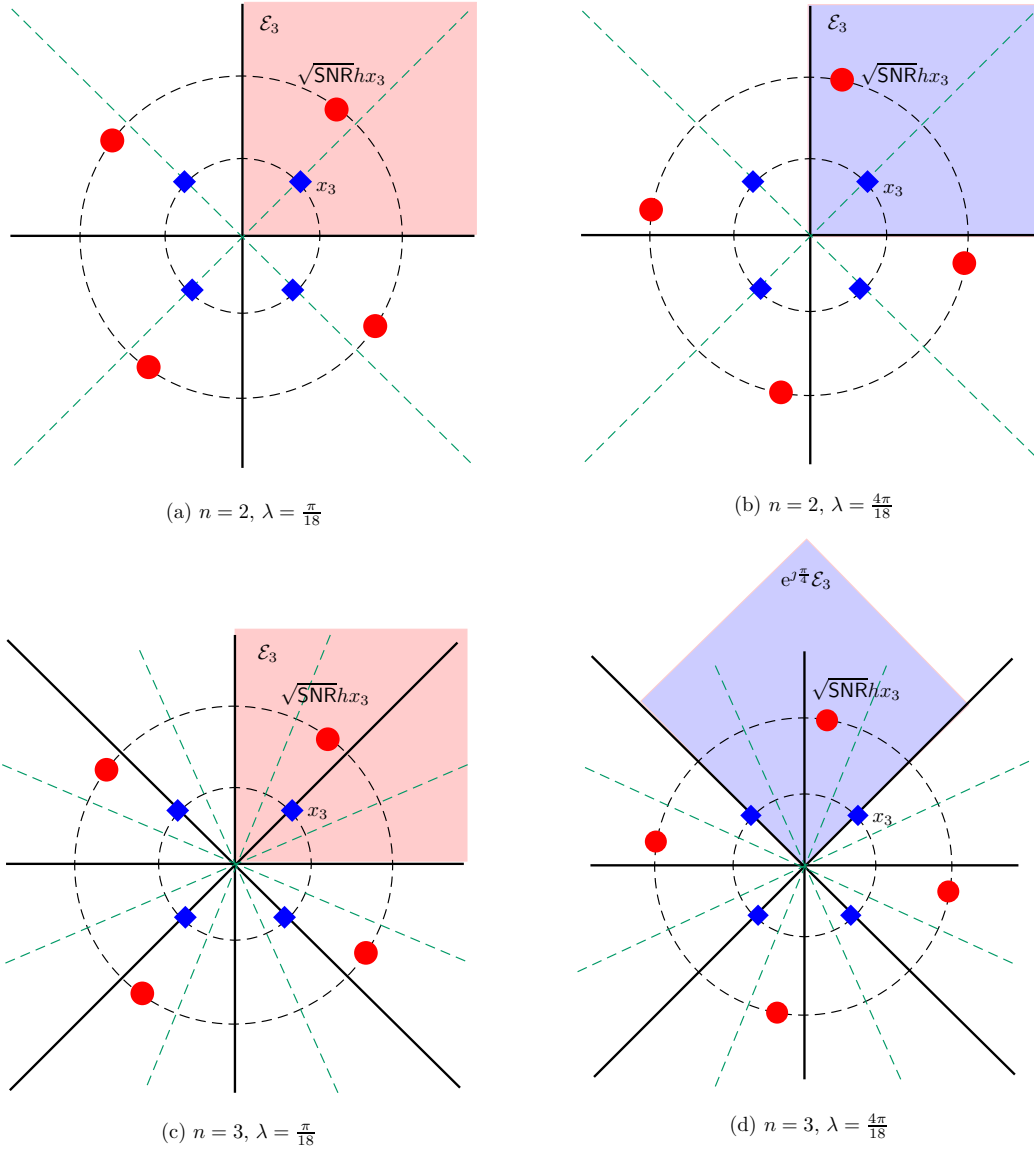


Figure 2.3: An illustration of the centering property for QPSK modulation with 2-bit and 3-bit quantization. Original signal points are indicated by ' $\diamond$ ', whereas the rotated ones after multiplication with $\sqrt{\text{SNR}}$ and h are indicated by ' $\bullet$ '. Quantization region boundaries and the corresponding bisectors are indicated in solid black lines and green dash lines, respectively. The shaded area represents the region of attraction of the transmitted symbol x_3 .

away from the error-prone decision boundaries to improve $p(\text{SNR})$. Similarly, when the number of bits in the quantizer continues to increase, the quantization regions will become thinner, and hence the regions of attraction will be better centered around the received signal points. This is the fundamental phenomenon that explains why the average SEP improves with a larger number of quantization bits.

2.5 The Decay Exponent for the Average Symbol Error Probability

In this section, we will analyze the communication robustness that can be achieved with low-resolution ADCs by focusing on the decay exponent for $p(\text{SNR})$, which is given by DVO defined in (2.5). Following the convention in the field, we will call DVO the *diversity order*, although there is only a single diversity branch in our system. It should be noted that Nakagami- m amplitude distribution can be obtained as the envelope distribution of m independent Rayleigh faded signals for integer values of m [75]. Hence, visualizing a Nakagami- m wireless channel as a pre-detection analog square-law diversity combiner will put the results of this section into context. We devote the rest of the current section to the proof of this important finding. We will first start with a definition that will simplify the notation below.

Definition 2.1. We say a function f is exponentially equal to SNR^d if $\lim_{\text{SNR} \rightarrow \infty} \frac{\log f(\text{SNR})}{\log \text{SNR}} = d$ for some $d \in \mathbb{R}$. We write $f(\text{SNR}) \stackrel{e}{=} \text{SNR}^d$ to indicate exponential equality whenever this limit exists. Similarly, we also write $f(\text{SNR}) \stackrel{e}{\leq} \text{SNR}^d$ and $f(\text{SNR}) \stackrel{e}{\geq} \text{SNR}^d$ if $\lim_{\text{SNR} \rightarrow \infty} \frac{\log f(\text{SNR})}{\log \text{SNR}} \leq d$ and $\lim_{\text{SNR} \rightarrow \infty} \frac{\log f(\text{SNR})}{\log \text{SNR}} \geq d$, respectively.

The following lemma establishes two important properties for exponential equality.

Lemma 2.2. Let $f(\text{SNR}) \stackrel{e}{=} \text{SNR}^d$ and $f_i(\text{SNR}) \stackrel{e}{=} \text{SNR}^{d_i}$ for $i \in [1 : N]$. Then,

- (i) For any $\alpha > 0$, $\alpha f(\text{SNR}) \stackrel{e}{=} \text{SNR}^d$ (i.e., invariance with scaling property).
- (ii) $\sum_{i=1}^N f_i(\text{SNR}) \stackrel{e}{=} \text{SNR}^{d_{\max}}$, where $d_{\max} = \max_{i \in [1:N]} d_i$ (i.e., summation property).

Proof: See Appendix 2.9.4. □

The next two lemmas establish the decay rates for $p_1(\text{SNR})$ and $p_2(\text{SNR})$ in Theorem 2.3 in terms of exponential equalities.

Lemma 2.3. For $M \geq 4$, $p_1(\text{SNR})$ is exponentially equal to

$$p_1(\text{SNR}) \stackrel{e}{=} \begin{cases} \text{SNR}^{-\frac{1}{2}} & \text{if } M = 4 \text{ and } n = 2, \\ \text{SNR}^{-m} & \text{if } M = 4 \text{ and } n > 2, \\ \text{SNR}^{-m} & \text{if } M > 4 \text{ and } n \geq \log_2 M. \end{cases} \quad (2.22)$$

Proof: See Appendix 2.9.5. □

Lemma 2.4. For $M \geq 4$, $p_2(\text{SNR})$ is exponentially equal to

$$p_2(\text{SNR}) \stackrel{e}{=} \begin{cases} \text{SNR}^{-\frac{1}{2}} & \text{if } n = \log_2 M, \\ \text{SNR}^{-m} & \text{if } n > \log_2 M. \end{cases} \quad (2.23)$$

Proof: See Appendix 2.9.6. □

The following lemma establishes lower and upper bounds on SEP in (2.11). We note that the bounds in Lemma 2.5 hold for all circularly-symmetric fading processes, including Nakagami- m magnitude pdf as a special case.

Lemma 2.5. For $M \geq 4$ and $n \geq \log_2 M$, let $L(\text{SNR}) = p_1(\text{SNR}) + \frac{1}{2}p_2(\text{SNR})$ and $U(\text{SNR}) = p_1(\text{SNR}) + 2p_2(\text{SNR})$. Then,

$$L(\text{SNR}) \leq p(\text{SNR}) \leq U(\text{SNR}). \quad (2.24)$$

Proof: See Appendix 2.9.7. □

The upper and lower bounds obtained in Lemma 2.5 are instrumental for us in providing proof for Theorem 2.4. In particular, they decay at the same rate as SNR grows to infinity. Technically speaking, they are *exponentially equal* to $p(\text{SNR})$ (i.e., see Definition 2.1), and this exponential equality property enables the derivation of ternary behaviour for DVO as established in Theorem 2.4. Further, our numerical analysis in Section 2.7 shows that the upper bound $U(\text{SNR})$ becomes also a very tight bound for $p(\text{SNR})$ for large SNR values in some certain circumstances.

Theorem 2.4. The DVO of a low-resolution ADC based receiver architecture with M -PSK modulation and Nakagami- m fading is given by

$$\text{DVO} = \begin{cases} \frac{1}{2} & n = \log_2 M, \\ m & n \geq \log_2 M + 1. \end{cases} \quad (2.25)$$

Proof: The proof for $M \geq 4$ directly follows from Lemmas 2.2, 2.3, 2.4 and 2.5.

For BPSK modulation (i.e., $M = 2$) and $n = 1$, we have

$$\begin{aligned} p(\text{SNR}) &= \frac{2^{n-1}m^m}{\pi^2} \int_0^{\frac{\pi}{2}} \int_0^{\pi} \left(\frac{\text{SNR}}{\sin^2 \beta} \sin^2 \theta + m \right)^{-m} d\theta d\beta \\ &= \frac{2^{n-1}m^m}{\pi^2} \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} \left(\frac{\text{SNR}}{\sin^2 \beta} \sin^2 \theta + m \right)^{-m} d\theta d\beta \\ &\quad + \frac{2^{n-1}m^m}{\pi^2} \int_0^{\frac{\pi}{2}} \int_{\frac{\pi}{2}}^{\pi} \left(\frac{\text{SNR}}{\sin^2 \beta} \sin^2 \theta + m \right)^{-m} d\theta d\beta \quad (2.26) \end{aligned}$$

By using the change of variables $\hat{\theta} = \theta - \frac{\pi}{2}$ in the second integral term of (2.26), we have

$$\begin{aligned} p(\text{SNR}) &= \frac{2^{n-1}m^m}{\pi^2} \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} \left(\frac{\text{SNR}}{\sin^2 \beta} \sin^2 \theta + m \right)^{-m} d\theta d\beta \\ &\quad + \frac{2^{n-1}m^m}{\pi^2} \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} \left(\frac{\text{SNR}}{\sin^2 \beta} \cos^2 \hat{\theta} + m \right)^{-m} d\hat{\theta} d\beta \quad (2.27) \end{aligned}$$

This expression is equivalent to $p_1(\text{SNR}) + p_2(\text{SNR})$ for $M = 4$ and $n = 2$. Hence, by using Lemma 2.2, we can conclude that

$$\lim_{\text{SNR} \rightarrow \infty} -\frac{\log(p(\text{SNR}))}{\log(\text{SNR})} = \frac{1}{2} \quad (2.28)$$

for BPSK modulation with 1-bit quantization and $m \geq \frac{1}{2}$.

For BPSK modulation with $n > \log_2(M)$, we have

$$p(\text{SNR}) = \frac{2^{n-1}m^m}{\pi^2} (\text{SNR})^{-m} \int_0^{\frac{\pi}{2}} \int_{\frac{\pi}{M} - \frac{\pi}{2^n}}^{\frac{\pi}{M} + \frac{\pi}{2^n}} \left(\frac{\sin^2 \theta}{\sin^2 \beta} + \frac{m}{\text{SNR}} \right)^{-m} d\theta d\beta \quad (2.29)$$

Therefore

$$\log(p(\text{SNR})) = c - m \log(\text{SNR}) + \log \left(\int_0^{\frac{\pi}{2}} \int_{\frac{\pi}{M} - \frac{\pi}{2^n}}^{\frac{\pi}{M} + \frac{\pi}{2^n}} \left(\frac{\sin^2 \theta}{\sin^2 \beta} + \frac{m}{\text{SNR}} \right)^{-m} d\theta d\beta \right), \quad (2.30)$$

where $c = \log \left(\frac{2^{n-1}m^m}{\pi^2} \right)$. Define the function $g_{\text{SNR}}(\theta, \beta) \triangleq \left(\frac{\sin^2 \theta}{\sin^2 \beta} + \frac{m}{\text{SNR}} \right)^{-m}$, indexed by SNR. Since it is positive and increases to the limiting function $g_{\infty}(\theta, \beta) = \left(\frac{\sin^2 \theta}{\sin^2 \beta} \right)^{-m}$

as SNR increases, we can use the monotone convergence theorem [86] to write

$$\begin{aligned} \lim_{\text{SNR} \rightarrow \infty} \log \left(\int_0^{\frac{\pi}{2}} \int_{\frac{\pi}{M} - \frac{\pi}{2^n}}^{\frac{\pi}{M} + \frac{\pi}{2^n}} \left(\frac{\sin^2 \theta}{\sin^2 \beta} + \frac{m}{\text{SNR}} \right)^{-m} d\theta d\beta \right) \\ = \log \left(\int_0^{\frac{\pi}{2}} \int_{\frac{\pi}{M} - \frac{\pi}{2^n}}^{\frac{\pi}{M} + \frac{\pi}{2^n}} \left(\frac{\sin^2 \theta}{\sin^2 \beta} \right)^{-m} d\theta d\beta \right). \end{aligned}$$

We note that the last integral is finite since $g_\infty(\theta, \beta)$ is continuous and finite over the range of integration. Therefore, after observing that the last term in (2.30) converges to a finite value for each fixed $m \geq \frac{1}{2}$, we get

$$\lim_{\text{SNR} \rightarrow \infty} -\frac{\log(p(\text{SNR}))}{\log(\text{SNR})} = m \quad (2.31)$$

for BPSK modulation with $n > \log_2(M)$. □

The DVO analysis above helps to discover the first-order effects of the low-resolution ADC based receivers on the SEP system performance. In particular, we observe that it is enough to use $\log_2 M + 1$ bits for quantizing the received signal to extract full diversity, which is equal to m for Nakagami- m faded wireless channels. Considering the fact that energy consumption increases exponentially with the number of quantization bits [11], this finding indicates that a significant energy saving is possible by means of low-resolution ADC based receivers without any (first order) loss in communication robustness.

We also observed that the DVO is only equal to $\frac{1}{2}$ when $n = \log_2 M$. Together with the universal bound obtained in Theorem 2.1, the discovered *ternary* behaviour has significant implications in terms of how to choose the number of quantization bits for low-resolution ADC based receivers. In particular, for fading environments with m close to $\frac{1}{2}$, a system designer may decide to trade off reliability for energy consumption, without having too much degradation in average SEP by using $\log_2 M$ bits. On the other hand, for fading environments with large m , it is more beneficial to use one extra bit to have a major improvement in average SEP.

2.6 Performance Analysis for QPSK modulation

In this section, we conduct performance analysis for QPSK modulation with low-resolution ADCs by using our results in previous sections. We first present a simplified version of the average SEP expression in (2.11) for QPSK modulation, and then we analyze the effect of low-resolution quantization under Rayleigh fading.

2.6.1 Symbol Error Probability for QPSK modulation

Nakagami-m Fading: In the special case of QPSK modulation (i.e., $M = 4$), the average SEP expression in (2.11) can be further simplified to produce

$$p(\text{SNR}) = p_1(\text{SNR}) + p_2(\text{SNR}) - p_3(\text{SNR}), \quad (2.32)$$

because $\tan\left(\frac{2\pi}{M}\right) = \infty$ for $M = 4$. By using hypergeometric function ${}_2F_1[\cdot]$ [84], we can simplify (2.32) for 2-bit quantization (i.e., $M = 4$ and $n = 2$) as

$$p(\text{SNR}) = \frac{2}{\pi} \int_0^{\frac{\pi}{2}} {}_2F_1\left[\frac{1}{2}, m, 1, \frac{-\text{SNR}}{m \sin^2 \beta}\right] d\beta \\ - \frac{m^m}{\pi} \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} \left(\frac{\text{SNR}}{\sin^2 \gamma} + m\right)^{-m} {}_2F_1\left[\frac{1}{2}, m, 1, z\right] d\beta d\gamma,$$

where $z = \frac{\text{SNR}(\sin^2 \beta - \sin^2 \gamma)}{\text{SNR} + m \sin^2 \gamma}$.

Rayleigh Fading: For special case of Rayleigh fading, which is obtained by setting $m = 1$, the expression in (2.32) can be re-expressed as (2.33).

$$p(\text{SNR}) = \frac{2^n}{\pi^2} \int_0^{\frac{\pi}{2}} \frac{\sin \beta}{\sqrt{\text{SNR} + \sin^2 \beta}} \arctan\left(\frac{2 \sin \beta \sqrt{\text{SNR} + \sin^2 \beta}}{\text{SNR} + 2 \sin^2 \beta} \tan\left(\frac{\pi}{2^{n-1}}\right)\right) d\beta \\ - \frac{2^{n-1}}{\pi^3} \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} \sqrt{\frac{\sin^2 \beta \sin^2 \gamma}{(\text{SNR} + \sin^2 \beta)(\text{SNR} + \sin^2 \gamma)}} \cdot \arctan(\vartheta) d\beta d\gamma, \quad (2.33)$$

where $\vartheta = \frac{2 \sin \beta \sin \gamma \sqrt{(\text{SNR} + \sin^2 \beta)(\text{SNR} + \sin^2 \gamma)}}{\text{SNR}(\sin^2 \beta + \sin^2 \gamma) + 2 \sin^2 \beta \sin^2 \gamma} \tan\left(\frac{\pi}{2^{n-1}}\right)$.

Furthermore, for 2-bit quantization with Rayleigh fading (i.e., $M = 4$, $n = 2$ and $m = 1$), we can obtain $p(\text{SNR})$ in closed form as

$$p(\text{SNR}) = \frac{2}{\pi} \arctan\left(\frac{1}{\sqrt{\text{SNR}}}\right) - \left(\frac{1}{\pi} \arctan\left(\frac{1}{\sqrt{\text{SNR}}}\right)\right)^2.$$

This closed-form analytical expression is very easy to compute without resorting to any numerical integration.

2.6.2 Analysis of Quantization Penalty for QPSK Modulation

By using the Taylor series expansion for high SNR values, we can re-express the average SEP expressions for QPSK modulation under Rayleigh fading given in (2.33) as (2.34).

$$p_A(\text{SNR}, n) = \begin{cases} \frac{2}{\pi} \text{SNR}^{-\frac{1}{2}} + o\left(\text{SNR}^{-\frac{1}{2}}\right) & n = 2 \\ \frac{2^{n-1}(4\pi-1)}{\pi^3} \tan\left(\frac{\pi}{2^{n-1}}\right) \text{SNR}^{-1} + o\left(\text{SNR}^{-1}\right) & n \geq 3. \end{cases} \quad (2.34)$$

While phase quantization with less number of quantization bits is desirable, due to less processing complexity at the receiver, it deteriorates the average SEP performance of the system. In the following, we quantify the increase in the average SEP as a quantization penalty defined as

$$\Psi(\text{SNR}, n) = 10 \log \left(\frac{p_A(\text{SNR}, n)}{p_A(\text{SNR}, \infty)} \right), \quad (2.35)$$

where $p_A(\text{SNR}, \infty)$ is the average SEP with infinite number of quantization bits. Based on (2.34), we can derive $p_A(\text{SNR}, \infty)$ as

$$p_A(\text{SNR}, \infty) = \left(\frac{4\pi-1}{\pi^2} \right) \text{SNR}^{-1} + o\left(\text{SNR}^{-1}\right), \quad (2.36)$$

where we have used the small-angle approximation $\tan(x) = x$ as $x \rightarrow 0$. Substituting (2.34) and (2.36) into (2.35) and doing some mathematical manipulations, we can derive

the quantization penalty in terms of average SEP with n -bit quantization as (2.37).

$$\Psi(\text{SNR}, n) = \begin{cases} 10 \log \left(\left(\frac{2\pi}{4\pi-1} \right) \text{SNR}^{\frac{1}{2}} \right) + o \left(\text{SNR}^{\frac{1}{2}} \right) & n = 2 \\ 10 \log \left(\frac{2^{n-1}}{\pi} \tan \left(\frac{\pi}{2^{n-1}} \right) \right) + o(1) & n \geq 3. \end{cases} \quad (2.37)$$

In Section 2.7, we use $\Psi(\text{SNR}, n)$ to quantify the increase in average SEP as we change from infinite-bit to n -bit quantization.

We further notice that, in order to achieve the same average SEP as with n -bit quantization, we need to transmit the signal using a higher power if we use only $(n-1)$ -bit quantization. In the following, we quantify the increase in the transmit power as another quantization penalty defined by

$$\Phi(\text{SEP}, n) = 10 \log \left(\frac{\text{SNR}_{n-1}}{\text{SNR}_n} \right), \quad (2.38)$$

where SNR_n and SNR_{n-1} are the SNR values required to achieve a certain average SEP with n and $n-1$ quantization bits, respectively. Substituting (2.34) into (2.38) and doing some mathematical manipulations, we can derive the quantization penalty with n -bit quantization as

$$\Phi(\text{SEP}, n) = \begin{cases} 10 \log \left(\frac{\pi^2}{2(4\pi-1)} (\text{SNR}_2)^{\frac{1}{2}} \right) & n = 3 \\ 10 \log \left(\frac{1}{2} \tan \left(\frac{\pi}{2^{n-2}} \right) \cot \left(\frac{\pi}{2^{n-1}} \right) \right) & n \geq 4. \end{cases} \quad (2.39)$$

In Section 2.7, we use $\Phi(\text{SEP}, n)$ to quantify the required transmit power increase as we change from n -bit to $(n-1)$ -bit quantization.

2.7 Numerical Results

In this section, we present analytical and simulated SEP results for M -PSK modulation with n -bit quantization. Channel fading is unit-power and circularly-symmetric with Nakagami- m distributed magnitude, and additive noise is complex Gaussian with zero mean and unit variance.

2.7.1 Effect of Quantization Bits on System Performance

First, we will investigate the effect of the number of quantization bits on the system performance and illustrate the accuracy of our analytical results obtained in previous sections. Fig. 2.4 plots the average SEP as a function of SNR for QPSK modulation with $n = 2, 3, 4$ -bit quantization under Nakagami- m fading with shape parameter $m = 1$ and 2. The simulated results are generated using Monte Carlo simulation, while the analytical results are generated using our expression in (2.11). As the plot illustrates, the analytical results accurately follow the simulated results for all cases. We observe a noteworthy improvement in the average SEP when n changes from 2 to 3-bit quantization for QPSK modulation in both $m = 1$ and 2. This jump in the average SEP performance is expected in the light of Theorem 2.4, which states that using one extra bit, on top of $\log_2 M$ bits, improves the DVO from $\frac{1}{2}$ to m . We also observe that the average SEP reduces as we increase n , but the amount by which it reduces also gets smaller as we increase n . This can be clearly observed from the zoomed-in section in Fig. 2.4. As expected, $\text{DVO} = m$ for all $n \geq 3$. Furthermore, $\text{DVO} = \frac{1}{2}$ for any m when $n = 2$.

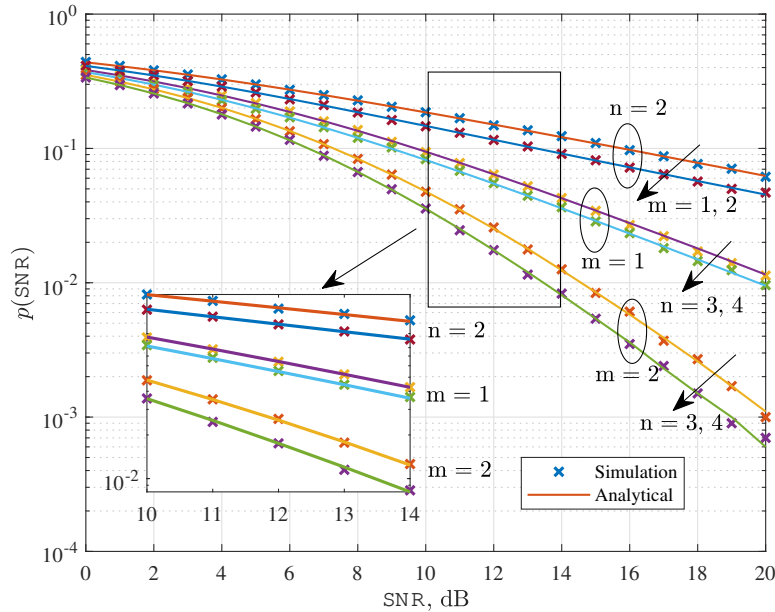


Figure 2.4: Average SEP curves as a function of SNR for QPSK modulation. $n = 2, 3, 4$ and $m = 1, 2$.

Fig. 2.5 plots the average SEP as a function of SNR for QPSK, 8-PSK and 16-PSK modulations schemes while keeping the Nakagami- m shape parameter fixed at $m = 1$, which is the classical Rayleigh fading scenario. We plot the average SEP for each modulation scheme by using $n = \log_2 M$, $\log_2 M + 1$ and $\log_2 M + 2$ bits. From the plots, we can clearly observe that QPSK with 2-bit, 8-PSK with 3-bit and 16-PSK with 4-bit quantization have a DVO of $\frac{1}{2}$. Further, we can observe that QPSK with 3 or more bits, 8-PSK with 4 or more bits, 16-PSK with 5 or more bits quantizations have a DVO of 1, which is equal to m in this case. To further emphasize this point, the zoomed-in section in Fig. 2.5 illustrates the asymptotic average SEP versus SNR for QPSK modulation. As stated in Theorem 2.4, these numerical observations clearly indicate the ternary behaviour in the decay exponent for $p(\text{SNR})$ depending on whether $n \geq \log_2 M + 1$, $n = \log_2 M$, or $n < \log_2 M$.

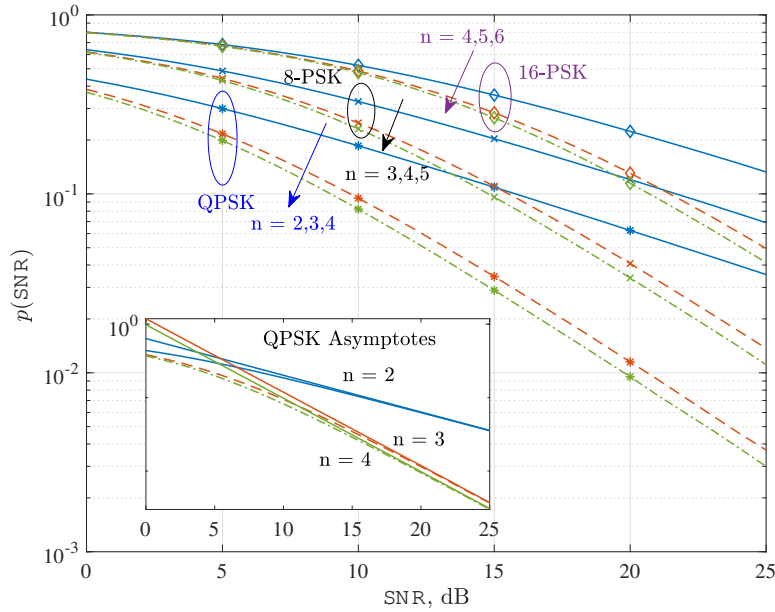


Figure 2.5: Average SEP curves as a function of SNR for different modulation schemes. $n = \log_2 M, \log_2 M + 1, \log_2 M + 2$ and $m = 1$.

In order to illustrate the accuracy of upper and lower bounds on $p(\text{SNR})$, derived in Lemma 2.5, in Fig. 2.6 we plot the expressions in (2.24), alongside the exact $p(\text{SNR})$ curve, as a function of SNR for QPSK, 8-PSK and 16-PSK modulations under Rayleigh fading (i.e., $m = 1$). This figure clearly shows that $U(\text{SNR})$ becomes a very tight upper

bound for 8-PSK and 16-PSK in the high SNR regime. The figure also confirms that the decay exponents of both $L(\text{SNR})$ and $U(\text{SNR})$ are the same as that of the $p(\text{SNR})$.

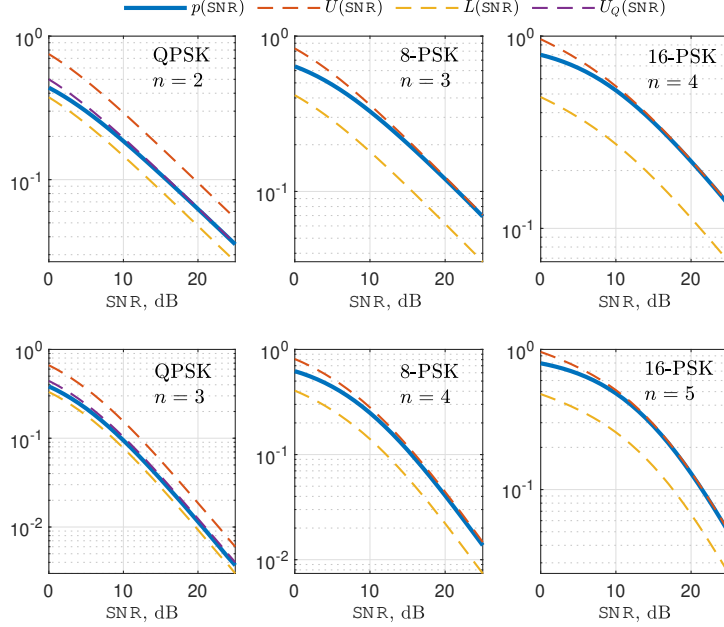


Figure 2.6: Upper and lower bounds on $p(\text{SNR})$ as a function of SNR for QPSK, 8-PSK and 16-PSK modulations. $n = 2, 3, 4, 5$ and $m = 1$.

Next, in Fig. 2.7, we plot the simulated average SEP curves as a function of SNR for 8-PSK modulation with 2-bit quantization and 16-PSK modulation with 2 and 3-bit quantization. We consider equiprobable transmitted symbols and Nakagami- m fading channel model with $m = 0.5, 1$ and 2 . The simulated results are again generated by using Monte Carlo simulations. We can clearly observe an error floor for high SNR values when $n < \log_2 M$, as established by Theorem 2.1. In particular, the average SEP for 8-PSK has a lower bound of 0.5 with 2-bit quantization. Similarly, the average SEP for 16-PSK has a lower bound of 0.75 with 2-bit quantization and a lower bound of 0.5 with 3-bit quantization. It should be noted that the error floor given in Theorem 2.1 is more conservative than those observed in Fig. 2.7. This is because it is a universal lower bound that holds for all modulation schemes, quantizer types and fading environments, not only for very specific ones used to plot average SEP curves in Fig. 2.7.

Fig. 2.8 illustrates the quantization penalty and plots the asymptotic average SEP curves as a function of SNR for QPSK modulation with $n = 2, 3, 4$ and ∞ under

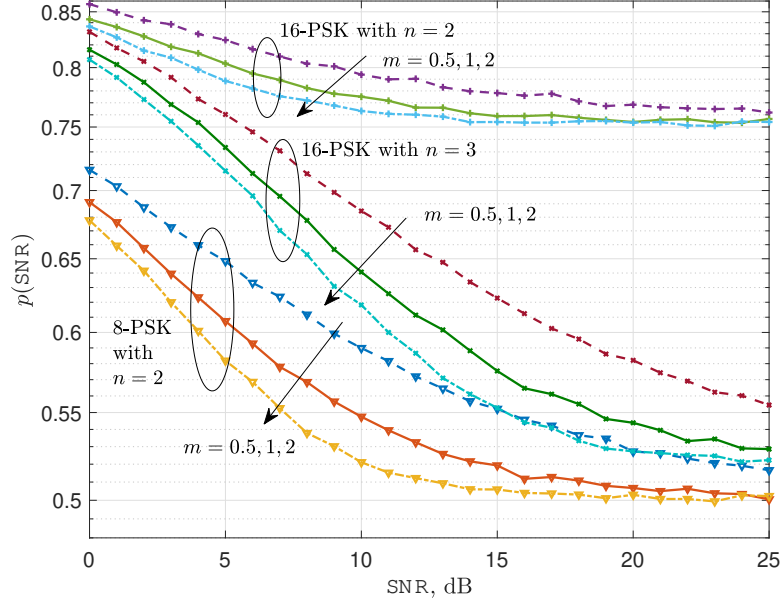


Figure 2.7: Average SEP as a function of SNR for 8-PSK and 16-PSK modulations. $n = 2 < \log_2 M$ and $m = 0.5, 1$ and 2 .

Rayleigh fading. The asymptotic plots are generated by using the expressions in (2.34). We observe that a DVO of half is achieved with 2-bit quantization, and the full DVO of one is achieved with $n > 2$. When the SNR is fixed at 18 dB, we observe a quantization penalty of $\Psi(18 \text{ dB}, 2) \approx 6.35 \text{ dB}$ as we change from $n = 2$ to ∞ , i.e., we get a 5-fold increase in the average SEP as we change from $n = 2$ to ∞ . When the average SEP is fixed at 0.015, we observe a quantization penalty of $\Phi(0.015, 4) \approx 0.8 \text{ dB}$ as we change from $n = 3$ to 4, i.e., to achieve an average SEP of 0.015 with 4-bit quantization, we need 0.8 dB more transmit power than what is required with 3-bit quantization.

2.7.2 Comparison with Independent I and Q Quantization

In this part, we will compare and contrast the system performance results obtained by using M -PSK modulation with the commonly adopted approach using independent I and Q channel quantization. To this end, Fig. 2.9 provides performance comparison curves for Rayleigh fading and QPSK modulation. We set the number of quantization bits to $n = 2, 3, 4$ and 6 by numerically optimizing the quantization regions for inde-

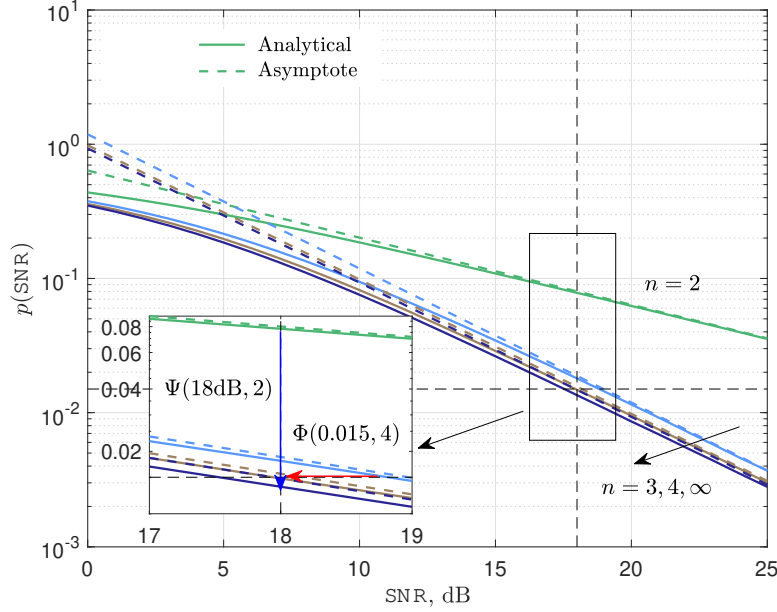


Figure 2.8: Quantization penalty for QPSK modulation and the asymptotic average SEP curves. $n = 2, 3, 4, \infty$ and at $m = 1$.

pendent I and Q quantization. In order to cover all possibilities in the latter case, we also consider, in addition to the identical bit allocation strategy, unbalanced allocation of quantization bits to the I and Q arms.

For $n = 2$, 2-bit phase quantization SEP performance coincides with that of using 1-bit quantization on each of the I and Q arms. The reason for this behaviour is that the optimized quantization regions coincide with the phase quantization regions, as introduced in Section 2.2 in the form of convex cones in $\mathbb{C}$. When $n = 3$, the 3-bit phase quantization clearly outperforms independent I and Q quantization with 2 bits on the I arm and 1 bit on the Q arm (and with optimized quantization regions). When $n = 4$, perhaps more interestingly, we observe that independent I and Q quantization with optimized quantization regions for 2 bits on both arms performs very similarly to the 3-bit phase quantization. Hence, for this particular allocation of quantization bits to the I and Q arms, the phase quantization approach introduced in this chapter has the advantage of saving one quantization bit when compared to the commonly used independent I and Q quantization.

To test the robustness of our observations above, we have also obtained the SEP

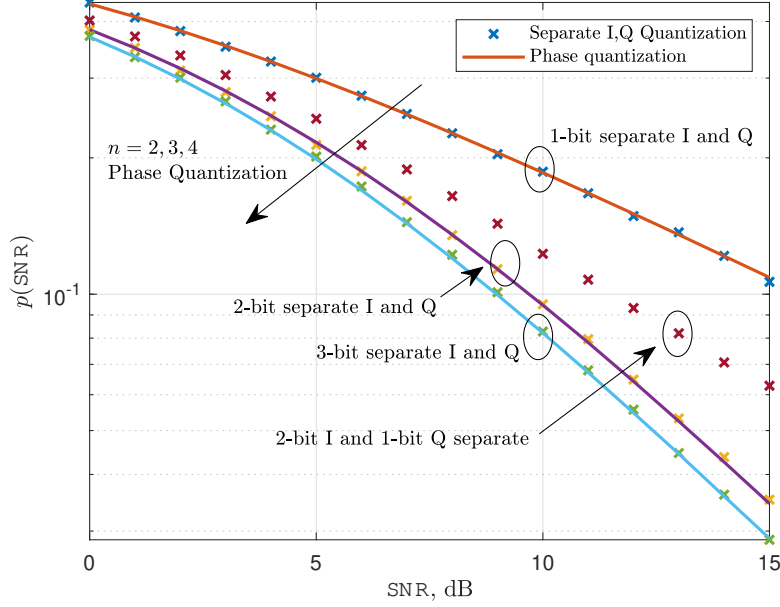


Figure 2.9: Average SEP as a function of SNR for QPSK modulation with phase quantization and separate I and Q quantization, $m = 1$.

curves for the 6-bit independent I and Q quantization, with 3 bits on each of the I and Q arms. Our results indicate that the SEP performance for 6-bit independent I and Q quantization is similar to 4-bit phase quantization. Surprisingly, we need two more additional bits in the case of independent I and Q quantization in order to match the SEP performance attained by 4-bit phase quantization. Similar results with an increased performance gap continue to hold with the unbalanced allocation of bits on the I and Q arms. While these results provide some initial insights into the bit-advantage of using phase quantization over independent I and Q quantization, a more thorough investigation is required to obtain general proofs establishing this phenomenon rigorously.

2.7.3 Effect of Channel Estimation Errors on DVO

In this part, we provide a robustness analysis for our analytical results under channel estimation errors. As we discussed above, we have assumed the availability of full-precision CSI at the receiver to derive the DVO results given in Theorem 2.4.

This assumption was motivated by the previous work establishing the feasibility of accurate channel estimation by means of carefully designed training sequences for low-resolution ADC receivers as well as by the availability of mixed-ADC architectures. Below, we will focus on Rayleigh fading ($m = 1$) and QPSK ($M = 4$) modulation but our conclusions continue to hold more generally.

In order to model channel estimation errors in the numerical study below, we first recall that our ML detector in Theorem 2.2 only requires the knowledge of the channel phase to determine the most likely transmitted symbol. Hence, we can model the channel estimation problem as that of estimating the phase information of H .⁴ For the purpose of analyzing the effect of channel estimation errors on the SEP performance, this shows that the channel estimation errors due to low-resolution quantization can be modelled as bounded uniform random variables with the MMSE estimator being the mid-point of the quantization regions (i.e., H has a circularly symmetric distribution). Hence, Figs. 2.10 and 2.11 are generated based on the assumption of bounded uniform channel estimation errors, with a span of $\frac{2\pi}{2^b}$ when b bits are used for channel estimation.⁵

A salient feature arising from the SEP curves presented in Figs. 2.10 and 2.11 is that the derived fundamental ternary behaviour for the decay exponent of $p(\text{SNR})$ in Theorem 2.4 stays the same if $n + 1$ bits are used during the channel estimation stage. We recall that n is the number of bits used for data decoding. Moreover, the performance difference between using n and $n + 1$ bits for channel estimation becomes negligibly small as n increases. We note that the use of a different number of bits for data decoding and channel estimation is possible by employing mixed-ADC receiver architectures commonly used in the literature [71]. Furthermore, the same effect to improve channel estimation accuracy can be achieved by using carefully designed training sequences to perform a bisection search in $[-\pi, \pi)$ during the channel estimation stage.

The robustness of our results against channel estimation errors in Figs. 2.10 and

⁴In addition to assisting in modeling channel estimation errors, this is also practically an important observation since it signifies that the same phase quantizer in a mixed-ADC architecture can be used for both channel estimation and data decoding by only dynamically changing the number of quantization bits.

⁵Here, we ignore the channel noise to focus on channel estimation errors due to quantization noise only.

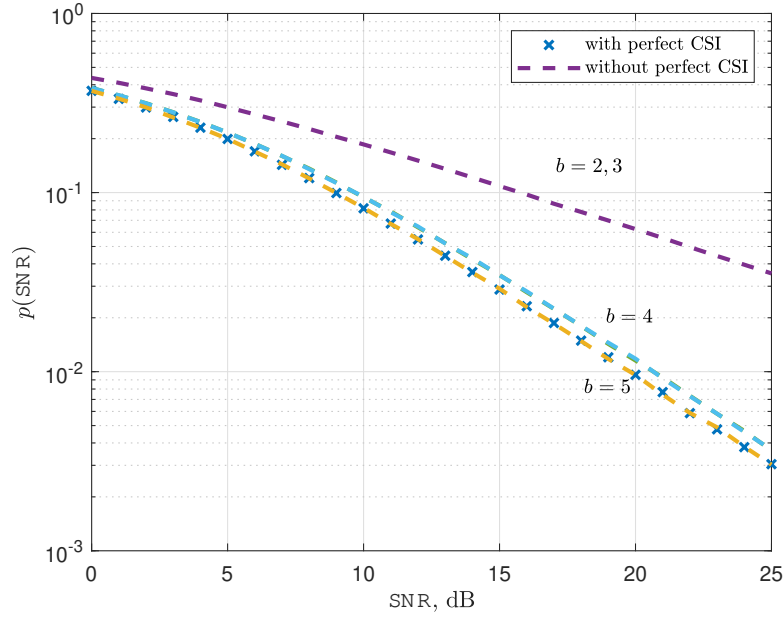


Figure 2.10: Average SEP as a function of SNR for QPSK modulation with 4-bit quantization and imperfect CSI, $m = 1$, b is the number of quantization bits used for channel estimation.

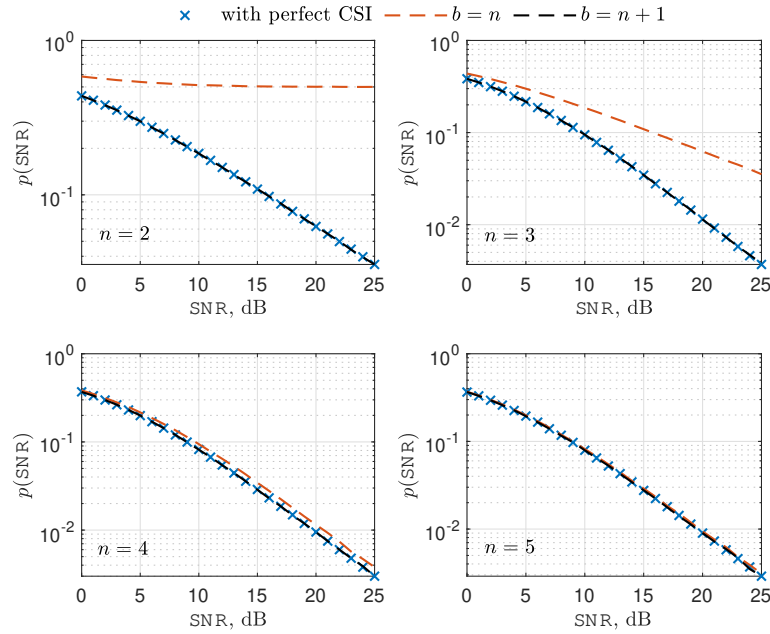


Figure 2.11: Average SEP as a function of SNR for QPSK modulation with n -bit quantization with and without perfect CSI, $m = 1$, $n = 2, 3, 4, 5$, $b = n, n + 1$, where b is the number of quantization bits used for channel estimation.

