

ISLAM TANASH

# Approximations for Performance Analysis in Wireless Communications and Applications to Reconfigurable Intelligent Surfaces



ISLAM TANASH

Approximations for Performance Analysis in  
Wireless Communications and Applications  
to Reconfigurable Intelligent Surfaces

ACADEMIC DISSERTATION

To be presented, with the permission of  
the Faculty of Information Technology and Communication Sciences  
of Tampere University,  
for public discussion in the Auditorium TB109  
of the Tietotalo, Korkeakoulunkatu 1, Tampere,  
on 25<sup>th</sup> of November 2022, at 12 o'clock.

ACADEMIC DISSERTATION  
Tampere University  
Faculty of Information Technology and Communication Sciences  
Finland

*Responsible  
supervisor  
and Custos*

Associate Professor  
Taneli Riihonen  
Tampere University  
Finland

*Pre-examiners*

Associate Professor  
F. Javier Lopez-Martinez  
Universidad de Granada  
Spain

Adjunct Professor Alexandros-  
Apostolos A. Boulogeorgos  
University of Western Macedonia  
Greece

*Opponent*

Professor Olav Tirkkonen  
Aalto University  
Finland

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Copyright ©2022 Islam Tanash

Cover design: Roihu Inc.

ISBN 978-952-03-2613-5 (print)

ISBN 978-952-03-2614-2 (pdf)

ISSN 2489-9860 (print)

ISSN 2490-0028 (pdf)

<http://urn.fi/URN:ISBN:978-952-03-2614-2>



Carbon dioxide emissions from printing Tampere University dissertations have been compensated.

PunaMusta Oy – Yliopistopaino  
Joensuu 2022

To my sons, *Ghassan*, with whom this journey started,  
and *Hamzah*, with whom it concludes,  
*“You are my beautiful life’s work.”*



# PREFACE

The incredible life-changing journey, which started in 2018 as a doctoral researcher at the Unit of Electrical Engineering, Tampere University, Tampere, Finland, is finally coming to an end in 2022. The financial support for this thesis was mainly provided by Tampere University's Doctoral School and the Academy of Finland, which partially supported this research work under grants 310991/326448, 315858, 341489, and 346622. This work would not have been possible without their financial contribution.

First and foremost, I would like to express my heartfelt gratitude to Prof. Taneli Riihonen, my thesis supervisor and role model, whom I can describe as an energetic, committed, and enthusiastic professor. He has always supported, guided, and inspired me throughout my academic research with his vast knowledge and wealth of experience. I will be forever grateful for that.

I would also like to express my appreciation to my thesis pre-examiners, Prof. F. Javier Lopez-Martinez (Universidad de Granada, Spain) and Prof. Alexandros-Apostolos A. Boulogeorgos (University of Western Macedonia, Greece), for their constructive feedback, which helped me improve the quality of the thesis. I also extend my appreciation and gratitude to Prof. Olav Tirkkonen (Aalto University, Finland), who agreed to act as my thesis opponent.

I am also thankful to all my colleagues from Tampere University for all the great moments and the exciting discussions we had. A special thanks to my officemate and dear friend, M.Sc. Niloofar Okati, and my colleagues, M.Sc. Ruben Morales, M.Sc. Nachiket Ayir, and M.Sc. Sahan Liyanaarachchi, for all the laughs, advice, and moments we have shared over the years. In addition, I would like to thank my dear friends, Mouna Khiriji and Emma Ruotanen, who have been like sisters to me. I do not know what I would do without you.

I am fortunate to have a supportive and loving family who believed in me and kept my spirits and motivation high during my research journey. For that, I

express my heartfelt thanks to my amazing parents, Mohammad and Maryam, my wonderful husband, Ahmad, who has always been there for me throughout everything, and my beautiful children, Ghassan and Hamzah.

Tampere, September 2022

*Islam Mohammad Tanash*



# ABSTRACT

In the last few decades, the field of wireless communications has witnessed significant technological advancements to meet the needs of today's modern world. The rapidly emerging technologies, however, are becoming increasingly sophisticated, and the process of investigating their performance and assessing their applicability in the real world is becoming more challenging. That has aroused a relatively wide range of solutions in the literature to study the performance of the different communication systems or even draw new results that were difficult to obtain. These solutions include field measurements, computer simulations, and theoretical solutions such as alternative representations, approximations, or bounds of classic functions that commonly appear in performance analyses. Field measurements and computer simulations have significantly improved performance evaluation in communication theory. However, more advanced theoretical solutions can be further developed in order to avoid using the expensive and time-consuming wireless communications measurements, replace the numerical simulations, which can sometimes be unreliable and suffer from failures in numerical evaluation, and achieve analytically simpler results with much higher accuracy levels than the existing theoretical ones.

To this end, this thesis firstly focuses on developing new approximations and bounds using unified approaches and algorithms that can efficiently and accurately guide researchers through the design of their adopted wireless systems and facilitate the conducted performance analyses in the various communication systems. Two performance measures are of primary interest in this study, namely the average error probability and the ergodic capacity, due to their valuable role in conducting a better understanding of the systems' behavior and thus enabling systems engineers to quickly detect and resolve design issues that might arise. In particular, several parametric expressions of different analytical forms are developed to approximate or bound the Gaussian  $Q$ -function,

which occurs in the error probability analysis. Additionally, any generic function of the  $Q$ -function is approximated or bounded using a tractable exponential expression. Moreover, a unified logarithmic expression is proposed to approximate or bound the capacity integrals that occur in the capacity analysis. A novel systematic methodology and a modified version of the classical Remez algorithm are developed to acquire optimal coefficients for the accompanying parametric approximation or bound in the minimax sense. Furthermore, the quasi-Newton algorithm is implemented to acquire optimal coefficients in terms of the total error. The average symbol error probability and ergodic capacity are evaluated for various applications using the developed tools.

Secondly, this thesis analyzes a couple of communication systems assisted with reconfigurable intelligent surfaces (RISs). RIS has been gaining significant attention lately due to its ability to control propagation environments. In particular, two communication systems are considered; one with a single RIS and correlated Rayleigh fading channels, and the other with multiple RISs and non-identical generic fading channels. Both systems are analyzed in terms of outage probability, average symbol error probability, and ergodic capacity, which are derived using the proposed tools. These performance measures reveal that better performance is achieved when assisting the communication system with RISs, increasing the number of reflecting elements equipped on the RISs, or locating the RISs nearer to either communication node.

In conclusion, the developed approximations and bounds, together with the optimized coefficients, provide more efficient tools than those available in the literature, with richer capabilities reflected by the more robust closed-form performance analysis, significant increase in accuracy levels, and considerable reduction in analytical complexity which in turns can offer more understanding into the systems' behavior and the effect of the different parameters on their performance. Therefore, they are expected to lay the groundwork for the investigation of the latest communication technologies, such as RIS technology, whose performance has been studied for some system models in this thesis using the developed tools.

# CONTENTS

1	Introduction . . . . .	1
1.1	Motivation and Scope of the Thesis . . . . .	2
1.2	Contributions of the Thesis . . . . .	5
1.3	Author's Contributions to the Publications . . . . .	6
1.4	Structure of the Thesis. . . . .	7
2	Background . . . . .	9
2.1	Error Probability Analysis. . . . .	10
2.1.1	Definition of $Q$ -Function and its Applications . . . . .	14
2.1.2	Existing Approximations/Bounds and Applications . . . . .	16
2.2	Ergodic Capacity Analysis. . . . .	20
2.3	Reconfigurable Intelligent Surfaces . . . . .	22
2.3.1	Conventional SISO System Model with a Single RIS . . . . .	24
2.3.2	Generic SISO System Model with Multiple RISs . . . . .	26
3	Approximation Theory and Optimization . . . . .	29
3.1	Minimax Error Optimization . . . . .	30
3.1.1	Methods of Implementation. . . . .	33
3.1.2	Lower and Upper Bounds . . . . .	39
3.2	Total Error Optimization . . . . .	40
4	Gaussian $Q$ -Function . . . . .	43
4.1	Exponential-Type Approximations . . . . .	43
4.1.1	Functions of $Q$ -Function . . . . .	44
4.1.2	Minimax Error-Based Solution . . . . .	45
4.1.3	Quadrature-Based Solution . . . . .	49
4.2	Generalized Karagiannidis–Lioumpas Approximations . . . . .	50
4.2.1	Mathematical Form and Origin. . . . .	51

4.2.2	Minimax Error-Based Solution . . . . .	52
4.2.3	Total Error-Based Solution . . . . .	54
4.3	Applications and Performance Analysis . . . . .	54
4.3.1	Applications . . . . .	55
4.3.2	Performance Evaluation . . . . .	57
5	Capacity Integrals . . . . .	61
5.1	Mathematical Form and Origin . . . . .	61
5.2	Developed Approximation and Bounds. . . . .	63
5.2.1	Nakagami and Lognormal Capacity Integrals . . . . .	64
5.2.2	Minimax Error-Based Solution . . . . .	65
5.2.3	Quadrature-Based Solution . . . . .	68
5.3	Applications. . . . .	69
5.4	Performance Evaluation . . . . .	72
6	Reconfigurable Intelligent Surfaces . . . . .	75
6.1	RIS-Aided System with Spatially Correlated Channels . . . . .	75
6.1.1	Performance Analysis . . . . .	79
6.1.2	Performance Evaluation . . . . .	81
6.2	Systems with Multiple RISs over $\kappa - \mu$ Fading Channels . . . . .	84
6.2.1	Performance Analysis . . . . .	87
6.2.2	Performance Evaluation . . . . .	89
7	Conclusions and Future Work. . . . .	93
7.1	Conclusions and Main Results. . . . .	93
7.2	Future Work . . . . .	95
	References . . . . .	97
	Publication 1. . . . .	115
	Publication 2. . . . .	129
	Publication 3. . . . .	137
	Publication 4. . . . .	145
	Publication 5. . . . .	151

Publication 6. . . . .	159
Publication 7. . . . .	177
Publication 8. . . . .	185



## ORIGINAL PUBLICATIONS

- P1 I. M. Tanash and T. Riihonen, “Global minimax approximations and bounds for the Gaussian  $Q$ -function by sums of exponentials,” *IEEE Transactions on Communications*, vol. 68, no. 10, pp. 6514–6524, Oct. 2020. DOI: 10.1109/TCOMM.2020.3006902.
- P2 I. M. Tanash and T. Riihonen, “Remez exchange algorithm for approximating powers of the  $Q$ -function by exponential sums,” in *Proc. IEEE Vehicular Technology Conference (VTC)*, Apr. 2021, pp. 1–6. DOI: 10.1109/VTC2021-Spring51267.2021.9448807.
- P3 I. M. Tanash and T. Riihonen, “Quadrature-based exponential-type approximations for the Gaussian  $Q$ -function,” in *Proc. IEEE Vehicular Technology Conference (VTC)*, Apr. 2021, pp. 1–5. DOI: 10.1109/VTC2021-Spring51267.2021.9448918.
- P4 I. M. Tanash and T. Riihonen, “Improved coefficients for the Karagiannidis–Lioumpas approximations and bounds to the Gaussian  $Q$ -function,” *IEEE Communications Letters*, vol. 25, no. 5, pp. 1468–1471, May 2021. DOI: 10.1109/LCOMM.2021.3052257.
- P5 I. M. Tanash and T. Riihonen, “Generalized Karagiannidis–Lioumpas approximations and bounds to the Gaussian  $Q$ -function with optimized coefficients,” *IEEE Communications Letters*, vol. 26, no. 3, pp. 513–517, Mar. 2022. DOI: 10.1109/LCOMM.2021.3139372.
- P6 I. M. Tanash and T. Riihonen, “Tight logarithmic approximations and bounds for generic capacity integrals and their applications to statistical analysis of wireless systems,” *IEEE Transactions on Communications*, 2022. DOI: 10.1109/TCOMM.2022.3198435, in press.

P7 I. M. Tanash and T. Riihonen, “Ergodic capacity analysis of RIS-aided systems with spatially correlated channels,” in *Proc. IEEE International Conference on Communications (ICC)*, May 2022, pp. 3293–3298. DOI: 10.1109/ICC45855.2022.9839244.

P8 I. M. Tanash and T. Riihonen, “Link performance of multiple reconfigurable intelligent surfaces and direct path in general fading,” in *Proc. International Conference on Signal Processing and Communication Systems (ICSPCS)*, Dec. 2021, pp. 1–6. DOI: 10.1109/ICSPCS53099.2021.9660225.

### List of Figures

2.1	Constellation diagram of BPSK modulation scheme. . . . .	13
2.2	BPSK modulation transmission link. . . . .	13
2.3	A SISO wireless system with a single RIS. The S–RIS and RIS–D links consist of multiple propagation paths through the $M$ REs . . .	25
2.4	A SISO wireless system with $M$ RISs. Each S–RIS $_l$ and RIS $_l$ –D link consists of multiple propagation paths through the $M_l$ REs. For simplicity, we illustrate the multipath components via two RISs only.. . . .	27
3.1	The uniform error function, $d(x)$ , that corresponds to the best minimax approximation with degree $\hat{D} = 4$ . . . . .	33
3.2	The uniform error function, $r(x)$ , that corresponds to the best minimax lower and upper bounds with degree $\hat{D} = 4$ . . . . .	41
4.1	Comparison between (4.3) and the references approximations [27]–[29] in terms of the absolute error. . . . .	58
4.2	Comparison between (4.11), (4.3), and the references approximations, [27], [35], [41] in terms of the relative error. . . . .	59
4.3	Relative error of the exponential (4.3) and GKL (4.11) approximations and bounds. . . . .	59



5.1	Effect of increasing $N$ on the accuracy of (5.8) for the Nakagami and lognormal capacity integrals in terms of global absolute error. .	73
6.1	The outage probability for different values of $M$ and $\Omega_u$ with $\gamma_{\text{th}} = 10$ dB . . . . .	82
6.2	The average SEP for different values of $M$ with $\Omega_u = -110$ dB for QPSK and $\aleph$ -ASK for different values of $\aleph$ . . . . .	82
6.3	The average SEP versus the strength of the direct path ( $\Omega_u$ ) for different values of $\gamma_0$ with $M = 100$ and $\aleph = 8$ . . . . .	90
6.4	The ergodic capacity for different values of $M$ with $N = 2$ and $\Omega_u = -72.5$ dB. . . . .	91

*List of Tables*

2.1	PDF of the SNR for some common fading models . . . . .	10
2.2	SEP for different modulation schemes with coherent detection [6] .	15



# ABBREVIATIONS

$\aleph$ -ASK	$\aleph$ -ary amplitude shift keying
$\aleph$ -QAM	$\aleph$ -ary quadrature amplitude modulation
3G	third generation
4G LTE	fourth generation long-term evolution
5G	fifth generation
6G	sixth generation
AF	amplify-and-forward
AWGN	additive white Gaussian noise
BEP	bit error probability
BER	bit error rate
BFGS	Broyden–Fletcher–Goldfarb–Shanno
BPSK	binary phase shift keying
CDF	cumulative distribution function
CLT	central limit theorem
CSI	channel state information
D	destination
DF	decode-and-forward
FPGA	field-programmable gate array
GKL	generalized Karagiannidis–Lioumpas
i.i.d.	independent and identically distributed
i.n.i.d.	independent but non-identically distributed

KL	Karagiannidis–Lioumpas
LoS	line-of-sight
MGF	moment generating function
MIMO	multiple-input multiple-output
MISO	multiple-input single-output
mmwave	millimetre wave
NOMA	non-orthogonal multiple-access
PDF	probability density function
QoS	quality of service
QPSK	quadrature phase shift keying
RE	reflecting element
RIS	reconfigurable intelligent surface
S	source
SEP	symbol error probability
SER	symbol error rate
SIMO	single-input multiple-output
SISO	single-input single-output
SNR	signal-to-noise ratio
THz	terahertz communication

# SYMBOLS

$A$	channel response
$C$	Shannon capacity
$C(\cdot)$	generic capacity function
$\tilde{C}(x)$	approximation of generic capacity function
$C_m(\cdot)$	Nakagami capacity integral
$C_\sigma(\cdot)$	lognormal capacity integral
$\bar{C}$	ergodic capacity
$d$	absolute error
$d_V$	vertical height of RE
$d_H$	horizontal height of RE
$D$	decision region
$D[\cdot]$	determinant of a matrix
$\hat{D}$	degree of approximating function
$e$	both absolute and relative error collectively
$e_{\max}$	minimax absolute/relative error
$e_{\text{tot}}$	total absolute/relative error
$\bar{e}_{\text{tot}}$	mean absolute/relative error
$E_s$	energy per symbol
$E_b$	energy per bit
$E[\cdot]$	expectation operator
$f_C(c)$	PDF of instantaneous capacity

$\tilde{f}(x)$	approximation of function $f(x)$
$\mathbb{F}_{\hat{D}}$	space of continuous functions of degree $\leq \hat{D}$
$\mathbf{g}$	channel vector of RIS–D link
$\mathbf{g}_l$	channel vector of RIS <sub>l</sub> –D link
$\mathbf{G}(\cdot)$	gradient vector
$\mathbf{h}$	channel vector of S–RIS link
$\mathbf{h}_l$	channel vector of S–RIS <sub>l</sub> link
$\tilde{\mathbf{H}}$	approximation to the Hessian matrix $\mathbf{H}$
$\mathbf{J}(\cdot)$	Jacobian matrix
$M$	number of REs equipped on the RIS
$M_\gamma(\cdot)$	MGF associated with the instantaneous SNR
$\mathcal{M}$	number of distributed RISs
$n_b$	number of binary bits per symbol
$P_s$	SEP
$P_s(E; \gamma)$	conditional SEP
$\bar{P}_s$	average SEP
$p(\mathbf{u}, x)$	approximating polynomial
$\mathbb{P}_N$	space of polynomials of degree $\leq N$
$P_O$	outage probability
$\text{Pr}(\cdot)$	probability operator
$Q(x)$	Gaussian $Q$ -function
$\tilde{Q}(\cdot)$	approximation of the $Q$ -function
$\tilde{Q}_p(\cdot)$	approximation of integer power $p$ of the $Q$ -function
$\tilde{Q}_\Omega(\cdot)$	approximation of polynomial of the $Q$ -function
$r$	relative error
$R$	range of values of interest in interval $[R_1, R_2]$
$\mathbf{R}$	correlation matrix

$s$	transmitted signal
$t$	counter of Remez outer iterations
$u$	fading coefficient of the direct S–D link
$\mathbf{u}$	set of coefficients to be optimized
$\mathbf{u}^*$	set of optimized coefficients
$\text{Var}[\cdot]$	variance operator
$w$	AWGN
$x_k$	location of the $k$ th extremum point
$y$	received signal
$\gamma$	instantaneous SNR
$\gamma_0$	transmit SNR
$\gamma_{\text{th}}(\cdot)$	predefined SNR threshold value
$\bar{\gamma}$	average SNR
$\delta$	iteration step size of quasi-Newton method
$\varepsilon$	predefined stopping threshold for Remez algorithm
$\zeta$	reflection coefficient
$\zeta_l$	reflection coefficient of the $l$ th RIS
$\theta_i$	phase shift of the $i$ th RE
$\theta_{l,i}$	phase shift of the $i$ th RE equipped on the $l$ th RIS
$\Theta$	diagonal phase-shift matrix of the RIS
$\Theta_l$	diagonal phase-shift matrix of the $l$ th RIS
$\iota_0$	reference path loss at $\varpi_0$
$\lambda$	wavelength
$\Lambda$	area of each RE
$\mu_h$	average intensity attenuation of S–RIS link
$\mu_g$	average intensity attenuation of RIS–D link
$\xi_j$	path loss exponent

$\varpi_0$	reference distance
$\varpi_j$	distance of the corresponding link in the RIS-aided system
$\tau$	counter of Newton–Raphson and quasi-Newton iterations
$\psi(\cdot)$	probability density function
$\Psi_\gamma$	cumulative distribution function
$\Omega(Q(\cdot))$	polynomial of the $Q$ -function with degree $P$
$\Omega_u$	large-scale fading coefficient of the direct S-D link
$\aleph$	number of modulation states
$\ \cdot\ _\infty$	uniform norm or supremum norm
$[\cdot]^T$	transpose operator



# 1 INTRODUCTION

Wireless technology has seen significant evolution, from the third generation (3G) to the fourth-generation long-term evolution (4G LTE), and then to the fifth-generation (5G) era. Nevertheless, it still has immense potential to revolutionize our daily life and create a fully connected world over the following few decades. In particular, the focus has recently turned to develop solutions beyond 5G, i.e., the sixth generation (6G) and beyond. Future wireless networks are expected to overcome the shortages of the current technologies and provide higher data rates, higher system capacity, lower latency, higher bandwidth, and improved quality of service (QoS). In addition, rapid major technological breakthroughs with increased complexity are coming to light regularly, such as millimeter wave (mmwave) technologies, terahertz (THz) communication, intelligent communication environments, pervasive artificial intelligence, large-scale network automation, ambient backscatter communications, and cell-free massive multiple-input multiple-output (MIMO) communication networks [1].

The accurate prediction of these technologies' performance is a critical factor in the timely adoption of these technologies in real-world systems. More specifically, measuring their different performance metrics is an essential step toward a better understanding of their behavior in the real world. The complexity of the analytical solutions' performance depends on the complexity of the encountered system and channel models. Most wireless communication systems encounter signal attenuation with different variables, including time, geographic location, and radio frequency. Thus, the received signals have different strengths and phases. This leads to the concept of *fading*, which is considered a random process. Therefore, the corresponding communication channel is referred to as a fading channel, and the corresponding performance analysis is referred to as the statistical performance analysis since it includes mathematical averaging over the statistical characterization of the fading channel. Statistical performance

measures include average signal-to-noise ratio (SNR), outage probability, average symbol error probability (SEP), etc. In many cases, when evaluating the different statistical performance measures of a communication system, complicated integrals that cannot be solved in closed form in terms of elementary functions occur.

Generally, performance assessment can be conducted through field measurements, computer simulations, or closed-form analytical solutions. However, wireless communications measurements are expensive, time-consuming, and necessitate the need for collaboration to construct comprehensive systems. On the other hand, numerical simulations performed on the different software packages can sometimes suffer from instability and oscillating issues that might cause wrong evaluations of the encountered integrals. Moreover, some formulations can also suffer from the issue of overflow and underflow of the very small or large floating point values, causing failures in numerical evaluation. This has raised the need for improved analytical tools to enable the study of the different communication systems and contribute to the communication fundamentals rather than using time-consuming computer simulations and expensive field measurements. It is much safer and more reliable to compute a performance measure through trusted tabulated functions tested and verified in the different software packages (instead of direct numerical integration). In addition, the closed-form analytical expressions can sometimes provide insightful observations into the effect of the different system parameters on its performance, especially when the results can be made simpler with some convenient approximations or bounds. It can also facilitate the design and optimization of various communication systems due to the availability of explicit analytical expressions.

## 1.1 Motivation and Scope of the Thesis

The general objective of this thesis is to facilitate statistical performance analysis, ease its calculations, render new analytical solutions that were previously deemed unfeasible and provide new theoretical insights into the effect of the different system's parameters on its performance. This requires creating and developing new mathematical tools that enable researchers to find closed-form approximations, bounds, or even exact expressions for measuring the perfor-

mance of wireless communication systems with enough accuracy and reasonable complexity. Furthermore, the applicability and reliability of the proposed tools need to be verified by implementing them in analyzing the most recent and promising technologies in wireless communication.

Among the different performance measures used to study the various communication systems, this thesis focuses on the average error probability and the ergodic capacity measures. This is because these two are the most common to appear in the broad literature of performance analysis, the most challenging to evaluate, and the most revealing about the system's behavior. Therefore, it is essential to have tools that enable their evaluation in a tractable closed form with high accuracy. The main source of difficulty in evaluating the average error probability and the ergodic capacity is that the conditional error probability on the fading channel and the instantaneous channel capacity are generally non-linear functions of the instantaneous SNR.

In particular, the conditional error probability for coherent detection under additive white Gaussian noise (AWGN) channel is usually a polynomial of the so-called *Gaussian Q-function*. The Gaussian  $Q$ -function measures the tail probability of a standard normal random variable  $X$  having unit variance and zero mean, i.e.,  $Q(x) = \Pr(X \geq x)$ . Since the  $Q$ -function is an integral that cannot be solved in closed form and the corresponding average error probability requires working with integrals involving it, the error probability most often cannot be expressed in closed form in terms of elementary functions. This leads to the first scope of this thesis which is the Gaussian  $Q$ -function for which the presented contributions are inclined toward developing approximations and bounds for the classical mathematical function in order to characterize the error probability performance of the different communication systems in a more desirable analytical form. The importance of the  $Q$ -function is not only limited to communication theory, but also to many other fields of statistical sciences such as diffusion problems in heat, mass, and momentum transfer applications and various branches of mathematical physics [2].

The second scope of the thesis is the ergodic capacity which specifies the maximum transmission rate of reliable communication that can be achieved over time-varying channels. It is calculated by taking the expectation of the instantaneous channel capacity. More specifically, the ergodic capacity for AWGN

channels is the mean of the Shannon capacity, which is a logarithmic function of the SNR. Therefore, computing the ergodic capacity might result in complicated integrals that cannot be solved in terms of elementary functions and are specific to the studied system. In fact, most of the approximations and bounds or even exact closed-form expressions available in the literature are unique to the system under study, so a complete analysis is required for each system independently and using different mathematical steps. Furthermore, as far as the author is aware, no unified approach for analyzing the performance of any communication system in terms of ergodic capacity exists. Therefore, the thesis' contributions for this scope are inclined toward developing unified approximations and bounds for the ergodic capacity of any communication system.

This study also focuses on the different communication systems whose analyses employ the developed mathematical tools, i.e., the novel approximations and bounds. In particular, it does not limit their applicability to classical wireless systems that have already been studied in the vast literature, but it also explores their importance in one of the most promising and revolutionizing techniques for 6G, namely, the *reconfigurable intelligent surfaces* (RISs) technology which is the third and last scope of this thesis. The importance of this technology comes from its ability to achieve more control over the wireless environment, which has long been recognized as an uncontrollable communication medium that chaotically reflects the transmitted signals. Therefore, it provides significantly improved spectral and energy efficiency. The thesis' contributions in this scope tend toward evaluating the performance of RIS-aided systems in terms of the different performance measures through the use of the developed tools. Other secondary concepts are also covered in this thesis and are used to achieve the targeted objectives.

The following are the main research questions that further elaborate on the objective and scope of the thesis.

1. How to enable the analysis of complicated systems and allow deriving new results that have been typically unobtainable?
2. How to analytically simplify error probability analysis to produce simple-form solutions with minimal accuracy loss?
3. How to unify capacity analysis in the different communication systems and render highly accurate yet tractable closed-form capacity expressions?

4. What benefits do the developed approximations and bounds have for communications, and what kind of communication systems can they be implemented at to enable performance evaluation?

## 1.2 Contributions of the Thesis

The general contribution of this study is a collection of tight and tractable approximations and bounds that facilitate statistical performance analyses, including the analysis of the RIS-aided systems. The thesis' major contributions can be summarized as follows.

Several tractable approximations and lower/upper bound of different mathematical forms are developed in this thesis for the Gaussian  $Q$ -function together with optimizing them in terms of the minimax absolute/relative error and the total absolute/relative error to sufficiently increase the approximations' or bounds' accuracy. A novel systematic methodology and a modified version of the classical Remez algorithm are developed in order to implement the minimax error optimization. Furthermore, simple approximations and bounds for integer powers and polynomials of the  $Q$ -function or any generic function of the  $Q$ -function that accepts a Taylor series expansion are developed.

Unified approximations/bounds that enable the evaluation of the ergodic capacity in any communication system are provided, together with redeveloping the novel systematic methodology and the modified Remez algorithm of the  $Q$ -function in such a way as to make them comply with the capacity analysis. The high efficiency and applicability of the proposed approximations/bounds are extensively validated through theoretical analyses, simulations, and application examples.

The proposed approximations/bounds are implemented in this thesis to analyze the performance of RIS-aided systems. Particularly, two different system models are considered with developing a different mathematical framework for each system setting to characterize each of the studied systems' end-to-end equivalent channels. Performance analysis for each system model is conducted in terms of outage probability, average SEP, and ergodic capacity, for which analytical expressions in closed form are derived, and the effect of the different system parameters is studied.

### 1.3 Author's Contributions to the Publications

This thesis comprises eight published scientific articles in total. The author of this thesis, referred to as the Author in what follows, is the primary contributor to all of the work presented herein while working under the supervision and guidance of Prof. Taneli Riihonen, who suggested the Author's research topic for her doctoral studies. Publications are divided into three categories based on the research challenge being addressed.

Publication [P1] was the starting point of this thesis, after which the publications [P2]–[P5] were produced to tackle the same research problem of finding approximations and bounds for the Gaussian  $Q$ -function using novel implementation methods. The original ideas of the publications [P1]–[P5] were developed by the Author with the help of Prof. Riihonen. The Author implemented the encountered optimization methodologies, acquired the data sets of optimized coefficients, conducted the theoretical and numerical analyses, and performed simulations for all of these publications. Prof. Riihonen supervised the Author during the whole writing process by providing important tips to improve the quality of the papers together with revising the manuscripts.

For publication [P6], Prof. Taneli Riihonen suggested to consider ergodic capacity after error probability analysis. The Author has considerably expanded this idea to its current form, in which a unified method is proposed to facilitate the capacity analysis of any wireless communication system. The Author was responsible for planning and implementing the study, in terms of methodology, analysis, acquiring the data sets, simulations, and writing, under the guidance of Prof. Riihonen, who also helped to copyedit the manuscript.

The topics of publications [P7] and [P8], which are concerned with applying the novel tools developed in publications [P1]–[P6], were proposed solely by the Author to analyze the performance of RIS-aided networks. The Author's contributions also include everything else, i.e., the development of the presented methodologies, their implementation, preparing numerical results, and writing the whole papers, which were only reviewed internally by Prof. Taneli Riihonen for valuable feedback that helped strengthen the quality of the papers.

## 1.4 Structure of the Thesis

The remainder of the thesis is organized as follows.

**Chapter 2** defines some important concepts that are needed in the technical chapters that follow. It also provides a brief overview of the different approximations and bounds available in the literature for the Gaussian  $Q$ -function along with simple tractability and accuracy comparisons and a few application examples from the literature. In addition, it presents a survey on related exact and approximated capacity analyses of various communication systems pertinent to the scope of this research. Finally, this chapter overviews the most relevant RIS-aided systems from the literature to this study and presents their system models.

**Chapter 3** introduces two optimization criteria to be adopted in the following two chapters for constructing optimal approximations of the considered performance measures. Specifically, this chapter starts by introducing the minimax approximation theory, as well as proposing two novel iterative implementation approaches and a brief discussion of the initial guesses required to initiate these methods. These approaches are then reformulated to construct novel lower and upper bounds. The concept of total error optimization and how to implement it is also presented in this chapter.

**Chapter 4** studies the Gaussian  $Q$ -function for which the main novel contribution lies in developing tight approximations and bounds for it or its polynomials. The chapter presents two types of approximations/bounds for which the mathematical frameworks contributed in Chapter 3 are used herein to optimize their coefficients in terms of minimax and total errors. It also presents an overview and a performance comparison of all the commonly-used numerical integration techniques to approximate the  $Q$ -function.

**Chapter 5** studies the ergodic capacity performance metric for which novel approximations and bounds are developed to enable the accurate evaluation of ergodic capacity in any communication system in a unified form. To optimize the corresponding coefficients, the mathematical frameworks presented in Chapter 3 are also employed herein. Lastly, an extensive overview of the wide range of fundamental and recent applications is presented to demonstrate their applicability.

**Chapter 6** studies the performance of two reconfigurable intelligent surfaces-aided system models in terms of outage probability, average SEP, and ergodic capacity, for which analytical expressions in closed form are derived using the developed tools proposed in the previous two chapters.

**Chapter 7** concludes the thesis with a discussion that summarizes the research's primary outcomes and contributions. Future research directions are suggested.



## 2 BACKGROUND

This chapter provides an overview of the existing approximations and bounds for the Gaussian  $Q$ -function and their applications in calculating error probabilities for various communication systems. It also presents an overview of the most relevant capacity analyses conducted in the literature, as well as a brief overview of the RIS technology, which is a critical and relevant application of the proposed novel tools. Some baseline concepts regarding signal and fading models are also clarified.

In particular, statistical performance analysis is often performed in the presence of fading, as discussed in Chapter 1. Fading is a random process that occurs in wireless links due to some impairments such as multipath propagation and shadowing that effect the transmitted radio signal strength and attenuate it. Wireless channels that have these properties are called fading channels. Table 2.1 lists the probability density function (PDF) denoted by  $\psi(\cdot)$  of the instantaneous SNR per symbol denoted by  $\gamma$ , with average SNR denoted by  $\bar{\gamma}$ , for the most relevant fading models to the contents of this dissertation, namely Rayleigh, Nakagami- $m$ , Rician,  $\eta$ - $\mu$ ,  $\kappa$ - $\mu$ ,  $\alpha$ - $\mu$ , and lognormal fading models.

For Nakagami- $m$ ,  $\frac{1}{2} \leq m \leq \infty$  is the fading parameter, for the Rice fading,  $K$  is the Rician factor, and for Fisher-Snedecor  $\mathcal{F}$ ,  $m$  is the fading severity parameter,  $m_s$  is the shadowing parameter and  $B(\cdot, \cdot)$  is the beta function [3]. For  $\eta$ - $\mu$  fading of format 1,  $0 < \eta < \infty$  is the scattered wave power ratio between the inphase and quadrature components of each multipath cluster,  $h = (2 + \eta^{-1} + \eta)/4$  and  $H = (\eta^{-1} - \eta)/4$ , whereas for  $\eta$ - $\mu$  fading of format 2,  $-1 < \eta < 1$  is the correlation coefficient between the inphase and quadrature components of each multipath cluster,  $h = \frac{1}{(1-\eta^2)}$  and  $H = \eta/(1 - \eta^2)$  with  $\mu$  being the number of multipath clusters in the fading environment for both formats. For  $\kappa$ - $\mu$  and  $\alpha$ - $\mu$  fading,  $\kappa > 0$  is the ratio between the total power of the dominant components and the total power of the scattered waves,  $\alpha$  is

**Table 2.1** PDF of the SNR for some common fading models

Fading Model	PDF ( $\psi_\gamma(\gamma)$ )
Rayleigh [5]	$\frac{1}{\bar{\gamma}} \exp\left(-\frac{\gamma}{\bar{\gamma}}\right)$
Nakagami- $m$ [6]	$\frac{m^m \gamma^{m-1}}{\bar{\gamma}^m \Gamma(m)} \exp\left(-\frac{m\gamma}{\bar{\gamma}}\right)$
Rician [6]	$\frac{(1+k)e^{-k}}{\bar{\gamma}} \exp\left(-\frac{(1+k)\gamma}{\bar{\gamma}}\right) I_0\left(2\sqrt{\frac{k(1+k)\gamma}{\bar{\gamma}}}\right)$
Fisher-Snedecor $\mathcal{F}$ [7]	$\frac{m^m (m_s-1)^{m_s} \bar{\gamma}^{m_s} \gamma^{m-1}}{B(m, m_s) (m\gamma + (m_s-1)\bar{\gamma})^{m+m_s}}$
$\eta - \mu$ [8]	$\frac{2\sqrt{\pi}\mu^{\mu+\frac{1}{2}} h^\mu \gamma^{\mu-\frac{1}{2}}}{\Gamma(\mu) H^{\mu-\frac{1}{2}} \bar{\gamma}^{\mu+\frac{1}{2}}} \exp\left(-\frac{2\mu\gamma h}{\bar{\gamma}}\right) I_{\mu-\frac{1}{2}}\left(\frac{2\mu H\gamma}{\bar{\gamma}}\right)$
$\kappa - \mu$ [8]	$\frac{\mu(1+\kappa)^{\frac{\mu+1}{2}} \gamma^{\frac{\mu-1}{2}}}{\kappa^{\frac{\mu-1}{2}} \exp(\mu\kappa) \bar{\gamma}^{\frac{\mu+1}{2}}} \exp\left(-\frac{\mu(1+\kappa)\gamma}{\bar{\gamma}}\right) I_{\mu-1}\left(2\mu\sqrt{\frac{\kappa(1+\kappa)\gamma}{\bar{\gamma}}}\right)$
$\alpha - \mu$ [9]	$\frac{\alpha}{2\Gamma(\mu)} \left(\frac{\Gamma(\mu+\frac{\alpha}{2})}{\bar{\gamma}\Gamma(\mu)}\right)^{\frac{\alpha\mu}{2}} \gamma^{\frac{\alpha\mu}{2}-1} \exp\left(-\gamma^\alpha \left(\frac{\Gamma(\mu+\frac{\alpha}{2})}{\bar{\gamma}\Gamma(\mu)}\right)^{\alpha/2}\right)$
lognormal [10]	$\frac{1}{\sqrt{2\pi}\sigma\gamma} \exp\left(-\frac{(\log_e(\gamma)-\eta_\gamma)^2}{2\sigma^2}\right)$

the nonlinearity parameter, and  $\mu > 0$  is the number of multipath clusters and describes the severity of fading process. For the lognormal fading,  $\eta_\gamma$  and  $\sigma$  are the mean and the standard deviation of the corresponding instantaneous SNR's natural logarithm, respectively. The function  $I_\alpha(\cdot)$  in the Rician,  $\eta - \mu$  and  $\kappa - \mu$  fading is the modified Bessel function of the first kind and order  $\alpha$  [4, Eq. 9.6.12].

## 2.1 Error Probability Analysis

In general, the error probability of a communication system is defined as the probability that a random variable  $X$  exceeds a certain value  $\epsilon$ , i.e.,  $\Pr(X \geq \epsilon)$  [11]. Thus, the symbol error probability (SEP) denoted by  $P_s$  is the error probability in the transmission of a single symbol. Since this study is concerned with error performance analyses over AWGN channels, a brief discussion of various types of digital modulation schemes with coherent detection is given in this section. In particular, modulation is the act of changing one or more features of a periodic waveform, known as the carrier signal, in accordance with a distinct signal, known as the modulation or message signal, which often carries

information to be sent. The goal of modulation is to imprint information on the carrier wave, which is then utilized to transport the information to another site. Modulation schemes can be analog or digital based on the message/modulation signal being continuous (sine wave) or discrete (square wave) [11].

On the other hand, detection or, alternatively, demodulation is the process of recovering message signals from the received signals. Coherent detection refers to the case where the optimal receiver, which minimizes the possibility of disagreement between the sent and detected messages, has perfect knowledge of the amplitude, phase, frequency, or any combination of them based on the modulation scheme being used, i.e., based on the attributes of the carrier being modulated [6]. The derivation of the SEP performance with coherent detection for one of the simplest digital modulation schemes, namely the binary phase shift keying (BPSK), is presented herein in detail to serve as a stepping stone for understanding the derivation of SEP for other relatively complex modulation schemes.

The constellation diagram of the BPSK system shown in Fig. 2.1 consists of two signal waveforms  $s_0$  and  $s_1$  that correspond to two phases separated by  $180^\circ$  to map the binary digits 0 and 1, respectively. The two signal waveforms, which have equal transmission probabilities ( $\Pr(s_0) = \Pr(s_1) = \frac{1}{2}$ ), are located at equal distance of  $\sqrt{E_s}$  from the origin which forms the decision boundary of zero between the decision regions  $D_1$  and  $D_2$ . The parameter  $E_s$  is the energy per symbol, which is equal herein to the energy per bit  $E_b$  since each symbol consists of one bit only.

The transmission link between transmitter and receiver in a BPSK system is illustrated in Fig. 2.2. The received signal is

$$y = s + w, \quad (2.1)$$

where  $s \in \{s_0, s_1\}$ , and  $w$  is the AWGN with zero mean and variance  $\frac{N_0}{2}$ . The PDFs of  $y$  conditioned on the two transmitted signal waveforms are

$$\psi(y | s_0) = \frac{1}{\sqrt{\pi N_0}} e^{-\frac{(y + \sqrt{E_b})^2}{N_0}}, \quad (2.2)$$

$$\psi(y | s_1) = \frac{1}{\sqrt{\pi N_0}} e^{-\frac{(y - \sqrt{E_b})^2}{N_0}}. \quad (2.3)$$

In particular, an error occurs in BPSK when  $y$  is not in  $D_0$  while  $s_0$  is transmitted, or when  $y$  is not in  $D_1$  while  $s_1$  is transmitted. The symbol error probability is calculated using [11, Eq. 4.1-13] as

$$P_s = \Pr(s_0) \Pr(e | s_0) + \Pr(s_1) \Pr(e | s_1), \quad (2.4)$$

where  $\Pr(e | s_0)$  and  $\Pr(e | s_1)$  denote respectively the error probability when  $s_0$  and  $s_1$  are transmitted, and are calculated using (2.2) and (2.3) according to [11, Eq. 4.1-14] as

$$\Pr(e | s_0) = \int_0^{\infty} \psi(y | s_0) dy = \frac{1}{\sqrt{\pi N_0}} \int_0^{\infty} e^{-\frac{(y+\sqrt{E_b})^2}{N_0}} dy = Q\left(\sqrt{\frac{2E_b}{N_0}}\right), \quad (2.5)$$

and

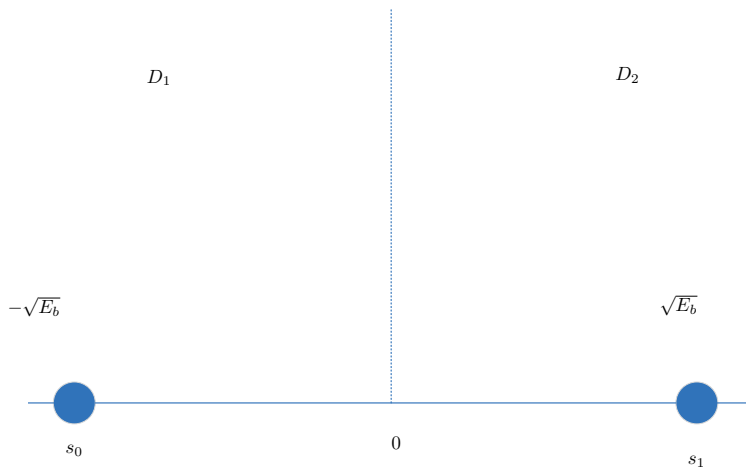
$$\Pr(e | s_1) = \int_{-\infty}^0 \psi(y | s_1) dy = \frac{1}{\sqrt{\pi N_0}} \int_{-\infty}^0 e^{-\frac{(y-\sqrt{E_b})^2}{N_0}} dy = Q\left(\sqrt{\frac{2E_b}{N_0}}\right). \quad (2.6)$$

After substituting (2.5), (2.6), and  $\Pr(s_0) = \Pr(s_1) = \frac{1}{2}$  in (2.4),

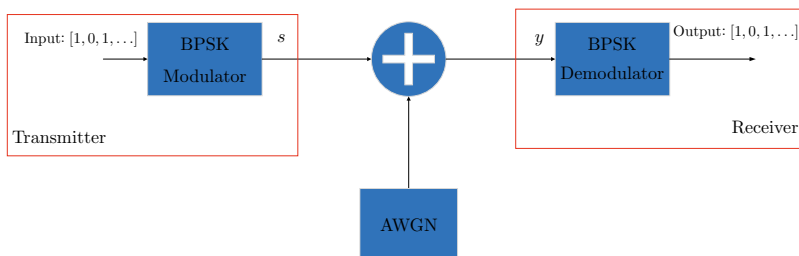
$$P_s = Q\left(\sqrt{\frac{2E_b}{N_0}}\right). \quad (2.7)$$

The function  $Q(\cdot)$  in (2.7) is the Gaussian  $Q$ -function which is the first scope of this study and is defined next in Section 2.1.1. The SEP for other digital communication systems that employ coherent detection, namely  $\aleph$ -ary amplitude shift keying ( $\aleph$ -ASK),  $\aleph$ -ary quadrature amplitude modulation ( $\aleph$ -QAM), quadrature phase shift keying (QPSK), differentially encoded BPSK, and differentially encoded QPSK, are given in Table 2.2, for which  $\aleph = 2^{n_b}$  is the number of modulation states with  $n_b$  being the number of binary bits per symbol. Following this discussion, it is evident the crucial importance of the Gaussian  $Q$ -function in communication theory where the Gaussian/normal distribution is frequently encountered. This importance can be best seen in error probability analysis for various communication systems such as those in Table 2.2.

When fading is present, the received instantaneous signal power is attenuated by the square of the fading amplitude. Therefore, conditioned on the



**Figure 2.1** Constellation diagram of BPSK modulation scheme.



**Figure 2.2** BPSK modulation transmission link.

fading, the SEP of any of the modulations schemes considered in Table 2.2 is obtained by replacing  $E_s/N_0$  by the instantaneous SNR per symbol,  $\gamma$ . The SEP over the non-fading AWGN channel is actually equivalent to the conditional SEP in the presence of fading and is denoted as  $P_s(E; \gamma)$ . Thus, the average SEP over fading is calculated as

$$\bar{P}_s = \int_0^\infty P_s(E; \gamma) \psi_\gamma(\gamma) d\gamma, \quad (2.8)$$

where  $\psi_\gamma(\gamma)$  is the PDF of the instantaneous SNR. In general, the average SEP decreases with increasing  $\bar{\gamma}$ . For fading channels, the change in slope of the probability of error defines the diversity gain, which occurs due to the usage of some diversity scheme. More specifically, diversity gain is the reduction in the transmitted power required to achieve a certain performance criterion, e.g., a certain SEP level, when using a diversity scheme [12]. A diversity scheme refers to any method that enhances the reliability of a communication signal by utilizing two or more communication channels with different properties, such as the RIS technology.

Since  $P_s(E; \gamma)$  in (2.8) for coherent detection is generally a polynomial of the Gaussian  $Q$ -function, which itself is integral that does not have a closed-form solution, complicated integrals will occur when evaluating the average error probabilities for the different modulation/detection schemes and the various fading channel models. This motivates the need for approximations and bounds to substitute the Gaussian  $Q$ -function with closed-form expressions that can ultimately allow for the evaluation of the encountered error probability measure in closed form.

### 2.1.1 Definition of $Q$ -Function and its Applications

The  $Q$ -function is often referred to as the Gaussian probability integral since it is the complement of the cumulative distribution function (CDF) corresponding to the standard Gaussian random variable  $X$  and is defined as

**Table 2.2** SEP for different modulation schemes with coherent detection [6]

Modulation Scheme	SEP ( $P_s$ )
$\aleph$ -ASK	$2 \left( \frac{\aleph-1}{\aleph} \right) Q \left( \sqrt{\frac{6E_s}{N_0(\aleph^2-1)}} \right)$
$\aleph$ -QAM	$4 \left( \frac{\sqrt{\aleph}-1}{\sqrt{\aleph}} \right) Q \left( \sqrt{\frac{3E_s}{N_0(\aleph-1)}} \right) - 4 \left( \frac{\sqrt{\aleph}-1}{\sqrt{\aleph}} \right)^2 Q^2 \left( \sqrt{\frac{3E_s}{N_0(\aleph-1)}} \right)$
QPSK	$2Q \left( \sqrt{\frac{E_s}{N_0}} \right) - Q^2 \left( \sqrt{\frac{E_s}{N_0}} \right)$
Differentially encoded BPSK	$2Q \left( \sqrt{\frac{2E_b}{N_0}} \right) - 2Q^2 \left( \sqrt{\frac{2E_b}{N_0}} \right)$
Differentially encoded QPSK	$4Q \left( \sqrt{\frac{E_s}{N_0}} \right) - 8Q^2 \left( \sqrt{\frac{E_s}{N_0}} \right) + 8Q^3 \left( \sqrt{\frac{E_s}{N_0}} \right) - 4Q^4 \left( \sqrt{\frac{E_s}{N_0}} \right)$

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{1}{2}t^2\right) dt \quad (2.9a)$$

$$= \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \exp\left(-\frac{1}{2\sin^2\theta}x^2\right) d\theta \quad [\text{for } x \geq 0]. \quad (2.9b)$$

For the Gaussian  $Q$ -function, the case  $x \geq 0$  is presumed throughout the thesis since the results can usually be extended to the negative real axis using the relation  $Q(x) = 1 - Q(-x)$ . The latter integral (2.9b) is the so-called Craig's formula [13], [14] obtained by manipulating the original results of [15], [16] to eliminate the function's argument from the lower limit of the integral in (2.9a) and thus reduce the analytical difficulty when this argument is dependent on other random parameters that eventually require statistical averaging over their probability distributions, i.e., solving multiple integrals. The  $Q$ -function is related to the error function  $\text{erf}(\cdot)$ , and the complementary error function  $\text{erfc}(\cdot)$ , which have equal importance in communication theory, respectively by

$$Q(x) = \frac{1}{2} - \frac{1}{2} \text{erf}\left(\frac{x}{\sqrt{2}}\right), \quad (2.10)$$

$$Q(x) = \frac{1}{2} \text{erfc}\left(\frac{x}{\sqrt{2}}\right). \quad (2.11)$$

Following the discussion in Section 2.1, the Gaussian  $Q$ -function is not only limited to error probability analysis of basic modulation schemes such as those mentioned in Table 2.2, but it also has a variety of applications in statistical performance analysis for various system models and fading channels [17]–[22]. For example, it occurs in the bit error probability (BEP) expressions for  $\mathfrak{N}$ -PSK with Gray, natural binary or folded binary bit-mappings in [17], [18], in bit error rate (BER) approximations for  $\mathfrak{N}$ -PSK and  $\mathfrak{N}$ -QAM with Gray code bit mapping based on a geometric approach in [19], in average BER expressions of  $\mathfrak{N}$ -QAM in flat Rayleigh fading with imperfect channel estimates in [20], in BER expressions of dual-hop orthogonal frequency division multiplexing-based amplify-and-forward (AF) relay system in [21], and in average SEP asymptotic expressions for composite lognormal with any small-scale fading in [22], as well as in measuring the performance of energy detectors for cognitive radio applications in [23], [24].

These applications often require working with complex integrals of the  $Q$ -function, mainly in the presence of fading, such as that in (2.8) whose  $P_s(E; \gamma)$  is actually a polynomial of the  $Q$ -function for coherent detection. In general, complicated integrals of the form

$$\int F(Q(f(\gamma))) Y(\gamma) d\gamma, \quad (2.12)$$

are expected to be encountered in the performance analysis of many communication systems. Above,  $Y(\gamma)$  is some integrable function and  $F(Q(f(\gamma)))$  is some well-behaved function of the  $Q$ -function that accepts a Taylor series expansion for  $0 \leq Q(f(\gamma)) \leq \frac{1}{2}$ . Since the  $Q$ -function itself cannot be expressed in closed form, (2.12) cannot be represented in closed form using elementary functions as well. As a result, finding tractable approximations and bounds for the  $Q$ -function becomes essential in order to make expression manipulations easier and to apply it to a wider range of analytical research. Several approximations and bounds are already available in the literature to meet this demand.

### 2.1.2 Existing Approximations/Bounds and Applications

A brief summary of the approximations and bounds already available in the literature for the Gaussian  $Q$ -function is provided herein with a basic assessment



of tractability and accuracy that are related through a trade-off relationship. The first and simplest known substitution for the  $Q$ -function is the so-called *Chernoff bound* which is in the form of a single exponential function [25] that was later tightened in [26] to become

$$Q(x) \leq \frac{1}{2} \exp\left(-\frac{x^2}{2}\right). \quad (2.13)$$

This bound was further extended by Chiani *et al.* in [27] to include multiple exponential functions with the aim of improving its accuracy to approximate or bound  $Q(x)$  as

$$Q(x) \approx \sum_{n=1}^N a_n \exp\left(-b_n x^2\right), \quad (2.14)$$

for which, Chiani *et al.* found the coefficients  $a_n$  and  $b_n$  for two exponential terms ( $N = 2$ ) using the trapezoidal integration rule with optimized mean relative error and for any  $N$  using the rectangular integration rule with non-optimized equispaced points, that despite them giving better accuracy than (2.13), their accuracy still not quite adequate for statistical analysis.

Other works have also considered developing more accurate exponential approximations and bounds of the same form as (2.14) with different approaches [28]–[31]. A sum of two or three exponentials, which is known as the Prony approximation, is proposed in [28] together with an iterative procedure for obtaining its parameters. Another approximation of the exponential form that is easily invertible and shows a good trade-off between computational efficiency and mathematical accuracy is proposed in [29]. The composite trapezoidal rule with an optimally chosen number of sub-intervals is used in [30] to realize (2.14). A single-term exponential lower bound is introduced in [31] by bounding from above the logarithmic function with a tangent line at some point that sets the limit's tightness.

More complicated approximations and bounds are also available in the literature. The authors in [32] propose a relatively tractable and tight mathematical

expression based on a second-order exponential function as

$$Q(x) \approx \exp(ax^2 + bx + c), \quad (2.15)$$

where  $a$ ,  $b$ , and  $c$  are numerically-calculated fitting parameters. The semi-infinite Gauss–Hermite quadrature rule is used in [33] to realize a new expression in the form of summation of (2.15). Moreover, tight upper and lower bounds are presented in [34] as a sum of two exponentials with respective constant and rational factors as

$$Q(x) \leq \frac{1}{50} \exp(-x^2) + \frac{1}{2(x+1)} \exp\left(\frac{-x^2}{2}\right), \quad (2.16)$$

$$Q(x) \geq \frac{1}{12} \exp(-x^2) + \frac{1}{\sqrt{2\pi}(x+1)} \cdot \exp\left(\frac{-x^2}{2}\right). \quad (2.17)$$

More accurate approximation for  $Q(x)$  that guarantees sufficient accuracy for all positive values of  $x$  is provided in [35], [36] as

$$Q(x) \approx \frac{1}{\sqrt{2\pi}} \frac{\left(1 - \exp\left(-A \frac{x}{\sqrt{2}}\right)\right)}{Bx} \exp\left(-\frac{x^2}{2}\right), \quad (2.18)$$

for which  $A$  and  $B$  are optimized in order to minimize the mean relative error. This approximation was later modified to derive an upper bound in [37] and to derive a simpler approximation that does not have the argument  $x$  in the denominator using Taylor series expansion in [38] as

$$Q(x) \approx \frac{1}{\sqrt{2\pi}} \left[ \sum_{n=1}^{n_a} \frac{(-1)^{n+1} (A)^n}{B(\sqrt{2})^n n!} x^{n-1} \right] \exp(-x^2/2), \quad (2.19)$$

with  $A$  and  $B$  being the same as for (2.18). A complicated truncated-infinite series expression that is more accurate for large values of  $x$  is derived in [39] as

$$Q(x) \approx \frac{hx \exp\left(-\frac{x^2}{2}\right)}{2\pi} \left( \frac{2}{x^2} + 2 \sum_{n=1}^N \frac{\exp\left(-n^2 h^2\right)}{n^2 h^2 + \frac{x^2}{2}} \right), \quad (2.20)$$

with  $h$  and  $N$  being empirically determined to achieve relative error less than  $10^{-10}$ .

The authors in [40] present an accurate polynomial approximation based on the observation that a Gaussian random variable can be approximated well by a sum of uniform random variables as

$$Q(x) \approx 1 - \sum_{m=0}^n \sum_{p=0}^n \frac{(-1)^{m+p} \binom{n}{p}}{m!(n-m)!} \left(\frac{n}{12}\right)^{p/2} \left(\frac{n}{2} - m\right)^{n-p} x^p \\ \times u \left[ x - \sqrt{\frac{12}{n}} \left(\frac{n}{2} - m\right) \right], \quad (2.21)$$

where  $u(\cdot)$  is the unit step function. The authors in [41] introduce a more complicated but tighter approximation in the form

$$Q(x) \approx \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{x^2 + 1}} \exp\left(-\frac{x^2}{2}\right). \quad (2.22)$$

More complex approximations and bounds that are not very suitable for algebraic manipulation related to communication systems' performance analysis are also derived in the same paper as

$$Q(x) \approx \frac{1}{\sqrt{2\pi}} \frac{1}{(1-a)x + a\sqrt{x^2 + b}} \exp\left(-\frac{x^2}{2}\right), \quad (2.23)$$

for which the accuracy can be controlled by the choice of the two parameters  $a$  and  $b$  using any numerical optimization procedure in order to minimize the global relative error. It is worth noting that as the mathematical form of the approximation/bound becomes more sophisticated, the accuracy of the approximation/bound is likely to improve. A detailed comparison in terms of accuracy, tractability as well as computational complexity between the different approximations and bounds accessible in the literature can be found in [42].

The approximations and bounds mentioned above have been implemented in the different areas of communication theory to evaluate systems' performance over fading channels in closed form. A few application examples from the literature are provided herein. The exponential approximation/bound in (2.14) is used to compute error probabilities for space-time codes and phase-shift

keying in [27], to derive the average BER for free-space optical systems in [43] and to evaluate the symbol error rate (SER) of phase-shift keying under Rician fading in [44]. Moreover, the frame error rate for a two-way decode-and-forward (DF) relay link is derived in [45], and the average BER of adaptive Walsh–Hadamard transform-aided quadrature amplitude modulation is derived in [46], using (2.19) for both. Furthermore, the average of integer powers of the  $Q$ -function over  $\eta - \mu$  and  $\kappa - \mu$  fading whose PDFs are defined in Table 2.1, is derived in [47], and the probability of detection in an energy detector under Rayleigh fading in cognitive radio networks is evaluated in closed form in [32] by utilizing (2.15) for both applications.

## 2.2 Ergodic Capacity Analysis

Channel capacity is a crucial and fundamental measure to analyze and study the performance of the various wireless communication systems [48]. In fact, whatever wireless technology is in use, all communication techniques have a tight upper bound on the rate at which information can be reliably transmitted over a communication channel, which is known as channel capacity, i.e., channel capacity is the maximum mutual information of the wireless channel. In the presence of additive white Gaussian noise, the channel capacity is calculated according to Shannon–Hartley theorem as

$$C = \log_2 (1 + \text{SNR}) \quad [\text{bit/s/Hz}], \quad (2.24)$$

and is referred to as the Shannon capacity. On the other hand, in the presence of multipath fading, the channel capacity varies with the channel state (SNR). Thus, the metric *ergodic capacity* is used to express the capacity of the fading channel, which is defined as the statistical average of the mutual information. The ergodic capacity is calculated by taking the expectation or mean of the random capacity that results from the random channel and is calculated for the AWGN channel as

$$\bar{C} \triangleq \text{E}[C] = \text{E} [\log_2 (1 + \text{SNR})] \quad [\text{bit/s/Hz}]. \quad (2.25)$$

Establishing closed-form expressions for the ergodic capacity is essential in

communication theory since it allows us to obtain scientific knowledge of how communication systems behave and how their parameters affect performance. The ergodic capacity has been studied well in the literature and exact closed-form formulas for it for various transmission schemes and under different assumptions on transmitter and receiver channel knowledge for several fading distributions have been derived [49]–[65]. In particular, the ergodic capacity over Rayleigh fading has been derived for numerous single-antenna systems and multi-antenna systems, namely multiple-input single-output (MISO), single-input multiple-output (SIMO), and MIMO systems for correlated and non-correlated channels with different combining techniques at the receiver in [49]–[59]. The ergodic capacity has also been derived for single-antenna and multi-antenna systems with non-correlated channels under Nakagami and Rician fading in [60]–[64]. Furthermore, it has been derived in [65] for single-input single-output (SISO) system under  $\kappa - \mu$  fading whose PDF is defined in Table 2.1.

In general, the exact evaluation of the ergodic capacity in terms of analytical functions is quite challenging. More specifically, since the ergodic capacity is the expectation of the Shannon capacity, it is then calculated as the integral

$$\bar{C} = \int_0^{\infty} \log_2(1 + \gamma) \psi_{\gamma}(\gamma) d\gamma, \quad (2.26)$$

which can be referred to as *capacity integral* whose analytical evaluation depends on the system’s complexity and the specific channel characteristics. Therefore, analytical tools are needed to facilitate the conducted capacity analysis and derive closed-form expressions. For that, many approximations and bounds are available in the literature, from which the most relevant ones to the scope of the thesis are [66]–[72]. The authors in [66] give a lower bound for the ergodic capacity of MIMO Rayleigh channels with frequency-selective fading and/or channel correlation, as well as an asymptotic estimate of the ergodic capacity over flat fading.

In [67], asymptotic results are presented for particular multi-antenna scenarios with channel information first at the receiver, then at the transmitter. In [68], various bounds are presented for the encountered channel capacity computations under fast Rayleigh fading, perfect channel knowledge at the receiver,

and with/without channel information at the transmitter. Two tractable yet relatively tight approximations for the ergodic capacity that enable the construction of analytical resource allocation techniques in Rayleigh MIMO systems are derived in [69]. In [70], the authors suggest two simple and reliable approximations for the ergodic capacity in the low-SNR area. In [71] and [72], complicated analytical bounds for the ergodic capacity in dual-hop fixed-gain AF relay networks are presented for Rayleigh and Nakagami fading channels, respectively.

In addition to the small-scale fading, approximations and bounds are also developed for systems experiencing the shadowing effect, which is normally modeled by lognormal distribution whose PDF is defined in Table 2.1. In particular, closed-form expressions for the ergodic capacity of the different communication systems over lognormal fading channels do not exist. For that, number of approximations and bounds that enable its evaluation in closed form are available in the literature [10], [73]–[76]. Alouini *et al.* in [10] were the first to propose lower and upper bounds for the ergodic capacity under lognormal fading channels. However, these bounds are very simple and thus loose, especially for the lower range of the SNR. More accurate approximations for SISO systems were later presented in [73], [74] and the results were extended to approximate the capacity of diversity combining techniques with or without channel correlation. Furthermore, the authors in [75] propose accurate closed-form approximations for the ergodic capacity for SISO and MIMO indoor ultra-wideband systems under lognormal fading. Closed-form approximations for the ergodic capacity of multiple adaptive transmission schemes are derived in [76].

## 2.3 Reconfigurable Intelligent Surfaces

The RIS is a large two-dimensional meta-surface that consists of small, low-cost, almost passive reflecting elements (REs) that can be intelligently controlled by a smart controller, e.g., a field-programmable gate array (FPGA), to collectively steer the incident electromagnetic signals into the desired direction. This is achieved by adjusting the phase shifts of the REs to regulate the directionality of the dispersed signals. The motivation behind using the RIS technology can be summarized by the following points [77]:

- **Spectrum efficiency improvement:** Since the RISs are able to reconfigure the wireless environment, virtual line-of-sight (LoS) links between base stations and users can be formed by controlling the direction of the reflected signals, allowing for improvements in the received signal-to-interference-plus-noise ratio which in turn implies enhanced diversity gain which has been introduced in Section 2.1.
- **Energy efficiency improvement:** Since the RIS does not need energy-hungry hardware components and is able to shape the incoming signal instead of using a power amplifier, RIS is more energy-efficient than conventional AF and DF systems which are two common relaying schemes. The former amplifies and retransmits the received signal without decoding, while the latter decodes, re-encodes, and retransmits the signal.
- **Easy to implement:** The RIS is flexible to be deployed and extended to many structures, e.g., buildings and traffic signs, due to its compatible size, low-cost elements, and minimal digital signal processing requirements. Nevertheless, the limited signal processing capabilities cause the acquisition of channel state information (CSI) to be quite difficult since the RIS consists of passive elements that are not capable of transmitting training sequences independently [78]. Therefore, the transmitter has to perform CSI estimation. Several estimation strategies for the RIS-aided systems are available in the literature [78]–[81].

Extensive research has been carried out to investigate the design [82]–[84], optimization [85]–[87] and applications [88]–[90] of RIS-aided systems. Specifically, the authors in [82] focus on the key mechanisms and typologies in designing the reflectarray and array lens technologies to adopt them in realizing reconfigurable designs. A digitally controlled RIS whose REs can be independently controlled is designed in [83] to dynamically manipulate the electromagnetic waves and thus achieve more versatility. In [84], a tunable metasurface is designed to function as a spatial microwave modulator with energy feedback, and it has been demonstrated to efficiently shape the complex existing microwave fields in reverberating indoor environments.

Earlier works have additionally examined optimizing the performance of RIS-aided systems. In [85], the authors tackle a non-convex optimization problem during the process of optimizing the phase shifts and the downlink transmit

powers in such a way as to efficiently maximize the system's energy. In [86], the authors optimize the discrete phase shifts, as well as the transmit beamforming of a multi-antenna base station, in order to minimize the transmitted power. Furthermore, in [87], the authors employ the RIS technology at the edge of cells to improve the downlink transmission for cell-edge users. In addition, they optimize the active precoding matrices of the transmitter, as well as the RIS phase shifts, in order to maximize the weighted sum rate of all users.

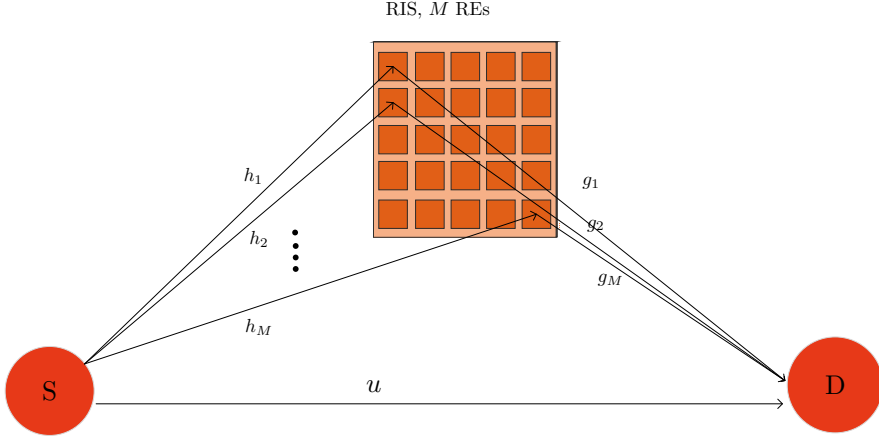
The usefulness and efficiency of the RIS technology can be best seen in its diverse applications that span the different areas of wireless communications. Among the various applications found in the literature, few are given herein. In particular, the RIS is adopted to improve the communication in unmanned aerial vehicles-assisted wireless systems in [88] and to assist the data transmission between a base station and a single-antenna receiver in a mmwave system in [89]. This technology can also be implemented in the different wireless systems to improve physical layer security, as seen in [90].

In general, the fundamental performance measures of RIS-aided systems are not fully investigated yet. In fact, due to the difficulties in assessing the statistical characterization of the received SNR, relatively few research works have been conducted to study the performance of these systems. As a result, this thesis focuses on the theoretical study of the RIS-aided systems and mainly on the SISO system with single RIS and multiple RISs. In the literature, a number of approximations, bounds, or asymptotic analysis techniques have been devised to analyze the different RIS-aided systems. For example, the average BER of a RIS-aided non-orthogonal multiple-access (NOMA) system is approximated in [91] by utilizing the central limit theorem (CLT), and the average error probability of a RIS-based wireless system with phase errors is investigated in [92] after modeling the transmission through a RIS by a direct channel with Nakagami scalar fading.

### 2.3.1 Conventional SISO System Model with a Single RIS

A conventional RIS-aided SISO transmission link is illustrated in Fig. 2.3 that consists of a single-antenna source (S), a single-antenna destination (D), and a RIS equipped with  $M$  REs. As depicted in the figure, the REs on the RIS receive the superposed multipath signals from S and then scatter the combined signal





**Figure 2.3** A SISO wireless system with a single RIS. The S–RIS and RIS–D links consist of multiple propagation paths through the  $M$  REs

towards D after adjusting the amplitude and phases as if it was transmitted from a single point source [93].

### Signal Model

For the system model in Fig 2.3, D can overhear the transmitted signal from S through the RIS as well as the direct path. Therefore, the received signal at D consists of three components, namely the reflected signal, the direct-path signal, and the channel noise. Thus, the received signal is written as

$$y = \underbrace{\mathbf{g}^T \mathbf{\Theta} \mathbf{h}}_{\text{Reflected Signal}} s + \underbrace{u s}_{\text{Direct-Path Signal}} + \underbrace{w}_{\text{Noise}} = A s + w, \quad (2.27)$$

for which  $A = \mathbf{g}^T \mathbf{\Theta} \mathbf{h} + u$  is the channel response,  $s$  is the transmitted signal whose transmitted power is  $E_s = E[|s|^2]$ ,  $w \sim \mathcal{N}_C(0, N_0)$  is the AWGN that is circularly symmetric Gaussian distributed with zero mean and variance  $N_0 = E[|w|^2]$ , and  $\mathbf{\Theta} = \zeta \text{diag}(\mathbf{e}^{j\theta_1}, \dots, \mathbf{e}^{j\theta_M})$  is the diagonal phase-shift matrix for which  $\zeta = [0, 1]$  denotes the reflection coefficient and controls the reflected signal's amplitude, whereas  $\{\theta_i\}_{i=1}^M$  denotes the set of phase shifts of the  $M$  REs. The channel

vectors  $\mathbf{h} = [h_1, \dots, h_M]^T \in \mathbb{C}^M$  and  $\mathbf{g} = [g_1, \dots, g_M]^T \in \mathbb{C}^M$  correspond to the links S–RIS and RIS–D, respectively, for which the fading coefficients  $\{h_i\}_{i=1}^M = \{|h_i|e^{j\angle h_i}\}_{i=1}^M$  and  $\{g_i\}_{i=1}^M = \{|g_i|e^{j\angle g_i}\}_{i=1}^M$ . The variable  $u = |u|e^{j\angle u} \in \mathbb{C}$  is the fading coefficient of the direct S–D link.

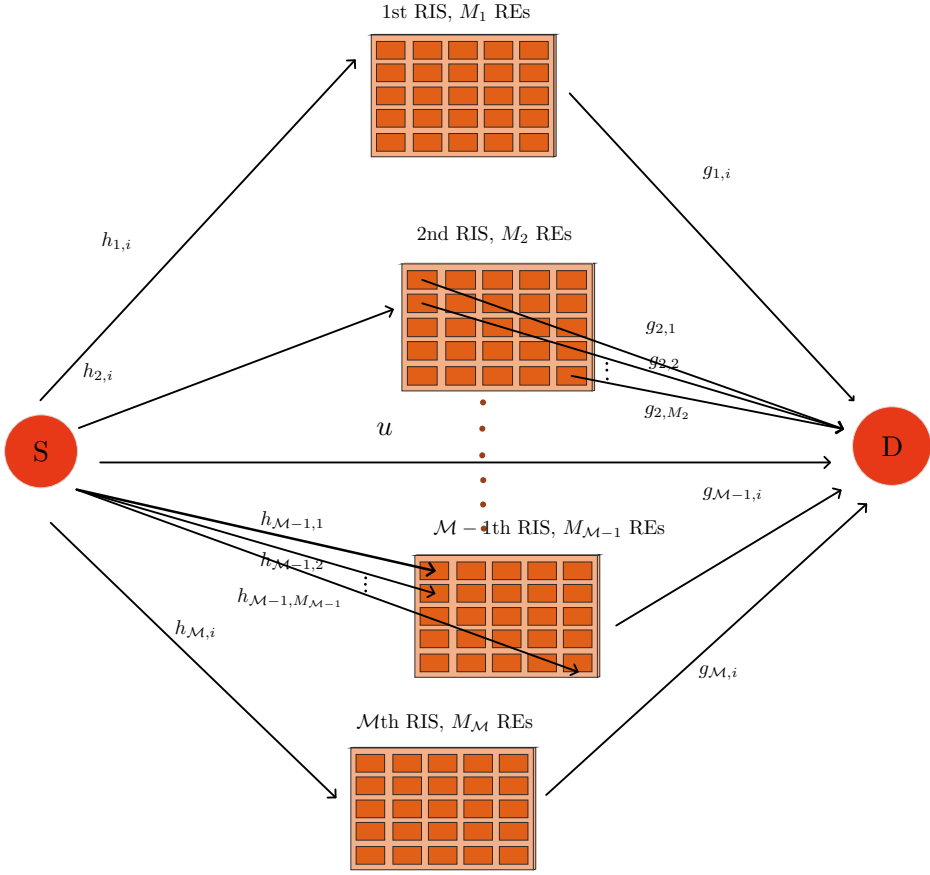
### Related Work from the Literature

Several research papers in the literature have analyzed the performance of SISO systems without a direct link. In particular, the CLT scheme is used to approximate the PDF of the equivalent channel and thus derive the different performance measures in [94], [95] under Rayleigh distribution. A different approximating approach is also developed under Rayleigh distribution in [96] and under Rician distribution in [97] to approximate the different performance measures with high accuracy despite the number of equipped REs on the RIS opposing to [94], and [95] which have low accuracy for the lower number of the REs. The same system model is again investigated in [98] under Rician distribution but with a direct link between S and D, and approximations are derived for the ergodic capacity and outage probability.

Other research papers have considered analyzing the performance of multi-antenna systems with a single RIS, from which a few examples are mentioned herein. In [99], the statistical properties of the received SNR are characterized and used to derive closed-form approximations for the different performance measures for a MISO system aided by a RIS. In [100], the performance of a  $2 \times 2$  MIMO system assisted with a RIS with optimized phase shifts over Rayleigh fading is investigated in terms of outage probability and throughput, whereas in [101], the ergodic capacity of a downlink multi-user MISO system is derived and used together with CSI to propose new RIS configuration algorithm.

### 2.3.2 Generic SISO System Model with Multiple RISs

A more generic SISO system that is aided by multiple RISs is shown in Fig. 2.4 and consists of a single-antenna source, a single-antenna destination, and  $\mathcal{M}$  geographically distributed RISs, with the  $l$ th RIS equipped with  $M_l$  REs.



**Figure 2.4** A SISO wireless system with  $\mathcal{M}$  RISs. Each S–RIS $_l$  and RIS $_l$ –D link consists of multiple propagation paths through the  $M_l$  REs. For simplicity, we illustrate the multipath components via two RISs only.

### Signal Model

For the system model in Fig 2.4, all the  $\mathcal{M}$  RISs collaboratively participate in the transmission process between S and D, in addition to the direct path. Therefore, the received signal at D consists of the reflected signal, direct-path signal, and the channel noise and can be written as

$$y = \underbrace{\sum_{l=1}^{\mathcal{M}} \mathbf{g}_l^T \mathbf{\Theta}_l \mathbf{h}_l}_{\text{Reflected Signal}} s + \underbrace{u s}_{\text{Direct-Path Signal}} + \underbrace{w}_{\text{Noise}} = \mathbf{A} s + w, \quad (2.28)$$

for which  $A = \sum_{l=1}^M \mathbf{g}_l^T \mathbf{\Theta}_l \mathbf{h}_l + u$  is the combined channel response, and  $\mathbf{\Theta}_l = \zeta_l \text{diag}(e^{j\theta_{l,1}}, \dots, e^{j\theta_{l,M_l}})$  is the diagonal phase-shift matrix of the  $l$ th RIS for which  $\zeta_l$  and  $\{\theta_{l,i}\}_{i=1}^{M_l}$  denote respectively the reflection coefficient of the  $l$ th RIS and the set of phase shifts of the  $M_l$  REs equipped on the  $l$ th RIS. The channels vectors  $\mathbf{h}_l = [h_{l,1}, \dots, h_{l,M_l}]^T \in \mathbb{C}^{M_l}$  and  $\mathbf{g}_l = [g_{l,1}, \dots, g_{l,M_l}]^T \in \mathbb{C}^{M_l}$  correspond to the links S–RIS $_l$  and RIS $_l$ –D, respectively, for which  $\{h_{l,i}\}_{i=1}^{M_l} = \{|h_{l,i}|e^{j\angle h_{l,i}}\}_{i=1}^{M_l}$  and  $\{g_{l,i}\}_{i=1}^{M_l} = \{|g_{l,i}|e^{j\angle g_{l,i}}\}_{i=1}^{M_l}$ .

## Related Work from the Literature

The generic SISO system with multiple RISs in Fig. 2.4 is studied in [102]–[106] and its performance is investigated after approximating the channel statistics using different approaches. In particular, the authors in [102] use a mathematical derivation that involves the CLT to tightly approximate the received SNR over independent and identically distributed (i.i.d.) Nakagami- $m$  fading channels, whereas the authors in [103] demonstrate that the received SNR over i.i.d. Rayleigh fading channels can be approximated by a non-central chi-square distribution assuming no direct path exists, and only the RIS with the highest transmitted SNR is selected to assist transmission. In [104], the authors present several RIS selection strategies based on the location information of the RISs whose number of REs can be arbitrarily adjusted with the assumption that the channel coefficients associated with various RISs are i.i.d. RVs.

Moreover, the authors in [105] propose two transmission models over multiple RISs for indoor and outdoor environments and analyze their performance, as well as develop novel selection strategies. In addition, a practical multi-RISs system is considered in [106] with two selection schemes; one scheme includes all the RISs in the communication process, while the other selects the best RIS in terms of maximum received SNR at the destination. Two approximating schemes are used to model the received SNR, which is used afterward to derive approximations for the different performance measures.

### 3 APPROXIMATION THEORY AND OPTIMIZATION

Halfway through the twentieth century, the field of approximation theory was born in the hands of the Russian mathematician Pafnuty Lvovich Chebyshev [107]. Approximation theory, in broad terms, is an area of mathematics concerned with the issue of approximating a given function  $f(x)$  by some other simpler function  $\tilde{f}(x)$ . In the scope of this thesis,  $f(x)$  would be the Gaussian  $Q$ -function defined in (2.9b) or the capacity integral defined in (2.26). An example of  $\tilde{f}(x)$  for the  $Q$ -function would be the exponential approximating function defined in (2.14). Other approximating functions are introduced in the following chapters for both the  $Q$ -function and the capacity integral, all of which are parametric functions. A parametric function refers to a mathematical expression that consists of one or more variables called parameters or coefficients, for which the choice of the coefficients affects the accuracy of the approximation.

In general, formulating tight approximations needs a careful choice of the corresponding coefficients. An efficient choice would be through optimization. This chapter introduces several optimization methodologies to be adopted in the following chapters by the developed approximation or bound for the investigated performance measure to produce a highly accurate analytical performance analysis in terms of the minimax or total error that are to be defined shortly. In particular, two novel approaches are presented to minimize the maximum error, while the quasi-Newton algorithm is adopted to minimize the total error. The contents of this chapter are elaborated in publications [P1], [P2],[P4]–[P7].

The most relevant errors of measurement for the context of this thesis are the absolute and relative error functions which are defined respectively as

$$d(x) \triangleq \tilde{f}(x) - f(x), \quad (3.1)$$

$$r(x) \triangleq \frac{d(x)}{f(x)} = \frac{\tilde{f}(x)}{f(x)} - 1, \quad (3.2)$$

where  $\tilde{f}(x)$  is the approximation of some function  $f(x)$ . The shorthand  $e \in \{d, r\}$  represents both error measures collectively in what follows. On the other hand, the most relevant optimization criteria for the context of this thesis are the minimax absolute/relative error optimization and minimum total absolute/relative error optimization for which the maximum and total errors are defined, respectively as

$$e_{\max} \triangleq \max_{R_1 \leq x \leq R_2} |e(x)|, \quad (3.3)$$

and

$$e_{\text{tot}}(\mathbf{u}) = \int_{R_1}^{R_2} |e(x)| dx. \quad (3.4)$$

Another related optimization criterion is the minimum mean error, whose mean error is denoted by  $\bar{e}_{\text{tot}}$  and is obtained by simply dividing the total error in (3.4) by the range of values of interest  $R = R_2 - R_1$  as

$$\bar{e}_{\text{tot}} = \frac{e_{\text{tot}}}{R}. \quad (3.5)$$

### 3.1 Minimax Error Optimization

When establishing an approximation for a complicated function  $f(x)$ , a proper coefficients choice for the considered approximation could be obtained by optimizing it in such a way as to minimize the corresponding maximum error in order to give sufficient accuracy for the whole argument range of interest. Pafnuty Lvovich Chebyshev [107] was the first to present the best minimax approximation or, alternatively, the *Chebyshev approximation*. The theory of the minimax approximation started by approximating a given function  $f(x)$  by

a polynomial

$$p(\mathbf{u}, x) = \sum_{n=1}^N a_n x^n, \quad (3.6)$$

where  $p(\mathbf{u}, x) \in \mathbb{P}_N$  with  $\mathbb{P}_N$  being the space of polynomials of degree  $\leq N$  and  $\mathbf{u} = \{a_n\}_{n=1}^N$  is the set of coefficients to be optimized.

In particular, for a continuous function  $f(x)$  on the closed interval  $[R_1, R_2]$ , i.e.,  $f \in C[R_1, R_2]$  where  $C$  is a function space of all continuous functions defined on a closed interval  $[R_1, R_2]$ , its best minimax approximation  $p(\mathbf{u}^*, x)$  from the set  $\mathbb{P}_N$  with the optimized coefficients  $\mathbf{u}^* = \{a_n^*\}_{n=1}^N$  must satisfy

$$\|f - p^*\|_{\infty} \leq \|f - p\|_{\infty}, \quad (3.7)$$

for all other polynomials  $p \in \mathbb{P}_N$ , or alternatively

$$\max_{R_1 \leq x \leq R_2} |f(x) - p(\mathbf{u}^*, x)| \leq \max_{R_1 \leq x \leq R_2} |f(x) - p(\mathbf{u}, x)|, \quad (3.8)$$

where  $\|\cdot\|_{\infty}$  denotes the uniform norm or the supremum norm.

The presented theory is not limited to ordinary polynomials only, but it can also be applied to approximate functions by generalized polynomials of the form

$$\tilde{f}(\mathbf{u}, x) = \sum_{n=1}^N a_n g_n(x), \quad (3.9)$$

for which  $\tilde{f}(x) \in \mathbb{F}_{\hat{D}}$ , with  $\mathbb{F}_{\hat{D}}$  being the space of continuous functions of degree  $\leq \hat{D}$ , and  $\{g_n(x)\}_{n=1}^N$  is a system of continuous functions that can implicitly include other coefficients to be optimized, hence  $\mathbf{u}$  will include  $\{a_n\}_{n=1}^N$  and any other coefficients encountered by  $\{g_n(x)\}_{n=1}^N$ . The degree of a generalized polynomial is the number of its corresponding coefficients which are to be optimized, i.e., size of  $\mathbf{u}$  [108]. For example, the degree of the ordinary polynomial in (3.6) is  $\hat{D} = N$ .

A key point to discuss in the minimax approximation theory is the uniqueness of the approximation, where the best minimax approximation is always unique. The uniqueness condition is met when the approximation  $\tilde{f}(\mathbf{u}, x)$  to

a function  $f \in C[R_1, R_2]$  satisfies the Haar condition, i.e. when the system of continuous functions  $\{g_n\}_{n=1}^N$  in (3.9) satisfies the Haar condition [109] and thus is called a Chebyshev system. In particular, the system  $\{g_n\}_{n=1}^N$  is a Chebyshev system (meets the Haar condition) if each nontrivial linear combination of the form (3.9) has at most  $N - 1$  distinct zeros on  $[R_1, R_2]$  or equivalently when the determinant

$$D[x_1, \dots, x_N] = \begin{vmatrix} g_1(x_1) & \cdots & g_N(x_1) \\ \vdots & \ddots & \vdots \\ g_1(x_N) & \cdots & g_N(x_N) \end{vmatrix} \quad (3.10)$$

is nonzero whenever  $\{x_n\}_{n=1}^N$  are all distinct. Some approximation families that satisfy the Haar condition can be found in [110].

A best minimax approximation of a certain function family with degree  $\hat{D}$  always results in an absolute error function  $d(x)$  (defined in (3.1)) that uniformly alternates  $\hat{D}$  times between  $\hat{D} + 1$  extrema points of the same value of error and alternating signs [111]. Extrema points can be critical points where the corresponding function changes from decreasing to increasing or vice versa, and the function's derivative disappears, or they can be endpoints. An illustration example of the expected uniform shape of the absolute error function with  $\hat{D} = 4$  and  $[R_1, R_2] = [0, \infty)$  is given in Fig. 3.1. In fact, Fig. 3.1 illustrates  $d(x)$  defined in (3.1) for  $f(x) = Q(x)$  and  $\tilde{f}(x)$  being the exponential approximation in (2.14) with  $N = 2$ . In particular,  $\tilde{f}(\mathbf{u}^*, x)$  with degree  $\hat{D}$  is the best approximation for  $f(x)$  if and only if there exist  $\hat{D} + 1$  points,  $\{x_k\}_{k=1}^{\hat{D}+1}$  with  $R_1 \leq x_1 < \dots < x_k < \dots < x_{\hat{D}+1} \leq R_2$ , such that

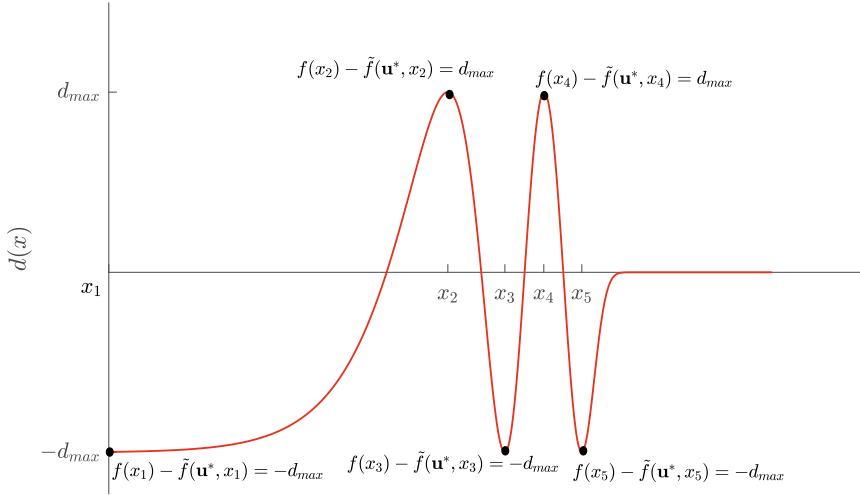
$$f(x_k) - \tilde{f}(\mathbf{u}^*, x_k) = (-1)^k d_{\max}, \quad (3.11)$$

for which

$$d_{\max} \triangleq \max_{R_1 \leq x \leq R_2} |d(x)| = \|d\|_{\infty}, \quad (3.12)$$

is the maximum absolute error. Although the theory on the minimax approximation was first proposed to minimize the maximum absolute error as discussed above, it can also be extended to minimize the maximum relative error defined in (3.2) as done in [112], for which the same minimax theory principles apply.





**Figure 3.1** The uniform error function,  $d(x)$ , that corresponds to the best minimax approximation with degree  $\hat{D} = 4$ .

### 3.1.1 Methods of Implementation

The best approximation is the member of a certain function family that is the tightest of them all and always occurs with optimal set of coefficients  $\mathbf{u}^*$  that minimizes the maximum absolute or relative error over the range of values of interest  $[R_1, R_2]$ , as

$$\mathbf{u}^* \triangleq \arg \min_{\mathbf{u}} e_{\max}, \quad (3.13)$$

where  $e_{\max}$  is defined in (3.3). New approaches are proposed to solve the minimax optimization problem and thus calculate the corresponding coefficients in (3.13).

#### New Scheme: Non-Linear System of Equations

The weighted sum of continuous functions with degree  $\hat{D}$  in (3.9) is adopted to derive minimax approximations for some function  $f(x)$ . The original idea in this scheme is to describe the uniformly oscillating error function between  $\hat{D} + 1$  maximum and minimum values of equal magnitude and alternating signs

at  $\{x_k\}_{k=1}^{\hat{D}+1}$  extremum points by a set of non-linear equations, for which the number of equations equals the number of unknowns. The complete set of equations that describes the minimax error function can be split into three subsets as follows:

**The first subset** consists of the equations that describe the error value at the critical points, which are defined herein as the local extrema points, excluding those that might occur at one or both endpoints of the considered range  $[R_1, R_2]$ .

**The second subset** consists of the equations that describe the zero derivative of the error function at the critical points.

**The third subset** pertains to the endpoints of the considered range  $[R_1, R_2]$  after evaluating their corresponding error values either directly or by taking the limit, and they could be maxima or minima. The endpoints can only contribute with one equation describing their corresponding error value. For the functions  $f(x)$  considered in this study and the proposed approximating expressions  $\tilde{f}(x)$ , the following characteristics are noted. When  $x$  tends to  $R_1$ , the error function normally converges to a constant value that is to be fixed to the equalized error value of the local extrema points,  $-e_{\max}$ , for both the absolute and relative errors; otherwise, a condition is imposed to bound the error which in turns decrease the degree of the approximating function by one. Therefore,  $x_1 = R_1$  for both error measures. Nevertheless, when  $x$  tends to  $R_2$ , the absolute error function converges to zero and hence  $x_{\hat{D}+1} < R_2$ . As for the relative error when  $x$  tends to  $R_2$ , the error function either converges to a constant value that is to be fixed to  $-r_{\max}$  or converges to infinity or  $-1$  for which a finite interval on the  $x$ -axis is chosen. Therefore,  $x_{\hat{D}+1} = R_2$  always for the relative error in this study.

The ultimate goal is to find the best set of coefficients,  $\mathbf{u}^*$ , which solves the following minimax set of equations that stems from the aforementioned three subsets for  $e \in \{d, r\}$  as

$$\begin{cases} f_k(\mathbf{v}) = e(x_k) + (-1)^{k+1} e_{\max} = 0, & \text{for } k = 1, 2, 3, \dots, \hat{D} + 1, \\ f'_k(\mathbf{v}) = e'(x_k) = 0, & \text{for } k = 2, 3, 4, \dots, L \end{cases} \quad (3.14)$$

for which  $L = \hat{D}$  for  $e = r$  and  $L = \hat{D}+1$  for  $e = d$ , and  $\mathbf{v}$  is a vector of the approximation's coefficients with  $e_{\max}$  which are to be optimized, i.e.,  $\mathbf{v} = [\mathbf{u}, e_{\max}]$ . Although only the set  $\mathbf{u}$  is needed to construct the minimax error function, other unknowns will also appear when solving the optimization problem in (3.14), which are  $\{x_k\}_{k=1}^{\hat{D}+1}$  and  $e_{\max}$ . In the rest of this chapter, the problem formulations cover the relevant functions to the scope of this thesis and their approximations. These problem formulations can be, however, generalized to other functions out of the scope of this thesis by simply constructing  $f_k$  for all the extrema points and constructing  $f'_k$  for all the critical points.

After formulating (3.14), the non-linear set of equations with an equal number of equations and unknowns can be solved using any numerical tool such as `fsolve` command in Matlab to yield the set of optimized coefficients  $\mathbf{u}^*$  which, when substituted in (3.9), yields the best minimax approximation for all  $x \in [R_1, R_2]$ . The key difficulty in implementing this scheme is its necessity for good initial guesses for the unknowns that can ultimately converge to the optimized solution. An alternative method that is less sensitive to the right choice of the initial guesses is introduced next.

### Modified Remez Algorithm

Another method for solving (3.13) is developing the well-known Remez exchange algorithm established by Evgeny Yakovlevich Remez in 1934. The Remez algorithm is an iterative procedure used to establish the best minimax approximation using non-linear approximating functions of the form (3.9), which are typically Chebyshev systems that satisfy the Haar condition. The best minimax approximation is characterized by the uniform alternation of the corresponding error function as shown in Fig. 3.1 and explained previously. The original Remez algorithm encounters solving a linear system of equations. The linearity of the resulting system of equations depends on the mathematical form of the chosen approximating function. For example, the ordinary polynomial in (3.6) results in a linear system. Nevertheless, some approximating functions can result in a non-linear system of equations. For that, a variation of the Remez exchange algorithm that complies with the nonlinearity that might occur from the approximating function is introduced herein. The modified Remez algorithm is less with  $\hat{D} - 1$  equations for  $e = r$  and with  $\hat{D}$  equations for  $e = d$

than the non-linear system of equations method. The absence of the derivative equations makes it less sensitive to the right choice of initial guesses.

The Remez algorithm can be implemented to find the optimized set of coefficients  $\mathbf{u}^*$  of the minimax approximation (3.9) for some function  $f(x)$  by following the steps summarized in Algorithm 1. In particular, a system of  $\hat{D} + 1$  simultaneous non-linear equations that describe the  $\hat{D} + 1$  equalized extrema, including those that might occur at the endpoints of the considered range, is constructed as

$$f_k(\mathbf{v}) = e(x_k) + (-1)^{k+1} e_{\max} = 0, \text{ for } k = 1, 2, 3, \dots, \hat{D} + 1, \quad (3.15)$$

for which  $\mathbf{v}$  is the vector of unknowns of length  $\hat{D} + 1$  and is defined in (3.14). The locations of the extrema points  $\{x_k\}_{k=1}^{\hat{D}+1}$  are then initialized. In fact, the location of the extrema points that occur at the endpoints ( $x_1 = R_1$  for  $e \in \{d, r\}$  and  $x_{\hat{D}+1} = R_2$  for  $e = r$ ) are always fixed to those values. The iterations  $t$  of the Remez algorithm begin after that and are referred to as the outer iterations, and in each iteration,  $\mathbf{f} = [f_1(\mathbf{v}), f_2(\mathbf{v}), \dots, f_{\hat{D}+1}(\mathbf{v})]^T$  is solved for  $\mathbf{v}$  using any numerical tool. Following each iteration, the new extrema points of the resulting error function are located and used to initialize the following Remez iteration. This procedure is repeated until the difference between the old, and new  $\hat{D} + 1$  extrema lies below a predefined threshold value  $\varepsilon$ , i.e., until the extrema are of equal values.

---

**Algorithm 1** Remez Exchange Algorithm

---

Initialize  $\{x_k^{(0)}\}_{k=2}^L$ , with  $L = \hat{D}$  for  $e = r$  and  $L = \hat{D} + 1$  for  $e = d$   
Set  $t \leftarrow 0$ ,  $x_1 \leftarrow R_1$  for  $e \in \{d, r\}$ ,  $x_{\hat{D}+1} \leftarrow R_2$  for  $e = r$   
**repeat**  
Solve  $\mathbf{f} = [f_1(\mathbf{v}), f_2(\mathbf{v}), \dots, f_{\hat{D}+1}(\mathbf{v})]^T$  for the vector of unknowns  $\mathbf{v}$ , using any numerical method for system of non-linear equations.  
Locate new  $\{x_k^{(t)}\}_{k=2}^L$ , with  $L = \hat{D}$  for  $e = r$  and  $L = \hat{D} + 1$  for  $e = d$   
Replace  $x_k^{(t-1)}$  by  $x_k^{(t)}$  for all the extrema points  
 $t \leftarrow t + 1$   
**until**  $\left| \{x_k^{(t)}\}_{k=1}^{\hat{D}+1} - \{x_k^{(t-1)}\}_{k=1}^{\hat{D}+1} \right| < \varepsilon$   
Best minimax approximation is obtained

---

## Newton–Raphson Method

Both the non-linear system of equations method and the modified Remez algorithm encounter non-linear sets of equations that can be solved numerically using any numerical-analysis software. One explicit example of a numerical method that the software can use is the Newton–Raphson method. In particular, Newton–Raphson is used to solve the whole system of equations in (3.14) for  $\mathbf{w} = [\mathbf{v}, x_2, x_3, \dots, x_L]$  with  $L = \hat{D}$  for  $e = r$  and  $L = \hat{D} + 1$  for  $e = d$ . On the other hand, it is used to solve (3.15) for  $\mathbf{w} = [\mathbf{v}]$  after initializing  $x_k$ ,  $k = 2, 3, 4, \dots, L$ , in the outer iterations of the Remez algorithm.

The Newton–Raphson method is a root-finding technique that is quadratically convergent while approaching the root, making it a fairly optimal solver for this system of non-linear equations. This method also starts from good initial guesses for the unknowns and is based on approximating a continuous and differentiable function by a straight line tangent to it, which results when applied to both of the considered non-linear systems of equations (3.14) and (3.15) in the iteration

$$\mathbf{w}^{(\tau+1)} = \mathbf{w}^{(\tau)} - \left[ \mathbf{J}^{(\tau)} \left( \mathbf{w}^{(\tau)} \right) \right]^{-1} \mathbf{f} \left( \mathbf{w}^{(\tau)} \right), \quad (3.16)$$

where  $\tau$  is its counter,  $\mathbf{J}(\cdot)$  is the Jacobian matrix defined as

$$\mathbf{J}(\mathbf{w}) = \left[ \frac{\partial \mathbf{f}}{\partial w_1}, \frac{\partial \mathbf{f}}{\partial w_2}, \dots, \frac{\partial \mathbf{f}}{\partial w_L} \right],$$

with  $L = \hat{D}$  when  $e = r$  and  $L = \hat{D} + 1$  when  $e = d$  for the non-linear system of equations method, whereas  $L = \hat{D} + 1$  for the modified Remez algorithm when  $e \in \{d, r\}$ .

On the other hand,

$$\mathbf{f}(\mathbf{w}) \triangleq \left[ f_1(\mathbf{w}), f_2(\mathbf{w}), \dots, f_{\hat{D}+1}(\mathbf{w}), f'_2(\mathbf{w}), f'_3(\mathbf{w}), \dots, f'_L(\mathbf{w}) \right]^T$$

for the non-linear system of equations method, whereas

$$\mathbf{f}(\mathbf{w}) \triangleq \left[ f_1(\mathbf{w}), f_2(\mathbf{w}), \dots, f_{\hat{D}+1}(\mathbf{w}) \right]^T$$

for the modified Remez algorithm. The iterations of the Newton–Raphson method in (3.16) are repeated until the differences between the values of  $\mathbf{w}$  of two successive iterations are smaller than a predefined threshold value. It should be noted that the Newton–Raphson method is implemented on the modified Remez algorithm to find the vector of unknowns in every outer iteration of the Remez algorithm.

### Initial Guesses

Before starting to implement the non-linear system of equations or the modified Remez algorithm, along with the Newton-Raphson method, one must have good initial guesses for the unknowns, namely,  $\mathbf{u}$ ,  $e_{\max}$  and  $x_k, k = 2, 3, 4, \dots, L$ . Although the Remez method, like the non-linear system of equations, requires initial guesses for the unknowns, it is far more resistant to the accuracy of the used guesses and converges to the optimal solution more rapidly. A general approach was employed throughout this study to achieve good initial guesses with certain differences specific to the investigated approximation. The good initial guesses lead to quick convergence to the optimum values that can accomplish the requisite uniform shape for the related error function.

In particular, for the lower values of the summation terms ( $N$ ) in (3.9), different random values are assigned repeatedly for  $\mathbf{u}$  from which  $e(x)$  is calculated per (3.1) or (3.2) for each  $N$ . Once any  $e(x)$  with the correct number of  $\hat{D} + 1$  extrema occurs,  $\{x_k\}_{k=1}^{\hat{D}+1}$  and  $e_{\max}$  are calculated and used together with the corresponding  $\mathbf{u}$  to solve the considered optimization problem (3.14) or (3.15) to find the optimal solution. After reaching certain  $N$ , curve fitting techniques have been used for the higher values of  $N$  to formulate equations that can give good initial values for  $\mathbf{u}$  or at least to work as mean values around which small random variance is introduced, and from which  $e(x)$  is plotted and  $\{x_k\}_{k=1}^{\hat{D}+1}$  and  $e_{\max}$  are calculated thereof. The choice of the initial guesses that is specific to the studied approximation is more elaborated in Chapter 4 together with publications [P1]–[P5] for the Gaussian  $Q$ -function, and in Chapter 5 together with publication [P6] for the ergodic capacity.

### 3.1.2 Lower and Upper Bounds

The minimax approximation theory introduced in Section 3.1 can be extended to derive upper and lower bounds of the same form as (3.9) with degree  $\hat{D}$  for a given function  $f(x)$  rather than approximations by controlling the corresponding coefficients. In fact, the same uniform shape for the corresponding error function  $e(x)$  is expected, but with a different number of extrema points and different placement of the error function about the  $x$ -axis. Furthermore, the non-linear system of equations and the modified Remez algorithm described in Section 3.1.1, which were originally proposed to implement the minimax approximation theory, can also be used to implement the derived bounds.

In particular, the same approach as for the approximations is used herein with ensuring that  $e(x) \leq 0$  and  $e(x) \geq 0$  for the lower and upper bounds, respectively, when  $x \in [R_1, R_2]$ . Both types of bounds will also alternate in sign between  $\hat{D} + 1$  extrema points. However, the alternation will be either above (upper bound) or below (lower bound) the  $x$ -axis, not crossing it as for the approximations. In addition, one should be extra careful when evaluating the limits at the right endpoint of the considered interval  $[R_1, R_2]$ , where  $R_2$  is considered as an extremum point only if the error function converges to a constant value or it converges to infinity or  $-1$  for which a finite interval on the  $x$ -axis is chosen. Otherwise, if the error function converges to zero, no extremum occurs at  $R_2$ . More specifically, for the absolute error, which converges to zero when  $x$  tends to  $R_2$  as discussed in the third subset in Section 3.1.1, the last extrema will never occur at  $R_2$ .

The optimized set of coefficients for the lower bound can be obtained by implementing the modified Remez algorithm to solve the following system of non-linear equations that describe the  $\hat{D}+1$  extrema, including those that might occur at the endpoints of the considered range as

$$\begin{cases} f_k(\mathbf{v}) = e(x_k) + e_{\max} = 0, & \text{for } k = 1, 3, 5, \dots, \\ f_k(\mathbf{v}) = e(x_k) = 0, & \text{for } k = 2, 4, 6, \dots, \end{cases} \quad (3.17)$$

whereas the optimized set of coefficients for the upper bound can be obtained

using the modified Remez algorithm by solving

$$\begin{cases} f_k(\mathbf{v}) = e(x_k) = 0, & \text{for } k = 1, 3, 5, \dots, \\ f_k(\mathbf{v}) = e(x_k) - e_{\max} = 0, & \text{for } k = 2, 4, 6, \dots \end{cases} \quad (3.18)$$

Alternatively, the non-linear system of equations can be used by adding into the above system of equations in (3.17) and (3.18) the set of equations that describes the zero-derivative at the extrema points, excluding those at the endpoints, which are described by

$$f'_k(\mathbf{v}) = e'(x_k) = 0, \text{ for } k = 2, 3, 4, \dots, L, \quad (3.19)$$

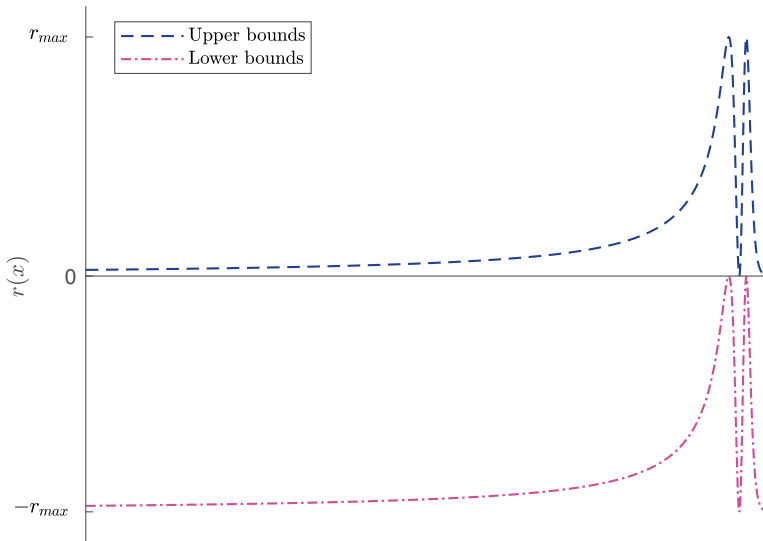
for which  $L = \hat{D}$  for  $e = r$  and  $L = \hat{D} + 1$  for  $e = d$ .

An illustration example of the expected uniform shape of the relative error function with  $\hat{D} = 4$  and  $[R_1, R_2] = [0, \infty)$  is given in Fig. 3.2. In fact, this figure illustrates  $r(x)$  defined in (3.2) for  $f(x)$  being the capacity integral in (2.26) for a SISO system under Rayleigh fading and  $\tilde{f}(x)$  being a novel logarithmic bound that is to be proposed in Chapter 5. It should be mentioned that some lower bounds can start the alternation from zero instead of  $-e_{\max}$  and hence the extrema points at the odd abscissa would be equal to zero, and those at the even abscissa would be equal to  $-e_{\max}$ . The same applies for the upper bounds, where it is possible that the upper bounds would start the alternation from  $e_{\max}$  instead of zero, and hence the extrema points at the odd abscissa would be equal to  $e_{\max}$ , and those at the even abscissa would be equal to zero.

## 3.2 Total Error Optimization

When approximating a complicated function  $f(x)$  with a simpler function  $\tilde{f}(\mathbf{u}, x)$  of degree  $\hat{D}$  in (3.9), a proper coefficients choice for the considered approximation could be obtained by optimizing the set of coefficients  $\mathbf{u}$  in such a way as to minimize the corresponding total absolute/relative error. In addition to the minimax error optimization discussed in the previous section, total error optimization is the second most relevant optimization criterion for the context of this thesis to improve the overall approximation accuracy. This optimization criterion is widely used in the literature to optimize the target function and to





**Figure 3.2** The uniform error function,  $r(x)$ , that corresponds to the best minimax lower and upper bounds with degree  $\hat{D} = 4$ .

evaluate the achieved accuracy of the studied systems [27], [35].

The total absolute/relative error optimization is performed as

$$\mathbf{u}^* \triangleq \arg \min_{\mathbf{u}} e_{\text{tot}}, \quad (3.20)$$

where  $e_{\text{tot}}$  is defined in (3.4). The total error optimization can be implemented using the numerical quasi-Newton algorithm. Generally, the quasi-Newton algorithm is an iterative technique used to find the roots of a certain differentiable function. Nevertheless, this algorithm can also be utilized in the context of optimization by implementing it to the derivative of a twice-differentiable target function. This will result in the optimized roots of the function's derivative. In particular, the quasi-Newton algorithm is implemented in this study to minimize the target function  $e_{\text{tot}}$ , which corresponds to an approximating function of degree  $\hat{D}$ . It starts from good initial guesses for the vector of unknowns  $\mathbf{u}$ , of size  $\hat{D}$ , to converge ultimately to the optimized one. The optimized coefficients result in the target function having the smallest possible value.