2.11 arise from the following two key properties of the approach employed in the paper. The first property leading to the observed robustness behaviour is the fact that the optimum ML detector we obtained in Theorem 2.2 is only dependent on the channel phase information. The second one is the centering property achieved by the number of quantization bits steering the received symbols away from the decision boundaries and leading to the observed ternary behaviour for the decay exponent of $p(\text{SNR})$ with full-precision CSI, as explained in Section 2.4.2 in detail. The same effect continues to hold for the channel estimation, with one more additional bit providing the required accuracy in channel phase estimation and thereby preventing the bounded channel phase estimation errors from causing the received symbols to flip over the adjacent quantization regions.

2.8 Conclusions and Future Generalizations

In this chapter, we performed a theoretical analysis of a low-resolution based ADC communication system and obtained fundamental performance limits, optimum ML detectors and a general analytical expression for the average SEP for M -PSK modulation with n -bit quantization. These results were further investigated for Nakagami- m fading model in detail. We conducted an asymptotic analysis to show that the decay exponent for the average SEP is the same and equal to m with infinite-bit and n -bit quantizers for $n \geq \log_2 M + 1$. We also performed an extensive numerical study to illustrate the accuracy of the derived analytical expressions, comparison with independent I and Q channel quantization and robustness under channel estimation errors.

In most parts of the chapter, we have focused on phase modulated communications. Phase modulation has an important and practical layering feature enabling the quantizer and detector design separation in low-resolution ADC communications. For a given number of bits, the quantizer needs to be designed only once and can be kept constant for all channel realizations. The detector can be implemented digitally as a table look-up procedure using channel knowledge and quantizer output. On the other hand, this feature is lost in joint phase and amplitude modulation schemes such as

QAM. The quantizer needs to be dynamically updated for each channel realization in low-resolution ADC based QAM systems. This is because the fading channel amplitude may vary over a wide range, but the phase always varies over $[-\pi, \pi)$. However, phase modulation is historically known to be optimum only up to modulation order 16 under peak power limitations [87]. Hence, it is a notable future research direction to extend the results of this chapter to higher order phase and amplitude modulations by taking practical design considerations into account.

A major result of this chapter is the discovery of a ternary SEP behaviour, indicating the sufficiency of $\log_2 M + 1$ bits for achieving asymptotically optimum M -ary communication reliability. Hence, without modifying the conventional RF chain, we can use one extra bit and still achieve asymptotically optimum communication performance. Another important future research direction is to compare and contrast the backwards-compatible receiver design approach of using one extra bit proposed in this chapter with other approaches that can potentially modify the conventional RF chain and manipulate the received signals in the waveform domain by introducing extra analog components. This study needs to be done in detail by considering the accuracy and agility of analog domain operations, the energy consumption of analog and digital circuit components, different modulation schemes and the average SEP performance curves resulting from different low-resolution ADC based receiver architectures. Similarly, utilizing the results of this chapter, a further detailed study on the receiver architecture design to determine where to place the diversity combiner (before or after quantizer or detector) and its type is needed when multiple diversity branches are available for data reception.

2.9 Appendix

2.9.1 Proof of Theorem 2.1

Let us consider a class of hypothetical genie-aided detectors $g : \mathbb{C}^2 \times [0 : 2^n - 1] \rightarrow [0 : M - 1]$ that has the knowledge of channel noise $W \in \mathbb{C}$, fading coefficient $H \in \mathbb{C}$ and

quantizer output $Q(Y) \in [0 : 2^{n-1}]$. We also let $\mathcal{S}_{w,h,k} = \{x \in \mathcal{C} : \sqrt{\text{SNR}}hx + w \in \mathcal{R}_k\}$ be the set of transmitted signal points resulting in $Q(Y) = k$ for particular realizations of $H = h$ and $W = w$. We first observe that since $n < \log_2 M$, there exists at least one quantization region $\mathcal{R}_{\tilde{k}}$ (depending on w and h) such that $\mathcal{S}_{w,h,\tilde{k}}$ contains at least $\frac{M}{2^n}$ signal points. We note that $\frac{M}{2^n}$ is always an integer greater than or equal to 2 since M is assumed to be an integer power of 2. Then, the conditional SEP of any detector g given $W = w$ and $H = h$, which we will denote by $p_g(\text{SNR}, h, w)$, can be lower-bounded as (2.40).

$$\begin{aligned} p_g(\text{SNR}, h, w) &\geq p_{\min} \sum_{x_i \in \mathcal{S}_{w,h,\tilde{k}}} \Pr\{g(h, w, \tilde{k}) \neq x_i \mid W = w, H = h, X = x_i\} \\ &\geq p_{\min} \left(\frac{M}{2^n} - 1 \right) \\ &= \frac{M - 2^n}{2^n} p_{\min}, \end{aligned} \tag{2.40}$$

By averaging with respect to w and h , we have $p_g(\text{SNR}) \geq \frac{M-2^n}{2^n} p_{\min}$, where $p_g(\text{SNR})$ is the average SEP corresponding to detector g . This concludes the proof since the obtained lower bound does not depend on the choice of modulation scheme, quantizer structure and detector rule, and hence holds for detectors not utilizing the knowledge of W for any choice of modulation scheme and quantizer structure.

2.9.2 Proof of Theorem 2.2

To prove Theorem 2.2, we will first obtain the following result.

Lemma 2.6. *Let $\mathcal{R}$ be a convex cone given by $\mathcal{R} = \{z \in \mathbb{C} : \alpha_1 \leq \text{Arg}(z) \leq \alpha_2\}$ for $\alpha_1, \alpha_2 \in [-\pi, \pi)$, and $W_1 \sim \mathcal{CN}(\mu_1, 1)$ and $W_2 \sim \mathcal{CN}(\mu_2, 1)$ be proper complex Gaussian random variables with means satisfying $|\mu_1| = |\mu_2| = r$ for some $r > 0$. Then, $\Pr\{W_1 \in \mathcal{R}\} \geq \Pr\{W_2 \in \mathcal{R}\}$ if $|\mu_1 - z_{\text{mid}}| \leq |\mu_2 - z_{\text{mid}}|$, where $z_{\text{mid}} = re^{j\frac{\alpha_1 + \alpha_2}{2}}$.*

Proof: It is enough to show this result only for $\alpha_2 = -\alpha_1 = \alpha$. Otherwise, we can first rotate W_1, W_2 and $\mathcal{R}$ with $e^{-j\frac{\alpha_1 + \alpha_2}{2}}$ and repeat the same calculations below. Let $g(\mu_i) = \Pr\{W_i \in \mathcal{R}\}$ for $i = 1, 2$, and assume $|\mu_1 - z_{\text{mid}}| \leq |\mu_2 - z_{\text{mid}}|$. There are

multiple cases in which the inequality $|\mu_1 - z_{\text{mid}}| \leq |\mu_2 - z_{\text{mid}}|$ holds, which we will analyze one-by-one below.

First, we will consider the case in which both μ_1 and μ_2 lie outside $\mathcal{R}^\circ$, where $\mathcal{R}^\circ$ is the set of interior points of $\mathcal{R}$. This is the case shown in Fig. 2.12. To start with, we

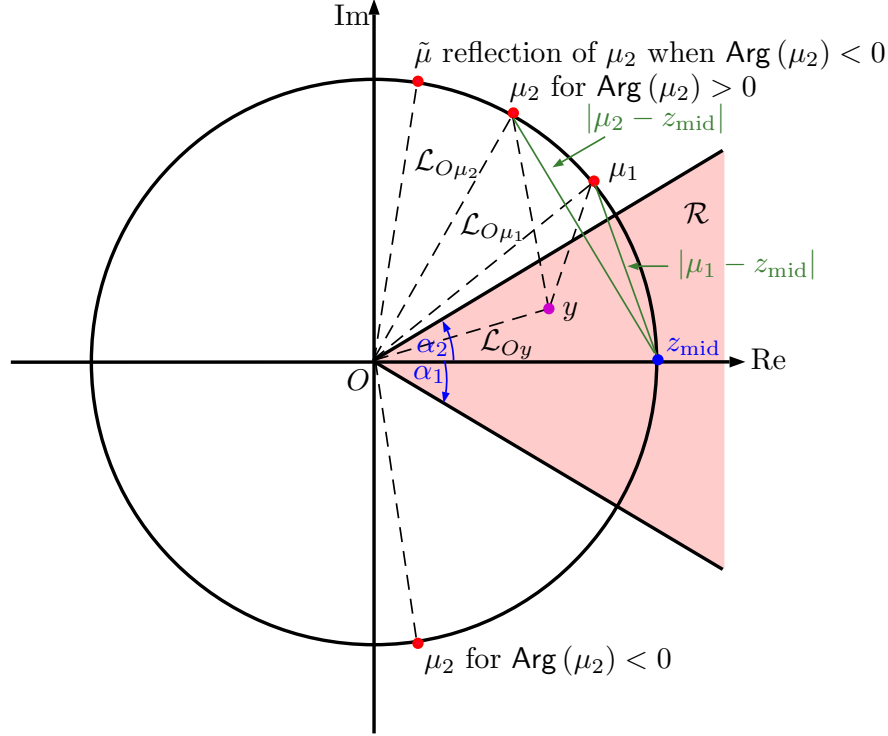


Figure 2.12: An illustration for the proof of Lemma 2.6 when μ_1 and μ_2 lie outside $\mathcal{R}^\circ$. $|\mu_1| = |\mu_2| = r$, $|\mu_1 - z_{\text{mid}}| \leq |\mu_2 - z_{\text{mid}}|$ and $\alpha_2 = -\alpha_1 = \alpha$.

will assume $0 \leq \text{Arg}(\mu_1) \leq \text{Arg}(\mu_2) < \pi$. Then, for any $y \in \mathcal{R}$, the angle between the line segments $\mathcal{L}_{Oy}$ and $\mathcal{L}_{O\mu_1}$ is smaller than the one between the line segments $\mathcal{L}_{Oy}$ and $\mathcal{L}_{O\mu_2}$.⁶ Hence, applying the cosine rule for the triangle formed by O, y and μ_1 , and for the triangle formed by O, y and μ_2 , it can be seen that $|y - \mu_1| \leq |y - \mu_2|$ for all $y \in \mathcal{R}$.⁷ Therefore, $g(\mu_1) = \frac{1}{\pi} \int_{\mathcal{R}} \exp(-|y - \mu_1|^2) dy \geq \frac{1}{\pi} \int_{\mathcal{R}} \exp(-|y - \mu_2|^2) dy = g(\mu_2)$. Next, we assume $\text{Arg}(\mu_2) \in [-\pi, 0)$ and $0 \leq \text{Arg}(\mu_1) \leq |\text{Arg}(\mu_2)| \leq \pi$. Let $\tilde{W}$ be the auxiliary random variable distributed according to $\mathcal{CN}(\tilde{\mu}, 1)$ with $\tilde{\mu} = re^{j|\text{Arg}(\mu_2)|}$, i.e., $\tilde{\mu}$ is the reflection of μ_2 around the real line. Symmetry around the real line implies that

⁶The line segment $\mathcal{L}_{z_1 z_2}$ between the points $z_1 \in \mathbb{C}$ and $z_2 \in \mathbb{C}$ is defined as $\mathcal{L}_{z_1 z_2} = \{(1-t)z_1 + tz_2 : t \in [0, 1]\}$.

⁷This statement is correct even when both y and μ_1 lies on the boundary of $\mathcal{R}$ and the triangle formed by O, y and μ_1 reduces to a line segment.

$g(\mu_2)$ is equal to $g(\tilde{\mu}) = \Pr\{\tilde{W} \in \mathcal{R}\}$, which is less than $g(\mu_1)$ due to our arguments above. For $\text{Arg}(\mu_1) \in [-\pi, 0)$, the same analysis still holds after reflecting μ_1 around the real line, leading to $g(\mu_1) \geq g(\mu_2)$ for all $\mu_1, \mu_2 \notin \mathcal{R}^\circ$ satisfying $|\mu_1 - z_{\text{mid}}| \leq |\mu_2 - z_{\text{mid}}|$.

Second, we consider the case where $\mu_1 \in \mathcal{R}^\circ$ but $\mu_2 \notin \mathcal{R}^\circ$. This is the case shown in Fig. 2.13. It is enough to establish the desired result only for $0 \leq \text{Arg}(\mu_1) \leq \text{Arg}(\mu_2) <$

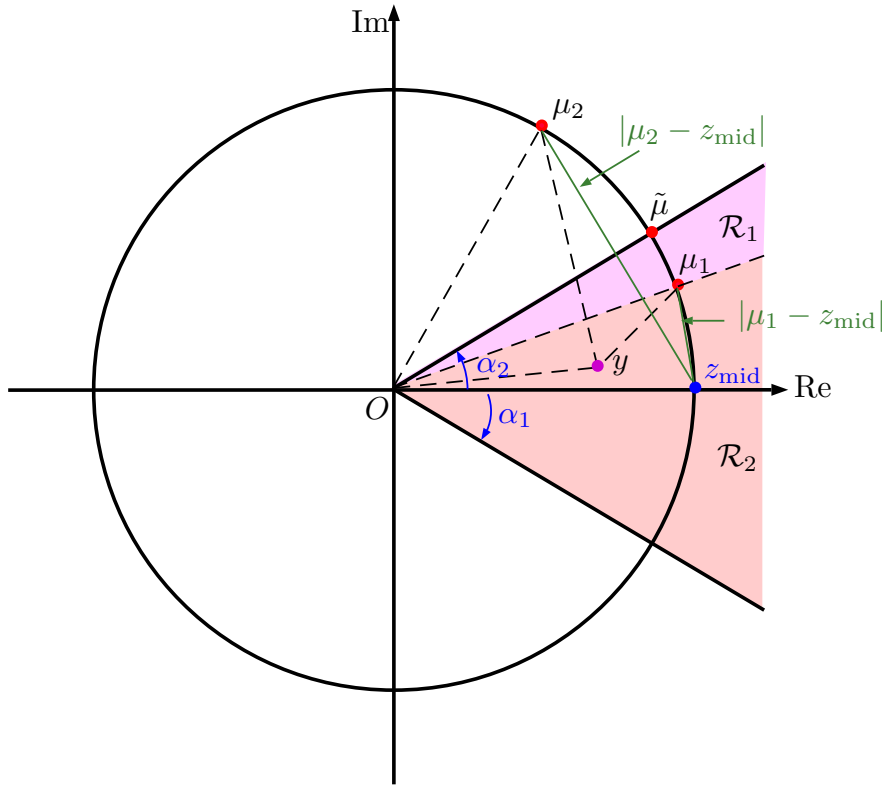


Figure 2.13: An illustration for the proof of Lemma 2.6 when $\mu_1 \in \mathcal{R}^\circ$ and $\mu_2 \notin \mathcal{R}^\circ$. $|\mu_1| = |\mu_2| = r$, $|\mu_1 - z_{\text{mid}}| \leq |\mu_2 - z_{\text{mid}}|$ and $\alpha_2 = -\alpha_1 = \alpha$.

π . When μ_1 or μ_2 has a negative phase angle, the same analysis below still holds after reflecting the mean with negative phase around the real line. Let $\tilde{W}$ be the auxiliary random variable distributed according to $\mathcal{CN}(\tilde{\mu}, 1)$ with $\tilde{\mu} = re^{j\alpha}$, i.e., $\tilde{\mu}$ is located at the upper boundary of $\mathcal{R}$. Our analysis in the first case shows that $g(\tilde{\mu}) = \Pr\{\tilde{W} \in \mathcal{R}\} \geq g(\mu_2)$ since both $\tilde{\mu}$ and μ_2 are outside $\mathcal{R}^\circ$ and $0 \leq \text{Arg}(\tilde{\mu}) \leq \text{Arg}(\mu_2) < \pi$. We next divide $\mathcal{R}$ into two disjoint regions: $\mathcal{R}_1 = \{z \in \mathbb{C} : \text{Arg}(\mu_1) \leq \text{Arg}(z) \leq \alpha\}$ and $\mathcal{R}_2 = \{z \in \mathbb{C} : -\alpha \leq \text{Arg}(z) < \text{Arg}(\mu_1)\}$. Then, we have $\Pr\{\tilde{W} \in \mathcal{R}_1\} = \Pr\{W_1 \in \mathcal{R}_1\}$

due to symmetry around the line bisecting $\mathcal{R}_1$ and $\Pr \{ \tilde{W} \in \mathcal{R}_2 \} \leq \Pr \{ W_1 \in \mathcal{R}_2 \}$ since $|y - \mu_1| \leq |y - \tilde{\mu}|$ for all $y \in \mathcal{R}_2$. Hence, $g(\mu_1) \geq g(\tilde{\mu}) \geq g(\mu_2)$. This establishes the desired results for all $\mu_1 \in \mathcal{R}^\circ, \mu_2 \notin \mathcal{R}^\circ$ satisfying $|\mu_1 - z_{\text{mid}}| \leq |\mu_2 - z_{\text{mid}}|$.

Finally, we will consider the third case where both μ_1 and μ_2 lie inside $\mathcal{R}^\circ$. This is the case shown in Fig. 2.14. Similar to the first two cases, it is enough to focus only on

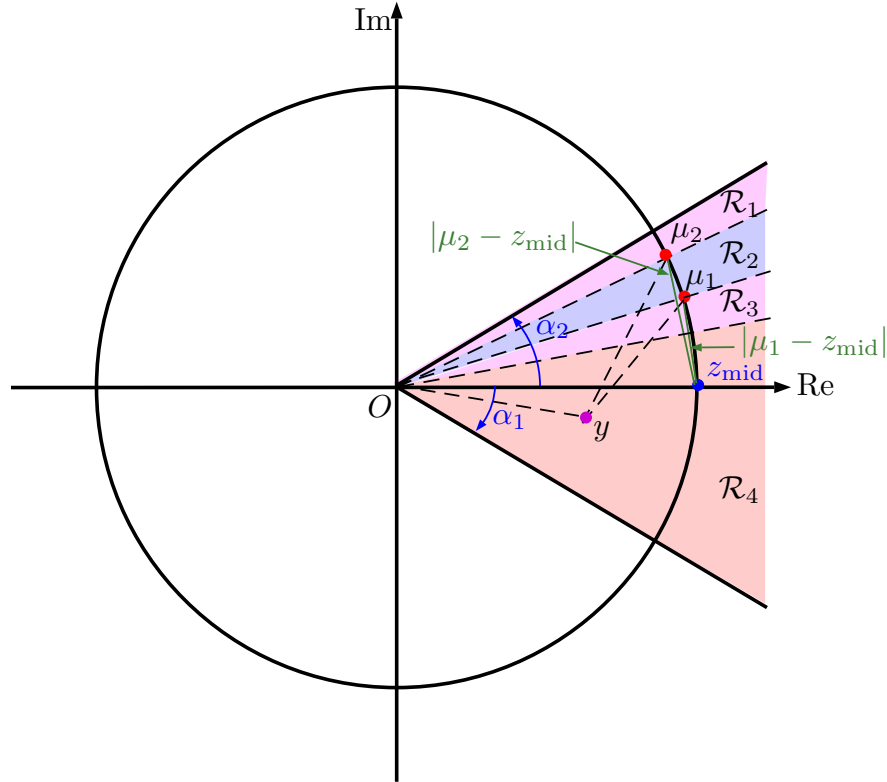


Figure 2.14: An illustration for the proof of Lemma 2.6 when μ_1 and μ_2 lie inside $\mathcal{R}^\circ$. $|\mu_1| = |\mu_2| = r$, $|\mu_1 - z_{\text{mid}}| \leq |\mu_2 - z_{\text{mid}}|$ and $\alpha_2 = -\alpha_1 = \alpha$.

$0 \leq \text{Arg}(\mu_1) \leq \text{Arg}(\mu_2) \leq \alpha$. We divide $\mathcal{R}$ into four disjoint regions given by

$$\mathcal{R}_1 = \{z \in \mathbb{C} : \text{Arg}(\mu_2) \leq \text{Arg}(z) \leq \alpha\}$$

$$\mathcal{R}_2 = \{z \in \mathbb{C} : \text{Arg}(\mu_1) \leq \text{Arg}(z) < \text{Arg}(\mu_2)\}$$

$$\mathcal{R}_3 = \{z \in \mathbb{C} : \text{Arg}(\mu_1) + \text{Arg}(\mu_2) - \alpha \leq \text{Arg}(z) < \text{Arg}(\mu_1)\}$$

$$\mathcal{R}_4 = \{z \in \mathbb{C} : -\alpha \leq \text{Arg}(z) < \text{Arg}(\mu_1) + \text{Arg}(\mu_2) - \alpha\}$$

Using the symmetry in the problem, we have $\Pr \{W_1 \in \mathcal{R}_1\} = \Pr \{W_2 \in \mathcal{R}_3\}$, $\Pr \{W_1 \in \mathcal{R}_2\} =$

$\Pr\{W_2 \in \mathcal{R}_2\}$ and $\Pr\{W_1 \in \mathcal{R}_3\} = \Pr\{W_2 \in \mathcal{R}_1\}$. On the other hand, $\Pr\{W_1 \in \mathcal{R}_4\} \geq \Pr\{W_2 \in \mathcal{R}_4\}$ since $|y - \mu_1| \leq |y - \mu_2|$ for all $y \in \mathcal{R}_4$. Hence, $g(\mu_1) \geq g(\mu_2)$ when both μ_1 and μ_2 lie inside $\mathcal{R}^\circ$, which completes the proof. $\square$

Now, we will utilize Lemma 2.6 to prove Theorem 2.2. For $Q(Y) = k$, the ML detector given in (2.7) reduces to finding a signal point in $\mathcal{C}$ maximizing the probability $\Pr\{\sqrt{\text{SNR}}hx + W \in \mathcal{R}_k\}$, i.e.,

$$\hat{x}(k, h) \in \arg \max_{x \in \mathcal{C}} \Pr\{\sqrt{\text{SNR}}hx + W \in \mathcal{R}_k\}.$$

By Lemma 2.6, $\hat{x}(k, h)$ is the signal point in $\mathcal{C}$ such that $h\hat{x}(k, h)$ is closest to $z_k = re^{j(\frac{2\pi}{2^n}k + \frac{\pi}{2^n})}$, where $r = |h|$. Further, $\hat{x}(k, h)$ is unique with probability one due to the continuity assumption of the fading distribution. Consider now the semi-circle $\mathcal{S} = \{z \in \mathbb{C} : |z| = r \text{ and } (\frac{2\pi}{2^n}k + \frac{\pi}{2^n} - \frac{\pi}{2}) \leq \text{Arg}(z) \leq (\frac{2\pi}{2^n}k + \frac{\pi}{2^n} + \frac{\pi}{2})\}$ centered around z_k and having $\mathcal{H}_k$ as its bisector. The semi-circle $\mathcal{S}$ is illustrated in Fig 2.15. Let $x^* \in \arg \min_{x \in \mathcal{C}} \text{dist}(hx, \mathcal{H}_k)$. For the M -PSK modulation scheme ($M \geq 2$) with regularly spaced signal points on the unit circle, we always have $hx^* \in \mathcal{S}$ and $h\hat{x}(k, h) \in \mathcal{S}$. Take now another signal point $x \in \mathcal{C}$ different than x^* and satisfying $hx \in \mathcal{S}$. Consider the triangle formed by $0, z_k$ and hx^* , and the one formed by $0, z_k$ and hx . We first observe that the area of the first triangle is smaller than the area of the second one since they share the line segment $\mathcal{L}_{0z_k}$ as their common base but the height of the first one $\text{dist}(hx^*, \mathcal{H}_k)$ corresponding to this base is smaller than the height of the second one $\text{dist}(hx, \mathcal{H}_k)$ corresponding to the same base. This is also illustrated in Fig. 2.15. This observation, in turn, implies $|hx^* - z_k| \leq |hx - z_k|$ because the remaining side lengths of both triangles are equal to r . Since this is correct for any $x \in \mathcal{C}$ satisfying $hx \in \mathcal{S}$, we conclude that x^* is unique and equal to $x^* = \hat{x}(k, h)$.

2.9.3 Proof of Lemma 2.1

The proof of Lemma 2.1 is based on an application of the law of total probability [88]. To this end, we consider a partition $\{\mathcal{D}_k\}_{k=0}^{2^n-1}$ of $\mathbb{C}$, where each element of this partition is given by $\mathcal{D}_k = \{z \in \mathbb{C} : (2k-1)\frac{\pi}{2^n} \leq \text{Arg}(z) + \pi < (2k+1)\frac{\pi}{2^n}\}$ for $k \in [1 : 2^n - 1]$,

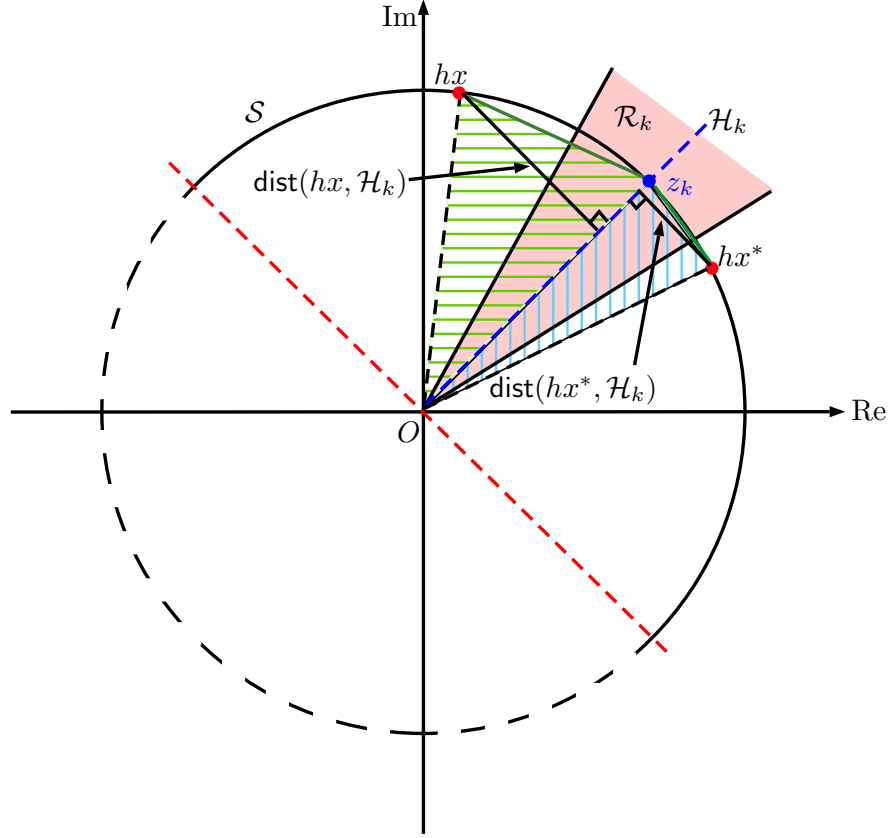


Figure 2.15: An illustration for the proof of Theorem 2.2 where $|hx^* - z_k| \leq |hx - z_k|$.

and $\mathcal{D}_0 = \{z \in \mathbb{C} : \pi - \frac{\pi}{2^n} \leq \text{Arg}(z) < \pi\} \cup \{z \in \mathbb{C} : -\pi \leq \text{Arg}(z) < \frac{\pi}{2^n} - \pi\}$. Let $x_i = e^{j\pi(\frac{2i+1}{M}-1)}$ be the i th signal point in the constellation set $\mathcal{C}$ for $i \in [0 : M-1]$. Then, we can express $p(\text{SNR})$ according to (2.41).

$$p(\text{SNR}) = \frac{1}{M} \sum_{i=0}^{M-1} \sum_{k=0}^{2^n-1} \int_{\mathcal{D}_k} \Pr\{x_i \neq \hat{x}(Q(Y), h) | H = h, X = x_i\} f_H(h) dh. \quad (2.41)$$

We will show that all the terms in (2.41) are equal to each other. Next, we define $\mathcal{E}_i = \{z \in \mathbb{C} : \text{Arg}(x_i) - \frac{\pi}{M} \leq \text{Arg}(z) < \text{Arg}(x_i) + \frac{\pi}{M}\}$ for $i \in [0 : M-1]$. Note that $\mathcal{E}_i$ contains all $\mathcal{H}_k$'s (i.e., bisectors of quantization regions) to which x_i is the closest signal point since x_i 's are uniformly spaced on the unit circle in $\mathbb{C}$. Furthermore, this statement continues to be true for $\sqrt{\text{SNR}}hx_i$ as long as $\text{Arg}(h) \in [-\frac{\pi}{2^n}, \frac{\pi}{2^n})$ since the angular spacing between $\mathcal{H}_k$'s is uniform and equal to $\frac{2\pi}{2^n}$. Notice that $\text{Arg}(h) \in [-\frac{\pi}{2^n}, \frac{\pi}{2^n})$ if and only if $h \in \mathcal{D}_{2^n-1}$. On the other hand, if $\text{Arg}(h) \in [\frac{\pi}{2^n}, \frac{3\pi}{2^n})$, the region

$e^{j\frac{2\pi}{2^n}} \mathcal{E}_i = \left\{ e^{j\frac{2\pi}{2^n}} z \in \mathbb{C} : \text{Arg}(x_i) - \frac{\pi}{M} \leq \text{Arg}(z) < \text{Arg}(x_i) + \frac{\pi}{M} \right\}$ contains all $\mathcal{H}_k$'s to which $\sqrt{\text{SNR}}hx_i$ is the closest. Notice also that $\text{Arg}(h) \in \left[\frac{\pi}{2^n}, \frac{3\pi}{2^n} \right)$ if and only if $h \in \mathcal{D}_{2^{n-1}+1}$. Similarly, $e^{-j\frac{2\pi}{2^n}} \mathcal{E}_i$ contains all $\mathcal{H}_k$'s to which $\sqrt{\text{SNR}}hx_i$ is closest if $\text{Arg}(h) \in \left[-\frac{3\pi}{2^n}, -\frac{\pi}{2^n} \right)$, and $\text{Arg}(h) \in \left[-\frac{3\pi}{2^n}, -\frac{\pi}{2^n} \right)$ if and only if $h \in \mathcal{D}_{2^{n-1}-1}$. The same idea extends to any $\mathcal{D}_k$, and we define

$$\mathcal{E}_{i,k} \triangleq \exp \left(j \left(k - 2^{n-1} \right) \frac{2\pi}{2^n} \right) \mathcal{E}_i, \quad (2.42)$$

for $i \in [0 : M-1]$ and $k \in [0 : 2^n - 1]$. We will use the sets defined in (2.42) to show that all the terms in (2.41) are equal.

To complete the proof, we let $p_{i,k} = \int_{\mathcal{D}_k} \Pr \{x_i \neq \hat{x}(Q(Y), h) | H = h, X = x_i\} f_H(h) dh$ for $i \in [0 : M-1]$ and $k \in [0 : 2^n - 1]$. We also define

$$\theta'_i = -\pi \left(\frac{2i}{M} - 1 \right), \theta''_k = - \left(k - 2^{n-1} \right) \frac{2\pi}{2^n}, \text{ and } \theta_{i,k} = \theta'_i + \theta''_k$$

for $i \in [0 : M-1]$ and $k \in [0 : 2^n - 1]$. We first observe that $e^{j\theta_{i,k}} \mathcal{E}_{i,k} = \mathcal{E}_{\frac{M}{2}}$ since multiplication with $e^{j\theta'_i}$ rotates the i th signal point to $x_{\frac{M}{2}}$ and multiplication with $e^{j\theta''_k}$ removes the effect of partition selection for h . Secondly, we observe that when $h \in \mathcal{D}_k$, the event $\{x_i \neq \hat{x}(Q(Y), h)\}$ is equivalent to $\{Y \notin \mathcal{E}_{i,k}\}$ since $\mathcal{E}_{i,k}$ contains all bisectors to which $\sqrt{\text{SNR}}hx_i$ is closest for this range of h values. Hence, the following chain of equalities hold:

$$\begin{aligned} p_{i,k} &\stackrel{(a)}{=} \int_{\mathcal{D}_k} \Pr \left\{ \sqrt{\text{SNR}}hx_i + W \notin \mathcal{E}_{i,k} \right\} f_H(h) dh \\ &= \int_{\mathcal{D}_k} \Pr \left\{ W \notin \mathcal{E}_{i,k} - \sqrt{\text{SNR}}hx_i \right\} f_H(h) dh \\ &= \int_{\mathcal{D}_k} \Pr \left\{ e^{j\theta_{i,k}} W \notin e^{j\theta_{i,k}} \mathcal{E}_{i,k} - e^{j\theta_{i,k}} \sqrt{\text{SNR}}hx_i \right\} f_H(h) dh \\ &\stackrel{(b)}{=} \int_{\mathcal{D}_k} \Pr \left\{ W \notin \mathcal{E}_{\frac{M}{2}} - \sqrt{\text{SNR}}e^{j\theta''_k} h x_{\frac{M}{2}} \right\} f_H(h) dh, \end{aligned}$$

where (a) follows from the independence of W , H and X , and (b) follows from above observations and the circular symmetry property of W . Let us now define $z = e^{j\theta''_k} h$ in (2.43). Since multiplication with a unit magnitude complex number is a unitary

transformation (i.e., rotation) over the complex plane, we have

$$\begin{aligned}
 p_{i,k} &= \int_{\mathcal{D}_k} \Pr \left\{ W \notin \mathcal{E}_{\frac{M}{2}} - \sqrt{\text{SNR}} e^{j\theta_k''} h x_{\frac{M}{2}} \right\} f_H(h) dh \\
 &= \int_{e^{j\theta_k''} \mathcal{D}_k} \Pr \left\{ W \notin \mathcal{E}_{\frac{M}{2}} - \sqrt{\text{SNR}} z x_{\frac{M}{2}} \right\} f_H(e^{-j\theta_k''} z) dz \\
 &\stackrel{(a)}{=} \int_{\mathcal{D}_{2^{n-1}}} \Pr \left\{ W \notin \mathcal{E}_{\frac{M}{2}} - \sqrt{\text{SNR}} z x_{\frac{M}{2}} \right\} f_H(z) dz \\
 &\stackrel{(b)}{=} \int_{\mathcal{D}_{2^{n-1}}} \Pr \left\{ \sqrt{\text{SNR}} z x_{\frac{M}{2}} + W \notin \mathcal{E}_{\frac{M}{2}, 2^{n-1}} \right\} f_H(z) dz \\
 &\stackrel{(c)}{=} p_{\frac{M}{2}, 2^{n-1}}, \tag{2.43}
 \end{aligned}$$

where (a), (b) and (c) follow from the circular symmetry of H [89, 90] and the corresponding definitions for $\mathcal{D}_k$, $\mathcal{E}_{i,k}$ and $p_{i,k}$ for $i \in [0 : M-1]$ and $k \in [0 : 2^n - 1]$. This shows $p(\text{SNR}) = 2^n p_{\frac{M}{2}, 2^{n-1}}$. For a circularly-symmetric pdf $f_H(h)$, it is well-known that $r f_H(r \cos \lambda, r \sin \lambda) = \frac{1}{2\pi} f_R(r)$ [91]. Switching to polar coordinates, and using the identities $r f_H(r \cos \lambda, r \sin \lambda) = \frac{1}{2\pi} f_R(r)$, $x_{\frac{M}{2}} = e^{j\frac{\pi}{M}}$ and $\mathcal{E}_{\frac{M}{2}, 2^{n-1}} = \mathcal{E}$, we have

$$p(\text{SNR}) = \frac{2^{n-1}}{\pi} \int_{-\frac{\pi}{2^n}}^{\frac{\pi}{2^n}} \int_0^\infty \Pr \left\{ \sqrt{\text{SNR}} r e^{j(\frac{\pi}{M} + \lambda)} + W \notin \mathcal{E} \right\} f_R(r) dr d\lambda.$$

By using the change of variables $\theta = \frac{\pi}{M} + \lambda$, we conclude the proof.

2.9.4 Proof of Lemma 2.2

The proof of part (i) follows immediately from Definition 2.1:

$$\begin{aligned}
 \lim_{\text{SNR} \rightarrow \infty} \frac{\log(\alpha f(\text{SNR}))}{\log \text{SNR}} &= \lim_{\text{SNR} \rightarrow \infty} \frac{\log \alpha}{\log \text{SNR}} + \lim_{\text{SNR} \rightarrow \infty} \frac{\log f(\text{SNR})}{\log \text{SNR}}, \\
 &= d_i.
 \end{aligned}$$

For the proof of part (ii), given any $\epsilon > 0$, let $c > 0$ be such that

$$\text{SNR}^{d_i - \epsilon} \leq f_i(\text{SNR}) \leq \text{SNR}^{d_i + \epsilon} \tag{2.44}$$

for all $\text{SNR} \geq c$ and $i \in [1 : N]$. Let $i' = \arg \max_{i \in [1:N]} d_i$. Then, as $\text{SNR} \rightarrow \infty$, we can write

$$\begin{aligned} \log \left(\sum_{i=1}^N f_i(\text{SNR}) \right) &\leq \log \left(\sum_{i=1}^N \text{SNR}^{d_i + \epsilon} \right) \\ &= \log \text{SNR}^{d_{\max} + \epsilon} + \log \left(1 + \sum_{\substack{i=1 \\ i \neq i'}}^N \text{SNR}^{d_i - d_{\max}} \right) \\ &= (d_{\max} + \epsilon) \log \text{SNR} + o(1), \end{aligned}$$

which implies $\sum_{i=1}^N f_i(\text{SNR}) \stackrel{e}{\leq} d_{\max} + \epsilon$. Since $\epsilon > 0$ is arbitrary, we conclude that $\sum_{i=1}^N f_i(\text{SNR}) \stackrel{e}{\leq} d_{\max}$. The other direction $\sum_{i=1}^N f_i(\text{SNR}) \stackrel{e}{\geq} d_{\max}$ follows from the same arguments, which completes the proof.

2.9.5 Proof of Lemma 2.3

We start with the case $M = 4$ and $n = 2$, and obtain upper and lower bounds on $p_1(\text{SNR})$ that will lead to the same exponential equality. For the upper bound, we write $p_1(\text{SNR})$ as

$$\begin{aligned} p_1(\text{SNR}) &= \frac{2m^m}{\pi^2} \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} \left(\frac{\text{SNR}}{\sin^2 \beta} \cos^2 \theta + m \right)^{-m} d\theta d\beta \\ &\leq \frac{m^m}{\pi} \int_0^{\frac{\pi}{2}} (\text{SNR} \cos^2 \theta + m)^{-m} d\theta. \end{aligned}$$

Let $\theta^*(\text{SNR})$ be such that $\cos(\theta^*(\text{SNR})) = \text{SNR}^{-\frac{1}{2}}$. Then, $\theta^*(\text{SNR}) = \arccos(\text{SNR}^{-\frac{1}{2}}) = \frac{\pi}{2} - \text{SNR}^{-\frac{1}{2}} - o(\text{SNR}^{-\frac{1}{2}})$ as $\text{SNR} \rightarrow \infty$. Using the fact that the line $1 - \frac{2}{\pi}\theta$ is a lower bound for $\cos \theta$ for $\theta \in [0, \frac{\pi}{2}]$, we have

$$\begin{aligned} p_1(\text{SNR}) &\leq \frac{m^m}{\pi} \int_0^{\theta^*(\text{SNR})} (\text{SNR} \cos^2 \theta + m)^{-m} d\theta + \frac{m^m}{\pi} \int_{\theta^*(\text{SNR})}^{\frac{\pi}{2}} (\text{SNR} \cos^2 \theta + m)^{-m} d\theta \\ &\leq \frac{m^m}{\pi} \text{SNR}^{-m} \int_0^{\theta^*(\text{SNR})} \left(1 - \frac{2}{\pi}\theta \right)^{-2m} d\theta + \frac{1}{\pi} \text{SNR}^{-\frac{1}{2}} (1 + o(1)) \\ &= -\frac{m^m}{2} \text{SNR}^{-m} \int_1^{1 - \frac{2}{\pi}\theta^*(\text{SNR})} u^{-2m} du + \frac{1}{\pi} \text{SNR}^{-\frac{1}{2}} (1 + o(1)) \end{aligned}$$

Hence, we have (2.45)

$$p_1(\text{SNR}) = \begin{cases} \text{SNR}^{-\frac{1}{2}} O(\log \text{SNR}) & \text{if } m = \frac{1}{2} \\ \text{SNR}^{-\frac{1}{2}} O(1) & \text{if } m > \frac{1}{2} \end{cases} \quad (2.45)$$

for large values of SNR. Equation (2.45) shows that $p_1(\text{SNR}) \stackrel{e}{\leq} \text{SNR}^{-\frac{1}{2}}$ for all $m \geq \frac{1}{2}$ when $M = 4$ and $n = 2$. For the other direction, we obtain a lower bound on $p_1(\text{SNR})$ as (2.46).

$$\begin{aligned} p_1(\text{SNR}) &= \frac{2m^m}{\pi^2} \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} (\sin \beta)^{2m} (\text{SNR} \cos^2 \theta + m \sin^2 \beta)^{-m} d\theta d\beta \\ &\geq \frac{2m^m}{\pi^2} \int_0^{\frac{\pi}{2}} (\sin \beta)^{2m} d\beta \int_0^{\frac{\pi}{2}} (\text{SNR} \cos^2 \theta + m)^{-m} d\theta \end{aligned} \quad (2.46)$$

We observe that $\int_0^{\frac{\pi}{2}} (\sin \beta)^{2m} d\beta = \frac{\sqrt{\pi} \Gamma(m + \frac{1}{2})}{2\Gamma(m+1)}$ for $m > -\frac{1}{2}$ and let $c = \frac{m^m}{\pi^{1.5}} \frac{\Gamma(m + \frac{1}{2})}{\Gamma(m+1)}$. We first consider the case $m = \frac{1}{2}$. Then, as $\text{SNR} \rightarrow \infty$, we have

$$\begin{aligned} p_1(\text{SNR}) &\geq c \text{SNR}^{-\frac{1}{2}} \int_0^{\frac{\pi}{2}} (\cos^2 \theta + 1)^{-\frac{1}{2}} d\theta \\ &= \text{SNR}^{-\frac{1}{2}} \Omega(1). \end{aligned} \quad (2.47)$$

For $m > \frac{1}{2}$, we define $\theta^*(\text{SNR})$ as above and lower bound $p_1(\text{SNR})$ for large values of SNR as

$$\begin{aligned} p_1(\text{SNR}) &\geq c \text{SNR}^{-m} \int_{\theta^*(\text{SNR})}^{\frac{\pi}{2}} \left(\cos^2 \theta + \frac{m}{\text{SNR}} \right)^{-m} d\theta \\ &\geq c \text{SNR}^{-m} \int_{\theta^*(\text{SNR})}^{\frac{\pi}{2}} \left(\frac{1}{\text{SNR}} + \frac{m}{\text{SNR}} \right)^{-m} d\theta \\ &= \text{SNR}^{-\frac{1}{2}} c (1+m)^{-m} (1+o(1)) \\ &= \text{SNR}^{-\frac{1}{2}} \Omega(1). \end{aligned} \quad (2.48)$$

Using (2.47) and (2.48), we conclude that $p_1(\text{SNR}) \stackrel{e}{\geq} \text{SNR}^{-\frac{1}{2}}$ for all $m \geq \frac{1}{2}$ when $M = 4$ and $n = 2$. Since $p_1(\text{SNR})$ also satisfies $p_1(\text{SNR}) \stackrel{e}{\leq} \text{SNR}^{-\frac{1}{2}}$ in this case, we have $p_1(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-\frac{1}{2}}$ for $m \geq \frac{1}{2}$, $M = 4$ and $n = 2$.