Each iteration of this algorithm is carried out as follows

$$\mathbf{u}^{(\tau+1)} = \mathbf{u}^{(\tau)} - \delta \left[ \tilde{\mathbf{H}}^{(\tau)} \right]^{-1} \mathbf{G}^{(\tau)} \left( \mathbf{u}^{(\tau)} \right), \quad (3.21)$$

for which  $\tau$  is the iteration counter,  $0 < \delta \leq 1$  is the iteration step size,  $\mathbf{G}(\cdot)$  is the gradient vector calculated as

$$\mathbf{G} = \left[ \frac{\partial e_{\text{tot}}(\mathbf{u})}{\partial u_1} \quad \frac{\partial e_{\text{tot}}(\mathbf{u})}{\partial u_2} \quad \dots \quad \frac{\partial e_{\text{tot}}(\mathbf{u})}{\partial u_D} \right]^T,$$

and  $\tilde{\mathbf{H}}$  is an approximation to the Hessian matrix  $\mathbf{H}$  which is calculated as

$$\mathbf{H} = \left[ \frac{\partial \mathbf{G}(\mathbf{u})}{\partial u_1} \quad \frac{\partial \mathbf{G}(\mathbf{u})}{\partial u_2} \quad \dots \quad \frac{\partial \mathbf{G}(\mathbf{u})}{\partial u_D} \right].$$

To compute  $\tilde{\mathbf{H}}$ , many algorithms have been devised from which the well-known Broyden–Fletcher–Goldfarb–Shanno (BFGS) method is used herein. The BFGS method begins with some symmetric positive-definite matrix  $\tilde{\mathbf{H}}^{(0)}$ , which is updated in consecutive iterations as

$$\tilde{\mathbf{H}}^{(\tau+1)} = \tilde{\mathbf{H}}^{(\tau)} + \frac{\Delta \mathbf{G}^{(\tau)} [\Delta \mathbf{G}^{(\tau)}]^T}{[\Delta \mathbf{G}^{(\tau)}]^T \Delta \mathbf{u}^{(\tau)}} - \frac{\tilde{\mathbf{H}}^{(\tau)} \Delta \mathbf{u}^{(\tau)} [\Delta \mathbf{u}^{(\tau)}]^T [\tilde{\mathbf{H}}^{(\tau)}]^T}{[\Delta \mathbf{u}^{(\tau)}]^T \tilde{\mathbf{H}}^{(\tau)} \Delta \mathbf{u}^{(\tau)}}, \quad (3.22)$$

where  $[\cdot]^T$  denotes the transpose,  $\Delta \mathbf{G}^{(\tau)} = \mathbf{G}^{(\tau+1)} \left( \mathbf{u}^{(\tau+1)} \right) - \mathbf{G}^{(\tau)} \left( \mathbf{u}^{(\tau)} \right)$  and  $\Delta \mathbf{u}^{(\tau)} = \mathbf{u}^{(\tau+1)} - \mathbf{u}^{(\tau)}$ . This method's iterations are repeated until the difference between the values of  $\mathbf{u}$  of two subsequent iterations is less than some preset threshold value.

Due to the recent advancements in the different programming languages, the quasi-Newton algorithm can be implemented directly using built-in functions. For example, Matlab software can be used herein through the `fminunc` command with setting its corresponding algorithm to 'quasi-newton' and choosing good initial values for  $\mathbf{u}$  to find  $\mathbf{u}^*$  in (3.20). If some constraints are added to the optimization problem, the command `fmincon` is used instead.

## 4 GAUSSIAN $Q$ -FUNCTION

Having defined the Gaussian  $Q$ -function, emphasized its importance in communication theory, and listed its approximations and bounds from the literature in Chapter 2, the next consideration is to develop new approximations and bounds for the  $Q$ -function that are generally better than the existing ones in terms of accuracy and analytical complexity. This chapter is divided into three main sections with each of the first two sections introducing novel approximation or bound for the  $Q$ -function with the optimization methods presented in Chapter 3 implemented in such a way as to optimize the performance. The third section introduces the various applications of the proposed approximations/bounds and investigates their performance. This chapter is based on publications [P1]–[P5].

### 4.1 Exponential-Type Approximations

The approximations and bounds presented in [32], [33], [37], [39]–[41] which have rather difficult mathematical forms, achieve high accuracy and can sometimes lead to closed-form expressions that would otherwise be impossible to solve. For example, the polynomial approximation in [40] succeeds in analytically evaluating the average SER of pulse amplitude modulation in lognormal channels whose PDF is defined in Table 2.1. However, their mathematical complexity makes them not preferable for algebraic manipulations in statistical performance analysis despite being accurate. For example, the approximation provided by Börjesson and Sundberg in [41] is extremely sophisticated and is best suited for programming applications.

As a result, Chiani *et al.* in [27] suggested the simplest known family in the form of a sum of exponentials (2.14). In particular, they apply the trapezoidal integration rule along with optimizing the center point to minimize the mean error defined in (3.5) in order to derive a new exponential approximation with

two terms ( $N = 2$ ). They also apply the rectangular integration rule on (2.9b) to produce an exponential upper bound that requires a large number of terms to achieve appropriate precision.

Nevertheless, neither their approximation nor bound is as accurate as it could be in order to be used as a reliable substitute for the Gaussian  $Q$ -function. In addition, the elegance and potentials of the exponential expression in (2.14) can still be exploited further, where not only the  $Q$ -function can be approximated by the exponential sum, but also its powers, polynomials, or many well-behaved functions of the  $Q$ -function. This is because the integer powers of the  $Q$ -function or any linear combination will result in the same exponential form as in (2.14). This has motivated the need for improved accuracy and higher versatility for this tractable family to be used most effectively in statistical performance analysis. Therefore, this section aims to develop new tight exponential approximations and bounds for the Gaussian  $Q$ -function and its functions, including powers and polynomials of the  $Q$ -function, that do not exist in the literature.

#### 4.1.1 Functions of $Q$ -Function

The exponential family can be used to approximate or bound many well-behaved functions of the  $Q$ -function, i.e.,  $F(Q(x))$ , by implementing Taylor series expansion. A well-behaved function means that both the function and its derivatives are defined and continuous in the range of the expansion around some point  $q_0$ . Taylor series is an expansion of an infinitely differentiable function  $F(q)$  around  $q_0$  by a polynomial of infinite degree as follows [3, Eq. 0.317.1]

$$F(q) = \sum_{p=0}^{\infty} \frac{F^{(p)}(q_0)}{p!} (q - q_0)^p, \quad (4.1)$$

where  $F^{(p)}(q_0)$  denotes the  $p$ th derivative of  $F(\cdot)$  evaluated at point  $q_0$ . Taylor polynomial approximation of degree  $P$  is obtained by truncating the infinite expansion in (4.1). In general, the Taylor series can be applied to approximate a  $P$ -times differentiable function  $F(Q(x))$  around a given point,  $0 \leq Q(x_0) \leq \frac{1}{2}$ ,

by a polynomial of the  $Q$ -function,  $\Omega(Q(x))$ , of degree  $P$  that is defined as

$$\Omega(Q(x)) \triangleq \sum_{p=0}^P c_p Q^p(x), \quad (4.2)$$

where  $\{c_p\}_{p=0}^P$  are constants and called the polynomial coefficients.

The exponential expression in (2.14) is used to directly approximate or bound any polynomial of the  $Q$ -function by

$$\tilde{Q}_\Omega(x) = \sum_{n=1}^N a_n \exp(-b_n x^2), \quad (4.3)$$

and thus approximating any function of the  $Q$ -function that accepts a Taylor series expansion. In addition, the polynomial (4.2) comprises the integer powers of the  $Q$ -function as special cases, including the first power ( $p = 1$ ), where (4.2) is a linear combination of non-negative integer powers of the  $Q$ -function. Therefore, instead of limiting the exponential approximations/bounds to the  $Q$ -function or its integer powers only, they can be sufficiently generalized to polynomials or even functions of the  $Q$ -function. For that purpose, the focus is on the research problem of finding new, improved coefficients for the exponential sum (4.3).

#### 4.1.2 Minimax Error-Based Solution

Using the theory presented in Section 3.1 on the best Chebyshev approximation, the polynomial function (4.2) or any of the nested special cases, which are all defined on the interval  $[0, \infty)$ , are approximated by the best exponential approximation of the form (4.3). In particular, the approximating function (4.3), which is an example on the generalized polynomial defined in (3.9), is of degree  $\hat{D} = 2N$  and thus  $\mathbf{u} = \{a_n, b_n\}_{n=1}^N$  is the set of coefficients to be optimized. The best approximation is derived by finding the optimized coefficients  $\mathbf{u}^* = \{a_n^*, b_n^*\}_{n=1}^N$  that satisfies (3.13) together with (3.3).

The best exponential approximation with optimized coefficients is unique since the system of continuous functions  $\{\exp(-b_n x^2)\}_{n=1}^N$  satisfies the Haar condition, where each nontrivial linear combination of the form (4.3) has at most  $N - 1$  distinct zeros on  $[0, \infty)$ . The minimax exponential approximation

will result in an error function that uniformly alternates  $2N$  times between  $2N + 1$  extrema points of the same value of error and alternating signs. In addition, the absolute error converges to zero when  $x$  tends to infinity, i.e.,  $\lim_{x \rightarrow \infty} d(x) = 0$ , whereas for the relative error,

$$\lim_{x \rightarrow \infty} r(x) = \begin{cases} \infty, & \text{when } \min\{b_n\}_{n=1}^N = \frac{1}{2}, \\ -1, & \text{otherwise.} \end{cases} \quad (4.4)$$

This renders some specific restrictions for all upper bounds and optimization w.r.t. the relative error as is shortly observed.

The minimax error optimization methods presented in Section 3.1.1, namely the non-linear system of equations method and the modified Remez algorithm, are implemented to optimize the sets of coefficients  $\{(a_n, b_n)\}_{n=1}^N$  that correspond to (4.3) or any of the nested special cases denoted for simplicity by  $\tilde{Q}_p(x)$  for integer powers of  $Q$ -function and  $\tilde{Q}$  for the Gaussian  $Q$ -function (its first power) in publications [P1], [P2].

#### Minimax solution based on the non-linear system of equations

This solution is based on describing the expected uniform shape of the corresponding error function  $e(x)$ , which is defined herein as

$$d(x) \triangleq \tilde{Q}_\Omega(x) - \sum_{p=0}^P c_p Q^p(x), \quad (4.5)$$

for the absolute error and as

$$r(x) \triangleq \frac{d(x)}{\sum_{p=0}^P c_p Q^p(x)} = \frac{\tilde{Q}_\Omega(x)}{\sum_{p=0}^P c_p Q^p(x)} - 1, \quad (4.6)$$

for the relative error. The above two equations are obtained by substituting (4.2) and (4.3) in the definitions of errors of measurement in (3.1) and (3.2), respectively. The three equations subsets presented in Section 3.1.1 are used to formulate the following system of equations that corresponds to the exponential

approximation

$$\left\{ \begin{array}{ll} f_k(\mathbf{v}) = e(x_k) + (-1)^{k+1} e_{\max} = 0, & \text{for } k = 2, 3, \dots, L, \\ f'_k(\mathbf{v}) = e'(x_k) = 0, & \text{for } k = 2, 3, \dots, L, \\ f_1(\mathbf{v}) = \sum_{n=1}^N a_n - \frac{1}{2} + d_{\max} = 0, & \text{for } e = d, \\ \left\{ \begin{array}{l} f_1(\mathbf{v}) = \sum_{n=1}^N a_n - \frac{1}{2} + \frac{1}{2} r_{\max} = 0, \\ f_{2N+1}(\mathbf{v}) = r(x_{2N+1}) + r_{\max} = 0. \end{array} \right. & \text{for } e = r \end{array} \right. \quad (4.7)$$

for which  $L$  and  $\mathbf{v}$  are defined generally for (3.14). More specifically, they are defined for the considered exponential approximation as  $L = 2N$  for  $e = r$  and  $L = 2N + 1$  for  $e = d$ , and  $\mathbf{v} = [a_1, a_2, \dots, a_N, b_1, b_2, \dots, b_N, e_{\max}]$ .

The equations  $f_1$  and  $f_{2N+1}$  in (4.7) pertain to the third subset which describes the endpoints that can be extrema points. In particular,  $f_1$  results from substituting  $Q(0) = \frac{1}{2}$  and  $\tilde{Q}_\Omega(0) = \sum_{n=1}^N a_n$  in the corresponding error function, whereas  $f_{2N+1}$  results from choosing a finite interval on the  $x$ -axis since the relative error function does not converge to zero or to a constant value when  $x$  tends to infinity as illustrated by (4.4). Thus, the right boundary of the relative error function is equal to  $x_{2N+1}$  and the relative error function is minimized globally over  $[0, x_{2N+1}]$  instead of  $[0, \infty)$  like for the absolute error.

The formulated system of equations has  $4N + 1$  equations for  $e = d$  and  $4N$  equations for  $e = r$ , that is equal to the number of unknowns, namely  $\mathbf{v}$  and  $\{x_k\}_{k=2}^L$ . System (4.7) can be solved by implementing the iterative Newton–Raphson method discussed in Section 3.1.1 to acquire the optimized solution or by directly implementing any numerical software tool, e.g., Matlab. Nevertheless, good initial guesses are needed to start any numerical iterative solver. Detailed information about the used initial guesses for the developed exponential approximations can be found in publication [P1].

### Minimax solution based on the modified Remez algorithm

The optimized set of coefficients  $\{(a_n^*, b_n^*)\}_{n=1}^N$  for (4.3) can be found by implementing the modified Remez algorithm developed in Section 3.1.1 according to the steps summarized in Algorithm 1. The modified Remez algorithm solution is less with  $2N - 1$  equations for  $e = r$  and with  $2N$  equations for  $e = d$  than

the non-linear system of equations solution. In particular, a system of  $2N + 1$  simultaneous non-linear equations that describes the  $2N + 1$  equalized extrema including those that might occur at the endpoints of the considered range is constructed as  $\mathbf{f} = [f_1(\mathbf{v}), f_2(\mathbf{v}), \dots, f_{2N+1}(\mathbf{v})]^T$  for which  $\mathbf{f}_l, l = 1, 2, \dots, 2N + 1$  and  $\mathbf{v}$  are defined in (4.7).

In each outer iteration of the Remez algorithm, the Newton–Raphson method is implemented to solve  $\mathbf{f}$  and find  $\mathbf{v}$ , after which the locations of the extrema points are updated, and the following Remez iteration takes place. Detailed information about one possible heuristic method to find good initial guesses for the unknowns, i.e.,  $\{x_k\}_{k=2}^L$  for the outer Remez iterations and  $\mathbf{v}$  for the inner Newton–Raphson iterations, can be found in publication [P2].

### Lower and Upper bounds for the Gaussian $Q$ -function

As explained in Section 3.1.2, it is possible to use the minimax approximation theory to derive lower and upper bounds for (4.2) using (4.3) by imposing additional constraints on the expected shape of the corresponding error function. In particular, the best lower exponential bound of degree  $D = 2N$ , which alternates between  $2N + 1$  extrema, is constructed by numerically finding the optimized coefficients that solve the same system of equations as that of (4.7) but with replacing

$$f_k(\mathbf{v}) = e(x_k) + (-1)^{k+1} e_{\max} = 0, \text{ for } k = 2, 3, \dots, L,$$

which ensures that  $e(x)$  alternates around the  $x$ -axis for the approximation, by

$$\begin{cases} f_k(\mathbf{v}) = e(x_k) + e_{\max} = 0, & \text{for } k = 3, 5, \dots, L - 1 \text{ for } e = r \text{ or } L \text{ for } e = d, \\ f_k(\mathbf{v}) = e(x_k) = 0, & \text{for } k = 2, 4, \dots, 2N, \end{cases} \quad (4.8)$$

which ensures that  $e(x) < 0$  for the lower bound. The parameter  $L = 2N$  for  $e = r$  and  $L = 2N + 1$  for  $e = d$ , and  $\mathbf{v} = [a_1, a_2, \dots, a_N, b_1, b_2, \dots, b_N, e_{\max}]$ . The modified Remez algorithm can also be used to find the optimized coefficients after deleting  $f'_k, k = 2, 3, 4, \dots, L$ , from the system of equations.



On the other hand, for the upper bound, the lowest value in the set  $\{b_n\}_{n=1}^N$  is forced to be  $\frac{1}{2}$  in order to get positive  $e(x)$  as concluded from (4.4). Otherwise,  $r(x)$  will converge to a negative value and  $d(x)$  would be negative for large  $x$  too. The best upper exponential bound of degree  $\hat{D} = 2N - 1$  (since  $\min\{b_n\}_{n=1}^N = \frac{1}{2}$ ), which alternates between  $2N$  extrema, is constructed by numerically finding the optimized coefficients that solves the following system of equations and ensures  $e(x) > 0$

$$\begin{cases} f_k(\mathbf{v}) = e(x_k) = 0, & \text{for } k = 3, 5, \dots, 2N - 1, \\ f_k(\mathbf{v}) = e(x_k) - e_{\max} = 0, & \text{for } k = 2, 4, \dots, L - 1 \text{ for } e = r \text{ or } L \text{ for } e = d, \\ f'_k(\mathbf{v}) = e'(x_k) = 0, & \text{for } k = 2, 3, 4, \dots, L, \\ f_1(\mathbf{v}) = \sum_{n=1}^N a_n - \frac{1}{2} = 0, & \text{for } e \in \{d, r\}, \\ f_{2N}(\mathbf{v}) = r(x_{2N}) - r_{\max} = 0 & \text{for } e = r \end{cases} \quad (4.9)$$

for which  $L = 2N - 1$  for  $e = r$  and  $L = 2N$  for  $e = d$ . The modified Remez algorithm can also be used by simply deleting the zero-derivative equations, i.e.,  $f'_k$ ,  $k = 2, 3, 4, \dots, L$ , from (4.9).

### 4.1.3 Quadrature-Based Solution

The exponential approximation in (4.3) for the Gaussian  $Q$ -function results from applying the different numerical integration methods [113], which are also referred to as the quadrature rules, to approximate Craig's formula in (2.9b) with numerical coefficients instead of the optimized minimax coefficients discussed above. More specifically, any integral of the form  $\int_{\mathcal{S}}^v W(\theta) f(\theta) d\theta$ , can be represented by a finite sum of the form [113]

$$\int_{\mathcal{S}}^v W(\theta) f(\theta) d\theta \approx \sum_{n=1}^N w_n f(\theta_n), \quad (4.10)$$

for which  $W(\theta)$  is some weighting function,  $[\mathcal{S}, v]$  is the integration domain,  $\{\theta_n\}_{n=1}^N$  are the quadrature points/nodes and  $\{w_n\}_{n=1}^N$  are the quadrature weights. Hence, for  $W(\theta) = 1$ ,  $f(\theta)$  equals the integrand in (2.9b) and  $[\mathcal{S}, v] = [0, \pi/2]$ ,

the Gaussian  $Q$ -function can be numerically approximated by (4.3) whose set  $\{(a_n, b_n)\}_{n=1}^N$  is the set of numerical coefficients, which depends on the applied numerical integration technique.

The quadrature integration techniques can be categorized into Newton–Cotes and Gaussian quadrature formulas. In general, as the number of quadrature points increases, the accuracy of the corresponding quadrature method increases. However, the instability of higher-order numerical methods increases too, especially with the Newton–Cotes rules which have negative weights that can result in subtractive cancellation. Therefore, the composite integration rule can be used as an efficient technique for approximation. For the composite rule, the integration interval,  $[\varsigma, \nu] = [0, \pi/2]$ , can be divided into smaller uniform or non-uniform sub-intervals at which simpler integration rules with a lower number of nodes can be used for each sub-interval.

A comprehensive overview of all the possible numerical integration techniques that can be applied to approximate the Gaussian  $Q$ -function is presented in publication [P3], together with a unified method for optimizing the coefficients of the resulting exponential approximation for any number of exponentials, any optimization criterion, and using any numerical quadrature rule.

## 4.2 Generalized Karagiannidis–Lioumpas Approximations

Karagiannidis and Lioumpas in [35] proposed a tight, although analytically tractable, approximation for the Gaussian  $Q$ -function as (2.18) with  $A$  and  $B$  being optimized in order to minimize the mean relative error defined in (3.5). Although the Karagiannidis–Lioumpas (KL) approximation has received some early criticism in [36], it has still established itself as one of the most usable alternative representations of the  $Q$ -function in the different problems of communication theory, where it has received a large number of citations. However, this approximation has only been optimized in terms of one optimization criterion, which limits its versatile potential. Therefore, this section aims at making the best use of the KL approximation not only by optimizing its coefficients in terms of other criteria for better accuracy depending on the application but also by repurposing it to derive lower and upper bounds and, most importantly, to generalize it into a new expression for approximating or bounding

the  $Q$ -function, for which the original KL approximation is a special case, with optimizing its coefficients in terms of other criteria for significantly improved accuracy.

#### 4.2.1 Mathematical Form and Origin

Based on the Karagiannidis–Lioumpas (KL) approximation, a new expression to approximate or bound the  $Q$ -function is derived as

$$\tilde{Q}(x) \triangleq \frac{1 - \exp(-cx)}{x} \cdot \sum_{n=1}^N a_n \exp(-b_n x^2), \quad (4.11)$$

which is referred to as the generalized KL (GKL) expression since (4.11) yields the original KL expression in (2.18) when  $N = 1$ . The GKL expression is derived using a similar approach as that in [35], where the  $Q$ -function is first approximated by the exponential sum as in (2.14). This results in an unbounded relative error for the higher argument values  $x$  and, thus, lower accuracy. To overcome this accuracy issue, the exponential sum is then multiplied by  $\frac{1 - \exp(-cx)}{x}$  in order to bound the relative error from the right side with setting  $b_1 \triangleq \min\{b_n\}_{n=1}^N = \frac{1}{2}$ .

The new approximation (4.11) is well exploited in terms of accuracy by optimizing it in terms of the minimax or total errors using the theory already presented in Section 3.1 and Section 3.2, respectively. For that purpose, the focus is on the research problem of finding optimized coefficients,  $\{(a_n^*, b_n^*)\}_{n=1}^N$  and  $c^*$ , for the GKL approximation. The optimized coefficients may be constrained by some conditions depending on the behavior of the GKL expression at the endpoints of the range  $[0, \infty)$  which is described by

$$d_0 \triangleq \lim_{x \rightarrow 0} d(x) = c \sum_{n=1}^N a_n - \frac{1}{2}, \quad (4.12)$$

$$\lim_{x \rightarrow \infty} d(x) = 0, \quad (4.13)$$

$$r_0 \triangleq \lim_{x \rightarrow 0} r(x) = 2c \sum_{n=1}^N a_n - 1, \quad (4.14)$$

$$\lim_{x \rightarrow \infty} r(x) = \begin{cases} \infty, & \text{if } b_1 < \frac{1}{2}, \\ \sqrt{2\pi} a_1 - 1, & \text{if } b_1 = \frac{1}{2}, \\ -1, & \text{if } b_1 > \frac{1}{2}, \end{cases} \quad (4.15)$$

where  $a_1$  is the counterpart of  $b_1 \triangleq \min\{b_n\}_{n=1}^N$ . Based on (4.15), it can be concluded that for  $b_1 = \frac{1}{2}$  only, there are global approximations and bounds in terms of relative error. On the other hand, the absolute error function is always bounded regardless of the value of  $b_1$ . Therefore, two variations of approximations w.r.t. absolute error are considered, i.e., first variation with  $b_1 < \frac{1}{2}$ , which provide higher accuracy, and second variation with  $b_1 = \frac{1}{2}$ .

The proposed GKL expression in (4.11) together with the optimized coefficients, can also be expanded in the same way as done in [114] for the original KL expression. In particular, the presence of the argument  $x$  in the denominator is avoided by implementing Taylor series expansion [3, Eq. 0.317.1] on the term  $\exp(-cx)$  and truncating it to several terms,  $\mathcal{L}$ , which results in the following approximation that has the same analytical complexity as the original one in [114]

$$\tilde{Q}(x) = \sum_{l=1}^{\mathcal{L}} \sum_{n=1}^N \frac{(-1)^{l+1} a_n c^l}{l!} \exp(-b_n x^2) x^{l-1}. \quad (4.16)$$

## 4.2.2 Minimax Error-Based Solution

The Gaussian  $Q$ -function is approximated by the best approximation of the form (4.11) using the minimax approximation theory in the same way as for the exponential approximating function. In particular, (4.11) is of degree  $\hat{D} = 2N+1$  for first variation of  $e = d$ , while  $\hat{D} = 2N$  for the second variation of  $e = d$  and for  $e = r$ . The unique best approximation is derived by finding the optimized coefficients  $\mathbf{u}^* = \{\{a_n^*, b_n^*\}_{n=1}^N, c^*\}$  that satisfies (3.13) and result in a uniformly oscillating error function between  $2N + 2$  extrema points for first variation of  $e = d$ , and between  $2N + 1$  extrema points for the second variation of  $e = d$  and for  $e = r$ . The non-linear system of equations method and the modified Remez algorithm can be implemented herein to optimize the sets of coefficients. In

fact, the detailed formulation of the associated systems of equations for the best GKL approximation, their solutions, and the used initial guesses are presented in publication [P5] for both methods of implementations for which the same general theory presented in Section 3.1 is followed.

The GKL expression can also be repurposed to derive lower and upper bounds for the Gaussian  $Q$ -function, whose mathematical problem formulation is stated herein explicitly since they are not covered in detail for the bounds in [P5]. In particular, for the lower bound,  $b_1 = \frac{1}{2}$  is always used for both absolute and relative errors in order to ensure that  $e(x) < 0$ , and thus  $\hat{D} = 2N$ . The solution is based on describing the expected uniform shape of the error function  $e(x)$ , defined herein as  $d(x) \triangleq \tilde{Q}(x) - Q(x)$  for  $e = d$  and as  $r(x) \triangleq \frac{\tilde{Q}(x)}{Q(x)} - 1$  for  $e = r$ , for which  $\tilde{Q}(x)$  and  $Q(x)$  are defined respectively by (4.11) and (2.9a).

The formulated system of the equation whose solution minimizes the error globally on  $[0, \infty)$ , is given as

$$\left\{ \begin{array}{ll} \mathbf{f}_k(\mathbf{v}) = e(x_k) = 0, & \text{for } k = 2, 4, \dots, 2N, \\ \mathbf{f}_k(\mathbf{v}) = e(x_k) + e_{\max} = 0, & \text{for } k = 3, 5, \dots, L - 1 \text{ for } e = r \text{ or } L \text{ for } e = d, \\ \mathbf{f}'_k(\mathbf{v}) = e'(x_k) = 0, & \text{for } k = 2, 3, 4, \dots, L, \\ \mathbf{f}_1(\mathbf{v}) = e_0 + e_{\max} = 0, & \text{for } e \in \{d, r\}, \\ \mathbf{f}_{2N+1}(\mathbf{v}) = a_1 + \frac{r_{\max}-1}{\sqrt{2\pi}} = 0, & \text{for } e = r, \end{array} \right. \quad (4.17)$$

where  $L = 2N$  for  $e = r$  and  $L = 2N + 1$  for  $e = d$ , and

$$\mathbf{v} = [a_1, a_2, \dots, a_N, b_2, \dots, b_N, c, e_{\max}].$$

The equations  $\mathbf{f}_1$  and  $\mathbf{f}_{2N}$  pertain to the endpoints whose corresponding error values are evaluated by taking the limits in (4.15). System (4.17) can be solved by implementing any numerical software tool or by implementing Newton–Raphson method. Alternatively, the optimized coefficients can be acquired by implementing the modified Remez algorithm to (4.17) with excluding  $\mathbf{f}'_k(\mathbf{v}), k = 2, 3, 4, \dots, L$ .

On the other hand, the problem formulation for the upper bounds, whose degree  $\hat{D} = 2N + 1$  for  $e = d$  and  $\hat{D} = 2N$  for  $e = r$  since  $\min\{b_n\}_{n=1}^N = \frac{1}{2}$ , is

formulated as

$$\left\{ \begin{array}{ll} f_k(\mathbf{v}) = e(x_k) - e_{\max} = 0, & \text{for } k = 2, 4, \dots, L, \\ f_k(\mathbf{v}) = e(x_k) = 0, & \text{for } k = 3, 5, \dots, L - 1, \\ f'_k(\mathbf{v}) = e'(x_k) = 0, & \text{for } k = 2, 3, 4, \dots, L, \\ f_1(\mathbf{v}) = e_0 - e_{\max} = 0, & \text{for } e \in \{d, r\}, \\ f_{2N+1}(\mathbf{v}) = a_1 - \frac{1}{\sqrt{2\pi}} = 0, & \text{for } e = r, \end{array} \right. \quad (4.18)$$

for which  $L = 2N$  for  $e = r$  and  $L = 2N + 2$  for  $e = d$ . The modified Remez algorithm can also be used by simply deleting the zero-derivative equations, i.e.,  $f'_k$ ,  $k = 2, 3, 4, \dots, L$ , from (4.18).

### 4.2.3 Total Error-Based Solution

The GKL approximation can be optimized in terms of the total error defined in (3.4) with  $R = [0, \infty)$  for the absolute error since  $d(x)$  converges to zero when  $x$  tends to infinity and with  $R = [0, R_2]$ , where  $R_2$  is some constant, for the relative error with  $b_1 = \frac{1}{2}$  since  $r(x)$  converges to a constant value when  $x$  tends to infinity. Starting from good initial guesses, which could be obtained using the corresponding optimized coefficients of the minimax optimization as mean values around which small random variance is introduced, the quasi-Newton algorithm can be implemented according to the theory presented in Section 3.2.

## 4.3 Applications and Performance Analysis

The developed approximations and bounds can be implemented in the different fields of communication theory. In general, the exponential approximation (4.3) and the GKL approximation (4.11) with the optimized coefficients can be used as substitutions to the intractable Gaussian  $Q$ -function whenever the application's mathematics defines them as tractable alternatives for it. Therefore, this section demonstrates the wide applicability of the proposed approximations and bounds in communication theory. In addition, it validates their accuracy by comparing their performance to key reference cases.

### 4.3.1 Applications

Whenever (4.3) or (4.11) is preferred, the optimized coefficients with the generalized expressions offer variety to tailor accuracy for the application or to use bounds. The most popular example of applications on the  $Q$ -function is the error probability analysis. More specifically, calculating the bit, symbol, and block error probabilities for various digital modulation schemes and coherent detection under AWGN channels.

On the one hand, the optimized coefficients for the exponential approximation can be used as one-to-one replacements for those used in the different applications utilizing (2.14) such as [27], [43], [44]. In particular, the elegance of the exponential approximation in the error performance analysis which encounters the  $Q$ -function, its integer powers, or its polynomials, is emphasized by unifying the evaluation of the average error probability of digital communication systems over the different fading channels in terms of the channel's moment generating function (MGF). In particular, the average SEP in coherent detection is calculated according to (2.8) whose conditional SEP is a polynomial of the  $Q$ -function which can be substituted directly by (4.3) and result in the unified MGF-based expression as

$$\bar{P}_s \approx \sum_{n=1}^N a_n \int_0^{\infty} \exp(-b_n \alpha^2 \gamma) \psi_{\gamma}(\gamma) d\gamma \quad (4.19)$$

$$= \sum_{n=1}^N a_n M_{\gamma}(-b_n \alpha^2), \quad (4.20)$$

where  $M_{\gamma}(s) = \int_0^{\infty} \exp(s\gamma) \psi_{\gamma}(\gamma) d\gamma$  is the MGF associated with the instantaneous SNR ( $\gamma$ ) and  $\alpha$  is a constant that depends on the digital modulation scheme.

Closed-form expressions for the average SEP are derived in [P2] using (4.20) together with the optimized coefficients over different fading channels, namely Nakagami- $m$ , Fisher–Snedecor  $\mathcal{F}^1$ ,  $\eta - \mu$ , and  $\kappa - \mu$  fading channels whose PDFs are defined in Table 2.1. It is important to note that increasing the number of exponentials in the summation (4.3) will not usually cause increased analytical complexity while increasing the accuracy sufficiently since summation and integration can be reordered in the expression as in (4.19) under certain conditions and thus, the integral is solved only once. In addition, the computational and/or analytical complexity of the exponential approximations/bounds for the polynomials of the  $Q$ -function and integer powers thereof is significantly less than using any other form of approximations from the literature including the exponential ones in [27]–[30]. This is because the reference approximations need to use linear combinations of multinomial expansions of the implemented approximation in order to evaluate the conditional SEP first (which is generally a polynomial of the  $Q$ -function), where none of the references has presented direct approximations or bounds for the polynomials and powers thereof of the  $Q$ -function.

On the other hand, the proposed GKL expression can be applied to substitute the original KL expression (2.18) in [116]–[120]. More specifically, it can be used to derive the sampling BER of BPSK in [116], to approximate the phase noise PDF in the system of [117], and to derive the coherent LoRa<sup>®</sup> SER under AWGN channel in [118]. Another application example is given in [P5], in which the average SEP is evaluated for coherent detection over  $\kappa - \mu$  fading. Moreover, GKL approximation/bound can also be used beyond communications. For example, it enables approximating the distribution functions of particles experiencing compound subdiffusion [119] and calculating the predictive error of the probability of failure [120]. It should be noted that the original

---

<sup>1</sup>It is worth mentioning that the general average SEP integral under Fisher–Snedecor  $\mathcal{F}$  derived in the attached publication [P2, Section IV.B] is slightly modified in this thesis summary part using [7, Eq. 10] as  $I_P(\alpha) \approx \sum_{n=1}^N a_n {}_1F_1\left(m; 1 - m_s; \frac{b_n \alpha^2 \bar{\gamma}(m_s - 1)}{m}\right) + \frac{\Gamma(-m_s)}{\beta(m, m_s)} \left(\frac{b_n \alpha^2 \bar{\gamma}(m_s - 1)}{m}\right)^{m_s} {}_1F_1\left(m + m_s; 1 + m_s; \frac{b_n \alpha^2 \bar{\gamma}(m_s - 1)}{m}\right)$ , for which  $\beta(\cdot, \cdot)$  and  ${}_1F_1(\cdot; \cdot; \cdot)$  denote beta and Kummer confluent hypergeometric functions, respectively, with  $m_s > 1$ . This minor error occurred in [P2] due to using the highly cited MGF [115], which was later pointed out and corrected in a newer publication [7] through only a footnote, which made this minor mistake hard to spot. The modification does not cause any visible effect on numerical results.



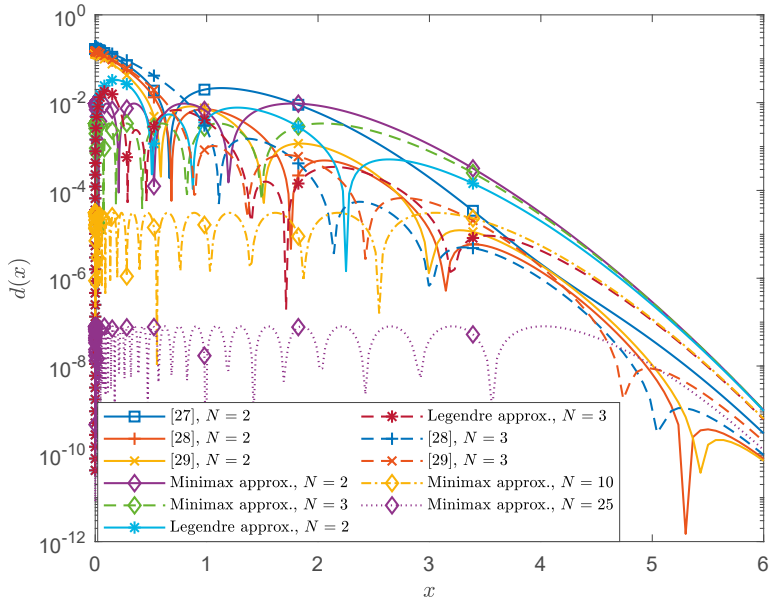
KL expression and the GKL expression have the same analytical complexity.

In addition, for the special case of the GKL expression with  $N = 1$  that results in the same original KL expression (2.18), the set of optimized coefficients can work as one-to-one substitutions for the original ones when utilized on the different applications such as those in [121]–[123] to achieve the same analytical results with significantly improved accuracy. More specifically, it can be used in [121] to tightly approximate the average probability of error for a MIMO recurrent neural network predictor, in [122] to derive the spectral efficiency of the round-robin scheduler, and in [123] to derive the caching distribution with arbitrary noise, path loss exponent of 4 and under Nakagami or Rician fading. Furthermore, since (4.16) can be used as a direct substitute for [114, Eq. 3], the novel GKL approximations are also useful to increase the accuracy for those applications that use [114, Eq. 3] such as [124]–[126]. In particular, they can be used in [124] to derive the packet error probability in a transmit beamforming system with imperfect feedback, and in [125] to derive the average SER of a DF cooperative transmission scenario with relay selection, and in [126] to evaluate the average SER in an unmanned aerial vehicles IRS-assisted communication system with imperfect phase compensation.

### 4.3.2 Performance Evaluation

The performance of the proposed approximations and bounds in (4.3) and (4.11) is evaluated in terms of accuracy. More specifically, the accuracy is interpreted by plotting the absolute and relative errors, which are defined in (3.1) and (3.2), respectively, for the proposed approximations together with the most relevant approximations which have been overviewed in Section 2.1.2.

The absolute error comparison in Fig. 4.1 illustrates the considerably gained accuracy by the exponential approximation (4.3) when using the optimized set of coefficients obtained by either implementing the minimax optimization presented in Section 4.1.2 or when using the optimized quadrature Legendre rule in terms of the mean error. The Quadrature Legendre rule is considered to formulate one of the tightest approximations among the various quadrature rules, as concluded in [P3]. The new coefficients for  $N = 2$  and  $N = 3$  not only provide lower minimax/global error but also adequate accuracy for the whole considered  $x$ -range, opposing to the exponential reference cases, which have

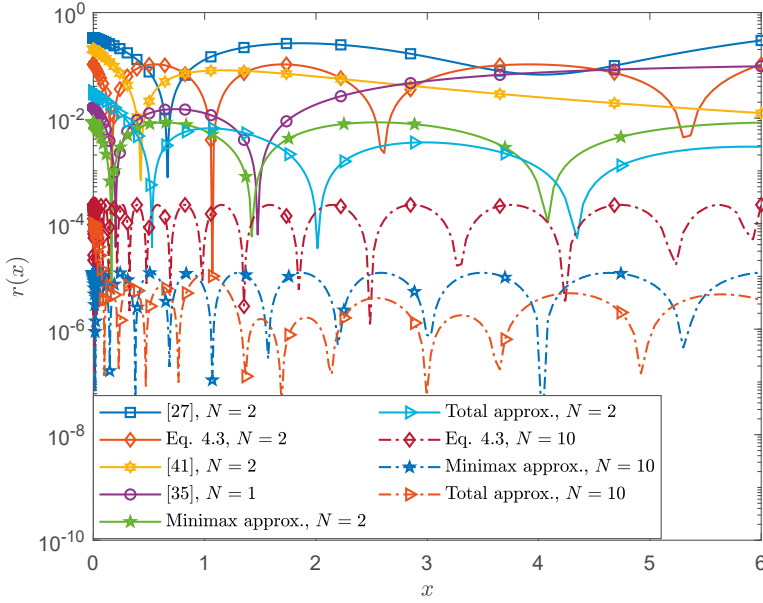


**Figure 4.1** Comparison between (4.3) and the references approximations [27]–[29] in terms of the absolute error.

poor accuracy for the lower values of the argument  $x$ . This accuracy can be increased significantly by simply increasing the number of exponential terms, like for  $N = 10$  and  $N = 25$ , without affecting the analytical complexity.

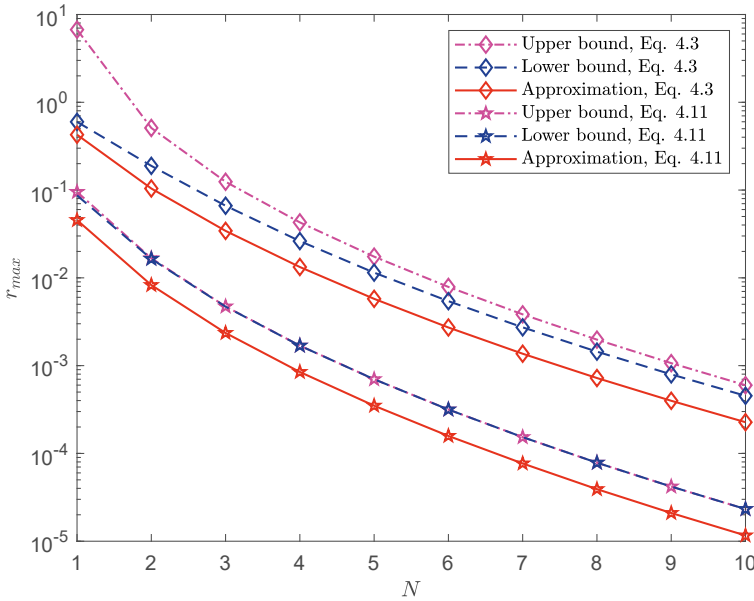
The relative error comparison in Fig. 4.2 depicts the high accuracy of the novel GKL approximation (4.11) with optimized coefficients in terms of the minimax and total error, where it outperforms all the considered key existing approximations, in addition to the novel exponential approximations with the minimax coefficients. The accuracy increases significantly when increasing  $N$ , as seen when comparing  $N = 2$  and  $N = 10$ . This observation is further investigated in Fig. 4.3 for both the exponential and the GKL expression, whose accuracy increases with increasing number of terms used in the expression, not only for the approximations but also for the lower and upper bounds.

In addition, the accuracy of the exponential approximations when integrated with another function, i.e., accuracy over fading, is investigated in [P2, Fig. 4]. This figure depicts a very small error and thus high accuracy. Furthermore, the performance of the developed exponential bounds is studied in [P1, Fig. 5] for the whole considered range of the argument. It shows that the proposed bounds not only outperform the other exponential bounds over the whole argument's



**Figure 4.2** Comparison between (4.11), (4.3), and the references approximations, [27], [35], [41] in terms of the relative error.

range but also outperform the other more complicated ones from the literature. More accuracy comparisons for both types of approximations can be found in [P



**Figure 4.3** Relative error of the exponential (4.3) and GKL (4.11) approximations and bounds.



## 5 CAPACITY INTEGRALS

Perusing the key references related to capacity analysis as summarized in Section 2.2, it is noted that the ergodic capacity has been studied thoroughly but using different comprehensive studies and analyses that are specific for each considered system separately and using different mathematical steps. The main limitation of the earlier research works is the lack of presence of a unified approach for analyzing the performance of any communication system in terms of ergodic capacity. As a result, this chapter's contributions are oriented toward developing unified tractable and highly accurate approximations and bounds for any communication system's ergodic capacity.

### 5.1 Mathematical Form and Origin

Since the instantaneous capacity,  $C$ , of any wireless system always exists and is calculated according to (2.24), an effective SNR ( $\gamma_{\text{eff}}$ ) whose average is  $\bar{\gamma}_{\text{eff}} \triangleq \text{E}[\gamma_{\text{eff}}]$  can always be calculated from  $C \triangleq \log_2(1 + \gamma_{\text{eff}})$ . The effective SNR could be the same as the actual SNR of the communication system. Assuming the instantaneous capacity with its PDF denoted by  $f_C(c)$  is conditioned on fading states, the ergodic capacity is calculated according to (2.25) as

$$\bar{C} = \text{E}[C] = \int_0^\infty c f_C(c) dc. \quad (5.1)$$

By changing the integration variable to  $z \triangleq (2^c - 1)/\bar{\gamma}_{\text{eff}}$  which corresponds to the random variable  $Z \triangleq \frac{2^C - 1}{\bar{\gamma}_{\text{eff}}} = \frac{\gamma_{\text{eff}}}{\bar{\gamma}_{\text{eff}}}$  whose PDF is denoted by  $f_Z(z)$ ,  $\bar{C}$  can be rewritten as

$$\bar{C} = \int_0^\infty \underbrace{\frac{\bar{\gamma}_{\text{eff}} f_C(\log_2(1 + \bar{\gamma}_{\text{eff}} z))}{\log_e(2) (1 + \bar{\gamma}_{\text{eff}} z)}}_{\triangleq f_Z(z)} \log_2(1 + \bar{\gamma}_{\text{eff}} z) dz. \quad (5.2)$$

The integral in (5.2) can be expanded using the well-known Riemann sum rule for which the integration interval  $[0, \infty)$  is divided into an infinite number of partitions that is then truncated into  $N$  partitions, each of length  $\delta$ . This results in approximating (5.2) as a sum of logarithmic functions as

$$\bar{C} \approx \sum_{n=1}^N a_n \log_2 (1 + b_n \bar{\gamma}_{\text{eff}}), \quad (5.3)$$

with

$$a_n \triangleq \frac{\delta \bar{\gamma}_{\text{eff}} f_C(\log_2(1 + \bar{\gamma}_{\text{eff}} n \delta))}{\log_e(2) (1 + \bar{\gamma}_{\text{eff}} n \delta)}, \quad (5.4)$$

and

$$b_n \triangleq n \delta. \quad (5.5)$$

It is noted from (5.4) that the set of coefficients  $\{a_n\}_{n=1}^N$  generally depends on  $\bar{\gamma}_{\text{eff}}$ . Nevertheless,  $f_C(c)$  can be represented in terms of  $f_Z(z)$  as

$$f_C(c) = \frac{2^c \log_e(2)}{\bar{\gamma}_{\text{eff}}} f_Z\left(\frac{2^c - 1}{\bar{\gamma}_{\text{eff}}}\right). \quad (5.6)$$

This results in  $a_n = \delta f_Z(n \delta)$  after substituting (5.6) into (5.4). Thus, if  $f_Z(\cdot)$  is independent of  $\bar{\gamma}_{\text{eff}}$ ,  $a_n$  is also independent of  $\bar{\gamma}_{\text{eff}}$ , which allow for the same coefficients per studied system to be used for the logarithmic approximation regardless of the value of  $\bar{\gamma}_{\text{eff}}$ . In practice, the majority of applications meet this condition, as has been noted from the wide range of applications studied in this thesis. The unified logarithmic approximation is also valid for systems where  $a_n$  depends on  $\bar{\gamma}_{\text{eff}}$ . However, different coefficients  $\{(a_n, b_n)\}_{n=1}^N$  are needed for each value of  $\bar{\gamma}_{\text{eff}}$ .

The proposed approximation (5.3) with  $N$  logarithmic terms is, in fact, a weighted sum of the Shannon capacities defined in (2.24) of basic static AWGN channels. Therefore, in terms of capacity, any communication system in the presence of fading is equivalent to a system with a scheduler that employs randomly one of  $N + 1$  parallel static channels for the transmission of each data block, or alternatively, with a scheduler that employs the parallel channels sequentially for data blocks with relative durations  $a_n$ ,  $n = 0, 1, \dots, N$ . For

this equivalent system, any Channel  $n$  have SNR of  $b_n \bar{\gamma}_{\text{eff}}$  and is selected with probability  $a_n$ , whereas Channel 0 represents a blocked channel ( $b_0 = 0$ ), with probability  $a_0 = 1 - \sum_{n=1}^N a_n$ . This interpretation of the ergodic capacity is illustrated in [P6, Fig. 1].

## 5.2 Developed Approximation and Bounds

The ergodic capacity of any communication system can be approximated using (5.3) by choosing appropriate values for the coefficients  $\{(a_n, b_n)\}_{n=1}^N$ . One possible yet inefficient choice is the Riemann sum coefficients which are given explicitly in (5.4) and (5.5) and lead to the usage of a very high number of logarithmic terms in order to achieve adequate accuracy. Therefore, (5.3) needs to be developed into an efficient and useful tool to be used in performance statistical analysis. This can be done by optimizing the coefficients  $\{(a_n, b_n)\}_{n=1}^N$  that correspond to the logarithmic expression (5.3) to directly approximate the ergodic capacity  $\bar{C} = C(1/\bar{\gamma}_{\text{eff}})/\log_e(2)$  of any communication system with high accuracy. The function  $C(\cdot)$  is referred to as the *generic capacity function*, which can be of any mathematical form. Nevertheless, for the analysis of most communication systems,  $C(\cdot)$  can be represented as the *generic capacity integral* as

$$C(x) \triangleq \int_0^\infty \log_e \left( 1 + \frac{t}{x} \right) f_Z(t) dt, \quad (5.7)$$

where  $f_Z(\cdot)$  is defined in (5.2).

Based on (5.3), a new family of simple functions is developed as

$$\tilde{C}(x) \triangleq \sum_{n=1}^N a_n \log_e \left( 1 + \frac{b_n}{x} \right), \quad (5.8)$$

for approximating or bounding  $C(x)$  by  $\tilde{C}(x)$  with proper coefficients choice. The approximation (5.8) directly stems from (5.3) where  $\tilde{C}(1/\bar{\gamma}_{\text{eff}})/\log_e(2)$  results in (5.3).

## 5.2.1 Nakagami and Lognormal Capacity Integrals

An alternative way for approximating the ergodic capacity is through developing approximations for specific capacity integrals that frequently appear as intermediate steps when evaluating the capacity of more complicated wireless systems. In other words, logarithmic approximations with optimized coefficients are calculated and used as building blocks to derive the capacity integral of many complicated communication systems [49]–[52], [54]–[65], [127]–[133]. Two specific capacity integrals are developed in this study, namely Nakagami capacity integral which originates from evaluating (5.7) for Nakagami fading, and lognormal capacity integral which originates from evaluating (5.7) for lognormal fading. Both Nakagami- $m$  and lognormal distributions are defined in Table 2.1.

Following the above discussion, the Nakagami capacity integral denoted by  $C_m(\cdot)$  is defined as

$$\begin{aligned} C_m(x) &\triangleq \int_0^\infty \frac{m^m}{\Gamma(m)} \log_e \left( 1 + \frac{t}{x} \right) t^{m-1} \exp(-m t) dt \\ &= \exp(mx) \sum_{k=0}^{m-1} \Gamma(-k, mx) (mx)^k, \end{aligned} \quad (5.9)$$

for  $x > 0$  [60, Eqs. 21 and 23] and  $C_m(x)$  is approximated by  $\tilde{C}_m(x)$  in (5.8). The function  $\Gamma(\zeta, x) = \int_x^\infty t^{\zeta-1} \exp(-t) dt$  is the upper incomplete gamma function [4, Eq. 6.5.3]. The closed-form expression above is only valid for the integer values of the fading parameter  $m$ . For  $m = 1$ , the Rayleigh capacity integral is obtained as

$$\begin{aligned} C_1(x) &= \int_0^\infty \log_e \left( 1 + \frac{t}{x} \right) \exp(-t) dt \\ &= \exp(x) E_1(x), \end{aligned} \quad (5.10)$$

for  $x > 0$  [49, Eqs. 4 and 5]. The function  $E_1(x) = \int_x^\infty \exp(-t)/t dt$  is the exponential integral [4, Eq. 5.1.1].

On the other hand, the lognormal capacity integral denoted by  $C_\sigma(\cdot)$  is



defined for  $\bar{\gamma} = \exp(\eta_\gamma + \frac{\sigma^2}{2})$  as

$$C_\sigma(x) \triangleq \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} \log_e \left( 1 + \frac{1}{x} \exp \left( \sqrt{2\sigma^2} t - \frac{\sigma^2}{2} \right) \right) \exp(-t^2) dt, \quad (5.11)$$

for  $x > 0$  [10, Eq. 29] and  $C_\sigma(x)$  is approximated by  $\tilde{C}_\sigma(x)$  in (5.8). This integral cannot be evaluated in closed form in terms of elementary functions.

## 5.2.2 Minimax Error-Based Solution

Using the theory presented in Section 3.1 on the best Chebyshev approximation, the generic capacity function  $C(\cdot)$  as well as the generic, Nakagami, and lognormal capacity integrals in (5.7), (5.9), and (5.11), respectively, which are all defined on the interval  $(0, \infty)$ , are approximated by the best logarithmic approximation of the form (5.8). In particular, the corresponding absolute and relative error functions which are defined using (3.1) and (3.2) respectively as

$$d(x) \triangleq \tilde{C}(x) - C(x), \quad (5.12)$$

$$r(x) \triangleq \frac{d(x)}{C(x)} = \frac{\tilde{C}(x)}{C(x)} - 1, \quad (5.13)$$

are expected to be bounded, uniform, and oscillating between equalized extrema of alternating signs. However, it is noted that for Nakagami and lognormal capacity integrals, and for all the considered applications in this chapter,  $d(x)$  actually diverges from the left, i.e.,  $\lim_{x \rightarrow 0} d(x) = \pm\infty$  which is equivalent to  $\lim_{\bar{\gamma}_{\text{eff}} \rightarrow \infty} d\left(\frac{1}{\bar{\gamma}_{\text{eff}}}\right)$  since  $x = \frac{1}{\bar{\gamma}_{\text{eff}}}$  as noted from approximating  $\bar{C} = C(1/\bar{\gamma}_{\text{eff}})/\log_e(2)$  by  $\tilde{C}(1/\bar{\gamma}_{\text{eff}})/\log_e(2)$ , unless the condition  $\sum_{n=1}^N a_n = 1$  holds. When this condition is met,  $d(x)$  will converge toward a constant value. The relative error function is always bounded.

The approximating function (5.8), which is an example of the generalized polynomial defined in (3.9), is of degree  $\hat{D} = 2N - 1$  for the absolute error since there is an imposed condition, while  $\hat{D} = 2N$  for the relative error. Thus  $\mathbf{u} = \{a_n, b_n\}_{n=1}^N$  is the set of coefficients to be optimized. The best approximation is derived by finding the optimized coefficients  $\mathbf{u}^* = \{a_n^*, b_n^*\}_{n=1}^N$

that satisfy (3.13) together with (3.3). The best logarithmic approximation with optimized coefficients is unique since the system of continuous functions  $\{\log_e \left(1 + \frac{b_n}{x}\right), n = 1, 2, \dots, N\}$  satisfies the Haar condition on  $(0, \infty)$  with a null set  $\{\infty\}$ , where each nontrivial linear combination of the form (5.8) has at most  $N - 1$  distinct zeros on  $(0, \infty)$ . The minimax logarithmic approximation will result in an error function that uniformly alternates between  $2N$  extrema points for  $e = d$  and between  $2N + 1$  extrema points for  $e = r$  [110], [134].

The two methods of implementation of the minimax error optimization, which are presented in Section 3.1.1 can be used to optimize the sets of coefficients  $\{(a_n, b_n)\}_{n=1}^N$  that correspond to (5.8). On the one hand, for the non-linear system of equations method, the solution is based on describing the expected uniform error function using the three equations subsets presented in Section 3.1.1 to formulate the following system of equations

$$\begin{cases} f_k(\mathbf{v}) = e(x_k) + (-1)^{k+\Xi} e_{\max} = 0, & \text{for } k = 2, 3, \dots, 2N, \\ f'_k(\mathbf{v}) = e'(x_k) = 0, & \text{for } k = 2, 3, \dots, 2N, \\ f_1(\mathbf{v}) = \lim_{x \rightarrow 0} e(x) + (-1)^{1+\Xi} e_{\max} = 0, & \text{for } e \in \{d, r\}, \\ \begin{cases} f_{2N+1}(\mathbf{v}) = \sum_{n=1}^N a_n - 1 = 0, & \text{for } e = d, \\ f_{2N+1}(\mathbf{v}) = \sum_{n=1}^N a_n b_n + r_{\max} - 1 = 0, & \text{for } e = r, \end{cases} \end{cases} \quad (5.14)$$

for which  $\Xi = 0$  for  $e = d$ ,  $\Xi = 1$  for  $e = r$  and

$$\mathbf{v} = [a_1, a_2, \dots, a_N, b_1, b_2, \dots, b_N, e_{\max}].$$

The equations  $f_1$  and  $f_{2N+1}$  in (5.14) pertain to the third subset which describes the endpoints that can be extrema points. In particular,  $f_1$  expresses the first extrema point, which occurs asymptotically at zero, i.e.,  $x_1$  is chosen to be a very small value near zero. On the other hand,  $f_{2N+1}$  expresses the imposed condition when  $e = d$ , while it expresses the  $(2N + 1)$ th extrema when  $e = r$ , where the relative error  $\sum$  converges to a constant value when  $x$  tends to infinity opposing to  $d(x)$  which converges to zero, i.e.,  $\lim_{x \rightarrow \infty} r(x) = \sum_{n=1}^N a_n b_n - 1 = -r_{\max}$ . The formulated system (5.14) can be solved by implementing the iterative Newton–Raphson method discussed in Section 3.1.1 to acquire the optimized solu-

tion or by directly implementing any numerical software tool, e.g., Matlab. Nevertheless, good initial guesses are needed to start any numerical iterative solver. Detailed information about the used initial guesses for the developed logarithmic approximations can be found in publication [P6].

The optimized coefficients corresponding to Nakagami (including Rayleigh) and lognormal capacity integrals are calculated for this study using the above methodology for a wide range of the parameters' values  $m$  and  $\sigma$ . The presented tool, together with the proposed methodology for obtaining the coefficients, can be repurposed to derive lower and upper bounds for the ergodic capacity using the presented framework in Section 3.1.2.

In particular, the optimized set of coefficients for the lower bound can be obtained by solving the following system of non-linear equations that describes the  $2N$  extrema for  $e = d$  and the  $2N+1$  extrema for  $e = r$ , including those that might occur at the endpoints of the considered range as

$$\left\{ \begin{array}{ll} f_k(\mathbf{v}) = e(x_k) + e_{\max} = 0, & \text{for } k = 2 + \Xi, 4 + \Xi, \dots, 2N, \\ f'_k(\mathbf{v}) = e'(x_k) = 0, & \text{for } k = 2, 3, \dots, 2N, \\ f_1(\mathbf{v}) = \lim_{x \rightarrow 0} e(x) + \Xi e_{\max} = 0, & \text{for } e \in \{d, r\}, \\ \left\{ \begin{array}{ll} f_{2N+1}(\mathbf{v}) = \sum_{n=1}^N a_n - 1 = 0, & \text{for } e = d, \\ f_{2N+1}(\mathbf{v}) = \sum_{n=1}^N a_n b_n + r_{\max} - 1 = 0, & \text{for } e = r, \end{array} \right. \end{array} \right. \quad (5.15)$$

for which  $\Xi = 0$  for  $e = d$  and  $\Xi = 1$  for  $e = r$ .

On the other hand, the optimized set of coefficients for the upper bound can be obtained by solving the following system of non-linear equations that describes the  $2N-1$  extrema for  $e = d$  and the  $2N+1$  extrema for  $e = r$  including

those that might occur at the endpoints of the considered range as

$$\left\{ \begin{array}{ll} f_k(\mathbf{v}) = e(x_k) - e_{\max} = 0, & \text{for } k = 2 + \Xi, 4 + \Xi, \dots, 2N, \\ f'_k(\mathbf{v}) = e'(x_k) = 0, & \text{for } k = 2, 3, \dots, 2N - 1 \text{ for } e = d \text{ or } 2N \text{ for } e = r, \\ f_1(\mathbf{v}) = \lim_{x \rightarrow 0} e(x) - \Xi e_{\max} = 0, & \text{for } e \in \{d, r\}, \\ f_{2N+1}(\mathbf{v}) = \sum_{n=1}^N a_n b_n - 1 = 0, & \text{for } e \in \{d, r\}, \\ f_{2N+2}(\mathbf{v}) = \sum_{n=1}^N a_n - 1 = 0, & \text{for } e = d, \end{array} \right. \quad (5.16)$$

for which  $\Xi = 1$  for  $e = d$  and  $\Xi = 0$  for  $e = r$ . It should be mentioned that for the absolute error, an extra condition  $\sum_{n=1}^N a_n b_n - 1 = 0$  is added in order to obtain the best upper bound.

The modified Remez algorithm, together with the Newton–Raphson method presented in Section 3.1.1, can also be used to calculate the optimized coefficients for the approximations, lower bounds, and upper bounds. This can be done by implementing Algorithm 1 on the system of equations in (5.14), (5.15), and (5.16) with excluding the equations describing the zero derivative of the corresponding error function at the extrema points.

### 5.2.3 Quadrature-Based Solution

As explained in Section 5.1, the logarithmic approximation (5.3) for the ergodic capacity (5.1), and (5.8) for generic capacity integral (5.7) thereof, result from applying the Riemann sum expansion. Nevertheless, the same logarithmic approximation but with more direct and efficient numerical coefficients, can be obtained from applying the different numerical quadrature integration methods [113] according to (4.10), for which the weighting function is  $W(t) = 1$  and  $f(t)$  is given in (5.7), (5.9), and (5.11) for the generic, Nakagami, and lognormal capacity integrals, respectively. The argument  $\theta$  used in (4.10) is exchanged with argument  $t$  herein in order to preserve manuscript's cohesiveness.

Therefore, the numerical coefficients corresponding to (5.8) are given for the

three capacity integrals, respectively as

$$\{(a_n, b_n)\}_{n=1}^N = \{(w_n f_Z(t_n), t_n)\}_{n=1}^N, \quad (5.17)$$

$$\{(a_n, b_n)\}_{n=1}^N = \left\{ \left( w_n \frac{m^m}{\Gamma(m)} t_n^{m-1} \exp(-m t_n), t_n \right) \right\}_{n=1}^N, \quad (5.18)$$

and

$$\{(a_n, b_n)\}_{n=1}^N = \left\{ \left( \frac{w_n}{\sqrt{\pi}} \exp(-t_n^2), \exp\left(\sqrt{2} \sigma^2 t_n - \frac{\sigma^2}{2}\right) \right) \right\}_{n=1}^N, \quad (5.19)$$

for which  $\{t_n\}_{n=1}^N$  are the nodes and  $\{w_n\}_{n=1}^N$  are the quadrature weights of the corresponding numerical integration method.

In particular, since the capacity integral is an improper convergent integral, it can be directly approximated using Gauss–Laguerre or Gauss–Hermite quadrature rules. Alternatively, the infinite integration interval can be limited by taking a large yet finite interval or by using variable transformation and then implementing the various numerical integration techniques including Newton–Cotes and Gaussian quadrature methods [4]. However, the logarithmic approximations and bounds with numerical coefficients are generally loose, especially in terms of global error. Therefore, they require a high number of logarithmic terms to achieve adequate accuracy.

### 5.3 Applications

The logarithmic approximation (5.8), together with the minimax optimization methodology, can be implemented directly to derive the ergodic capacity of any communication system. On the other hand, the logarithmic approximation of the Nakagami and lognormal capacity integrals ( $\tilde{C}_m(x)$  and  $\tilde{C}_\sigma(x)$ ) together with calculated optimized coefficients can be used as building blocks for deriving the ergodic capacity whenever possible in the different communication scenarios and will mostly result in the same logarithmic form as that of (5.8) as end expressions. While this chapter does introduce a general framework to directly implement the proposed tool for any communication system, its main focus is on the second approach, which takes advantage of the available opti-

mized coefficients for the Nakagami and lognormal capacity integrals and how to efficiently implement them for capacity analyses without the need to redo the optimization methodology to derive specific coefficients for each studied system.

In particular, a similar integral to that of the Nakagami capacity integral of the form

$$\begin{aligned} I_{m,\phi}(x) &\triangleq \int_0^\infty \log_e(1+xt) t^{m-1} \exp(-\phi t) dt \\ &= \phi^{-m} \Gamma(m) C_m(\phi/(xm)) \approx \phi^{-m} \Gamma(m) \tilde{C}_m(\phi/(xm)), \end{aligned} \quad (5.20)$$

is frequently seen in the intermediate steps when evaluating the ergodic capacity of many wireless communication systems [49]–[52], [54]–[65], [127]–[133]. Therefore,  $\tilde{C}_m(x)$  can be used in the process of evaluating the ergodic capacity in the different communications systems over small-scale fading which encounters  $I_{m,\phi}$ , after which  $\tilde{C}_\sigma(x)$  can be used to approximate the ergodic capacity when lognormal shadowing is present in the system.

More specifically, since the modified Bessel function of the first kind is included in the PDF of many of the fading distributions as noted from Table 2.1 and can be expanded as a power series [4, Eq. 9.6.12],  $\tilde{C}_m$  can be used then to approximate the ergodic capacity of SISO systems over the more complicated distributions as

$$\bar{C} \approx \frac{1}{\log_e(2)} \sum_{j=0}^{\infty} \Phi_j \tilde{C}_{m_j} \left( \frac{\theta_j}{\bar{\gamma}} \right) \approx \sum_{n=1}^N a_n \log_2(1 + b_n \bar{\gamma}), \quad (5.21)$$

where  $\Phi_j$ ,  $j = 0, 1, \dots$ , are constants whose values are listed in [P6, Table 1] together with  $m_j$  and  $\theta_j$  for the different fading distributions. The infinite series in (5.21) can be truncated to a few terms that achieve adequate accuracy, and the double-summation, when including the approximation sum, can be rearranged into a single summation which has the same form as in (5.3). Moreover,  $\tilde{C}_m(x)$  can be used to derive the ergodic capacity of a wide variety of point-to-point multi-antenna systems that encounter similar integrals as  $I_{m,\phi}(x)$  in (5.20). The ergodic capacity for some application examples such as SIMO, MISO, and MIMO systems with different diversity, combining, and multiplexing schemes are stated in [P6, Table 2] in closed form.

Another aspect of novelty about this unified tool is that it evaluates the ergodic capacity of various communication systems in the presence of shadowing together with small-scale fading and yields the same approximation as (5.3). More specifically, for a composite fading channel whose SNR is  $\gamma_{\text{eff}} = \psi s$ , for which  $\psi$  is a random variable that models small-scale fading and  $s$  is a random variable that models lognormal fading with  $\psi$  and  $s$  being independent, the average SNR of the small-scale fading is lognormally distributed. Therefore, the capacity integral can be evaluated as

$$\bar{C} = E_{\gamma_{\text{eff}}}[\log_2(1 + \gamma_{\text{eff}})] = E_s[E_{\gamma_{\text{eff}}|s}[\log_2(1 + \gamma_{\text{eff}})]]. \quad (5.22)$$

Above,  $\tilde{C}_m(x)$  can be used to evaluate the inner expectation, which is linked to small-scale fading. This yields a similar expression as in (5.11) when considering the shadowing effect represented by the outer expectation, which is then evaluated using  $\tilde{C}_\sigma(x)$ . This concept is not only limited to the lognormal shadowing, but it can also be applied to the other distributions that could model the shadowing effect, including inverse Gaussian, inverse Gamma, and Gamma distributions. It is worth noting that  $\tilde{C}_m$  can be used directly for the Gamma shadowing.

Multiple examples on deriving the ergodic capacity for some single-antenna and multi-antenna systems using  $\tilde{C}_m(x)$  and  $\tilde{C}_\sigma(x)$  for the concatenation concept presented in (5.22) are given in [P6]. In particular, a generic expression for the ergodic capacity of the more complicated small-scale distributions with lognormal shadowing in SISO systems is derived in [P6, Eq. 27], the ergodic capacity of MIMO spatial multiplexing [128], [129] is derived in [P6, Eq. 28], and of cooperative spatial multiplexing [130] over Rayleigh fading channels with lognormal shadowing is derived in [P6, Eq. 29].

The applicability and usefulness of the proposed approximations and bounds are not only limited to fundamental applications but also to the most timely communication systems. This importance is illustrated in [P6, Eqs. 30 and 31] by directly approximating the ergodic capacity of a NOMA system over the  $\alpha - \mu$  fading distribution [135] whose PDF is given in Table 2.1 by (5.8) together with implementing the minimax optimization method to calculate the optimized coefficients. The Nakagami capacity integral is not used for this application since its capacity analysis does not encounter  $I_{m,\phi}(x)$ . Furthermore,

$\tilde{C}_m(x)$  and  $\tilde{C}_\sigma(x)$  are used to derive the ergodic capacity expressions for a system with coordinated multipoint reception for mmwave uplink with blockages and Nakagami- $m$  fading [131], for a mmwave downlink NOMA system over fluctuating two-ray channels under general power allocation [132], and for RIS-assisted SISO system with correlated channels [133] in [P6, Eq. 32, Eq. 33, and Eq. 34, respectively].

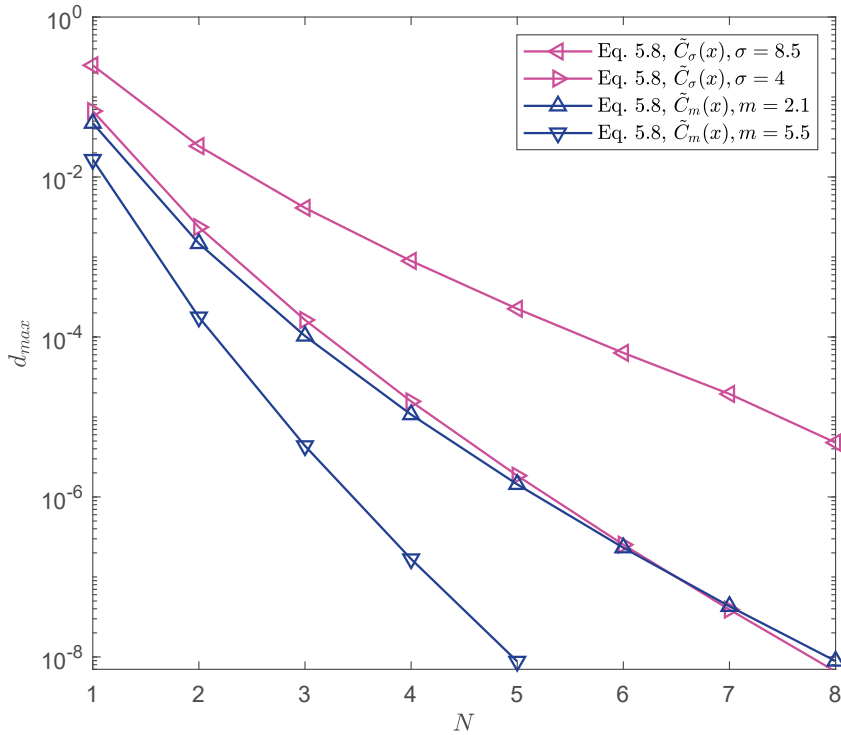
## 5.4 Performance Evaluation

The performance of the proposed logarithmic approximations and bounds in (5.8) is evaluated in terms of tractability and accuracy. Their mathematical tractability, which reveals insightful observations about the behavior of the communication system in terms of ergodic capacity, is assessed for the different communication systems in comparison with the corresponding capacity expressions derived in the literature. It has been illustrated in [P6, Section IV.E] that the proposed tool results in a unified, elegant, and simple expression for many communication systems such as those of the single-antenna and multi-antenna systems under small-scale fading, e.g., [P6, Table 2] or when combined with lognormal shadowing. The references, however, admit different analytical expressions for their ergodic capacity and utilize different mathematical evaluations.

In particular, the ergodic capacity is written in terms of complicated functions in the references such as exponential integral function, incomplete gamma function, Meijer  $G$ -function, Gaussian  $Q$ -function, or combinations of them. The timely applications considered in [P6] strongly support the significant gain in analytical tractability of the proposed tool, where the ergodic capacity expressions can be evaluated in this study in unified logarithmic form as opposed to in the references where they are evaluated using Meijer  $G$ -function and Fox  $H$ -function, both of which are unsolvable integrals. In contrast to the capacity expressions that consist of complex special functions, the tractable mathematical form of the proposed approximations reflects direct or even visual insights into the system's behavior, e.g., [P6 Table 2], draws some behavioral patterns about the effect of the different systems' parameters on their performance.

On the other hand, the accuracy of the proposed tool for the different com-





**Figure 5.1** Effect of increasing  $N$  on the accuracy of (5.8) for the Nakagami and lognormal capacity integrals in terms of global absolute error.

munication systems is assessed by mainly comparing the absolute error (5.12) resulting from applying the logarithmic approximations and bounds with those resulting from the previously derived capacity expressions in the corresponding references, in addition to the most relevant numerical approximations. More specifically, for the Nakagami capacity integral, Gauss–Laguerre quadrature rule is used to obtain the numerical coefficients, whereas for the lognormal capacity integral, Gauss–Hermite quadrature rule is used.

The extensive set of numerical results presented in [P6, Section V], confirms the high accuracy of (5.8) with the optimized coefficients for the various systems considered in this study. In particular, the sufficiently higher accuracy of the proposed approximations for both capacity integrals (5.9) and (5.11) is illustrated in [P6, Figs. 2 and 3] in terms of global absolute error and over the whole range of the argument, when compared to the numerical and refer-

ence approximations. Therefore, the same high accuracy is expected for the other communications systems whose ergodic capacity is evaluated using the Nakagami and lognormal capacity integrals. This is indeed confirmed by evaluating the performance of some example applications using the proposed tool in [P6, Fig. 5] for Rician fading channel with lognormal shadowing and in [P6, Fig. 6] for  $2 \times 2$  MIMO network over a shadowed-Rayleigh channel. It should be mentioned that this higher accuracy compared to the reference cases is achieved with much less analytical tractability.

The accuracy of the direct application of (5.8) to approximate the ergodic capacity in a NOMA system over  $\alpha - \mu$  fading distribution with two users is also investigated in terms of the absolute error in [P6, Fig. 7] and it depicts virtual alignment with the exact capacity evaluated in [135]. In addition, Fig. 5.1 herein shows that as the number of logarithmic terms in (5.8) increases, the accuracy of the approximation increases sufficiently for both capacity integrals (5.9) and (5.11), and for different values of the parameters  $m$  and  $\sigma$ .

## 6 RECONFIGURABLE INTELLIGENT SURFACES

This chapter analyzes the performance of two RIS-aided systems by deriving closed-form expressions for the outage probability, average symbol error probability, and ergodic capacity via employing the novel mathematical tools developed in Chapters 4 and 5, as well as in [P1]–[P6]. In addition, it investigates the behavior of the considered system models and draws some observations about the effect of the different systems' parameters on their performance.

### 6.1 RIS-Aided System with Spatially Correlated Channels

Most of the related theoretical studies accessible in the literature focus on the performance analysis when the fading channels are assumed to be i.i.d. [94]–[97]. Nevertheless, the REs of the RIS are closely located to each other. Hence, dependence is expected between the encountered REs' channels. Björnson and Sanguinetti in [136] introduce a more realistic SISO RIS-aided system setup, namely a spatially correlated Rayleigh fading system model, which can serve as a baseline for the theoretical research on RIS-aided communications. However, the performance of this model has not been investigated yet. Therefore, the first RIS-aided system studied in this thesis is the conventional system model depicted in Fig. 2.3 with the adopted correlated Rayleigh channels from [136]. The derived expressions in this section stem from the system and signal model introduced in Section 2.3.1. The developed tools in Chapters 4 and 5 are used to derive the different performance measures, namely the outage probability, average SEP, and ergodic capacity.

Based on the adopted spatially correlated Rayleigh fading model, the faded S-D, S-RIS, and RIS-D links are characterized as follows

$$u \sim \mathcal{N}_{\mathbb{C}}(0, \Omega_u), \mathbf{h} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \Lambda\mu_h\mathbf{R}), \mathbf{g} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \Lambda\mu_g\mathbf{R}), \quad (6.1)$$

for which their large-scale fading coefficients (path strength) are referred to respectively as  $\Omega_u$ ,  $\Lambda\mu_h$  and  $\Lambda\mu_g$ , with  $\Lambda = d_H \times d_V$  being the area of each RE whose horizontal width is denoted by  $d_H$  and its vertical height is denoted by  $d_V$ , whereas  $\mu_h$  and  $\mu_g$  being the average intensity attenuation of the corresponding link. The spatial channel correlation of the channels  $\mathbf{h}$  and  $\mathbf{g}$  is characterized by the spatial correlation matrix  $\mathbf{R} \in \mathbb{C}^{M \times M}$  which is assumed to be the same for both RIS links. Its elements are calculated according to [136, Eq. 10] as

$$\mathbf{R}_{n,m} = \text{sinc}\left(\frac{2\|\mathbf{a}_n - \mathbf{a}_m\|}{\lambda}\right), \quad n, m = 1, \dots, M, \quad (6.2)$$

for which  $\lambda$  is the wavelength, and  $\mathbf{a}_n$  and  $\mathbf{a}_m$  are the respective locations of the  $n$ th and  $m$ th elements w.r.t. the origin.

In order to maximize the end-to-end SNR when assuming perfect CSI at the RIS, optimal phase configuration is considered for the signal model introduced in Section 2.3.1 for this system by choosing reflection coefficient  $\zeta = 1$  and

$$\theta_i = \angle u - (\angle h_i + \angle g_i), \quad i = 1, \dots, M. \quad (6.3)$$

Hence,

$$|A| = \sum_{i=1}^M |h_i g_i| + |u|. \quad (6.4)$$

The analytical evaluation of the different performance measures normally includes dealing with the PDF of the end-to-end SNR ( $\gamma$ ), which is defined at the receiver for the considered system as

$$\gamma = \gamma_0 |A|^2, \quad (6.5)$$

for which  $\gamma_0 = E_s/N_0$  denotes the transmit SNR. From (6.5), it follows that finding the exact distribution density of  $\gamma$  is infeasible due to its complicated structure. Therefore, an approximating methodology is followed to determine the PDF of  $\gamma$  ( $\psi_\gamma$ ). In particular, two approximating schemes are adopted for the studied system model, namely the non-central chi-square and the Gamma distribution.

## Scheme 1: Non-Central Chi-Square Distributed Approximation

The channel's response in (6.4) is a sum of an independent random variable  $|u|$  plus weakly correlated/dependent random variables, i.e., weak correlation exists between the multiplicative terms  $|h_i g_i|, i = 1, 2, \dots, M$  although the channel vectors  $\mathbf{h}$  and  $\mathbf{g}$  are independent of each other. Therefore, (6.4) converges toward a Gaussian random variable according to the central limit theorem (CLT), and thus the end-to-end SNR in (6.5) is distributed according to the non-central chi-square distribution with one degree-of-freedom as

$$\psi_\gamma(x) \simeq \frac{1}{2\gamma_0\ell^2} \left(\frac{x}{\gamma_0\rho}\right)^{-\frac{1}{4}} \exp\left(-\frac{x+\rho\gamma_0}{2\gamma_0\ell^2}\right) I_{-\frac{1}{2}}\left(\frac{\sqrt{\rho x}}{\sqrt{\gamma_0\ell^2}}\right), \quad (6.6)$$

where  $\rho = (\mathbb{E}[|A|])^2$  and  $\ell^2 = \text{Var}[|A|]$  are evaluated respectively as

$$\rho = \left(\frac{M\pi\Lambda}{4}\sqrt{\mu_h\mu_g} + \frac{\sqrt{\pi\Omega_u}}{2}\right)^2, \quad (6.7)$$

and

$$\ell^2 = \sum_{i=1}^M \sum_{j=1}^M \frac{\Psi\Lambda^2\mu_h\mu_g\mathbf{R}_{i,j}}{4} \left[\Psi\mathbf{R}_{i,j} + \pi\right] + \frac{\Psi\Omega_u}{2}, \quad (6.8)$$

for which  $\Psi = \left(\frac{4-\pi}{2}\right)$ . Interested readers are kindly requested to refer to [P7] for the complete derivations of (6.7) and (6.8).

## Scheme 2: Gamma Distributed Approximation

The PDF of the non-central chi-square distributed end-to-end SNR for Scheme 1 above has a similar shape to that of the Gaussian PDF, where both have a single maximum, and their tails extend to infinity from the right side but are truncated to zero from the left side. As a result, the first term of a Laguerre series expansion, as stated in [137], can be used to closely approximate the PDF

of the non-central chi-square distribution as

$$\psi_\gamma(x) \simeq \frac{x^{\alpha-1}}{(\gamma_0 \beta)^\alpha \Gamma(\alpha)} \exp\left(-\frac{x}{\gamma_0 \beta}\right), \quad (6.10)$$

with

$$\alpha = \frac{(\mathbb{E}[|A|^2])^2}{\text{Var}[|A|^2]} \quad \text{and} \quad \beta = \frac{\text{Var}[|A|^2]}{\mathbb{E}[|A|^2]}, \quad (6.11)$$

where

$$\begin{aligned} \mathbb{E}[|A|^2] &= \sum_{i=1}^M \sum_{j=1}^M \Psi \frac{\Lambda^2 \mu_h \mu_g \mathbf{R}_{i,j}}{4} \left[ \Psi \mathbf{R}_{i,j} + \pi \right] \\ &\quad + \frac{\pi^2 \Lambda^2 M^2 \mu_h \mu_g}{16} + \frac{M \pi \Lambda \sqrt{\mu_h \mu_g \pi \Omega_u}}{4} + \Omega_u, \end{aligned} \quad (6.12)$$

and

$$\begin{aligned} \text{Var}[|A|^2] &= \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^M \sum_{m=1}^M \left[ \frac{\Psi^4 \Lambda^4 \mu_h^2 \mu_g^2}{16} \hat{R}^2 + \frac{\Psi^3 \Lambda^4 \mu_h^2 \mu_g^2 \pi}{16} \hat{R} \tilde{R} \right. \\ &\quad \left. + \frac{\Psi^2 \Lambda^4 \mu_h^2 \mu_g^2 \pi^2}{32} \left( \hat{R} + \frac{\tilde{R}^2}{2} \right) + \frac{\Psi \Lambda^4 \mu_h^2 \mu_g^2 \pi^3}{64} \tilde{R} + \frac{\Lambda^4 \mu_h^2 \mu_g^2 \pi^4}{256} \right] \\ &\quad + 2\sqrt{\pi \Omega_u} \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^M \left[ \left( \frac{\pi \Lambda}{4} \right)^3 (\mu_h \mu_g)^{\frac{3}{2}} + \Psi \frac{\pi^2 \Lambda^3 (\mu_h \mu_g)^{\frac{3}{2}}}{16} \bar{R} \right. \\ &\quad \left. + \Psi^2 \frac{\Lambda^3 (\mu_h \mu_g)^{\frac{3}{2}} \pi}{16} \bar{R}^2 \right] + 6\Omega_u \sum_{i=1}^M \sum_{j=1}^M \left[ \frac{\Psi^2 \Lambda^2 \mu_h \mu_g \mathbf{R}_{i,j}^2}{4} + \frac{\Psi \pi \Lambda^2 \mu_h \mu_g \mathbf{R}_{i,j}}{4} \right. \\ &\quad \left. + \frac{\pi^2 \Lambda^2 \mu_h \mu_g}{16} \right] + M \pi \Lambda \sqrt{\mu_h \mu_g} \beta^{\frac{3}{2}} \Gamma\left(\frac{5}{2}\right) - (\mathbb{E}[|A|^2])^2. \end{aligned} \quad (6.13)$$

Above  $\hat{R} = [\mathbf{R}_{i,j} \mathbf{R}_{k,m} + \mathbf{R}_{i,k} \mathbf{R}_{j,m} + \mathbf{R}_{i,m} \mathbf{R}_{j,k}]$ ,  $\tilde{R} = [\mathbf{R}_{i,j} + \mathbf{R}_{k,m} + \mathbf{R}_{i,k} + \mathbf{R}_{j,m} + \mathbf{R}_{i,m} + \mathbf{R}_{j,k}]$ , and  $\bar{R} = [\mathbf{R}_{j,k} + \mathbf{R}_{i,k} + \mathbf{R}_{i,j}]$ . Interested readers are kindly requested to refer to [P7] for the complete derivations of (6.12) and (6.13). This thesis uses slightly different notations than the original publication [P7] for consistency.

### 6.1.1 Performance Analysis

The performance of the conventional SISO RIS-aided system is investigated in terms of outage probability, average SEP, and ergodic capacity.

#### Outage Probability

Outage probability is the probability that the end-to-end instantaneous SNR takes a value less than a predefined threshold value ( $\gamma_{\text{th}}(\cdot)$ ). It is mathematically calculated as

$$P_O = \Pr(\gamma \leq \gamma_{\text{th}}) = \Psi_\gamma(\gamma_{\text{th}}), \quad (6.14)$$

where  $\Psi_\gamma$  is the cumulative distribution function (CDF) of the end-to-end instantaneous SNR. Therefore, the outage probability is calculated using the PDF of  $\gamma$  as

$$P_O = \int_0^{\gamma_{\text{th}}} \psi_\gamma(x) dx. \quad (6.15)$$

Here, two expressions for the outage probability are obtained. The first expression results from substituting the non-central chi-square distributed PDF (6.6) in (6.15) and using [138, Eq. 1] as

$$P_{O_{\chi^2}} = 1 - Q_{\frac{1}{2}}\left(\frac{\sqrt{\rho}}{\ell}, \frac{\sqrt{\gamma_{\text{th}}}}{\sqrt{\gamma_0 \ell}}\right), \quad (6.16)$$

for which  $Q_\nu(\cdot, \cdot)$  is the Marcum  $Q$ -function [138]. The second expression results from substituting the Gamma distributed PDF (6.10) in (6.15) and using [3, Eq. 8.350.1] as

$$P_{O_\Gamma} = \frac{\gamma\left(\alpha, \frac{x}{\gamma_0 \beta}\right)}{\Gamma(\alpha)}, \quad (6.17)$$

for which  $\gamma(\cdot, \cdot)$  is the lower incomplete Gamma function [3, Eq. 8.350.1].

## Symbol Error Probability

The average SEP under fading is generally calculated using (2.8), for which its conditional SEP for coherent detection is a polynomial of the  $Q$ -function of the form (4.2). Among the different modulation schemes available in Table 2.2, the average SEP for QPSK system is calculated using the non-central chi-square distribution and the developed exponential approximation (4.3), which can directly substitute the conditional SEP in (2.8). This leads to the average SEP to be evaluated in terms of the MGF of the non-central chi-square distribution [11] using (4.20) as

$$\bar{P}_{s_{\chi^2}} = \sum_{n=1}^N a_n \left( \frac{1}{1 + 2\gamma_0 b_n \ell^2} \right)^{\frac{1}{2}} \exp \left( - \frac{\rho \gamma_0 b_n}{1 + 2\gamma_0 b_n \ell^2} \right). \quad (6.18)$$

The corresponding coefficients  $\{(a_n, b_n)\}_{n=1}^N$  are the optimized coefficients in terms of the minimax or total error as reported in Section 4.1.

On the other hand, the average SEP for  $\aleph$ -ASK system is calculated using the Gamma distribution and the novel GKL approximation (4.11), which substitutes the conditional SEP from Table 2.2 in (2.8). This results in

$$\begin{aligned} \bar{P}_{s_{\Gamma}} = & \sum_{n=1}^N 2a_n \left( \frac{\aleph - 1}{\aleph} \right) \sqrt{\frac{(\aleph^2 - 1)}{6}} \frac{1}{(\gamma_0 \beta)^\alpha \Gamma(\alpha)} c_1^{-c_2 - \frac{3}{2}} \left( \sqrt{c_3} \Gamma \left( c_2 + \frac{3}{2} \right) \right. \\ & \left. \times {}_1F_1 \left( c_2 + \frac{3}{2}; \frac{3}{2}; \frac{c_3}{4c_1} \right) - \sqrt{c_1} \Gamma(c_2 + 1) \left( {}_1F_1 \left( c_2 + 1; \frac{1}{2}; \frac{c_3}{4c_1} \right) - 1 \right) \right), \end{aligned} \quad (6.19)$$

where  $c_1 = \frac{6b_n}{(\aleph^2 - 1)} + \frac{1}{\gamma_0 \beta}$ ,  $c_2 = \alpha - \frac{3}{2}$ ,  $c_3 = \frac{6c^2}{(\aleph^2 - 1)}$ , and  ${}_1F_1(\cdot; \cdot; \cdot)$  is the confluent hypergeometric function [3, Eq. 9.21]. The coefficients  $\{(a_n, b_n)\}_{n=1}^N$  and  $c$  are the optimized coefficients in terms of the minimax or total error as reported in Section 4.2.

## Ergodic Capacity

The ergodic capacity for this system can be evaluated based on the non-central chi-square distribution by substituting (6.6) in (2.26). This results in the same



mathematical expression as [139, Eq. 21] which is rewritten in [P7], with substituting  $\rho$  and  $\ell^2$  by the novel expressions (6.7) and (6.8), respectively. On the other hand, the ergodic capacity can be evaluated based on the Gamma distribution by substituting (6.10) in (2.26) as

$$\bar{C}_\Gamma = \frac{\exp\left(\frac{1}{\gamma_0\beta}\right)}{\log(2)} \sum_{j=1}^{\alpha} \Gamma\left(-\alpha + j, \frac{1}{\gamma_0\beta}\right) (\gamma_0\beta)^{j-\alpha}. \quad (6.20)$$

The ergodic capacity in (6.20) is valid for the integer values of  $\alpha$  only, leaving the case with non-integer values of  $\alpha$  intractable.

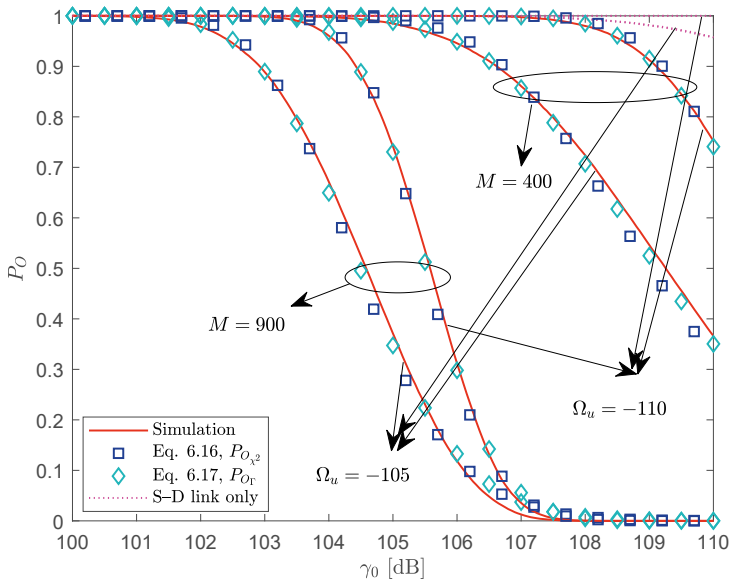
The two exact ergodic capacity expressions with respect to the adopted distribution are very complicated, and that in (6.20) is not even valid for all values of  $\alpha$ . As a result, the novel logarithmic expression proposed in Chapter 5 is adopted for this system for both distributions in order to derive unified and tight yet tractable approximations for  $\bar{C}_{\chi^2}$  and  $\bar{C}_\Gamma$  as

$$\tilde{C}(\gamma_0) \triangleq \sum_{n=1}^N a_n \log_2(1 + b_n \gamma_0). \quad (6.21)$$

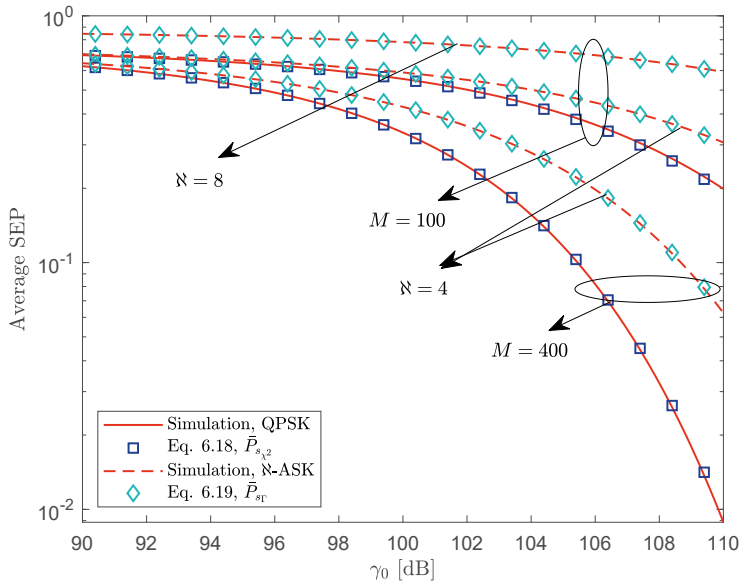
The coefficients  $\{(a_n, b_n)\}_{n=1}^N$  are found in terms of the minimax or total absolute/relative error. More information about calculating these coefficients in the minimax sense can be found in [P7].

### 6.1.2 Performance Evaluation

In this section, the analytical expressions derived above for evaluating the performance of the studied RIS-aided system with spatially correlated channels are verified and the effect of the different system's parameters on its performance is investigated. In particular, the adopted two approximating schemes (6.6) and (6.10), namely the non-central chi-square and the Gamma distribution show high agreement with the true PDF for different values of  $M$  as seen in [P7, Fig. 2]. Therefore, the derived expressions for the outage capacity, average SEP, and ergodic capacity for both distributions are expected to be well corroborated with the exact/simulated expressions. This is indeed confirmed by Fig. 6.1 for the outage probability and Fig. 6.2 for the average SEP and by



**Figure 6.1** The outage probability for different values of  $M$  and  $\Omega_u$  with  $\gamma_{th} = 10$  dB



**Figure 6.2** The average SEP for different values of  $M$  with  $\Omega_u = -110$  dB for QPSK and  $\aleph$ -ASK for different values of  $\aleph$ .

[P7, Fig. 3] for the ergodic capacity.

The outage probability derived in (6.16) for the non-central chi-square distribution and in (6.17) for the Gamma distribution is plotted in Fig. 6.1. It can be depicted from the figure that assisting the communication system with a RIS decreases the outage probability significantly. It also decreases further with increasing the number of reflecting elements equipped on the RIS, i.e., increasing  $M$  will decrease the transmitted power needed to achieve a certain outage probability. For example, at  $\Omega_u = -105$  dB, an outage probability of approximately 36% occurs with  $M = 400$  at  $\gamma_0 = 110$  dB, whereas it occurs at  $\gamma_0 = 105$  dB with  $M = 900$  dB. Thus, using 500 REs more saves 5 dB of transmitted power. In addition, as the strength of the direct path increases, the outage probability decreases sufficiently, e.g., for  $M = 900$  and  $\gamma_0 = 105$  dB, as  $\Omega_u$  increases from  $-110$  to  $-105$  dB, the outage probability improves by approximately 50%.

The average SEP derived in (6.18) based on the non-central chi-square distribution and using the developed exponential approximation (4.3) of the  $Q$ -function for QPSK modulation scheme is plotted in Fig. 6.2, together with the average SEP derived in (6.19) based on the Gamma distribution and using the novel GKL approximation (4.11) for  $\aleph$ -ASK modulation scheme. The figure shows that as  $M$  increases, the average SEP decreases indicating better performance. Moreover, as the modulation order of the  $\aleph$ -ASK increases, the average SEP achieved at a certain transmitted power increases indicating worse performance, e.g., for  $\gamma_0 = 110$  dB and  $M = 100$ , as  $\aleph$  changes from 4 to 8, the average SEP increases by about 93%. This is because higher-order modulation schemes provide higher data rates within a given bandwidth at the expense of reduced robustness to noise and interference, which in turn increases error probabilities.

In [P7], the ergodic capacity is thoroughly investigated for which the unified logarithmic approximation in (6.21) is derived for both distributions with optimized coefficients. This approximation shows excellent agreement with exact expressions with respect to the adopted distribution and with the simulated results. In general, the significant increase in ergodic capacity that occurs when a RIS is imposed on the communication system and when the number of REs equipped on the RIS is increased is confirmed in [P7, Figs. 3, 4, and 5]. In

addition, the capacity increases with increasing the transmitted power,  $\gamma_0$ , and with increasing the strength of the direct path. Furthermore, the impact of using a RIS in the communication process is best demonstrated when the direct path is weaker. More detailed observations about the behavior of the studied RIS-aided system can be found in [P7].

## 6.2 Systems with Multiple RISs over $\kappa - \mu$ Fading Channels

The single-RIS-aided systems have been investigated extensively in the literature [94]–[97], [136], [140]. On the other hand, limited efforts have been made to study the systems with wireless links that are aided by multiple RISs [102], [103]. The conducted studies for this generic system in the literature not only assumed i.i.d. fading channels within the same RIS, but also assumed i.i.d. fading channels across the different RISs. Nevertheless, this is not a practical assumption because RISs may be spread throughout a large geographical region. As a result, various RISs are likely to encounter non-identical and independent channels with the same or different fading distribution. On the other hand, the channels encountered by the REs per RIS may be considered identical since they are arranged on the same surface, i.e., the REs of a single RIS are located very close to each other. It should be mentioned that there could be a dependency between the encountered channels by the REs on each RIS. Nevertheless, this is left as a future research direction, for which, in this study, they are assumed to be independent per RIS for simplicity.

As a result, this section focuses on studying the performance of a relatively more realistic SISO system model that is aided by multiple RISs and a direct path. More specifically, it studies the generic RISs-based system model discussed in Section 2.3 and illustrated in Fig. 2.4, with independently but non-identically distributed (i.n.i.d.) fading channels across the geographically separated and distributed RISs. This implies that each RIS or even each hop (S-RIS<sub>*l*</sub> and RIS<sub>*l*</sub>-D hops) may experience different fading distributions. For that, a generic  $\kappa$ - $\mu$  fading model is adopted, whose PDF is defined in Table 2.1 and consists of the most widely-used fading scenarios, namely, Rayleigh, Rice, Nakagami- $m$ , and one-sided Gaussian distribution. The developed tools in Chapters 4 and 5 are used in this section to derive the different performance

measures, namely the outage probability, average SEP, and ergodic capacity.

For the generic RISs-aided system, the entries of the system's channels vectors  $\mathbf{h}_l = [h_{l,1}, \dots, h_{l,M_l}]^T$  and  $\mathbf{g}_l = [g_{l,1}, \dots, g_{l,M_l}]^T$  are slowly varying flat fading coefficients that are assumed to be statistically independent and identical per  $l$ th RIS. Their envelopes  $|h_{l,i}|$  and  $|g_{l,i}|$  follow the generalized  $\kappa$ - $\mu$  fading distribution. Nevertheless, the entries of  $\{\mathbf{h}_l\}_{l=1}^M$  and  $\{\mathbf{g}_l\}_{l=1}^M$  are nonidentical among the different distributed RISs. The average gains of the envelopes of the  $M_l$  fading coefficients per RIS $_l$  link are equal and are defined as

$$\sigma_{h_l}^2 = \text{E} [|h_{l,i}|^2] = \iota_0 \left( \frac{\varpi_0}{\varpi_{h_l}} \right)^{\xi_{h_l}}, \text{ for the S-RIS}_l \text{ link,}$$

and

$$\sigma_{g_l}^2 = \text{E} [|g_{l,i}|^2] = \iota_0 \left( \frac{\varpi_0}{\varpi_{g_l}} \right)^{\xi_{g_l}}, \text{ for the RIS}_l\text{-D link,}$$

whereas the average gain of the envelope of the S-D link is defined as

$$\sigma_u^2 = \text{E} [|u|^2] = \iota_0 \left( \frac{\varpi_0}{\varpi_u} \right)^{\xi_u},$$

where  $\iota_0$  is the reference path loss at the reference distance  $\varpi_0$ , and  $\varpi_j$  and  $\xi_j$ ,  $j \in \{h_l, g_l, u\}$  denote respectively the distance and path loss exponent of the corresponding link. The direct path follows Rayleigh distribution since no line-of-sight (LoS) in the S-D link is assumed.

Like in Section 6.1, the end-to-end SNR is maximized by choosing  $\zeta = 1$  and

$$\theta_{l,i} = \angle u - (\angle h_{l,i} + \angle g_{l,i}) \quad (6.22)$$

Hence,

$$|A| = \sum_{l=1}^M \sum_{i=1}^{M_l} |h_{l,i}| |g_{l,i}| + |u|. \quad (6.23)$$

By substituting (6.23) in (6.5), the end-to-end SNR ( $\gamma$ ) can be obtained for the studied system. Finding the exact PDF of  $\gamma$  is infeasible due to its significantly complicated structure. For that, an approximating methodology is followed to determine the PDF of  $\gamma$  ( $\psi_\gamma$ ). Since the first summation in (6.23) comprises

$M_l$  identical double  $\kappa$ - $\mu$  random variables, which all are positive, continuous, and independent, it converges toward a Gaussian random variable according to CLT. This causes the second summation in (6.23) to be a sum of the  $M$  resulting Gaussian variables plus a single Rayleigh random variable that will also converge toward a normally distributed random variable.

As a result, the corresponding PDF will have a single maximum, and its tails extend to infinity from the right side but is truncated to zero from the left side which allows this PDF to be further tightly approximated by the first term of a Laguerre-series expansion according to [137] as

$$\psi_{|A|}(x) \simeq \frac{x^\alpha}{\beta^{\alpha+1} \Gamma(\alpha + 1)} \exp\left(-\frac{x}{\beta}\right). \quad (6.24)$$

Therefore, the end-to-end SNR can be approximated by

$$\psi_\gamma(x) \simeq \frac{1}{2\beta^{\alpha+1} \Gamma(\alpha + 1)} \gamma_0^{-\frac{\alpha+1}{2}} x^{\frac{\alpha-1}{2}} \exp\left(-\sqrt{\frac{x}{\beta^2 \gamma_0}}\right), \quad (6.25)$$

where

$$\alpha = \frac{(\mathbb{E}[|A|])^2}{\text{Var}[|A|]} - 1, \quad (6.26)$$

$$\beta = \frac{\text{Var}[|A|]}{\mathbb{E}[|A|]}. \quad (6.27)$$

The mean and variance of  $|A|$  are given, respectively as

$$\begin{aligned} \mathbb{E}[|A|] &= \sum_{l=1}^M M_l \frac{\sigma_{h_l} \Gamma\left(\mu_{h_l} + \frac{1}{2}\right) \exp(-\kappa_{h_l} \mu_{h_l})}{\Gamma(\mu_{h_l}) ((1 + \kappa_{h_l}) \mu_{h_l})^{\frac{1}{2}}} \frac{\sigma_{g_l} \Gamma\left(\mu_{g_l} + \frac{1}{2}\right) \exp(-\kappa_{g_l} \mu_{g_l})}{\Gamma(\mu_{g_l}) ((1 + \kappa_{g_l}) \mu_{g_l})^{\frac{1}{2}}} \\ &\times {}_1F_1\left(\mu_{h_l} + \frac{1}{2}; \mu_{h_l}; \kappa_{h_l} \mu_{h_l}\right) {}_1F_1\left(\mu_{g_l} + \frac{1}{2}; \mu_{g_l}; \kappa_{g_l} \mu_{g_l}\right) + \sqrt{\frac{\pi \sigma_u^2}{4}}, \quad (6.28) \end{aligned}$$

and

$$\begin{aligned} \text{Var}[|A|] = & \sum_{l=1}^M M_n \left[ \sigma_{h_l}^2 \sigma_{g_l}^2 - \frac{\sigma_{h_l}^2 \Gamma^2\left(\mu_{h_l} + \frac{1}{2}\right) \exp(-2\kappa_{h_l} \mu_{h_l})}{\Gamma^2(\mu_{h_l}) (1 + \kappa_{h_l}) \mu_{h_l}} \right. \\ & \times \frac{\sigma_{g_l}^2 \Gamma^2\left(\mu_{g_l} + \frac{1}{2}\right) \exp(-2\kappa_{g_l} \mu_{g_l})}{\Gamma^2(\mu_{g_l}) (1 + \kappa_{g_l}) \mu_{g_l}} {}_1F_1^2\left(\mu_{h_l} + \frac{1}{2}; \mu_{h_l}; \kappa_{h_l} \mu_{h_l}\right) \\ & \left. \times {}_1F_1^2\left(\mu_{g_l} + \frac{1}{2}; \mu_{g_l}; \kappa_{g_l} \mu_{g_l}\right) \right] + \frac{4 - \pi}{4} \sigma_u^2, \end{aligned} \quad (6.29)$$

for which  $\kappa_{h_l}$  and  $\mu_{h_l}$  are the fading parameters of the S-RIS<sub>l</sub> hop and  $\kappa_{g_l}$  and  $\mu_{g_l}$  of the RIS<sub>l</sub>-D hop, while  ${}_1F_1(\cdot; \cdot; \cdot)$  is the confluent hypergeometric function of the first kind [3, Eq. 9.210.1]. The exact derivations of (6.25), (6.28), and (6.29) can be found in [P8].

### 6.2.1 Performance Analysis

The performance of the generic SISO RISs-aided system is investigated in terms of outage probability, average SEP, and ergodic capacity.

#### Outage Probability

The outage probability for a communication system assisted with multiple RISs over  $\kappa - \mu$  fading channels is calculated by deriving the CDF of  $\gamma$  first and then using (6.14) to obtain

$$P_O \simeq \frac{\gamma\left(\alpha + 1, \frac{1}{\beta} \sqrt{\frac{\gamma_{\text{th}}}{\gamma_0}}\right)}{\Gamma(\alpha + 1)}, \quad (6.30)$$

where  $\alpha$  and  $\beta$  are defined respectively in (6.26) and (6.27). The detailed derivation of (6.30) is available in [P8].

#### Symbol Error Probability

The average SEP of  $\aleph$ -QAM modulation scheme whose SEP is given in Table 2.2 is calculated as an example herein using the exponential approximation (4.3).