Next, we consider $M = 4$ and $n > 2$. In this case, we write $p_1(\text{SNR})$ as

$$p_1(\text{SNR}) = \frac{2^{n-1}m^m}{\pi^2} \text{SNR}^{-m} \int_0^{\frac{\pi}{2}} \int_{\frac{\pi}{4}-\frac{\pi}{2^n}}^{\frac{\pi}{4}+\frac{\pi}{2^n}} \left(\frac{\cos^2 \theta}{\sin^2 \beta} + \frac{m}{\text{SNR}} \right)^{-m} d\theta d\beta. \quad (2.49)$$

Let $g_{\text{SNR}}(\theta, \beta) = \left(\frac{\cos^2 \theta}{\sin^2 \beta} + \frac{m}{\text{SNR}} \right)^{-m}$ be a collection of functions indexed by SNR. These functions increase to $g_{\infty}(\theta, \beta) = \left(\frac{\cos^2 \theta}{\sin^2 \beta} \right)^{-m}$ as $\text{SNR} \rightarrow \infty$. Further, $\int_0^{\frac{\pi}{2}} \int_{\frac{\pi}{4}-\frac{\pi}{2^n}}^{\frac{\pi}{4}+\frac{\pi}{2^n}} g_{\infty}(\theta, \beta) d\theta d\beta < \infty$. Hence, as $\text{SNR} \rightarrow \infty$, we conclude that

$$p_1(\text{SNR}) = \text{SNR}^{-m} \Theta(1) \quad (2.50)$$

by using the monotone convergence theorem [86], which implies $p_1(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-m}$. The proof for $M > 4$ and $n \geq \log_2 M$ is similar, and we omit it to avoid repetition.

2.9.6 Proof of Lemma 2.4

For $M = 4$ and $n = 2$, $p_2(\text{SNR}) = p_1(\text{SNR})$, and hence the proof of $p_2(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-\frac{1}{2}}$ in this case directly follows from Lemma 2.3. For $M > 4$ and $n > \log_2 M$, a similar argument using the monotone convergence theorem as in the proof of Lemma 2.3 readily shows that $p_2(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-m}$ for all $m \geq \frac{1}{2}$. Therefore, we will only focus on the case $M > 4$ and $n = \log_2 M$ to complete the proof of Lemma 2.4.

For $M > 4$ and $n = \log_2 M$, we will obtain upper and lower bounds on $p_2(\text{SNR})$ leading to the same exponential equality. For the upper bound, we have

$$\begin{aligned} p_2(\text{SNR}) &= \frac{2^{n-1}m^m}{\pi^2} \int_0^{\frac{\pi}{2}} \int_0^{\frac{2\pi}{M}} \left(\frac{\text{SNR}}{\sin^2 \beta} \sin^2 \theta + m \right)^{-m} d\theta d\beta \\ &\leq \frac{2^{n-1}m^m}{\pi^2} \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} \left(\frac{\text{SNR}}{\sin^2 \beta} \sin^2 \theta + m \right)^{-m} d\theta d\beta \\ &\stackrel{e}{=} \text{SNR}^{-\frac{1}{2}}, \end{aligned} \quad (2.51)$$

where the last equality follows from the fact that $p_2(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-\frac{1}{2}}$ for $M = 4$ and $n = 2$ and Lemma 2.2.

For the lower bound, we have (2.52) where $c = \frac{2^{n-2}m^m}{\pi^{1.5}} \frac{\Gamma(m+\frac{1}{2})}{\Gamma(m+1)}$.

$$\begin{aligned} p_2(\text{SNR}) &= \frac{2^{n-1}m^m}{\pi^2} \int_0^{\frac{\pi}{2}} \int_0^{\frac{2\pi}{M}} (\sin \beta)^{2m} (\text{SNR} \sin^2 \theta + m \sin^2 \beta)^{-m} d\theta d\beta \\ &\geq c \text{SNR}^{-m} \int_0^{\frac{2\pi}{M}} \left(\sin^2 \theta + \frac{m}{\text{SNR}} \right)^{-m} d\theta, \end{aligned} \quad (2.52)$$

Let $\theta^*(\text{SNR})$ be such that $\sin(\theta^*(\text{SNR})) = \text{SNR}^{-\frac{1}{2}}$. Then, $\theta^*(\text{SNR}) = \arcsin(\text{SNR}^{-\frac{1}{2}}) = \text{SNR}^{-\frac{1}{2}} + o(\text{SNR}^{-\frac{1}{2}})$ as $\text{SNR} \rightarrow \infty$. Hence, as $\text{SNR} \rightarrow \infty$, we have

$$\begin{aligned} p_2(\text{SNR}) &\geq c \text{SNR}^{-m} \int_0^{\theta^*(\text{SNR})} \left(\sin^2 \theta + \frac{m}{\text{SNR}} \right)^{-m} d\theta \\ &\geq c \text{SNR}^{-m} \int_0^{\theta^*(\text{SNR})} \left(\frac{1}{\text{SNR}} + \frac{m}{\text{SNR}} \right)^{-m} d\theta \\ &= \text{SNR}^{-\frac{1}{2}} \Omega(1), \end{aligned} \quad (2.53)$$

which implies $p_2(\text{SNR}) \stackrel{e}{\geq} \text{SNR}^{-\frac{1}{2}}$. Since $p_2(\text{SNR})$ also satisfies $p_2(\text{SNR}) \stackrel{e}{\leq} \text{SNR}^{-\frac{1}{2}}$ in this case, we have $p_2(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-\frac{1}{2}}$ for $m \geq \frac{1}{2}$, $M > 4$ and $n = \log_2 M$.

2.9.7 Proof of Lemma 2.5

We will prove this lemma for general circularly-symmetric fading processes. To this end, let $H = Re^{j\Theta}$ be the circularly-symmetric fading coefficient with the joint phase and magnitude pdf $f_{R,\Theta}(r, \theta) = \frac{1}{2\pi} f_R(r)$ for $\theta \in [-\pi, \pi)$ and $r \geq 0$. In the proof of Theorem 2.3, we obtained the expression $p(\text{SNR}, h)$ with $h = re^{j\theta}$ given by (2.54).

$$\begin{aligned} p(\text{SNR}, h) &= \mathcal{Q}(\sqrt{2\text{SNR}}r \cos \theta) + \mathcal{Q}(\sqrt{2\text{SNR}}r \sin \theta) \\ &\quad - \mathcal{Q}(\sqrt{2\text{SNR}}r \cos \theta) \mathcal{Q}(\sqrt{2\text{SNR}}r \sin \theta) \\ &\quad + \frac{1}{\sqrt{\pi}} \int_{-\sqrt{\text{SNR}}r \cos \theta}^{\infty} \mathcal{Q}\left(\sqrt{2\text{SNR}} \sec\left(\frac{2\pi}{M}\right) \sin\left(\frac{2\pi}{M} - \theta\right) + \sqrt{2} \tan\left(\frac{2\pi}{M}\right) w\right) e^{-w^2} dw. \end{aligned} \quad (2.54)$$

Below, we will always use $re^{j\theta}$ as the polar coordinate representation of h , i.e., $r = |h|$ and $\theta = \text{Arg}(h)$. Integrating $2^n p(\text{SNR}, h)$ with respect to $f_{R,\Theta}(r, \theta)$ for $\theta \in$

$\left[\frac{\pi}{M} - \frac{\pi}{2^n}, \frac{\pi}{M} + \frac{\pi}{2^n}\right)$ and $r \geq 0$, and using the Nakagami- m pdf for $f_R(r)$ together with Craig's formula, we obtained the resulting $p(\text{SNR})$ expression in Theorem 2.3. Here, we will not assume any specific functional form for $f_R(r)$.

We start with obtaining the lower bound $L(\text{SNR})$ on $p(\text{SNR})$. Let

$$\begin{aligned} p_1(\text{SNR}, h) &= Q\left(\sqrt{2\text{SNR}}r \cos \theta\right) \\ p_2(\text{SNR}, h) &= Q\left(\sqrt{2\text{SNR}}r \sin \theta\right) \\ p_3(\text{SNR}, h) &= Q\left(\sqrt{2\text{SNR}}r \cos \theta\right) Q\left(\sqrt{2\text{SNR}}r \sin \theta\right) \end{aligned}$$

and $p_4(\text{SNR}, h)$ be the last integral term in (2.54). For $i \in [1 : 4]$, $p_i(\text{SNR})$ is defined to be the integral of $2^n p_i(\text{SNR}, h)$ with respect to $f_{R,\Theta}(r, \theta)$ for $\theta \in \left[\frac{\pi}{M} - \frac{\pi}{2^n}, \frac{\pi}{M} + \frac{\pi}{2^n}\right)$ and $r \geq 0$. For the given integration range, $p_3(\text{SNR}, h) \leq \frac{1}{2} p_2(\text{SNR}, h)$ since the argument of the Q -function is always positive. Hence, we have

$$\begin{aligned} p(\text{SNR}, h) &\geq p_1(\text{SNR}, h) + p_2(\text{SNR}, h) - p_3(\text{SNR}, h) \\ &\geq p_1(\text{SNR}, h) + \frac{1}{2} p_2(\text{SNR}, h). \end{aligned} \quad (2.55)$$

After scaling with 2^n and integrating (2.55) with respect to $f_{R,\Theta}(r, \theta)$ over the above integration range, we have

$$\begin{aligned} p(\text{SNR}) &\geq p_1(\text{SNR}) + \frac{1}{2} p_2(\text{SNR}) \\ &= L(\text{SNR}). \end{aligned} \quad (2.56)$$

Next, we establish that $U(\text{SNR}) = p_1(\text{SNR}) + 2p_2(\text{SNR})$ is an upper bound on $p(\text{SNR})$. To this end, we will show that $p_4(\text{SNR}) \leq p_2(\text{SNR})$ for all $M \geq 4$. For $M = 4$, this is trivial since $p_4(\text{SNR}, h) = 0 \leq p_2(\text{SNR}, h)$. For $M > 4$, we define

$$p_5(\text{SNR}, h) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} Q(\varphi) e^{-w^2} dw,$$

where $\varphi = \sqrt{2\text{SNR}}r \sec\left(\frac{2\pi}{M}\right) \sin\left(\frac{2\pi}{M} - \theta\right) + \sqrt{2}w \tan\left(\frac{2\pi}{M}\right)$. We also define $p_5(\text{SNR})$ to be the integral of $2^n p_5(\text{SNR}, h)$ with respect to $f_{R,\Theta}(r, \theta)$ for $\theta \in \left[\frac{\pi}{M} - \frac{\pi}{2^n}, \frac{\pi}{M} + \frac{\pi}{2^n}\right)$ and

$r \geq 0$.

We observe that $p_4(\text{SNR}) \leq p_5(\text{SNR})$ since the integrands are always positive and the integral with respect to w is over the whole real line for $p_5(\text{SNR}, h)$. Thus, it will be enough to show $p_2(\text{SNR}) = p_5(\text{SNR})$ to conclude the proof. For $\text{SNR} = 0$, this can be verified by using the identity $Q(x) = 1 - Q(-x)$. To prove the equality for all $\text{SNR} \geq 0$, we define the function $f(\text{SNR}) = p_2(\text{SNR}) - p_5(\text{SNR})$. It is enough to show that the derivative of $f(\text{SNR})$, which we represent by $f'(\text{SNR})$, is equal to zero everywhere in order to show $p_2(\text{SNR}) = p_5(\text{SNR})$. This is because if $f'(\text{SNR})$ is equal to zero for all $\text{SNR} \geq 0$, then $f(\text{SNR})$ must be a constant function. Since $f(0) = 0$, we have $f(\text{SNR}) = p_2(\text{SNR}) - p_5(\text{SNR}) = 0$ for all $\text{SNR} \geq 0$. We devote the rest of the proof to showing that $f'(\text{SNR}) = 0$ for all $\text{SNR} \geq 0$.

Using the definition of the Q -function, the derivative of $p_2(\text{SNR})$ with respect to SNR , which we represent by $p'_2(\text{SNR})$, is given by

$$\begin{aligned} p'_2(\text{SNR}) &= \frac{2^{n-1}}{\pi} \int_{\frac{\pi}{M} - \frac{\pi}{2^n}}^{\frac{\pi}{M} + \frac{\pi}{2^n}} \int_0^\infty \frac{dp_2(\text{SNR}, re^{j\theta})}{d\text{SNR}} f_R(r) dr d\theta \\ &= \frac{2^{n-1}}{\pi} \int_{\frac{\pi}{M} - \frac{\pi}{2^n}}^{\frac{\pi}{M} + \frac{\pi}{2^n}} \int_0^\infty \frac{-r \sin \theta}{2\sqrt{\pi \text{SNR}}} e^{-\text{SNR} r^2 \sin^2 \theta} f_R(r) dr d\theta. \end{aligned}$$

Similarly, $p'_5(\text{SNR})$ can be written as

$$\begin{aligned} p'_5(\text{SNR}) &= \frac{2^{n-1}}{\pi} \int_{\frac{\pi}{M} - \frac{\pi}{2^n}}^{\frac{\pi}{M} + \frac{\pi}{2^n}} \int_0^\infty \frac{dp_5(\text{SNR}, re^{j\theta})}{d\text{SNR}} f_R(r) dr d\theta \\ &= \frac{2^{n-1}}{\pi \sqrt{\pi}} \int_{\frac{\pi}{M} - \frac{\pi}{2^n}}^{\frac{\pi}{M} + \frac{\pi}{2^n}} \int_0^\infty \frac{-A(r, \theta)}{2\sqrt{\pi \text{SNR}}} I(r, \theta) f_R(r) dr d\theta, \end{aligned} \quad (2.57)$$

where $A(r, \theta) = r \sec\left(\frac{2\pi}{M}\right) \sin\left(\frac{2\pi}{M} - \theta\right)$, $I(r, \theta) = \int_{-\infty}^\infty e^{-\left(w^2 + (A(r, \theta)\sqrt{\text{SNR}} + Bw)^2\right)} dw$ and $B = \tan\left(\frac{2\pi}{M}\right)$. After completing the term in the exponent in $I(r, \theta)$ to square and using the affinity of the resulting expression to a Gaussian pdf, $I(r, \theta)$ can be shown to be equal to

$$I(r, \theta) = \sqrt{\pi} \cos\left(\frac{2\pi}{M}\right) e^{-\text{SNR} r^2 \sin^2\left(\frac{2\pi}{M} - \theta\right)}. \quad (2.58)$$

Using (2.58) in (2.57), we have

$$p'_5(\text{SNR}) = \frac{2^{n-1}}{\pi} \int_{\frac{\pi}{M} - \frac{\pi}{2^n}}^{\frac{\pi}{M} + \frac{\pi}{2^n}} \int_0^\infty \frac{-r \sin\left(\frac{2\pi}{M} - \theta\right)}{2\sqrt{\pi \text{SNR}}} e^{-\text{SNR} r^2 \sin^2\left(\frac{2\pi}{M} - \theta\right)} f_R(r) dr d\theta. \quad (2.59)$$

Change of variables $u = \frac{2\pi}{M} - \theta$ in (2.59) shows that $p'_2(\text{SNR}) = p'_5(\text{SNR})$, and hence $f'(\text{SNR}) = 0$ as desired.

Chapter 3

Low-Resolution Quantization with Non-coherent Detection

This chapter studies optimum detectors and error rate analysis for wireless systems with low-resolution quantizers in the presence of fading and noise. In the case of DQPSK, the ML detector is derived with and without the CSI at the receiver. Utilizing the structure of the derived detector, a general average SEP expression for DQPSK modulation with n -bit quantization is obtained when the wireless channel is subject to Rayleigh fading. It is shown that a transceiver architecture with n -bit quantization is asymptotically optimum in terms of communication reliability if $n \geq 4$ when there is no channel knowledge available at the receiver. That is, the decay exponent for the average SEP is the same and equal to 1 with infinite-bit and n -bit quantizers for $n \geq 4$. On the other hand, it is only equal to $\frac{1}{2}$ and 0 for $n = 2, 3$ and $n < 2$, respectively. On the other hand, it is shown that a transceiver architecture with n -bit quantization is asymptotically optimum in terms of communication reliability if $n \geq 3$ when the channel knowledge is available at the receiver. That is, the decay exponent for the average SEP is the same and equal to 1 with infinite-bit and n -bit quantizers for $n \geq 3$. On the other hand, it is only equal to $\frac{1}{2}$ and 0 for $n = 2$ and $n < 2$, respectively. Therefore, when the channel knowledge is not available at the receiver, the quantizer has to use one additional bit to achieve optimum communication robustness. An extensive simulation study is performed to illustrate the accuracy of the derived results.

3.1 Introduction

3.1.1 Background and Motivation

All the coherent detectors that we discussed in Section 1.2.1 require CSI to be available at the receiver. However, due to the non-linear nature of quantization, channel estimation has become one of the major challenges associated with low-resolution quantization

based communication systems. As presented in Section 1.2.2, researchers have focused on several approaches to estimate CSI as accurate as possible with low-resolution ADCs.

In this work, we take a different approach to handle this challenge by adopting non-coherent data detection. More specifically, we adopt the DQPSK modulation method to differentially encode the transmit data and at the receiver side, employ non-coherent detection that does not require instantaneous CSI. We consider a communication system in which the receiver is equipped with low-resolution ADCs that quantize the phase of the received signal.

To draw fundamental insights, we consider an abstract SISO system that provides the building block for more complicated MIMO systems. The error probability performance of such a system is analysed in detail in the previous chapter under coherent detection where we established a fundamental ternary behaviour on the average error probability with low-resolution ADCs. In this chapter, we focus on non-coherent detection and conduct a thorough mathematical analysis to quantify the difference in SEP performance between coherent and non-coherent detection. In other words, we quantify the penalty we have to pay in terms of the number of quantization bits, for not using the CSI at the receiver with low-resolution ADCs.

3.1.2 Main Contributions

In this chapter, we consider a point-to-point wireless communication system, in which the receiver is equipped with a low-resolution ADC that quantizes the phase of the received signal. Our main contributions are summarized as follows.

- Motivated by the capacity achieving property of circularly symmetric input distributions for low-resolution ADCs [77], we adopt DQPSK modulation and derive the optimum ML detection rule for signal detection with low-resolution ADCs. We first consider the case where CSI is not available at the detector and then CSI is available. These results are presented in Theorems 3.1 and 3.4 respectively.

- For the general n -bit quantization, we obtain exact analytical expressions for the average SEP attained by the derived ML rule when the wireless channel is subjected to Rayleigh fading. We use extensive numerical examples to compare the analytical expressions against the Monte-Carlo simulations. These results are presented in Theorems 3.2 and 3.5 respectively.
- We establish a fundamental ternary behaviour in the average SEP performance of a wireless communication system with low-resolution ADCs and DQPSK modulation under non-coherent detection. Based on the Rayleigh fading model, we show that the decay exponent of the average SEP is the same as that of an infinite-bit quantization, which is equal to 1 when n is larger than or equal to 4. We also show that it is equal to $\frac{1}{2}$ and 0 for $n = 2, 3$ and $n < 2$, respectively. This result is presented in Theorem 3.3.
- We establish a fundamental ternary behaviour in the average SEP performance of a wireless communication system with low-resolution ADCs and DQPSK modulation under coherent detection. Based on the Rayleigh fading model, we show that the decay exponent of the average SEP is the same as that of an infinite-bit quantization, which is equal to 1 when n is larger than or equal to 3. We also show that it is equal to $\frac{1}{2}$ and 0 for $n = 2$ and $n < 2$, respectively. This result is presented in Theorem 3.7.
- By analysing the error probability performance in the high-SNR regime, we characterize the DVO for low-resolution ADC based communication systems. From a system design point of view, our results show that by using 4-bit quantization, we can achieve optimum communication robustness in the high-SNR regime when the channel knowledge is not available at the receiver. We also analyse the asymptotic behaviour of the SEP performance by focusing on the high SNR regime when the channel knowledge is available at the receiver and conclude that if the instantaneous CSI is available at the receiver we need one less quantization bit to achieve the full DVO when compared to the number of quantization bits required for non-coherent detection.

- We perform numerical analysis to compare the performance of DQPSK modulation to the more common QPSK modulation when low-resolution quantization is used. The QPSK results are generated using [92] under coherent detection.

3.2 System Setup

We consider a classical point-to-point wireless communication system with DQPSK modulation. We adopt a block fading model where the channel is assumed to remain constant at least over two symbol periods.

3.2.1 Transmitter Architecture

The transmitter is composed of two blocks, which we call the *message source* and *modulator* as shown in Fig. 3.1. The time index is taken to be a non-negative integer and will be denoted by t in the remainder of this chapter. Message source generates equiprobable iid messages from an alphabet $\mathcal{M} = \{m_0, m_1, m_2, m_3\}$ according to $\Pr \{M_t = m_i\} = \frac{1}{4}$ for all $t \geq 1$ and $i = 0, 1, 2, 3$.

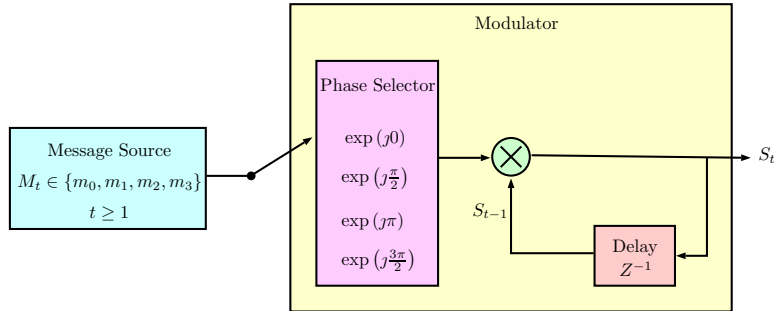


Figure 3.1: The transmitter architecture with the DQPSK modulation scheme. The modulator generates the symbol to be transmitted at time t according to (3.1).

The output of the modulator or the transmitted symbol at time t , which will be denoted by S_t , will be governed by the message M_t and symbol S_{t-1} for all $t \geq 1$. To initiate the communication process at time $t = 0$, the modulator generates a random signal S_0 uniformly at random from $\mathcal{C} = \left\{ e^{j\frac{\pi}{4}(2i+1)} \right\}_{i=0}^3$. For $t \geq 1$, we can express the

Message (M_t)	00	10	01	11	10	01	10	00
Transmitted Relative Phase	0	$-\frac{\pi}{2}$	0	π	$\frac{\pi}{2}$	π	$\frac{\pi}{2}$	$\frac{\pi}{2}$
Transmitted symbol (S_t)	00	10	00	11	01	11	01	01

Table 3.1: An example of the DQPSK modulation scheme.

differentially encoded transmitted symbol S_t at time t as

$$S_t = S_{t-1} e^{j\frac{\pi}{2}} \quad \text{if } M_t = m_\lambda. \quad (3.1)$$

We note that $\Pr \left\{ S_t = e^{j\frac{\pi}{4}(2i+1)} \right\} = \frac{1}{4}$ for $i = 0, 1, 2, 3$.

The following example further elaborates on the DQPSK modulation scheme. Let the information symbols 00, 01, 11 and 10 correspond to relative phases of 0, $\frac{\pi}{2}$, π and $-\frac{\pi}{2}$ respectively. Then Table 3.1 shows the encoded symbol sequence for a given information sequence with respect to the above-mentioned relative phases.

3.2.2 Channel Model

We consider a fading wireless channel in which fading process changes at a slower time scale than the input signalling rate. Therefore, the input-output relationship of the channel can be represented as

$$Y_t = \sqrt{\text{SNR}} H S_t + W_t, \quad (3.2)$$

where $t \geq 0$ is the time index, $H \in \mathbb{C}$ is the Rayleigh fading coefficient that stays constant over a large block of input signals, W_t is the circularly-symmetric zero-mean unit-variance additive white Gaussian noise (AWGN), i.e., $W_t \sim \mathcal{CN}(0, 1)$. For a given channel realization $H = h$, $\{Y_t\}_{t=1}^\infty$ is an iid sequence of random variables since the input sequence $\{S_t\}_{t=1}^\infty$ is iid.

3.2.3 Receiver Architecture

The receiver architecture is based on a low-resolution ADC. As illustrated in Fig. 3.2, the received signal Y_t is first sent through a low-resolution quantizer, and then the

resulting quantized signal information is used to determine the transmitted message M_t . Since information is encoded in the phase difference of S_{t-1} and S_t with the above choice of constellation points, we choose the same quantization process given in Section 2.2.4 to quantize each received signal Y_t .

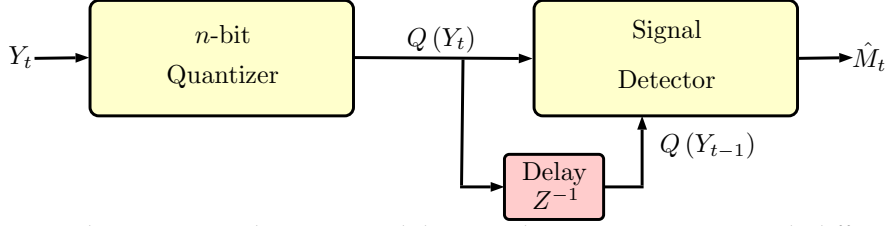


Figure 3.2: The receiver architecture with low-resolution quantization with differential modulation. The signal detector observes only the n -bit quantized versions of Y_t and Y_{t-1} to estimate the transmitted message signal $\hat{M}_t$.

We assume that no CSI is available at the receiver. Therefore, the signal detector estimates the transmitted message symbol non-coherently. Since the message M_t is encoded in the phase difference of the two consecutive transmitted signals S_{t-1} and S_t , the detector has to use at least two consecutive quantized received signals $Q(Y_{t-1})$ and $Q(Y_t)$ to estimate the transmitted message symbol $\hat{M}_t$.

We define the *block-2* detector as follows.

Definition 3.1. A *block-2* detector g is defined as a mapping

$$g : [0 : 2^n - 1]^2 \rightarrow \{m_0, m_1, m_2, m_3\}, \quad (3.3)$$

which takes two consecutive quantized observations to produce a single estimated message value. Here, the first quantized observation is used as a reference for decoding. We can illustrate the operation of the block-2 detector to estimate $\hat{M}_t$ as in Fig. 3.3.

In this work, we consider block-2 detectors for our analysis. This forms the baseline model for our analysis since this case is a direct extension of the detector available for unquantized differentially encoded systems. We also note that this detector is not optimum in general as it considers only two consecutive quantized observations $Q(Y_{t-1})$ and $Q(Y_t)$ to estimate the message $\hat{M}_t$ while neglecting all the other quantized observations before $t - 1$.

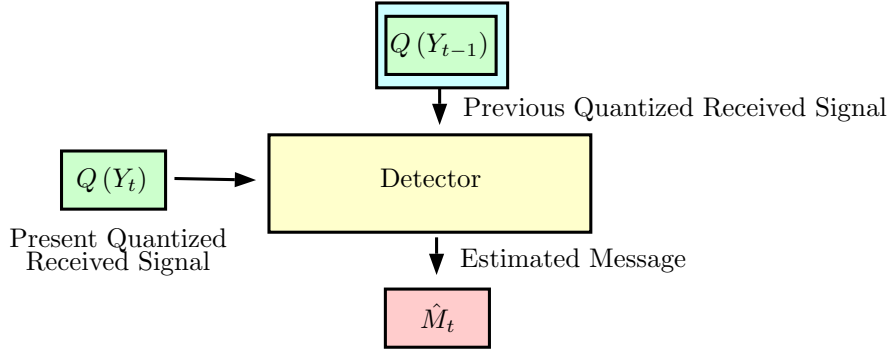


Figure 3.3: The block-2 detector architecture with low-resolution quantization with differential modulation. The signal detector uses two consecutive n -bit quantized received signals $Q(Y_{t-1})$ and $Q(Y_t)$ to estimate the message $\hat{M}_t$.

3.3 Optimum Signal Detection

The detector aims to minimize the SEP by using the knowledge of $Q(Y_{t-1})$ and $Q(Y_t)$ which can be represented as selecting signal point $\hat{M}_t(k_{t-1}, k_t)$ satisfying

$$\hat{M}_t(k_{t-1}, k_t) \in \arg \max_{m \in \mathcal{M}} \Pr \{M_t = m \mid Q(Y_{t-1}) = k_{t-1}, Q(Y_t) = k_t\}, \quad (3.4)$$

for $k_{t-1}, k_t \in [0 : 2^n - 1]$. The main performance figure of merit for the optimum detector is the average SEP given by

$$p(\text{SNR}) = \Pr \{M_t \neq \hat{M}_t(Q(Y_{t-1}), Q(Y_t))\}. \quad (3.5)$$

It is important to note that $p(\text{SNR})$ depends on SNR as well as the number of quantization bits.

The ML detector for the above-mentioned signal detection problem can be given as in Theorem 3.1.

Theorem 3.1. *The ML detection rule for the low-resolution quantization based receiver architecture with block-2 detector $g(k_{t-1}, k_t)$ can be given as*

$$\hat{M}_t(k_{t-1}, k_t) = \begin{cases} m_\lambda & \text{if } \lambda \neq 4 \\ m_0 & \text{if } \lambda = 4, \end{cases} \quad (3.6)$$

where k_t is the quantizer output at time $t \geq 1$. λ can be using either the detection rule 1 where

$$\lambda = \begin{cases} \left\lceil \frac{k_t - k_{t-1}}{2^{n-2}} \right\rceil & \text{if } k_t \geq k_{t-1} \\ \left\lceil \frac{k_t - k_{t-1}}{2^{n-2}} + 4 \right\rceil & \text{if } k_t < k_{t-1}, \end{cases} \quad (3.7)$$

with $\lceil a \rceil$ representing the ceil of a or using the detection rule 2 where

$$\lambda = \begin{cases} \left\lfloor \frac{k_t - k_{t-1}}{2^{n-2}} \right\rfloor & \text{if } k_t \geq k_{t-1} \\ \left\lfloor \frac{k_t - k_{t-1}}{2^{n-2}} + 4 \right\rfloor & \text{if } k_t < k_{t-1}, \end{cases} \quad (3.8)$$

with $\lfloor a \rfloor$ representing the floor of a .

Proof: See Appendix 3.9.1.

□

Before moving to analyze the SEP performance with respect to the above derived ML detection rule, we would like to note down an important observation with respect to the above-mentioned ML detector. According to the ML detector given, there are two possible detection rules that maximize the given likelihood function. For $n = 2$, both the detection rules have the same output and hence the given likelihood function will be maximized by only one element in $\mathcal{M}$. However, for $n \geq 3$, the given likelihood function will be maximized by two elements in $\mathcal{M}$. As an example, for $n = 3$ both m_0 and m_3 maximize the likelihood function when $Y_{t-1} \in \mathcal{R}_0$ and $Y_t \in \mathcal{R}_7$. Those are given in the detection rule 1 and 2 respectively. Similarly for $n = 3$, both m_1 and m_0 maximize the given likelihood function when $Y_{t-1} \in \mathcal{R}_0$ and $Y_t \in \mathcal{R}_1$. Those are given in the detection rule 1 and 2 respectively. Therefore, the detector has to choose one of these two detection rules when estimating the message. The detector can use any randomized strategy to select one out of these two detection rules. Therefore, we can have infinitely many ML detectors. In this case, we consider only one of them. Here onward, we assume that the detector has chosen the detection rule 1.

3.4 Symbol Error Probability

In this section, we will obtain a general $p(\text{SNR})$ expression for the optimum ML detector with a block length of 2. We will only consider the case $n \geq 2$ due to the existence of an error floor for $n < 2$ with DQPSK modulation. This is because with the DQPSK modulation the detector requires at least 4 quantization regions to differentiate the transmitted signals.

Considering these advantages, we will focus on the Rayleigh fading model for H to derive integral expressions for $p(\text{SNR})$ in the remainder of this chapter. This will be done so for all parameter combinations of $n \geq 2$ with DQPSK modulation. It will be seen that the derived integral expressions are easy to calculate numerically. Further, we will also show that using 4 bits is enough to achieve the maximum communication robustness achieved by using an infinite number of quantization bits. In comparison with coherent QPSK modulation with low-resolution quantization [92], this is one bit additional where we only require 3-bit quantization to achieve the maximum communication robustness achieved by using an infinite number of quantization bits.

Theorem 3.2. *Assume H is a unit-power Rayleigh fading coefficient. Then, for $n \geq 2$ with DQPSK modulation, $p(\text{SNR})$ is given according to*

$$p(\text{SNR}) = \frac{2^n}{\pi^2 (2\text{SNR} + 1)} \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{\infty} \int_{y_{t-1}^{\text{re}} \tan(\frac{\pi}{4} - \psi)}^{y_{t-1}^{\text{re}} \tan(\frac{\pi}{4} + \frac{\pi}{2^{n-1}} - \psi)} \mathcal{G}(\mathbf{y}) dy_{t-1}^{\text{im}} dy_{t-1}^{\text{re}} dy_t^{\text{re}} dy_t^{\text{im}} \\ + \frac{2^n}{\pi^2 (2\text{SNR} + 1)} \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} \int_{y_{t-1}^{\text{re}} \tan(\frac{\pi}{4} - \psi)}^{y_{t-1}^{\text{re}} \tan(\frac{\pi}{4} + \frac{\pi}{2^{n-1}} - \psi)} \mathcal{G}(\mathbf{y}) dy_{t-1}^{\text{im}} dy_{t-1}^{\text{re}} dy_t^{\text{re}} dy_t^{\text{im}} \quad (3.9)$$

where $\mathcal{G}(\mathbf{y}) = \exp \left\{ -\frac{\text{SNR}+1}{2\text{SNR}+1} \|\mathbf{y}\|_2^2 + \frac{2\text{SNR}}{2\text{SNR}+1} (y_{t-1}^{\text{re}} y_t^{\text{re}} + y_{t-1}^{\text{im}} y_t^{\text{im}}) \right\}$, $\|\mathbf{y}\|_2$ is the Euclidean norm of the vector $\mathbf{y} = [y_{t-1}^{\text{re}} \quad y_{t-1}^{\text{im}} \quad y_t^{\text{re}} \quad y_t^{\text{im}}]^T$, and

$$\psi = \begin{cases} \frac{\pi}{4} & \text{if } n = 2 \\ 0 & \text{if } n \geq 3. \end{cases}$$

Proof: See Appendix 3.9.2. □

For the special case of $n = 2$ and 3, Theorem 3.2 can be presented compactly as

given in the following corollary.

Corollary 3.1. *Assume H is a unit-power Rayleigh fading coefficient. Let $\mathcal{Q}(\cdot)$ be the complementary distribution function of the standard normal random variable. Then, for $n = 2$ and 3 with DQPSK modulation, $p(\text{SNR})$ is given according to*

$$p(\text{SNR}) = 1 - \frac{4}{\pi} \int_0^{2\pi} \int_0^\infty \mathcal{Q}^2\left(-\sqrt{2\text{SNR}}r \cos \theta\right) \mathcal{Q}^2\left(-\sqrt{2\text{SNR}}r \sin \theta\right) r e^{-r^2} dr d\theta. \quad (3.10)$$

Proof: See Appendix 3.9.3. □

3.5 The Decay Exponent for the Average Symbol Error Probability

In this section, we will analyze the communication robustness that can be achieved with low-resolution ADCs by focusing on the decay exponent for $p(\text{SNR})$, which is given by (2.5). Following the convention in the field, we will call DVO the *diversity order*, although there is only a single diversity branch in our system. We devote the rest of the current section to the proof of this important finding. We will use Definition 2.1 to simplify the notation. The following theorem establishes an important result for the DVO for our system.

Theorem 3.3. *The DVO of a low-resolution ADC based receiver architecture with DQPSK modulation and Rayleigh fading is given by*

$$\text{DVO} = \begin{cases} \frac{1}{2} & n = 2, 3 \\ 1 & n \geq 4. \end{cases} \quad (3.11)$$

Proof: See Appendix 3.9.4. □

3.6 Low-Resolution quantized systems with DQPSK modulation under coherent detection

In the above sections, we focused on a low-resolution quantization based system with DQPSK modulation when there is no channel information available at the receiver. That is, we used non-coherent detection. In this section, we analyze the same communication system when perfect knowledge of the channel is available at the receiver to obtain a lower bound on the error probability. We assume the same transmitter architecture and the channel model as provided in Section 3.2. However, the receiver architecture and the optimum detector will be different due to the availability of CSI.

3.6.1 Receiver Architecture with Coherent Detection

The receiver architecture is based on a low-resolution quantizer as illustrated in Fig. 3.4, where the quantization process is the same as the one used in Section 3.2. In this case, the message at time t will be estimated in two steps. The signal detector will use the quantized observations available at time t along with CSI to first estimate the transmitted signal $\hat{S}_t$. Then, by comparing $\hat{S}_t$ with the estimated transmitted signal in the previous time interval, $\hat{S}_{t-1}$, estimated message $\hat{M}_t$ will be decided.

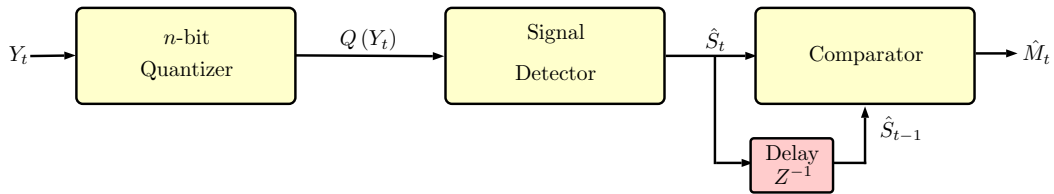


Figure 3.4: The receiver architecture with low-resolution quantization with differential modulation when perfect channel knowledge is available at the receiver. The message at time t , $\hat{M}_t$ will be decided after two steps. The signal detector first uses the n -bit quantized versions of Y_{t-1} and Y_t to estimate the transmitted signals $\hat{S}_{t-1}$ and $\hat{S}_t$ respectively. These estimated transmitted signals will be compared to estimate the message $\hat{M}_t$.

Similar to our previous work, we consider two consecutive quantized observations $Q(Y_{t-1})$ and $Q(Y_t)$ to decide on the estimated message $\hat{M}_t$. Therefore, the detector aims to minimize the SEP by using the knowledge of $Q(Y_{t-1}) = k_{t-1}$, $Q(Y_t) = k_t$ and

$H = h$ which can be represented as selecting the signal point $\hat{M}_t(k_{t-1}, k_t, h)$ satisfying

$$\hat{M}_t(k_{t-1}, k_t, h) \in \arg \max_{m \in \mathcal{M}} \Pr \{M_t = m \mid Q(Y_{t-1}) = k_{t-1}, Q(Y_t) = k_t, H = h\}, \quad (3.12)$$

for $k_{t-1}, k_t \in [0 : 2^n - 1]$ and $h \in \mathbb{C}$.

3.6.2 Optimum Signal Detection with Coherent Detection

The ML detector for the above-mentioned signal detection problem is given by the theorem below.

Theorem 3.4. *Assume $H \in \mathbb{C}$ has a continuous pdf. Then, $\hat{M}(k_{t-1}, k_t, h)$ is unique with probability one, and the ML detection rule for the low-resolution ADC based receiver architecture can be given as*

$$\hat{M}(k_{t-1}, k_t, h) = m_\lambda \quad \text{if } \hat{s}_t = \hat{s}_{t-1} \exp\left(\lambda \frac{j\pi}{2}\right) \quad (3.13)$$

where $\lambda \in \{0, 1, 2, 3\}$ and

$$\hat{s}_t(k_t, h) = \arg \min_{s_t \in \mathcal{C}} \text{dist}\left(\sqrt{\text{SNR}}hs_t, \mathcal{H}_{k_t}\right), \quad (3.14)$$

in which $k_t \in [0 : 2^n - 1]$, $\text{dist}(z, \mathcal{A}) \triangleq \inf_{s \in \mathcal{A}} |z - s|$ is the distance between a point $z \in \mathbb{C}$ and a set $\mathcal{A} \subseteq \mathbb{C}$ and $\mathcal{H}_{k_t} = \{z \in \mathbb{C} : \text{Arg}(z) = \frac{\pi}{2^n}(2k + 1)\}$ is the bisector of $\mathcal{R}_k$ where $Y_t \in \mathcal{R}_k$.

Proof: See Appendix 3.9.5. □

We first note that although the ML detection rule given in Theorem 3.4 is written as dependent on the full channel knowledge, we only need to know the phase of H in order to determine $\hat{s}_t(k_t, h)$. This can be seen by means of simple scaling arguments as the set $\mathcal{H}_{k_t}$ is invariant under any amplitude scaling. The half-hyperplane $\mathcal{H}_{k_t}$ in Theorem 3.4 bisects the k th quantization region $\mathcal{R}_k$ at time t into two symmetric regions. This is an important observation since it indicates that we are only required to estimate the channel phase to implement the ML detector. For phase estimation with circularly symmetric fading distributions, the estimation errors due to low-resolution

quantization can be modelled as bounded uniform random variables, which lead to a key robustness property for our analytical results under channel estimation errors as explained in [92].

Secondly, given $S_t = s$ and $H = h$, the conditional probability of the event $Q(Y_t) = k_t$ corresponds to the integral of unit-variance and mean $\sqrt{\text{SNR}}hs$ proper complex Gaussian distribution over $\mathcal{R}_k$ at time t . As we vary s in $\mathcal{C}$, we change the mean value $\sqrt{\text{SNR}}hs$. Hence, Theorem 3.4 indicates that the most probability mass is accumulated in the region $\mathcal{R}_k$ at time t when the unit-variance proper complex Gaussian distribution with mean $\sqrt{\text{SNR}}hs$, $s \in \mathcal{C}$, closest to $\mathcal{H}_{k_t}$ is integrated over $\mathcal{R}_k$, which coincides with the intuition. We will use the structure of the ML detection rule to derive integral expressions for $p(\text{SNR})$ for $n \geq 2$ in Section 3.6.3. Further, in order to characterize the communication robustness with low-resolution ADCs in the high SNR regime, we will also provide a detailed analysis of the asymptotic decay exponent of $p(\text{SNR})$ in Section 3.6.4.

3.6.3 Symbol Error Probability with Coherent Detection

In this section, we will obtain a general $p(\text{SNR})$ expression for the optimum ML detector that we presented above in the presence of Rayleigh fading when $n \geq 2$. It will be seen that the derived integral expressions are easy to calculate numerically. Further, we will also show that using 3-bit quantization is enough to achieve the maximum communication robustness achieved by using an infinite number of quantization bits when perfect channel information is available at the receiver. Recalling that we wanted to have at least 4 bits at the quantizer to achieve the full communication robustness with DQPSK modulation when there is no channel knowledge at the receiver, we can conclude that we need 1 additional bit at the quantizer when the channel knowledge is not available at the receiver. In other words, we need to have one more bit at the quantizer to compensate for the non-availability of the channel knowledge at the receiver.

Theorem 3.5. *Assume H is a unit-power Rayleigh fading coefficient. Let $\mathcal{Q}(\cdot)$ be the complementary distribution function of the standard normal random variable. Then, for $n \geq 2$ with*

DQPSK modulation under coherent detection, $p(\text{SNR})$ is given according to

$$p(\text{SNR}) = 1 - \frac{2^n}{\pi} \int_{\frac{\pi}{4} - \frac{\pi}{2^n}}^{\frac{\pi}{4} + \frac{\pi}{2^n}} \int_0^\infty \left\{ \sum_{i=1}^4 Q^2 \left(\rho_i^{\text{re}} \sqrt{2\text{SNR}} r \cos \theta \right) Q^2 \left(\rho_i^{\text{im}} \sqrt{2\text{SNR}} r \sin \theta \right) \right\} r \exp(-r^2) dr d\theta, \quad (3.15)$$

where $\rho_i^{\text{re}} = (-1)^i$ for $i = 1, 2, 3, 4$, $\rho_i^{\text{im}} = -1$ for $i = 1, 2$ and $\rho_i^{\text{im}} = 1$ for $i = 3, 4$.

Proof: See Appendix 3.9.6. □

The following corollary establishes an important comparison for SEP when $n = 2$ case with coherent and non-coherent detection.

Corollary 3.2. Assume H is a unit-power Rayleigh fading coefficient. Let $\mathcal{Q}(\cdot)$ be the complementary distribution function of the standard normal random variable. Then, for $n = 2$ with DQPSK modulation under both coherent and non-coherent detection, $p(\text{SNR})$ is the same and given according to

$$p(\text{SNR}) = 1 - \frac{4}{\pi} \int_0^{2\pi} \int_0^\infty Q^2 \left(-\sqrt{2\text{SNR}} r \cos \theta \right) Q^2 \left(-\sqrt{2\text{SNR}} r \sin \theta \right) r e^{-r^2} dr d\theta. \quad (3.16)$$

Proof: See Appendix 3.9.7. □

According to Corollary 3.2, the SEP performance of the detector will not change depending on whether the channel knowledge is available or not for $n = 2$. This means that the SEP for any block- L detector, $L \geq 2$, with non-coherent detection where $L \geq 2$ will be given by (3.16) and the block-2 detector that we considered is the optimum one for $n = 2$. This result is presented in the following Theorem.

Theorem 3.6. The ML block-2 detector given in Theorem 3.1 is the optimum detector among the class of block- L detectors when $n = 2$.

Proof: We proved that the SEP with non-coherent detection using the optimum block-2 detector is the same as that achieved by the optimum detector with coherent detection. This result is presented in Corollary 3.2. Since the optimum coherent detection forms a lower bound for the SEP performance of any block- L detector with

non-coherent detection, *whilst* the optimum non-coherent block-2 detection forms an upper bound for the same and these upper and lower bounds coincide with each other as established in Corollary 3.2, it can be concluded that the ML block-2 detector studied in the previous section is the optimum detector among the class of block- L detectors when $n = 2$. $\square$

3.6.4 The Decay Exponent for the Average Symbol Error Probability under Coherent Detection

The following theorem establishes the DVO for DQPSK modulation with coherent detection under Rayleigh fading.

Theorem 3.7. *The DVO of a low-resolution ADC based receiver architecture with DQPSK modulation and Rayleigh fading with coherent detection is given by*

$$\text{DVO} = \begin{cases} \frac{1}{2} & n = 2 \\ 1 & n \geq 3. \end{cases} \quad (3.17)$$

Proof: See Appendix 3.9.8. $\square$

By using the SEP and DVO results in the previous and current sections, we can draw several important insights regarding the block-2 detectors that we analyzed in the previous section. From Theorem 3.3, we know that the DVO is the same as that of an infinite-bit quantization, which is equal to 1 when $n \geq 4$ under non-coherent detection. Since this is the optimum DVO, we can conclude that it is enough to use a block-2 detector to achieve the optimum DVO when the quantizer is equipped with at least 4 bits and channel knowledge is not available at the receiver. In this case, the SEP achieved by the optimum block-2 detector will be an upper bound to the SEP achieved by the optimum block- L detector for $L \geq 3$. In this work, we did not focus on the optimum block- L detectors for $L \geq 3$ due to the problem complexity.

On the other hand, we have an important and somehow surprising result when $n = 2$. For $n = 2$, we proved that the SEP with non-coherent detection using the optimum block-2 detector is the same as that achieved by the optimum detector with coherent detection. This result is presented in Corollary 3.2. Since the optimum coherent

detection forms a lower bound for the SEP performance of any block- L detector with non-coherent detection, *whilst* the optimum non-coherent block-2 detection forms an upper bound for the same and these upper and lower bounds coincide with each other as established in Corollary 3.2, it can be concluded that the ML block-2 detector studied in the previous section is the optimum detector among the class of block- L detectors when $n = 2$. Although this result cannot be generalized to any n greater than 2, we still know that the ML block-2 detector is still the optimum detector in the DVO-sense for $n \geq 4$.

3.7 Numerical Results

In this section, we present analytical and simulated error probability performance results for DQPSK modulation with n -bit quantization and evaluate the performance of non-coherent detection with low-resolution quantization. Channel fading is unit-power and circularly-symmetric with Rayleigh distributed magnitude, and additive noise is complex Gaussian with zero mean and unit variance.

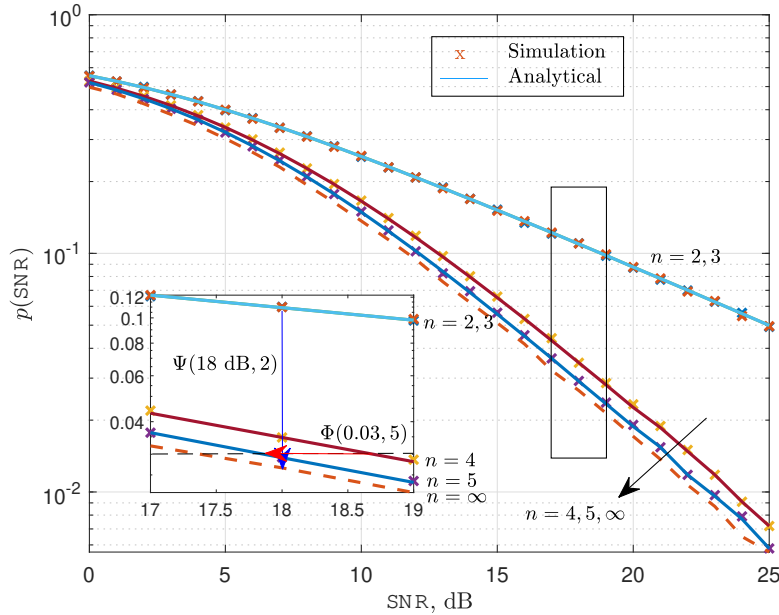


Figure 3.5: Average SEP curves as a function of SNR for DQPSK modulation with non-coherent detection. $n = 2, 3, 4, 5, \infty$.

Fig. 3.5 plots the average SEP as a function of SNR for DQPSK modulation with $n = 2, 3, 4, 5$ and ∞ -bit quantization under Rayleigh fading with non-coherent detection. The simulated results are generated using Monte Carlo simulation, while the analytical results are generated using our expression in (3.9). As the plot illustrates, the analytical results accurately follow the simulated results for all cases. As mentioned in Theorem 3.2, we observe similar SEP curves for both $n = 2$ and 3-bit quantization. We also observe a noteworthy improvement in the average SEP when n changes from 3 to 4-bit quantization for DQPSK modulation. This jump in the average SEP performance is expected in the light of Theorem 3.3, which states that using two extra bits, on top of $\log_2 M$ bits, improves the DVO from $\frac{1}{2}$ to 1 which is the full DVO under Rayleigh fading. We also observe that the average SEP reduces as we increase n , but the amount by which it reduces also gets smaller as we increase n . This can be clearly observed from the zoomed-in section in Fig. 3.5 which indicates that $\text{DVO} = 1$ for all $n \geq 4$ and $\text{DVO} = \frac{1}{2}$ for $n = 2, 3$, as expected.

Fig. 3.5 also illustrates the quantization penalty using (2.35) and (2.38) for DQPSK modulation under Rayleigh fading and non-coherent detection. When the SNR is fixed at 18 dB, we observe a quantization penalty of $\Psi(18 \text{ dB}, 2) \approx 6.14 \text{ dB}$ as we change from $n = 2, 3$ to ∞ , i.e., we get a 5-fold increase in the average SEP as we change from $n = 2, 3$ to ∞ . When the average SEP is fixed at 0.03, we observe a quantization penalty of $\Phi(0.03, 5) \approx 0.8 \text{ dB}$ as we change from $n = 4$ to 5, i.e., to achieve an average SEP of 0.03 with 5-bit quantization, we need 0.8 dB more transmit power than what is required with 4-bit quantization.

Fig. 3.6 plots the average SEP as a function of SNR for DQPSK modulation with $n = 2, 3, 4, 5$ and ∞ -bit quantization under Rayleigh fading with coherent detection. The simulated results are generated using Monte Carlo simulation, while the analytical results are generated using our expression in (3.15). As the plot illustrates, the analytical results accurately follow the simulated results for all cases. Under coherent detection, we observe a noteworthy improvement in the average SEP when n changes from 2 to 3-bit quantization. This jump in the average SEP performance is expected in the light of Theorem 3.7, which states that using one extra bit, on top of $\log_2 M$ bits, improves

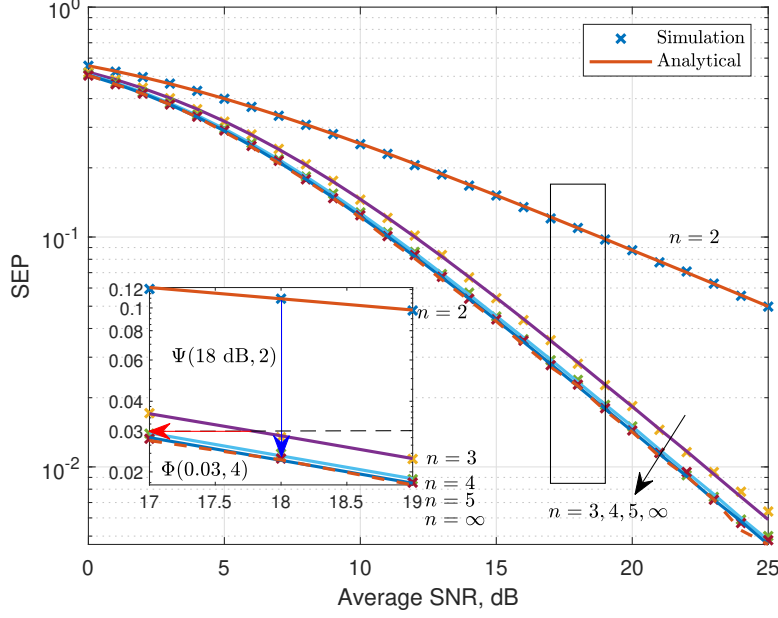


Figure 3.6: Average SEP curves as a function of SNR for DQPSK modulation with coherent detection. $n = 2, 3, 4, 5, \infty$.

the DVO from $\frac{1}{2}$ to 1 which is the full DVO under Rayleigh fading. Similar to Fig. 3.5, we observe that the average SEP reduces as we increase n , but the amount by which it reduces also gets smaller as we increase n . This can be clearly observed from the zoomed-in section in Fig. 3.6 which indicates that $DVO = 1$ for all $n \geq 3$ and $DVO = \frac{1}{2}$ for $n = 2$.

Fig. 3.6 also illustrates the quantization penalty using (2.35) and (2.38) for DQPSK modulation under Rayleigh fading and coherent detection. When the SNR is fixed at 18 dB, we observe a quantization penalty of $\Psi(18 \text{ dB}, 2) \approx 6.86 \text{ dB}$ as we change from $n = 2$ to ∞ , i.e., we get a 5-fold increase in the average SEP as we change from $n = 2$ to ∞ . When the average SEP is fixed at 0.03, we observe a quantization penalty of $\Phi(0.03, 4) \approx 0.8 \text{ dB}$ as we change from $n = 3$ to 4, i.e., to achieve an average SEP of 0.03 with 4-bit quantization, we need 0.8 dB more transmit power than what is required with 3-bit quantization.

Fig. 3.7 plots the average SEP as a function of SNR for DQPSK modulations scheme with coherent and non-coherent detection. We plot the average SEP for both the cases by using $n = \log_2 M, \log_2 M + 1, \log_2 M + 2$ and $\log_2 M + 3$ bits. From the plots, we

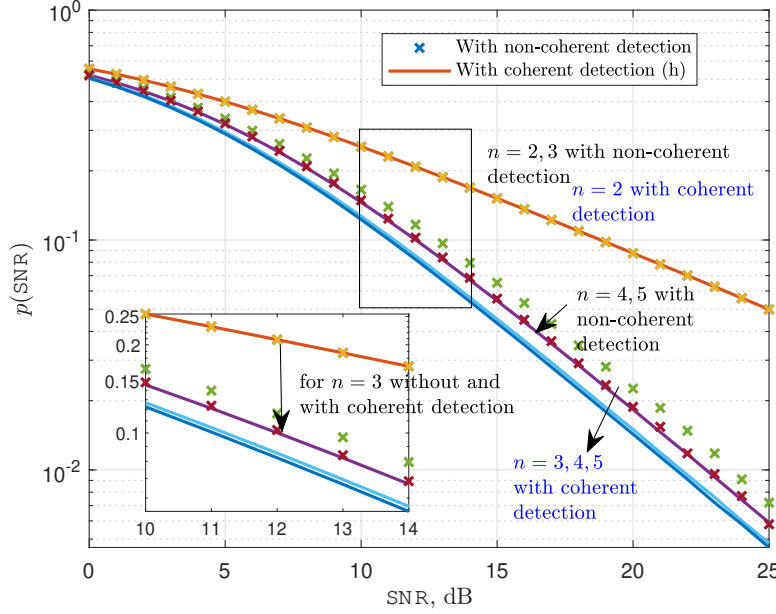


Figure 3.7: Average SEP curves as a function of SNR for coherent and non-coherent detection with DQPSK modulation. $n = 2, 3, 4, 5$. For $n = 2$ -bit quantization, both coherent and non-coherent detection show the same SEP behaviour with respect to average SNR. It is also visible that the full DVO is achieved with $n = 3$ under coherent detection. On the other hand, it requires $n = 4$ for non-coherent detection to achieve full DVO.

can clearly observe that DQPSK modulation with 2-bit quantization shows a similar SEP curve for both non-coherent and coherent detection with a DVO of $\frac{1}{2}$. DQPSK modulation with non-coherent detection with 2 and 3-bit quantization have a DVO of $\frac{1}{2}$ whereas DQPSK modulation with coherent detection with 2-bit quantization has a DVO of $\frac{1}{2}$. Further, we can observe that DQPSK with 4 or more bits with non-coherent detection and DQPSK with 3 or more bits with coherent detection have a DVO of 1, which is equal to the full diversity under the Rayleigh fading. To further emphasize this point, the zoomed-in section in Fig. 3.7 illustrates the average SEP versus SNR for DQPSK modulation under the two different detection methods. As stated in Theorem 3.3, these numerical observations clearly indicate the ternary behaviour in the decay exponent for $p(\text{SNR})$ under non-coherent detection, i.e., the $\text{DVO} = 1, \frac{1}{2}$ or 0 depending on whether $n \geq \log_2 M + 2$, $n = \log_2 M, \log_2 M + 1$, or $n < \log_2 M$, respectively. Similarly, as stated in Theorem 3.7, these numerical observations clearly indicate the ternary behaviour in the decay exponent for $p(\text{SNR})$ under coherent detection, i.e.,

$\text{DVO} = 1, \frac{1}{2}$ or 0 depending on whether $n \geq \log_2 M + 1$, $n = \log_2 M$, or $n < \log_2 M$, respectively under coherent detection.

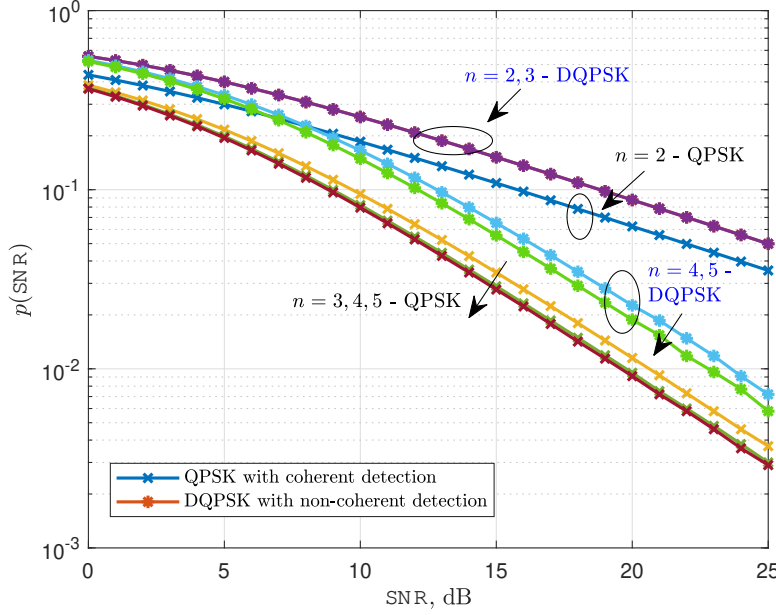


Figure 3.8: Average SEP curves as a function of SNR for QPSK modulation and DQPSK modulation with non-coherent detection. $n = 2, 3, 4, 5$.

Fig. 3.8 plots the average SEP as a function of SNR for QPSK and DQPSK modulation schemes with n -bit quantization. We adopt coherent and non-coherent detection for QPSK and DQPSK modulation schemes, respectively. We plot the average SEP for each modulation scheme by using $n = \log_2 M$, $\log_2 M + 1$, $\log_2 M + 2$ and $\log_2 M + 3$ bits. The analytical results for the average SEP of QPSK modulation under coherent detection were generated using Theorem 3 of [92]. The plot clearly illustrates that QPSK modulation with 2-bit and DQPSK modulation with 2 and 3-bit quantization have a DVO of $\frac{1}{2}$. Further, we can observe that QPSK modulation with 3 or more bits and DQPSK modulation with 4 or more bits quantizations have a DVO of 1, which is equal to the full DVO under Rayleigh fading.

Fig. 3.9 plots the average SEP as a function of SNR for QPSK and DQPSK modulation schemes with n -bit quantization under the Rayleigh fading scenario. Here we use coherent detection for both QPSK and DQPSK modulation schemes. Similar to Fig. 3.8, the analytical results for the average SEP of QPSK modulation under coherent

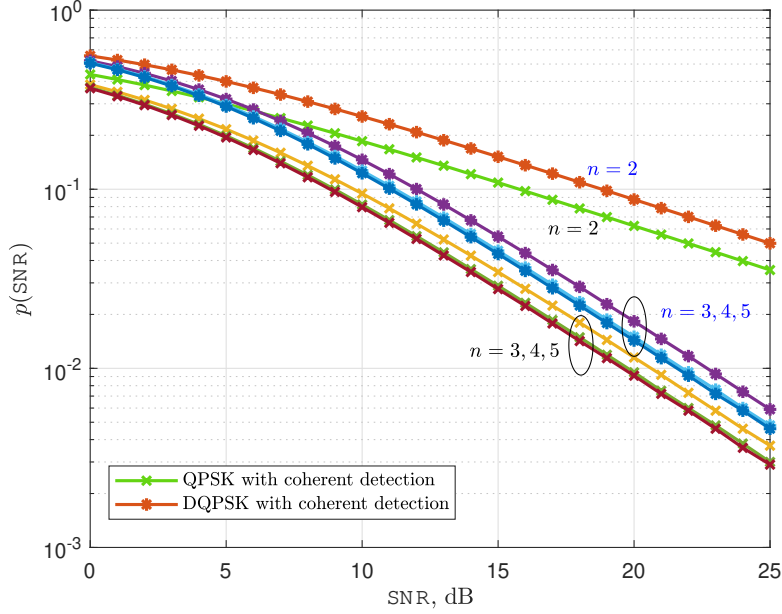


Figure 3.9: Average SEP curves as a function of SNR for QPSK modulation and DQPSK modulation with coherent detection. $n = 2, 3, 4, 5$.

detection was generated using Theorem 3 of [92] while that for DQPSK modulation under coherent detection was generated using (3.39). We plot the average SEP for each modulation scheme by using $n = \log_2 M$, $\log_2 M + 1$, $\log_2 M + 2$ and $\log_2 M + 3$ bits. From the plots, we can clearly observe that QPSK modulation with 2-bit and DQPSK modulation with 2-bit quantization has a DVO of $\frac{1}{2}$ under coherent detection. Further, we can observe that QPSK and DQPSK modulation schemes with 3 or more bit quantization have a DVO of 1, which is equal to the full DVO under coherent detection.

3.8 Conclusions and Future Generalizations

In this chapter, we performed a theoretical analysis of a low-resolution based ADC communication system and obtained fundamental performance limits, optimum ML detectors and a general analytical expression for the average SEP for DQPSK modulation with n -bit quantization. These results were further investigated for the Rayleigh fading model in detail. We conducted an asymptotic analysis to show that the decay exponent for the average SEP is the same and equal to 1 with infinite-bit and n -bit

quantizers for $n \geq 4$ under non-coherent detection. Similarly, the decay exponent for the average SEP is the same and equal to 1 with infinite-bit and n -bit quantizers for $n \geq 3$ under coherent detection. We also performed an extensive numerical study to illustrate the accuracy of the derived analytical expressions.

Channel estimation has been one of the major challenges associated with low-resolution quantization based communication systems as all these coherent detection techniques require CSI to be available at the receiver. In this work, we took a different approach to handle this challenge by adopting non-coherent data detection. We adopt the DQPSK modulation to differentially encode the transmit data and at the receiver side, employ non-coherent detection that does not require instantaneous CSI. Phase modulation has an important and practical layering feature enabling the quantizer and detector design separation in low-resolution ADC communications. For a given number of bits, the quantizer needs to be designed only once and can be kept constant for all channel realizations. The detector can be implemented digitally as a table look-up procedure using channel knowledge and quantizer output [92]. On the other hand, this feature is lost in joint phase and amplitude modulation schemes such as QAM. The quantizer needs to be dynamically updated for each channel realization in low-resolution ADC based QAM systems. This is because the fading channel amplitude may vary over a wide range, but the phase always varies over $[-\pi, \pi)$. However, phase modulation is historically known to be optimum only up to modulation order 16 under peak power limitations [87]. Hence, it is a notable future research direction to extend the results of this chapter to higher order phase and amplitude modulations by taking practical design considerations into account.

A major result of this work is the discovery of a ternary SEP behaviour, indicating the sufficiency of 4 bits for achieving asymptotically optimum DQPSK communication reliability under non-coherent detection. This is one extra bit more than to achieve the asymptotically optimum communication performance when the channel knowledge is available at the receiver. This is the penalty we have to pay in terms of the number of quantization bits for not having the CSI at the receiver of a low-resolution quantization based communication system. Another important future research direction is to extend

the result to general M -ary differential modulation schemes under Nakagami- m fading. Similarly, utilizing the results of this chapter, a further detailed study on the receiver architecture design to determine where to place the diversity combiner (before or after quantizer or detector) and its type is needed when multiple diversity branches are available for data reception.

3.9 Appendix

3.9.1 Proof of Theorem 3.1

We will first focus on $n = 2$ case. For $n = 2$, we note that both the detection rules 1 and 2 are overlap with each other. In this case, without loss of generality, we will assume that $k_{t-1} = k_t = 0$. The proof for different combinations of k_{t-1} and k_t is similar and will be omitted to avoid repetitions.

In order to obtain the ML rule for the setup described above, we have to maximize the likelihood function $\mathcal{L}(s_{t-1}, s_t)$ given by

$$\begin{aligned}\mathcal{L}(s_{t-1}, s_t) &= \Pr\{Y_{t-1} \in \mathcal{R}_0, Y_t \in \mathcal{R}_0 \mid S_{t-1} = s_{t-1}, S_t = s_t\} \\ &= \Pr\left\{\sqrt{\text{SNR}} H s_{t-1} + W_{t-1} \in \mathcal{R}_0, \sqrt{\text{SNR}} H s_t + W_t \in \mathcal{R}_0 \mid S_{t-1} = s_{t-1}, S_t = s_t\right\}.\end{aligned}\quad (3.18)$$

We note that given S_{t-1}, S_t , the message transmitted at time t is completely determined according to the encoding rule in (3.1). In particular, M_t depends only on the phase difference between S_{t-1} and S_t . Based on this observation, it will be enough to find the optimum phase difference between s_{t-1} and s_t maximizing $\mathcal{L}(s_{t-1}, s_t)$. To this end, we will write $\mathcal{L}(s_{t-1}, s_t)$ as

$$\begin{aligned}\mathcal{L}(s, \phi) &= \Pr\{Y_{t-1} \in \mathcal{R}_0, Y_t \in \mathcal{R}_0 \mid S_{t-1} = s, S_t = s e^{j\phi}\} \\ &= \Pr\left\{\sqrt{\text{SNR}} H s + W_{t-1} \in \mathcal{R}_0, \sqrt{\text{SNR}} H s e^{j\phi} + W_t \in \mathcal{R}_0 \mid S_{t-1} = s, S_t = s e^{j\phi}\right\}\end{aligned}\quad (3.19)$$

for a given $s \in \mathcal{C}$ and $\phi \in \Phi = \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$. We can graphically represent an example case for the decoding process of the detector with 2-bit quantization as in Fig. 3.10. The

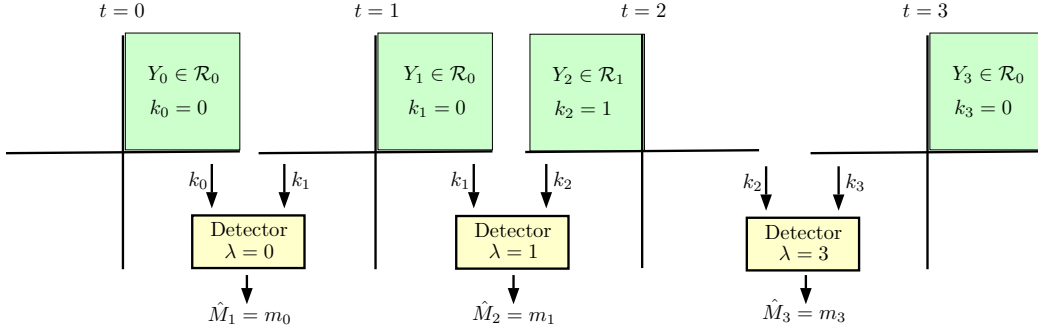


Figure 3.10: The decoding process of the ML detector with 2-bit quantization with differential modulation.

detector compares two consecutive quantized observations to estimate each message.

The critical step to obtain the pair (s^*, ϕ^*) maximizing $\mathcal{L}(s, \phi)$ in (3.19) is to obtain the joint distribution for the vector $[Y_{t-1}, Y_t]^T$. To this end, we will utilize the fact that the channel is Rayleigh distributed and slowly varying in our setup. We first observe that the received signals $Y_{t-1} = \sqrt{\text{SNR}} Hs + W_{t-1}$ and $Y_t = \sqrt{\text{SNR}} Hse^{j\phi} + W_t$ are identically distributed circularly-symmetric zero-mean complex Gaussian random variables at times $t-1$ and t . However, they are not independent. The slowly-varying fading process introduces dependency between them. It will be enough to obtain their covariance matrix $\mathbf{K}$ obtain the joint distribution of Y_{t-1} and Y_t since they are jointly Gaussian. The derivation of $\mathbf{K}$ is given as below.

We first define the vectors $\mathbf{Y}_{t-1} = [Y_{t-1}^{\text{re}} \ Y_{t-1}^{\text{im}}]^T$, $\mathbf{Y}_t = [Y_t^{\text{re}} \ Y_t^{\text{im}}]^T$, $\mathbf{W}_{t-1} = [W_{t-1}^{\text{re}} \ W_{t-1}^{\text{im}}]^T$, $\mathbf{W}_t = [W_t^{\text{re}} \ W_t^{\text{im}}]^T$, $\mathbf{H} = [H^{\text{re}} \ H^{\text{im}}]^T$ and $\mathbf{s} = [s^{\text{re}} \ s^{\text{im}}]^T$, where Z^{re} and Z^{im} represent the real and imaginary parts of the complex variable $Z \in \mathbb{C}$ (i.e., $Z^{\text{re}} = \text{Re}(Z)$ and $Z^{\text{im}} = \text{Im}(Z)$). With these definitions, we can write

$$\begin{bmatrix} Y_{t-1}^{\text{re}} \\ Y_{t-1}^{\text{im}} \end{bmatrix} = \sqrt{\text{SNR}} \begin{bmatrix} H^{\text{re}} s^{\text{re}} - H^{\text{im}} s^{\text{im}} \\ H^{\text{re}} s^{\text{im}} + H^{\text{im}} s^{\text{re}} \end{bmatrix} + \begin{bmatrix} W_{t-1}^{\text{re}} \\ W_{t-1}^{\text{im}} \end{bmatrix} \quad (3.20)$$

and

$$\begin{bmatrix} Y_t^{\text{re}} \\ Y_t^{\text{im}} \end{bmatrix} = \sqrt{\text{SNR}} \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix} \begin{bmatrix} H^{\text{re}} s^{\text{re}} - H^{\text{im}} s^{\text{im}} \\ H^{\text{re}} s^{\text{im}} + H^{\text{im}} s^{\text{re}} \end{bmatrix} + \begin{bmatrix} W_t^{\text{re}} \\ W_t^{\text{im}} \end{bmatrix}, \quad (3.21)$$

where $W_{t-1}^{\text{re}}, W_{t-1}^{\text{im}}, W_t^{\text{re}}$, and W_{t-1}^{im} are all independent random variables with a Gaussian distribution of zero mean and a variance of $\frac{1}{2}$ (i.e., $W_{t-1}^{\text{re}}, W_{t-1}^{\text{im}}, W_t^{\text{re}}, W_{t-1}^{\text{im}} \sim \mathcal{N}(0, \frac{1}{2})$). Further defining $\mathbf{Y} = [Y_{t-1}^{\text{re}} \ Y_{t-1}^{\text{im}} \ Y_t^{\text{re}} \ Y_t^{\text{im}}]^T$ and taking $\mathbf{K} = \mathbb{E}[\mathbf{Y}\mathbf{Y}^T]$ (i.e., $\mathbb{E}[\mathbf{Y}] = \mathbf{0}$), we can obtain

$$\mathbf{K} = \frac{1}{2} \begin{bmatrix} \text{SNR} + 1 & 0 & \text{SNR} \cos \phi & \text{SNR} \sin \phi \\ 0 & \text{SNR} + 1 & -\text{SNR} \sin \phi & \text{SNR} \cos \phi \\ \text{SNR} \cos \phi & -\text{SNR} \sin \phi & \text{SNR} + 1 & 0 \\ \text{SNR} \sin \phi & \text{SNR} \cos \phi & 0 & \text{SNR} + 1 \end{bmatrix}$$

after some matrix calculations.

Therefore we can write the pdf of $\mathbf{Y}$ using multi-variate Gaussian distribution [88] as

$$f_Y^{(\phi)}(\mathbf{y}) = \frac{1}{4\pi^2 \sqrt{|\mathbf{K}|}} \exp \left\{ -\frac{1}{2} \mathbf{y}^T \mathbf{K}^{-1} \mathbf{y} \right\} = \frac{1}{\pi^2 (2\text{SNR} + 1)} \exp \{G(\mathbf{y}, \phi)\}, \quad (3.22)$$

where $|\mathbf{K}| = \left(\frac{2\text{SNR}+1}{4}\right)^2$ is the determinant of the covariance matrix for all $\phi \in \Phi$ and $G(\mathbf{y}, \phi) = -\frac{1}{2} \mathbf{y}^T \mathbf{K}^{-1} \mathbf{y}$ is given as

$$G(\mathbf{y}, \phi) = \begin{cases} A + B(-y_{t-1}^{\text{re}} y_t^{\text{re}} - y_{t-1}^{\text{im}} y_t^{\text{im}}) & \phi = 0, \\ A + B(-y_{t-1}^{\text{re}} y_t^{\text{im}} + y_{t-1}^{\text{im}} y_t^{\text{re}}) & \phi = \frac{\pi}{2}, \\ A + B(y_{t-1}^{\text{re}} y_t^{\text{re}} + y_{t-1}^{\text{im}} y_t^{\text{im}}) & \phi = \pi, \\ A + B(y_{t-1}^{\text{re}} y_t^{\text{im}} - y_{t-1}^{\text{im}} y_t^{\text{re}}) & \phi = \frac{3\pi}{2}, \end{cases}$$

for different values of $\phi \in \Phi$. Here $A = -\frac{\text{SNR}+1}{2\text{SNR}+1} \left\{ (y_{t-1}^{\text{re}})^2 + (y_{t-1}^{\text{im}})^2 + (y_t^{\text{re}})^2 + (y_t^{\text{im}})^2 \right\}$ and $B = \frac{2\text{SNR}}{2\text{SNR}+1}$.

Using the pdf in (3.22), we now can evaluate the likelihood function in (3.19) for

any $S_{t-1} = s \in \mathcal{C}$ as

$$\mathcal{L}(s, \phi) = \frac{1}{\pi^2 (2\text{SNR} + 1)} \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty \exp \{G(\mathbf{y}, \phi)\} dy_{t-1}^{\text{re}} dy_{t-1}^{\text{im}} dy_t^{\text{re}} dy_t^{\text{im}}. \quad (3.23)$$

The limits of the integrals came because $Y_{t-1}, Y_t \in \mathcal{R}_0$ where $Y_{t-1}^{\text{re}} \geq 0, Y_{t-1}^{\text{im}} \geq 0, Y_t^{\text{re}} \geq 0$ and $Y_t^{\text{im}} \geq 0$. By comparing the likelihood function for different values of $\phi \in \Phi$, we can conclude that

$$\mathcal{L}(s, 0) \geq \mathcal{L}\left(s, \frac{\pi}{2}\right) = \mathcal{L}\left(s, \frac{3\pi}{2}\right) \geq \mathcal{L}(s, \pi), \quad (3.24)$$

for any $s \in \mathcal{C}$. Therefore the maximum likelihood occurs when $\phi^* = 0$ which corresponds to the estimated message $\hat{M}_t = m_0$. This is the same as the output of the detector presented in Theorem 3.1 for $n = 2$. The analysis can be simply extended for all the cases with different combinations of k_{t-1} and k_t in a similar manner.

We now focus on $n \geq 3$ case. In this case, the detector has to choose which detection rule to use first. We first provide the proof for the detection rule 1 given as (3.7). The steps of the proof for the detection rule 2 is similar and we will brief it at the end of this proof. We will also provide a graphical example at the end of the proof to showcase the difference between the two detection rules.

According to the detection rule 1 given in (3.6), if $Y_{t-1} \in \mathcal{R}_k$ and $Y_t \in \mathcal{E}_k$ where

$$\mathcal{E}_k = \left\{ z \in \mathbb{C} : \frac{\pi}{2^{n-1}}k - \frac{\pi}{4} \leq \text{Arg}(z) < \frac{\pi}{2^{n-1}}k + \frac{\pi}{4} \right\}$$

for $k \in [0 : 2^{n-1}]$, the detector outputs m_0 as the estimated message. Without loss of generality, in this proof, we assume that $Y_t \in \mathcal{R}_0$ (i.e. $k_{t-1} = 0$) and $Y_t \in \mathcal{E}_0$. As an example, for $n = 3$, if $k_t = 0$ or $k_t = 7$ given that $k_{t-1} = 0$, the detector outputs m_0 as the estimated message. The proof for different combinations of $Y_{t-1} \in \mathcal{R}_k$ and $Y_t \in \mathcal{E}_k$ is similar and will be omitted to avoid repetitions.

In order to obtain the ML rule for the setup described above, we have to maximize

the likelihood function $\mathcal{L}(s_{t-1}, s_t)$ given by

$$\begin{aligned}\mathcal{L}(s_{t-1}, s_t) &= \Pr\{Y_{t-1} \in \mathcal{R}_0, Y_t \in \mathcal{E}_0 \mid S_{t-1} = s_{t-1}, S_t = s_t\} \\ &= \Pr\left\{\sqrt{\text{SNR}} Hs_{t-1} + W_{t-1} \in \mathcal{R}_0, \sqrt{\text{SNR}} Hs_t + W_t \in \mathcal{E}_0 \mid S_{t-1} = s_{t-1}, S_t = s_t\right\}.\end{aligned}$$

Since M_t depends only on the phase difference between S_{t-1} and S_t , it is enough to find the optimum phase difference between s_{t-1} and s_t maximizing $\mathcal{L}(s_{t-1}, s_t)$. Therefore, we can write

$$\begin{aligned}\mathcal{L}(s, \phi) &= \Pr\{Y_{t-1} \in \mathcal{R}_0, Y_t \in \mathcal{E}_0 \mid S_{t-1} = s, S_t = se^{j\phi}\} \\ &= \Pr\left\{\sqrt{\text{SNR}} Hs + W_{t-1} \in \mathcal{R}_0, \sqrt{\text{SNR}} Hse^{j\phi} + W_t \in \mathcal{E}_0 \mid S_{t-1} = s, S_t = se^{j\phi}\right\},\end{aligned}\tag{3.25}$$

for a given $s \in \mathcal{C}$ and $\phi \in \Phi$.

By using similar steps to the analysis provided for $n = 2$ case, we can express the likelihood function as

$$\mathcal{L}(s, \phi) = \int_{\mathcal{R}_0 \times \mathcal{E}_0} f_Y^{(\phi)}(\mathbf{y}) d\mathbf{y}.$$

However, in this case, to compare the values of the likelihood function for different ϕ values, we have to re-arrange the likelihood function in terms of the pdf of Y_{t-1} and the conditional pdf of Y_t given Y_{t-1} . Further, only the latter pdf depends on ϕ . Then we write

$$\begin{aligned}\mathcal{L}(s, \phi) &= \int_{\mathcal{R}_0 \times \mathcal{E}_0} f_{Y_{t-1}}(\mathbf{y}_{t-1}) f_{Y_t|Y_{t-1}}^{(\phi)}(\mathbf{y}_t | \mathbf{y}_{t-1}) d\mathbf{y}_{t-1} d\mathbf{y}_t \\ &= \int_{\mathcal{R}_0} f_{Y_{t-1}}(\mathbf{y}_{t-1}) \Pr\{Y_t \in \mathcal{E}_0 \mid Y_{t-1} = \mathbf{y}_{t-1}, S_{t-1} = s, S_t = se^{j\phi}\} d\mathbf{y}_{t-1}.\end{aligned}\tag{3.26}$$

We now need to evaluate the probability $\Pr\{Y_t \in \mathcal{E}_0 \mid Y_{t-1} = \mathbf{y}_{t-1}, S_{t-1} = s, S_t = se^{j\phi}\}$

for $\phi \in \Phi$. To that end, we will first find the conditional distribution of $\mathbf{Y}_t$ given $\mathbf{Y}_{t-1}$ as

$$f_{\mathbf{Y}_t | \mathbf{Y}_{t-1}}^{(\phi)}(\mathbf{y}_t | \mathbf{y}_{t-1}) = \frac{f_{\mathbf{Y}}^{(\phi)}(\mathbf{y})}{f_{\mathbf{Y}_{t-1}}(\mathbf{y}_{t-1})},$$

where $f_{\mathbf{Y}}^{(\phi)}(\mathbf{y})$ is given in (3.22) and the marginal distribution of $\mathbf{Y}_{t-1}$ is given as

$$f_{\mathbf{Y}_{t-1}}(\mathbf{y}_{t-1}) = \frac{1}{\pi(\text{SNR} + 1)} \exp\left(-\frac{1}{\text{SNR} + 1} \|\mathbf{y}_{t-1}\|_2^2\right). \quad (3.27)$$

Therefore, we can write

$$f_{\mathbf{Y}_t | \mathbf{Y}_{t-1}}^{(\phi)}(\mathbf{y}_t | \mathbf{y}_{t-1}) = \frac{\text{SNR} + 1}{\pi(2\text{SNR} + 1)} \exp\left(-\frac{\text{SNR} + 1}{2\text{SNR} + 1} \left\| \mathbf{y}_t - \frac{\text{SNR}e^{j\phi}}{\text{SNR} + 1} \mathbf{y}_{t-1} \right\|_2^2\right). \quad (3.28)$$

By inspecting the above conditional pdf, we observe that $\mathbf{Y}_t$ is Gaussian distributed and its conditional mean $\mathbb{E}[\mathbf{Y}_t | \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}] = \frac{\text{SNR}e^{j\phi}}{\text{SNR} + 1} \mathbf{y}_{t-1}$. This means that the mean of $\mathbf{Y}_t$ lies in $\mathcal{R}_0 e^{j\phi}$ given that $\mathbf{Y}_{t-1} \in \mathcal{R}_0$ for $\phi \in \Phi$. Hence, by using Lemma 6 of [92], we can conclude that

$$0 = \arg \max_{\phi \in \Phi} \Pr\{\mathbf{Y}_t \in \mathcal{E}_0 | \mathbf{Y}_{t-1} \in \mathcal{R}_0, S_{t-1} = s, S_t = se^{j\phi}\}, \quad (3.29)$$

since $\mathbb{E}[\mathbf{Y}_t | \mathbf{Y}_{t-1}]$ is closest to the bisector $\mathcal{H}_{\mathcal{E}_0}$ of $\mathcal{E}_0$ when $\phi = 0$. We use Fig. 3.11 as a visual guide to further illustrate this.

Therefore the maximum likelihood value of (3.26) occurs when $\phi = 0$ which corresponds to the estimated message $\hat{M}_t = m_0$. This coincides with the output of the detector presented in Theorem 3.1. With similar steps, we can prove the theorem for all the combinations of $\mathbf{Y}_{t-1} \in \mathcal{R}_k$ and $\mathbf{Y}_t \in \mathcal{E}_k$ for all $k \in [0 : 2^{n-1}]$.

We use Fig. 3.12 as a visual guide to illustrate how Fig. 3.11 would be changed if we select the detection rule 2 given in (3.8). According to the detection rule 2 given in (3.6), if $\mathbf{Y}_{t-1} \in \mathcal{R}_k$ and $\mathbf{Y}_t \in \mathcal{E}_k$ where

$$\mathcal{E}_k = \left\{ z \in \mathbb{C} : \frac{\pi}{2^{n-1}}(k+1) - \frac{\pi}{4} \leq \text{Arg}(z) < \frac{\pi}{2^{n-1}}(k+1) + \frac{\pi}{4} \right\}$$

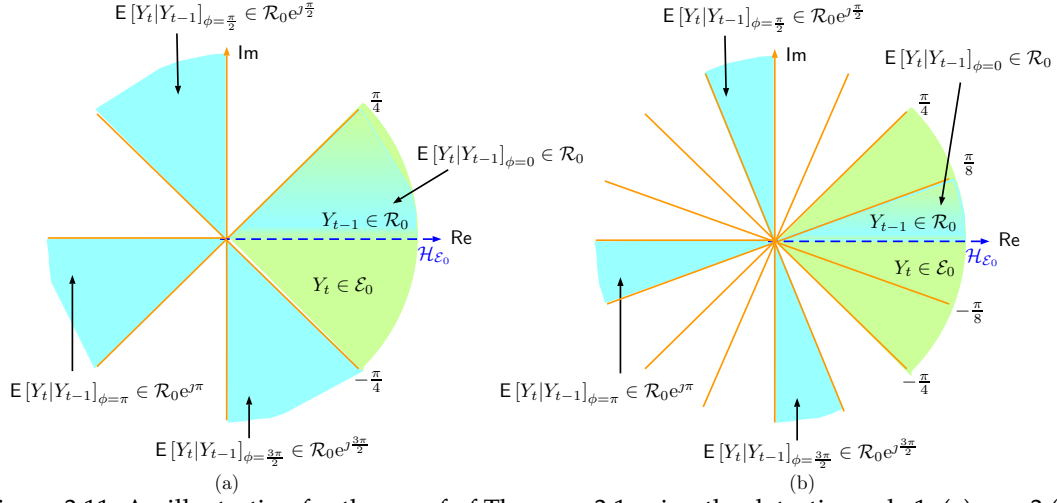


Figure 3.11: An illustration for the proof of Theorem 3.1 using the detection rule 1. (a) $n = 3$ (b) $n = 4$. Given that, $Y_{t-1} \in \mathcal{R}_0$, $E[Y_t|Y_{t-1}]$ should be in either one of the blue coloured regions depending on the value for ϕ . Given $Y_t \in \mathcal{E}_0$, $E[Y_t|Y_{t-1}]$ become closest to the bisector $\mathcal{H}_{\mathcal{E}_0}$ of the region $\mathcal{E}_0$ when $\phi = 0$.