This leads to the average SEP to be calculated using (4.19). The closed-form expression is obtained by substituting (6.25) in (4.19) and using [3, Eq. 3.462.1] as

$$\begin{aligned} \bar{P}_s = & \frac{1}{2\beta^{\alpha+1}\Gamma(\alpha+1)} \sum_{n=1}^N a_n \left( \gamma_0 b_n \left( \frac{3}{\aleph-1} \right) \right)^{-\frac{\alpha+1}{2}} \left( \Gamma \left( \frac{\alpha+1}{2} \right) \right. \\ & \times {}_1F_1 \left( \frac{\alpha+1}{2}, \frac{1}{2}, \frac{1}{4\beta^2 \gamma_0 b_n \left( \frac{3}{\aleph-1} \right)} \right) - \left( \beta^2 \gamma_0 b_n \left( \frac{3}{\aleph-1} \right) \right)^{-\frac{1}{2}} \\ & \left. \times \Gamma \left( \frac{\alpha}{2} + 1 \right) {}_1F_1 \left( \frac{\alpha}{2} + 1, \frac{3}{2}, \frac{1}{4\beta^2 \gamma_0 b_n \left( \frac{3}{\aleph-1} \right)} \right) \right), \end{aligned} \quad (6.31)$$

for which  $\alpha$  and  $\beta$  are defined respectively in (6.26) and (6.27).

### Ergodic Capacity

The ergodic capacity for the studied system is derived by substituting (6.25) in (2.26). This results in the same analytical form as [97, Eq. 11] which is rewritten in [P8] with utilizing the novel expressions of  $\alpha$  and  $\beta$ , which are calculated for this system model using the derived mean and variance of the combined channel response in (6.28) and (6.29), respectively.

On the other hand, the ergodic capacity can also be tightly approximated herein using the unified logarithmic approximation proposed in Chapter 5 as

$$\tilde{C}(\gamma_0) \triangleq \sum_{n=1}^N a_n \log_2 (1 + b_n \gamma_0), \quad (6.32)$$

for which the coefficients  $\{(a_n, b_n)\}_{n=1}^N$  are found in terms of the minimax or total absolute/relative error by implementing the non-linear system of equations method or the modified Remez algorithm for the former, and quasi-Newton algorithm for the latter.

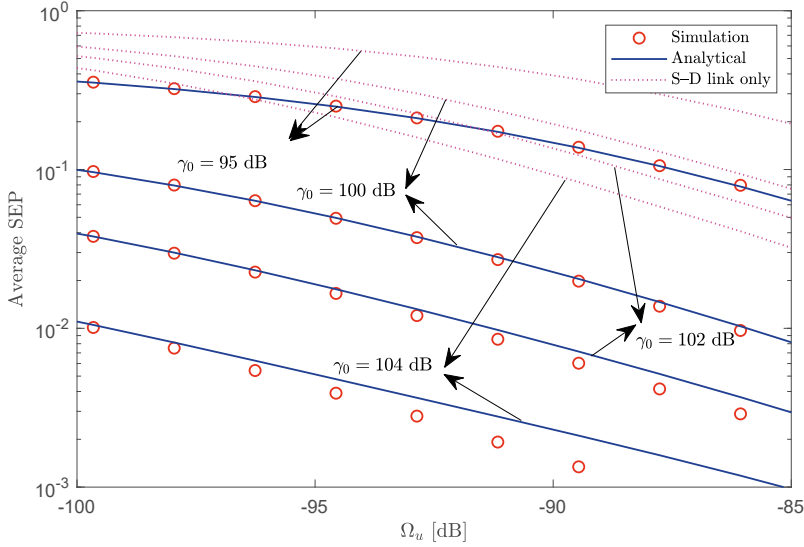


## 6.2.2 Performance Evaluation

In this section, the analytical expressions derived above for evaluating the performance of the studied RISs-aided system with  $\kappa - \mu$  fading channels are verified, and the effect of the different system parameters on its performance is investigated. In general, the results for the communication system that is assisted by multiple RISs over a generic fading distribution confirm those of the communication system that is assisted by a single RIS with correlated channels in Section 6.1. In particular, the adopted approximating scheme (6.24) coincides well with the true PDF for different values of  $M$  and different combinations of the fading distributions as seen in [P8, Fig. 2]. Therefore, the derived expressions for the outage probability, average SEP, and ergodic capacity are expected to be well corroborated with the exact/simulated expressions. This is indeed confirmed by [P8, Fig. 3].

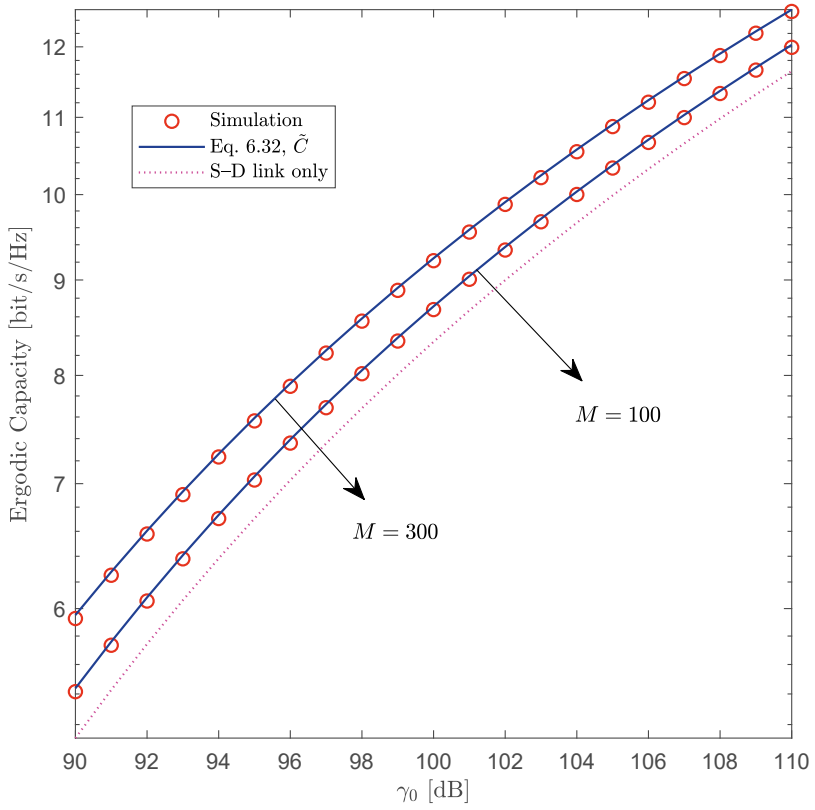
More specifically, it can be depicted from [P8, Fig. 3] that assisting the communication process with multiple RISs, improves the three performance measures. This can be noted by comparing the performance achieved when transmission takes place through the direct path only with that when the transmission is assisted with the RISs. The performance is improved even further by increasing the total number of REs ( $M$ ) equipped on the distributed RISs. In addition, [P8, Fig. 3 (b)] shows that as the order of the modulation scheme increases, the transmitted power  $\gamma_0$  needs to be increased to achieve a certain SEP requirement, e.g., to achieve an average SEP of 10% with  $M = 200$ , as  $\aleph$  changes from 16 to 64 for the  $\aleph$ -QAM,  $\gamma_0$  needs to be increased by approximately 6.4 dB.

Furthermore, the increase of the rate of change in the slope of the outage probability and the average SEP in Fig. 3 with increasing  $M$  indicates a higher diversity gain. The impact of the horizontal and vertical placement of the distributed RISs between S and D on the system's performance is studied in [P8, Fig. 4]. In general, it is noted that as the RISs are horizontally or vertically nearer to either S or D, the performance enhances, whereas the performance degrades when placing them near halfway between S and D, where path losses are maximized, and thus the RISs contribute less efficiently to the communication process.



**Figure 6.3** The average SEP versus the strength of the direct path ( $\Omega_u$ ) for different values of  $\gamma_0$  with  $M = 100$  and  $\aleph = 8$ .

Additionally, the effect of the strength of the direct path on the system's performance is studied in Fig. 6.3. The same system setup as that in [P8] is used herein but with  $d_u = 50$  and the  $x$  cartesian coordinate of each RIS ( $d_{x_n}$ ) divided by 2. Ranging the direct path loss exponent  $\xi_u$  from 3.2 to 4.2 results in decreasing the path strength  $\Omega_u$  from nearly  $-85$  dB to  $-100$  dB. It can be depicted from the figure that as the strength of the direct path increases, the average SEP improves sufficiently. The accuracy of the proposed logarithmic sum approximation of the ergodic capacity in (6.32) is illustrated in Fig. 6.4. The tractable approximation shows excellent agreement with the simulated ergodic capacity for different values of  $M$  with two logarithmic terms only whose coefficients are optimized in terms of the minimax absolute error.



**Figure 6.4** The ergodic capacity for different values of  $M$  with  $N = 2$  and  $\Omega_u = -72.5$  dB.



## 7 CONCLUSIONS AND FUTURE WORK

This thesis' final chapter summarizes the most important findings and outcomes and possible future research directions. As this study provides general methodologies and frameworks for approximation but limits their use to error probability and capacity analyses, the presented theory can be further extended to generate new tools for many other performance measures and special functions to make statistical performance analysis easier not only in the wireless communication field but also in many other statistical sciences.

### 7.1 Conclusions and Main Results

The primary goal of this thesis, which was to facilitate statistical performance analysis in terms of closed-form expressions, is achieved by developing several mathematical tools that can be efficiently implemented when analyzing systems' performance. The developed tools were intended to achieve very high accuracy levels while keeping their analytical forms as simple as possible. This necessitated certain types of optimization in order to select the appropriate coefficients to meet those requirements. Chapter 3 together with publications [P1]–[P6] presented two optimization criteria that are targeted in order to optimize the developed approximations, alongside their methods of implementation and the needed initial guesses. In addition, a generalization of these implementation methods was presented in order to generate new lower and upper bounds of the same mathematical form as the approximation under study. The provided general theory set the foundation for the following chapters.

Chapter 4, together with publications [P1]–[P5], answers the first and second research questions introduced in Section 1.1 where it presented two main parametric approximations/bounds for the Gaussian  $Q$ -function with multiple choices for the corresponding coefficients that stem from either implementing

the optimization methods presented in Chapter 3 or from applying the different quadrature numerical rules with or without optimizing them with respect to any optimization criterion. One of the approximations was extended to approximate integer powers, polynomials, or any well-behaved function of the  $Q$ -function with maintaining the same analytical form of the approximation, while the other was extended to approximate the  $Q$ -function with a simpler mathematical form using Taylor series expansion. Both types of approximations were evaluated for the absolute/relative minimax or total error by comparing them to relevant existing or numerical approximations. Numerical results proved the significant gain in accuracy achieved by the tractable approximating expressions, whose accuracy increases even further with increasing the number of approximating terms while keeping the analytical complexity the same. Their importance in communication theory was illustrated by implementing them in calculating the average SEP in closed form for many wireless communication systems, thus partially answering this thesis' fourth research question.

Chapter 5 and publication [P6] answer the first and third research questions introduced in Section 1.1, where a novel expression was proposed to directly approximate or bound the ergodic capacity of any communication system in a unified form without restrictions on the PDF of the channel. While this approach can be implemented in any communication system, it has been particularly implemented in this thesis to derive tight approximations/bounds for the Nakagami and lognormal distributions since they most frequently appear as building blocks for many complex communication systems. Therefore, instead of calculating the optimized coefficients for each studied system using the presented optimization methodologies in Chapter 3, the Nakagami and lognormal approximations/bounds with their optimized coefficients are used to evaluate the ergodic capacity of various communication systems that experience small-scale fading with or without the lognormal shadowing and allow for simplifying the complicated integrals encountered when evaluating the ergodic capacity in different communication scenarios. Furthermore, the applicability of the proposed approximations and bounds was verified by applying them to a wide range of classical and timely applications in communication theory, for which they illustrated sufficient accuracy where they virtually coincide with the true measures, thus partially answering this thesis' fourth research question.

Chapter 6 of this thesis, as well as publications [P7] and [P8], completes the answer to the fourth research question where it studied a very timely and promising technology as a use-case for the developed approximations and bounds, namely RIS-aided system with correlated channels, and multiple RISs-aided system with non-identical channels. Since the performance analysis of these systems is quite difficult, the proposed approximations are used to derive tight closed-form formulas for the average SEP and the ergodic capacity. Additionally, closed-form expressions for the outage probability were also acquired for the two system models. The derived expressions were plotted and demonstrated excellent agreement with the true measures. Moreover, they showed that better performance is achieved by increasing the transmitted SNR, the number of REs equipped on the RISs, the strength of the direct path, and placing the RIS closer to either the source or destination. Moreover, the results showed that the effect of using a RIS to aid the communication process is best depicted when the direct path is weak.

In summary, this thesis shows that statistical performance analysis of the various communication systems is indeed simplified with high levels of accuracy by using the proposed mathematical tools in this study. In fact, they even lead to closed-form solutions in situations where they are typically unobtainable. In addition, insightful observations are sometimes gained from the derived expressions.

## 7.2 Future Work

This thesis' results open up a number of new research areas. Firstly, the general approach of approximating any integral by a sum of the integrand evaluated at specific points (quadrature nodes) and multiplied by constants (weights), can be used to approximate other special functions than the ones studied in this thesis by using the novel methodologies of optimization presented in this study in order to replace the quadrature coefficients, i.e., nodes and weights, with new optimized coefficients that enable these approximations to be highly efficient and accurate tools for statistical evaluations.

In addition, most of the approximations available in the literature for the Gaussian  $Q$ -function can still be improved in terms of accuracy while keeping

the analytical complexity the same using a similar approach as that done for the KL approximation, for example. Moreover, the potentials of the novel logarithmic tool for evaluating the ergodic capacity can be further examined and extended to systems with non-Gaussian or non-additive noise, too, since this logarithmic approximation actually holds with any noise model, and probably nothing prevents one from applying the provided optimization methodologies to find the optimal coefficients.

In addition, the potential of the proposed tools can be further exploited by implementing them in the recent and future technologies that are rapidly emerging, including grant-free and semi-grant free NOMA, physical layer security, wireless power transfer, unmanned aerial vehicles assisted wireless communications systems, etc. Their applications do not only include wireless communication, but they can be implemented in many other scientific fields such as geothermal energy, astrophysics, and thermal vision. At last, the performance of the RIS-aided systems can be further explored for many system setups and assumptions using the new mathematical tools, e.g., systems with phase errors and a single RIS, systems with multiple RISs in the presence of dependency between the REs per RIS together with phase errors, and RIS-aided multi-antenna systems with beamforming. In conclusion, while significant progress has already been made in this field of research, there is still much potential to efficiently develop this research direction of mathematical analysis, for which further research is likely to be beneficial.



## REFERENCES

- [1] I. F. Akyildiz, A. Kak, and S. Nie, “6G and beyond: The future of wireless communications systems,” *IEEE Access*, vol. 8, pp. 133 995–134 030, Jul. 2020.
- [2] M. Cimmino, “An approximation of the finite line source solution to model thermal interactions between geothermal boreholes,” *International Communications in Heat and Mass Transfer*, vol. 127, p. 105 496, Oct. 2021.
- [3] I. Gradshteyn and I. Ryzhik, *Table of integrals, series, and products*, 7th ed. Elsevier/Academic Press, 2007.
- [4] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications, 1972.
- [5] B. Sklar, “Rayleigh fading channels in mobile digital communication systems. I: Characterization,” *IEEE Communications Magazine*, vol. 35, no. 7, pp. 90–100, Jul. 1997.
- [6] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels*, 2nd. John Wiley and Sons, Inc., Jan. 2005.
- [7] S. K. Yoo, P. C. Sofotasios, S. L. Cotton, S. Muhaidat, F. J. Lopez-Martinez, J. M. Romero-Jerez, and G. K. Karagiannidis, “A comprehensive analysis of the achievable channel capacity in  $\mathcal{F}$  composite fading channels,” *IEEE Access*, vol. 7, pp. 34 078–34 094, Feb. 2019.
- [8] N. Ermolova, “Moment generating functions of the generalized  $\eta - \mu$  and  $\kappa - \mu$  distributions and their applications to performance evaluations of communication systems,” *IEEE Communications Letters*, vol. 12, no. 7, pp. 502–504, Jul. 2008.

- [9] P. Ramirez-Espinosa, J. M. Moualeu, D. B. da Costa, and F. J. Lopez-Martinez, "The  $\alpha - \kappa - \mu$  shadowed fading distribution: Statistical characterization and applications," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Dec. 2019, pp. 1–6.
- [10] M.-S. Alouini and A. J. Goldsmith, "Area spectral efficiency of cellular mobile radio systems," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 4, pp. 1047–1066, Jul. 1999.
- [11] J. Proakis and M. Salehi, *Digital Communications*, 5th. McGraw-Hill Education, 2007.
- [12] D. Brennan, "Linear diversity combining techniques," *Proceedings of the IRE*, vol. 47, no. 6, pp. 1075–1102, Jun. 1959.
- [13] J. Craig, "A new, simple and exact result for calculating the probability of error for two-dimensional signal constellations," in *Proc. IEEE Military Communications Conference (MILCOM)*, vol. 2, Nov. 1991, pp. 571–575.
- [14] F. J. López-Martínez, R. Pawula, E. Martos-Naya, and J. Paris, "A clarification of the proper-integral form for the Gaussian  $Q$ -function and some new results involving the  $F$ -function," *IEEE Communications Letters*, vol. 18, no. 9, pp. 1495–1498, Sep. 2014.
- [15] F. Weinstein, "Simplified relationships for the probability distribution of the phase of a sine wave in narrow-band normal noise," *IEEE Transactions on Information Theory*, vol. 20, no. 5, pp. 658–661, Sep. 1974.
- [16] R. Pawula, S. Rice, and J. Roberts, "Distribution of the phase angle between two vectors perturbed by Gaussian noise," *IEEE Transactions on Communications*, vol. 30, no. 8, pp. 1828–1841, Aug. 1982.
- [17] P. Lee, "Computation of the bit error rate of coherent  $M$ -ary PSK with Gray code bit mapping," *IEEE Transactions on Communications*, vol. 34, no. 5, pp. 488–491, May 1986.
- [18] M. Irshid and I. Salous, "Bit error probability for coherent  $M$ -ary PSK systems," *IEEE Transactions on Communications*, vol. 39, no. 3, pp. 349–352, Mar. 1991.

- [19] J. Lu, K. Letaief, J. Chuang, and M. Liou, “ $M$ -PSK and  $M$ -QAM BER computation using signal-space concepts,” *IEEE Transactions on Communications*, vol. 47, no. 2, pp. 181–184, Feb. 1999.
- [20] X. Tang, M.-S. Alouini, and A. Goldsmith, “Effect of channel estimation error on  $M$ -QAM BER performance in Rayleigh fading,” *IEEE Transactions on Communications*, vol. 47, no. 12, pp. 1856–1864, Dec. 1999.
- [21] H. Suraweera and J. Armstrong, “Performance of OFDM-based dual-hop amplify-and-forward relaying,” *IEEE Communications Letters*, vol. 11, no. 9, pp. 726–728, Sep. 2007.
- [22] B. Zhu, “Asymptotic performance of composite lognormal- $x$  fading channels,” *IEEE Transactions on Communications*, vol. 66, no. 12, pp. 6570–6585, Dec. 2018.
- [23] A. Taherpour, M. Nasiri-Kenari, and S. Gazor, “Multiple antenna spectrum sensing in cognitive radios,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 2, pp. 814–823, Feb. 2010.
- [24] A. Mariani, A. Giorgetti, and M. Chiani, “Effects of noise power estimation on energy detection for cognitive radio applications,” *IEEE Transactions on Communications*, vol. 59, no. 12, pp. 3410–3420, Dec. 2011.
- [25] H. Chernoff, “A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations,” *The Annals of Mathematical Statistics*, vol. 23, no. 4, pp. 493–507, Dec. 1952.
- [26] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications*. Computer Science Press, 1985, vol. I.
- [27] M. Chiani, D. Dardari, and M. Simon, “New exponential bounds and approximations for the computation of error probability in fading channels,” *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 840–845, Jul. 2003.
- [28] P. Loskot and N. Beaulieu, “Prony and polynomial approximations for evaluation of the average probability of error over slow-fading

- channels,” *IEEE Transactions on Vehicular Technology*, vol. 58, no. 3, pp. 1269–1280, Mar. 2009.
- [29] O. Olabiya and A. Annamalai, “Invertible exponential-type approximations for the Gaussian probability integral  $Q(x)$  with applications,” *IEEE Wireless Communications Letters*, vol. 1, no. 5, pp. 544–547, Oct. 2012.
- [30] D. Sadhwani, R. Yadav, and S. Aggarwal, “Tighter bounds on the Gaussian  $Q$ -function and its application in Nakagami- $m$  fading channel,” *IEEE Wireless Communications Letters*, vol. 6, no. 5, pp. 574–577, Oct. 2017.
- [31] M. Wu, Y. Li, M. Gurusamy, and P. Kam, “A tight lower bound on the Gaussian  $Q$ -function with a simple inversion algorithm, and an application to coherent optical communications,” *IEEE Communications Letters*, vol. 22, no. 7, pp. 1358–1361, Jul. 2018.
- [32] M. López-Benítez and F. Casadevall, “Versatile, accurate, and analytically tractable approximation for the Gaussian  $Q$ -function,” *IEEE Transactions on Communications*, vol. 59, no. 4, pp. 917–922, Apr. 2011.
- [33] Q. Shi, “Novel approximation for the Gaussian  $Q$ -function and related applications,” in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2011, pp. 2030–2034.
- [34] G. Abreu, “Very simple tight bounds on the  $Q$ -function,” *IEEE Transactions on Communications*, vol. 60, no. 9, pp. 2415–2420, Sep. 2012.
- [35] G. Karagiannidis and A. Lioumpas, “An improved approximation for the Gaussian  $Q$ -function,” *IEEE Communications Letters*, vol. 11, no. 8, pp. 644–646, Aug. 2007.
- [36] J. Dyer and S. Dyer, “Corrections to, and comments on, An improved approximation for the Gaussian  $Q$ -function,” *IEEE Communications Letters*, vol. 12, no. 4, p. 231, Apr. 2008.

- [37] W. Jang, "A simple upper bound of the Gaussian  $Q$ -function with closed-form error bound," *IEEE Communications Letters*, vol. 15, no. 2, pp. 157–159, Feb. 2011.
- [38] Y. Isukapalli and B. Rao, "An analytically tractable approximation for the Gaussian  $Q$ -function," *IEEE Communications Letters*, vol. 12, no. 9, pp. 669–671, Sep. 2008.
- [39] C. Tellambura and A. Annamalai, "Efficient computation of  $\operatorname{erfc}(x)$  for large arguments," *IEEE Transactions on Communications*, vol. 48, no. 4, pp. 529–532, Apr. 2000.
- [40] Y. Chen and N. Beaulieu, "A simple polynomial approximation to the Gaussian  $Q$ -function and its application," *IEEE Communications Letters*, vol. 13, no. 2, pp. 124–126, Feb. 2009.
- [41] P. Borjesson and C. Sundberg, "Simple approximations of the error function  $Q(x)$  for communications applications," *IEEE Transactions on Communications*, vol. 27, no. 3, pp. 639–643, Mar. 1979.
- [42] V. N. Bao, L. Tuyen, and T. Huynh, "A survey on approximations of one-dimensional Gaussian  $Q$ -function," *REV Journal on Electronics and Communications*, vol. 5, Jan. 2015.
- [43] T. Tsiftsis, H. Sandalidis, G. Karagiannidis, and M. Uysal, "Optical wireless links with spatial diversity over strong atmospheric turbulence channels," *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 951–957, Feb. 2009.
- [44] M. McKay, A. Zanella, I. Collings, and M. Chiani, "Error probability and SINR analysis of optimum combining in Rician fading," *IEEE Transactions on Communications*, vol. 57, no. 3, pp. 676–687, Mar. 2009.
- [45] Q. Zhou, Y. Li, F. Lau, and B. Vucetic, "Decode-and-forward two-way relaying with network coding and opportunistic relay selection," *IEEE Transactions on Communications*, vol. 58, no. 11, pp. 3070–3076, Nov. 2010.

- [46] B. Choi and L. Hanzo, "Adaptive WHT aided QAM for fading channels subjected to impulsive noise," *IEEE Communications Letters*, vol. 17, no. 7, pp. 1317–1320, Jul. 2013.
- [47] M. López-Benítez, "Average of arbitrary powers of Gaussian  $Q$ -function over  $\eta$ - $\mu$  and  $\kappa$ - $\mu$  fading channels," *Electronics Letters*, vol. 51, no. 11, pp. 869–871, May 2015.
- [48] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [49] W. C. Y. Lee, "Estimate of channel capacity in Rayleigh fading environment," *IEEE Transactions on Vehicular Technology*, vol. 39, no. 3, pp. 187–189, Aug. 1990.
- [50] H. Zhang and T. Gulliver, "Capacity and error probability analysis for space-time block codes over fading channels," in *Proc. IEEE Pacific Rim Conference on Communications Computers and Signal Processing (PACRIM)*, vol. 1, Aug. 2003, pp. 102–105.
- [51] H. Zhang and T. A. Gulliver, "Closed form capacity expressions for space time block codes over fading channels," in *Proc. International Symposium on Information Theory (ISIT)*, Jun. 2004, p. 411.
- [52] L. Musavian, M. Nakhai, and A. Aghvami, "Capacity of space time block codes with adaptive transmission in correlated Rayleigh fading channels," in *Proc. IEEE Vehicular Technology Conference (VTC)*, vol. 3, May 2006, pp. 1511–1515.
- [53] M. Vu, "MISO capacity with per-antenna power constraint," *IEEE Transactions on Communications*, vol. 59, no. 5, pp. 1268–1274, May 2011.
- [54] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, Jun. 2003.
- [55] I. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–596, Sep. 1999.

- [56] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering, and keyhole," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2636–2647, Oct. 2003.
- [57] M.-S. Alouini and A. J. Goldsmith, "Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 4, pp. 1165–1181, Jul. 1999.
- [58] M. Kang and M.-S. Alouini, "Capacity of correlated MIMO Rayleigh channels," *IEEE Transactions on Wireless Communications*, vol. 5, no. 1, pp. 143–155, Jan. 2006.
- [59] A. Maaref and S. Aissa, "Closed-form expressions for the outage and ergodic Shannon capacity of MIMO MRC systems," *IEEE Transactions on Communications*, vol. 53, no. 7, pp. 1092–1095, Jul. 2005.
- [60] M.-S. Alouini and A. Goldsmith, "Capacity of Nakagami multipath fading channels," in *Proc. IEEE Vehicular Technology Conference (VTC). Technology in Motion*, vol. 1, May 1997, pp. 358–362.
- [61] G. Fraidenraich, O. Leveque, and J. M. Cioffi, "On the MIMO channel capacity for the Nakagami- $m$  channel," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3752–3757, Aug. 2008.
- [62] I.-S. Koh and T. Hwang, "Simple expression of ergodic capacity for Rician fading channel," *IEICE Transactions on Communications*, vol. 93-B, no. 6, pp. 1594–1596, Jul. 2010.
- [63] M. Kang and M.-S. Alouini, "Capacity of MIMO Rician channels," *IEEE Transactions on Wireless Communications*, vol. 5, no. 1, pp. 112–122, Jan. 2006.
- [64] D. E. Kontaxis, G. V. Tsoulos, and S. Karaboyas, "Ergodic capacity optimization for single-stream beamforming transmission in MISO Rician fading channels," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 2, pp. 628–641, Feb. 2013.

- [65] C. García-Corrales, F. J. Cañete Corripio, and J. Paris, “Capacity of  $\kappa - \mu$  shadowed fading channels,” *International Journal of Antennas and Propagation*, vol. 2014, pp. 1–8, Jul. 2014.
- [66] O. Oyman, R. U. Nabar, H. Bolcskei, and A. J. Paulraj, “Tight lower bounds on the ergodic capacity of Rayleigh fading MIMO channels,” in *Proc. Global Telecommunications Conference (GLOBECOM)*, vol. 2, Nov. 2002, pp. 1172–1176.
- [67] E. Gauthier, A. Yongacoglu, and J.-Y. Chouinard, “Capacity of multiple antenna systems in Rayleigh fading channels,” in *Proc. Canadian Conference on Electrical and Computer Engineering*, vol. 1, May 2000, pp. 275–279.
- [68] G. Alex, “Rayleigh fading multi-antenna channels,” *EURASIP Journal on Advances in Signal Processing*, vol. 2002, no. 3, pp. 316–329, Mar. 2002.
- [69] M. Dohler and H. Aghvami, “On the approximation of MIMO capacity,” *IEEE Transactions on Wireless Communications*, vol. 4, no. 1, pp. 30–34, Jan. 2005.
- [70] B. Banerjee, A. Abu Al Haija, C. Tellambura, and H. A. Suraweera, “Simple and accurate low SNR ergodic capacity approximations,” *IEEE Communications Letters*, vol. 22, no. 2, pp. 356–359, Feb. 2018.
- [71] O. Waqar, M. Ghogho, and D. McLernon, “Tight bounds for ergodic capacity of dual-hop fixed-gain relay networks under Rayleigh fading,” *IEEE Communications Letters*, vol. 15, no. 4, pp. 413–415, Apr. 2011.
- [72] D. B. da Costa and S. Aissa, “Capacity analysis of cooperative systems with relay selection in Nakagami- $m$  fading,” *IEEE Communications Letters*, vol. 13, no. 9, pp. 637–639, Sep. 2009.
- [73] A. Laourine, A. Stephenne, and S. Affes, “Estimating the ergodic capacity of log-normal channels,” *IEEE Communications Letters*, vol. 11, no. 7, pp. 568–570, Jul. 2007.
- [74] ———, “On the capacity of log-normal fading channels,” *IEEE Transactions on Communications*, vol. 57, no. 6, pp. 1603–1607, Jun. 2009.



- [75] F. Heliot, X. Chu, R. Hoshyar, and R. Tafazolli, "A tight closed-form approximation of the log-normal fading channel capacity," *IEEE Transactions on Wireless Communications*, vol. 8, no. 6, pp. 2842–2847, Jun. 2009.
- [76] G. Pan, E. Ekici, and Q. Feng, "Capacity analysis of log-normal channels under various adaptive transmission schemes," *IEEE Communications Letters*, vol. 16, no. 3, pp. 346–348, Mar. 2012.
- [77] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, Aug. 2019.
- [78] Z. Mao, M. Peng, and X. Liu, "Channel estimation for reconfigurable intelligent surface assisted wireless communication systems in mobility scenarios," *China Communications*, vol. 18, no. 3, pp. 29–38, Mar. 2021.
- [79] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization," *IEEE Wireless Communications Letters*, vol. 9, no. 4, pp. 518–522, Apr. 2020.
- [80] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent reflecting surface meets OFDM: Protocol design and rate maximization," *IEEE Transactions on Communications*, pp. 1–1, Mar. 2020.
- [81] Z.-Q. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive MIMO," *IEEE Wireless Communications Letters*, vol. 9, pp. 210–214, May 2020.
- [82] S. V. Hum and J. Perruisseau-Carrier, "Reconfigurable reflectarrays and array lenses for dynamic antenna beam control: A review," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 1, pp. 183–198, Jan. 2014.
- [83] H. Yang *et al.*, "A programmable metasurface with dynamic polarization, scattering and focusing control," *International Journal of Scientific Reports.*, vol. 6, Oct. 2016.

- [84] N. Kaina, M. Dupré, G. Lerosey, and M. Fink, “Shaping complex microwave fields in reverberating media with binary tunable metasurfaces,” *International Journal of Scientific Reports.*, vol. 4, no. 6693, Oct. 2014.
- [85] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, “Reconfigurable intelligent surfaces for energy efficiency in wireless communication,” *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 4157–4170, Jun. 2019.
- [86] Q. Wu and R. Zhang, “Beamforming optimization for intelligent reflecting surface with discrete phase shifts,” in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2019, pp. 7830–7833.
- [87] C. Pan, H. Ren, K. Wang, W. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, “Multicell MIMO communications relying on intelligent reflecting surfaces,” *IEEE Transactions on Wireless Communications*, Jun. 2020.
- [88] M. Al-Jarrah, A. Al-Dweik, E. Alsusa, Y. Iraqi, and M.-S. Alouini, “IRS-assisted UAV communications with imperfect phase compensation,” *IEEE Transactions on Wireless Communications*, 2021.
- [89] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, “Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 14 960–14 973, Oct. 2020.
- [90] A. Almohamad, A. M. Tahir, A. Al-Kababji, H. M. Furqan, T. Khat-tab, M. O. Hasna, and H. Arslan, “Smart and secure wireless communications via reflecting intelligent surfaces: A short survey,” *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1442–1456, Sep. 2020.
- [91] V. C. Thirumavalavan and T. S. Jayaraman, “BER analysis of reconfigurable intelligent surface assisted downlink power domain NOMA system,” in *Proc. International Conference on Communication Systems and Networks (COMSNETS)*, Mar. 2020, pp. 519–522.

- [92] M. Badiu and J. P. Coon, “Communication through a large reflecting surface with phase errors,” *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 184–188, Feb. 2020.
- [93] Q. Wu and R. Zhang, “Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network,” *IEEE Communications Magazine*, vol. 58, no. 1, pp. 106–112, Jan. 2020.
- [94] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M. Alouini, and R. Zhang, “Wireless communications through reconfigurable intelligent surfaces,” *IEEE Access*, vol. 7, pp. 116 753–116 773, Aug. 2019.
- [95] D. Kudathanthirige, D. Gunasinghe, and G. Amarasuriya, “Performance analysis of intelligent reflective surfaces for wireless communication,” in *Proc. IEEE International Conference on Communications (ICC)*, Jun. 2020.
- [96] A.-A. A. Boulogeorgos and A. Alexiou, “Performance analysis of reconfigurable intelligent surface-assisted wireless systems and comparison with relaying,” *IEEE Access*, vol. 8, pp. 94 463–94 483, May 2020.
- [97] A. Salhab and M. Samuh, “Accurate performance analysis of reconfigurable intelligent surfaces over Rician fading channels,” *IEEE Wireless Communications Letters*, vol. 10, no. 5, pp. 1051–1055, May 2021.
- [98] Q. Tao, J. Wang, and C. Zhong, “Performance analysis of intelligent reflecting surface aided communication systems,” *IEEE Communications Letters*, vol. 24, no. 11, pp. 2464–2468, Nov. 2020.
- [99] N. K. Kundu and M. R. McKay, “RIS-assisted MISO communication: Optimal beamformers and performance analysis,” in *IEEE Global Communications Workshops (GC Wkshps)*, Dec. 2020, pp. 1–6.
- [100] P. Dharmawansa, S. Atapattu, and M. D. Renzo, “Performance analysis of a two-tile reconfigurable intelligent surface assisted  $2 \times 2$  MIMO system,” *IEEE Wireless Communications Letters*, vol. 10, no. 3, pp. 493–497, Mar. 2021.
- [101] G. Xiao, T. Yang, C. Huang, X. Wu, H. Feng, and B. Hu, “Average rate approximation and maximization for RIS-assisted multi-user

- MISO system,” *IEEE Wireless Communications Letters*, vol. 11, no. 1, pp. 173–177, Jan. 2022.
- [102] D. L. Galappaththige, D. Kudathanthirige, and G. Amarasuriya, “Performance analysis of distributed intelligent reflective surface aided communications,” in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Dec. 2020.
- [103] L. Yang, Y. Yang, D. B. da Costa, and I. Trigui, “Outage probability and capacity scaling law of multiple RIS-aided networks,” *IEEE Wireless Communications Letters*, vol. 10, no. 2, pp. 256–260, Feb. 2021.
- [104] Y. Fang, S. Atapattu, H. Inaltekin, and J. Evans, “Optimum reconfigurable intelligent surface selection for indoor and outdoor communications,” Dec. 2020.
- [105] I. Yildirim, A. Uyrus, and E. Basar, “Modeling and analysis of reconfigurable intelligent surfaces for indoor and outdoor applications in future wireless networks,” *IEEE Transactions on Communications*, vol. 69, no. 2, pp. 1290–1301, Feb. 2021.
- [106] T. N. Do, G. Kaddoum, T. L. Nguyen, D. B. da Costa, and Z. J. Haas, “Multi-RIS-aided wireless systems: Statistical characterization and performance analysis,” *IEEE Transactions on Communications*, vol. 69, no. 12, pp. 8641–8658, Dec. 2021.
- [107] N. I. Akhiezer and C. J. Hyman, *Digital Communication over Fading Channels*, 6th. New York: F. Ungar Pub. Co., 1956.
- [108] C. B. Dunham, “Chebyshev approximation by exponential-polynomial sums,” *Journal of Computational and Applied Mathematics*, vol. 5, no. 1, pp. 53–57, Mar. 1979.
- [109] A. Haar, “Die minkowskische geometrie and die annäherung an stetige funktionen,” *Mathematische Annalen*, no. 78, pp. 249–311, Dec. 1917.
- [110] C. Dunham, “Families satisfying the Haar condition,” *Journal of Approximation Theory*, vol. 12, pp. 291–298, Nov. 1974.
- [111] M. J. D. Powell, *Approximation Theory and Methods*. Cambridge University Press, May 1981.

- [112] W. Cody, W. Fraser, and J. Hart, “Handbook series methods of approximation. rational Chebyshev approximation using linear equations.,” *Numerische Mathematik*, vol. 12, pp. 242–251, 1968.
- [113] P. Davis and P. Rabinowitz, *Methods of Numerical Integration*, 2nd ed. Academic Press, 1984.
- [114] Y. Isukapalli and B. D. Rao, “An analytically tractable approximation for the Gaussian  $Q$ -function,” *IEEE Communications Letters*, vol. 12, no. 9, pp. 669–671, Sep. 2008.
- [115] S. Yoo, S. Cotton, P. Sofotasios, M. Matthaiou, M. Valkama, and G. Karagiannidis, “The Fisher–Snedecor  $\mathcal{F}$  distribution: A simple and accurate composite fading model,” *IEEE Communications Letters*, vol. 21, no. 7, pp. 1661–1664, Jul. 2017.
- [116] W. M. Jang, “A simple performance approximation of general-order rectangular QAM with MRC in Nakagami- $m$  fading channels,” *IEEE Transactions on Vehicular Technology*, vol. 62, no. 7, pp. 3457–3463, Sep. 2013.
- [117] S. Lin, Z. Wang, J. Xiong, Y. Fu, J. Jiang, Y. Wu, Y. Chen, C. Lu, and Y. Rao, “Rayleigh fading suppression in one-dimensional optical scatters,” *IEEE Access*, vol. 7, pp. 17 125–17 132, Jan. 2019.
- [118] O. Afisiadis, S. Li, J. Tapparel, A. Burg, and A. Balatsoukas-Stimming, “On the advantage of coherent LoRa detection in the presence of interference,” *IEEE Internet of Things Journal*, vol. 8, no. 14, pp. 11 581–11 593, Jul. 2021.
- [119] A. Shalchi and V. Arendt, “Distribution functions of energetic particles experiencing compound subdiffusion,” *Astrophys. J.*, vol. 890, no. 2, p. 147, Feb. 2020.
- [120] J. Wauters, I. Couckuyt, and J. Degroote, “A new surrogate-assisted single-loop reliability-based design optimization technique,” *Structural and Multidisciplinary Optimization*, vol. 63, no. 6, pp. 2653–2671, Apr. 2021.

- [121] C. Potter, G. Venayagamoorthy, and K. Kosbar, “RNN based MIMO channel prediction,” *Signal Processing*, vol. 90, no. 2, pp. 440–450, Feb. 2010.
- [122] J. Wu, N. B. Mehta, A. F. Molisch, and J. Zhang, “Unified spectral efficiency analysis of cellular systems with channel-aware schedulers,” *IEEE Transactions on Communications*, vol. 59, no. 12, pp. 3463–3474, Dec. 2011.
- [123] D. Malak, M. Al-Shalash, and J. Andrews, “Optimizing content caching to maximize the density of successful receptions in device-to-device networking,” *IEEE Transactions on Communications*, vol. 64, no. 10, pp. 4365–4380, Oct. 2016.
- [124] Y. Isukapalli and B. D. Rao, “Packet error probability of a transmit beamforming system with imperfect feedback,” *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2298–2314, Apr. 2010.
- [125] M. Seyfi, S. Muhaidat, J. Liang, and M. Dianati, “Effect of feedback delay on the performance of cooperative networks with relay selection,” *IEEE Transactions on Wireless Communications*, vol. 10, no. 12, pp. 4161–4171, Dec. 2011.
- [126] M. A. Al-Jarrah, E. Alsusa, A. Al-Dweik, and M.-S. Alouini, “Performance analysis of wireless mesh backhauling using intelligent reflecting surfaces,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 6, pp. 3597–3610, Jun. 2021.
- [127] D. Castanheira and A. Gameiro, “Distributed MISO system capacity over Rayleigh flat fading channels,” in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2008, pp. 1–5.
- [128] L. Yang, “On the capacity of MIMO Rayleigh fading channels with log-normal shadowing,” in *Proc. Congress on Image and Signal Processing*, vol. 5, May 2008, pp. 479–482.
- [129] D. Shen, A. Lu, Y. Cui, F. Kuang, X. Zhang, K. Wu, and J. Yao, “On the channel capacity of MIMO Rayleigh-lognormal fading channel,” in *Proc. International Conference on Microwave and Millimeter Wave Technology (ICMMT)*, May 2010, pp. 156–159.

- [130] T. Q. Duong and H.-J. Zepernick, “On the ergodic capacity of cooperative spatial multiplexing systems in composite channels,” in *Proc. IEEE Radio and Wireless Symposium (RWS)*, Jan. 2009, pp. 175–178.
- [131] B. Maham and P. Popovski, “Capacity analysis of coordinated multipoint reception for mmwave uplink with blockages,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 16 299–16 303, Dec. 2020.
- [132] Y. Tian, G. Pan, and M.-S. Alouini, “On NOMA-based mmwave communications,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 15 398–15 411, Dec. 2020.
- [133] T. Van Chien, A. K. Papazafeiropoulos, L. T. Tu, R. Chopra, S. Chatzinotas, and B. Ottersten, “Outage probability analysis of IRS-assisted systems under spatially correlated channels,” *IEEE Wireless Communications Letters*, vol. 10, no. 8, pp. 1815–1819, Aug. 2021.
- [134] C. B. Dunham, “Chebyshev approximation by logarithmic families,” *Zeitschrift Angewandte Mathematik und Mechanik*, vol. 53, no. 5, pp. 352–353, Jan. 1973.
- [135] A. Alqahtani, E. Alsusa, A. Al-Dweik, and M. Al-Jarrah, “Performance analysis for downlink NOMA over  $\alpha - \mu$  generalized fading channels,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 7, pp. 6814–6825, Jul. 2021.
- [136] E. Björnson and L. Sanguinetti, “Rayleigh fading modeling and channel hardening for reconfigurable intelligent surfaces,” *IEEE Wireless Communications Letters*, vol. 10, no. 4, pp. 830–834, Apr. 2021.
- [137] S. Primak, *Stochastic Methods and Their Applications to Communications: Stochastic Differential Equations Approach*, eng. Wiley, 2004.
- [138] A. Nuttall, “Some integrals involving the  $Q$ -function,” *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 95–96, Jan. 1975.
- [139] Y. Zhang, J. Zhang, M. D. Renzo, H. Xiao, and B. Ai, “Performance analysis of RIS-aided systems with practical phase shift and amplitude response,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 5, pp. 4501–4511, May 2021.

- [140] Q.-U.-A. Nadeem, H. Alwazani, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, “Intelligent reflecting surface-assisted multi-user MISO communication: Channel estimation and beamforming design,” *IEEE Open Journal of the Communications Society*, vol. 1, pp. 661–680, May 2020.



## PUBLICATIONS



# PUBLICATION

1

**Global minimax approximations and bounds for the Gaussian  
 $Q$ -function by sums of exponentials**

I. M. Tanash and T. Riihonen



*IEEE Transactions on Communications*, vol. 68, no. 10, pp. 6514–6524

DOI: 10.1109/TCOMM.2020.3006902

**Publication reprinted with the permission of the copyright holders.**



# Global Minimax Approximations and Bounds for the Gaussian $Q$ -Function by Sums of Exponentials

Islam M. Tanash  and Taneli Riihonen , *Member, IEEE*

**Abstract**—This paper presents a novel systematic methodology to obtain new simple and tight approximations, lower bounds, and upper bounds for the Gaussian  $Q$ -function, and functions thereof, in the form of a weighted sum of exponential functions. They are based on minimizing the maximum absolute or relative error, resulting in globally uniform error functions with equalized extrema. In particular, we construct sets of equations that describe the behaviour of the targeted error functions and solve them numerically in order to find the optimized sets of coefficients for the sum of exponentials. This also allows for establishing a trade-off between absolute and relative error by controlling weights assigned to the error functions' extrema. We further extend the proposed procedure to derive approximations and bounds for any polynomial of the  $Q$ -function, which in turn allows approximating and bounding many functions of the  $Q$ -function that meet the Taylor series conditions, and consider the integer powers of the  $Q$ -function as a special case. In the numerical results, other known approximations of the same and different forms as well as those obtained directly from quadrature rules are compared with the proposed approximations and bounds to demonstrate that they achieve increasingly better accuracy in terms of the global error, thus requiring significantly lower number of sum terms to achieve the same level of accuracy than any reference approach of the same form.

**Index Terms**—Gaussian  $Q$ -function, error probability, minimax approximation, bounds, quadrature amplitude modulation (QAM), statistical performance analysis.

## I. INTRODUCTION

THE Gaussian  $Q$ -function and the related error function  $\text{erf}(\cdot)$  are ubiquitous in and fundamental to communication theory, not to mention all other fields of statistical sciences where the Gaussian/normal distribution is often encountered. In particular, the  $Q$ -function measures the tail probability of a standard normal random variable  $X$  having unit variance and zero mean, i.e.,  $Q(x) = \text{Prob}(X \geq x)$ , by which

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{1}{2}t^2\right) dt \quad (1a)$$

$$= \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \exp\left(-\frac{1}{2\sin^2\theta}x^2\right) d\theta \quad [\text{for } x \geq 0]. \quad (1b)$$

The latter integral is the so-called Craig's formula [1], [2], obtained by manipulating the original results of [3], [4].

Manuscript received June 20, 2019; revised November 8, 2019, March 6, 2020, and May 26, 2020; accepted June 22, 2020. This work was supported by the Academy of Finland under the grant 310991/326448. The associate editor coordinating the review of this paper and approving it for publication was F. J. López-Martínez. (Corresponding author: Islam M. Tanash.)

The authors are with Unit of Electrical Engineering, Faculty of Information Technology and Communication Sciences, Tampere University, FI-33720 Tampere, Finland (e-mail: islam.tanash@tuni.fi; taneli.riihonen@tuni.fi).

Digital Object Identifier X

The Gaussian  $Q$ -function has many applications in statistical performance analysis such as evaluating bit, symbol, and block error probabilities for various digital modulation schemes and different fading models [5]–[11], and evaluating the performance of energy detectors for cognitive radio applications [12], [13], whenever noise and interference or a channel can be modelled as a Gaussian random variable. However, in many cases formulating such probabilities will result in complicated integrals of the  $Q$ -function that cannot be expressed in a closed form in terms of elementary functions. Therefore, finding tractable approximations and bounds for the  $Q$ -function becomes a necessity in order to facilitate expression manipulations and enable its application over a wider range of analytical studies. Toward this demand, several approximations and bounds are already available in the literature.

## A. Approximations and Bounds for the $Q$ -Function

A brief overview on the existing approximations and bounds for the Gaussian  $Q$ -function is presented herein with the focus on those with the exponential form. The approximations and bounds presented in [14]–[26] have relatively complex mathematical forms and achieve high accuracy. Although some of them may lead to closed-form expressions, which would be otherwise impossible to solve, e.g., the polynomial approximation in [21] succeeds in analytically evaluating the average symbol error rate of pulse amplitude modulation in log-normal channels, the mathematical complexity of the aforementioned approximations make them still not quite convenient for algebraic manipulations in statistical performance analysis despite being accurate. For example, the approximation proposed by Börjesson and Sundberg in [15] is very complicated and best suitable for programming purposes. Therefore, the simplest known family with the form of a sum of exponentials was proposed by Chiani *et al.* [27], to provide bounds and approximations based on the Craig's formula.

The expression for approximating or bounding  $Q(x)$  by  $\tilde{Q}(x)$  that is generally suitable for applications, where one needs to express average error probabilities for fading distributions with adequate accuracy, is written as [27, Eq. (8)]

$$\tilde{Q}(x) \triangleq \sum_{n=1}^N a_n \exp\left(-b_n x^2\right) \quad [\text{for } x \geq 0 \text{ only}]. \quad (2)$$

Chiani *et al.* use the monotonically increasing property of the integrand in (1b) and apply the rectangular integration rule to derive exponential upper bounds. Moreover, when using the trapezoidal rule with optimizing the center point to minimize

the integral of relative error in an argument range of interest, an approximation with two exponential terms,  $N = 2$ , is obtained.

Other exponential approximations and bounds are also available [28]–[32]. A coarse single-term exponential approximation is presented in [28] based on the Chernoff bound, and a sum of two or three exponentials is proposed in [29], which is known as the Prony approximation. Another approximation of the exponential form that shows good trade-off between computational efficiency and mathematical accuracy is proposed in [30]. In [31], the composite trapezoidal rule with optimally chosen number of sub-intervals is used. The authors in [32] introduce a single-term exponential lower bound by using a tangent line to upper-bound the logarithmic function at some point which defines the tightness of the bound.

All of the aforementioned references propose approximations and bounds for the Gaussian  $Q$ -function and they can be also used as building blocks to approximate the powers or polynomials thereof. However, none of them directly derived approximations or bounds to evaluate the powers or polynomials of the  $Q$ -function, which arise frequently when analyzing various communication systems, e.g., error probability in quadrature amplitude modulation (QAM).

### B. Applications of the Approximations and Bounds

The above approximations and bounds have been implemented in the different areas of communication theory. We provide herein few examples from the literature. The approximations from [19] and [24] are used respectively to derive the frame error rate for a two-way decode-and-forward relay link in [33], and to analytically evaluate the average of integer powers of the  $Q$ -function over  $\eta$ - $\mu$  and  $\kappa$ - $\mu$  fading in [34]. As for the exponential form, it is used in [27] to compute error probabilities for space-time codes and phase-shift keying. Furthermore, (2) is used to derive the average bit-error rate for free-space optical systems in [35] and the symbol error rate of phase-shift keying under Rician fading in [36].

In general, the elegance of the exponential approximation in (2) can be illustrated by

$$\int F(Q(f(\gamma))) Y(\gamma) d\gamma \approx \sum_n a_n \int \exp(-b_n [f(\gamma)]^2) Y(\gamma) d\gamma,$$

where  $Y(\gamma)$  is some integrable function and  $F(Q(f(\gamma)))$  is some well-behaved function of the  $Q$ -function that accepts a Taylor series expansion for  $0 \leq Q(f(\gamma)) \leq \frac{1}{2}$ . Above, the polynomial of  $Q(f(\gamma))$  from the Taylor series of  $F(Q(f(\gamma)))$  is approximated by (2), either directly or indirectly (by first approximating  $Q(f(\gamma))$  by  $\tilde{Q}(f(\gamma))$  and then expanding the polynomial of the sum), which results in the latter sum.

Evaluating the integral in the above summation is usually much easier than evaluating the integral in the original expression at the left-hand side. This idea is applied in [37], when evaluating the average block error rate for Gamma-Gamma turbulence models under coherent binary phase-shift keying. Taylor series can also be used to approximate  $Y(\gamma)$  or parts of it [9], [37], eventually leading to closed-form expressions. Finally, it is worth mentioning that increasing the number of exponential terms in the summation (2) will typically

not increase the analytical complexity since summation and integration can be reordered in the expression under certain conditions and, hence, the integral is solved only once.

### C. Contributions and Organization of the Paper

The objective of this paper is to develop new accurate approximations and bounds for the Gaussian  $Q$ -function and functions thereof. To that end, we adopt the exponential sum expression originally proposed in [27] and restated in (2) and focus on the research problem of finding new, improved coefficients for it.<sup>1</sup> The coefficients developed herein will work as one-to-one replacements to those available in existing literature [27]–[32], but they offer significantly better accuracy and flexibility as well as generalization to various cases that have not been addressed before.

The major contributions of this paper are detailed as follows:

- We propose an original systematic methodology to optimize the set of coefficients  $\{(a_n, b_n)\}_{n=1}^N$  of (2) to obtain increasingly accurate but tractable approximations for the  $Q$ -function with any  $N$  in terms of the absolute or relative error, based on the minimax approximation theory, by which the global error is minimized when the corresponding error function is uniform.
- We further repurpose the methodology to find new exponential lower and upper bounds with very high accuracy that is comparable to, or even better than, the accuracy of other bounds of more complicated forms.
- We generalize our approximations and bounds to apply to polynomials and integer powers of the  $Q$ -function, or even implicitly to any generic function of the  $Q$ -function that accepts a Taylor series expansion.
- We show that the proposed minimax procedure reflects high flexibility in allowing for lower absolute or relative error at the expense of the other, or in allowing for higher accuracy in a specified range at the expense of less accuracy in the remaining ranges and a worse global error, by controlling weights assigned to the resulting non-uniform error function's extrema.

These contributions are verified by means of an extensive set of numerical results and an application example illustrating their accuracy and significance in communication theory.