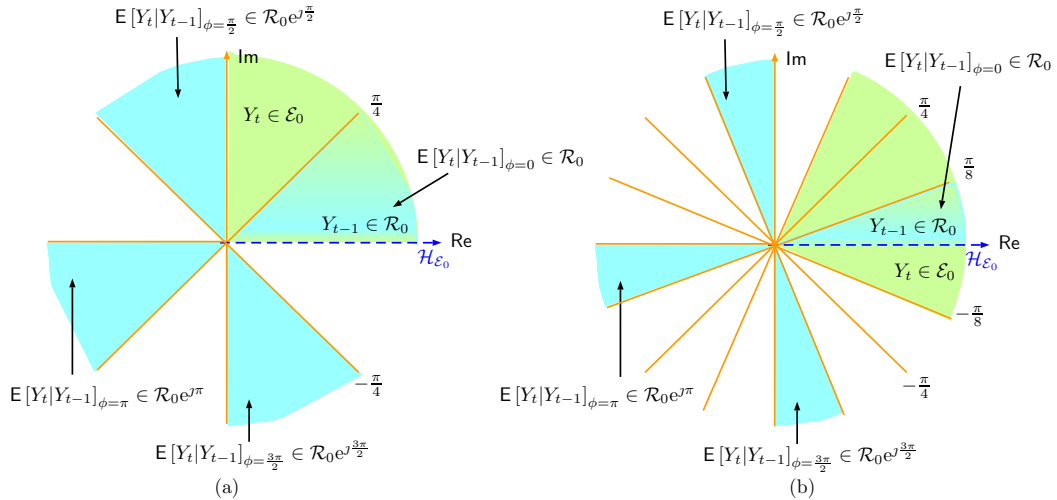


Figure 3.12: An illustration for the proof of Theorem 3.1 using the detection rule 2. (a) $n = 3$ (b) $n = 4$. Given that, $Y_{t-1} \in \mathcal{R}_0$, $E[Y_t|Y_{t-1}]$ should be in either one of the blue coloured regions depending on the value for ϕ . Given $Y_t \in \mathcal{E}_0$, $E[Y_t|Y_{t-1}]$ become closest to the bisector $\mathcal{H}_{\mathcal{E}_0}$ of the region $\mathcal{E}_0$ when $\phi = 0$.

for $k \in [0 : 2^{n-1}]$, the detector outputs m_0 as the estimated message. As an example, for $n = 3$, if $k_t = 0$ or $k_t = 1$ given that $k_{t-1} = 0$, the detector outputs m_0 as the estimated message. The proof with detection rule 2 is similar to that of detection rule 1 and will be omitted to avoid repetitions.

3.9.2 Proof of Theorem 3.2

We first focus on $n = 2$ case. In this case, without loss of generality, let us assume that $M_t = m_0$ (i.e., $S_{t-1} = S_t$). Then, there will not be any error in estimating the message $\hat{M}_t$ if the two consecutive received signals land in the same quantization region (i.e., $Q(Y_{t-1}) = Q(Y_t)$). Therefore, we can write the SEP as

$$\begin{aligned} p(\text{SNR}) &= \sum_{k=0}^{2^n-1} \Pr\{Y_{t-1} \in \mathcal{R}_k, Y_t \notin \mathcal{R}_k \mid S_{t-1} = S_t\}, \\ &\stackrel{(a)}{=} 2^n \Pr\{Y_{t-1} \in \mathcal{R}_0, Y_t \notin \mathcal{R}_0 \mid S_{t-1} = S_t\}, \end{aligned} \quad (3.30)$$

where (a) follows due to the circular symmetry of the problem.

We can write $\Pr\{Y_{t-1} \in \mathcal{R}_0, Y_t \notin \mathcal{R}_0 \mid S_{t-1} = S_t\}$ by integrating the pdf in (3.22) as

$$\begin{aligned} \Pr\{Y_{t-1} \in \mathcal{R}_0, Y_t \notin \mathcal{R}_0 \mid S_{t-1} = S_t\} &= \\ &= \frac{1}{\pi^2 (2\text{SNR} + 1)} \int_{-\infty}^{\infty} \int_0^{-\infty} \int_0^{\infty} \int_0^{\infty} \mathcal{G}(\mathbf{y}) \, dy_{t-1}^{\text{im}} \, dy_{t-1}^{\text{re}} \, dy_t^{\text{re}} \, dy_t^{\text{im}} \\ &+ \frac{1}{\pi^2 (2\text{SNR} + 1)} \int_0^{-\infty} \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} \mathcal{G}(\mathbf{y}) \, dy_{t-1}^{\text{im}} \, dy_{t-1}^{\text{re}} \, dy_t^{\text{re}} \, dy_t^{\text{im}}, \end{aligned} \quad (3.31)$$

where $\mathcal{G}(\mathbf{y})$ is as given in Theorem 3.2.

We will now focus on $n \geq 3$ case. In this case, without loss of generality, let us assume that the message $M_t = m_0$ is transmitted. That is $S_{t-1} = S_t$. According to the ML detector given in Theorem 3.1, there will be no error in detecting the message if the received signal $Y_t \in \mathcal{E}_k$ given that $Y_{t-1} \in \mathcal{R}_k$ where

$$\mathcal{E}_k = \left\{ z \in \mathbb{C} : \frac{\pi}{2^{n-1}}k - \frac{\pi}{4} \leq \text{Arg}(z) < \frac{\pi}{2^{n-1}}k + \frac{\pi}{4} \right\}$$

for $k \in [0 : 2^{n-1}]$. Here we use the detection rule 1 given in (3.7). There will be no difference in SEP expression even if we use the detection rule 2 given in (3.8). We use Fig. 3.13 as a visual guide to illustrate an example for $n = 3$ and $n = 4$ cases.

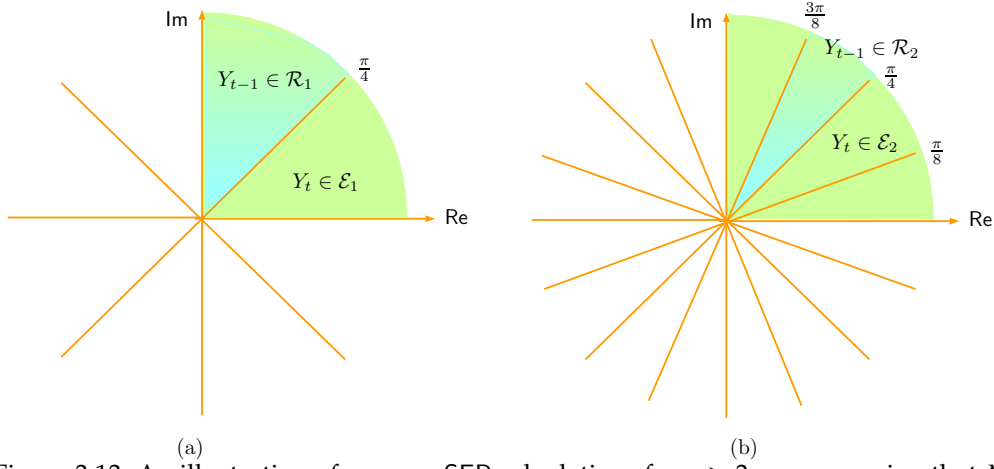


Figure 3.13: An illustration of average SEP calculations for $n \geq 3$ case assuming that $M_t = m_0$. (a) $n = 3$. Given that, $Y_{t-1} \in \mathcal{R}_1$, if $Y_t \in \mathcal{E}_1$, there will be no error in the estimating the message $\hat{M}_t$. (b) $n = 4$. Given that, $Y_{t-1} \in \mathcal{R}_2$, if $Y_t \in \mathcal{E}_2$, there will be no error in the estimating the message $\hat{M}_t$. Without loss of generality, we choose k' in which $Y_{t-1} \in \mathcal{R}_{k'}$ such that $Y_t \in \mathcal{E}_{k'}$ where $\mathcal{E}_{k'} = \{z \in \mathbb{C} : 0 \leq \text{Arg}(z) < \frac{\pi}{2}\}$.

Then, we can write the SEP as

$$\begin{aligned}
 p(\text{SNR}) &= \sum_{k=0}^{2^n-1} \Pr\{Y_{t-1} \in \mathcal{R}_k, Y_t \notin \mathcal{E}_k \mid S_{t-1} = S_t\}, \\
 &\stackrel{(a)}{=} 2^n \Pr\{Y_{t-1} \in \mathcal{R}_{k'}, Y_t \notin \mathcal{E}_{k'} \mid S_{t-1} = S_t\}, \tag{3.32}
 \end{aligned}$$

where (a) follows due to the circular symmetry of the problem. Without loss of generality, we choose $\mathcal{R}_{k'} = \{z \in \mathbb{C} : \frac{\pi}{4} \leq \text{Arg}(z) < \frac{\pi}{4} + \frac{\pi}{2^{n-1}}\}$ and $\mathcal{E}_{k'} = \{z \in \mathbb{C} : 0 \leq \text{Arg}(z) < \frac{\pi}{2}\}$. As illustrated in Fig. 3.13, $k' = 1$ and 2 respectively for $n = 3$ and 4 cases.

Then, we can write $\Pr\{Y_{t-1} \in \mathcal{R}_{k'}, Y_t \notin \mathcal{E}_{k'} \mid S_{t-1} = S_t\}$ by integrating the pdf in (3.22) as

$$\begin{aligned}
 \Pr\{Y_{t-1} \in \mathcal{R}_{k'}, Y_t \notin \mathcal{E}_{k'} \mid S_{t-1} = S_t\} &= \\
 &\frac{1}{\pi^2 (2\text{SNR} + 1)} \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{\infty} \int_{y_{t-1}^{\text{re}}}^{y_{t-1}^{\text{re}} \tan\left(\frac{\pi}{4} + \frac{\pi}{2^{n-1}}\right)} \mathcal{G}(\mathbf{y}) \, dy_{t-1}^{\text{im}} \, dy_{t-1}^{\text{re}} \, dy_t^{\text{re}} \, dy_t^{\text{im}} \\
 &+ \frac{1}{\pi^2 (2\text{SNR} + 1)} \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} \int_{y_{t-1}^{\text{re}}}^{y_{t-1}^{\text{re}} \tan\left(\frac{\pi}{4} + \frac{\pi}{2^{n-1}}\right)} \mathcal{G}(\mathbf{y}) \, dy_{t-1}^{\text{im}} \, dy_{t-1}^{\text{re}} \, dy_t^{\text{re}} \, dy_t^{\text{im}}
 \end{aligned}$$

where $\mathcal{G}(\mathbf{y})$ is as given in Theorem 3.2. Therefore, we get the SEP expression in

Theorem 3.2 for $n \geq 3$.

3.9.3 Proof of Corollary 3.1

We first focus on the $n = 2$ case and start the proof by defining the SEP conditioned on a particular realization of the fading coefficient. To that end, let us assume $H = h = re^{j\theta_h}$ and $S_{t-1} = S_t = s = e^{j\theta_s}$, where $|h| = r$, $\text{Arg}(h) = \theta_h$ and $\text{Arg}(s) = \theta_s$. Then, with a slight abuse of notation, we write the conditional SEP as

$$\begin{aligned} p(\text{SNR}, h) &= 1 - 4\Pr\left\{\sqrt{\text{SNR}}r e^{j\theta} + W_{t-1} \in \mathcal{R}_0, \sqrt{\text{SNR}}r e^{j\theta} + W_t \in \mathcal{R}_0\right\}, \\ &= 1 - 4Q^2\left(-\sqrt{2\text{SNR}}r \cos \theta\right) Q^2\left(-\sqrt{2\text{SNR}}r \sin \theta\right) \end{aligned} \quad (3.33)$$

where $\theta = \theta_h + \theta_s$ and $Q^2(\cdot)$ is the second power of $Q(\cdot)$. By averaging the above expression with respect to Rayleigh fading distribution $f_{|H|}(r) = 2re^{-r^2}$ for $r \geq 0$ and $f_{\Theta}(\theta) = \frac{1}{2\pi}$, we get the expression in (3.10) for $n = 2$.

We now move to the proof for $n = 3$ case. In this case, we can write the conditional SEP

$$\begin{aligned} p(\text{SNR}, h) &= 1 - \sum_{k=0}^{2^n-1} \Pr\{Y_{t-1} \in \mathcal{R}_k, Y_t \in \mathcal{E}_k\}, \\ &\stackrel{(a)}{=} 1 - 4(\Pr\{Y_{t-1} \in \mathcal{R}_0, Y_t \in \mathcal{E}_0\} + \Pr\{Y_{t-1} \in \mathcal{R}_1, Y_t \in \mathcal{E}_1\}), \end{aligned}$$

where (a) follows due to the symmetry of the problem. Further, it is evident that

$$\Pr\{Y_{t-1} \in \mathcal{R}_1, Y_t \in \mathcal{E}_1\} = \Pr\{Y_{t-1} \in \mathcal{R}_1, Y_t \in \mathcal{E}_0\}$$

due to the symmetrical nature of the problem as illustrated in Fig. 3.14. Therefore, we have

$$\begin{aligned} p(\text{SNR}, h) &= 1 - 4(\Pr\{Y_{t-1} \in \mathcal{R}_0, Y_t \in \mathcal{E}_0\} + \Pr\{Y_{t-1} \in \mathcal{R}_1, Y_t \in \mathcal{E}_0\}), \\ &= 1 - 4\Pr\{Y_{t-1} \in \mathcal{R}_0 \cup \mathcal{R}_1, Y_t \in \mathcal{E}_0\}, \\ &= 1 - 4\Pr\left\{\sqrt{\text{SNR}}r e^{j\theta} + W_{t-1} \in \mathcal{R}_0 \cup \mathcal{R}_1, \sqrt{\text{SNR}}r e^{j\theta} + W_t \in \mathcal{E}_0\right\}, \end{aligned}$$

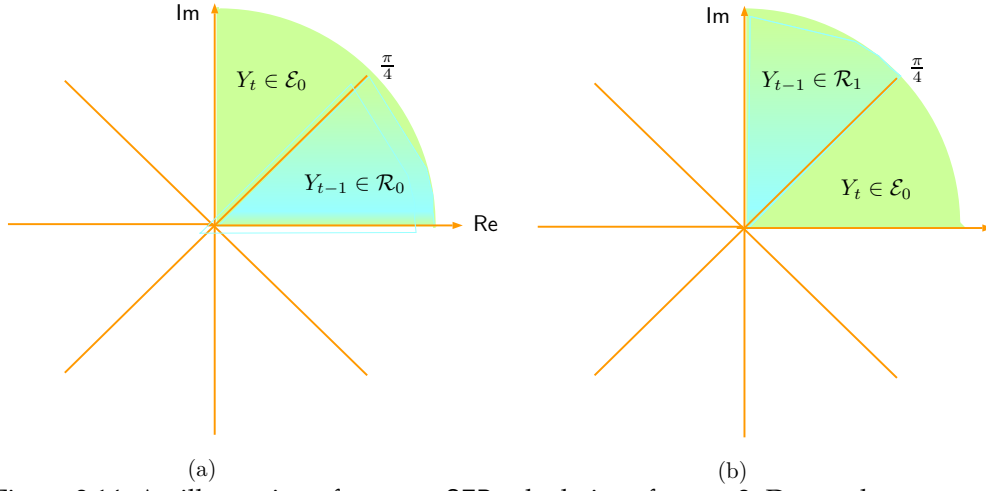


Figure 3.14: An illustration of average SEP calculations for $n = 3$. Due to the symmetry of the problem, $\Pr \{Y_{t-1} \in \mathcal{R}_1, Y_t \in \mathcal{E}_1\} = \Pr \{Y_{t-1} \in \mathcal{R}_1, Y_t \in \mathcal{E}_0\}$. By combining (a) and (b) together, we evaluate the probability of $\Pr \{Y_{t-1} \in \mathcal{R}_0 \cup \mathcal{R}_1, Y_t \in \mathcal{E}_0\}$.

which is the same conditional expression that we evaluated for $n = 2$ case. Hence, by averaging the conditional SEP with the fading distribution, we get the same SEP as given in Corollary 3.1.

3.9.4 Proof of Theorem 3.3

We will first focus on $n = 2$ and 3 cases. As given in Corollary 3.1, SEP expressions for both cases are the same. There are two methods that we can use to prove Theorem 3.3 for $n = 2$ and 3 cases. We can either work with the SEP expression given in Theorem 3.2 or the one given in Corollary 3.1.

Method 1:

In this case, we use the SEP expression given in Theorem 3.2 to prove the DVO result for $n = 2$ and 3. To that end, we can convert the express (3.9) to

$$p(\text{SNR}) = 1 - 4 \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty f_Y(\mathbf{y}) dy_{t-1}^{\text{re}} dy_{t-1}^{\text{im}} dy_t^{\text{re}} dy_t^{\text{im}},$$

where

$$f_Y(\mathbf{y}) = \frac{1}{\pi^2 (2\text{SNR} + 1)} \exp \left\{ -\frac{\text{SNR} + 1}{2\text{SNR} + 1} \|\mathbf{y}\|^2 + \frac{2\text{SNR}}{2\text{SNR} + 1} (y_{t-1}^{\text{re}} y_t^{\text{re}} + y_{t-1}^{\text{im}} y_t^{\text{im}}) \right\}$$

In order to find the decay rate of the above integral expression, we first define $\mathbf{Y}_{\text{re}} = [Y_{t-1}^{\text{re}} \ Y_t^{\text{re}}]^T$ and $\mathbf{Y}_{\text{im}} = [Y_{t-1}^{\text{im}} \ Y_t^{\text{im}}]^T$ such that $f_Y(\mathbf{y}) = f_{Y_{\text{re}}}(\mathbf{y}_{\text{re}}) f_{Y_{\text{im}}}(\mathbf{y}_{\text{im}})$ where $f_{Y_{\text{re}}}(\mathbf{y}_{\text{re}})$ and $f_{Y_{\text{im}}}(\mathbf{y}_{\text{im}})$ are the pdf of $\mathbf{Y}_{\text{re}}$ and $\mathbf{Y}_{\text{im}}$ respectively, given as

$$f_{Y_{\text{re}}}(\mathbf{y}_{\text{re}}) = \frac{1}{\pi\sqrt{2\text{SNR}+1}} \exp\left(-\frac{c_1}{2} \|\mathbf{y}_{\text{re}}\|_2^2 - c_2 y_{t-1}^{\text{re}} y_t^{\text{re}}\right),$$

$$f_{Y_{\text{im}}}(\mathbf{y}_{\text{im}}) = \frac{1}{\pi\sqrt{2\text{SNR}+1}} \exp\left(-\frac{c_1}{2} \|\mathbf{y}_{\text{im}}\|_2^2 - c_2 y_{t-1}^{\text{im}} y_t^{\text{im}}\right),$$

where $c_1 = 2\frac{\text{SNR}+1}{2\text{SNR}+1}$ and $c_2 = -\frac{2\text{SNR}}{2\text{SNR}+1}$. Since the limits of all the variables of the above expressions are similar, we can interchange the argument of $f_{Y_{\text{re}}}(\mathbf{y}_{\text{re}})$ with $\mathbf{y}_{\text{im}}$ and produce the same expression. Therefore, $f_{Y_{\text{re}}}(\mathbf{y}_{\text{re}}) = f_{Y_{\text{im}}}(\mathbf{y}_{\text{im}})$ and hence $f_Y(\mathbf{y}) = (f_{Y_{\text{re}}}(\mathbf{y}_{\text{re}}))^2$.

By doing some mathematical manipulations, we can re-express $f_{Y_{\text{re}}}(\mathbf{y}_{\text{re}})$ as

$$f_{Y_{\text{re}}}(\mathbf{y}_{\text{re}}) = \frac{1}{2\pi\sqrt{|\mathbf{\Sigma}|}} \exp\left(-\frac{1}{2} \mathbf{y}_{\text{re}}^T \mathbf{\Sigma}^{-1} \mathbf{y}_{\text{re}}\right),$$

where $\mathbf{\Sigma}^{-1} = \begin{bmatrix} c_1 & c_2 \\ c_2 & c_1 \end{bmatrix}$. Therefore, we can identify $f_{Y_{\text{re}}}(\mathbf{y}_{\text{re}})$ as bivariate Gaussian distribution with zero mean and covariance matrix $\mathbf{\Sigma}$.

Therefore, we have

$$p(\text{SNR}) = 1 - \left(2 \int_0^\infty \int_0^\infty f_{Y_{\text{re}}}(\mathbf{y}_{\text{re}}) dy_{t-1}^{\text{re}} dy_t^{\text{re}}\right)^2.$$

By defining $p'(\text{SNR}) = 2 \int_0^\infty \int_0^\infty f_{Y_{\text{re}}}(\mathbf{y}_{\text{re}}) dy_{t-1}^{\text{re}} dy_t^{\text{re}}$, we then focus on finding the decay rate of $p'(\text{SNR})$. To that end, we use the transformation $\begin{bmatrix} u \\ v \end{bmatrix} = \mathbf{\Sigma}^{-\frac{1}{2}} \begin{bmatrix} y_{t-1}^{\text{re}} \\ y_t^{\text{re}} \end{bmatrix}$ to find the asymptotic behaviour of $p'(\text{SNR})$ as $\text{SNR} \rightarrow \infty$.

We then have,

$$\begin{aligned}
\lim_{\text{SNR} \rightarrow \infty} p'(\text{SNR}) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(u^2 + v^2)\right) du dv \\
&\quad - \frac{1}{\pi} \int_{-\infty}^0 \int_{-v}^{\left(-1 - \frac{\sqrt{2}}{\sqrt{\text{SNR}}}\right)v} \exp\left(-\frac{1}{2}(u^2 + v^2)\right) du dv \\
&\quad - \frac{1}{\pi} \int_0^{\infty} \int_{\left(-1 + \frac{\sqrt{2}}{\sqrt{\text{SNR}}}\right)v}^{-v} \exp\left(-\frac{1}{2}(u^2 + v^2)\right) du dv, \\
&\stackrel{(a)}{=} 1 - \frac{2}{\pi} \left(\tan^{-1} \left(-1 + \frac{\sqrt{2}}{\sqrt{\text{SNR}}} \right) + \frac{\pi}{4} \right) \\
&\stackrel{(b)}{=} 1 - \frac{2}{\pi} \left(-\frac{\pi}{4} + \frac{\sqrt{2}}{2\sqrt{\text{SNR}}} + O(\text{SNR}^{-1}) + \frac{\pi}{4} \right) \\
&= 1 - \frac{\sqrt{2}}{\pi\sqrt{\text{SNR}}} + O(\text{SNR}^{-1}), \tag{3.34}
\end{aligned}$$

where (a) follows by converting the following integral to polar coordinates as

$$\begin{aligned}
\int_0^{\infty} \int_{\left(-1 + \frac{\sqrt{2}}{\sqrt{\text{SNR}}}\right)v}^{-v} \exp\left(-\frac{1}{2}(u^2 + v^2)\right) du dv &= \\
&= - \int_0^{\infty} \int_{\tan^{-1}\left(-1 + \frac{\sqrt{2}}{\sqrt{\text{SNR}}}\right)}^{-\frac{\pi}{4}} \lambda \exp\left(-\frac{\lambda^2}{2}\right) d\lambda d\beta = \tan^{-1} \left(-1 + \frac{\sqrt{2}}{\sqrt{\text{SNR}}} \right) + \frac{\pi}{4},
\end{aligned}$$

in which $\lambda^2 = u^2 + v^2$ and $\beta = \tan^{-1} \left(\frac{u}{v} \right)$ [84]. With similar steps, we can prove that

$$\int_{-\infty}^0 \int_{-v}^{\left(-1 - \frac{\sqrt{2}}{\sqrt{\text{SNR}}}\right)v} \exp\left(-\frac{1}{2}(u^2 + v^2)\right) du dv = \tan^{-1} \left(-1 + \frac{\sqrt{2}}{\sqrt{\text{SNR}}} \right) + \frac{\pi}{4}.$$

Then, (b) follows from the Taylor series expansion and hence, when $\text{SNR} \rightarrow \infty$, we have

$$p(\text{SNR}) = \frac{2\sqrt{2}}{\pi} \text{SNR}^{-\frac{1}{2}} + O(\text{SNR}^{-1}). \tag{3.35}$$

Therefore, $p(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-\frac{1}{2}}$, and hence the DVO = $\frac{1}{2}$ for $n = 2$ and 3 .

Method 2:

We can prove the DVO result, starting with the SEP expression given in Corollary

3.1 as well. To that end, we can re-express the conditional pdf given in (3.33) as

$$p(\text{SNR}, h) = 4\mathcal{Q}\left(-\sqrt{2\text{SNR}}r \cos \theta\right) \mathcal{Q}\left(-\sqrt{2\text{SNR}}r \sin \theta\right) \bar{p}(\text{SNR}, h),$$

where

$$\begin{aligned} \bar{p}(\text{SNR}, h) = & \mathcal{Q}\left(\sqrt{2\text{SNR}}r \cos \theta\right) \mathcal{Q}\left(-\sqrt{2\text{SNR}}r \sin \theta\right) + \\ & \mathcal{Q}\left(\sqrt{2\text{SNR}}r \cos \theta\right) \mathcal{Q}\left(\sqrt{2\text{SNR}}r \sin \theta\right) + \mathcal{Q}\left(-\sqrt{2\text{SNR}}r \cos \theta\right) \mathcal{Q}\left(\sqrt{2\text{SNR}}r \sin \theta\right). \end{aligned}$$

By using the fact that $\mathcal{Q}(-a) = 1 - \mathcal{Q}(a)$, we can expand the above expression as

$$p(\text{SNR}, h) = 4\{1 - \Lambda(\text{SNR}, h)\} \Lambda(\text{SNR}, h),$$

where

$$\begin{aligned} \Lambda(\text{SNR}, h) = & \mathcal{Q}\left(-\sqrt{2\text{SNR}}r \cos \theta\right) + \mathcal{Q}\left(-\sqrt{2\text{SNR}}r \sin \theta\right) \\ & - \mathcal{Q}\left(-\sqrt{2\text{SNR}}r \cos \theta\right) \mathcal{Q}\left(-\sqrt{2\text{SNR}}r \sin \theta\right). \end{aligned}$$

Therefore, we can evaluate the decay rate of each term of the integral expression as

$$\begin{aligned} p(\text{SNR}) &= \frac{1}{\pi} \int_0^{2\pi} \int_0^\infty p(\text{SNR}, h) r e^{-r^2} dr d\theta, \\ &= \frac{4}{\pi} \int_0^{2\pi} \int_0^\infty \left\{ \mathcal{Q}\left(\sqrt{2\text{SNR}}r \cos \theta\right) + \mathcal{Q}\left(\sqrt{2\text{SNR}}r \sin \theta\right) - \right. \\ &\quad \left. \mathcal{Q}\left(\sqrt{2\text{SNR}}r \cos \theta\right) \mathcal{Q}\left(\sqrt{2\text{SNR}}r \sin \theta\right) - \Lambda^2(\text{SNR}, h) \right\} r e^{-r^2} dr d\theta, \end{aligned}$$

where $\Lambda^2(\text{SNR}, h)$ contains the higher order terms of $\mathcal{Q}(\cdot)$ which we can omit. Then,

by using Lemmas 3 and 4 in [92], we can conclude that

$$\begin{aligned} \int_0^{2\pi} \int_0^\infty \mathcal{Q}(\sqrt{2\text{SNR}}r \cos \theta) r e^{-r^2} dr d\theta &\stackrel{e}{=} \text{SNR}^{-\frac{1}{2}}, \\ \int_0^{2\pi} \int_0^\infty \mathcal{Q}(\sqrt{2\text{SNR}}r \sin \theta) r e^{-r^2} dr d\theta &\stackrel{e}{=} \text{SNR}^{-\frac{1}{2}}, \\ \int_0^{2\pi} \int_0^\infty \mathcal{Q}(\sqrt{2\text{SNR}}r \cos \theta) \mathcal{Q}(\sqrt{2\text{SNR}}r \sin \theta) r e^{-r^2} dr d\theta &\stackrel{e}{=} \text{SNR}^{-1}. \end{aligned}$$

With a similar analysis, we can find the decay rate of the rest of the terms and show that they are decaying to zero at a rate greater than $\text{SNR}^{-\frac{1}{2}}$. Hence, we conclude that $p(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-\frac{1}{2}}$ for $n = 2$ and 3.

Now we will move to the proof for $n \geq 4$ case. In this case, we write

$$\begin{aligned} p(\text{SNR}) &= \frac{2^n}{\pi^2 (2 + \frac{1}{\text{SNR}})} \text{SNR}^{-1} \int_{-\infty}^\infty \int_0^\infty \int_0^\infty \int_{y_{t-1}^{\text{re}} \tan(\frac{\pi}{4} - \psi)}^{y_{t-1}^{\text{re}} \tan(\frac{\pi}{4} + \frac{\pi}{2^{n-1}} - \psi)} \\ &\quad g_{\text{SNR}}(\mathbf{y}) dy_{t-1}^{\text{im}} dy_{t-1}^{\text{re}} dy_t^{\text{re}} dy_t^{\text{im}} \\ &+ \frac{2^n}{\pi^2 (2 + \frac{1}{\text{SNR}})} \text{SNR}^{-1} \int_0^\infty \int_0^\infty \int_0^\infty \int_{y_{t-1}^{\text{re}} \tan(\frac{\pi}{4} - \psi)}^{y_{t-1}^{\text{re}} \tan(\frac{\pi}{4} + \frac{\pi}{2^{n-1}} - \psi)} \\ &\quad g_{\text{SNR}}(\mathbf{y}) dy_{t-1}^{\text{im}} dy_{t-1}^{\text{re}} dy_t^{\text{re}} dy_t^{\text{im}}, \end{aligned}$$

where $g_{\text{SNR}}(\mathbf{y}) = \exp \left\{ -\frac{1 + \frac{1}{\text{SNR}}}{2 + \frac{1}{\text{SNR}}} \|\mathbf{y}\|_2^2 + \frac{2}{2 + \frac{1}{\text{SNR}}} (y_{t-1}^{\text{re}} y_t^{\text{re}} + y_{t-1}^{\text{im}} y_t^{\text{im}}) \right\}$ is a collection of functions indexed by SNR. As $\text{SNR} \rightarrow \infty$ the function g_{SNR} goes to

$$g_\infty(\mathbf{y}) = \exp \left\{ -\frac{1}{2} \|\mathbf{y}\|_2^2 + (y_{t-1}^{\text{re}} y_t^{\text{re}} + y_{t-1}^{\text{im}} y_t^{\text{im}}) \right\}.$$

Further,

$$\int_{-\infty}^\infty \int_0^\infty \int_0^\infty \int_{y_{t-1}^{\text{re}} \tan(\frac{\pi}{4} - \psi)}^{y_{t-1}^{\text{re}} \tan(\frac{\pi}{4} + \frac{\pi}{2^{n-1}} - \psi)} g_\infty(\mathbf{y}) dy_{t-1}^{\text{im}} dy_{t-1}^{\text{re}} dy_t^{\text{re}} dy_t^{\text{im}} < \infty$$

and

$$\int_0^\infty \int_0^\infty \int_0^\infty \int_{y_{t-1}^{\text{re}} \tan(\frac{\pi}{4} - \psi)}^{y_{t-1}^{\text{re}} \tan(\frac{\pi}{4} + \frac{\pi}{2^{n-1}} - \psi)} g_\infty(\mathbf{y}) dy_{t-1}^{\text{im}} dy_{t-1}^{\text{re}} dy_t^{\text{re}} dy_t^{\text{im}} < \infty.$$

Hence, as $\text{SNR} \rightarrow \infty$, we conclude that

$$p(\text{SNR}) = \text{SNR}^{-1} \Theta(1) \quad (3.36)$$

by using the monotone convergence theorem [86], which implies $p(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-1}$. Therefore, the DVO = 1 for $n \geq 4$.

3.9.5 Proof of Theorem 3.4

In order to obtain the ML rule for the setup described above, we have to maximize the likelihood function $\mathcal{L}(s_{t-1}, s_t, h)$ given by

$$\begin{aligned} \mathcal{L}(s_{t-1}, s_t, h) &= \Pr \{ Y_{t-1} \in \mathcal{R}_{k_{t-1}}, Y_t \in \mathcal{R}_{k_t} \mid S_{t-1} = s_{t-1}, S_t = s_t, H = h \}, \\ &\stackrel{(a)}{=} \Pr \left\{ \sqrt{\text{SNR}} h s_{t-1} + W_{t-1} \in \mathcal{R}_{k_{t-1}} \mid S_{t-1} = s_{t-1}, H = h \right\} \\ &\quad \Pr \left\{ \sqrt{\text{SNR}} h s_t + W_t \in \mathcal{R}_{k_t} \mid S_t = s_t, H = h \right\}, \end{aligned}$$

where (a) follows due to the independence of two received signals given the channel and $s_{t-1}, s_t \in \mathcal{C}$. Therefore, in order to maximize the above likelihood function, we have to maximize each of the two probabilities separately. Let $\hat{s}(k, h)$ be defined as

$$\hat{s}(k, h) = \arg \max_{s \in \mathcal{C}} \Pr \left\{ \sqrt{\text{SNR}} h s + W \in \mathcal{R}_k \mid S = s, H = h \right\}.$$

By solving this optimization problem for each probability in $\mathcal{L}(\hat{s}_{t-1}, \hat{s}_t, h)$, we can maximize the likelihood function.

The structure of the above optimization problem is exactly in the same form studied in [92], and using Theorem 2 in this paper, we can write

$$\hat{s}(k, h) = \arg \min_{s \in \mathcal{C}} \text{dist} \left(\sqrt{\text{SNR}} h s, \mathcal{H}_k \right),$$

where $\mathcal{H}_k$ is defined as given in Theorem 3.4.

After estimating $\hat{s}_{t-1}$ and $\hat{s}_t$, the estimated message at time t is completely determined by reversing the encoding rule given in (3.1). This is because the transmitted

message depends only on the phase difference of $\hat{s}_{t-1}$ and $\hat{s}_t$. Therefore, the ML estimator of the message at time t is given by

$$\hat{M}(k_{t-1}, k_t, h) = m_\lambda \quad \text{if } \hat{s}_t = \hat{s}_{t-1} \exp\left(\lambda \frac{j\pi}{2}\right), \quad (3.37)$$

which concludes the proof.

3.9.6 Proof of Theorem 3.5

Without loss of generality, we will assume that $M_t = m_0$ is the message at time t . That is, $S_t = S_{t-1}$. Then, we can write the SEP

$$p(\text{SNR}) = \Pr\{\hat{S}_t \neq \hat{S}_{t-1} \mid S_{t-1} = S_t = s\} \quad (3.38)$$

Here, we first present the proof for the $n = 2$ case. To that end, with a slight abuse of notation, we write the SEP conditioned on fading as

$$\begin{aligned} p(\text{SNR}, h) &= 1 - \sum_{k=0}^3 \Pr\{Y_{t-1} \in \mathcal{R}_k, Y_t \in \mathcal{R}_k \mid S_{t-1} = S_t = s, H = h\}, \\ &= 1 - \sum_{k=0}^3 \Pr\left\{\sqrt{\text{SNR}}re^{j\theta} + W_{t-1} \in \mathcal{R}_k \mid S_{t-1} = s, H = h\right\} \\ &\quad \Pr\left\{\sqrt{\text{SNR}}re^{j\theta} + W_t \in \mathcal{R}_k \mid S_t = s, H = h\right\}, \\ &= 1 - Q^2\left(-\sqrt{2\text{SNR}}r \cos \theta\right) Q^2\left(-\sqrt{2\text{SNR}}r \sin \theta\right) \\ &\quad - Q^2\left(\sqrt{2\text{SNR}}r \cos \theta\right) Q^2\left(-\sqrt{2\text{SNR}}r \sin \theta\right) \\ &\quad - Q^2\left(\sqrt{2\text{SNR}}r \cos \theta\right) Q^2\left(\sqrt{2\text{SNR}}r \sin \theta\right) \\ &\quad - Q^2\left(-\sqrt{2\text{SNR}}r \cos \theta\right) Q^2\left(\sqrt{2\text{SNR}}r \sin \theta\right). \end{aligned} \quad (3.39)$$

By averaging (3.39) with respect to the fading distribution, we get the SEP expression given in Theorem 3.4 for $n = 2$.

We now move to the proof for $n \geq 3$ case. By using Lemma 1 and its proof, we

know that when $\theta \in \left[\frac{\pi}{4} - \frac{\pi}{2^n}, \frac{\pi}{4} + \frac{\pi}{2^n}\right]$, we have the same conditional SEP as given in (3.39). Hence, by averaging with respect to the fading distribution, we get the general SEP expression for $n \geq 2$ as in (3.15). We avoid giving these steps to avoid repetitions.

3.9.7 Proof of Corollary 3.2

Under coherent detection, in order to get the SEP expression (3.15), we have

$$\begin{aligned} p(\text{SNR}) &= 1 - \frac{4}{\pi} \int_0^{\frac{\pi}{2}} \int_0^\infty \sum_{k=0}^3 \Pr\left\{\sqrt{\text{SNR}}re^{j\theta} + W_{t-1} \in \mathcal{R}_k \mid S_{t-1} = s, H = h\right\} \\ &\quad \Pr\left\{\sqrt{\text{SNR}}re^{j\theta} + W_t \in \mathcal{R}_k \mid S_t = s, H = h\right\} re^{-r^2} dr d\theta, \\ &= 1 - \frac{4}{\pi} \int_0^{\frac{\pi}{2}} \int_0^\infty \{\Lambda_1 + \Lambda_2 + \Lambda_3 + \Lambda_4\} re^{-r^2} dr d\theta, \end{aligned}$$

where

$$\begin{aligned} \Lambda_1 &= Q^2\left(-\sqrt{2\text{SNR}}r \cos \theta\right) Q^2\left(-\sqrt{2\text{SNR}}r \sin \theta\right), \\ \Lambda_2 &= Q^2\left(\sqrt{2\text{SNR}}r \cos \theta\right) Q^2\left(-\sqrt{2\text{SNR}}r \sin \theta\right), \\ \Lambda_3 &= Q^2\left(\sqrt{2\text{SNR}}r \cos \theta\right) Q^2\left(\sqrt{2\text{SNR}}r \sin \theta\right), \\ \Lambda_4 &= Q^2\left(-\sqrt{2\text{SNR}}r \cos \theta\right) Q^2\left(\sqrt{2\text{SNR}}r \sin \theta\right). \end{aligned}$$

By using the change of variables $\bar{\theta} = \theta + k\frac{\pi}{2}$ for each integral term $\int_0^{\frac{\pi}{2}} \int_0^\infty \Lambda_k re^{-r^2} dr d\theta$, we get $\int_0^{\frac{\pi}{2}} \int_0^\infty \Lambda_k re^{-r^2} dr d\theta = \int_{k\frac{\pi}{2}}^{k\frac{\pi}{2} + \frac{\pi}{2}} \int_0^\infty \Lambda_1 re^{-r^2} dr d\theta$ and hence we can convert the $p(\text{SNR})$ expression to

$$\begin{aligned} p(\text{SNR}) &= 1 - \frac{4}{\pi} \int_0^{\frac{\pi}{2}} \int_0^\infty \Lambda_1 re^{-r^2} dr d\theta - \frac{4}{\pi} \int_{\frac{\pi}{2}}^\pi \int_0^\infty \Lambda_1 re^{-r^2} dr d\theta \\ &\quad - \frac{4}{\pi} \int_\pi^{\frac{3\pi}{2}} \int_0^\infty \Lambda_1 re^{-r^2} dr d\theta - \frac{4}{\pi} \int_{\frac{3\pi}{2}}^{2\pi} \int_0^\infty \Lambda_1 re^{-r^2} dr d\theta, \\ &= 1 - \frac{4}{\pi} \int_0^{2\pi} \int_0^\infty \Lambda_1 re^{-r^2} dr d\theta. \end{aligned}$$

Therefore we get the SEP expression in (3.16) and it is the same as (3.10) which give

the SEP under non-coherent detection. Therefore, we conclude that for $n = 2$, the SEP is same for both coherent and non-coherent detection.

3.9.8 Proof of Theorem 3.7

In this case, we will first re-express the SEP given (3.15) as

$$p(\text{SNR}) = \frac{2^n}{\pi} \int_{\frac{\pi}{4}-\frac{\pi}{2^n}}^{\frac{\pi}{4}+\frac{\pi}{2^n}} \int_0^\infty \sum_{i=1}^4 \mathcal{Q}(\rho_i^{\text{re}} \sqrt{2\text{SNR}} r \cos \theta) \mathcal{Q}(\rho_i^{\text{im}} \sqrt{2\text{SNR}} r \sin \theta) p_i(\text{SNR}, h) r \exp(-r^2) dr d\theta$$

where $\rho_i^{\text{re}} = (-1)^i$ for $i = 1, 2, 3, 4$, $\rho_i^{\text{im}} = -1$ for $i = 1, 2$ and $\rho_i^{\text{im}} = 1$ for $i = 3, 4$, and

$$\begin{aligned} p_1(\text{SNR}, h) &= \mathcal{Q}(z^{\text{re}}) \mathcal{Q}(-z^{\text{im}}) + \mathcal{Q}(z^{\text{re}}) \mathcal{Q}(z^{\text{im}}) + \mathcal{Q}(-z^{\text{re}}) \mathcal{Q}(z^{\text{im}}) \\ p_2(\text{SNR}, h) &= \mathcal{Q}(-z^{\text{re}}) \mathcal{Q}(-z^{\text{im}}) + \mathcal{Q}(z^{\text{re}}) \mathcal{Q}(z^{\text{im}}) + \mathcal{Q}(-z^{\text{re}}) \mathcal{Q}(z^{\text{im}}) \\ p_3(\text{SNR}, h) &= \mathcal{Q}(-z^{\text{re}}) \mathcal{Q}(-z^{\text{im}}) + \mathcal{Q}(z^{\text{re}}) \mathcal{Q}(-z^{\text{im}}) + \mathcal{Q}(-z^{\text{re}}) \mathcal{Q}(z^{\text{im}}) \\ p_4(\text{SNR}, h) &= \mathcal{Q}(-z^{\text{re}}) \mathcal{Q}(-z^{\text{im}}) + \mathcal{Q}(z^{\text{re}}) \mathcal{Q}(-z^{\text{im}}) + \mathcal{Q}(z^{\text{re}}) \mathcal{Q}(z^{\text{im}}), \end{aligned}$$

in which $z^{\text{re}} = \sqrt{2\text{SNR}} r \cos \theta$ and $z^{\text{im}} = \sqrt{2\text{SNR}} r \sin \theta$.

By using the fact that $\mathcal{Q}(-a) = 1 - \mathcal{Q}(a)$, we can expand the above expression and evaluate the decay rate of each term of the integral expression as

$$p(\text{SNR}) = \frac{2^n}{\pi} \int_{\frac{\pi}{4}-\frac{\pi}{2^n}}^{\frac{\pi}{4}+\frac{\pi}{2^n}} \int_0^\infty \left\{ c_1 \mathcal{Q}(\sqrt{2\text{SNR}} r \cos \theta) + c_2 \mathcal{Q}(\sqrt{2\text{SNR}} r \sin \theta) - c_3 \mathcal{Q}(\sqrt{2\text{SNR}} r \cos \theta) \mathcal{Q}(\sqrt{2\text{SNR}} r \sin \theta) + \dots \right\} r e^{-r^2} dr d\theta,$$

where $c_1, c_2, c_3, \dots$ are constants and we omit the higher order terms of $\mathcal{Q}(\cdot)$. Then, by using Lemmas 3 and 4 in [92], we can conclude that

$$\int_{\frac{\pi}{4}-\frac{\pi}{2^n}}^{\frac{\pi}{4}+\frac{\pi}{2^n}} \int_0^\infty \mathcal{Q}(\sqrt{2\text{SNR}} r \cos \theta) r e^{-r^2} dr d\theta \stackrel{e}{=} \begin{cases} \text{SNR}^{-\frac{1}{2}} & \text{if } n = 2 \\ \text{SNR}^{-1} & \text{if } n \geq 3 \end{cases},$$

$$\int_{\frac{\pi}{4}-\frac{\pi}{2^n}}^{\frac{\pi}{4}+\frac{\pi}{2^n}} \int_0^\infty \mathcal{Q}(\sqrt{2\text{SNR}}r \sin \theta) r e^{-r^2} dr d\theta \stackrel{e}{=} \begin{cases} \text{SNR}^{-\frac{1}{2}} & \text{if } n = 2 \\ \text{SNR}^{-1} & \text{if } n \geq 3 \end{cases},$$

and

$$\int_{\frac{\pi}{4}-\frac{\pi}{2^n}}^{\frac{\pi}{4}+\frac{\pi}{2^n}} \int_0^\infty \mathcal{Q}(\sqrt{2\text{SNR}}r \cos \theta) \mathcal{Q}(\sqrt{2\text{SNR}}r \sin \theta) r e^{-r^2} dr d\theta \stackrel{e}{=} \text{SNR}^{-1} \quad \text{if } n \geq 2.$$

With a similar analysis, we can find the decay rate of the rest of the terms and can show that they are decaying to zero at a rate greater than SNR^{-1} . Hence, we conclude that $p(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-\frac{1}{2}}$ for $n = 2$ and $p(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-1}$ for $n \geq 3$. Therefore the DVO result in Theorem 3.7.

Chapter 4

Low-Resolution Quantization in Multi-Antenna Systems

This chapter focuses on optimum and sub-optimum detectors and error rate analysis for wireless systems with low-resolution quantizers in the presence of fading and noise for a wireless communication network consisting of one transmit antenna and N receive antennas. In the case of QPSK, the ML detector is derived. Based on selection combining, we present three antenna selection strategies that pick the best antennas based on the channels and the quantization regions in which the received signals land. We derive exact analytical expressions for the average SEP attained by each antenna selection strategy when the wireless channel is corrupted by Rayleigh fading and AWGN. For QPSK modulation, it is shown that a transceiver architecture with n -bit quantization and N receiver antennas is asymptotically optimum in terms of communication reliability if $n \geq 3$. That is, with the optimum ML detector, the decay exponent for the average SEP is the same and equals to N either with infinite-bit quantizers or with n -bit quantizers where $n \geq 3$. On the other hand, it is only equal to $\frac{N}{2}$ and 0 for $n = 2$ and $n < 2$, respectively. We show that the strategy that selects the diversity branch with channel rotated constellation points being furthest away from the decision boundary is the best strategy in terms of the symbol error probability performance. An extensive simulation study is performed to illustrate the accuracy of the derived results.

4.1 Introduction

4.1.1 Background and Motivation

In this chapter, we extend the work presented in Chapter 2 to low-resolution quantization based communication systems with multiple antennas. In Chapter 2, we analyzed the error probability performance of a SISO communication system when the receiver

is equipped with a low-resolution phase quantizer. Considering that a SISO system acts as a basic building block for more complex systems, in this chapter, we focus on analyzing the error probability performance of a SIMO system with low-resolution quantization. The performance results obtained in this chapter hold in the general case of n -bit quantization, QPSK modulation and Rayleigh fading channels.

4.1.2 Main Contributions

In this chapter, we consider a wireless communication system with 1 transmit antenna and N receive antennas, in which the receiver is equipped with low-resolution ADCs that quantizes the phase of the received signal. Our main contributions are summarized as follows.

- Motivated by the capacity achieving property of circularly symmetric input distributions for low-resolution ADCs [77], we adopt QPSK modulation and derive the optimum ML detection rule for signal detection with low-resolution ADCs in a SIMO system. For the general n -bit quantization, we analyze the average SEP attained by the derived ML rule when the wireless channel is subjected to Rayleigh fading through Monte-Carlo simulations. These results are presented in Theorems 4.1 and 4.2.
- Based on selection combining, we present three antenna selection strategies that pick the best antennas based on the channels and the quantization regions in which the received signals land. These strategies are presented in Sections 4.4.1, 4.4.2 and 4.4.3. We derive exact analytical expressions for the average SEP attained by each antenna selection strategy when the wireless channel is corrupted by Rayleigh fading and AWGN. We use extensive numerical examples to compare the analytical expressions against the Monte-Carlo simulations. These results are presented in Theorems 4.3, 4.4 and 4.5, respectively.
- Under the selection strategy given in Section 4.4.1, we establish a fundamental ternary behaviour in the average SEP performance of a wireless communication system with low-resolution ADCs and QPSK modulation. Based on the Rayleigh

fading model, we show that the decay exponent of the average SEP is the same as that of an infinite-bit quantization, which is equal to N when $n \geq 3$. We also show that it is equal to $\frac{1}{2}$ and 0 for 2 and $n < 2$ bit quantization, respectively. This result is presented in Theorem 4.6.

- Under the selection strategy given in Section 4.4.2, we establish a binary behaviour in the average SEP performance of a wireless communication system with low-resolution ADCs and QPSK modulation. Based on the Rayleigh fading model, we show that the decay exponent of the average SEP is 1 for $N \geq 2$ and $n \geq 2$. We also show that it is equal to 0 for $n < 2$ bit quantization. This result is presented in Theorem 4.7.
- Under the selection strategy given in Section 4.4.3, we establish a fundamental ternary behaviour in the average SEP performance of a wireless communication system with low-resolution ADCs and QPSK modulation. Based on the Rayleigh fading model, we show that the decay exponent of the average SEP is the same as that of an infinite-bit quantization, which is equal to N when n is larger than or equal to 3 using numerical results. We also show that it is equal to $\frac{N}{2}$ for 2-bit quantization.
- By analysing the error probability performance in the high-SNR regime, we characterize the DVO for low-resolution ADC based communication systems. From a system design point of view, our results show that the best sub-optimum detection rule is the one given in Section 4.4.3. The results show that using one additional bit on top of 2-bit quantization can achieve optimum communication robustness in the high-SNR regime. In particular, for Rayleigh fading, the simple sub-optimum detection rule given in Section 4.4.3 is optimum in terms of the DVO.

4.2 System Setup

4.2.1 Channel Model and Signal Modulation

We consider a wireless communication network consisting of a transmitter with one antenna and a receiver with N antennas. The received signals at the receiver antenna array can be given in vector form as

$$\mathbf{Y} = \sqrt{\text{SNR}}\mathbf{H}X + \mathbf{W}, \quad (4.1)$$

where $X \in \mathcal{C}$ is the transmitted signal with $\mathcal{C}$ denoting the constellation set of information signals in $\mathbb{C}$, SNR is the ratio of the transmitted signal power to the AWGN power, $\mathbf{H} = [H_0 \ H_1 \ \cdots \ H_{N-1}]^T$ with $H_l \in \mathbb{C}$ is the unit power channel gain between the transmitter and the l -th receiver antenna, and $\mathbf{W} = [W_0 \ W_1 \ \cdots \ W_{N-1}]^T$ with W_l is the circularly-symmetric zero-mean unit-variance AWGN, i.e., $W_l \sim \mathcal{CN}(0, 1)$ at the l -th receiver antenna. Then $\mathbf{Y} = [Y_0 \ Y_1 \ \cdots \ Y_{N-1}]^T$ with

$$Y_l = \sqrt{\text{SNR}}H_lX + W_l, \quad l = 0, 1, \dots, N-1 \quad (4.2)$$

where Y_l represents the received signal at the receiver antenna l . Assuming that the transmitted symbols are modulated using QPSK, we set $\mathcal{C} = \left\{ e^{j\pi(\frac{2i+1}{4}-1)} \right\}_{i=0}^3$. While the analysis presented in this chapter is restricted to QPSK modulation it can be extended to M -PSK modulation.

Under high-resolution quantization where the receiver can fully observe $\mathbf{Y} = \mathbf{y}$ with perfect channel knowledge $\mathbf{H} = \mathbf{h}$, the symbol detected by the optimum ML detector can be given as

$$\hat{x}_{\text{high}}(\mathbf{y}, \mathbf{h}) = \arg \min_{x \in \mathcal{C}} \left\| \mathbf{y} - \sqrt{\text{SNR}}\mathbf{h}x \right\|_2^2, \quad (4.3)$$

where $\|(\cdot)\|_2$ represents the Euclidean norm of the vector $(\cdot)$ [93]. In our work, however, we are focusing on the case where the receiver quantizes each incoming signal and conveys the quantized received signal vector $Q(\mathbf{Y}) = [Q(Y_0), Q(Y_1), \dots, Q(Y_{N-1})]^T$

to the detector. By using these quantized observations, the detector has to estimate the transmitted signal X .

4.2.2 Receiver Architecture

The receiver architecture is based on a low-resolution ADC. As illustrated in Fig. 4.1, each received signal is sent through a low-resolution quantizer, and then the resulting quantized received signals are used to determine the transmitted symbol X . More specifically, the signal Y_l received from antenna l goes through an n -bit low-resolution quantizer for all $l = 0, 1, \dots, N-1$. Then the resulting quantized signals $Q(Y_0), Q(Y_1), \dots, Q(Y_{N-1})$ are used to estimate the transmitted signal X . Since information is encoded in the phase of X with the above choice of constellation points, we use the same quantization process as given in Section 2.2.4 to quantize each received signal Y_l .

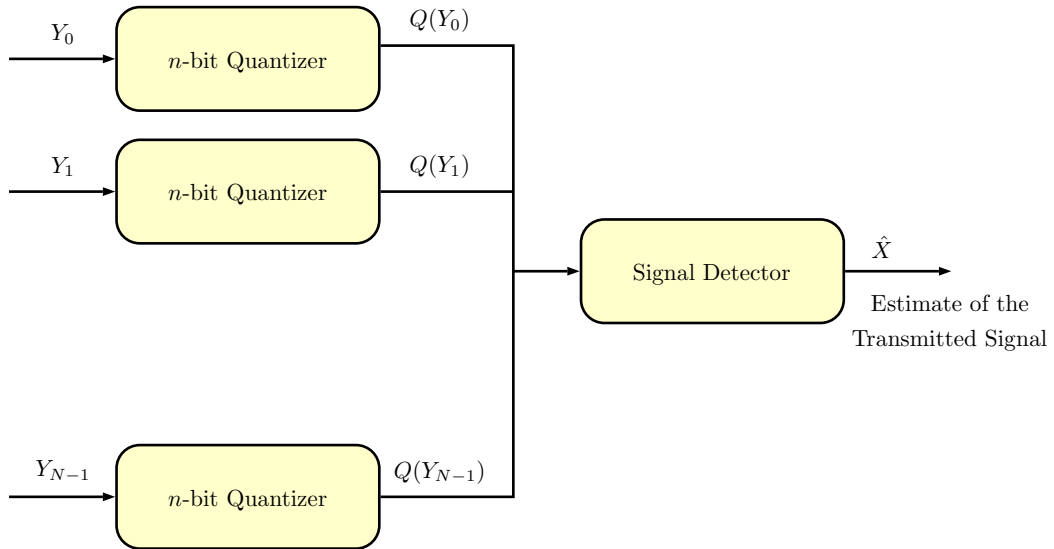


Figure 4.1: The receiver architecture with low-resolution quantization. The signal detector observes only the n -bit quantized versions of Y_l to estimate the transmitted signal. $l = 0, 1, \dots, N-1$

We also assume that full CSI is available at the receiver. The motivation behind our model with full-precision CSI is the same as given in Section 2.2.5.

4.3 Optimum Signal Detection

With slight abuse of notation, we first define $Q(\mathbf{Y}) = [Q(Y_0), Q(Y_1), \dots, Q(Y_{N-1})]^T$ for vector notation of the quantized observations of the received signals. Then, the aim of the detector is to minimize the SEP by using the knowledge of $Q(\mathbf{Y})$ and CSI $\mathbf{H}$, which can be represented as selecting a signal point $\hat{x}(\mathbf{k}, \mathbf{h})$ satisfying

$$\hat{x}(\mathbf{k}, \mathbf{h}) \in \arg \max_{x \in \mathcal{C}} \Pr \{X = x | Q(\mathbf{Y}) = \mathbf{k}, \mathbf{H} = \mathbf{h}\}, \quad (4.4)$$

where $\mathbf{k} = [k_0, k_1, \dots, k_{N-1}]^T$ with $Q(Y_l) = k_l$ and $\mathbf{h} = [h_0, h_1, \dots, h_{N-1}]^T$ with $h_l \in \mathbb{C}$ for all $l \in [0 : N-1]$ and $k_l \in [0 : 2^n - 1]$. The main performance figure of merit for our detection problem is the average SEP given by

$$p(\text{SNR}) = \Pr \{X \neq \hat{x}(Q(\mathbf{Y}), \mathbf{H})\}. \quad (4.5)$$

In the following, we present five detection techniques that can be used in the process of estimating the transmitted symbols in this setting. Then we analyze the SEP performance of them under n -bit quantization. We start with the optimum ML detector and since the complexity of the ML detector increases with the number of receiving antennas, we next focus on detectors based on selection combining and majority combining.

4.3.1 ML Detector with 2-bit Quantization

In the following, we present the ML detection rule for the special case with $n = 2$ which is later generalized to $n \geq 3$.

Theorem 4.1. *Assume H_l has a continuous pdf for all $l = 0, 1, \dots, N-1$ and $\mathcal{Q}(\cdot)$ is the complementary distribution function of the standard normal random variable. Then, $\hat{x}(\mathbf{k}, \mathbf{h})$ is unique with probability one because the set of $\mathbf{h}$ values for which*

$$\arg \max_{x \in \mathcal{C}} \Pr \{Q(\mathbf{Y}) = \mathbf{k} | X = x, \mathbf{H} = \mathbf{h}\}$$

is singleton has probability one. Then, the ML detection rule for low-resolution ADC based receiver architecture with 2-bit quantization can be derived as

$$\hat{x}(\mathbf{k}, \mathbf{h}) = \arg \max_{x \in \mathcal{C}} \prod_{l=0}^{N-1} \mathcal{Q} \left(-\text{sgn}(\text{Re}(Y_l)) \sqrt{\text{SNR}} r_l \cos \{\text{Arg}(h_l) + \text{Arg}(x)\} \right) \mathcal{Q} \left(-\text{sgn}(\text{Im}(Y_l)) \sqrt{\text{SNR}} r_l \sin \{\text{Arg}(h_l) + \text{Arg}(x)\} \right), \quad (4.6)$$

where $r_l = |h_l|$ is the magnitude of h_l , $\text{Arg}(h_l)$ and $\text{Arg}(x)$ are the phase angles of h_l and x respectively and $\text{sgn}(\cdot)$ is signum function given as (1.2).

Proof: See Appendix 4.9.1. □

4.3.2 ML Receiver with n -bit Quantization

In the following, we generalize the ML detection rule for $n \geq 3$ and present the following theorem.

Theorem 4.2. Assume H_l has a continuous PDF for all $l = 0, 1, \dots, N-1$ and $\mathcal{Q}(\cdot)$ is the complementary distribution function of the standard normal random variable. Then, $\hat{x}(\mathbf{k}, \mathbf{h})$ is unique with probability one and the ML detection rule for low-resolution ADC based receiver architecture with $n \geq 3$ bit quantization can be derived as

$$\hat{x}(\mathbf{k}, \mathbf{h}) = \arg \max_{x \in \mathcal{C}} \prod_{l=0}^{N-1} \left(\frac{1}{2^n} \exp(-\text{SNR} r_l^2) + g(\text{Arg}(x), k_l, h_l) \right), \quad (4.7)$$

where

$$g(\text{Arg}(x), k_l, h_l) = \sqrt{\frac{\text{SNR}}{\pi}} r_l \int_{k_l \frac{\pi}{2^{n-1}}}^{(k_l+1) \frac{\pi}{2^{n-1}}} \left[\exp \{ -\text{SNR} r_l^2 \sin^2(\alpha - \text{Arg}(h_l) - \text{Arg}(x)) \} \cos(\alpha - \text{Arg}(h_l) - \text{Arg}(x)) \mathcal{Q} \left(-\sqrt{2\text{SNR}} r_l \cos(\alpha - \text{Arg}(h_l) - \text{Arg}(x)) \right) \right] d\alpha,$$

where $r_l = |h_l|$ is the magnitude of h_l , $\text{Arg}(h_l)$ is the phase angles of h_l and $k_l = Q(Y_l)$.

Proof: See Appendix 4.9.2. □

4.4 Antenna Selection and Signal Detection

We can see that evaluating the integral expressions in the ML detector given in Theorem 4.7 is not easy. As we increase N , the number of integral expressions in the likelihood function also increases. Therefore, in the following, we focus on sub-optimum detectors based on the selection combining and derive easy-to-evaluate expressions for the error performance. Therefore, we next move to sub-optimum detectors based on selection combining to check whether there are ways to reduce the complexity.

In selection combining, we pick a path L^* according to a given selection criterion which can be defined as follows.

Definition 4.1. A selection criterion s is defined as a mapping

$$s : \mathbb{C}^N \rightarrow [0 : N - 1], \quad (4.8)$$

which takes N fading coefficients as input and outputs the index of the selected path. That is, $s(\mathbf{H}) = L^*$ where $\mathbf{H} = [H_0 \ H_1 \ \cdots \ H_{N-1}]^T$ with $H_l \in \mathbb{C}$ for all $l \in [0 : N - 1]$.

Once the path L^* is selected, the detector only uses the information available at antenna L^* to estimate the transmitted signal. Thus, the detector has to minimize the SEP using the knowledge of $Q(Y_{L^*})$ and H_{L^*} , which can be represented as selecting a signal point $\hat{x}(k^*, h^*)$ satisfying

$$\hat{x}(k^*, h^*) \in \arg \max_{x \in \mathcal{C}} \Pr \{X = x | Q(Y_{L^*}) = k^*, H_{L^*} = h^*\},$$

for all $h^* \in \mathbb{C}$ and $k^* \in [0 : 2^n - 1]$. As such, the SIMO system considered in this chapter reduces to a point-to-point system once a path is selected. Therefore, we can directly use Theorem 2.3 of Chapter 2 to express the ML detection rule under a given selection criterion as

$$\hat{x}(k^*, h^*) = \arg \min_{x \in \mathcal{C}} \text{dist} \left(\sqrt{\text{SNR}} h^* x, \mathcal{H}_{k^*} \right), \quad (4.9)$$

where $\text{dist}(z, \mathcal{A}) \triangleq \inf_{s \in \mathcal{A}} |z - s|$ is the distance between a point $z \in \mathbb{C}$ and a set $\mathcal{A} \subseteq \mathbb{C}$,

and $\mathcal{H}_{k^*} = \{z \in \mathbb{C} : \text{Arg}(z) + \pi = \frac{\pi}{2^i} (2k^* + 1)\}$ is the bisector of $\mathcal{R}_{k^*}$ with $Y_{L^*} \in \mathcal{R}_{k^*}$.

In this work, we consider three different criteria to select a path from N independent paths as

1. Selection Criterion 1: Selecting the path having the channel with the maximum amplitude.
2. Selection Criterion 2: Selecting the path having the channel with the minimum absolute phase angle.
3. Selection Criterion 3: Selecting the path having the channel that locates the rotated constellation point furthest away from the decision boundary.

These selection criteria are discussed in detail in the following. We also analyse the SEP performance of each method.

4.4.1 Selection Criterion 1

In this case, the detector chooses the path with the maximum channel magnitude as the path L^* . Thus, the path selection criterion can be given as

$$L^* = \arg \max_{l \in [0:N-1]} R_l, \quad (4.10)$$

where $R_l = |H_l|$ is the channel magnitude of path l . The main idea behind this selection criterion is to choose the path with the channel that locates the rotated constellation signal (i.e., $\sqrt{\text{SNR}} H_l X$) furthest away from the origin of the constellation plane. Hence, it maximizes the *effective* SNR of the system. Albeit its simplicity, this selection criterion does not take into account the phase of the channel when selecting the path L^* . The observations available in the selected path L^* is used in (4.9) to estimate the transmitted symbol $\hat{x}(k^*, h^*)$.

4.4.2 Selection Criterion 2

In this case, the detector chooses the path with the minimum absolute phase angle of the channel as the path L^* . Then the path selection criterion can be given as

$$L^* = \arg \min_{l \in [0:N-1]} |\Theta_l|, \quad (4.11)$$

where $\Theta_l = \text{Arg}(H_l)$ is the phase angle of the channel in path l . The main idea behind this selection technique is to choose the channel which makes the rotated constellation signal (i.e., $\sqrt{\text{SNR}}H_l X$) closest to the bisector $\mathcal{H}_{k_l}$ of the quantization region $\mathcal{R}_{k_l}$ where $Y_l \in \mathcal{R}_{k_l}$. This selection criterion is motivated by Theorem 2.3 in Chapter 2 which indicates that most probability mass is accumulated in the region $\mathcal{R}_{k_l}$ when the unit-variance proper complex Gaussian distribution with mean $\sqrt{\text{SNR}}h_l x$, $x \in \mathcal{C}$, closest to $\mathcal{H}_{k_l}$ is integrated over $\mathcal{R}_{k_l}$. The observations available in the selected path L^* is used in (4.9) to estimate the transmitted symbol $\hat{x}(k^*, h^*)$.

4.4.3 Selection Criterion 3

In this case, the detector chooses the path L^* that locates the rotated constellation point $\sqrt{\text{SNR}}H_l X$ furthest away from the boundary of the region in which the received signal has to be landed for correct decoding of the transmitted signal X . We call this the *region of attraction* and its boundaries the *decision boundaries*. The main idea behind this selection technique is to choose the channel that will minimize the probability of making an error. Let us use Fig. 4.2 to illustrate this criterion assuming that the region of attraction is the first quadrant of the constellation plane. Let x_0 be the transmitted symbol. Given a channel h_l , we make an error if the rotated constellation point $(\sqrt{\text{SNR}}h_l x)$ is dragged to a region other than $\mathcal{R}_0$, by AWGN. Therefore, by selecting the channel which makes the rotated constellation point furthest away from the decision boundary, we reduce the probability of making an error.

As illustrated in Fig. 4.2, the distances to the decision boundaries are $\left| \text{Re} \left(\sqrt{\text{SNR}}H_l X \right) \right|$ and $\left| \text{Im} \left(\sqrt{\text{SNR}}H_l X \right) \right|$ for each path l . Therefore the detector chooses the smaller distance out of these two distances for each path l

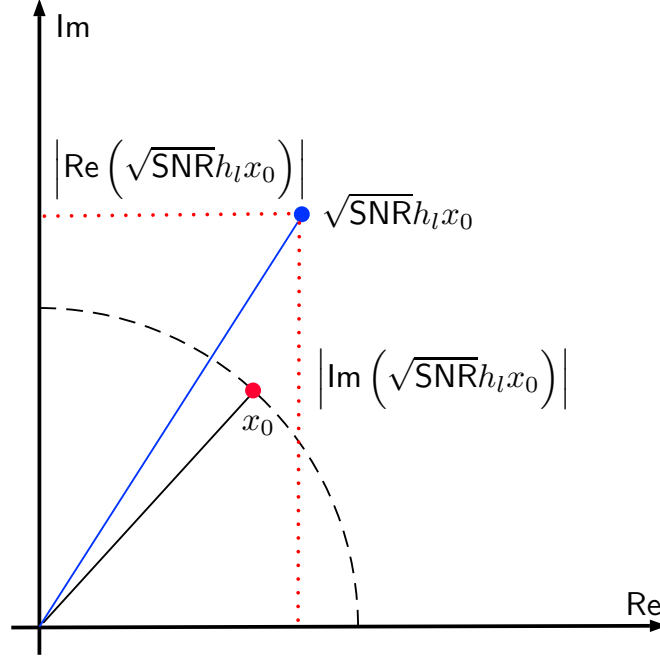


Figure 4.2: This is an example case for $n = 2$ case. We assume $x = x_0$ (indicated by the red dot) has been transmitted and $\text{Arg}(H_l) \in [-\frac{\pi}{4}, \frac{\pi}{4}]$. The transmit signal x_0 has been moved to $\sqrt{\text{SNR}}h_l x_0$ (indicated by the blue dot) by the channel $H_l = h_l$ in path l . Since, in this case the region of attraction is $\mathcal{R}_0$, the distances to the decision boundaries are $|\text{Re}(\sqrt{\text{SNR}}h_l x_0)|$ and $|\text{Im}(\sqrt{\text{SNR}}h_l x_0)|$. The detector compares these distances for all l and chooses the path according to (4.12).

(i.e., $\min \{ |\text{Re}(\sqrt{\text{SNR}}H_l X)|, |\text{Im}(\sqrt{\text{SNR}}H_l X)| \}$). Then the detector chooses the path that maximizes the minimum distance as the path L^* as given by

$$L^* = \arg \max_{l \in [0:N-1]} \{ \min (|H_l^{\text{re}} - H_l^{\text{im}}|, |H_l^{\text{re}} + H_l^{\text{im}}|) \} \quad (4.12)$$

where H_l^{re} and H_l^{im} denote the real and the imaginary parts of H_l , respectively. The observations available in the selected path L^* are used in (4.9) to estimate the transmitted symbol $\hat{x}(k^*, h^*)$.

4.5 Average Symbol Error Probability

In this section, we derive analytical expressions for the SEP when the sub-optimum detection criteria introduced in Section 4.4 are adopted.

Under selection combining we select the path L^* , which could be different based

on the detection criteria that we use. Once the path L^* is selected, the setup reduces to a point-to-point system in which $H_{L^*} = h^*$ and $k_{L^*} = k^*$. The pdf of H_{L^*} is different from one detection criteria to another. With a slight abuse of notation, we first define the SEP conditioned on the channel fading as

$$p(\text{SNR}, h^*) = \Pr \left\{ \sqrt{\text{SNR}} r^* e^{j\theta^*} + W^* \notin \mathcal{E}^* \right\}, \quad (4.13)$$

where $r^* e^{j\theta^*} = h^* x$, $W^* = W_{L^*}$ and the set $\mathcal{E}^*$ is defined as $\mathcal{E}^* = \{z \in \mathbb{C} : 0 \leq \text{Arg}(z) < \frac{\pi}{2}\}$. The probability in (4.13) can be calculated by conditioning on the real part of W^* , which is denoted by W_{re}^* as,

$$\Pr \left\{ \sqrt{\text{SNR}} r^* e^{j\theta^*} + W^* \notin \mathcal{E}^* \mid W_{\text{re}}^* = w \right\} = \mathcal{Q} \left(\sqrt{2\text{SNR}} r^* \sin \theta^* \right), \quad (4.14)$$

for $w \geq -\sqrt{\text{SNR}} r^* \cos \theta^*$. Similarly, for $w < -\sqrt{\text{SNR}} r^* \cos \theta^*$, we get

$$\Pr \left\{ \sqrt{\text{SNR}} r^* e^{j\theta^*} + W^* \notin \mathcal{E}^* \mid W_{\text{re}}^* = w \right\} = 1. \quad (4.15)$$

Integrating (4.14) and (4.15) with respect to the pdf of W_{re}^* , which is given by $f_{W_{\text{re}}^*}(w) = \frac{1}{\sqrt{\pi}} e^{-w^2}$, we obtain $p(\text{SNR}, h^*)$ as

$$\begin{aligned} p(\text{SNR}, h^*) &= \mathcal{Q} \left(\sqrt{2\text{SNR}} r^* \cos \theta^* \right) + \mathcal{Q} \left(\sqrt{2\text{SNR}} r^* \sin \theta^* \right) \\ &\quad - \mathcal{Q} \left(\sqrt{2\text{SNR}} r^* \cos \theta^* \right) \mathcal{Q} \left(\sqrt{2\text{SNR}} r^* \sin \theta^* \right). \end{aligned} \quad (4.16)$$

We also highlight that (4.16) is a special case of (2.19) in Chapter 2 which is derived for a point-to-point system with M -PSK modulation and Nakagami- m fading channels. Thus, more details on the derivation of (4.16) are available under the proof of Lemma 2.1 and Theorem 2.3 of Chapter 2. With Rayleigh fading, we can write the pdf of the fading magnitude for each path l as $f_{R_l}(r) = 2r e^{-r^2}$ for $r \geq 0$. Based on this, we next derive the distribution of H_{l^*} under each selection criteria and proceed to derive an expression for the average SEP.

4.5.1 SEP with Selection Criterion 1

Under selection criterion 1 given in Section 4.4.1, where the detector chooses the path with the maximum channel magnitude as the path L^* to estimate the transmitted signal $\hat{x}(k^*, h^*)$, the average SEP can be derived as given in the following Theorem.

Theorem 4.3. Assume H_l is a unit-power Rayleigh fading coefficient. Under selection criterion 1, for a system with 1 transmit antenna and N receive antennas with QPSK modulation and $n \geq 2$ bit quantization, the average SEP is given according to

$$p(\text{SNR}) = \frac{2^n N}{\pi} \int_{\frac{\pi}{4} - \frac{\pi}{2^n}}^{\frac{\pi}{4} + \frac{\pi}{2^n}} \int_0^\infty p(\text{SNR}, h^*) r^* \exp\left\{- (r^*)^2\right\} \left[1 - \exp\left\{- (r^*)^2\right\}\right]^{N-1} dr^* d\theta^*, \quad (4.17)$$

where $r^* = \max_{l \in [0:N-1]} r_l$, with $|h_l| = r_l$ and $p(\text{SNR}, h^*)$ is given in (4.16).

Proof: See Appendix 4.9.3. □

4.5.2 SEP with Selection Criterion 2

Under selection criterion 2 given in Section 4.4.2, where the detector chooses the path with the minimum absolute channel phase angle as the path L^* to estimate the transmitted signal $\hat{x}(k^*, h^*)$, the average SEP can be derived as given in the following Theorem 4.4.

Theorem 4.4. Assume H_l is a unit-power Rayleigh fading coefficient. Under selection criterion 2, for a system with 1 transmit antenna and N receive antennas with QPSK modulation and $n \geq 2$ bit quantization, the average SEP is given according to

$$p(\text{SNR}) = \frac{2^{n+1} N}{\pi} \int_{\frac{\pi}{4}}^{\frac{\pi}{4} + \frac{\pi}{2^n}} \int_0^\infty p(\text{SNR}, h^*) r^* \exp\left\{- (r^*)^2\right\} \left(1 + 2^{n-2} - \frac{2^n \theta^*}{\pi}\right)^{N-1} dr^* d\theta^*, \quad (4.18)$$

where $p(\text{SNR}, h^*)$ is given in (4.16).

Proof: See Appendix 4.9.4. □

4.5.3 SEP with Selection Criterion 3

Under selection criterion 3 given in Section 4.4.3, where the detector chooses the path L^* with the channel which makes the rotated constellation point furthest away from the decision boundaries, the average SEP can be derived as given in the following Theorem 4.5.

Theorem 4.5. *Assume H_l is a unit-power Rayleigh fading coefficient. Let $\mathcal{Q}(\cdot)$ be the complementary distribution function of the standard normal random variable [84]. Under selection criterion 3, for a system with 1 transmit antenna and N receive antennas with QPSK modulation, the average SEP is given according to*

$$p(\text{SNR}) = \frac{2^{n+1}N}{\pi} \int_{\frac{\pi}{4}}^{\frac{\pi}{4} + \frac{\pi}{2^n}} \int_0^\infty p(\text{SNR}, h^*) \left\{ G\left(\sqrt{\text{SNR}} r^* \cos \theta^*\right) \right\}^{N-1} r \exp\left\{-\left(r^*\right)^2\right\} dr^* d\theta^* \quad (4.19)$$

where $p(\text{SNR}, h^*)$ is given in (4.16) and

$$G\left(\sqrt{\text{SNR}} r^* \cos \theta^*\right) = g\left(\sqrt{2} r^* \cos \theta^*\right) - 2^{n-2} \left\{ 1 - 2\mathcal{Q}\left(\sqrt{2} r^* \cos \theta^*\right) \right\}^2$$

with

$$g(\alpha) = \frac{2^n}{\sqrt{\pi \text{SNR}}} \int_0^\alpha \left\{ 1 - 2\mathcal{Q}\left(\frac{\sqrt{2}t}{\sqrt{\text{SNR}}} \tan\left(\frac{\pi}{4} + \frac{\pi}{2^n}\right)\right) \right\} \exp\left(-\frac{t^2}{\text{SNR}}\right) dt.$$

Proof: See Appendix 4.9.5. □

4.6 The Decay Exponent for the Average Symbol Error Probability

In this section, we will analyze the communication robustness that can be achieved with low-resolution ADCs by focusing on the decay exponent for $p(\text{SNR})$, which is given by (2.5). We will use Definition 2.1 to simplify the notation. The following Theorems will establish some important results for DVO for our system under the three selection

criteria we presented above.

4.6.1 DVO with Selection Criterion 1

Under selection criterion 1 given in Section 4.4.1, where the detector chooses the path with the maximum channel magnitude as the path L^* to estimate the transmitted signal $\hat{x}(k^*, h^*)$, the DVO can be derived as given in the following Theorem 4.6.

Theorem 4.6. *The DVO of a low-resolution ADC based receiver architecture under selection criteria 1 with QPSK modulation and Rayleigh fading is given by*

$$\text{DVO} = \begin{cases} \frac{1}{2} & n = 2 \\ N & n \geq 3, \end{cases} \quad (4.20)$$

where N is the number of receive antennas.

Proof: See Appendix 4.9.6. □

4.6.2 DVO with Selection Criterion 2

Under selection criterion 2 given in Section 4.4.2, where the detector chooses the path with the minimum absolute channel phase angle as the path L^* to estimate the transmitted signal $\hat{x}(k^*, h^*)$, the DVO can be derived as given in the following Theorem 4.7.

Theorem 4.7. *The DVO of a low-resolution ADC based receiver architecture under selection criteria 2 with QPSK modulation and Rayleigh fading is given by*

$$\text{DVO} = 1 \quad (4.21)$$

where $N \geq 2$ is the number of receive antennas and $n \geq 2$.

Proof: See Appendix 4.9.7. □

4.6.3 DVO with Selection Criterion 3

Under selection criterion 3 given in Section 4.4.3, the detector chooses the path L^* with the channel which makes the rotated constellation point furthest away from the decision boundaries. In this case, we only obtain the DVO through numerical simulations. Based on the numerical simulations given in Section 4.7.4, we can obtain that

$$\text{DVO} = \begin{cases} \frac{N}{2} & n = 2 \\ N & n \geq 3. \end{cases} \quad (4.22)$$

Further, for the special case of $N = 2$ and $n = 2$, we find an upper and a lower bounded for DVO as given in Appendix 4.9.8.

For $n = 2$ case, the DVO of a low-resolution ADC based receiver architecture under selection criterion 3 with QPSK modulation and Rayleigh fading is difficult to derive. Thus, we take a bounded approach and show that the DVO is lower bounded by $\frac{1}{2}$ and upper bounded by $\frac{3}{2}$ for the specific case of $N = 2$. More details about the derivation of these bounds are included at the end of Appendix 4.9.8. However, based on simulations we show that the DVO in this system is $\frac{N}{2}$ where N is the number of receive antennas. We have added a detailed discussion on the DVO for $n = 2$ case in Section 4.7.4.

4.7 Numerical Results

In this section, we present numerical examples to illustrate the accuracy of SEP analysis for QPSK modulation with n -bit phase quantization. Channel fading is unit-power and circularly-symmetric with Rayleigh distributed magnitude, and additive noise is complex Gaussian with zero mean and unit variance.

4.7.1 System Performance with ML Detector

First, we investigate the system performance with the optimum ML detector presented in Sections 4.3.1 and 4.3.2. Fig. 4.3 plots the average SEP as a function of SNR for QPSK modulation with $n = 2, 3, 4$ and ∞ -bit quantization under Rayleigh fading. We change

the number of receive antennas as $N = 2, 3, 4$ and simulate the SEP results based on the detection rule provided in Theorem 4.1 and 4.2. We use the ML detector given in (4.3) to plot the average SEP for the unquantized case ($n = \infty$). We observe a noteworthy improvement in the average SEP when n changes from 2 to 3-bit quantization for QPSK modulation. This states that using one extra bit, on top of $\log_2 M$ bits, improves SEP significantly. By using the simulation results, we can figure out the DVO and it moves from $\frac{N}{2}$ to N when n changes from 2 to 3 or more, i.e., we observe full DVO when the number of quantization bits is $\log_2 M + 1$. We also observe that the average SEP reduces as we increase n , but the amount by which it reduces also gets smaller as we increase n . As expected, $\text{DVO} = N$ for all $n \geq 3$. Furthermore, $\text{DVO} = \frac{N}{2}$ when $n = 2$.

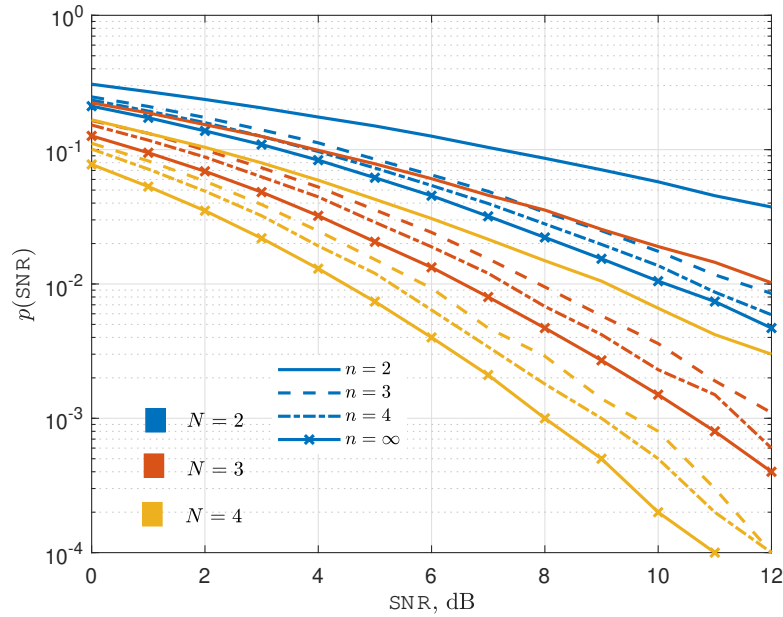


Figure 4.3: Average SEP curves as a function of SNR with QPSK modulation and ML detection. $n = 2, 3, 4, \infty$ and $N = 2, 3, 4$. We note that the non-smooth variations of the curves in the high SNR regime are a result of numerical errors in the simulation.

Next, we investigate the effect of the number of quantization bits on the system performance with selection combining.

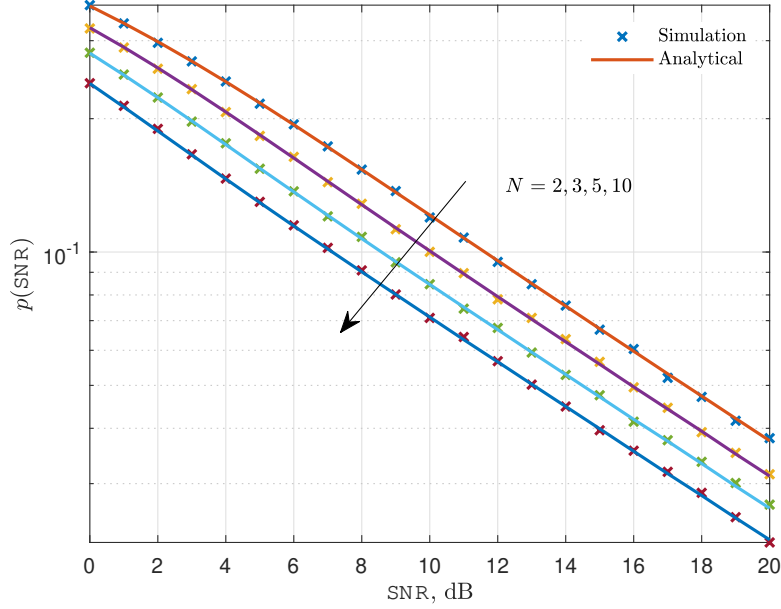


Figure 4.4: Average SEP curves as a function of SNR with QPSK modulation and selection combining with selection criterion 1. $n = 2$ and $N = 2, 3, 5, 10$.

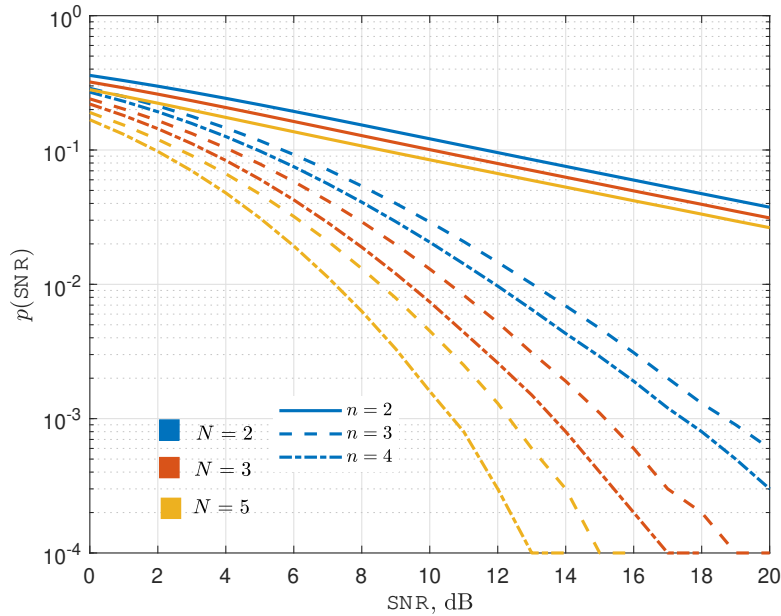


Figure 4.5: Average SEP curves as a function of SNR with QPSK modulation and selection combining with selection criterion 1. $n = 2, 3, 4$ and $N = 2, 3, 5$.

4.7.2 System Performance under Selection Criterion 1

Fig. 4.4 plots the average SEP as a function of SNR for QPSK modulation with $n = 2$ bit quantization under Rayleigh fading. We change the number of receive antennas as $N = 2, 3, 5, 10$. The simulated results are generated using Monte Carlo simulation, while the analytical results are generated using our expression in (4.17). As the plot illustrates, the analytical results accurately follow the simulated results for all four cases. As established by Theorem 4.6, we observe that with 2-bit quantization, the DVO remains $\frac{1}{2}$ for all N . We also observe that the average SEP reduces as we increase N , but the amount by which it reduces also gets smaller as we increase N .

Fig. 4.5 plots the average SEP as a function of SNR for QPSK modulation with $n = 2, 3, 4$ bit quantization under Rayleigh fading. We also change the number of antennas as $N = 2, 3, 5$. The plots are generated using the analytical expression in (4.17). We observe a noteworthy improvement in the average SEP when n changes from 2 to 3-bit quantization for all $N = 2, 3, 5$. This is expected in light of Theorem 4.6, which states that using one extra bit, on top of 2-bits, improves the DVO from $\frac{1}{2}$ to N . We also observe that the average SEP reduces as we increase n , but the amount by which it reduces also gets smaller as we increase n . As expected, $\text{DVO} = N$ for all $n \geq 3$ and $\text{DVO} = \frac{1}{2}$ when $n = 2$.

4.7.3 System Performance under Selection Criterion 2

Fig. 4.6 plots the average SEP as a function of SNR for QPSK modulation with $n = 2$ bit quantization under Rayleigh fading. We also change the number of antennas as $N = 2, 3, 5, 10$. The simulated results are generated using Monte Carlo simulation, while the analytical results are generated using our expression in (4.18). As the plot illustrates, the analytical results accurately follow the simulated results for all cases. As established by Theorem 4.7, we observe that the $\text{DVO} = 1$ for all $n \geq 2$ and $N \geq 2$. We also observe that the average SEP reduces as we increase N , but the amount by which it reduces also gets smaller as we increase N .

Fig. 4.7 plots the average SEP as a function of SNR for QPSK modulation with

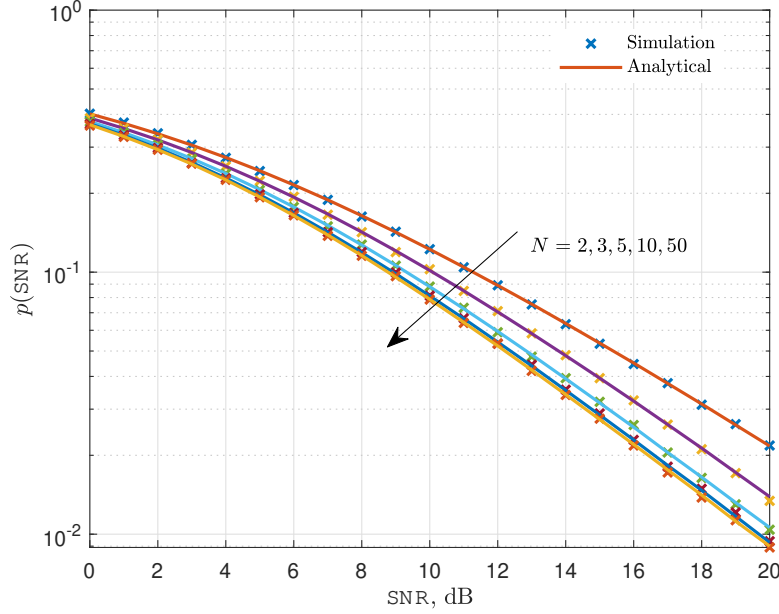


Figure 4.6: Average SEP curves as a function of SNR with QPSK modulation and selection combining with selection criterion 2. $n = 2$ and $N = 2, 3, 5, 10, 50$.