The remainder of this paper is organized as follows. In Section II, we present the mathematical preliminaries needed for the formulation of the research problem and proposed solutions. Section III introduces our new approximations and bounds for the  $Q$ -function. Section IV presents our new approximations and bounds for the polynomials of the  $Q$ -function. The increasing accuracy of the novel solutions is demonstrated as well as comparisons with the best numerical alternatives and other known approximations having the same exponential form are presented in Section V. Concluding remarks are given in Section VI.

<sup>1</sup>Throughout the paper, when referring to 'our approximation/bound', we mean the existing sum expression (2) from [27] with our new coefficients.

## II. PRELIMINARIES

The case  $x \geq 0$  is presumed throughout this article. The results can be usually extended to the negative real axis using the relation  $Q(x) = 1 - Q(-x)$ . Likewise, the following discussions focus solely on the Gaussian  $Q$ -function but the results directly apply also to the related error function  $\operatorname{erf}(\cdot)$  and the complementary error function  $\operatorname{erfc}(\cdot)$  through the identity  $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x) = 2Q(\sqrt{2}x)$ , as well as to the cumulative distribution function  $\Phi(\cdot)$  of a normal random variable with mean  $\mu$  and standard deviation  $\sigma$  through the identity  $\Phi(x) = 1 - Q\left(\frac{x-\mu}{\sigma}\right)$ , which can be extended to  $x < \mu$  using the relation  $\Phi(x) = 1 - \Phi(2\mu - x) = Q\left(\frac{\mu-x}{\sigma}\right)$ .

The approximations and bounds will be optimized shortly in terms of the absolute or relative error using the minimax approach, in which the possible error in the worst-case scenario (i.e., the maximum error over all  $x$ ) is minimized. The baseline absolute and relative error functions<sup>2</sup> are defined as

$$d(x) \triangleq \tilde{Q}(x) - Q(x), \quad (3)$$

$$r(x) \triangleq \frac{d(x)}{Q(x)} = \frac{\tilde{Q}(x)}{Q(x)} - 1, \quad (4)$$

respectively, and the shorthand  $e \in \{d, r\}$  represents both of them collectively in what follows. In particular, the tightness of some approximation or bound  $\tilde{Q}(x)$  over the range  $[x_0, x_\infty]$  is measured as

$$e_{\max} \triangleq \max_{x_0 \leq x \leq x_\infty} |e(x)|, \quad (5)$$

and the approximations and bounds for minimax error optimization are solved as

$$\{(a_n^*, b_n^*)\}_{n=1}^N \triangleq \arg \min_{\{(a_n, b_n)\}_{n=1}^N} e_{\max}, \quad (6)$$

where  $e(x) \geq 0$  for upper bounds and  $e(x) \leq 0$  for lower bounds when  $x \geq 0$ .

Our optimization method depends on the extrema of the error function (cf. Fig. 1), which occur at points  $x_k$  where  $e'(x_k) = 0$ , for which the derivatives are given by

$$d'(x) = \tilde{Q}'(x) - Q'(x), \quad (7)$$

$$r'(x) = \frac{\tilde{Q}'(x)Q(x) - \tilde{Q}(x)Q'(x)}{[Q(x)]^2}. \quad (8)$$

The derivatives of the approximation/bound in (2) and of the  $Q$ -function in (1) are

$$\tilde{Q}'(x) = -2 \cdot \sum_{n=1}^N a_n b_n x \exp(-b_n x^2), \quad (9)$$

$$Q'(x) = -\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right), \quad (10)$$

respectively. Let us also note that the absolute error converges to zero when  $x$  tends to infinity, i.e.,  $\lim_{x \rightarrow \infty} d(x) = 0$ , whereas for the relative error, we have

$$\lim_{x \rightarrow \infty} r(x) = \begin{cases} \infty, & \text{when } \min\{b_n\}_{n=1}^N = \frac{1}{2}, \\ -1, & \text{otherwise.} \end{cases} \quad (11)$$

<sup>2</sup>These should not be confused with the error function  $\operatorname{erf}(\cdot)$ .

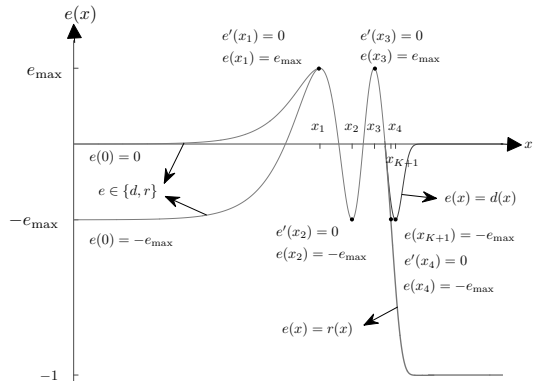


Fig. 1. The optimized minimax error function starts either from  $e(0) = 0$  or from  $e(0) = -e_{\max}$  and oscillates between local maximum and minimum values of equal magnitude; when considering relative error, this is possible only in a finite range of  $x$  as opposed to global bounds obtained w.r.t. absolute error. The minimax criterion implies uniform error function with  $w_k = 1$ .

This renders some specific restrictions for all upper bounds and optimization w.r.t. the relative error as is shortly observed.

For reference, the Craig's formula in (1b) can also be approximated using various numerical integration techniques [38]. This results in low-accuracy approximations or bounds of the same form as (2) with numerical coefficients that can be directly calculated from the weights and nodes of the corresponding numerical method.

## III. MINIMAX APPROXIMATIONS AND BOUNDS FOR THE GAUSSIAN $Q$ -FUNCTION

We adopt the weighted sum of exponential functions in (2) to express global minimax approximations and bounds for the Gaussian  $Q$ -function. In particular, according to Kammler in [39, Theorem 1], the best approximation in which the maximum value of the corresponding error function is minimized to reach its minimax error, occurs when the error function is uniformly oscillating between maximum and minimum values of equal magnitude, as illustrated in Fig. 1.

The original idea in our work is that one can describe the minimax error function by a set of equations, where the number of equations is equal to the number of unknowns. These equations describe the error function at the extrema points in which all of them have the same value of error and the derivative of the error function at these points is equal to zero. Our ultimate goal is then to find the optimized set of coefficients,  $\{(a_n^*, b_n^*)\}_{n=1}^N$ , that solves the formulated set of equations. In general, for problem formulation of  $e \in \{d, r\}$ ,

$$\begin{cases} e'(x_k) = 0, & \text{for } k = 1, 2, 3, \dots, K, \\ e(x_k) = (-1)^{k+1} w_k e_{\max}, & \text{for } k = 1, 2, 3, \dots, K, \end{cases} \quad (12)$$

where  $w_k$  is a potential weight for error at  $x_k$  (set  $w_k = 1$  as default for uniform approximations/bounds) and  $K$  is the number of the error function's extrema excluding the endpoints.

TABLE I

NUMBER OF ERROR EXTREMA EXCLUDING ENDPOINTS NEEDED TO FORMULATE THE PROBLEM IN TERMS OF ABSOLUTE OR RELATIVE ERROR.

Error measure, $e$	Type	Number of extrema
Absolute error, $d$	Upper bound	$K = 2N - 1$
	Approximation	$K = 2N$
	Lower bound	$K = 2N$
Relative error, $r$	Upper bound	$K = 2N - 2$
	Approximation	$K = 2N - 1$
	Lower bound	$K = 2N - 1$

Table I summarizes the values of  $K$  in terms of the number of sum terms  $N$  for the different cases considered next.

In this study, we aim to minimize the global error over the whole positive  $x$ -axis, which is possible in terms of the absolute error. However, the relative error does not converge to zero when  $x$  tends to infinity as seen in (11). Thus, we must choose a finite interval on the  $x$ -axis, in which its right boundary,  $x_\infty$ , is equal to  $x_{K+1}$  as will be discussed later. On the other hand, the left boundary of the  $x$ -range,  $x_0$ , is equal to zero for both error measures. In addition to  $w_k$ ,  $k = 1, 2, \dots, K$ , the weight set also includes  $w_0$  which occurs at  $x_0$ , and  $w_{K+1}$  which occurs at  $x_{K+1}$  for the relative error.

Although the minimum global absolute or relative error is obtained when the error function is uniform, the weight set that can be controlled is added throughout this article when formulating the approximation or bound problem to facilitate a compromise between  $d_{\max}$  and  $r_{\max}$  when tailoring it specifically for some application. The weight set can be even controlled to obtain better accuracy in some specified range of the argument. It should be mentioned that, in these cases, at least one of the weights has to be equal to one, representing the maximum error, and the remaining should be smaller and positive. When all of the weights are equal to one, the approximations and bounds are called uniform and they achieve the global minimax error as discussed earlier.

Two variations of equations can be formulated depending on whether the error starts from  $e(0) = 0$  or  $e(0) = -w_0 e_{\max}$  as seen in Fig. 1. The importance of the former case comes from the fact that such approximation or upper bound gives the exact same value as the  $Q$ -function at  $x = 0$ , resulting in a continuous function when extending it to the negative values of  $x$ . The latter case gives slightly better accuracy at the expense of the discontinuity that occurs at  $x = 0$ .

#### A. Problem Formulation in Terms of Absolute Error

Here we describe the formulation of the approximations and bounds of the  $Q$ -function when minimizing the global absolute error according to (5) and (6). The corresponding set of coefficients,  $\{(a_n, b_n)\}_{n=1}^N$ , in (2) are optimized as follows:

$$\{(a_n^*, b_n^*)\}_{n=1}^N \triangleq \arg \min_{\{(a_n, b_n)\}_{n=1}^N} \max_{x \geq 0} |\tilde{Q}(x) - Q(x)|. \quad (13)$$

1) *Approximations*: The approximation's maximum absolute error is globally minimized when all local error extrema are equal to the global error extrema. The extrema occur where the derivative of the absolute error function is zero. For the produced error, all positive and negative extrema have the same

value of error, i.e.,  $d_{\max}$ . Moreover, we optimize (3) at  $x_0 = 0$  for two variations:  $d(0) = 0$  or  $d(0) = -w_0 d_{\max}$ , where  $Q(0) = \frac{1}{2}$  and  $\tilde{Q}(0) = \sum_{n=1}^N a_n$ .

Therefore, we can formulate the approximation problem as

$$\begin{cases} d'(x_k) = 0, & \text{for } k = 1, 2, 3, \dots, K, \\ d(x_k) = (-1)^{k+1} w_k d_{\max}, & \text{for } k = 1, 2, 3, \dots, K, \\ \begin{cases} \sum_{n=1}^N a_n = \frac{1}{2}, & \text{when } d(0) = 0, \\ \sum_{n=1}^N a_n = \frac{1}{2} - w_0 d_{\max}, & \text{when } d(0) = -w_0 d_{\max}. \end{cases} \end{cases} \quad (14)$$

Although only the set  $\{(a_n^*, b_n^*)\}_{n=1}^N$  is needed to construct the minimax absolute error function indicated by  $e \in \{d, r\}$  together with  $e(x) = d(x)$  in Fig. 1, other unknowns will also appear when solving the optimization problem in (13), which are  $\{x_k\}_{k=1}^K$  and  $d_{\max}$  for the uniform approximations and bounds.

The number of equations throughout this paper is always equal to the number of unknowns. For the minimax approximation in terms of absolute error, a set of  $4N + 1$  equations is constructed to solve  $4N + 1$  unknowns using  $2N$  extrema points according to Table I. Each extremum yields two equations; one expresses its value, and the other expresses the derivative of the error function at that point. An additional equation originates from evaluating the error function at  $x_0$ . This corresponds to either  $e(0) = 0$  or  $e(0) = -e_{\max}$  as indicated in Fig. 1. For any  $N$ , a solution to the system of equations yields  $\{(a_n^*, b_n^*)\}_{n=1}^N$  that defines the minimax approximation, and we prove by construction that it exists.

2) *Bounds*: For the bounds, we use the same approach as for the approximations with ensuring that  $d(x) \leq 0$  and  $d(x) \geq 0$  for the lower and upper bounds, respectively, when  $x \geq 0$ . The former results in  $4N + 1$  equations, with the optimized absolute error function starting from  $d(0) = -w_0 d_{\max}$ , the maxima equal to zero and the minima equal to  $-w_k d_{\max}$ . On the other hand, the latter results in  $4N$  equations with the corresponding error function starting from  $d(0) = 0$ , the maxima equal to  $w_k d_{\max}$  and the minima equal to zero, with forcing the lowest value in the set  $\{b_n\}_{n=1}^N$  to be  $\frac{1}{2}$ , so that both error measures are always positive. Otherwise  $r(x)$  will converge to a negative value as shown in (11),  $d(x)$  would be negative for large  $x$  too, and we could not find an upper bound of the  $Q$ -function. Moreover, the derivative of the corresponding error function is equal to zero at all the  $K$  extrema points for both types of bounds.

#### B. Problem Formulation in Terms of Relative Error

Here we describe the formulation of the exponential approximations and bounds of the  $Q$ -function when minimizing the global relative error defined by (4). We optimize the corresponding set of coefficients,  $\{(a_n, b_n)\}_{n=1}^N$ , as follows:

$$\{(a_n^*, b_n^*)\}_{n=1}^N \triangleq \arg \min_{\{(a_n, b_n)\}_{n=1}^N} \max_{0 \leq x \leq x_{K+1}} \left| \frac{\tilde{Q}(x)}{Q(x)} - 1 \right|. \quad (15)$$

Unlike the absolute error, the relative error does not converge to zero when  $x$  tends to infinity as shown in (11). This is why we must limit the minimax approximation in terms of



the relative error to the finite range by choosing  $x_\infty = x_{K+1}$ , as opposed to  $x_\infty \rightarrow \infty$  in the case of absolute error. This yields

$$\begin{cases} r(x_{K+1}) = w_{K+1} r_{\max}, & \text{for upper bounds,} \\ r(x_{K+1}) = -w_{K+1} r_{\max}, & \text{otherwise.} \end{cases} \quad (16)$$

Hence, the relative error function is minimized globally over  $[0, x_{K+1}]$ . This can be seen by the case where  $e(x) = r(x)$  in Fig. 1, in which the point  $x_{K+1}$  is chosen so that its corresponding error value is equal to  $-r_{\max}$ .

1) *Approximations*: In regard to the relative error, the same approach as for the absolute error is implemented herein in order to construct the minimax approximations with the corresponding uniform error function illustrated by  $e \in \{d, r\}$  together with  $e(x) = r(x)$  in Fig. 1. A set of  $4N$  equations originates from the  $2N - 1$  extrema and the two endpoints, which are  $x_0$  and  $x_{K+1}$ . It is noted that,  $r'(x_{K+1}) \neq 0$  and only one equation can be acquired from this point, since the minimax approximation herein is limited to the range  $0 \leq x \leq x_{K+1}$ . Therefore, the optimized coefficients for the two variations are found by solving the following set of equations:

$$\begin{cases} r'(x_k) = 0, & \text{for } k = 1, 2, 3, \dots, K, \\ r(x_k) = (-1)^{k+1} w_k r_{\max}, & \text{for } k = 1, 2, 3, \dots, K, \\ \begin{cases} \sum_{n=1}^N a_n = \frac{1}{2}, \\ \sum_{n=1}^N a_n = \frac{1}{2} - \frac{1}{2} w_0 r_{\max}, \end{cases} & \text{when } r(0) = 0, \\ r(x_{K+1}) = -w_{K+1} r_{\max}. & \text{when } r(0) = -w_0 r_{\max}, \end{cases} \quad (17)$$

2) *Bounds*: We optimize the lower and upper bounds for  $0 \leq x \leq x_{K+1}$  in terms of the relative error using the same problem formulation as for the absolute bounds but with  $4N$  equations in case of lower bounds, and  $4N - 1$  equations in case of upper bounds, and by substituting  $d$  by  $r$ , in addition to enforcing (16) that describes the error function at  $x_{k+1}$ .

### C. Proof by Construction: Solutions for $N = 1, 2, 3, \dots, 25$

We prove the existence of the proposed solutions to (13) and (15) by construction, i.e., numerically solving (14) and (16), (17). In particular, we implemented the set of equations of each of the considered variations in Matlab and used the `fsolve` command with equal number of equations and unknowns to find the optimized set of coefficients  $\{(a_n^*, b_n^*)\}_{n=1}^N$ , where the main challenge was to choose heuristic initial guesses. For the initial guesses of lower values of  $N$ , we used iteratively random values for  $e_{\max}$ ,  $\{(a_n, b_n)\}_{n=1}^N$  and  $\{x_k\}_{k=1}^K$  with  $K$  as given in Table I, along the process of finding their optimal values that solve the proposed research problem. After reaching certain  $N$  which is enough to form a relation between the previous values, we constructed a pattern to predict their successive values for higher values of  $N$ .

The sets of optimized coefficients are solved herein up to  $N = 25$  for the novel minimax approximations and bounds as well as released to public domain in a supplementary digital file with  $x_{K+1}$  ranging from 1 to 10 in steps of 0.1 for the relative error. Nevertheless, let us illustrate the sets of

optimized coefficients of the absolute error for  $d(0) = -d_{\max}$  and  $N = 2, 3, 4$  in Table II, in addition to the set of optimized coefficients of the relative error in the case where  $r(0) = 0$ ,  $x_{K+1} = 6$  and  $N = 20$ , for quick reference.

Our optimized coefficients yield very accurate approximations that outperform all the existing ones in terms of the global error. For example, for  $N = 2$ , our approximation yields  $d_{\max} = 9.546 \cdot 10^{-3}$  and the reference approximations [27], [29] and [30], yield  $d_{\max} = 1.667 \cdot 10^{-1}, 1.450 \cdot 10^{-1}$  and  $1.297 \cdot 10^{-1}$ , respectively. The accuracy can be increased even further by increasing  $N$ . For example, the tabulated coefficients of the relative error for  $N = 20$  render a tight uniform approximation in terms of the relative error while satisfying  $\bar{Q}(0) = Q(0) = \frac{1}{2}$ . Namely,  $|r(x)| \leq r_{\max}^* < 2.831 \cdot 10^{-6}$  when  $x \leq 6$  and  $|r(x_k)| = r_{\max}^*$  at all the  $K = 39$  local maximum error points. This approximation is also tight in terms of the absolute error since  $|d(x)| \leq d_{\max} < 1.416 \cdot 10^{-6}$  for all  $x \geq 0$  and the largest local error maxima are observed when  $x \ll 1$  while  $|d(x)| \ll d_{\max}$  for  $x > 1$ .

TABLE II  
THE SET OF OPTIMIZED COEFFICIENTS OF THE ABSOLUTE ERROR FOR  $d(0) = -d_{\max}$  AND  $N = 2, 3, 4$ , AND THE SET OF OPTIMIZED COEFFICIENTS OF THE RELATIVE ERROR FOR  $r(0) = 0$ ,  $x_{K+1} = 6$  AND  $N = 20$ .

$N$	$n$	$a_n^*$	$b_n^*$
2	1	3.736889599671366e-1	8.179084584179674e-1
	2	1.167651897698837e-1	1.645047046852372e+1
3	1	3.259195350781647e-1	7.051797307608448e-1
	2	1.302528627687561e-1	5.489376068647640e+0
	3	4.047435009465072e-2	1.335391071637174e+2
4	1	2.936683276537767e-1	6.517755981618476e-1
	2	1.357580421878250e-1	3.250040490513459e+0
	3	5.245255757691102e-2	3.186882707224491e+1
	4	1.673209873360605e-2	7.786613983601425e+2
20	1	7.558818716991463e-2	5.071654316592885e-1
	2	7.283303478836754e-2	5.678040654656637e-1
	3	6.886155063785772e-2	7.104625738749141e-1
	4	6.439172935348138e-2	9.994060383297402e-1
	5	5.779242444673264e-2	1.601184575755943e+0
	6	4.808415837769939e-2	2.928772702717808e+0
	7	3.692309273438261e-2	6.019071014437780e+0
	8	2.656563850645104e-2	1.358210951915055e+1
	9	1.820530043799255e-2	3.304520236491907e+1
	10	1.201348364882034e-2	8.584892772825742e+1
	11	7.675500579336059e-3	2.375751011169581e+2
	12	4.755522827095319e-3	7.025476884457923e+2
	13	2.853832378872099e-3	2.237620299200472e+3
	14	1.652925274323080e-3	7.776239381556935e+3
	15	9.183202474880042e-4	3.007617539336614e+4
	16	4.846308477760495e-4	1.334789827558299e+5
	17	2.391717111298367e-4	7.146006517383908e+5
	18	1.074573496224467e-4	4.056149657406912e+6
	19	4.174113678130675e-5	5.790627530626244e+7
	20	1.229754587599716e-5	2.138950747557404e+9

## IV. APPROXIMATIONS AND BOUNDS FOR POLYNOMIALS OF THE $Q$ -FUNCTION

In this section, we generalize the novel minimax optimization method presented in Section III, to derive approximations and bounds for any polynomial of the  $Q$ -function and any integer power of the  $Q$ -function as a special case. In fact, this method can be applied to expressing approximations and

bounds for many well-behaved functions of the  $Q$ -function using Taylor series expansion, in which it is represented as an infinite sum of terms. Therefore, Taylor series is a polynomial of infinite degree [40] that one needs to truncate to get a Taylor polynomial approximation of degree  $P$ .

In general, any  $P$ th degree polynomial of the  $Q$ -function is expressed as

$$\Omega(Q(x)) \triangleq \sum_{p=0}^P c_p Q^p(x), \quad (18)$$

where  $\{c_p\}_{p=0}^P$  are constants and called the polynomial coefficients. In particular, the novel optimization methodology is extended to such polynomials by directly approximating/bounding  $\Omega(Q(x))$  by  $\tilde{Q}_\Omega(x)$  that has the same exponential form as  $Q(x)$  in (2). We optimize the coefficient set,  $\{(a_n, b_n)\}_{n=1}^N$ , in order to minimize the maximum absolute or relative error of the polynomial, which results in a uniform error function as described before.

The absolute and relative error functions for any polynomial of the  $Q$ -function are defined respectively as

$$d_\Omega(x) \triangleq \tilde{Q}_\Omega(x) - \sum_{p=0}^P c_p Q^p(x), \quad (19)$$

$$r_\Omega(x) \triangleq \frac{d_\Omega(x)}{\sum_{p=0}^P c_p Q^p(x)} = \frac{\tilde{Q}_\Omega(x)}{\sum_{p=0}^P c_p Q^p(x)} - 1. \quad (20)$$

The derivatives of the error functions are

$$d'_\Omega(x) = \tilde{Q}'_\Omega(x) - \sum_{p=1}^P p c_p Q^{p-1} Q'(x), \quad (21)$$

$$r'_\Omega(x) = \frac{\tilde{Q}'_\Omega(x) \sum_{p=0}^P c_p Q^p(x) - \tilde{Q}_\Omega(x) \sum_{p=1}^P p c_p Q^{p-1} Q'(x)}{\left[ \sum_{p=0}^P c_p Q^p(x) \right]^2}, \quad (22)$$

where  $\tilde{Q}'_\Omega(x)$  has the same expression as  $\tilde{Q}'(x)$  in (9) and  $Q'(x)$  is given by (10).

Following the procedure explained in Section III, and using the mentioned definitions, approximations/bounds for polynomials of the  $Q$ -function are formulated in terms both error measures. More specifically, what applies to error functions with the  $Q$ -function described by (13)–(17) also applies herein, with replacing  $\sum_{n=1}^N a_n = \frac{1}{2}$  by  $\sum_{n=1}^N a_n = \sum_{p=0}^P (\frac{1}{2})^p c_p$  for the absolute and relative errors of the approximations that start from  $e(0) = 0$  and for the upper bounds. Furthermore, one should replace  $\sum_{n=1}^N a_n = \frac{1}{2} - w_0 d_{\max}$  by  $\sum_{n=1}^N a_n = \sum_{p=0}^P (\frac{1}{2})^p c_p - w_0 d_{\max}$  for the absolute error and  $\sum_{n=1}^N a_n = \frac{1}{2} - \frac{1}{2} w_0 r_{\max}$  by  $\sum_{n=1}^N a_n = \sum_{p=0}^P (\frac{1}{2})^p c_p - \sum_{p=0}^P (\frac{1}{2})^p c_p w_0 r_{\max}$  for the relative error of the approximations that start from  $e(0) = -w_0 e_{\max}$  and for lower bounds.

#### A. Special Case: Integer Powers of the $Q$ -Function

In general, any polynomial of the  $Q$ -function as per (18) is a linear combination of non-negative integer powers of the  $Q$ -function. The integer powers themselves are important special cases in communication theory, where they appear frequently on their own. To that end, one may derive the

optimized approximations and bounds for them by simply setting the coefficient  $c_p$  of the required power  $p$  in (18)–(22) to one and the remaining to zero while following exactly the same optimization procedure as explained above for the general case of polynomials. It should also be mentioned that, for the upper bounds,  $\min\{b_n\}_{n=1}^N = \frac{p}{2}$ . We refer to the approximations and bounds of this special case by  $\tilde{Q}_p(\cdot)$  to differentiate it from the general case of polynomials.

In the coefficient data that we release to public domain along with this paper, the sets of optimized coefficients  $\{(a_n^*, b_n^*)\}_{n=1}^N$  for the approximations/bounds of the exponential form shown in (2) are numerically solved with  $p = 1, 2, 3, 4$  and  $N = 1, 2, \dots, 25$  for the novel minimax approximations and bounds with  $x_{K+1}$  ranging from 1 to 10 in steps of 0.1 for the relative error. However, the provided approximations and bounds can be extended to any value of  $p$ .

If not approximating directly, the approximations/bounds for any polynomial of the  $Q$ -function with  $N$  terms can be obtained by using the integer powers' approximations/bounds (including the first power) as follows:

$$\begin{aligned} \Omega(Q(x)) &\approx \sum_{p=0}^P c_p \tilde{Q}_{p, N_p}(x) \\ &= \sum_{p=0}^P c_p \prod_{l=1}^L \tilde{Q}_{p_l, N_{p_l}}(x) \\ &= \sum_{p=0}^P c_p \sum_{n_{p_1}=1}^{N_{p_1}} \sum_{n_{p_2}=1}^{N_{p_2}} \dots \sum_{n_{p_L}=1}^{N_{p_L}} \prod_{l=1}^L a_{n_{p_l}}[l] \\ &\quad \times \exp\left(-\left(\sum_{l=1}^L b_{n_{p_l}}[l]\right) x^2\right), \end{aligned} \quad (23)$$

where  $\sum_{l=1}^L p_l = p$ ,  $\prod_{l=1}^L N_{p_l} = N_p$ ,  $\sum_{p=0}^P N_p = N$ , and  $a_{n_p}[l]$ ,  $b_{n_p}[l]$  are the coefficients of  $\tilde{Q}_{p_l, N_{p_l}}(x)$ . The ultimate number of terms in (23) may be less than  $N$  if some of them can be combined. The above implies also that the approximations/bounds of any integer power of the  $Q$ -function with  $N_p$  terms can be obtained using the product rule.

#### B. Application Example: Evaluation of the Average SEP in Optimal Detection of 4-QAM in Nakagami- $m$ Fading

Let us emphasize on the elegance of (2) for approximating or bounding the  $Q$ -function, its integer powers or any polynomial thereof by giving an application example of average error probabilities over fading channels. In general, they are obtained for coherent detection in most cases by evaluating

$$\bar{P}_E = \int_0^\infty \Omega(Q(\alpha\sqrt{\gamma})) \psi_\gamma(\gamma) d\gamma, \quad (24)$$

where  $\Omega(Q(\alpha\sqrt{\gamma}))$  is some polynomial of the  $Q$ -function as per (18) and refers to the error probability conditioned on the instantaneous signal-to-noise ratio (SNR), i.e.,  $\gamma$ , with  $\psi_\gamma(\gamma)$  being its probability density function, and  $\alpha$  is a constant that

TABLE III  
THE SET OF OPTIMIZED COEFFICIENTS OF THE ABSOLUTE ERROR FOR  
 $d_{\Omega}(0) = -d_{\max}$  AND  $N = 5$ .

$n$	$a_n^*$	$b_n^*$
1	4.920547396876422e-1	5.982476003750250e-1
2	1.587491012166297e-1	2.024383866054074e+0
3	6.460001610510117e-2	1.323465438792062e+1
4	2.567521272080907e-2	1.314581690889673e+2
5	8.236936034796302e-3	3.211202445024321e+3

depends on the digital modulation and detection techniques. Substituting our approximation into the above equation yields

$$\bar{P}_E \approx \sum_{n=1}^N a_n \int_0^{\infty} \exp(-b_n \alpha^2 \gamma) \psi_{\gamma}(\gamma) d\gamma \quad (25a)$$

$$= \sum_{n=1}^N a_n \Theta_{\gamma}(-b_n \alpha^2), \quad (25b)$$

where  $\Theta_{\gamma}(s) = \int_0^{\infty} \exp(s\gamma) \psi_{\gamma}(\gamma) d\gamma$  is the moment generating function associated with the random variable  $\gamma$ .

Let us next evaluate the average symbol error probability (SEP) in optimal detection of 4-QAM over Nakagami- $m$  fading channels, under which it is often hard to derive closed-form expressions for error probabilities if  $m$  is not an integer. Thus, we first solve exponential approximations and bounds for the conditional SEP in 4-QAM that is a second-order polynomial of the  $Q$ -function as follows [5, Eq. 8.20]:

$$P_E(\gamma) = 2Q(\sqrt{\gamma}) - Q^2(\sqrt{\gamma}). \quad (26)$$

By comparing to (18),  $c_0 = 0$ ,  $c_1 = 2$ , and  $c_2 = -1$ . This SEP is approximated by  $\tilde{Q}_{\Omega}(x)$  as described above. Finally, we substitute the gamma probability distribution in (25a) and evaluate the integral using [41, Eq. 3.351.3] as

$$\begin{aligned} \bar{P}_E &= \frac{m^m}{\bar{\gamma}^m \Gamma(m)} \sum_{n=1}^N a_n \int_0^{\infty} \gamma^{m-1} \exp\left(-\gamma\left(b_n + \frac{m}{\bar{\gamma}}\right)\right) d\gamma \\ &= \frac{m^m}{\bar{\gamma}^m} \sum_{n=1}^N a_n \left(b_n + \frac{m}{\bar{\gamma}}\right)^{-m}, \end{aligned} \quad (27)$$

where  $m$  defines the fading parameter, ranging from 0.5 to  $\infty$ ,  $\bar{\gamma}$  is the average SNR, and  $\Gamma(\cdot)$  denotes the gamma function.

The sets of optimized coefficients  $\{(a_n^*, b_n^*)\}_{n=1}^N$  for the approximations and bounds of the conditional SEP in 4-QAM were solved for  $N = 1, 2, \dots, 25$  for the minimax approach in terms of both error measures. Table III shows an example of the coefficients optimized in terms of the absolute error in the case where  $d_{\Omega}(0) = -d_{\max}$  and  $N = 5$ . These render a tight uniform approximation with  $|d_{\Omega}(x)| \leq d_{\max}^* < 6.84 \cdot 10^{-4}$ .

The computational and/or analytical complexity using our approximations and bounds for the integer powers and the polynomials of the  $Q$ -function is much less than using any other approximation from the literature, in which none of them has proposed approximations or lower/upper bounds for the powers or the polynomials of the  $Q$ -function. Therefore, directly substituting the SEP polynomial by our exponential approximations is more tractable than evaluating it by applying reference approximations to (26).

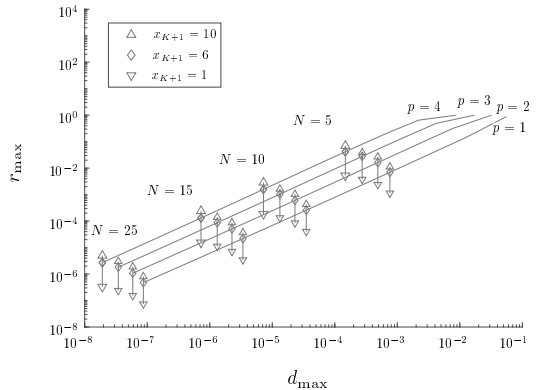


Fig. 2. Optimal absolute error versus optimal relative error for the first four powers of the  $Q$ -function for the approximations starting from  $e(0) = 0$ . The two-sided vertical arrows indicate  $r_{\max}$  for  $x_{K+1}$  ranging from 1 to 10.

## V. NUMERICAL RESULTS AND DISCUSSION

Let us next compare the proposed approximations and bounds with the existing ones having the same exponential form, in addition to the best approximations among the different numerical integration techniques. The optimized sets of coefficients,  $\{(a_n^*, b_n^*)\}_{n=1}^{25}$  for the cases considered in this paper, all in terms of both absolute and relative error, are constructed in this paper to form round 37 000 coefficient sets in total. Due to Matlab's fixed (64-bit) floating-point precision, some other programming software with adjustable precision is required to pursue the proposed minimax approach for finding approximations and bounds for values of  $N$  much beyond 25. This is because some  $a_n$  become very small when the corresponding  $b_n$  become very large resulting in underflow when computing  $a_n \exp(-b_n x^2)$  numerically for (2).

To begin, we plot the minimax absolute error versus minimax relative error for  $p = 1, 2, 3, 4$ , and  $N = 1, 2, 3, \dots, 25$  of the approximation starting from  $e(0) = 0$ , in Fig. 2, with showing  $x_{K+1}$  ranging from 1 to 10 for  $N = 5, 10, 15, 25$  in terms of relative error. The other types of approximations and the lower/upper bounds follow similar behaviour as the one shown in Fig. 2. It is clear from the figure that, as the number of exponential terms increases, the minimax absolute and relative error decrease significantly.

For reference, we have investigated the different numerical integration techniques and their  $h$ -point composites (up to  $h = 4$ ) that can be implemented to approximate the Gaussian  $Q$ -function as a weighted sum of exponentials in terms of both absolute and relative errors. However, we only include the Legendre rule and its four-point composite formula in Fig. 3, where they achieve the least global error among all the other numerical methods and their composites, respectively, along with the two types of the proposed minimax approximations. In addition, the global error values of the existing approximations of the same form are also calculated and plotted in the same figure for specific number of terms, namely,  $N = 1, 2, 3, 4$ , where  $\tilde{Q}(\cdot)$  is expressed using one

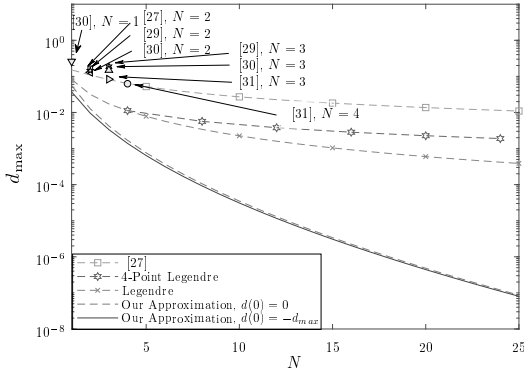


Fig. 3. Comparison of the absolute error between our approximations and those obtained using [27], [29], [30] and [31], as well as those calculated using Legendre rule and its 4-point composite version.

exponential in [30], two exponentials in [27], [29] and [30], three exponentials in [29], [30] and [31], and four exponentials in [31]. The composite right-rectangular rule, which was used to approximate  $Q(\cdot)$  in [27], is also plotted for comparison. In Fig. 3, we only include the absolute error since the relative error illustrates similar results, and only the maximum error over  $x \geq 0$  is compared.

It is evident from the figure that our approximations outperform all of the existing approximations as well as those obtained from numerical integration in terms of the global error, and as the number of terms increases, even better accuracy is obtained. In contrast, we can see that the numerical methods are converging slowly, causing the number of terms required by the numerical integration to be much higher than that required by our approximations in order to achieve the same level of error.

Table IV compares the values of  $N$  between the proposed approximations and the best integration rules that achieve certain absolute error levels. Clearly, our approximations are much more tractable than any other numerical approximation in terms of the global error, where only a few exponential terms are needed to achieve high accuracy. For the non-composite Legendre rule, when applied to approximate the  $Q$ -function, the error will start to oscillate for  $N > 41$  and eventually converge to infinity. This implies that Legendre approximations are not reliable and cannot achieve high level of accuracy. After illustrating the efficiency of our proposed approximations in terms of the global error, we further verify the accuracy for the whole considered range of the positive argument by comparing the relative error function obtained when applying our approximations and the existing ones for  $N = 2$  and  $N = 4$  as shown in Fig. 4. In addition to the fact that our approximations have the least global error, their accuracy surpass all the reference approximations over the range  $[0, 0.4]$  and attain comparable accuracy for  $x > 0.4$ .

For the ranges, where other approximations have better accuracy, the error function can be reshaped in such a way that the accuracy over the specified range is improved at the

TABLE IV  
COMPARISON BETWEEN  $N$  VALUES FOR THE PROPOSED APPROXIMATIONS AND BOTH COMPOSITE AND NON-COMPOSITE LEGENDRE INTEGRATION RULES THAT ACHIEVE CERTAIN ABSOLUTE ERROR LEVEL.

Absolute error	$N$ for approx. with $d(0) = 0$	$N$ for composite Legendre rule	$N$ for non-composite Legendre rule
$1 \cdot 10^{-2}$	2	4	4
$1 \cdot 10^{-3}$	4	44	15
$1 \cdot 10^{-4}$	8	452	41
$1 \cdot 10^{-5}$	12	3504	—

cost of less accuracy in the other ranges and, hence, increased global error. We do that by controlling the weights of the error function's extrema of our approximations when setting the problem conditions.

As an example, let us consider the problem conditions in (17) that formulates the relative error shown in Fig. 4. We can increase the accuracy of the approximation which has three extrema for  $N = 2$  and starts from  $r(0) = -w_0 r_{\max}$  over the range  $[-2, 14]$ , by controlling the weights of the extrema to be  $w_0 = 1, w_1 = w_2 = w_3 = w_4 = 1/10$ ,  $w_4$  is the weight at the right boundary of the interval of optimization. This example is illustrated in the figure by the solid-diamond line. We can see that the error has decreased to be more accurate in the specified range and outperforms the other reference approximations over most of the range. However, the global error has increased substantially. This demonstrates how our approximations' and bounds' accuracy can be tailored for specific ranges of values, depending on their application.

The accuracy of our upper and lower bounds was investigated in terms of both error measures but only the relative error is shown in Fig. 5 to save space. It is obvious that our bounds not only have the least global error but they also outperform the other exponential bounds presented in [27], [28]. Moreover, over a wide range of the argument, our bounds have even better accuracy than the other bounds of more complicated forms. For instance, our lower bound is the best over the whole positive range  $x > 0$ . On the other hand, our upper bound has better accuracy than that of [22] and comparable accuracy to [26], although [23] is more accurate over the range  $[0, 3.5]$ , where it has a more complex form.

As mentioned earlier, we can achieve better absolute or relative error at the expense of the other by controlling the weights of the extrema. We test the trade-off behaviour herein by starting from the uniform relative error with equal weights and gradually decreasing the weights' values,  $w_k$ , for  $k = 1, 2, \dots, K$  while maintaining  $w_{K+1} = 1$ . The maximum obtained relative and absolute error values are measured and plotted in Fig. 6 for  $N = 1, 2, \dots, 10$ . The cross marker in the figure refers to the minimax error obtained when formulating the minimax approximation in terms of absolute error, for any  $N$ . In the same way, the plus marker refers to the minimax error obtained when formulating the minimax approximation in terms of relative error, for any  $N$ . We can see from Fig. 6 that as the absolute error decreases, the relative error increases, forming smooth transition and a trade-off between the two error measures. Other transition lines can be formed between the extremes based on how the weight set is controlled.

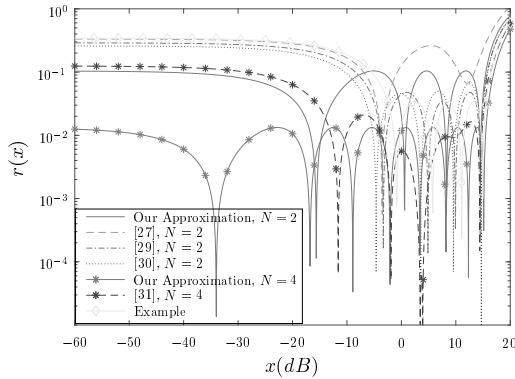


Fig. 4. Comparison among our approximations and the references approximations, [27], [29], [30] and [31] for  $N = 2$  and  $N = 4$  in terms of the relative error.

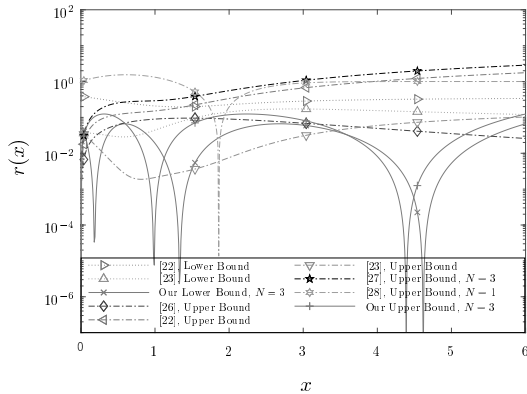


Fig. 5. Comparison between our bounds and the references bounds [22], [23], [26]–[28] in terms of relative error.

For the relative error, the effect of changing the value of  $x_{K+1}$  is illustrated in Fig. 7. As the value of  $x_{K+1}$  increases,  $r_{max}$  increases too for the approximations and bounds, achieving worse accuracy. Furthermore, like noted before, higher values of  $N$  result in highly improved accuracy as can be seen in the figure, in which the relative error for  $N = 25$  is several orders of magnitude lower than for  $N = 5$ .

In Fig. 8, we compare the absolute error of the proposed approximations and bounds for the third power of the  $Q$ -function, with the error calculated using (23) for all  $N$ . The minimum error among all errors obtained using all the possible combinations of  $N_{p_l}$ ,  $l = 1, \dots, L$  is considered in this comparison for each combination set of the  $Q$ -function whose powers add to three. It is noted that representing the integer powers of the  $Q$ -function as weighted sum of exponentials using (2), is more accurate and simpler than representing it using the different combinations.

Finally, approximating SEP in (26) directly using (2) in the coherent detection of 4-QAM is compared in Fig. 9(a) with

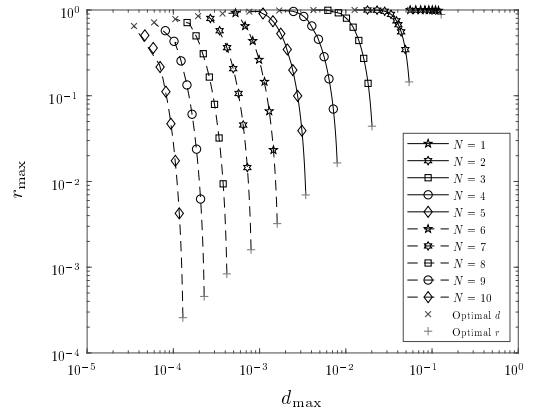


Fig. 6. The trade-off between the absolute and relative error. To obtain better absolute error than obtained when optimizing the relative error, the weight set when formulating the optimization problem can be controlled to achieve less absolute error but with increased relative error, and vice versa.

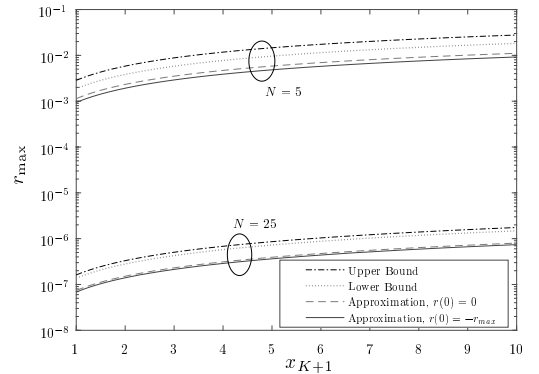


Fig. 7. The effect of changing  $x_{K+1}$  on the relative error of the proposed approximations and bounds, for  $N = 5$  and  $N = 25$ .

those obtained using the different combinations when applying (23). The direct solutions give increasingly higher absolute and relative accuracy as expected. Figure 9(b) together with Table V compares the accuracy of the corresponding average SEP in 4-QAM when evaluated for different values of the fading parameter  $m$  using our exponential approximation,  $\tilde{Q}_\Omega(\cdot)$ , with the optimized coefficients that are listed in Table III, and the other reference exponential approximations. The results demonstrate excellent agreement over the entire range of average SNR between the exact average SEP and our approximation that is very tight even for lower values of SNR, in contrast to the references that are accurate only at higher SNRs. Furthermore, the tightness of our approximation is preserved when changing the value of  $m$ , while the approximation from [27] is accurate only for small values of  $m$ . It should be noted that, when we substitute the reference approximations with two terms in (26), we get a five-

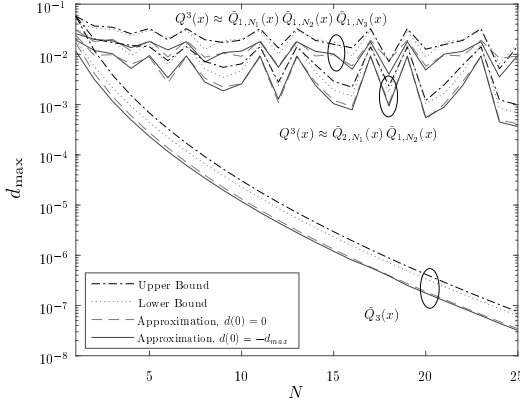


Fig. 8. Comparison of the absolute error of the proposed exponential approximations and bounds for  $p = 3$ , with the minimum error among all errors obtained using all the possible combinations as given in (23).

TABLE V  
COMPARISON OF ACCURACY OF AVERAGE SEP FOR 4-QAM OVER  
NAKAGAMI- $m$  FADING.

For $m = 0.8$				
Exact	0.530436	0.379629	0.216681	0.101863
[27], $N = 5$	0.478317	0.368463	0.220893	0.106127
[29], $N = 5$	0.487660	0.363137	0.211014	0.099769
[30], $N = 5$	0.499954	0.369262	0.213148	0.100468
Our approx., $N = 5$	0.530440	0.379629	0.216629	0.101753
Our approx., $N = 10$	0.530436	0.379629	0.216680	0.101859
For $m = 1.9$				
Exact	0.509397	0.333819	0.142200	0.034658
[27], $N = 5$	0.474948	0.346634	0.160587	0.040565
[29], $N = 5$	0.482288	0.333805	0.144435	0.035039
[30], $N = 5$	0.493865	0.337280	0.143837	0.034678
Our approx., $N = 5$	0.509432	0.333780	0.142188	0.034474
Our approx., $N = 10$	0.509398	0.333819	0.142200	0.034652
$\bar{\gamma}$ (in dB)	-5	0	5	10

term exponential approximation for the SEP. As the number of exponential terms increases, our approximation becomes virtually exact, outperforming all the existing approximations as seen in Table V with already  $N = 10$ .

## VI. CONCLUSIONS

This paper proposed accurate and tractable approximations, lower bounds and upper bounds for the Gaussian  $Q$ -function and any polynomial of the  $Q$ -function as a weighted sum of exponential functions. The novel sets of coefficients of the sum terms are optimally solved in minimax sense to minimize the global absolute or relative error of approximations/bounds, where in the limit of a larger number of terms, they approach very close to their corresponding exact functions. Moreover, we show that the weights set to the extrema of the error function can be controlled to compromise between the absolute and the relative error. The significantly (i.e., by several orders of magnitude) improved accuracy of the proposed expressions with optimized coefficients has been demonstrated by compar-

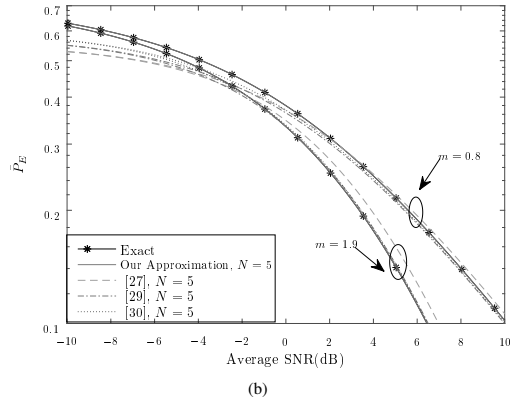
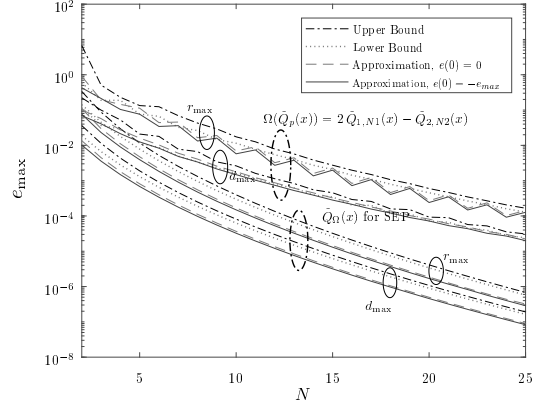


Fig. 9. (a) Comparison of both absolute and relative error of the proposed exponential approximations and bounds for SEP in 4-QAM, with those obtained by applying (23). (b) Average SEP plots for 4-QAM over Nakagami- $m$  using our approximation and the reference exponential approximations for  $N = 5$ .

ing the results with approximations from numerical integration and other existing approaches.

## REFERENCES

- [1] J. Craig, "A new, simple and exact result for calculating the probability of error for two-dimensional signal constellations," in *Proc. IEEE Mil. Commun. Conf.*, vol. 2, Nov. 1991, pp. 571–575.
- [2] F. J. López-Martínez, R. Pawula, E. Martos-Naya, and J. Paris, "A clarification of the proper-integral form for the Gaussian  $Q$ -function and some new results involving the  $F$ -function," *IEEE Commun. Lett.*, vol. 18, no. 9, pp. 1495–1498, Sep. 2014.
- [3] F. Weinstein, "Simplified relationships for the probability distribution of the phase of a sine wave in narrow-band normal noise," *IEEE Trans. Inf. Theory*, vol. 20, no. 5, pp. 658–661, Sep. 1974.
- [4] R. Pawula, S. Rice, and J. Roberts, "Distribution of the phase angle between two vectors perturbed by Gaussian noise," *IEEE Trans. Commun.*, vol. 30, no. 8, pp. 1828–1841, Aug. 1982.
- [5] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels*, 2nd ed. John Wiley and Sons, Inc., Jan. 2005.
- [6] M. Irshid and I. Salous, "Bit error probability for coherent  $M$ -ary PSK systems," *IEEE Trans. Commun.*, vol. 39, no. 3, pp. 349–352, Mar. 1991.

- [7] P. Lee, "Computation of the bit error rate of coherent  $M$ -ary PSK with Gray code bit mapping," *IEEE Trans. Commun.*, vol. 34, no. 5, pp. 488–491, May. 1986.
- [8] H. Suraweera and J. Armstrong, "Performance of OFDM-based dual-hop amplify-and-forward relaying," *IEEE Commun. Lett.*, vol. 11, no. 9, pp. 726–728, Sep. 2007.
- [9] B. Zhu, "Asymptotic performance of composite lognormal- $x$  fading channels," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6570–6585, Dec. 2018.
- [10] X. Tang, M.-S. Alouini, and A. Goldsmith, "Effect of channel estimation error on  $M$ -QAM BER performance in Rayleigh fading," *IEEE Trans. Commun.*, vol. 47, no. 12, pp. 1856–1864, Dec. 1999.
- [11] J. Lu, K. Letief, J. Chuang, and M. Liou, " $M$ -PSK and  $M$ -QAM BER computation using signal-space concepts," *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 181–184, Feb. 1999.
- [12] A. Taherpour, M. Nasiri-Kenari, and S. Gazor, "Multiple antenna spectrum sensing in cognitive radios," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 814–823, Feb. 2010.
- [13] A. Mariani, A. Giorgetti, and M. Chiani, "Effects of noise power estimation on energy detection for cognitive radio applications," *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3410–3420, Dec. 2011.
- [14] W. Cody, "Rational Chebyshev approximations for the error function," *Math. Comp.*, vol. 23, no. 107, pp. 631–637, Jul. 1969.
- [15] P. Borjesson and C. Sundberg, "Simple approximations of the error function  $Q(x)$  for communications applications," *IEEE Trans. Commun.*, vol. 27, no. 3, pp. 639–643, Mar. 1979.
- [16] J. Nikolić, Z. Perić, and A. Marković, "Proposal of simple and accurate two-parametric approximation for the  $Q$ -function," *Int. J. Electron.*, vol. 100, no. 4, pp. 1–10, Dec. 2017.
- [17] G. Karagiannidis and A. Lioumpas, "An improved approximation for the Gaussian  $Q$ -function," *IEEE Commun. Lett.*, vol. 11, no. 8, pp. 644–646, Aug. 2007.
- [18] J. Dyer and S. Dyer, "Corrections to, and comments on, An improved approximation for the Gaussian  $Q$ -function," *IEEE Commun. Lett.*, vol. 12, no. 4, p. 231, Apr. 2008.
- [19] Y. Isukapalli and B. Rao, "An analytically tractable approximation for the Gaussian  $Q$ -function," *IEEE Commun. Lett.*, vol. 12, no. 9, pp. 669–671, Sep. 2008.
- [20] C. Tellambura and A. Annamalai, "Efficient computation of  $\operatorname{erfc}(x)$  for large arguments," *IEEE Trans. Commun.*, vol. 48, no. 4, pp. 529–532, Apr. 2000.
- [21] Y. Chen and N. Beaulieu, "A simple polynomial approximation to the Gaussian  $Q$ -function and its application," *IEEE Commun. Lett.*, vol. 13, no. 2, pp. 124–126, Feb. 2009.
- [22] G. Abreu, "Jensen–Cotes upper and lower bounds on the Gaussian  $Q$ -function and related functions," *IEEE Trans. Commun.*, vol. 57, no. 11, pp. 3328–3338, Nov. 2009.
- [23] G. Abreu, "Very simple tight bounds on the  $Q$ -function," *IEEE Trans. Commun.*, vol. 60, no. 9, pp. 2415–2420, Sep. 2012.
- [24] M. López-Benítez and F. Casadevall, "Versatile, accurate, and analytically tractable approximation for the Gaussian  $Q$ -function," *IEEE Trans. Commun.*, vol. 59, no. 4, pp. 917–922, Apr. 2011.
- [25] Q. Shi, "Novel approximation for the Gaussian  $Q$ -function and related applications," in *Proc. 22nd IEEE PIMRC*, Sep. 2011, pp. 2030–2034.
- [26] W. Jang, "A simple upper bound of the Gaussian  $Q$ -function with closed-form error bound," *IEEE Commun. Lett.*, vol. 15, no. 2, pp. 157–159, Feb. 2011.
- [27] M. Chiani, D. Dardari, and M. Simon, "New exponential bounds and approximations for the computation of error probability in fading channels," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 840–845, Jul. 2003.
- [28] S. Chang, P. Cosman, and L. Milstein, "Chernoff-type bounds for the Gaussian error function," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 2939–2944, Nov. 2011.
- [29] P. Loskot and N. Beaulieu, "Prony and polynomial approximations for evaluation of the average probability of error over slow-fading channels," *IEEE Trans. Veh. Technol.*, vol. 58, no. 3, pp. 1269–1280, Mar. 2009.
- [30] O. Olabiyi and A. Annamalai, "Invertible exponential-type approximations for the Gaussian probability integral  $Q(x)$  with applications," *IEEE Wireless Commun. Lett.*, vol. 1, no. 5, pp. 544–547, Oct. 2012.
- [31] D. Sadhwani, R. Yadav, and S. Aggarwal, "Tighter bounds on the Gaussian  $Q$  function and its application in Nakagami- $m$  fading channel," *IEEE Wireless Commun. Lett.*, vol. 6, no. 5, pp. 574–577, Oct. 2017.
- [32] M. Wu, Y. Li, M. Gurusamy, and P. Kam, "A tight lower bound on the Gaussian  $Q$ -function with a simple inversion algorithm, and an application to coherent optical communications," *IEEE Commun. Lett.*, vol. 22, no. 7, pp. 1358–1361, Jul. 2018.
- [33] Q. Zhou, Y. Li, F. Lau, and B. Vucetic, "Decode-and-forward two-way relaying with network coding and opportunistic relay selection," *IEEE Trans. Commun.*, vol. 58, no. 11, pp. 3070–3076, Nov. 2010.
- [34] M. López-Benítez, "Average of arbitrary powers of Gaussian  $Q$ -function over  $\eta$ - $\mu$  and  $\kappa$ - $\mu$  fading channels," *Electron. Lett.*, vol. 51, no. 11, pp. 869–871, May 2015.
- [35] T. Tsiftsis, H. Sandalidis, G. Karagiannidis, and M. Uysal, "Optical wireless links with spatial diversity over strong atmospheric turbulence channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 951–957, Feb. 2009.
- [36] M. McKay, A. Zanella, I. Collings, and M. Chiani, "Error probability and SINR analysis of optimum combining in Rician fading," *IEEE Trans. Commun.*, vol. 57, no. 3, pp. 676–687, Mar. 2009.
- [37] Q. Zhang, J. Cheng, and G. Karagiannidis, "Block error rate of optical wireless communication systems over atmospheric turbulence channels," *IET Commun.*, vol. 8, no. 5, pp. 616–625, March 2014.
- [38] P. Davis and P. Rabinowitz, *Methods of Numerical Integration*, 2nd ed. Academic Press, 1984.
- [39] D. Kammler, "Chebyshev approximation of completely monotonic functions by sums of exponentials," *SIAM Journal on Numerical Analysis*, vol. 13, no. 5, pp. 761–774, 1976.
- [40] D. Widder, "A generalization of Taylor's series," *Trans. Am. Math. Soc.*, vol. 30, no. 1, pp. 126–154, Jan. 1928.
- [41] I. Gradshteyn and I. Ryzhik, *Table of integrals, series, and products*, 7th ed. Elsevier/Academic Press, 2007.



**Islam M. Tanash** received the B.Sc. and M.Sc. degrees in electrical engineering from Jordan University of Science and Technology (JUST), Irbid, Jordan in 2014 and 2016, respectively. She is currently a PhD student and doctoral researcher at the Faculty of Information Technology and Communication Sciences, Tampere University, Finland. Her research interests include the areas of communications theory, wireless networks, and wireless systems security.



**Taneli Riihonen** (S'06–M'14) received the D.Sc. degree in electrical engineering (with distinction) from Aalto University, Helsinki, Finland, in August 2014. He is currently an Assistant Professor (tenure track) at the Faculty of Information Technology and Communication Sciences, Tampere University, Finland. He held various research positions at Aalto University School of Electrical Engineering from September 2005 through December 2017. He was a Visiting Associate Research Scientist and an Adjunct Assistant Professor at Columbia University in the City of New York, USA, from November 2014 through December 2015. He has been nominated eleven times as an Exemplary/Top Reviewer of various IEEE journals and is serving as an Editor for IEEE WIRELESS COMMUNICATIONS LETTERS since May 2017. He has previously served as an Editor for IEEE COMMUNICATIONS LETTERS from October 2014 through January 2019. He received the Finnish technical sector's award for the best doctoral dissertation of the year and the EURASIP Best PhD Thesis Award 2017. His research activity is focused on physical-layer OFDM(A), multi-antenna, relaying and full-duplex wireless techniques with current interest in the evolution of beyond 5G systems.





# PUBLICATION

2

**Remez exchange algorithm for approximating powers of the  
 $Q$ -function by exponential sums**

I. M. Tanash and T. Riihonen

In *Proc. IEEE Vehicular Technology Conference (VTC)*, Apr. 2021, pp. 1–6

DOI: 10.1109/VTC2021-Spring51267.2021.9448807

**Publication reprinted with the permission of the copyright holders.**



# Remez Exchange Algorithm for Approximating Powers of the $Q$ -Function by Exponential Sums

Islam M. Tanash and Taneli Riihonen

Faculty of Information Technology and Communication Sciences, Tampere University, Finland

e-mail: {islam.tanash, taneli.riihonen}@tuni.fi

**Abstract**—In this paper, we present simple and tight approximations for the integer powers of the Gaussian  $Q$ -function, in the form of exponential sums. They are based on optimizing the corresponding coefficients in the minimax sense using the Remez exchange algorithm. In particular, the best exponential approximation is characterized by the alternation of its absolute error function, which results in extrema that alternate in sign and have the same magnitude of error. The extrema are described by a system of nonlinear equations that are solved using Newton–Raphson method in every iteration of the Remez algorithm, which eventually leads to a uniform error function. This approximation can be employed in the evaluation of average symbol error probability (ASEP) under additive white Gaussian noise and various fading models. Especially, we present several application examples on evaluating ASEP in closed forms with Nakagami- $m$ , Fisher–Snedecor  $\mathcal{F}$ ,  $\eta - \mu$ , and  $\kappa - \mu$  channels. The numerical results show that our approximations outperform the existing ones with the same form in terms of the global error. In addition, they achieve high accuracy for the whole range of the argument with and without fading, and it can even be improved further by increasing the number of exponential terms.

## I. INTRODUCTION

The Gaussian  $Q$ -function and the directly related error function  $\text{erf}(\cdot)$  are of fundamental importance to communication theory—and many other statistical sciences—whenever noise and interference or a channel can be modelled as a Gaussian random variable. This importance is reflected by the different applications in statistical performance analysis including the evaluation of error probabilities for various digital modulation schemes and different fading models [1]. The  $Q$ -function does not have an exact closed expression and it usually exists as a built-in numerical function in most of the software programs. Nevertheless, many of the  $Q$ -function applications encounter complicated integrals of it that cannot be simplified to closed-form expressions in terms of elementary functions.

Therefore, several approximations and bounds are available in [2]–[13]. The authors in [2] and [3] have proposed relatively complicated, but highly accurate, approximations and bounds that are impractical for actual evaluation of systems’ performance and more suitable for improving the calculation efficiency. More accurate approximations for the  $Q$ -function are provided in [4], [5]. The approximation of the first power in [4] is later simplified in [6] using Taylor series expansion. An accurate polynomial approximation for  $Q(x)$  is derived in

[7]. A single-term exponential approximation with polynomial argument of the second degree is presented in [8]. The simplest form of exponential approximations and bounds were first proposed by Chiani *et al.* in [9], and other ones are also developed using different approaches in [10]–[13].

The aforementioned approximations and bounds find applications in various communication problems. For example, the approximations in [7] are applied to analytically calculate the average symbol error rate of pulse amplitude modulation in log-normal channels. In [8], the authors derive the probability of detection for an energy detector over a Rayleigh fading channel. Moreover, the exponential approximations in [9] are implemented to compute error probabilities for space–time codes and phase-shift keying.

The aim of this work is to develop new accurate approximations for the Gaussian  $Q$ -function and its integer powers by adopting the simple exponential form originally proposed in [9] with acquiring novel, improved coefficients for it. The work in [9] is limited to two-term approximation without methodology for optimally extending it to higher number of terms and integer powers. In particular, we minimize the maximum absolute difference between the exponential sum (3) and  $Q^p(x)$  to obtain the best global minimax approximation for any number of terms like we did in [13], but now we avoid complicated nonlinear equations thereof, numerical solving of which is very sensitive to the right choice of initial guesses.

We solve the coefficients by the Remez exchange algorithm and propose a new heuristic method to find the initial guesses needed for it. The resulting approximations render significantly higher accuracy in terms of global error and adequate accuracy for the whole range of the argument when compared to the existing ones of [9]–[11] with the same form and number of terms. The accuracy can even be increased further by increasing the number of exponential terms. Finally, some application examples on evaluating average error probabilities over different generalized fading distributions are provided to validate the high accuracy of the new approximations in comparison to the reference approximations.

## II. PROBLEM FORMULATION

The Gaussian  $Q$ -function is defined classically as

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{1}{2}t^2\right) dt. \quad (1)$$

An alternative representation in the polar domain was developed by Craig [14] for communication theory applications as

This research work was funded in part by the Academy of Finland under the grant 326448 “Generalized Fading Distributions and Matrix Functions for the Analysis of Wireless Communication Systems.”

$$Q(x) = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \exp\left(-\frac{1}{2 \sin^2 \theta} x^2\right) d\theta, \quad (2)$$

that is valid for  $x \geq 0$  only. Indeed, throughout this article, we shall confine our discussions to the domain  $x \geq 0$  since the results can be trivially extended to the negative real axis using the relation  $Q(x) = 1 - Q(-x)$ .

The weighted sum of exponential functions adopted herein for approximating  $Q^p(x)$  is written as [9, Eq. (8)]

$$\tilde{Q}_p(x) \triangleq \sum_{n=1}^N a_n \exp(-b_n x^2), \quad (3)$$

that is likewise valid for  $x \geq 0$  only. In [9], Chiani *et al.* use the trapezoidal integration rule to find  $\{(a_n, b_n)\}_{n=1}^N$  for  $N = 2$  by optimizing the center point of (2) to minimize the integral of relative error in an argument range of interest. Moreover, other approximations for any  $N$  are also derived using the rectangular rule with non-optimized equispaced points.

Our research problem is to optimize the coefficients of the approximation in the sense of minimax absolute error as

$$\{(a_n^*, b_n^*)\}_{n=1}^N \triangleq \arg \min_{\{(a_n, b_n)\}_{n=1}^N} d_{\max}, \quad (4)$$

in which  $d_{\max}$  refers to the global tightness of the approximation  $\tilde{Q}_p(x)$  over the range  $[0, \infty)$  and is measured as

$$d_{\max} \triangleq \max_{x \geq 0} |d(x)|. \quad (5)$$

The above absolute error function is defined as

$$d(x) \triangleq \tilde{Q}_p(x) - Q^p(x), \quad (6)$$

and it converges to zero when  $x$  tends to infinity, i.e.,  $\lim_{x \rightarrow \infty} d(x) = 0$ . Thus, in plain words, our goal that is expressed in (4) is to solve the optimized set of coefficients  $\{(a_n^*, b_n^*)\}_{n=1}^N$  to minimize  $d_{\max}$  given in (5), substitute them in (3), and so obtain increasingly accurate approximations not only for the  $Q$ -function but also for its integer powers.

### III. SOLUTION BY REMEZ EXCHANGE ALGORITHM

We solve (4) by applying the famous exchange algorithm established by Evgeny Remez in 1934. The Remez algorithm is an iterative methodology that can be used to derive the best approximation in the minimax sense using different nonlinear approximating functions (that are typically Chebyshev polynomials) and is characterized by the uniform alternation of the corresponding error function [15] as seen in Fig. 1 after the third iteration. In this paper, we use the sum of exponentials defined in (3) as the approximating function to obtain the best unique approximation for the power of the  $Q$ -function, since it is a completely monotonic function [16], [17]. The corresponding error function should alternate exactly  $2N$  times on  $[0, \infty)$  between maximum and minimum values of equal magnitude, resulting in a total of  $2N + 1$  extrema points. The exponential approximation also results in  $2N + 1$  unknowns, namely the  $2N$  coefficients of (3) and the global error per (5).

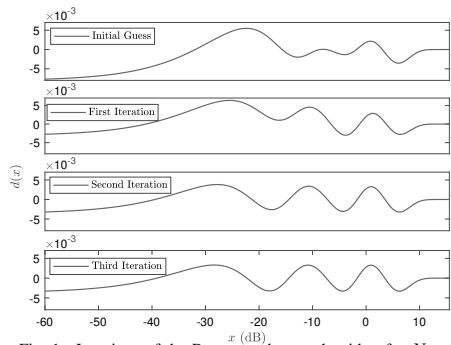


Fig. 1. Iterations of the Remez exchange algorithm for  $N = 3$ .

#### A. Algorithm Formulation

The steps for applying the Remez exchange algorithm to approximate the  $Q$ -function are summarized in Algorithm 1.

First, we construct a system of  $2N + 1$  simultaneous equations that describe the  $2N + 1$  extrema of the required uniform error function as

$$\mathbf{f}(\mathbf{r}) \triangleq \begin{bmatrix} f_0(\mathbf{r}) \\ f_1(\mathbf{r}) \\ \vdots \\ f_{2N}(\mathbf{r}) \end{bmatrix} \triangleq \begin{bmatrix} d(x_0) + d_{\max} \\ d(x_1) - d_{\max} \\ \vdots \\ d(x_k) + (-1)^k d_{\max} \\ \vdots \\ d(x_{2N}) + (-1)^{2N} d_{\max} \end{bmatrix} = \mathbf{0}, \quad (7)$$

where  $x_k$  is the abscissa value of the  $k$ th extremum of the error function and  $\mathbf{r} = [a_1, a_2, \dots, a_N, b_1, b_2, \dots, b_N, d_{\max}]^T$  is a vector of the unknowns. The first extremum occurs always at  $x_0 = 0$ , which results in  $d(0) = \sum_{n=1}^N a_n - (\frac{1}{2})^p$  since  $Q^p(0) = (\frac{1}{2})^p$  and  $\tilde{Q}_p(0) = \sum_{n=1}^N a_n$ . The adopted exponential approximation results in a nonlinear type of equations, opposing to the linear type which usually occur with the best polynomial approximations and often accompanied with the Remez algorithm whenever it is presented in the literature.

The Newton–Raphson method is a root finding technique that can be regarded as a somewhat ideal solver for this system of nonlinear equations since it is quadratically convergent when approaching the root. It is also an iterative method that requires initial guesses for the unknowns (roots) and we refer to its iterations as the inner iterations to differentiate them from the outer ones of the Remez algorithm. Furthermore, it is based on approximating a continuous and differentiable function by a straight line tangent to it, which results when applied on our system of equations (7) in

$$\mathbf{r}^{(v+1)} = \mathbf{r}^{(v)} - \left[ \mathbf{J}^{(v)}(\mathbf{r}^{(v)}) \right]^{-1} \mathbf{f}(\mathbf{r}^{(v)}), \quad (8)$$

where  $v$  is the inner-iteration counter, and  $\mathbf{J}(\cdot)$  is the Jacobian matrix that is calculated as

$$\mathbf{J}(\mathbf{r}) = \begin{bmatrix} \frac{\partial f_0(\mathbf{r})}{\partial r_0} & \frac{\partial f_0(\mathbf{r})}{\partial r_1} & \cdots & \frac{\partial f_0(\mathbf{r})}{\partial r_{2N}} \\ \frac{\partial f_1(\mathbf{r})}{\partial r_0} & \frac{\partial f_1(\mathbf{r})}{\partial r_1} & \cdots & \frac{\partial f_1(\mathbf{r})}{\partial r_{2N}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_{2N}(\mathbf{r})}{\partial r_0} & \frac{\partial f_{2N}(\mathbf{r})}{\partial r_1} & \cdots & \frac{\partial f_{2N}(\mathbf{r})}{\partial r_{2N}} \end{bmatrix}, \quad (9)$$

with  $[r_0, r_1, \dots, r_{2N}] = [a_1, a_2, \dots, a_N, b_1, b_2, \dots, b_N, d_{\max}]$ ,  $\frac{\partial f_k(\mathbf{r})}{\partial a_n} = \exp(-b_n x_k^2)$ ,  $\frac{\partial f_k(\mathbf{r})}{\partial b_n} = -a_n x_k^2 \exp(-b_n x_k^2)$ , and  $\frac{\partial f_k(\mathbf{r})}{\partial d_{\max}} = (-1)^k$ . This procedure is repeated until the differences between the values of  $\mathbf{r}$  of two successive iterations are smaller than a predefined threshold value. The Newton–Raphson method is implemented on (7) to find the vector of unknowns in every iteration of the Remez algorithm.

Assuming that we have a reasonably good initial guess for  $\{(a_n, b_n)\}_{n=1}^N$  that formulates the proposed approximation and enables the construction of the corresponding absolute error function, we can locate extrema points thereof and the value of global error and use them for initializing  $\{x_k\}_{k=1}^{2N}$  (but fixing  $x_0 = 0$ ) and  $d_{\max}$ , respectively. We start the iterative procedure by solving the nonlinear system of equations using the aforementioned Newton–Raphson method, together with the initialized vector of unknowns  $\mathbf{r}^{(0)}$ . The obtained error function that has the same error value at each of the initial extrema points with alternating signs does not (yet) necessarily give the minimax solution since these points may not be at the extrema of the error function. Therefore, we need to find the new set of  $\{x_k\}_{k=1}^{2N}$  by first locating the  $2N$  roots of  $d(x)$ , which we denote by  $\{z_i\}_{i=1}^{2N}$  using any root-finding numerical technique such as the bisection method or even the Newton–Raphson method yet again. Then we split the positive  $x$ -axis into  $2N + 1$  sub-intervals as  $[0, z_1], [z_1, z_2], \dots, [z_{2N}, \infty)$ .

For each sub-interval, we locate the point at which the error function attains its maximum magnitude by setting  $d'(x) = 0$ , for which the derivative is defined as

$$d'(x) = -2 \sum_{n=1}^N a_n b_n x \exp(-b_n x^2) + p \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) Q^{p-1}(x). \quad (10)$$

In particular, we numerically find  $x_k$  that meets  $d'(x_k) = 0$  after substituting the  $k$ th sub-interval in (10). If the root does not exist, we take the endpoint that gives the larger absolute value of the two.

Finally, we replace the previous extrema points by the new ones and continue repeating the above steps for a number of iterations until the difference between the previous extrema points and the new ones are below a predefined threshold  $\epsilon$ .

### B. Initial Guesses

Before we can start the Remez method, we must obtain good initial guesses for  $\{(a_n, b_n)\}_{n=1}^N$ . In this subsection, we describe one possible, heuristic method that works for the cases illustrated in this paper. In particular, we focus on finding

---

### Algorithm 1 Remez Exchange Algorithm

---

```

Initialize  $\{x_k\}_{k=1}^{2N}, \epsilon$ 
Set  $t \leftarrow 0, x_0 \leftarrow 0$ 
repeat
  Solve (7) for unknowns  $\{(a_n, b_n)\}_{n=1}^N, d_{\max}$  using
  Newton–Raphson method
  Find  $\{z_i\}_{i=1}^{2N}$ 
  Divide  $[0, \infty)$  into  $2N + 1$  sub-intervals by using  $\{z_i\}_{i=1}^{2N}$ 
  as boundaries
  for  $k \leftarrow 1$  to  $2N$  do
    Find the root of  $d'(x)$  in the  $k$ th sub-interval.
    if such root does not exist then
      Evaluate  $d(x)$  at endpoints and choose the point that
      gives the maximum
    end if
    Denote the obtained root or point by  $x_k^{t+1}$ 
  end for
  Set  $\{x_k^{t+1}\}_{k=1}^{2N}$  to  $\{x_k^t\}_{k=1}^{2N}$ 
   $t \leftarrow t + 1$ 
until  $|\{x_k^{t+1}\}_{k=1}^{2N} - \{x_k^t\}_{k=1}^{2N}| < \epsilon$ 
Best minimax approximation is obtained

```

---

initial guesses for the first power of the Gaussian  $Q$ -function, which we can use as basis for finding initial guesses for higher values of  $p$  as will be explained later in this subsection.

For  $p = 1$  and lower values of  $N$ , we assigned repeatedly different random values for  $\{(a_n, b_n)\}_{n=1}^N$  and calculated  $d(x)$  per (6) for each  $N$ . Once we were lucky enough to come across any  $d(x)$  that has the correct shape with  $2N + 1$  extrema (e.g., the initial guess in Fig. 1),  $\{x_k\}_{k=1}^{2N}$  and  $d_{\max}$  were calculated and used together with the corresponding  $\{(a_n, b_n)\}_{n=1}^N$  to solve the considered optimization problem (4) using Algorithm 1. This yields in a unique set of the optimized coefficients  $\{(a_n^*, b_n^*)\}_{n=1}^N$  which gives exactly the required uniform shape (e.g., the third iteration in Fig. 1).

After reaching certain  $N$ , we were able to use curve fitting techniques to formulate equations that can give good initial values for  $\{b_n\}_{n=1}^N$  and  $\{z_i\}_{i=1}^{2N}$  for  $N = 1, 2, 3, \dots, 10$ . Each  $b_n$ -coefficient of the proposed approximations with any  $N$  has been assigned an equation of the form  $b_n = A_n N^{B_n} + C_n$ , and  $A_n, B_n$  and  $C_n$  are given in Table I. Moreover, one equation is formulated to calculate all the initial guesses of  $z_i, i = 1, 2, 3, \dots, 2N$ , for any value of  $N$  as  $z_i = (0.4845 i^{-1.364} - 29.72) N^{(0.003752 i^{-1.122} + 0.4884)} + (105.9 i^{0.1924} - 94.83)$ . Next, the initial guesses for  $\{a_n\}_{n=1}^N$  are found by substituting the above calculated initial values in (6) to formulate a system of linear equations describing the absolute error function at its roots as  $d(z_i) = \sum_{n=1}^N c_n a_n - q_i = 0$ , where  $c_n = \exp(-b_n z_i^2)$  and  $q_i = Q(z_i)$  are constant. After solving the linear system of equations for the unknowns  $\{a_n\}_{n=1}^N$ , we can easily locate the initial guesses for  $d_{\max}$  and  $\{x_k\}_{k=0}^{2N}$  from  $d(x)$  that is numerically calculated using the initial guesses of  $\{(a_n, b_n)\}_{n=1}^N$ .

On the other hand, for higher values of  $p$ , we rely on

TABLE I

THE PARAMETERS OF THE POWER EQUATION THAT IS USED TO FIND AN INITIAL GUESS  $b_n = A_n N^{B_n} + C_n$  FOR  $n = 1, 2, \dots, N$  WITH  $N \leq 10$ .

$n$	$A_n$	$B_n$	$C_n$
1	6.514e-1	-1.075e+0	5.051e-1
2	2.389e+1	-1.658e+0	6.633e-1
3	6.908e+2	-2.481e+0	1.217e+0
4	6.699e+4	-3.983e+0	5.022e+0
5	3.002e+6	-4.959e+0	1.183e+1
6	2.793e+8	-6.244e+0	3.453e+1
7	1.063e+14	-1.129e+1	4.356e+2
8	7.474e+16	-1.315e+1	1.188e+3
9	3.721e+19	-1.478e+1	2.790e+3
10	1.048e+20	-1.384e+1	1.808e+4

the optimized coefficients  $\{b_n^*\}_{n=1}^N$  and the corresponding  $\{x_k\}_{k=0}^{2N}$  of the first power. We have found that they can be used to construct initial guesses for the higher powers through the relations  $x_{k,p} = x_k - 2p$ ,  $b_{n,p} = (2.25 + 1.65(p-2))b_n$ , and  $d_{max,p} = d_{max}$  where we use the subscripts  $p$  only herein in this equation to differentiate the coefficients of  $p > 1$  from those of the first power. The initial guesses for  $\{a_n\}_{n=1}^N$  can be easily found using the linear system of equations that solves  $d(x_k) = \sum_{n=1}^N c_n a_n - q_k = (-1)^{k+1} d_{max}$ , where  $c_n = \exp(-b_n x_k^2)$  and  $q_k = Q(x_k)$  are constant. It is worth mentioning that using these relations will directly give all the required initial guesses for  $p = 2, 3, 4$ . However, for  $p \geq 5$ , one might need to use the resulted values from applying the above relations as a mean value around which small random variance is introduced; this iterative process is repeated until the correct number of extrema is obtained.

### C. Proposed Approximations

The convergence of the algorithm is illustrated in Fig. 1, which shows an example of finding the uniform error function for  $N = 3$  that results in seven extrema points. The approximation converges to its minimax behaviour after three iterations starting from a non-uniform error function with the correct number of extrema and ending with all the extrema points having the same value of error.

The new sets of the optimized coefficients of the considered approximation (3) are solved herein for  $N = 1, 2, 3, \dots, 10$  and  $p = 1, 2, 3, 4$  in the minimax sense. In particular, we have calculated the required initial guesses using the heuristic method explained in the previous subsection and then applied the iterative Remez algorithm to obtain the uniform exponential approximation. In Fig. 2, we illustrate the achieved global absolute error,  $d_{max}$ , in all the considered cases. We can clearly see that as the number of terms increases, the global error decreases resulting in very high accuracy.

## IV. APPLICATION EXAMPLES

In general, the ASEP of most of the digital modulation techniques for coherent detection are linear combinations of integrals, whose integrand is the product of powers of the Gaussian  $Q$ -function and the fading probability density function (PDF) of the fading channel as follows:

$$I_p(\alpha) \triangleq \int_0^\infty Q^p(\alpha \sqrt{\gamma}) f_\gamma(\gamma) d\gamma, \quad (11)$$

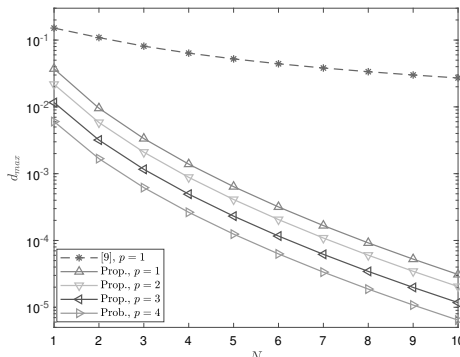


Fig. 2. The global absolute error when  $\tilde{Q}_p(x)$  is the minimax approximation of  $Q^p(x)$  for  $p = 1, 2, 3, 4$ , and when  $\tilde{Q}_1(x)$  is the non-optimized rectangular rule in [9], both for  $N = 1, 2, 3, \dots, 10$ .

where  $\gamma$  is the instantaneous signal-to-noise ratio (SNR), with  $f_\gamma(\gamma)$  being its PDF, and  $\alpha$  is a constant that depends on the digital modulation and detection techniques. For example, the conditional SEP in coherent detection of quadrature amplitude modulation (QAM) and differentially encoded quadrature phase-shift keying (DE-QPSK) are calculated by [1]

$$P_E = 2Q(\sqrt{\gamma}) - Q^2(\sqrt{\gamma}), \quad (12)$$

$$P_E = 4Q(\sqrt{\gamma}) - 8Q^2(\sqrt{\gamma}) + 8Q^3(\sqrt{\gamma}) - 4Q^4(\sqrt{\gamma}), \quad (13)$$

respectively, and the corresponding ASEPs in terms of (11) thus become  $\bar{P}_E = 2I_1(1) - I_2(1)$  for 4-QAM and  $\bar{P}_E = 4I_1(1) - 8I_2(1) + 8I_3(1) - 4I_4(1)$  for DE-QPSK.

Next we substitute the exponential approximation into (11) to obtain

$$\begin{aligned} I_p(\alpha) &\approx \sum_{n=1}^N a_n \int_0^\infty \exp(-b_n \alpha^2 \gamma) f_\gamma(\gamma) d\gamma \\ &= \sum_{n=1}^N a_n M_\gamma(-b_n \alpha^2), \end{aligned} \quad (14)$$

where  $M_\gamma(s) = \int_0^\infty \exp(s\gamma) f_\gamma(\gamma) d\gamma$  is the moment generating function (MGF) associated with the random variable  $\gamma$ . In what follows, we derive closed-form expressions for the general ASEP term defined in (11) over different fading channels, namely Nakagami- $m$ , Fisher-Snedecor  $\mathcal{F}$ ,  $\eta - \mu$ , and  $\kappa - \mu$  fading channels.

### A. Nakagami- $m$ Fading

For Nakagami- $m$  fading, we substitute the gamma MGF, i.e.,  $M_\gamma(s) = (1 - \frac{s\bar{\gamma}}{m})^{-m}$ , in (14) which yields directly

$$I_p(\alpha) \approx \sum_{n=1}^N a_n \left( 1 + \frac{b_n \alpha^2 \bar{\gamma}}{m} \right)^{-m}, \quad (15)$$

where  $m > 0$  is the fading parameter and  $\bar{\gamma}$  is the average SNR. The ASEP of 4-QAM and DE-QPSK over Nakagami- $m$  fading are calculated using (15) and the corresponding absolute error is illustrated in Fig. 3.

### B. Fisher–Snedecor $\mathcal{F}$ Fading

Next we find analytical results for (11) with Fisher–Snedecor  $\mathcal{F}$  distribution which is used to model the composite effects of both small and large scale fading (shadowing). The former is assumed to follow Nakagami- $m$  distribution, and the latter follows inverse Nakagami- $m$  distribution. We substitute the MGF derived in [18, Eq. 10] in (14), which yields

$$I_p(\alpha) \approx \sum_{n=1}^N a_n {}_1F_1\left(m; 1 - m_s; \frac{b_n \alpha^2 \bar{\gamma} m_s}{m}\right) + \frac{\Gamma(-m_s)}{\beta(m, m_s)} \\ \times \left(\frac{b_n \alpha^2 \bar{\gamma} m_s}{m}\right)^{m_s} {}_1F_1\left(m + m_s; 1 + m_s; \frac{b_n \alpha^2 \bar{\gamma} m_s}{m}\right),$$

where  $m$  is the fading severity parameter,  $m_s \neq \mathbb{N}$  is the shadowing parameter,  $\beta(\cdot, \cdot)$  and  ${}_1F_1(\cdot; \cdot; \cdot)$  denote beta and Kummer confluent hypergeometric functions, respectively.

### C. Generalized $\eta - \mu$ and $\kappa - \mu$ Fading

Finally, we evaluate the average of arbitrary powers of the  $Q$ -function in (11) over  $\eta - \mu$  and  $\kappa - \mu$  fading channels. The former fits well for non-line-of-sight applications and includes the Nakagami- $q$  (Hoyt) and Nakagami- $m$  fading as special cases while the latter fits better to line-of-sight applications and includes the Rice and Nakagami- $m$  fading as special cases. We calculate their MGFs from their PDFs [19, Eqs. 1, 4] and we substitute them in (14). Thus, under  $\eta - \mu$  fading we obtain

$$I_p(\alpha) \approx \frac{2\sqrt{\pi} \mu^{\mu+\frac{1}{2}} h^\mu}{\Gamma(\mu) H^{\mu-\frac{1}{2}} \bar{\gamma}^{\mu+\frac{1}{2}}} \sum_{u=0}^{\infty} \frac{\Gamma(2\mu + 2u)}{u! \Gamma(\mu - \frac{1}{2} + u + 1)} \\ \times \left(\frac{\mu H}{\bar{\gamma}}\right)^{\mu-\frac{1}{2}+2u} \sum_{n=1}^N a_n \left(b_n \alpha^2 + \frac{2\mu h}{\bar{\gamma}}\right)^{-(2\mu+2u)},$$

where  $\eta$  and  $\mu$  are the fading parameters,  $h = (2 + \eta^{-1} + \eta)/4$  and  $H = (\eta^{-1} - \eta)/4$  for Format 1 of the distribution and  $h = \frac{1}{(1-\eta^2)}$  and  $H = \eta/(1-\eta^2)$  for Format 2. On the other hand, for the  $\kappa - \mu$  fading model, we obtain

$$I_p(\alpha) \approx \frac{1}{\exp(\mu\kappa)} \sum_{u=0}^{\infty} \frac{\mu^{\mu+2u} \kappa^u (1+\kappa)^{\mu+u}}{\bar{\gamma}^{\mu+u} \Gamma(u+1)} \sum_{n=1}^N a_n \left(b_n \alpha^2 + \frac{\mu(1+\kappa)}{\bar{\gamma}}\right)^{-\mu-u}, \quad (16)$$

in which  $\kappa > 0$  is the ratio between the total power of the dominant components and the total power of the scattered waves, and  $\mu > 0$  is the number of multipath clusters.

## V. NUMERICAL RESULTS

Throughout this section, we will be dealing with the absolute error function obtained by subtracting the numerically calculated exact expression of  $I_p$  defined in (11), from the approximated one in (14). The same applies for ASEP which is a linear combination of  $I_p$ . In Fig. 3, we compare the absolute error calculated from the proposed approximations and the existing ones with the same form, for different values of  $m$ . It is observed that our approximation have the least global error and result in a tighter approximation of the ASEP over the whole range of the average SNR for  $m = 0.5$  as seen for 4-QAM plot. For higher values of  $m$ , some of the

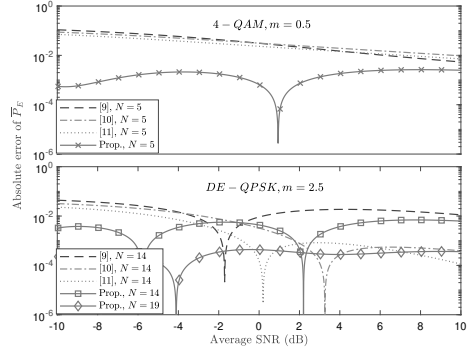


Fig. 3. The absolute error of ASEP for 4-QAM and DE-QPSK over Nakagami- $m$  using the proposed approximation and the reference exponential approximations.

existing approximations have higher accuracy with exactly the same number of exponential terms as seen for DE-QPSK plot. However, when increasing number of terms, the accuracy of our approximation increases significantly and outperforms the others for almost the whole range of average SNR. It should be mentioned that increasing the number of exponential terms does not affect the analytical complexity. Moreover, when substituting the reference approximations with two terms in (12) and (13), we get 5-term and 14-term approximations for the ASEP in 4-QAM and DE-QPSK, respectively.

Figure 4 compares the difference between the exact  $I_p$  in (11) and its approximations in Nakagami- $m$ , Fisher–Snedecor  $\mathcal{F}$ ,  $\eta - \mu$ , and  $\kappa - \mu$  fading channels presented in (15), after (15), before(16), and in (16), respectively, calculated using the existing and proposed approximations, for different values of the fading parameters and different integer powers. It is seen that the proposed approximations are tight even for lower SNR values, opposing to the existing ones. In particular, our approximation outperforms the others for a wide range of the argument using the same number of exponential terms and its accuracy can be increased even further by increasing the number of terms. The reference approximations are derived for a limited number of terms, namely  $N = 2, 3$  or  $4$  only, whereas our approximations are derived till  $N = 10$  to offer higher and adequate accuracy without affecting analytical complexity.

## VI. CONCLUSION

This paper proposed accurate and tractable approximations for the integer powers of the  $Q$ -function as a weighted sum of exponential functions. The novel sets of coefficients of the best exponential approximation are optimally solved using the Remez exchange algorithm to obtain uniform alternating absolute error function. We also considered the general problem of evaluating the ASEP over different fading channels, in which we implemented our approximations and showed that they render high accuracy in terms of global error and for the whole

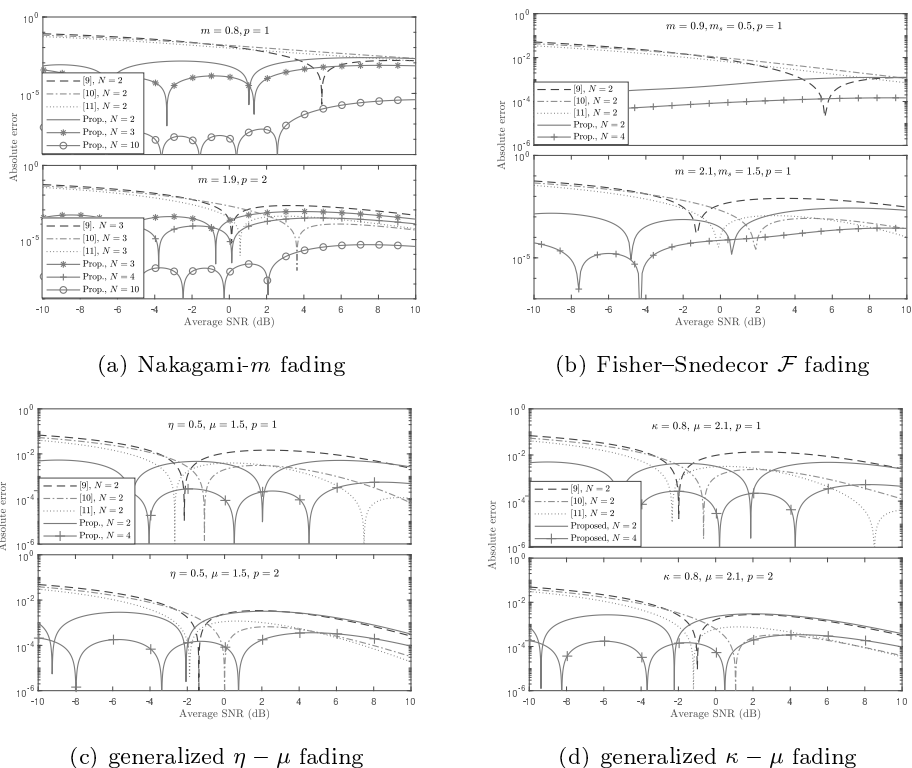


Fig. 4. The absolute error of  $I_p(\alpha)$  over several fading distributions, all with  $\alpha = 1$  and for different fading parameters.

range of the argument. Even higher accuracy can be achieved by simply increasing the number of exponential terms.

#### REFERENCES

- [1] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels*, 2nd ed. John Wiley and Sons, Inc., Jan. 2005.
- [2] W. Cody, "Rational Chebyshev approximations for the error function," *Math. Comp.*, vol. 23, no. 107, pp. 631–637, Jul. 1969.
- [3] P. Börjesson and C. Sundberg, "Simple approximations of the error function  $Q(x)$  for communications applications," *IEEE Trans. Commun.*, vol. 27, no. 3, pp. 639–643, Mar. 1979.
- [4] G. Karagiannidis and A. Lioumpas, "An improved approximation for the Gaussian  $Q$ -function," *IEEE Commun. Lett.*, vol. 11, no. 8, pp. 644–646, Aug. 2007.
- [5] I. M. Tanash and T. Riihonen, "Improved coefficients for the Karagiannidis-Lioumpas approximations and bounds to the Gaussian  $Q$ -function," *IEEE Commun. Lett.*, IEEE Early Access, 2021.
- [6] Y. Isukapalli and B. Rao, "An analytically tractable approximation for the Gaussian  $Q$ -function," *IEEE Commun. Lett.*, vol. 12, no. 9, pp. 669–671, Sep. 2008.
- [7] Y. Chen and N. Beaulieu, "A simple polynomial approximation to the Gaussian  $Q$ -function and its application," *IEEE Commun. Lett.*, vol. 13, no. 2, pp. 124–126, Feb. 2009.
- [8] M. López-Benítez and F. Casadevall, "Versatile, accurate, and analytically tractable approximation for the Gaussian  $Q$ -function," *IEEE Trans. Commun.*, vol. 59, no. 4, pp. 917–922, Apr. 2011.
- [9] M. Chiani, D. Dardari, and M. Simon, "New exponential bounds and approximations for the computation of error probability in fading channels," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 840–845, Jul. 2003.
- [10] P. Loskot and N. Beaulieu, "Prony and polynomial approximations for evaluation of the average probability of error over slow-fading channels," *IEEE Trans. Veh. Technol.*, vol. 58, no. 3, pp. 1269–1280, Mar. 2009.
- [11] O. Olabiya and A. Annamalai, "Invertible exponential-type approximations for the Gaussian probability integral  $Q(x)$  with applications," *IEEE Wireless Commun. Lett.*, vol. 1, no. 5, pp. 544–547, Oct. 2012.
- [12] D. Sadhwani, R. Yadav, and S. Aggarwal, "Tighter bounds on the Gaussian  $Q$  function and its application in Nakagami- $m$  fading channel," *IEEE Wireless Commun. Lett.*, vol. 6, no. 5, pp. 574–577, Oct. 2017.
- [13] I. M. Tanash and T. Riihonen, "Global minimax approximations and bounds for the Gaussian  $Q$ -function by sums of exponentials," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6514–6524, Oct. 2020.
- [14] J. Craig, "A new, simple and exact result for calculating the probability of error for two-dimensional signal constellations," in *Proc. IEEE Mil. Commun. Conf.*, vol. 2, Nov. 1991, pp. 571–575.
- [15] E. Remes, "Sur le calcul effectif des polynômes d'approximation de Techebychef," *C. R. Acad. Sci.*, pp. 337–340, 1934.
- [16] D. Kammler, "Chebyshev approximation of completely monotonic functions by sums of exponentials," *SIAM J. on Numer. Anal.*, vol. 13, no. 5, pp. 761–774, Oct. 1976.
- [17] R. McGlinn, "Uniform approximation of completely monotone functions by exponential sums," *J. Math. Anal. Appl.*, vol. 65, no. 1, pp. 211–218, Aug. 1978.
- [18] S. Yoo, S. Cotton, P. Sofotasios, M. Matthaiou, M. Valkama, and G. Karagiannidis, "The Fisher-Snedecor  $\mathcal{F}$  distribution: A simple and accurate composite fading model," *IEEE Commun. Lett.*, vol. 21, no. 7, pp. 1661–1664, Jul. 2017.
- [19] N. Ermolova, "Moment generating functions of the generalized  $\eta - \mu$  and  $\kappa - \mu$  distributions and their applications to performance evaluations of communication systems," *IEEE Commun. Lett.*, vol. 12, no. 7, pp. 502–504, Jul. 2008.



# PUBLICATION

3

## Quadrature-based exponential-type approximations for the Gaussian $Q$ -function

I. M. Tanash and T. Riihonen

In *Proc. IEEE Vehicular Technology Conference (VTC)*, Apr. 2021, pp. 1–5

DOI: 10.1109/VTC2021-Spring51267.2021.9448918

**Publication reprinted with the permission of the copyright holders.**



# Quadrature-Based Exponential-Type Approximations for the Gaussian $Q$ -Function

Islam M. Tanash and Taneli Riihonen

Faculty of Information Technology and Communication Sciences, Tampere University, Finland

e-mail: {islam.tanash, taneli.riihonen}@tuni.fi

**Abstract**—In this paper, we present a comprehensive overview of (perhaps) all possible approximations resulting from applying the most common numerical integration techniques on the Gaussian  $Q$ -function. We also present a unified method to optimize the coefficients of the resulting exponential approximation for any number of exponentials and using any numerical quadrature rule to produce tighter approximations. Two new tight approximations are provided as examples by implementing the Legendre numerical rule with Quasi-Newton method for two and three exponential terms. The performance of the different numerical integration techniques is evaluated and compared, and the accuracy of the optimized ones is verified for the whole argument-range of interest and in terms of the chosen optimization criterion.

## I. INTRODUCTION

The Gaussian  $Q$ -function  $Q(\cdot)$  and the related complementary error function  $\operatorname{erfc}(\cdot)$  represent significant value in the performance analysis of different communication systems, where noise or interference is typically modeled as a Gaussian random variable, such as evaluating error probabilities. As the Gaussian  $Q$ -function does not have an exact closed-form expression, several approximations and bounds have been proposed in the literature to facilitate its applications, especially when complicated integrals involving it are encountered.

The authors in [1] propose tight approximations for the  $Q$ -function using rational Chebyshev functions. In [2], the authors use integration by parts to derive new bounds and take their geometric mean to yield an approximation; another complex approximation with two controllable parameters that define the level of accuracy is also derived. A relatively accurate approximation for the  $Q$ -function and integer powers thereof are provided in [3], [4]. Based on [3], an upper bound is later developed in [5] and the approximation of the first power is modified in [6] using Taylor series expansion to get a simpler form. An infinite series expression for  $\operatorname{erfc}(x)$  that is more accurate for large values of  $x$  is derived in [7].

The authors in [8] present an accurate polynomial approximation. In [9], a relatively complicated new family of bounds is proposed using Jensen's inequality with the Cotes trapezoidal integration rule and convex-concave partitions of the integrand of the Craig's formula:

$$Q(x) \triangleq \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \exp\left(-\frac{1}{2\sin^2\theta} x^2\right) d\theta \quad [\text{for } x \geq 0]. \quad (1)$$

Moreover, tight bounds are presented in [10] as a sum of two exponentials with respective constant and rational factors. The authors in [11] propose a simple and accurate mathematical expression as an exponential function with a

polynomial argument of the second degree. In addition, [12] develops an efficient approximation based on a semi-infinite Gauss-Hermite quadrature rule that results in a finite sum of exponential functions.

Chiani *et al.* in [13] propose a simple family in the form of exponential sums, in which they apply the trapezoidal integration rule with optimizing the center point to minimize the integral of relative error in the range of interest. This family was later generalized in [14] to be applied to polynomials and integer powers of the  $Q$ -function, or even to generic functions thereof using an original minimax methodology. Other approximations and bounds of this form are presented in [15]–[19]. In [15], the composite trapezoidal rule is used with an optimized number of sub-intervals, whereas in [16], a coarse single-term approximation from the classic Chernoff bound is presented. Since the  $Q$ -function can be well approximated as an infinite sum of exponentials, [17] presents the Prony approximation. An invertible exponential approximation is presented in [18] and a single-term exponential lower bound is introduced in [19] by upper-bounding the logarithmic function.

The approximations and bounds in [1]–[12] have relatively complex mathematical forms, which makes them inconvenient for algebraic manipulations in statistical performance analysis despite being accurate. On the other hand, those with the exponential form [13], [15]–[19] are more suitable to be used due to their analytical tractability. However, they still provide an inadequate accuracy for some of the argument range of interest, i.e., some of them are suitable only for some certain range. This leads us toward the various numerical integration methods [20], i.e., quadrature rules, which can be implemented to approximate the Craig's form of the  $Q$ -function (1) to obtain a flexible approximations and bounds of the same form as in [13, Eq. (8)] with numerical coefficients instead. Based on the accuracy level required in a given range, the suitable numerical quadrature rule is selected to approximate the  $Q$ -function.

The goal of this paper is to present an overview of all the known numerical integration techniques that are commonly used to approximate the Gaussian  $Q$ -function and compare their performance. In particular, we consider Newton-Cotes formulas, Gaussian quadrature formulas and the composite integration rules. In addition, we generalize the work of Chiani *et al.* [13] from only  $N = 2$  exponential terms to any  $N$  and using any numerical integration technique by minimizing optimization criteria for composite intervals thereof. In terms of explicit expressions, we provide tight exponential approximations using the Legendre rule for quick reference.

## II. APPROXIMATIONS FROM NUMERICAL INTEGRATION

The quadrature integration techniques that can be used to approximate the  $Q$ -function (viz.  $Q(x) \approx \tilde{Q}(x)$ ) are categorized herein as Newton–Cotes formulas and Gaussian quadrature formulas. Due to the instability of higher-order numerical methods (especially with the Newton–Cotes rules, which have negative weights that can result in subtractive cancellation), the composite integration rules are also considered in this paper.

In general, any integral of the form  $\int_u^v W(\theta) f(\theta) d\theta$ , where  $W(\theta)$  is some weighting function and  $[u, v]$  is the domain of integration, can be rewritten as a finite sum of the form [20]

$$\int_u^v W(\theta) f(\theta) d\theta = \sum_{g=1}^G w_g f(\theta_g) + D(\xi), u < \xi < v, \quad (2)$$

where  $D(\xi)$  is the resulting error term,  $\{\theta_g\}_{g=1}^G$  are the nodes and  $\{w_g\}_{g=1}^G$  are the quadrature weights. Thus, the  $Q$ -function that is defined by (1) over the interval  $[0, \pi/2]$  can be numerically approximated after applying (2) as

$$Q(x) \approx \tilde{Q}(x) \triangleq \sum_{g=1}^G a_g \exp(-b_g x^2) \quad (3)$$

such that  $Q(x) = \tilde{Q}(x) + D(\xi, x)$  for some  $u < \xi < v$ , where  $x \geq 0$  and  $\{(a_g, b_g)\}_{g=1}^G$  is the set of numerical coefficients, which depends on the specific applied numerical integration technique as will be explained below.

### A. Newton–Cotes Numerical Integration

A Newton–Cotes formula can be either closed or open, depending on whether it uses the function values at the endpoints or not. The weights for Newton–Cotes rules are derived from Lagrange basis polynomials as

$$w_g = \int_u^v \prod_{\substack{t=1 \\ t \neq g}}^G \frac{\theta - \theta_t}{\theta_g - \theta_t} d\theta = c_g \Delta\theta, \quad (4)$$

where  $c_g, g = 1, 2, \dots, G$ , are constants that depend on the type of the applied Newton–Cotes rule and can be found in many mathematical books, e.g., [21], whereas  $\Delta\theta$  is the step size. For the Newton–Cotes rule, the nodes are always chosen uniformly in the integration interval. Therefore, when applied to the Gaussian  $Q$ -function, the numerical coefficients of the exponential summation in (3) are calculated as

$$\{(a_g, b_g)\}_{g=1}^G = \begin{cases} \left\{ \left\{ \left( \frac{c_g \Delta\theta}{\pi}, \frac{1}{2 \sin^2((g-1) \Delta\theta)} \right) \right\}_{g=1}^G \right\}, & \text{for closed types,} \\ \left\{ \left\{ \left( \frac{c_g \Delta\theta}{\pi}, \frac{1}{2 \sin^2(g \Delta\theta)} \right) \right\}_{g=1}^G \right\}, & \text{for open types.} \end{cases} \quad (5)$$

where  $\Delta\theta = \frac{\pi}{2(\tilde{G}-1)}$  and  $\Delta\theta = \frac{\pi}{2\tilde{G}+1}$  for the closed and open types, respectively.

The error term for the Newton–Cotes rules is

$$D(\xi, x) = \frac{f^{(G)}(\xi, x)}{G!} \int_0^{\pi/2} \prod_{g=1}^G (\theta - \theta_g) d\theta,$$

$0 < \xi < \pi/2, x \geq 0$ , which shows that there exists some (unknown) point  $\xi \in (0, \pi/2)$  for each  $x$ , for which the respective error has exactly the displayed form.

### B. Gaussian Quadrature Numerical Integration

Another type of numerical integration techniques is the Gaussian quadrature family. For this type, the domain of integration in (2) is  $[-1, 1]$ , but since the  $Q$ -function is defined over  $[u, v] = [0, \pi/2]$  for Craig's formula, a change of variables is needed. This results in multiplying the weights by  $\frac{v-u}{2}$  and transforming the nodes as  $\frac{v-u}{2} \theta_g + \frac{v+u}{2}$ . Thus, the numerical coefficients of the exponential sum in (3), when applying Gaussian quadrature numerical rules, are

$$\{(a_g, b_g)\}_{g=1}^G = \left\{ \left( \frac{1}{4} w_g, \frac{1}{2 \sin^2\left(\frac{\pi}{4} \theta_g + \frac{\pi}{4}\right)} \right) \right\}_{g=1}^G. \quad (6)$$

Five Gaussian rules are used herein for comparison purposes, namely Legendre, Chebyshev first and second kinds, Radau's, and Lobatto's rules. For the Chebyshev first and second kind rules, we should consider their weighting functions which result in  $a_g = \frac{1}{4} w_g \sqrt{1 - \theta_g^2}$  and  $a_g = \frac{w_g}{4 \sqrt{1 - \theta_g^2}}$ , respectively. Table I summarizes the expressions for finding  $w_g, \theta_g$  and  $D(\xi, x)$  of these five Gaussian rules [21] while  $\phi_G(\theta)$  in the table is the Legendre polynomial of degree  $G$ .