$n = 2, 3, 4$ bit quantization under Rayleigh fading. We also change the number of antennas as $N = 2, 3, 5$. The plots are generated using the analytical expression in (4.18). We do not observe a noteworthy improvement in the average SEP even when we use more quantization bits. This is expected in light of Theorem 4.7, which states that the DVO is 1 for all $N \geq 2$ and $n \geq 2$. This may be because selection criterion 2 does not take the advantage of better SNR values at other antennas. In the classical approach, more antenna branches improve DVO since some of them provide higher SNR either selected individually or maximally combined together. Here, there is no such SNR gain due to the SNR-ignorant operation of the system.

4.7.4 System Performance under Selection Criterion 3

Fig. 4.8 plots the average SEP as a function of SNR for QPSK modulation with $n = 2$ bit quantization under Rayleigh fading. We also change the number of antennas as $N = 2, 3, 4$. The simulated results are generated using Monte Carlo simulation, while the analytical results are generated using our expression in (4.19) for $n = 2$. As the plot

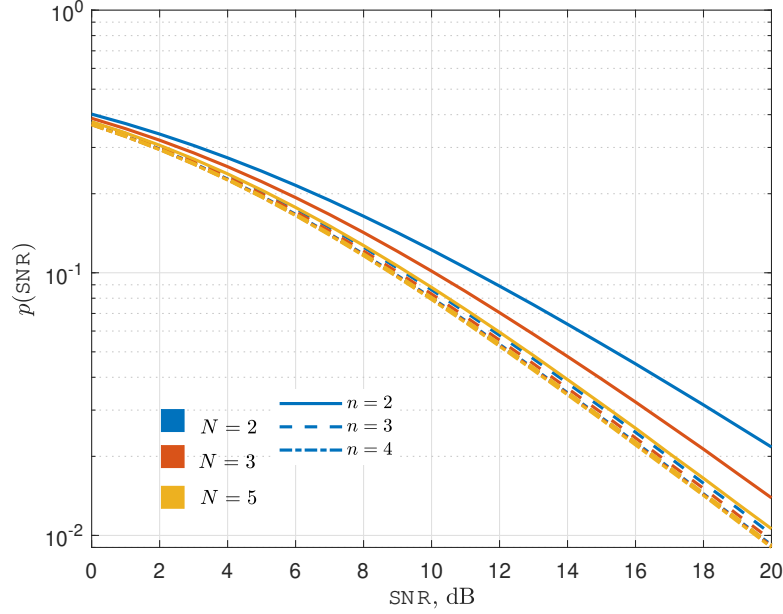


Figure 4.7: Average SEP curves as a function of SNR with QPSK modulation and selection combining with selection criterion 2. $n = 2, 3, 4$ and $N = 2, 3, 5, 10, 50$.

illustrates, the analytical results accurately follow the simulated results for all cases. We observe a noteworthy improvement in the average SEP when N changes. Using the simulations, with 2-bit quantization we get the $DVO = \frac{N}{2}$. That is, $DVO = 1, 1.5$ and 2 as we change N as $2, 3$, and 4 , respectively.

Fig. 4.9 plots the average SEP as a function of SNR for QPSK modulation with $n = 2, 3, 4$ -bit quantization under Rayleigh fading. Again we change the number of antennas as $N = 2, 3, 4$. The curves are generated using the analytical expression in (4.19) for $n \geq 3$. We observe a noteworthy improvement in the average SEP when n changes from 2 to 3-bit quantization for all $N = 2, 3, 4$. Through simulation results, we obtain $DVO = \frac{N}{2}$ for 2-bit quantization and $DVO = N$ for $n \geq 3$ bit quantization. This means that using one extra bit, on top of 2-bits, improves the DVO from $\frac{N}{2}$ to N . We also observe that the average SEP reduces as we increase n , but the amount by which it reduces also gets smaller as we increase n .

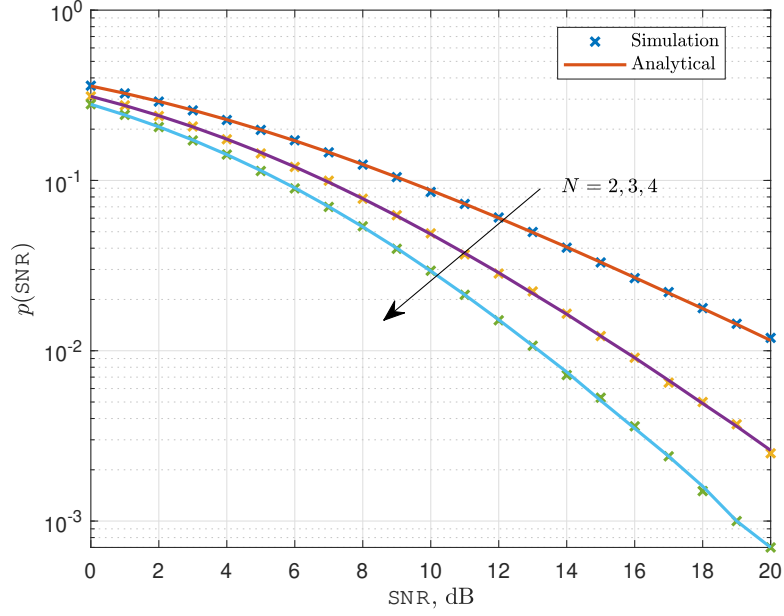


Figure 4.8: Average SEP curves as a function of SNR with QPSK modulation and selection combining with selection criterion 3. $n = 2$ and $N = 2, 3, 4$.

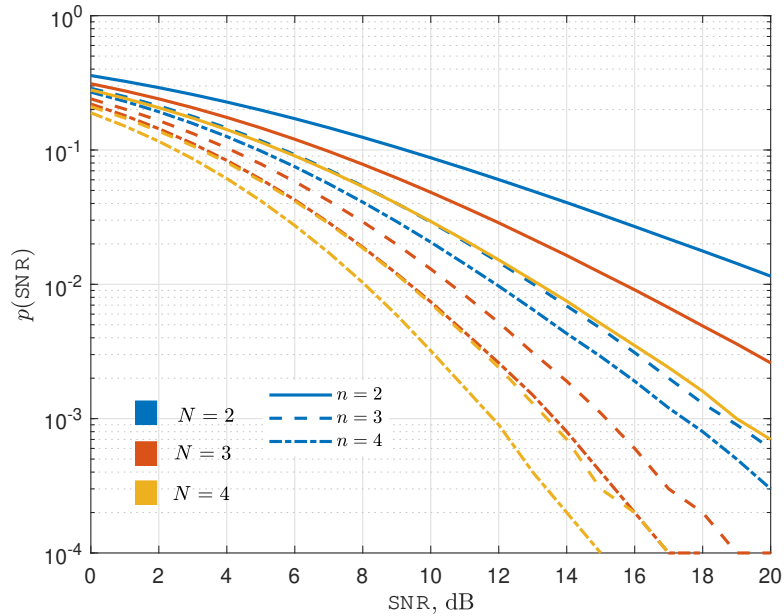


Figure 4.9: Average SEP curves as a function of SNR with QPSK modulation and selection combining with selection criterion 3. $n = 2, 3, 4$ and $N = 2, 3, 4$.

4.7.5 System Performance Comparison

Fig. 4.10 plots the average SEP as a function of SNR for QPSK modulation under Rayleigh fading with the selection criteria 1, 2 and 3 with $n = 2, 3, 4$ and $N = 2$. The curves are generated using Theorem 4.3, 4.4 and 4.5 respectively for selection criterion 1, 2 and 3. We observe that selection criterion 3 outperforms selection criterion 1 and 2. For $n = 2$ case, this performance difference is very clear. However, for $n \geq 3$, selection criterion 3 outperforms selection criterion 1 by a very small margin.

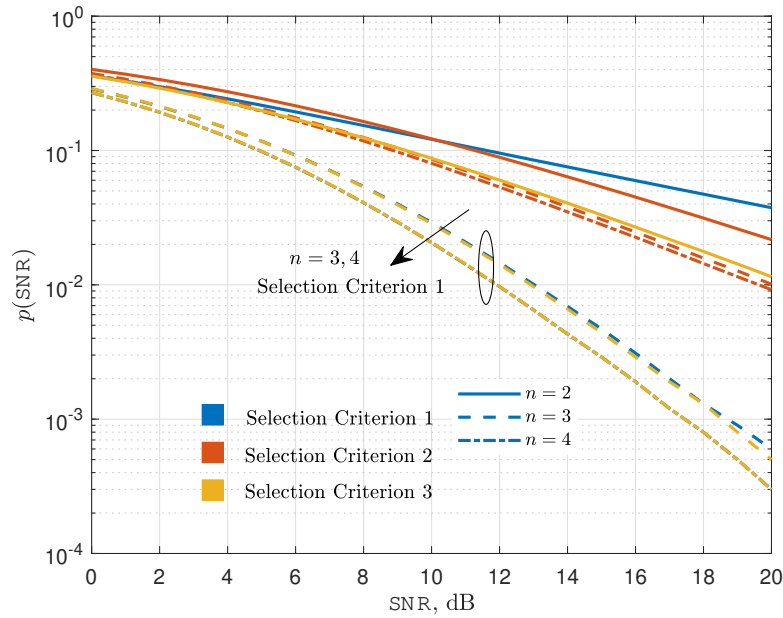


Figure 4.10: Average SEP curves as a function of SNR with QPSK modulation and selection combining with different selection criteria. $n = 2, 3, 4$ and $N = 2$.

Fig. 4.11 plots the average SEP as a function of SNR for QPSK modulation under Rayleigh fading with the ML detector and selection criterion 3 with $n = 2, 3, 4$ and $N = 2$. The curves related to the ML detector are generated using Monte Carlo simulation and the curves related to selection criterion 3 are generated using Theorem 4.5. We observe that selection criterion 3 shows the same DVO performance compared to the ML detector. That is, with 2-bit quantization the DVO is $\frac{N}{2}$ and with 3 or more bit quantization, the DVO is N .

Fig. 4.12 plots the average SEP as a function of SNR for QPSK modulation under Rayleigh fading with the ML detector, selection criterion 3 and majority combining

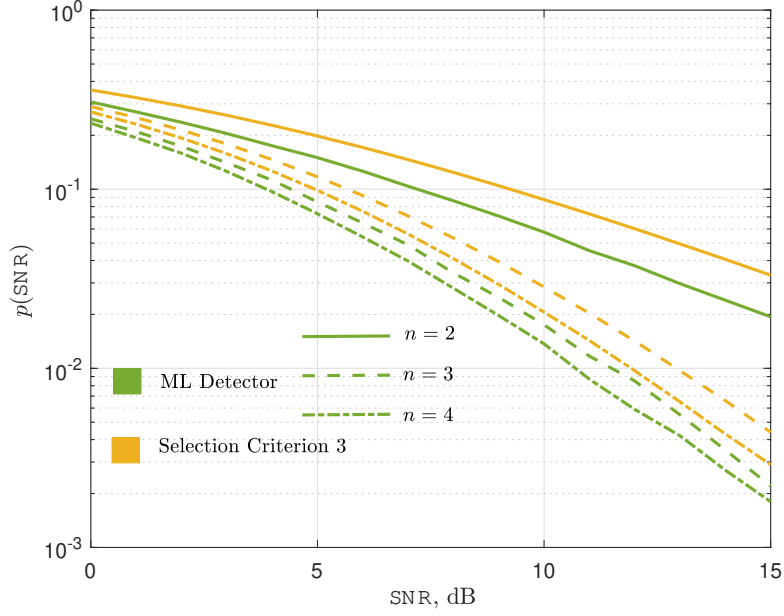


Figure 4.11: Average SEP curves as a function of SNR with QPSK modulation under the ML detector and selection combining with selection criterion 3. $n = 2, 3, 4$ and $N = 2$.

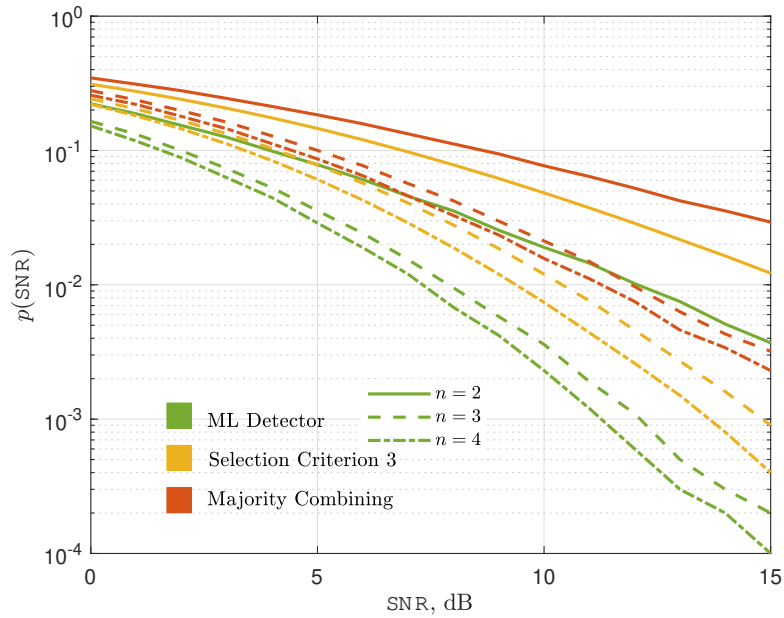


Figure 4.12: Average SEP curves as a function of SNR with QPSK modulation under the ML detector, selection combining with selection criterion 3 and majority combining. $n = 2, 3, 4$ and $N = 3$. We note that the non-smooth variations of the curves in the high SNR regime are a result of numerical errors in the simulation.

with $n = 2, 3, 4$ and $N = 3$. In majority combining, the detector first estimate the transmitted signal $\hat{x}_l(k_l, h_l)$ by applying Theorem 2.2 for each path $l \in [0 : N - 1]$. Then the estimated transmitted symbol $\hat{x}(k, h)$ will be decided based on the majority of $\hat{x}_l(k_l, h_l)$ for all $l \in [0 : N - 1]$. This can be given as

$$\hat{x}(k, h) = \text{majority of } \{\hat{x}_0(k_0, h_0), \hat{x}_1(k_1, h_1), \dots, \hat{x}_{N-1}(k_{N-1}, h_{N-1})\}.$$

If the majority is shared between two or more estimated transmitted signals, then the detector has to make a random guess to estimate the transmitted signal. While this is a sub-optimal approach, we only use this method for comparison and do not perform a mathematical analysis.

We observe that both the ML detector and selection criterion 3 outperform the majority combining based detector in terms of SEP for all $n = 2, 3, 4$ and $N = 3$. We also observe that the average SEP performance gap between the ML detector and the majority combining detector increases as we increase n . Through Monte Carlo simulation, we also observe that the DVO is less than $\frac{N}{2}$ for $n = 2$ and DVO is less than N for $n \geq 3$. Therefore, the proposed selection criterion 3 based sub-optimum detector is the closest to the ML detector in terms of SEP performance.

4.8 Conclusions and Future Generalizations

In this chapter, we performed a theoretical analysis of a low-resolution based ADC communication system with a single transmit antenna and multiple receive antennas and obtained fundamental performance limits, optimum ML detectors, sub-optimum detectors and a general analytical expression for the average SEP for QPSK modulation with n -bit quantization. These results were further investigated for the Rayleigh fading model in detail. We presented three antenna selection strategies and characterized their SEP performance limits. The average SEP expressions are derived for QPSK modulation with n -bit quantization. For the Rayleigh fading channel, it is shown that a transceiver architecture with n -bit quantization and N receiver antennas is asymptotically optimum in terms of communication reliability if $n \geq 3$. That is, with

the optimum ML detector, the decay exponent for the average SEP is the same and equals to N either with infinite-bit quantizers or with n -bit quantizers where $n \geq 3$. On the other hand, it is only equal to $\frac{N}{2}$ and 0 for $n = 2$ and $n < 2$, respectively. We also note that the simple sub-optimum detector that selects the path with the channel that locates the rotated constellation point furthest away from the decision boundary is *asymptotically* optimum in terms of communication reliability if $n \geq 3$.

A major result of this chapter is the discovery of a ternary SEP behaviour, indicating the sufficiency of 3-bits for achieving asymptotically optimum QPSK communication reliability. Hence, without modifying the conventional RF chain, we can use one extra bit and still achieve asymptotically optimum communication performance. Another important result is that the discovery of the new selection combining based sub-optimum detector that performs the same as the ML detector in terms of DVO.

We can extend this research to M -ary modulation schemes under the Nakagami- m fading as given in Chapter 2. This would allow generalizing our findings to broader cases. We then can extend this analysis to a low-resolution quantization based MIMO wireless communication systems. Further, as discussed in Chapter 2, another important future research direction is to compare and contrast the backwards-compatible receiver design approach of using one extra bit proposed in this chapter.

4.9 Appendix

4.9.1 Proof of Theorem 4.1

In order to obtain the ML rule for the setup described above, we have to maximize the likelihood function $\mathcal{L}(x)$ given by

$$\mathcal{L}(x) = \Pr\{Q(\mathbf{Y}) = \mathbf{k} \mid X = x, \mathbf{H} = \mathbf{h}\}. \quad (4.23)$$

First, we re-express (4.23) as a multiplication of N probabilities given by

$$\mathcal{L}(x) = \prod_{l=0}^{N-1} \Pr \left\{ \sqrt{\text{SNR}} h_l x + W_l \in \mathcal{R}_{k_l} \mid X = x, H_l = h_l \right\}. \quad (4.24)$$

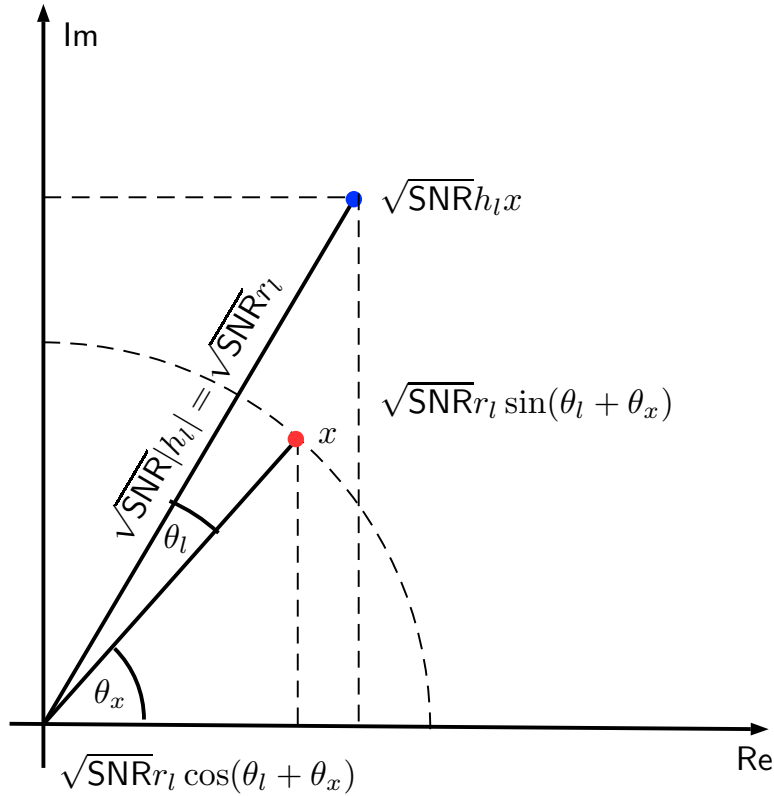


Figure 4.13: This is an example case. We assume the received signal at path l has been landed in $\mathcal{R}_0$. The transmitted signal x (denoted by the red dot) has been rotated by the channel h_l to the point $\sqrt{\text{SNR}}h_l x$ (denoted by the blue dot). Unless the noise W_l drags $\sqrt{\text{SNR}}h_l x$ to a different quantization region other than $\mathcal{R}_0$, there will be no errors in estimating the transmitted signal. As such, there will be no error if $Y_l^{\text{re}} \geq 0$ and $Y_l^{\text{im}} \geq 0$.

Then, we use Fig. 4.13 as a visual guide to provide the proof for ML detection rule when $n = 2$. By defining $Z^{\text{re}} = \text{Re}(Z)$ and $Z^{\text{im}} = \text{Im}(Z)$ for any $Z \in \mathbb{C}$ we can write the real and the imaginary parts of the signal received at antenna l as

$$\begin{aligned} Y_l^{\text{re}} &= \sqrt{\text{SNR}}r_l \cos(\theta_l + \theta_x) + W_l^{\text{re}}, \\ Y_l^{\text{im}} &= \sqrt{\text{SNR}}r_l \sin(\theta_l + \theta_x) + W_l^{\text{im}}, \end{aligned}$$

where $r_l = |h_l|$, $\theta_l = \text{Arg}(h_l)$ and $\theta_x = \text{Arg}(x)$. From Fig. 4.13 we observe that unless the noise W_l drags $\sqrt{\text{SNR}}h_l x$ to a different quantization region other than $\mathcal{R}_0$, there will

be no errors in estimating the transmitted signal. As such, we can re-express (4.24) as

$$\begin{aligned}
\mathcal{L}(x) &= \prod_{l=0}^{N-1} \Pr \left\{ \rho_l^{\text{re}} \sqrt{\text{SNR}} r_l \cos(\theta_l + \theta_x) + W_l^{\text{re}} \geq 0 \mid X = e^{j\theta_x}, H_l = r_l e^{j\theta_l} \right\} \\
&\quad \Pr \left\{ \rho_l^{\text{im}} \sqrt{\text{SNR}} r_l \sin(\theta_l + \theta_x) + W_l^{\text{im}} \geq 0 \mid X = e^{j\theta_x}, H_l = r_l e^{j\theta_l} \right\} \\
&= \prod_{l=0}^{N-1} \Pr \left\{ W_l^{\text{re}} \geq -\rho_l^{\text{re}} \sqrt{\text{SNR}} r_l \cos(\theta_l + \theta_x) \mid X = e^{j\theta_x}, H_l = r_l e^{j\theta_l} \right\} \\
&\quad \Pr \left\{ W_l^{\text{im}} \geq -\rho_l^{\text{im}} \sqrt{\text{SNR}} r_l \sin(\theta_l + \theta_x) \mid X = e^{j\theta_x}, H_l = r_l e^{j\theta_l} \right\}, \tag{4.25}
\end{aligned}$$

where the terms $\rho_l^{\text{re}} = \text{sgn}(\text{Re}(Y_l))$ and $\rho_l^{\text{im}} = \text{sgn}(\text{Im}(Y_l))$ are the signs of real and imaginary parts of the received signal in path l , respectively. Since $W_l \sim \mathcal{CN}(0, 1)$ is the AWGN at receiving antenna l we note that W_l^{re} and W_l^{im} are Gaussian random variables with zero mean and variance $\frac{1}{2}$ and the resulting probabilities can be written in terms of the $\mathcal{Q}$ function as

$$\mathcal{L}(x) = \prod_{l=0}^{N-1} \mathcal{Q} \left(-\rho_l^{\text{re}} \sqrt{\text{SNR}} r_l \cos(\theta_l + \theta_x) \right) \mathcal{Q} \left(-\rho_l^{\text{im}} \sqrt{\text{SNR}} r_l \sin(\theta_l + \theta_x) \right). \tag{4.26}$$

The ML detection rule given in (4.6) is achieved by maximizing the likelihood function in (4.26).

4.9.2 Proof of Theorem 4.2

In this section, we provide the proof of theorem 2. In order to obtain the ML rule for the setup described above, we have to maximize the likelihood function $\mathcal{L}(x)$ given by (4.24). To that end, we need to maximize the probability $\Pr \{Y_l \in \mathcal{R}_{k_l} \mid X = x, H_l = h_l\}$ which states the probability of landing in the quantization region $\mathcal{R}_{k_l}$ which is no longer a quadrant. In order to find this probability we first write the real and imaginary parts of the received signal at antenna l as

$$Y_l^{\text{re}} = \sqrt{\text{SNR}} r_l \cos(\theta_l + \theta_x) + W_l^{\text{re}}$$

and

$$Y_l^{\text{im}} = \sqrt{\text{SNR}} r_l \sin(\theta_l + \theta_x) + W_l^{\text{im}},$$

where $r_l = |h_l|$, $\theta_l = \text{Arg}(h_l)$ and $\theta_x = \text{Arg}(x)$. Since W_l^{re} and W_l^{im} are Gaussian random variables with zero mean and a variance of $\frac{1}{2}$, given $H_l = h_l$ we have $Y_l^{\text{re}} \sim \mathcal{N}\left(\sqrt{\text{SNR}} r_l \cos(\theta_l + \theta_x), \frac{1}{2}\right)$ and $Y_l^{\text{im}} \sim \mathcal{N}\left(\sqrt{\text{SNR}} r_l \sin(\theta_l + \theta_x), \frac{1}{2}\right)$. Then we proceed to find the joint pdf of Y_l^{re} and Y_l^{im} which can be expressed using a bivariate Gaussian distribution as

$$f_{Y_l^{\text{re}}, Y_l^{\text{im}}}(y_l^{\text{re}}, y_l^{\text{im}}) = \frac{1}{\pi} \exp \left\{ - \left(y_l^{\text{re}} - \sqrt{\text{SNR}} r_l \cos(\theta_l + \theta_x) \right)^2 - \left(y_l^{\text{im}} - \sqrt{\text{SNR}} r_l \sin(\theta_l + \theta_x) \right)^2 \right\}. \quad (4.27)$$

Thus, for each l we can find $\Pr\{Y_l \in \mathcal{R}_{k_l} | X = x, H_l = h_l\}$ by integrating the above pdf within appropriate regions. However, in our model, these regions are convex cones and to proceed with our analysis we convert the pdf into polar coordinates and write $\Pr\{Y_l \in \mathcal{R}_{k_l} | X = x, H_l = h_l\}$ as

$$\Pr\{Y_l \in \mathcal{R}_{k_l} | X = x, H_l = h_l\} = \int_{k_l \frac{\pi}{2^{n-1}}}^{(k_l+1) \frac{\pi}{2^{n-1}}} \int_0^\infty f_{Y_l^{\text{re}}, Y_l^{\text{im}}}(r \cos \theta, r \sin \theta) r dr d\theta. \quad (4.28)$$

By substituting (4.27) into (4.28) and using the integral identity given by

$$\begin{aligned} & \int z \exp(-az^2 + bz + c) dz \\ &= \frac{\sqrt{\pi}}{4a^{\frac{3}{2}}} b \exp\left(c + \frac{b^2}{4a}\right) \text{erf}\left(\frac{2az - b}{2\sqrt{a}}\right) - \frac{1}{2a} \exp(-az^2 + bz + c) + C, \end{aligned}$$

where C is a constant, we can reduce this integral expression to

$$\Pr\{Y_l \in \mathcal{R}_{k_l} | X = x, H_l = h_l\} = \frac{1}{2^n} \exp(-\text{SNR} r_l^2) + \sqrt{\frac{\text{SNR}}{\pi}} r_l g(\text{Arg}(x), k_l, h_l), \quad (4.29)$$

where $g(\text{Arg}(x), k_l, h_l)$ is given in Theorem 4.2. Substituting (4.29) into (4.24), the ML detection rule given in (4.7) is achieved.

4.9.3 Proof of Theorem 4.3

As the channel fading from the transmitter to each receiving antenna is independent, we can write the cumulative distribution function (cdf) of $R^* = |H_{l^*}|$ as

$$F_{R^*}(r^*) = \Pr\{R^* \leq r^*\} = \left[1 - \exp\left\{-(r^*)^2\right\}\right]^N,$$

and the resulting pdf can be given as

$$f_{R^*}(r^*) = 2Nr^* \exp\left\{-(r^*)^2\right\} \left[1 - \exp\left\{-(r^*)^2\right\}\right]^{N-1}. \quad (4.30)$$

Similar to the analysis given under the proof for Theorem 2.3 of Chapter 2, using the fact that θ^* lies uniformly between $(\frac{\pi}{4} - \frac{\pi}{2^n})$ and $(\frac{\pi}{4} + \frac{\pi}{2^n})$ for n bit quantization, we get the expression in (4.17) by averaging (4.16) with respect to the pdfs of R^* and θ^* .

4.9.4 Proof of Theorem 4.4

In order to find $p(\text{SNR})$, we can average (4.16) over the distributions of $R^* = |H_{l^*}|$ and Θ^* . As discussed in Theorem 2.3 in Chapter 2, without loss of generality let us assume that $x_0 = e^{j\frac{\pi}{4}}$ is transmitted and $\Theta_l \in [-\frac{\pi}{2^n}, \frac{\pi}{2^n}]$ for all $l \in [0 : N-1]$. By defining $\Omega_l = |\Theta_l|$, we can write the cdf of it as $F_{\Omega_l}(\omega) = \frac{2^n}{\pi}\omega$. According to the selection rule, the detector selects Ω^* in such a way that $\Omega^* = \min_{l \in [0:N-1]} \Omega_l$. Thus the cdf of Ω^* can be derived as

$$F_{\Omega^*}(\omega^*) = \Pr\{\Omega^* \leq \omega^*\} = 1 - \left(1 - \frac{2^n \omega^*}{\pi}\right)^N.$$

Therefore, in the selected path L^* we have $\Theta_{L^*} = \text{sgn}(\Theta_{L^*}) \Omega^*$ where $\text{sgn}(\cdot)$ is given as (1.2). Then we can write the pdf of Θ_{L^*} as

$$f_{\Theta_{L^*}}(\lambda) = \frac{2^{n-1}N}{\pi} \left(1 + \frac{2^n |\lambda|}{\pi}\right)^{N-1},$$

where $|\lambda|$ is the modulus of λ . Then the angle of the rotated constellation point ($H^* x_0$) will be $\Theta^* = \Theta_{L^*} + \frac{\pi}{4}$. Therefore, the pdf of the angle of the rotated constellation point

Θ^* can be given as

$$f_{\Theta^*}(\theta^*) = \begin{cases} \frac{2^{n-1}N}{\pi} \left(1 + 2^{n-2} - \frac{2^n\theta^*}{\pi}\right)^{N-1} & \text{if } \theta^* \geq \frac{\pi}{4} \\ \frac{2^{n-1}N}{\pi} \left(1 - 2^{n-2} + \frac{2^n\theta^*}{\pi}\right)^{N-1} & \text{if } \theta^* < \frac{\pi}{4}. \end{cases}$$

Since the pdf of Rayleigh fading magnitude in the selected path L^* is $f_{R^*}(r^*) = 2r^* \exp\{- (r^*)^2\}$, we can use $f_{R^*}(r^*)$ and $f_{\Theta^*}(\theta^*)$ to average $p(\text{SNR}, h^*)$ to get the SEP expression in (4.18).

4.9.5 Proof of Theorem 4.5

We will first obtain the following lemma as a supporting result.

Lemma 4.1. *Consider a finite sequence of random variables*

$$(U_0, V_0), (U_1, V_1) \cdots (U_{N-1}, V_{N-1}),$$

where $U_0, \dots, U_{N-1}, V_0, \dots, V_{N-1}$ are all iid with a joint pdf of

$$f_{U_l, V_l}(u, v) = \begin{cases} \frac{f(u)f(v)}{P_{\mathcal{E}}} & \text{if } (u, v) \in S_{\mathcal{E}} \\ 0 & \text{otherwise,} \end{cases}$$

where $f(u)$ is a common pdf, $S_{\mathcal{E}}$ is the convex cone given as

$$S_{\mathcal{E}} = \left\{ (u, v) \in \mathbb{R}^2 : \frac{\pi}{4} - \mathcal{E} \leq \tan^{-1}\left(\frac{v}{u}\right) \leq \frac{\pi}{4} + \mathcal{E} \right\},$$

and $P_{\mathcal{E}}$ is the normalization coefficient given by

$$\begin{aligned} P_{\mathcal{E}} &= \int_0^\infty \int_{u \tan(\frac{\pi}{4} - \mathcal{E})}^{u \tan(\frac{\pi}{4} + \mathcal{E})} f(v) f(u) dv du, \\ &= \int_0^\infty f(u) \left[F\left(u \tan\left(\frac{\pi}{4} + \mathcal{E}\right)\right) - F\left(u \tan\left(\frac{\pi}{4} - \mathcal{E}\right)\right) \right] du, \end{aligned}$$

where $F(u)$ is the corresponding cdf of $f(u)$. Let $Z_l^{\min} = \min\{U_l, V_l\}$, $Z_l^{\max} = \max\{U_l, V_l\}$

and then $Z^* = \max_{l \in [0:N-1]} Z_l^{\min}$, $L^* = \arg \max_{l \in [0:N-1]} Z_l^{\min}$, and $Z_{-0}^* = \max_{l \in [1:N-1]} Z_l^{\min}$. Further, let E_l

be the event such that $E_l \triangleq \{Z^* = Z_l^{\min}\}$ and $\Pr\{E_l\} = \frac{1}{N}$ for all $l \in [0 : N-1]$. Then let $g_{Z^*}(z)$ and $g_{Z_{-0}^*}(z)$ be the pdf of Z^* and Z_{-0}^* respectively. Then the following results hold.

1.

$$f_{Z_{L^*}^{\min}, Z_{L^*}^{\max}}(u, v) = \begin{cases} \frac{f(v)g_{Z^*}(u)}{F(u \tan(\frac{\pi}{4} + \mathcal{E})) - F(u)} & \text{if } v \geq u \\ 0 & \text{otherwise.} \end{cases} \quad (4.31)$$

2.

$$f_{Z_0^{\max}|E_0, Z_0^{\min}}(v|u) = \begin{cases} \frac{f(v)}{F(u \tan(\frac{\pi}{4} + \mathcal{E})) - F(u)} & \text{if } v \geq u \\ 0 & \text{otherwise.} \end{cases} \quad (4.32)$$

Proof: We first provide the proof for the first case. In this case, we can write

$$f_{Z_{L^*}^{\min}, Z_{L^*}^{\max}}(u, v) = \sum_{l=0}^{N-1} \Pr\{E_l\} f_{Z_l^{\min}, Z_l^{\max}|E_l}(u, v).$$

Since $\Pr\{E_l\} = \frac{1}{N}$ and $f_{Z_l^{\min}, Z_l^{\max}|E_l}(u, v) = f_{Z_0^{\min}, Z_0^{\max}|E_l}(u, v)$ for all $l \in [0 : N-1]$, we have $f_{Z_{L^*}^{\min}, Z_{L^*}^{\max}}(u, v) = f_{Z_0^{\min}, Z_0^{\max}|E_0}(u, v)$. That concludes the proof of the case 1.

Therefore, we then find $F_{Z_0^{\min}, Z_0^{\max}|E_0}(u, v) = \Pr\{Z_0^{\min} \leq u, Z_0^{\max} \leq v | E_0\}$ which is the joint cdf of $Z_0^{\min}$ and $Z_0^{\max}$ conditioned on the event E_0 . Then, for $v \geq u$, we have $F_{Z_0^{\min}, Z_0^{\max}|E_0}(u, v) = \Pr\{Z_0^{\min} \leq u, Z_0^{\max} \leq v | Z_0^{\min} \geq Z_{-0}^*\}$. By using the Bayes' theorem, we can write $F_{Z_0^{\min}, Z_0^{\max}|E_0}(u, v) = N \Pr\{Z_{-0}^* \leq Z_0^{\min} \leq u, Z_0^{\max} \leq v\}$. Further, by conditioning on the event $Z_{-0}^* = z$, we can write

$$F_{Z_0^{\min}, Z_0^{\max}|E_0}(u, v) = N \int_0^u \Pr\{z \leq Z_0^{\min} \leq u, Z_0^{\max} \leq v | Z_{-0}^* = z\} g_{Z_{-0}^*}(z) dz.$$

Due to the independence of Z_{-0}^* from (U_0, V_0) , we can re-write the above expression as

$$F_{Z_0^{\min}, Z_0^{\max}|E_0}(u, v) = N \int_0^u \Pr\{z \leq Z_0^{\min} \leq u, Z_0^{\max} \leq v\} g_{Z_{-0}^*}(z) dz.$$

Considering the symmetry of the problem, we note that

$$\Pr \{z \leq Z_0^{\min} \leq u, Z_0^{\max} \leq v\} = 2 \Pr \{z \leq U_0 \leq u, U_0 \leq V_0 \leq v\}.$$

Therefore, we have

$$F_{Z_0^{\min}, Z_0^{\max} | E_0}(u, v) = 2N \int_0^u \Pr \{z \leq U_0 \leq u, U_0 \leq V_0 \leq v\} g_{Z_{-0}^*}(z) dz.$$

After some mathematical manipulations, we then write $f_{Z_0^{\min}, Z_0^{\max} | E_0}(u, v)$ which is the joint pdf of $Z_0^{\min}$ and $Z_0^{\max}$ given E_0 as $f_{Z_0^{\min}, Z_0^{\max} | E_0}(u, v) = \frac{\partial^2}{\partial u \partial v} F_{Z_0^{\min}, Z_0^{\max} | E_0}(u, v)$ can be given by

$$f_{Z_0^{\min}, Z_0^{\max} | E_0}(u, v) = \begin{cases} \frac{2N}{P_{\mathcal{E}}} f(v) f(u) G_{Z_{-0}^*}(u) & \text{if } (u, v) \in S_{\mathcal{E}} \\ 0 & \text{otherwise,} \end{cases} \quad (4.33)$$

where $G_{Z_{-0}^*}(u) = \Pr \{Z_{-0}^* \leq u\}$ and $v \geq u$. Since $g_{Z^*}(u) = \frac{\partial}{\partial u} \Pr \{Z_0^{\min} \leq u\}^N$, we have $g_{Z^*}(u) = N G_{Z_{-0}^*}(u) f_{Z_0^{\min}}(u)$. We will then find $f_{Z_0^{\min}}(u)$ and to that end we first find the cdf $F_{Z_0^{\min}}(u)$ as follows.

$$\begin{aligned} F_{Z_0^{\min}}(u) &= \Pr \{Z_0^{\min} \leq u\} \\ &= 2 \Pr \{U_0 \leq u, V_0 \geq U_0\} \\ &= 2 \int_0^u \Pr \{V_0 \geq t | U_0 = t\} f(t) dt \\ &= \frac{2}{P_{\mathcal{E}}} \int_0^u \left[F \left(t \tan \left(\frac{\pi}{4} + \mathcal{E} \right) \right) - F(t) \right] f(t) dt. \end{aligned}$$

Therefore,

$$f_{Z_0^{\min}}(u) = \frac{2}{P_{\mathcal{E}}} f(u) \left[F \left(u \tan \left(\frac{\pi}{4} + \mathcal{E} \right) \right) - F(u) \right].$$

Therefore, we have

$$g_{Z^*}(u) = \frac{2N}{P_{\mathcal{E}}} G_{Z_{-0}^*}(u) f(u) \left[F \left(u \tan \left(\frac{\pi}{4} + \mathcal{E} \right) \right) - F(u) \right]. \quad (4.34)$$

By substituting $G_{Z_{-0}^*}(u)$ using the expression (4.34) in (4.33), we get the expression (4.31). That concludes the case 1 of the proof.

We then moved to the case 2 of the proof. In this case, we write

$$f_{Z_{0^*}^{\max}|E_0, Z_{0^*}^{\min}}(v|u) = \frac{f_{Z_{0^*}^{\min}, Z_{0^*}^{\max}|E_0}(u, v)}{f_{Z_{0^*}^{\min}|E_0}(u)} = \frac{f_{Z_{L^*}^{\min}, Z_{L^*}^{\max}}(u, v)}{g_{Z^*}(u)}.$$

Therefore, by substituting the expressions (4.31) and (4.34) in the above expression, we get (4.32). That concludes the case 2 of the proof. $\square$

We now use Lemma 4.1 to prove Theorem 4.5. To that end, let us first define $U_l = |\operatorname{Re}(\sqrt{\text{SNR}}H_l X)|$ and $V_l = |\operatorname{Im}(\sqrt{\text{SNR}}H_l X)|$ which are the distances to the relevant decision boundaries given that $(U_l, V_l) \in S_{\frac{\pi}{2^n}}$ with n -bit quantization. Therefore, under Rayleigh fading, we have $f(u) = \frac{2}{\sqrt{\pi}\text{SNR}} \exp\left(-\frac{u^2}{\text{SNR}}\right)$, $F(u) = 1 - 2Q\left(\frac{\sqrt{2}u}{\sqrt{\text{SNR}}}\right)$ and $P_{\frac{\pi}{2^n}} = \frac{1}{2^{n-2}}$. We use Fig. 4.14 as a visual guide.

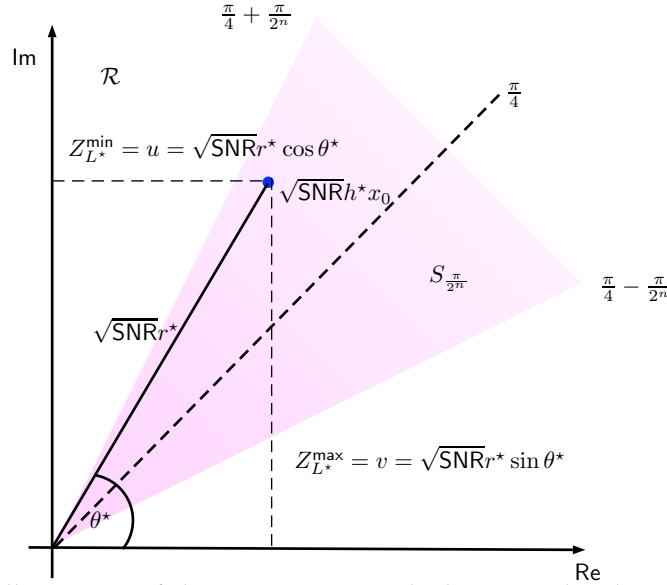


Figure 4.14: An illustration of the average SEP calculation under the selection strategy 3. Without loss of generality, let us assume that the x_0 is transmitted and the region of attraction is $\mathcal{R} = \{z \in \mathbb{C} : 0 \leq \operatorname{Arg}(z) \leq \frac{\pi}{2}\}$. Then, the distances to the decision boundaries are $Z_{L^*}^{\min} = u$ and $Z_{L^*}^{\max} = v$. Using polar coordinates, we can represent these distances as $Z_{L^*}^{\min} = \sqrt{\text{SNR}}r^* \cos \theta^*$ and $Z_{L^*}^{\max} = \sqrt{\text{SNR}}r^* \sin \theta^*$ where $r^* = |h^*|$ and $\theta^* = \operatorname{Arg}(h^*x_0)$.

By following similar steps to obtain (4.16), we can write

$$p(\text{SNR}, h^*) = Q(\sqrt{2}u) + Q(\sqrt{2}v) - Q(\sqrt{2}u)Q(\sqrt{2}v), \quad (4.35)$$

given that $Z_{L^*}^{\min} = u$ and $Z_{L^*}^{\max} = v$. Therefore, we then have to average the above expression by the joint pdf of $Z_{L^*}^{\min}$ and $Z_{L^*}^{\max}$ to obtain $p(\text{SNR})$. By substituting $g_{Z_0^*}(u)$ given in (4.34) to the expression (4.31), we write the joint pdf $Z_{L^*}^{\min}$ and $Z_{L^*}^{\max}$ as

$$f_{Z_{L^*}^{\min}, Z_{L^*}^{\max}}(u, v) = \begin{cases} \frac{2^{n+1}N}{\pi \text{SNR}} G_{Z_{-0}^*}(u) \exp\left(-\frac{u^2+v^2}{\text{SNR}}\right) & \text{if } (u, v) \in S_{\frac{\pi}{2^n}} \\ 0 & \text{otherwise,} \end{cases} \quad (4.36)$$

where $G_{Z_{-0}^*}(u) = \left\{F_{Z_0^{\min}}(u)\right\}^{N-1}$ with $F_{Z_0^{\min}}(u)$ being the cdf of $Z_0^{\min}$. We can find this cdf as

$$\begin{aligned} F_{Z_0^{\min}}(u) &= \frac{2}{P_{\frac{\pi}{2^n}}} \int_0^u \left\{F\left(t \tan\left(\frac{\pi}{4} + \frac{\pi}{2^n}\right) - F(t)\right)\right\} f(t) dt \\ &= \frac{2}{P_{\frac{\pi}{2^n}}} \int_0^u F\left(t \tan\left(\frac{\pi}{4} + \frac{\pi}{2^n}\right)\right) f(t) dt - \frac{F^2(u)}{P_{\frac{\pi}{2^n}}}. \end{aligned}$$

By substituting the corresponding functions, we get

$$\begin{aligned} F_{Z_0^{\min}}(u) &= \frac{2^n}{\sqrt{\pi \text{SNR}}} \int_0^u \left\{1 - 2Q\left(\frac{\sqrt{2}t}{\sqrt{\text{SNR}}} \tan\left(\frac{\pi}{4} + \frac{\pi}{2^n}\right)\right)\right\} \exp\left(-\frac{t^2}{\text{SNR}}\right) dt \\ &\quad - 2^{n-2} \left\{1 - 2Q\left(\frac{\sqrt{2}u}{\sqrt{\text{SNR}}}\right)\right\}^2 \end{aligned}$$

for general n -bit quantization. For $n = 2$ case, we can further simplify the above expression to

$$F_{Z_0^{\min}}(u) = 1 - 4Q^2\left(\frac{\sqrt{2}u}{\sqrt{\text{SNR}}}\right).$$

Therefore for $(u, v) \in S_{\frac{\pi}{2^n}}$, we have

$$f_{Z_{L^*}^{\min}, Z_{L^*}^{\max}}(u, v) = \frac{2^{n+1}N}{\pi \text{SNR}} \left\{F_{Z_0^{\min}}(u)\right\}^{N-1} \exp\left(-\frac{u^2+v^2}{\text{SNR}}\right). \quad (4.37)$$

By averaging the expression (4.35) over the above joint distribution, we get the $p(\text{SNR})$. To get the average SEP expression in Theorem 4.5 we use the change of variables $u = \sqrt{\text{SNR}}r^* \cos \theta^*$ and $v = \sqrt{\text{SNR}}r^* \sin \theta^*$.

4.9.6 Proof of Theorem 4.6

We start the proof by defining

$$p_1(\text{SNR}) = \int_{\frac{\pi}{4}-\frac{\pi}{2^H}}^{\frac{\pi}{4}+\frac{\pi}{2^H}} \int_0^\infty \mathcal{Q}\left(\sqrt{2\text{SNR}}r \cos \theta\right) r \exp(-r^2) [1 - \exp(-r^2)]^{N-1} dr d\theta,$$

$$p_2(\text{SNR}) = \int_{\frac{\pi}{4}-\frac{\pi}{2^H}}^{\frac{\pi}{4}+\frac{\pi}{2^H}} \int_0^\infty \mathcal{Q}\left(\sqrt{2\text{SNR}}r \sin \theta\right) r \exp(-r^2) [1 - \exp(-r^2)]^{N-1} dr d\theta,$$

and

$$p_3(\text{SNR}) = \int_{\frac{\pi}{4}-\frac{\pi}{2^H}}^{\frac{\pi}{4}+\frac{\pi}{2^H}} \int_0^\infty \mathcal{Q}\left(\sqrt{2\text{SNR}}r \cos \theta\right) \mathcal{Q}\left(\sqrt{2\text{SNR}}r \sin \theta\right) r \exp(-r^2) [1 - \exp(-r^2)]^{N-1} dr d\theta.$$

By using $\mathcal{Q}(x) = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \exp\left(-\frac{x^2}{2\sin^2 \beta}\right) d\beta$ from the Craig's formula and $[1 - \exp(-r^2)]^{N-1} = \sum_{i=0}^{N-1} \binom{N-1}{i} (-1)^i \exp(-ir^2)$ from the Binomial expansion, we can re-express $p_1(\text{SNR})$ as

$$p_1(\text{SNR}) = \int_0^{\frac{\pi}{2}} \int_{\frac{\pi}{4}-\frac{\pi}{2^H}}^{\frac{\pi}{4}+\frac{\pi}{2^H}} \int_0^\infty \sum_{i=0}^{N-1} \binom{N-1}{i} (-1)^i r \exp\left\{-\left(i+1 + \frac{\text{SNR} \cos^2 \theta}{\sin^2 \beta}\right) r^2\right\} dr d\theta d\beta,$$

where $\binom{N-1}{i} = \frac{(N-1)!}{i!(N-1-i)!}$. We further can reduce this integral using equation 3.321 of [84] as

$$p_1(\text{SNR}) = \int_0^{\frac{\pi}{2}} \int_{\frac{\pi}{4}-\frac{\pi}{2^H}}^{\frac{\pi}{4}+\frac{\pi}{2^H}} \sum_{i=0}^{N-1} \binom{N-1}{i} (-1)^i \frac{\sin^2 \beta}{2\{(i+1)\sin^2 \beta + \text{SNR} \cos^2 \theta\}} d\theta d\beta.$$

Using the expression 2.562 of [84] and after doing some mathematical manipulations we define $\rho_{\text{SNR}}(\beta)$ be a collection of functions indexed by SNR.

$$\rho_{\text{SNR}}(\beta) = \int_{\frac{\pi}{4}-\frac{\pi}{2^H}}^{\frac{\pi}{4}+\frac{\pi}{2^H}} \frac{1}{\{(i+1)\sin^2 \beta + \text{SNR} \cos^2 \theta\}} d\theta =$$

$$\frac{\tan^{-1} \left[\frac{\sqrt{(i+1)\sin^2 \beta \{(i+1)\sin^2 \beta + \text{SNR}\}} \left\{ \tan\left(\frac{\pi}{4} + \frac{\pi}{2^H}\right) - \tan\left(\frac{\pi}{4} - \frac{\pi}{2^H}\right) \right\}}{2(i+1)\sin^2 \beta + \text{SNR}} \right]}{\sqrt{(i+1)\sin^2 \beta \{(i+1)\sin^2 \beta + \text{SNR}\}}}.$$

Hence, we have

$$p_1(\text{SNR}) = \frac{1}{2} \int_0^{\frac{\pi}{2}} \sum_{i=0}^{N-1} \binom{N-1}{i} (-1)^i \sin^2 \beta \rho_{\text{SNR}}(\beta) d\beta.$$

Let us now, evaluate the above integral for $n = 2$ case first. When $n = 2$, the numerator of the function $\rho_{\text{SNR}}(\beta)$ becomes $\frac{\pi}{2}$ and $p_1(\text{SNR})$ reduces to

$$\begin{aligned} p_1(\text{SNR}) &= \frac{\pi}{4} \int_0^{\frac{\pi}{2}} \sum_{i=0}^{N-1} \binom{N-1}{i} (-1)^i \frac{\sin^2 \beta}{\sqrt{(i+1) \sin^2 \beta \{(i+1) \sin^2 \beta + \text{SNR}\}}} d\beta. \\ &= \text{SNR}^{-\frac{1}{2}} \frac{\pi}{4} \int_0^{\frac{\pi}{2}} \sum_{i=0}^{N-1} \binom{N-1}{i} (-1)^i \frac{\sin^2 \beta}{\sqrt{(i+1) \sin^2 \beta \left\{ \frac{(i+1) \sin^2 \beta}{\text{SNR}} + 1 \right\}}} d\beta. \end{aligned}$$

This function goes to $p_1(\text{SNR}) = \text{SNR}^{-\frac{1}{2}} \frac{\pi}{8} \sum_{i=0}^{N-1} \binom{N-1}{i} \frac{(-1)^i}{\sqrt{i+1}}$ as $\text{SNR} \rightarrow \infty$. Hence, as $\text{SNR} \rightarrow \infty$, we conclude that

$$p_1(\text{SNR}) = \text{SNR}^{-\frac{1}{2}} \Theta(1), \quad (4.38)$$

which implies $p_1(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-\frac{1}{2}}$ for $n = 2$.

We now consider the proof for $n \geq 3$ case. To that end, by using the Taylor series expansion for $\tan^{-1}(a) = a + \frac{1}{6}a^3 + \dots$ [84], we note that for high SNR regime,

$$\rho_{\text{SNR}}(\beta) = \frac{\tan\left(\frac{\pi}{4} + \frac{\pi}{2^n}\right) - \tan\left(\frac{\pi}{4} - \frac{\pi}{2^n}\right)}{2(i+1) \sin^2 \beta + \text{SNR}}, \quad (4.39)$$

by considering the first term of the series which has the smallest decay rate in term of SNR. Hence, we can simplify $p_1(\text{SNR})$ to

$$p_1(\text{SNR}) = \frac{1}{2} \left\{ \tan\left(\frac{\pi}{4} + \frac{\pi}{2^n}\right) - \tan\left(\frac{\pi}{4} - \frac{\pi}{2^n}\right) \right\} \int_0^{\frac{\pi}{2}} \frac{\prod_{i=0}^{N-2} \{2(i+1) \sin^2 \beta\}}{\prod_{i=0}^{N-1} \{2(i+1) \sin^2 \beta + \text{SNR}\}} d\beta,$$

after doing some mathematical manipulations. This function goes to

$$p_1(\text{SNR}) = \frac{1}{2} \left\{ \tan\left(\frac{\pi}{4} + \frac{\pi}{2^n}\right) - \tan\left(\frac{\pi}{4} - \frac{\pi}{2^n}\right) \right\} \text{SNR}^{-N} \int_0^{\frac{\pi}{2}} \prod_{i=0}^{N-2} \{2(i+1) \sin^2 \beta\} d\beta,$$

as $\text{SNR} \rightarrow \infty$. Further, $\int_0^{\frac{\pi}{2}} \prod_{i=0}^{N-2} \{2(i+1) \sin^2 \beta\} d\beta \leq \infty$ and hence, as $\text{SNR} \rightarrow \infty$, we conclude that

$$p_1(\text{SNR}) = \text{SNR}^{-N} \Theta(1), \quad (4.40)$$

which implies $p_1(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-N}$ for $n \geq 3$. The proof for $p_2(\text{SNR})$ and $p_3(\text{SNR})$ are similar, and we get

$$p_2(\text{SNR}) \stackrel{e}{=} \begin{cases} \text{SNR}^{-\frac{1}{2}} & n = 2 \\ \text{SNR}^{-N} & n \geq 3, \end{cases}$$

and

$$p_3(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-N},$$

for $n \geq 2$. However, we omit the proof to avoid repetition.

Since $p(\text{SNR}) = \frac{2^n N}{\pi} (p_1(\text{SNR}) + p_2(\text{SNR}) - p_3(\text{SNR}))$, we conclude that $p(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-\frac{1}{2}}$ for $n = 2$ and $p(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-N}$ for $n \geq 3$.

4.9.7 Proof of Theorem 4.7

We start the proof by defining,

$$p_1(\text{SNR}) = \int_{\frac{\pi}{4}}^{\frac{\pi}{4} + \frac{\pi}{2^n}} \int_0^\infty \mathcal{Q}(\sqrt{2\text{SNR}}r \cos \theta) r \exp(-r^2) \left(1 + 2^{n-2} - \frac{2^n \theta}{\pi}\right)^{N-1} dr d\theta,$$

$$p_2(\text{SNR}) = \int_{\frac{\pi}{4}}^{\frac{\pi}{4} + \frac{\pi}{2^n}} \int_0^\infty \mathcal{Q}(\sqrt{2\text{SNR}}r \sin \theta) r \exp(-r^2) \left(1 + 2^{n-2} - \frac{2^n \theta}{\pi}\right)^{N-1} dr d\theta,$$

and

$$p_3(\text{SNR}) = \int_{\frac{\pi}{4}}^{\frac{\pi}{4} + \frac{\pi}{2^n}} \int_0^\infty \mathcal{Q}(\sqrt{2\text{SNR}}r \cos \theta) \mathcal{Q}(\sqrt{2\text{SNR}}r \sin \theta) r \exp(-r^2) \left(1 + 2^{n-2} - \frac{2^n \theta}{\pi}\right)^{N-1} dr d\theta.$$

By using Craig's formula for $\mathcal{Q}(\cdot)$ and

$$\left(1 + 2^{n-2} - \frac{2^n \theta}{\pi}\right)^{N-1} = (1 + 2^{n-2})^{N-1} \sum_{i=0}^{N-1} \binom{N-1}{i} \left(-\frac{2^n \theta}{\pi(1 + 2^{n-2})}\right)^i$$

from the Binomial expansion, we can re-express $p_1(\text{SNR})$ as

$$\begin{aligned} p_1(\text{SNR}) &= (1 + 2^{n-2})^{N-1} \int_0^{\frac{\pi}{2}} \int_{\frac{\pi}{4}}^{\frac{\pi}{4} + \frac{\pi}{2^n}} \int_0^\infty \sum_{i=0}^{N-1} \binom{N-1}{i} \left(-\frac{2^n \theta}{\pi(1 + 2^{n-2})}\right)^i \\ &\quad r \exp\left\{-\left(\frac{\text{SNR} \cos^2 \theta}{\sin^2} + 1\right) r^2\right\} dr d\theta d\beta \\ &= \frac{(1 + 2^{n-2})^{N-1}}{2} \int_0^{\frac{\pi}{2}} \sin^2 \beta \int_{\frac{\pi}{4}}^{\frac{\pi}{4} + \frac{\pi}{2^n}} \sum_{i=0}^{N-1} \binom{N-1}{i} \left(-\frac{2^n}{\pi(1 + 2^{n-2})}\right)^i \\ &\quad \frac{\theta^i}{\text{SNR} \cos^2 \theta + \sin^2 \beta} d\theta d\beta. \end{aligned}$$

Let $g_{\text{SNR}}(\theta, \beta) = \frac{(1+2^{n-2})^{N-1}}{2} \sin^2 \beta \sum_{i=0}^{N-1} \binom{N-1}{i} \left(-\frac{2^n}{\pi(1+2^{n-2})}\right)^i \frac{\theta^i}{\text{SNR} \cos^2 \theta + \sin^2 \beta}$ be a collection of functions indexed by SNR. These functions increase to

$$g_\infty(\theta, \beta) = \text{SNR}^{-1} \frac{(1 + 2^{n-2})^{N-1}}{2} \sin^2 \beta \sum_{i=0}^{N-1} \binom{N-1}{i} \left(-\frac{2^n}{\pi(1 + 2^{n-2})}\right)^i \frac{\theta^i}{\cos^2 \theta}$$

as $\text{SNR} \rightarrow \infty$. Further, we have $\int_0^{\frac{\pi}{2}} \int_{\frac{\pi}{4}}^{\frac{\pi}{4} + \frac{\pi}{2^n}} g_\infty(\theta, \beta) d\theta d\beta < \infty$ by using the expression 2.643(9) of [84] for $N \geq 2$. Hence, as $\text{SNR} \rightarrow \infty$, we conclude that

$$p_1(\text{SNR}) = \text{SNR}^{-1} \Theta(1) \quad (4.41)$$

by using the monotone convergence theorem [86], which implies $p_1(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-1}$. The proof for $p_2(\text{SNR})$ and $p_3(\text{SNR})$ follow similar steps and we get $p_2(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-1}$ and $p_3(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-1}$. However, we omit the proof to avoid repetition.

Since $p(\text{SNR}) = \frac{2^{n+1}N}{\pi} (p_1(\text{SNR}) + p_2(\text{SNR}) - p_3(\text{SNR}))$, we conclude that $p(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-1}$ for $n \geq 2$ and $N \geq 2$.

We also note that, for $N = 1$ case, the system become a SISO and by following similar steps as given in the proof of Theorem 4 of [92], we have $\text{DVO} = \frac{1}{2}$ for $n = 2$

and $\text{DVO} = 1$ for $n \geq 3$.

4.9.8 DVO bounds for $N = 2$ and $n = 2$ case with Selection Criterion 3

Let us consider the case where $n = 2$. We start the proof by defining

$$p_1(\text{SNR}) = \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \int_0^\infty \mathcal{Q}(\sqrt{2\text{SNR}}r \cos \theta) \left[1 - 4\mathcal{Q}^2(\sqrt{2}r \cos \theta)\right]^{N-1} r \exp(-r^2) dr d\theta,$$

$$p_2(\text{SNR}) = \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \int_0^\infty \mathcal{Q}(\sqrt{2\text{SNR}}r \sin \theta) \left[1 - 4\mathcal{Q}^2(\sqrt{2}r \cos \theta)\right]^{N-1} r \exp(-r^2) dr d\theta,$$

and

$$p_3(\text{SNR}) = \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \int_0^\infty \mathcal{Q}(\sqrt{2\text{SNR}}r \cos \theta) \mathcal{Q}(\sqrt{2\text{SNR}}r \sin \theta) \left[1 - 4\mathcal{Q}^2(\sqrt{2}r \cos \theta)\right]^{N-1} r \exp(-r^2) dr d\theta.$$

By using $\mathcal{Q}(x) = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \exp\left(-\frac{x^2}{2\sin^2 \beta}\right) d\beta$, $\mathcal{Q}^2(x) = \frac{1}{\pi} \int_0^{\frac{\pi}{4}} \exp\left(-\frac{x^2}{2\sin^2 \beta}\right) d\beta$ [94] and $[1 - 4\mathcal{Q}^2(\sqrt{2}r \cos \theta)]^{N-1} = \sum_{i=0}^{N-1} \binom{N-1}{i} (-4)^i \mathcal{Q}^{2i}(\sqrt{2}r \cos \theta)$ from the Binomial expansion, we can re-express $p_1(\text{SNR})$ as

$$p_1(\text{SNR}) = \int_0^{\frac{\pi}{4}} \cdots \int_0^{\frac{\pi}{4}} \int_0^{\frac{\pi}{2}} \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \int_0^\infty \sum_{i=0}^{N-1} \binom{N-1}{i} \frac{(-4)^i}{\pi^{i+1}} r \exp\left\{-r^2 \left(1 + \frac{\text{SNR} \cos^2 \theta}{\sin^2 \beta} + \sum_{p=1}^i \frac{\text{sgn}(i) \cos^2 \theta}{\sin^2 \gamma_p}\right)\right\} dr d\theta d\beta d\gamma_1 \cdots d\gamma_i,$$

where $\text{sgn}(\cdot)$ is given as (1.2). Using [84], we can simplify the above expression as

$$p_1(\text{SNR}) = \int_0^{\frac{\pi}{4}} \cdots \int_0^{\frac{\pi}{4}} \int_0^{\frac{\pi}{2}} \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \sum_{i=0}^{N-1} \binom{N-1}{i} \frac{(-4)^i}{\pi^{i+1}} \frac{1}{2(1+B \cos^2 \theta)} d\theta d\beta d\gamma_1 \cdots d\gamma_i,$$

$$= \int_0^{\frac{\pi}{4}} \cdots \int_0^{\frac{\pi}{4}} \int_0^{\frac{\pi}{2}} \sum_{i=0}^{N-1} \binom{N-1}{i} \frac{(-4)^i}{\pi^{i+1}} \frac{\tan^{-1}(1+B)}{2\sqrt{1+B}} d\beta d\gamma_1 \cdots d\gamma_i,$$

where $B = \frac{\text{SNR}}{\sin^2 \beta} + \sum_{p=1}^i \frac{\text{sgn}(i)}{\sin^2 \gamma_p}$. Using the Taylor series expansion [84], in the high SNR

regime $\tan^{-1}(1+B)$ converges to $\frac{\pi}{2}$. Therefore,

$$p_1(\text{SNR}) = \int_0^{\frac{\pi}{4}} \cdots \int_0^{\frac{\pi}{4}} \int_0^{\frac{\pi}{2}} \sum_{i=0}^{N-1} \binom{N-1}{i} \frac{(-4)^i}{4\pi^i} \frac{1}{\sqrt{1+B}} d\beta d\gamma_1 \cdots d\gamma_i,$$

as $\text{SNR} \rightarrow \infty$.

Let us now consider the case for $N = 2$ case. Then the above expression becomes

$$p_1(\text{SNR}) = \frac{1}{4} \int_0^{\frac{\pi}{2}} \frac{1}{\sqrt{1 + \frac{\text{SNR}}{\sin^2 \beta}}} d\beta - \frac{1}{\pi} \int_0^{\frac{\pi}{4}} \int_0^{\frac{\pi}{2}} \frac{1}{\sqrt{1 + \frac{\text{SNR}}{\sin^2 \beta} + \frac{1}{\sin^2 \gamma_1}}} d\beta d\gamma_1.$$

Through numerical simulations, we can show that the decay rate of the above expression is 1 which is equal to $\frac{N}{2}$. However, it is difficult to show that the decay rate of this expression analytically. Therefore, we will provide an upper and lower bound to this expression in order to bound the decay rate. We will start with the upper bound by substituting the smallest value for $\sin^2 \gamma_1$ which is 0 in this case. Therefore, an upper bound $p_{1,u}(\text{SNR})$ for $p_1(\text{SNR})$ becomes

$$p_{1,u}(\text{SNR}) = \frac{1}{4} \int_0^{\frac{\pi}{2}} \frac{1}{\sqrt{1 + \frac{\text{SNR}}{\sin^2 \beta}}} d\beta,$$

and as $\text{SNR} \rightarrow \infty$

$$p_{1,u}(\text{SNR}) = \frac{1}{4} \text{SNR}^{-\frac{1}{2}} \int_0^{\frac{\pi}{2}} \sin \beta d\beta = \frac{1}{4} \text{SNR}^{-\frac{1}{2}}, \quad (4.42)$$

which concludes that

$$p_{1,u}(\text{SNR}) = \text{SNR}^{-\frac{1}{2}} \Theta(1), \quad (4.43)$$

which implies $p_{1,u}(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-\frac{1}{2}}$.

We now move to provide a lower bound $p_{1,l}(\text{SNR})$ for $p_1(\text{SNR})$ by substituting the

largest value for $\sin^2 \gamma_1$ which is $\frac{1}{2}$ in (4.42). Therefore, we have

$$p_{1,l}(\text{SNR}) = \frac{1}{\pi} \int_0^{\frac{\pi}{4}} \int_0^{\frac{\pi}{2}} \frac{1}{\sqrt{1 + \frac{\text{SNR}}{\sin^2 \beta}}} - \frac{1}{\sqrt{3 + \frac{\text{SNR}}{\sin^2 \beta}}} d\beta d\gamma_1.$$

By using several mathematical manipulations, we can write the above expression as

$$p_{1,l}(\text{SNR}) = \frac{1}{\pi} \int_0^{\frac{\pi}{4}} \int_0^{\frac{\pi}{2}} \frac{2 \sin^3 \beta d\beta d\gamma_1}{\sqrt{\text{SNR} + \sin^2 \beta} \sqrt{\text{SNR} + 3 \sin^2 \beta} \left(\sqrt{\text{SNR} + 3 \sin^2 \beta} + \sqrt{\text{SNR} + \sin^2 \beta} \right)}.$$

In the high SNR regime, we can write the above expression as

$$p_{1,l}(\text{SNR}) = \frac{1}{4} \text{SNR}^{-\frac{3}{2}} \int_0^{\frac{\pi}{2}} \sin^3 \beta d\beta = \frac{1}{6} \text{SNR}^{-\frac{3}{2}},$$

which concludes that

$$p_{1,l}(\text{SNR}) = \text{SNR}^{-\frac{3}{2}} \Theta(1), \quad (4.44)$$

which implies $p_{1,u}(\text{SNR}) \stackrel{e}{=} \text{SNR}^{-\frac{3}{2}}$.

Therefore, we have the DVO which is bounded as $\frac{1}{2} \leq \text{DVO} \leq \frac{3}{2}$ for $N = 2$ case. However, through simulations, we have the exact $\text{DVO} = 1$ for $N = 2$ case. In a similar manner, we can find lower bounds for different N values.

Chapter 5

Conclusion and Future Work

In this thesis, we presented a comprehensive and rigorous analytical investigation on the performance impact of using low-resolution phase quantization at the receiver of a wireless communication system when compared to traditional high-resolution systems. Our new analytical results accurately establish the factors that contribute to improve the performance of a low-resolution quantization based system.

5.1 Summary of Contributions

In Chapter 2, we considered a point-to-point wireless communication system with a low-resolution phase quantization. We derived the optimum ML detection rule for signal detection and obtained exact analytical expressions for the average SEP attained by the derived ML rule when the wireless channel is subjected to Nakagami- m fading. We established a fundamental ternary behaviour in the average SEP performance of a wireless communication system with low-resolution ADCs and M -PSK modulation. Based on the Nakagami- m fading model, we showed that the decay exponent of the average SEP is the same as that of an infinite-bit quantization, which is equal to m when n is larger than or equal to $\log_2 M + 1$. We also showed that it is equal to $\frac{1}{2}$ and 0 for $n = \log_2 M$ and $n < \log_2 M$, respectively.

In Chapter 3, We adopted the DQPSK modulation scheme to differentially encode the transmit data in a point-to-point wireless communication system with low-resolution phase quantization. We derived the optimum ML detection rule when the

detection is performed non-coherently by using two consecutive quantized observations and the exact analytical expressions for the average SEP when the channel is subjected to Rayleigh fading. We showed that we can achieve the same DVO as with infinite-bit quantization when the number of quantization bits $n \geq 4$ which is one bit additional compared to the case where full CSI is available at the receiver.

In Chapter 4, we considered a SIMO wireless communication system with low-resolution phase quantization. The wireless channel consists of a transmitter with one antenna and a receiver with N antennas. We derived the optimum ML detection rule and used Monte Carlo simulations to numerically analyze the average SEP performance under QPSK modulation when the channel is subjected to Rayleigh fading. We further proposed three sub-optimum but low-complexity antenna selection strategies and characterize their symbol error probability performance. We showed that the strategy that selects the diversity branch with channel rotated constellation points being furthest away from the decision boundary is the best strategy in terms of the SEP performance. We also establish that this selection strategy achieves the same diversity order with the optimum ML detector under the same operating conditions.

5.2 Future Research

Signal Detection

In this thesis, we considered a wireless channel model with flat-fading when deriving the detection rules. A potential direction for future research is to extend the ML and sub-optimum detection rules for a wireless channel model with frequency-selective fading. Further, low-complexity sub-optimum detectors which can reduce the computational complexity of the optimum ML detectors should be investigated.

In Chapter 4 of this thesis, we presented detection rules for a low-resolution quantization based SIMO system with selection combining. A potential direction for future research is to investigate the possibility of extending the work with maximal ratio combining and analyze the performance.

Channel Estimation

Developing accurate channel estimation techniques in low-resolution quantization based systems are essential to gain its full benefits. A potential direction for future research is to develop a channel estimation technique for a low-resolution based phase quantized system. Techniques to reduce the complexity of the estimator and the lengths of the training sequences should be important considerations. Further, low-complexity non-coherent detection techniques can be developed. The findings in Chapter 3 can be extended to differential M -ary modulation schemes under Nakagami- m fading to gain more insights in this regard.

Channel Capacity Analysis

The exact channel capacity of a low-resolution quantization based massive MIMO system is still unknown since the input distribution is unknown. To achieve this goal, a potential direction for future research is to find the channel capacity of a multi-bit phase quantized wireless communication system. To start with, one can use the system models provided in Chapters 2 and 4.

Multi-Antenna Systems

In this thesis, we considered SISO and SIMO wireless communication systems with low-resolution quantization. A potential direction for future research is to extend the analysis to MIMO, massive MIMO systems with multi-user detection. Further, in this thesis, we considered uncoded SEP for performance considerations. The analysis may be extended to coded systems as well.

Mixed-ADC Systems

In Chapter 2, we assumed full CSI available at the receiver. We suggested using a mixed-ADC architecture where high-resolution ADCs are used during the channel estimation stage and during the data transmission phase, the receiver switches to low-resolution operation by using fewer quantization bits. A potential direction for future research is to investigate the optimum way of deploying high-resolution

and low-resolution ADCs depending on SNRs while making full use of the available low-resolution quantizers during the channel estimation phase. Further, the energy efficiency trade-off should be considered in that sense.

Bibliography

- [1] D. M. Barnes, "The introduction of cellular radio in the United Kingdom," in *35th IEEE Vehicular Technology Conference*, vol. 35, 1985, pp. 147–152.
- [2] P. Gupta, "Evolution of mobile generations: 1G to 5G," in *International Journal For Technological Research in Engineering*, vol. 1, Nov. 2013, pp. 2347–4718.
- [3] "Ericsson mobility report," June 2020. [Online]. Available: <https://bit.ly/3kbcRxM>
- [4] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.
- [5] "Ericsson mobility report," Nov. 2017. [Online]. Available: <https://bit.ly/3k0b7qX>
- [6] Y. Li, C. Tao, A. Mezghani, A. L. Swindlehurst, G. S. Granados, and L. Liu, "Optimal design of energy and spectral efficiency tradeoff in one-bit massive MIMO systems," *CoRR*, vol. abs/1612.03271, 2016. [Online]. Available: <http://arxiv.org/abs/1612.03271>
- [7] R. Chen, L. Liu, K. Sayana, and H. Li, "Energy-efficient wireless communications with future networks and diverse devices," *Journal of Computer Networks and Communications*, 2013.
- [8] M. D. Renzo, H. Haas, A. Ghrayeb, S. Sugiura, and L. Hanzo, "Spatial modulation for generalized MIMO: Challenges, opportunities, and implementation," *Proceedings of the IEEE*, vol. 102, no. 1, pp. 56–103, Jan. 2014.

- [9] Q. Bai and J. A. Nossek, "Energy efficiency maximization for 5G multi-antenna receivers," *Trans. Emerging Telecommunications Technologies*, vol. 26, pp. 3–14, 2015.
- [10] B. Murmann, "ADC performance survey 1997-2020." [Online]. Available: <http://web.stanford.edu/~murmman/adcsurvey.html>
- [11] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.
- [12] Y. Li, B. Bakaloglu, and C. Chakrabartin, "A system level energy model and energy-quality evaluation for integrated transceiver front-ends," *IEEE Trans. Very Large Scale Integration Syst*, vol. 15, p. 90–103, Jan. 2007.
- [13] J. Zhang, L. Dai, X. Li, Y. Liu, and L. Hanzo, "On low-resolution ADCs in practical 5G millimeter-wave massive MIMO systems," *IEEE Commun. Mag.*, vol. 56, no. 7, pp. 205–211, Jul. 2018.
- [14] J. Choi, D. J. Love, D. R. Brown, and M. Boutin, "Quantized distributed reception for MIMO wireless systems using spatial multiplexing," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3537–3548, Jul. 2015.
- [15] J. Choi, J. Mo, and R. W. Heath, "Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2005–2018, May 2016.
- [16] J. Mo, P. Schniter, N. G. Prelcic, and R. W. Heath, "Channel estimation in millimeter wave MIMO systems with one-bit quantization," in *Proc. 48th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, California, Nov. 2014, pp. 957–961.
- [17] J. Choi, D. J. Love, and D. R. Brown, "Channel estimation techniques for quantized distributed reception in MIMO systems," in *Proc. 48th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, California, Nov. 2014, pp. 1066–1070.
- [18] M. T. Ivrlac and J. A. Nossek, "On MIMO channel estimation with single-bit signal-quantization," in *Proc. Int. ITG/IEEE Workshop on Smart Antennas (WSA)*, Feb. 2007, pp. 1–7.

- [19] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Aug 2017.
- [20] A. K. Saxena, I. Fijalkow, and A. L. Swindlehurst, "On one-bit quantized ZF precoding for the multiuser massive MIMO downlink," in *Proc. 2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Rio de Janeiro, Brazil, Jul. 2016, pp. 1–5.
- [21] —, "Analysis of one-bit quantized precoding for the multiuser massive MIMO downlink," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4624–4634, Sep. 2017.
- [22] A. Swindlehurst, A. Saxena, A. Mezghani, and I. Fijalkow, "Minimum probability-of-error perturbation precoding for the one-bit massive MIMO downlink," in *Proc. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, USA, Mar. 2017, pp. 6483–6487.
- [23] J. Mo and R. W. Heath, "High SNR capacity of millimeter wave MIMO systems with one-bit quantization," in *Proc. 2014 Information Theory and Applications Workshop (ITA)*, San Diego, California, USA, Feb. 2014, pp. 1–5.
- [24] —, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498–5512, Oct. 2015.
- [25] A. Mezghani and J. A. Nossek, "Analysis of 1-bit output noncoherent fading channels in the low SNR regime," in *Proc. 2009 IEEE International Symposium on Information Theory*, Seoul, Korea, Jun. 2009, pp. 1080–1084.
- [26] —, "On ultra-wideband MIMO systems with 1-bit quantized outputs: Performance analysis and input optimization," in *Proc. 2007 IEEE International Symposium on Information Theory*, Nice, France, Jun. 2007, pp. 1286–1289.

- [27] Y. Li, C. Tao, A. L. Swindlehurst, A. Mezghani, and L. Liu, "Downlink achievable rate analysis in massive MIMO systems with one-bit DACs," *IEEE Commun. Lett.*, vol. 21, no. 7, pp. 1669–1672, Jul. 2017.
- [28] S. Krone and G. Fettweis, "Capacity of communications channels with 1-bit quantization and oversampling at the receiver," in *2012 35th IEEE Sarnoff Symposium*, 2012, pp. 1–7.
- [29] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 4038–4051, 2017.
- [30] C. Wen, C. Wang, S. Jin, K. Wong, and P. Ting, "Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2541–2556, May 2016.
- [31] J. Singh and U. Madhow, "Phase-quantized block noncoherent communication," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2828–2839, Jul. 2013.
- [32] A. Wadhwa and U. Madhow, "Blind phase/frequency synchronization with low-precision ADC: A bayesian approach," in *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2013, pp. 181–188.
- [33] —, "Near-coherent QPSK performance with coarse phase quantization: A feedback-based architecture for joint phase/frequency synchronization and demodulation," *IEEE Trans. Signal Process.*, vol. 64, no. 17, pp. 4432–4443, 2016.
- [34] S. Wang, L. Zhang, Y. Li, J. Wang, and E. Oki, "Multiuser MIMO communication under quantized phase-only measurements," *IEEE Trans. Commun.*, vol. 64, no. 3, pp. 1083–1099, 2016.
- [35] A. Cay and D. C. Popescu, "On the probability of symbol error for two-dimensional signal constellations with non-uniform decision regions," in *2012 46th Annual Conference on Information Sciences and Systems (CISS)*, 2012, pp. 1–5.

- [36] P. Nazari, B. Chun, F. Tzeng, and P. Heydari, "Polar quantizer for wireless receivers: Theory, analysis, and CMOS implementation," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 3, pp. 877–887, 2014.
- [37] H. Wang, F. F. Dai, Z. Su, and Y. Wang, "Sub-sampling direct rf-to-digital converter with 1024-apsk modulation for high throughput polar receiver," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 4, pp. 1064–1076, 2020.
- [38] T. M. Lok and V. K.-W. Wei, "Channel estimation with quantized observations," in *Proc. 1998 IEEE International Symposium on Information Theory*, Cambridge, MA, USA, Aug. 1998, pp. 333–333.
- [39] J. Singh, S. Ponnuru, and U. Madhow, "Multi-Gigabit communication: the ADC bottleneck," in *2009 IEEE International Conference on Ultra-Wideband*, Vancouver, Canada, Sep. 2009, pp. 22–27.
- [40] O. Dabeer and U. Madhow, "Channel estimation with low-precision analog-to-digital conversion," in *2010 IEEE International Conference on Communications*, Cape Town, South Africa, May 2010, pp. 1–6.
- [41] "Common public radio interface (CPRI); interface specification, CPRI." [Online]. Available: <http://www.cpri.info/spec.html>
- [42] S. Park, O. Simeone, O. Sahin, and S. Shamai Shitz, "Fronthaul compression for cloud radio access networks: Signal processing advances inspired by network information theory," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 69–79, 2014.
- [43] "Mobile IoT in the 5G Future." [Online]. Available: <https://www.ericsson.com/assets/local/networks/documents/gsma-5g-mobile-iot.pdf>
- [44] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

- [45] J. H. Winters, J. Salz, and R. D. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Transactions on Communications*, vol. 42, no. 234, pp. 1740–1751, 1994.
- [46] A. Duel-Hallen, "Decorrelating decision-feedback multiuser detector for synchronous code-division multiple-access channel," *IEEE Trans. Commun.*, vol. 41, pp. 285–290, Feb. 1993.
- [47] X. Xie, R. Short, and C. Rushforth, "A family of suboptimum detectors for coherent multiuser communications," *IEEE J. Sel. Areas Commun.*, vol. 8, pp. 683–690, May 1990.
- [48] P. Sung and K. Chen, "A linear minimum mean square error multiuser receiver in Rayleigh-fading channels," *IEEE J. Sel. Areas Commun.*, vol. 14, pp. 1583–1594, Oct. 1996.
- [49] U. Madhow and M. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Commun.*, vol. 42, pp. 3178–3188, Dec. 1994.
- [50] H. Poor and S. Verdu, "Probability of error in MMSE multiuser detection," *IEEE Trans. Inf. Theory*, vol. 43, pp. 858–871, May 1997.
- [51] E. Andersen, "Asymptotic properties of conditional maximum likelihood estimators," *Journal of the Royal Statistical Society*, vol. 32, no. 2, pp. 283–301, 1970.
- [52] F. Edgeworth, "On the portable errors of frequency-constants," *Journal of the Royal Statistical Society*, vol. 71, pp. 499–512, Sep. 1908.
- [53] D. Seethaler, H. Artes, and F. Hlawatsch, "Detection techniques for MIMO spatial multiplexing systems," *e i Elektrotechnik und Informationstechnik*, vol. 122, Mar. 2005.
- [54] A. Mezghani, M. S. Khoufi, and J. A. Nossek, "Maximum likelihood detection for quantized MIMO systems," in *Proc. 2008 International ITG Workshop on Smart Antennas*, Feb. 2008, pp. 278–284.

- [55] S. Wang, Y. Li, and J. Wang, "Multiuser detection for uplink large-scale MIMO under one-bit quantization," in *Proc. 2014 IEEE International Conference on Communications (ICC)*, Jun. 2014, pp. 4460–4465.
- [56] —, "Multiuser detection in massive MIMO with quantized phase-only measurements," in *2015 IEEE International Conference on Communications (ICC)*, Jun. 2015, pp. 4576–4581.
- [57] —, "Convex optimization based multiuser detection for uplink large-scale MIMO under low-resolution quantization," in *Proc. 2014 IEEE International Conference on Communications (ICC)*, Jun. 2014, pp. 4789–4794.
- [58] Y. Jeon, N. Lee, S. Hong, and R. W. Heath, "One-bit sphere decoding for uplink massive MIMO systems with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4509–4521, 2018.
- [59] A. Mezghani, M. S. Khoufi, and J. A. Nossek, "A modified MMSE receiver for quantized MIMO systems," in *Proc. Int. ITG/IEEE Workshop on Smart Antennas (WSA)*, Vienna, Austria, 2007.
- [60] A. Mezghani, M. Rouatbi, and J. A. Nossek, "An iterative receiver for quantized MIMO systems," in *Proc. 2012 16th IEEE Mediterranean Electrotechnical Conference*, Yasmine Hammamet, Tunisia, Mar. 2012, pp. 1049–1052.
- [61] Y. Jeon, N. Lee, S. Hong, and R. Heath, "One-bit sphere decoding for uplink massive MIMO systems with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4509–4521, July 2018.
- [62] Y. Jeon, H. Do, S. Hong, and N. Lee, "Soft-output detection methods for sparse millimeter-wave MIMO systems with low-precision ADCs," *IEEE Trans. Commun.*, vol. 67, no. 4, pp. 2822–2836, April 2019.
- [63] Y. Jeon, N. Lee, and V. Poor, "Robust data detection for MIMO systems with one-bit ADCs: A reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1663–1676, March 2020.

- [64] Y. Jeon, H. Lee, and N. Lee, "Robust MLSD for wideband SIMO systems with one-bit ADCs: Reinforcement-learning approach," in *IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Kansas City, MO, USA, May 2018, pp. 1–6.
- [65] S. K. J. Chae and S. Hong, "Machine learning detectors for MU-MIMO systems with one-bit ADCs," *IEEE Access*, vol. 8, pp. 86 608–86 616, May 2020.
- [66] A. Mezghani, F. Antreich, and J. A. Nossek, "Multiple parameter estimation with quantized channel output," in *Proc. 2010 International ITG Workshop on Smart Antennas (WSA)*, Feb 2010, pp. 143–150.
- [67] J. Zhang, L. Dai, S. Sun, and Z. Wang, "On the spectral efficiency of massive MIMO systems with low-resolution ADCs," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 842–845, 2016.
- [68] J. Liu, Z. Luo, and X. Xiong, "Low-resolution ADCs for wireless communication: A comprehensive survey," *IEEE Access*, vol. 7, pp. 91 291–91 324, 2019.
- [69] J. Singh and U. Madhow, "Phase-quantized block noncoherent communication," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2828–2839, Jul. 2013.
- [70] H. Jedda, M. M. Ayub, J. Munir, A. Mezghani, and J. A. Nossek, "Power- and spectral efficient communication system design using 1-bit quantization," in *Proc. 2015 International Symposium on Wireless Communication Systems (ISWCS)*, Aug. 2015, pp. 296–300.
- [71] N. Liang and W. Zhang, "Mixed-ADC massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 983–997, April 2016.
- [72] L. Hanzo, M. El-Hajjar, and O. Alamri, "Near-capacity wireless transceivers and cooperative communications in the MIMO era: Evolution of standards, waveform design, and future perspectives," *Proc. IEEE Proc.*, vol. 99, no. 8, pp. 1343–1385, 2011.

- [73] S. Wang, Y. Li, and J. Wang, "Multiuser detection in massive spatial modulation MIMO with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2156–2168, 2015.
- [74] J. Zhang, L. Dai, Z. He, S. Jin, and X. Li, "Performance analysis of mixed-ADC massive MIMO systems over Rician fading channels," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1327–1338, 2017.
- [75] M. Nakagami, "The m -distribution - A general formula of intensity distribution of rapid fading," *Statistical Methods in Radio Wave Propagation - Pergamon Press*, pp. 7–36, 1960.
- [76] G. L. Stüber, *Principles of Mobile Communication (2Nd Ed.)*, 2nd ed. Norwell, MA, USA: Kluwer Academic Publishers, 2001.
- [77] S. Krone and G. Fettweis, "Fading channels with 1-bit output quantization: Optimal modulation, ergodic capacity and outage probability," in *2010 IEEE Information Theory Workshop*, Dublin, Ireland, Aug 2010, pp. 1–5.
- [78] R. G. Gallager, *Principles of Digital Communication*. New York, NY, USA: Cambridge University Press, 2008.
- [79] J. Mo, P. Schniter, and R. W. Heath, "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, March 2018.
- [80] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "One-bit massive MIMO: Channel estimation and high-order modulations," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, 2015, pp. 1304–1309.
- [81] Y. Li, C. Tao, L. Liu, G. Seco-Granados, and A. L. Swindlehurst, "Channel estimation and uplink achievable rates in one-bit massive MIMO systems," in *Proc. 2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Jul. 2016, pp. 1–5.

-
- [82] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: John Wiley & Sons, 1991.
- [83] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.
- [84] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, 7th ed. Academic Press, 2007.
- [85] N. C. Beaulieu and C. Cheng, "Efficient Nakagami-m fading channel simulation," *IEEE Trans. Veh. Technol.*, vol. 54, no. 2, pp. 413–424, Mar. 2005.
- [86] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York: McGraw-Hill, 1987.
- [87] R. W. Lucky and J. C. Hancock, "On the optimum performance of n -ary systems having two-degrees of freedom," *IRE Trans. Commun. Syst.*, vol. 10, no. 2, pp. 185–192, Jun. 1962.
- [88] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability*, 2nd ed. Nashua: Athena Scientific, 2008.
- [89] B. Picinbono, "On circularity," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3473–3482, Dec. 1994.
- [90] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5597–5625, Nov. 2012.
- [91] K. T. Fang, S. Kotz, and K. W. Ng, *Symmetric Multivariate and Related Distributions*, 1st ed. New York: CRC Press, 1990.
- [92] S. Gayan, R. Senanayake, H. Inaltekin, and J. Evans, "Low-resolution quantization in phase modulated systems: Optimum detectors and error rate analysis," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1000–1021, 2020.

-
- [93] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York, NY, USA: Cambridge University Press, 2005.
- [94] M. K. Simon, "Single integral representations of certain integer powers of the Gaussian Q-function and their application," *IEEE Commun. Lett.*, vol. 6, no. 12, pp. 532–534, 2002.