### C. Composite Integration

The composite quadrature rules are preferred to approximate the  $Q$ -function for higher orders due to the oscillatory nature of high-degree polynomials in non-composite rules. The integration interval,  $[u, v] = [0, \pi/2]$ , can be divided into  $M$  smaller uniform or non-uniform sub-intervals,  $[u_m, v_m], m = 1, 2, \dots, M$ , and simpler  $K$ -point integration rule is used for each sub-interval, where  $K$  is the number of nodes in each sub-interval. Therefore, the Gaussian  $Q$ -function is approximated by applying any composite integration rule as

$$\tilde{Q}(x) = \sum_{m=1}^M \sum_{k=1}^K a_{m,k} \exp(-b_{m,k} x^2). \quad (7)$$

The numerical coefficients for the  $m$ th sub-interval are given for the  $K$ -point composite Newton–Cotes rules as

$$\{(a_{m,k}, b_{m,k})\}_{k=1}^K = \begin{cases} \left\{ \left\{ \left( \frac{c_k \Delta\theta_m}{\pi}, \frac{1}{2 \sin^2(u_m + (k-1) \Delta\theta_m)} \right) \right\}_{k=1}^K \right\}, & \text{for closed types,} \\ \left\{ \left\{ \left( \frac{c_k \Delta\theta_m}{\pi}, \frac{1}{2 \sin^2(u_m + k \Delta\theta_m)} \right) \right\}_{k=1}^K \right\}, & \text{for open types,} \end{cases} \quad (8)$$

where  $\Delta\theta_m = \frac{v_m - u_m}{K-1}$  and  $\Delta\theta_m = \frac{v_m - u_m}{K+1}$  for the closed and open types, respectively. For the uniform sub-intervals,  $\Delta\theta_m = \frac{\pi}{2M(K-1)}$  and  $\Delta\theta_m = \frac{\pi}{2M(K+1)}$ , respectively.

On the other hand, for composite Gaussian quadratures, the numerical coefficients for the  $m$ th sub-interval are given as

$$\{(a_{m,k}, b_{m,k})\}_{k=1}^K = \left\{ \left\{ \left( \frac{(v_m - u_m) w_k}{2\pi}, \frac{1}{2 \sin^2\left(\frac{(v_m - u_m)}{2} \theta_k + \frac{(v_m + u_m)}{2}\right)} \right) \right\}_{k=1}^K \right\}, \quad (9)$$

TABLE I  
GAUSSIAN QUADRATURE NUMERICAL INTEGRATION METHODS.

Gaussian rule	$w_g$	$\theta_g$	$D(\xi, x), -1 < \xi < 1, x \geq 0$
Legendre rule	$\frac{2}{(1-\theta_g^2)\phi_G'(\theta_g)}$	$g^{\text{th}}$ zero of $\phi_G(\theta)$	$\frac{2^{2G+1}(G!)^4}{(2G+1)[(2G)!]^3} f^{(2G)}(\xi, x)$
Chebyshev first kind	$\frac{\pi}{G}$	$\cos\left(\frac{2g-1}{2G}\pi\right)$	$\frac{\pi}{(2G)2^{2G-1}} f^{(2G)}(\xi, x)$
Chebyshev second kind	$\frac{\pi}{G+1} \sin^2\left(\frac{g}{G+1}\pi\right)$	$\cos\left(\frac{g}{G+1}\pi\right)$	$\frac{\pi}{(2G)!2^{2G+1}} f^{(2G)}(\xi, x)$
Radau's rule	$\frac{1}{(1-\theta_g)\phi_{G-1}'(\theta_g)}$	$g^{\text{th}}$ zero of $\frac{\phi_{G-1}(\theta)+\phi_G(\theta)}{\theta+1}$	$\frac{2^{2G-1}G}{[(2G-1)!]^3} [(G-1)!]^4 f^{(2G-1)}(\xi, x)$
Lobatto's rule	$\frac{2}{G(G-1)\phi_{G-1}'(\theta_g)}$	$(g-1)^{\text{th}}$ zero of $\phi_{G-1}'(\theta)$	$\frac{-G(G-1)^2 2^{2G-1} [(G-2)!]^4}{(2G-1)[(2G-2)!]^3} f^{(2G-2)}(\xi, x)$

where  $w_k$  and  $\theta_k$  are the same weights and nodes as illustrated in Table I. When considering equal-spaced intervals, the numerical coefficients become

$$\{(a_{m,k}, b_{m,k})\}_{k=1}^K = \left\{ \left( \frac{w_k}{4M}, \frac{1}{2 \sin^2\left(\frac{\pi}{4M}\theta_k + \frac{(2m-1)\pi}{4M}\right)} \right) \right\}_{k=1}^K.$$

As a remark, one should note that the single-point Legendre rule is mathematically the same as the single-point open Newton–Cotes (a.k.a. rectangular) rule, and the two-point Lobatto's rule is the same as the two-point closed Newton–Cotes (a.k.a. trapezoidal) rule. In addition, the weighting functions of Chebyshev rules should be also considered here when calculating the numerical coefficients as explained above.

#### D. Implementation Aspects and Numerical Results

Accuracy comparison between the different composite and non-composite numerical integration techniques is presented herein for the same number of non-zero exponential terms which we refer to as  $N$ . In general, if the left endpoint of the integration domain is used in the summation in (3) as a node, i.e.,  $\theta_1 = 0$ , to evaluate the integration then the function's value at that node is equal to zero, i.e., integrand in (1) evaluates to zero, and hence the first term in the summation is neglected. In this case, in order to establish exactly  $N$  exponential terms in the summation,  $G = N + 1$ , since the first term in the  $(N + 1)$ -term summation is zero and hence the total number of exponential terms is  $N$ . On the other hand, if the left endpoint is not included in the summation, then  $G = N$ . In plain words,  $G$  refers to the total number of terms including zero if any, and  $N$  refers to that of the non-zero terms.

The same applies to the composite rules for which we want to construct an  $N$ -term expression in (7). Hence,  $N = MK$  for the composite open Newton–Cotes, Legendre and Chebyshev first and second type rules, while  $N = M(K - 1)$ , for the composite closed Newton–Cotes and Lobatto's rules since the two endpoints of each sub-interval are included, which results in adding the last term of the  $m$ th sub-interval to the first term of the  $(m + 1)$ th sub-interval to produce a single term (in addition to the fact that the first term of the first sub-interval is neglected since its  $b_{1,1} \rightarrow \infty$ ). This is also the reason why for the composite Radau's numerical rule  $N = MK - 1$ .

In Fig. 1, we illustrate the absolute relative error resulted from applying all the different types of numerical integration rules and their composites to approximate the Gaussian

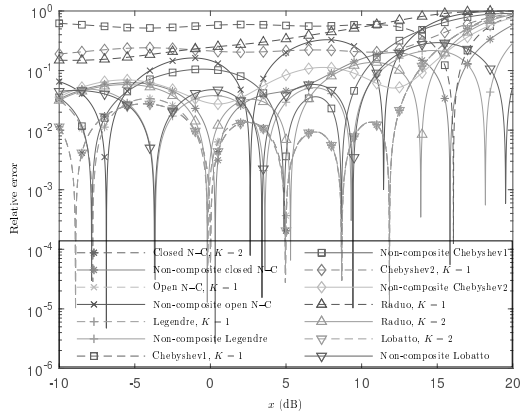


Fig. 1. Comparison among all the different numerical integration techniques for  $N = 3$ . The abbreviation N-C refers to Newton–Cotes rule.

$Q$ -function for  $N = 3$ . It is noted that the two-point closed Newton–Cotes (equivalent to two-point Lobatto's rule) and the single-point Legendre (equivalent to single-point open Newton–Cotes rule) rules are the most accurate for a wide argument range. However, some techniques outperform the others for different  $x$ -ranges, e.g., the two-point Radau's rule has the least error for  $13 \leq x \leq 15$ , whereas the four-point closed Newton–Cotes rule is more accurate for  $15 \leq x \leq 17$ .

### III. OPTIMIZED NUMERICAL APPROXIMATIONS

The seminal work in [13] has shown that the traditional trapezoidal rule can be optimized with respect to the integral of the relative error in the range of values of interest. This yields a tighter approximation with respect to the optimization criterion than the un-optimized one. This approach has also been implemented in [22] to propose a new sum for the trapezoidal approximation of three exponentials. In this section, we present a unified method to optimize any composite numerical rule with any  $N$  and in respect to any optimization criterion.

The composite rule is based on dividing the integration interval into smaller uniform or non-uniform sub-intervals. In the previous section, we give explicit expressions for the numerical coefficients of the uniform composite rules. On the other hand, for the non-uniform case, the integration interval can be arbitrarily partitioned and general expression

are presented for their corresponding coefficients. However, the division can actually be chosen optimally in order to increase the approximation's accuracy, e.g., like in [13].

We can conclude from the relations (8) and (9) that the optimized set of coefficients,  $\{(a_{m,k}, b_{m,k})\}_{k=1}^K$ , are obtained by optimizing the sub-intervals boundaries,  $[u_m, v_m]$ ,  $m = 1, 2, \dots, M$ , according to the chosen optimization criterion with keeping in mind that  $v_m = u_{m+1}$ ,  $m = 1, 2, \dots, M-1$ . Thus, the whole integration range has  $M+1$  boundary points in which for the Gaussian  $Q$ -function,  $u_m = 0$  and  $v_M = \pi/2$ . This will give a total of  $M-1$  boundary points to be optimized.

Any optimization criterion could be selected for calculating the approximation's corresponding optimized numerical coefficients. For consistency with [13], we consider the integral of the relative error in the range of values of interest  $[0, R]$

$$F(\mathbf{v}) = \frac{1}{R} \int_0^R \left| \frac{\tilde{Q}(x) - Q(x)}{Q(x)} \right| dx, \quad (10)$$

and the numerical approximation is optimized as

$$\mathbf{v}^* \triangleq \underset{\mathbf{v}}{\arg \min} F(\mathbf{v}), \quad (11)$$

where  $\mathbf{v} = [v_1, v_2, \dots, v_{M-1}]$  is the vector of unknowns (i.e., the boundary points to optimize),  $Q(x)$  is defined in (1) and  $\tilde{Q}(x)$  with the corresponding expressions of the numerical coefficients is defined in (7), (8) and (9), respectively. It should be noted that when using two-point closed Newton–Cotes rule, i.e., the trapezoidal rule, with  $M = 2$ , (7) will become [13, Eq. (12)] but in terms of the  $Q$ -function instead of  $\operatorname{erfc}(\cdot)$ .

We solve this optimization problem for all values of  $M$  and  $K$  and for an arbitrary integration rule by applying the Quasi-Newton method, which is an iterative technique for finding the roots of a given differentiable function. It can also be used in the context of optimization by applying it on the derivative of the target function which is differentiable twice. This will yield the optimized roots of the function's derivative.

In particular, we implement the Quasi-Newton optimization method herein to minimize the target function  $F(\mathbf{v})$  in (10). We start with some initial guesses for the  $M-1$  unknowns that converge eventually to the optimized values, which give the minimum possible value for the target function. The iteration process is performed as

$$\mathbf{v}^{(t+1)} = \mathbf{v}^{(t)} - \gamma \left[ \tilde{\mathbf{H}}^{(t)} \right]^{-1} \mathbf{J}^{(t)}(\mathbf{v}^{(t)}), \quad (12)$$

where  $t$  is the iteration counter,  $0 < \gamma \leq 1$  is the iteration step size,  $\mathbf{J}(\cdot)$  is the gradient vector calculated as  $\mathbf{J}(\mathbf{v}) = \left[ \frac{\partial F(\mathbf{v})}{\partial v_1} \quad \frac{\partial F(\mathbf{v})}{\partial v_2} \quad \dots \quad \frac{\partial F(\mathbf{v})}{\partial v_{M-1}} \right]$ , and  $\tilde{\mathbf{H}}$  is an approximation to the Hessian matrix. Among the various methods developed to calculate  $\tilde{\mathbf{H}}$ , we implement the well-known Broyden–Fletcher–Goldfarb–Shanno (BFGS) method which starts from some symmetric positive-definite matrix  $\tilde{\mathbf{H}}^{(0)}$  that is updated in the sequential iterations as

$$\tilde{\mathbf{H}}^{(t+1)} = \tilde{\mathbf{H}}^{(t)} + \frac{\Delta \mathbf{J}^{(t)} [\Delta \mathbf{J}^{(t)}]^T}{[\Delta \mathbf{J}^{(t)}]^T \Delta \mathbf{v}^{(t)}} - \frac{\tilde{\mathbf{H}}^{(t)} \Delta \mathbf{v}^{(t)} [\Delta \mathbf{v}^{(t)}]^T [\tilde{\mathbf{H}}^{(t)}]^T}{[\Delta \mathbf{v}^{(t)}]^T \tilde{\mathbf{H}}^{(t)} \Delta \mathbf{v}^{(t)}}, \quad (13)$$

where  $[\cdot]^T$  denotes the transpose,  $\Delta \mathbf{J}^{(t)} = \mathbf{J}^{(t+1)}(\mathbf{v}^{(t+1)}) - \mathbf{J}^{(t)}(\mathbf{v}^{(t)})$  and  $\Delta \mathbf{v}^{(t)} = \mathbf{v}^{(t+1)} - \mathbf{v}^{(t)}$ . The iterations of Quasi-Newton method are repeated until the difference between the values of  $\mathbf{v}$  of two successive iterations become smaller than some predefined threshold value. It is worth mentioning that, for practical implementation, we can directly use the `fminunc` command in Matlab with setting its corresponding algorithm to 'quasi-newton' and choosing initial values in the range  $(0, \frac{\pi}{2})$  for  $\mathbf{v}$  in order to eventually find  $\mathbf{v}^*$  in (11).

Let us next develop new tight exponential approximations by finding new optimized numerical coefficients using the approach explained above. Based on the observations concluded from Fig. 1, we use Legendre rule to formulate the new approximations as it is one of the approximations that has the least relative error among the various quadrature rules.

For  $N = 2$ , using the single-point Gauss Legendre rule ( $K = 1$ ), the integration interval is divided into  $M = 2$  sub-intervals which yields one unknown as  $\mathbf{v} = [v_1]$ . Thus, the approximated  $Q$ -function in (7) after using (9) and calculating  $w_k$  and  $\theta_k$  from Table I, will result in

$$\tilde{Q}(x) = \frac{v_1^*}{\pi} \exp\left(\frac{-x^2}{2 \sin^2(\frac{v_1^*}{2})}\right) + \left(\frac{1}{2} - \frac{v_1^*}{\pi}\right) \exp\left(\frac{-x^2}{2 \sin^2(\frac{\pi}{4} + \frac{v_1^*}{2})}\right). \quad (14)$$

The optimum value of the parameter  $v_1$  is calculated using Quasi-Newton optimization method with respect to (11) for the case of  $R = 13$  dB and is found to be  $v_1^* = 0.967$ .

For  $N = 3$ , the integration interval is divided into  $M = 3$  sub-intervals which yields two unknowns as  $\mathbf{v} = [v_1, v_2]$ . Thus, the  $Q$ -function's approximation is written as

$$\tilde{Q}(x) = \frac{v_1^*}{\pi} \exp\left(\frac{-x^2}{2 \sin^2(\frac{v_1^*}{2})}\right) + \frac{v_2^* - v_1^*}{\pi} \exp\left(\frac{-x^2}{2 \sin^2(\frac{v_2^* + v_1^*}{2})}\right) + \left(\frac{1}{2} - \frac{v_2^*}{\pi}\right) \exp\left(\frac{-x^2}{2 \sin^2(\frac{\pi}{4} + \frac{v_2^*}{2})}\right). \quad (15)$$

The optimum parameters are  $v_1^* = 0.5571$  and  $v_2^* = 1.0702$ .

In Fig. 2, we compare the accuracy of the optimized single-point Legendre rule with that of the trapezoidal rule, i.e., composite two-point closed Newton–Cotes rule. In particular, our optimized approximation outperforms that of Chiani *et al.* for  $N = 2$  over most of the  $x$ -range, and over the whole  $x$ -range for  $N = 3$  where they have the least relative error. The numerical approximations with  $N = 3$  are tighter than those with  $N = 2$ . The integral of the absolute relative error is also calculated and plotted in Fig. 3 for the optimized and un-optimized trapezoidal and single-point Legendre rules up to  $N = 10$ . As expected, the optimized approximation has a better total accuracy. In addition, it can be concluded from the decaying curves that, as the number of exponential terms increases, the accuracy of the numerical method increases.

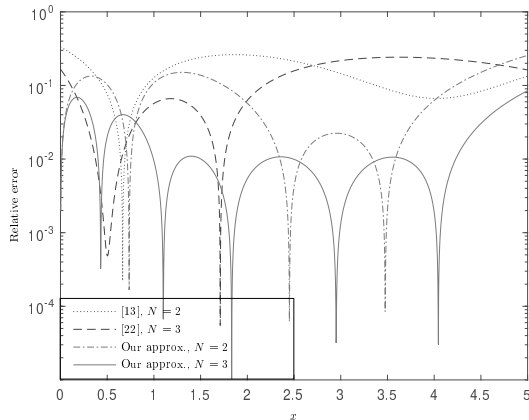


Fig. 2. Comparison between the optimized trapezoidal rule and the optimized Legendre rule.

#### IV. CONCLUSION

This paper provided a mathematical study on the different numerical integration techniques that can be applied on the Craig's form of the Gaussian  $Q$ -function to produce exponential approximations for it. We presented explicit expressions for the corresponding coefficients of all the possible approximations of composite and non-composite Newton-Cotes and Gaussian quadrature rules. We also contributed the optimization of the numerical rules to increase their tightness by adopting Quasi-Newton optimization method. The coefficients of the exponential approximation with two and three terms are reported based on the optimized composite Legendre rule.

#### ACKNOWLEDGMENT

This research work was funded in part by the Academy of Finland under the grant 326448 "Generalized Fading Distributions and Matrix Functions for the Analysis of Wireless Communication Systems."

#### REFERENCES

- [1] W. Cody, "Rational Chebyshev approximations for the error function," *Math. Comp.*, vol. 23, no. 107, pp. 631–637, Jul. 1969.
- [2] P. Börjesson and C. Sundberg, "Simple approximations of the error function  $Q(x)$  for communications applications," *IEEE Trans. Commun.*, vol. 27, no. 3, pp. 639–643, Mar. 1979.
- [3] G. Karagiannidis and A. Lioumpas, "An improved approximation for the Gaussian  $Q$ -function," *IEEE Commun. Lett.*, vol. 11, no. 8, pp. 644–646, Aug. 2007.
- [4] J. Dyer and S. Dyer, "Corrections to, and comments on, An improved approximation for the Gaussian  $Q$ -function," *IEEE Commun. Lett.*, vol. 12, no. 4, p. 231, Apr. 2008.
- [5] W. Jang, "A simple upper bound of the Gaussian  $Q$ -function with closed-form error bound," *IEEE Commun. Lett.*, vol. 15, no. 2, pp. 157–159, Feb. 2011.
- [6] Y. Isukapalli and B. Rao, "An analytically tractable approximation for the Gaussian  $Q$ -function," *IEEE Commun. Lett.*, vol. 12, no. 9, pp. 669–671, Sep. 2008.
- [7] C. Tellambura and A. Annamalai, "Efficient computation of  $\text{erfc}(x)$  for large arguments," *IEEE Trans. Commun.*, vol. 48, no. 4, pp. 529–532, Apr. 2000.

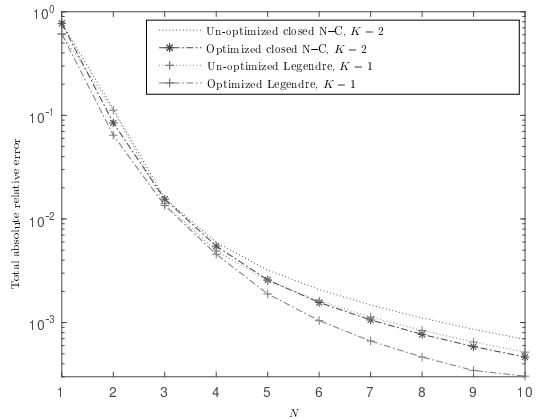


Fig. 3. The total absolute relative error of the optimized and un-optimized two-point closed Newton-Cotes approximation and single-point Legendre approximation for  $N = 1, 2, \dots, 10$ .

- [8] Y. Chen and N. Beaulieu, "A simple polynomial approximation to the Gaussian  $Q$ -function and its application," *IEEE Commun. Lett.*, vol. 13, no. 2, pp. 124–126, Feb. 2009.
- [9] G. Abreu, "Jensen-Cotes upper and lower bounds on the Gaussian  $Q$ -function and related functions," *IEEE Trans. Commun.*, vol. 57, no. 11, pp. 3328–3338, Nov. 2009.
- [10] G. Abreu, "Very simple tight bounds on the  $Q$ -function," *IEEE Trans. Commun.*, vol. 60, no. 9, pp. 2415–2420, Sep. 2012.
- [11] M. López-Benítez and F. Casadevall, "Versatile, accurate, and analytically tractable approximation for the Gaussian  $Q$ -function," *IEEE Trans. Commun.*, vol. 59, no. 4, pp. 917–922, Apr. 2011.
- [12] Q. Shi, "Novel approximation for the Gaussian  $Q$ -function and related applications," in *Proc. 22nd IEEE PIMRC*, Sep. 2011, pp. 2030–2034.
- [13] M. Chiani, D. Dardari, and M. Simon, "New exponential bounds and approximations for the computation of error probability in fading channels," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 840–845, Jul. 2003.
- [14] I. M. Tanash and T. Riihonen, "Global minimax approximations and bounds for the Gaussian  $Q$ -function by sums of exponentials," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6514–6524, Oct. 2020.
- [15] D. Sadhwani, R. Yadav, and S. Aggarwal, "Tighter bounds on the Gaussian  $Q$  function and its application in Nakagami- $m$  fading channel," *IEEE Wireless Commun. Lett.*, vol. 6, no. 5, pp. 574–577, Oct. 2017.
- [16] S. Chang, P. Cosman, and L. Milstein, "Chernoff-type bounds for the Gaussian error function," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 2939–2944, Nov. 2011.
- [17] P. Loskot and N. Beaulieu, "Prony and polynomial approximations for evaluation of the average probability of error over slow-fading channels," *IEEE Trans. Veh. Technol.*, vol. 58, no. 3, pp. 1269–1280, Mar. 2009.
- [18] O. Olabiyyi and A. Annamalai, "Invertible exponential-type approximations for the Gaussian probability integral  $Q(x)$  with applications," *IEEE Wireless Commun. Lett.*, vol. 1, no. 5, pp. 544–547, Oct. 2012.
- [19] M. Wu, Y. Li, M. Gurusamy, and P. Kam, "A tight lower bound on the Gaussian  $Q$ -function with a simple inversion algorithm, and an application to coherent optical communications," *IEEE Commun. Lett.*, vol. 22, no. 7, pp. 1358–1361, Jul. 2018.
- [20] P. Davis and P. Rabinowitz, *Methods of Numerical Integration*, 2nd ed. Academic Press, 1984.
- [21] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed. Dover Publications, Inc., 1965.
- [22] Q. Zhang, J. Cheng, and G. Karagiannidis, "Block error rate of optical wireless communication systems over atmospheric turbulence channels," *IET Commun.*, vol. 8, no. 5, pp. 616–625, March 2014.





# PUBLICATION

4

**Improved coefficients for the Karagiannidis–Lioumpas approximations and bounds to the Gaussian  $Q$ -function**

I. M. Tanash and T. Riihonen



*IEEE Communications Letters*, vol. 25, no. 5, pp. 1468–1471

DOI: 10.1109/LCOMM.2021.3052257

**Publication reprinted with the permission of the copyright holders.**



# Improved Coefficients for the Karagiannidis–Lioumpas Approximations and Bounds to the Gaussian $Q$ -Function

Islam M. Tanash  and Taneli Riihonen , *Member, IEEE*

**Abstract**—We revisit the Karagiannidis–Lioumpas (KL) approximation of the  $Q$ -function by optimizing its coefficients in terms of absolute error, relative error and total error. For minimizing the maximum absolute/relative error, we describe the targeted uniform error functions by sets of nonlinear equations so that the optimized coefficients are the solutions thereof. The total error is minimized with numerical search. We also introduce an extra coefficient in the KL approximation to achieve significantly tighter absolute and total error at the expense of unbounded relative error. Furthermore, we extend the KL expression to lower and upper bounds with optimized coefficients that minimize the error measures in the same way as for the approximations.

**Index Terms**—Communication theory, error probability.

## I. INTRODUCTION

KARAGIANNIDIS AND LIOUMPAS presented in [1] a relatively tight, yet analytically tractable, approximation for the Gaussian  $Q$ -function [2] as follows:

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{1}{2}t^2\right) dt \quad (1)$$

$$\approx a \exp\left(-bx^2\right) \cdot \frac{1 - \exp(-cx)}{x} \triangleq \tilde{Q}(x)$$

for which their original study sets  $(a, b, c) = (\frac{1}{B\sqrt{2\pi}}, \frac{1}{2}, \frac{A}{\sqrt{2}})$  and proposes for error minimization example coefficient values  $A = 1.98$  and  $B = 1.135$  rendering  $(a, c) \approx (0.3515, 1.4001)$ .

Despite drawing some criticism [3] shortly after publication, the ‘Karagiannidis–Lioumpas (KL) approximation’ has gradually established itself as one of the most usable substitutes for the Gaussian  $Q$ -function in communication theory problems and the paper [1] has received a large number of citations; it is only fitting to begin calling the expression after its inventors.

A diverse set of applications for the KL approximation can be found in [4]–[8] to name but a few prominent articles. In general, the approximation is often used in the calculation of average bit or symbol error probability as a tractable replacement for the Gaussian  $Q$ -function such that analysis can be carried out and completed in a closed form at the cost of making results tight approximations instead of exact ones. This usually involves integrating something like, e.g.,  $f(\tilde{Q}(x(y)))$ , where even simple functions  $f(q)$  and  $x(y)$ , which are derived from the communication system under study, may forbid exact

analysis using the actual  $Q$ -function [9]–[11]. One should note especially that the approximation is always used in an intermediate step of analytical derivations and it is not meant for numerical probability computations per se — instead, rational Chebyshev functions [12] are perfect to that end.

This Letter is inspired by the fact that the original study [1] presents explicit values of  $a$  and  $c$  for *only one approximation* (which has low integrated total error when  $b = \frac{1}{2}$ , to be exact). However, the KL approximation family is actually much more versatile, where new coefficients can be acquired in terms of other criteria for better accuracy depending on the application. The KL expression can be also repurposed to achieve lower and upper bounds (that are also tight approximations) and, in certain cases, coefficients admit explicit values. Furthermore, by introducing the extra coefficient  $b$  in (1) that originally was  $b = \frac{1}{2}$  and permitting  $b < \frac{1}{2}$ , we achieve significantly improved accuracy in terms of absolute and total error.

The objective of this Letter is to apply the KL expression of the  $Q$ -function to derive improved approximations and bounds which are global and tight over  $x \geq 0$  by optimizing the coefficients  $(a, b, c)$  in respect to their minimum global absolute or relative error or minimum integrated total error. Like [11] for another popular expression [9], we present new formulation that minimizes the maximum global error of (1) by constructing a set of equations, which describes the corresponding error function, and solve them numerically to find the optimized coefficients. The total error is optimized with exhaustive search for reference. In general, when optimizing one of the three criteria, better performance will be achieved at the expense of decreased accuracy in terms of the others.

The new coefficients solved herein are applicable as one-to-one replacements for the original ones of [1] adopted into the analysis of [4]–[8] and many other studies. Literature is rich in approximations/bounds for the  $Q$ -function and, typically, the application’s mathematics define, which ones are tractable for it. Whenever (1) is preferred, our coefficients offer variety to tailor accuracy for the application or to use bounds.

## II. PRELIMINARIES

The case  $x \geq 0$  is presumed throughout this Letter with little loss of generality because the relation  $Q(x) = 1 - Q(-x)$  extends all the considered functions to the negative real axis. In fact, this is the main motive for optimizing approximations and bounds also subject to an additional constraint  $\tilde{Q}(0) = \frac{1}{2}$  that makes their extensions continuous at the origin like  $Q(x)$ .

This study solves optimized approximations and bounds for three criteria and for combinations thereof, viz.  $\min_{(a,b,c)} d_{\max}$  (‘minimax absolute error’),  $\min_{(a,b,c)} r_{\max}$  (‘minimax relative

Manuscript received November 19, 2020; revised December 18, 2020 and January 8, 2021; accepted January 9, 2021. Date of publication January ??, 2021; date of current version January 17, 2022. This work was partially supported by the Academy of Finland under Grant 326448. The associate editor coordinating the review of this letter and approving it for publication was A.-A. A. Boulogeorgos. (Corresponding author: Islam M. Tanash.)

The authors are with Tampere University, Tampere 33720, Finland (e-mail: islam.tanash@tuni.fi; taneli.riihonen@tuni.fi).

Digital Object Identifier 10.1109/LCOMM.???????????

error') and  $\min_{(a,b,c)} d_{\text{tot}}$  ('integrated total error' [1]), where  $d_{\text{max}} \triangleq \max_{x \geq 0} |d(x)|$ ,  $r_{\text{max}} \triangleq \max_{x \geq 0} |r(x)|$ ,  $d_{\text{tot}} \triangleq \int_0^{\infty} |d(x)| dx$ , and the error functions are defined as

$$d(x) \triangleq \tilde{Q}(x) - Q(x), \quad (2)$$

$$r(x) \triangleq \frac{d(x)}{Q(x)} = \frac{\tilde{Q}(x)}{Q(x)} - 1. \quad (3)$$

For baseline reference, the coefficients originally given in [1] render  $d_{\text{max}} \approx 0.00789$ ,  $r_{\text{max}} \approx 0.119$ , and  $d_{\text{tot}} \approx 0.00385$ .

As implied above, the presented approximations and bounds will be global ones, i.e., tight over the whole non-negative real axis (for all  $x \geq 0$ ). The error functions converge to explicit values, which may be local extrema, at both ends of this range:

$$\lim_{x \rightarrow 0} d(x) = ac - \frac{1}{2}, \quad \lim_{x \rightarrow 0} r(x) = 2ac - 1, \quad (4)$$

$$\lim_{x \rightarrow \infty} d(x) = 0, \quad \lim_{x \rightarrow \infty} r(x) = \begin{cases} \infty, & \text{if } b < \frac{1}{2}, \\ a\sqrt{2\pi} - 1, & \text{if } b = \frac{1}{2}, \\ -1, & \text{if } b > \frac{1}{2}. \end{cases}$$

The last limit shows especially that global approximations and bounds in terms of relative error exist if and only if we set  $b = \frac{1}{2}$ . However, as a novel fact, our study demonstrates that absolute error and total error can be instead significantly reduced by permitting  $b < \frac{1}{2}$ . Therefore, two scenarios of approximations for the absolute and total error are considered in this Letter, i.e., approximations with  $b = \frac{1}{2}$  or  $b < \frac{1}{2}$ .

Local error extrema may occur also at critical points, where the derivatives of the continuous error functions vanish. Denoting differentiation with an apostrophe, they are given by  $d'(x) = \tilde{Q}'(x) - Q'(x)$ ,  $r'(x) = \frac{\tilde{Q}'(x)Q(x) - \tilde{Q}(x)Q'(x)}{[Q(x)]^2}$ , where

$$\tilde{Q}'(x) = -\frac{a \left( (2bx^2 + 1)(e^{cx} - 1) - cx \right) e^{-bx^2 - cx}}{x^2}, \quad (5)$$

$$Q'(x) = -\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right). \quad (6)$$

Two variations of approximations are considered herein:  $d(0) = r(0) = 0$  and  $d(0) = -d_{\text{max}}$  (resp.  $r(0) = -r_{\text{max}}$ ). The former case maintains the continuity of the  $Q$ -function when extending to  $x < 0$  and results in  $c = \frac{1}{2a}$ , when substituted in  $\lim_{x \rightarrow 0} d(x)$  (resp.  $\lim_{x \rightarrow 0} r(x)$ ) that is given in (4). The latter case provides slightly better accuracy at the cost of discontinuity occurring at  $x = 0$  and results in  $c = \sqrt{\frac{\pi}{2}}$  in the cases of relative error, by solving  $\lim_{x \rightarrow 0} r(x) = -r_{\text{max}}$  with  $\lim_{x \rightarrow \infty} r(x) = -r_{\text{max}}$  that are defined in (4).

### III. ALTERNATIVE IMPROVED COEFFICIENTS FOR (1)

In this section, we describe the methodologies to solve the new coefficients  $(a, b, c)$  for the KL expression. They are optimized either in the minimax sense or in terms of the integrated total error to yield an approximation, an upper bound or a lower bound. All the 17 thus-obtained improved/alternative coefficient sets and accuracy thereof are listed in Table I.

#### A. Global Uniform Approximations and Bounds

The minimax optimization problems are solved in terms of both absolute and relative errors defined in (2) and (3), respectively, by constructing a set of nonlinear equations. This set describes the resulting error function, which should be uniform with equal values for all the extrema points. Each extremum point yields two equations, where one expresses its value and the other sets the derivative of the error function to zero at that point. In addition, one equation (for  $d(x)$ ) or two equations (for  $r(x)$ ) is/are obtained from evaluating the limits at the two endpoints of the considered range,  $[0, \infty]$ , per (4).

The resulting sets of equations, which have equal number of equations and unknowns, can be solved straightforwardly by any numerical tool for the considered variations to find the optimized sets of coefficients that satisfy  $\min_{(a,b,c)} d_{\text{max}}$  for the absolute error and  $\min_{(a,b,c)} r_{\text{max}}$  for the relative error. We used iteratively random initial guesses for the unknowns in this approach, namely  $(a, b, c)$ ,  $d_{\text{max}}$  or  $r_{\text{max}}$ , and the location of the extrema ( $x_k$ ), until `fsolve` in Matlab converged to the solution, which is confirmed by substitution. The formulations for the minimax approximations/bounds are described below.

1) *Approximations in Terms of Absolute Error:* The coefficients  $(a, b, c)$  are optimized for approximations in terms of the absolute error by formulating a set of equations as

$$\begin{cases} d'(x_k) = 0, & \text{for } k = 1, 2 \text{ or } 1, 2, 3, \\ d(x_k) = (-1)^{k+1} d_{\text{max}}, & \text{for } k = 1, 2 \text{ or } 1, 2, 3, \\ \begin{cases} ac = \frac{1}{2}, & \text{when } d(0) = 0, \\ ac = \frac{1}{2} - d_{\text{max}}, & \text{when } d(0) = -d_{\text{max}}, \end{cases} \end{cases} \quad (7)$$

where  $x_k$  is an extremum point. The number of the error function's extrema depends on the value of  $b$ ; if  $b$  is fixed to  $\frac{1}{2}$ , then we have two extrema, whereas if  $b$  is allowed to be any positive value, then we need three separate extrema.

2) *Lower Bounds in Terms of Absolute Error:* For the lower bounds, we need to find the optimized coefficients which minimize the global absolute error for  $d(x) \leq 0$  when  $x \geq 0$ . The value of  $b$  must always equal to  $\frac{1}{2}$ . The tightest resulting uniform error function will start from  $d(0) = -d_{\text{max}}$ , with its maximum equal to zero and its minimum equal to  $-d_{\text{max}}$  so that we can formulate a set of equations as

$$\begin{cases} d'(x_1) = d'(x_2) = 0, \\ d(x_1) = 0, d(x_2) = -d_{\text{max}}, \\ ac = \frac{1}{2} - d_{\text{max}}. \end{cases} \quad (8)$$

When  $d(0) = 0$ , we get  $a = \sqrt{\frac{\pi}{32}}$  and  $c = \sqrt{\frac{8}{\pi}}$  by imposing  $d'(0) = 0$  (only in this case), which produces  $ac^2 = \sqrt{2/\pi}$ , and solving with  $c = \frac{1}{2a}$  that results from setting  $d(0) = 0$ .

3) *Upper Bounds in Terms of Absolute Error:* The set of equations becomes

$$\begin{cases} d'(x_1) = d'(x_2) = d'(x_3) = 0, \\ d(x_1) = d(x_3) = d_{\text{max}}, d(x_2) = 0, \\ ac = \frac{1}{2}. \end{cases} \quad (9)$$

In particular, we shape the uniform error function to have three extrema with  $d(x) \geq 0$  when  $x \geq 0$  in which its maxima

are equal to  $d_{\max}$  and its minimum is equal to zero. The corresponding error function must always start from  $d(0) = 0$ .

4) *Approximations in Terms of Relative Error*: The targeted uniform error function in terms of the relative error consists of only one maximum point and converges to  $-r_{\max}$  as  $x$  tends to infinity, which results in  $-r_{\max} = a\sqrt{2\pi} - 1$  according to (4). Therefore, we can formulate the set of equations as

$$\begin{cases} r'(x_1) = 0, r(x_1) = r_{\max}, \\ \begin{cases} ac = \frac{1}{2}, & \text{when } r(0) = 0, \\ ac = \frac{1-r_{\max}}{2}, & \text{when } r(0) = -r_{\max}, \end{cases} \\ a = \frac{1-r_{\max}}{\sqrt{2\pi}}. \end{cases} \quad (10)$$

5) *Lower Bounds in Terms of Relative Error*: We need to find the optimized coefficients,  $a$  and  $c$ , in the minimax sense for  $r(x) \leq 0$  when  $x \geq 0$  which converges to  $-r_{\max}$  as  $x$  tends to infinity. The resulting error function can either start from  $r(0) = -r_{\max}$  to formulate a set of equations as

$$\begin{cases} r'(x_1) = r(x_1) = 0, \\ ac = \frac{1-r_{\max}}{2}, a = \frac{1-r_{\max}}{\sqrt{2\pi}}, \end{cases} \quad (11)$$

or from  $r(0) = 0$  yielding  $a = \sqrt{\frac{\pi}{32}}$  and  $c = \sqrt{\frac{8}{\pi}}$  like with the corresponding lower bound in terms of absolute error.

6) *Upper Bound in Terms of Relative Error*: We must ensure that  $r(x) \geq 0$  when  $x \geq 0$  for the uniform error function. The resulting error function has only one maximum point and converges to zero as  $x$  tends to infinity. Therefore,  $a = \frac{1}{\sqrt{2\pi}}$  and  $c = \sqrt{\frac{\pi}{2}}$  as proposed earlier in [13] and  $b$  is known to be equal to  $\frac{1}{2}$ . The optimized upper bound in terms of relative error is also optimal in terms of absolute error and integrated total error for the case where  $b = \frac{1}{2}$ .

### B. Numerical Optimization in Terms of Total Error

Instead of defining  $d_{\text{tot}} \triangleq \int_0^R |d(x)| dx$  like in [1] and so making optimized coefficients specific to the value chosen for  $R$  and limited to the range  $[0, R]$ , we measure total error with  $R \rightarrow \infty$  and obtain globally optimized approximations and bounds. In particular, we optimized the coefficients for the two variations of the approximations with or without setting  $b = \frac{1}{2}$  by performing an extensive search, where we evaluated the target metric ( $d_{\text{tot}}$ ) over wide one/two/three-dimensional grids for the unknowns  $a$ ,  $(a, b)$ ,  $(a, c)$ , or  $(a, b, c)$  with granularity of 0.000001 and selected the grid point with the minimum total error for each variation. This renders four sets of optimized coefficients. Extra constraint checks guarantee  $d(x) < 0$  for the lower bound and  $d(x) > 0$  for the upper bound.

## IV. NUMERICAL RESULTS AND CONCLUSIONS

We summarize the improved coefficients for the minimax approximations and bounds and for the total absolute error in Table I and illustrate their error functions in Fig. 1, together with the original KL approximation from [1] and reference approximations and bounds from [9] and [10].<sup>1</sup>

<sup>1</sup>The labels having the form  $Xy-n$  in the results refer to the approximations and bounds as follows:  $X$  is U for upper bounds, A for approximations, and L for lower bounds; whereas  $y$  is d for absolute error, r for relative error, and t for total error; in addition,  $n$  refers to rank of the coefficients according to the accuracy of the absolute error of each variation in an ascending order.

TABLE I  
NEW COEFFICIENTS FOR (1) AND APPROXIMATION ERROR THEREOF

#*	type	a	b	c	$d_{\max}$	$r_{\max}$	$d_{\text{tot}}$
[1]	$\tilde{Q}(x) \approx Q(x)$	0.351491	1/2	1.400071	0.007887	0.1189	0.003847
Ud-1	$\tilde{Q}(x) \geq Q(x), \tilde{Q}(0) = Q(0)$	0.320848	0.467551	1/(2a)	0.000894	$\infty$	0.001638
U-2	$\tilde{Q}(x) \geq Q(x), \tilde{Q}(0) = Q(0)$	1/\sqrt{2\pi}	1/2	\sqrt{\pi/2}	0.019413	0.0953	0.023034
Ad-1	$\tilde{Q}(x) \approx Q(x)$	0.321272	0.471452	1.554646	0.000536	$\infty$	0.001130
Ad-2	$\tilde{Q}(x) \approx Q(x), \tilde{Q}(0) = Q(0)$	0.319695	0.469381	1/(2a)	0.000632	$\infty$	0.001330
Ad-3	$\tilde{Q}(x) \approx Q(x)$	0.335419	1/2	1.484436	0.002092	0.1592	0.003505
Ad-4	$\tilde{Q}(x) \approx Q(x), \tilde{Q}(0) = Q(0)$	0.332106	1/2	1/(2a)	0.002568	0.1675	0.004272
Ar-6	$\tilde{Q}(x) \approx Q(x)$	0.380797	1/2	\sqrt{\pi/2}	0.022742	0.0455	0.010439
Ar-5	$\tilde{Q}(x) \approx Q(x), \tilde{Q}(0) = Q(0)$	0.376056	1/2	1/(2a)	0.013787	0.0574	0.015096
Ld-1	$\tilde{Q}(x) \leq Q(x)$	0.329783	1/2	1.506303	0.003247	0.1734	0.004659
L-2	$\tilde{Q}(x) \leq Q(x), \tilde{Q}(0) = Q(0)$	\sqrt{\pi/32}	1/2	\sqrt{8/\pi}	0.007148	0.2146	0.010188
Lr-3	$\tilde{Q}(x) \leq Q(x)$	0.364230	1/2	\sqrt{\pi/2}	0.043505	0.0870	0.013683
Ur-1	$\tilde{Q}(x) \geq Q(x), \tilde{Q}(0) = Q(0)$	0.323300	0.472329	1/(2a)	0.001326	$\infty$	0.001454
Ar-2	$\tilde{Q}(x) \approx Q(x)$	0.326530	0.477951	1.523737	0.002454	$\infty$	0.000877
Ar-1	$\tilde{Q}(x) \approx Q(x), \tilde{Q}(0) = Q(0)$	0.322612	0.474260	1/(2a)	0.001126	$\infty$	0.001185
Ar-4	$\tilde{Q}(x) \approx Q(x)$	0.342771	1/2	1.437908	0.007127	0.1408	0.002881
Ar-3	$\tilde{Q}(x) \approx Q(x), \tilde{Q}(0) = Q(0)$	0.336219	1/2	1/(2a)	0.003519	0.1572	0.004058
Lr-1	$\tilde{Q}(x) \leq Q(x)$	0.339602	1/2	1.445957	0.008950	0.1505	0.003602

\*notes: U-2=Ud-2=Ur-2=U-2 [21], L-2=Ld-2=Lr-2=L-2, underlining indicates the error metric(s) that is/are minimized

The numerical results show that the improved coefficients of the proposed KL approximations and bounds are optimal subject to their optimization targets, yet expressed precisely in implicit form as solutions to systems of nonlinear equations as opposed to relying on numerical search to minimize error measures. In some specific cases, a part or even all of the three coefficients can be expressed as explicit constants. The best approximation/bound from Table I for a specific application is chosen by contrasting requirements against Fig. 1, provided that the KL expression (1) is suitable for it to begin with.

As an ultimate conclusion, the presented data suggests good alternatives to the original coefficients given in [1] for the case of  $b = \frac{1}{2}$ : In some applications, the accuracy of the KL approximation might be improved by choosing instead  $A = 1.95$ ,  $B = 1.113$  (a compromise between all  $Ay-n$ ) for decreasing both absolute and relative error by round 15% at the cost of increasing total error by round 65%; or  $A = 2.03$ ,  $B = 1.162$  (At-4) for decreasing absolute error and total error by round 10% and 25%, respectively, at the cost of increasing relative error by round 15%. Sometimes it may also be useful to choose  $A = B\sqrt{\pi} \approx 1.88$ ,  $B = 1.061$  (Ar-5) for minimizing relative error (with round 50% reduction) subject to zero error at the origin. In contrast, when primarily minimizing absolute error, accuracy can be improved significantly by generalizing the KL approximation to allow any positive  $b$ : Namely, the choice  $a = 0.32$ ,  $b = 0.4703$ ,  $c = 1.5625$  (Ad-5) guarantees zero error at the origin while decreasing absolute error and total error as much as round 90% and 65%, respectively, at the cost of making relative error unbounded for large arguments.

## REFERENCES

- [1] G. K. Karagiannidis and A. S. Lioumpas, "An improved approximation for the Gaussian  $Q$ -function," *IEEE Commun. Lett.*, vol. 11, no. 8, pp. 644–646, Aug. 2007.
- [2] S. Aggarwal, "A survey-cum-tutorial on approximations to Gaussian  $Q$  function for symbol error probability analysis over Nakagami- $m$  fading channels," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2195–2223, Jul.–Sep. 2019.
- [3] J. Dyer and S. Dyer, "Corrections to, and comments on, "An improved approximation for the Gaussian  $Q$ -function"," *IEEE Commun. Lett.*, vol. 12, no. 4, p. 231, Apr. 2008.

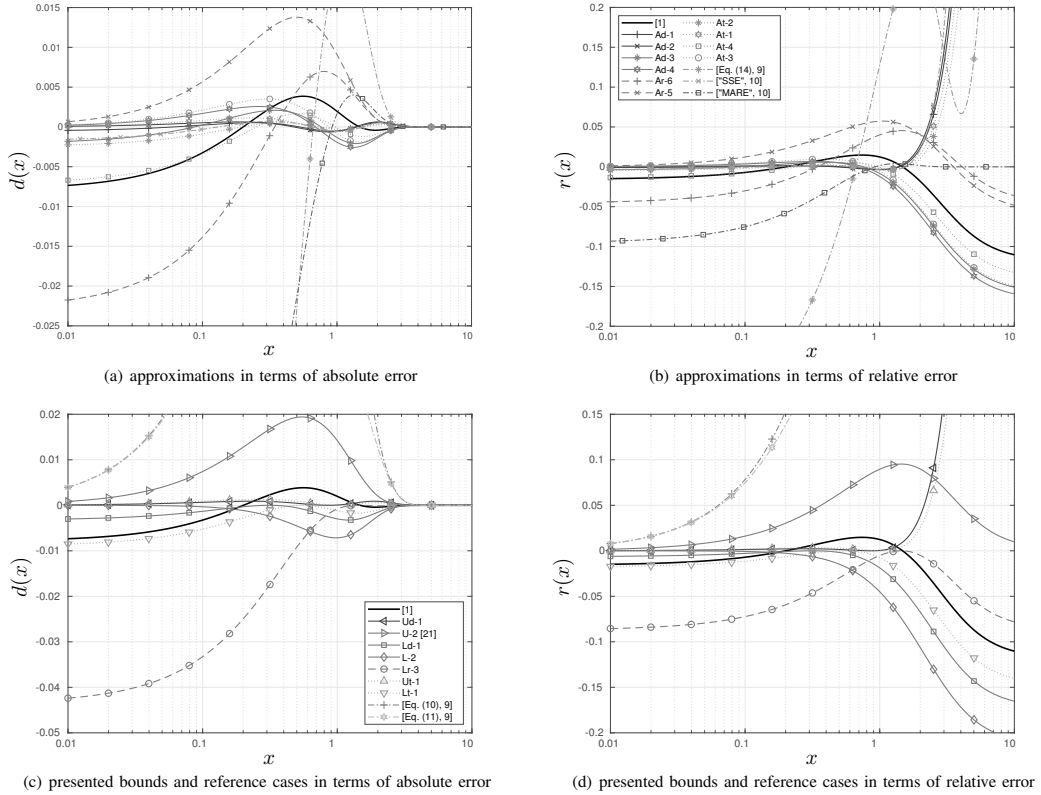


Fig. 1. The improved approximations and bounds compared to the KL approximation with the original coefficients [1] and to expressions from [9] and [10].

- [4] C. Potter, G. Venayagamoorthy, and K. Kosbar, "RNN based MIMO channel prediction," *Signal Process.*, vol. 90, no. 2, pp. 440–450, Feb. 2010.
- [5] L. Tan and L. Le, "Distributed MAC protocol for cognitive radio networks: Design, analysis, and optimization," *IEEE Trans. Veh. Technol.*, vol. 60, no. 8, pp. 3990–4003, Oct. 2011.
- [6] J. Wu *et al.*, "Unified spectral efficiency analysis of cellular systems with channel-aware schedulers," *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3463–3474, Dec. 2011.
- [7] D. Malak, M. Al-Shalash, and J. Andrews, "Optimizing content caching to maximize the density of successful receptions in device-to-device networking," *IEEE Trans. Commun.*, vol. 64, no. 10, pp. 4365–4380, Oct. 2016.
- [8] S. Lin *et al.*, "Rayleigh fading suppression in one-dimensional optical scatters," *IEEE Access*, vol. 7, pp. 17 125–17 132, Jan. 2019.
- [9] M. Chiani, D. Dardari, and M. K. Simon, "New exponential bounds and approximations for the computation of error probability in fading channels," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 840–845, Jul. 2003.
- [10] M. López-Benítez and F. Casadevall, "Versatile, accurate, and analytically tractable approximation for the Gaussian  $Q$ -function," *IEEE Trans. Commun.*, vol. 59, no. 4, pp. 917–922, Apr. 2011.
- [11] I. M. Tanash and T. Riihonen, "Global minimax approximations and bounds for the Gaussian  $Q$ -function by sums of exponentials," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6514–6524, Oct. 2020.
- [12] W. Cody, "Rational Chebyshev approximations for the error function," *Math. Comp.*, vol. 23, no. 107, pp. 631–637, Jul. 1969.
- [13] W. M. Jang, "A simple upper bound of the Gaussian  $Q$ -function with closed-form error bound," *IEEE Commun. Lett.*, vol. 15, no. 2, pp. 157–159, Feb. 2011.

# PUBLICATION

5

**Generalized Karagiannidis–Lioumpas approximations and bounds  
to the Gaussian  $Q$ -function with optimized coefficients**

I. M. Tanash and T. Riihonen

*IEEE Communications Letters*, vol. 26, no. 3, pp. 513–517



DOI: 10.1109/LCOMM.2021.3139372

**Publication reprinted with the permission of the copyright holders.**





# Generalized Karagiannidis–Lioumpas Approximations and Bounds to the Gaussian $Q$ -Function with Optimized Coefficients

Islam M. Tanash  and Taneli Riihonen , *Member, IEEE*

**Abstract**—We develop extremely tight novel approximations, lower bounds and upper bounds for the Gaussian  $Q$ -function and offer multiple alternatives for the coefficient sets thereof, which are optimized in terms of the four most relevant criteria: minimax absolute/relative error and total absolute/relative error. To minimize error maximum, we modify the classic Remez algorithm to comply with the challenging nonlinearity that pertains to the proposed expression for approximations and bounds. On the other hand, we minimize the total error numerically using the quasi-Newton algorithm. The proposed approximations and bounds are so well matching to the actual  $Q$ -function that they can be regarded as virtually exact in many applications since absolute and relative errors of  $10^{-9}$  and  $10^{-5}$ , respectively, are reached with only ten terms. The significant advance in accuracy is shown by numerical comparisons with key reference cases.

**Index Terms**—Gaussian  $Q$ -function, error probability.

## I. INTRODUCTION

THE Gaussian  $Q$ -function and the related complementary error function  $\text{erfc}(\cdot)$  are very important entities for communication theory (as well as in statistical sciences at large). They emerge often when noise, interference, or a signal is characterized by the normal distribution. Although the  $Q$ -function, which has no exact closed form, can be evaluated using many software packages, the literature is rich in several approximations and bounds [1]–[10] based on either the statistical definition [5, Eq. 1] or on the alternative representation proposed by Craig [11]. Their significant value is in facilitating closed-form calculations of error probabilities for different digital modulations and fading models [12]–[14], in which functions of  $Q$ -function usually appear in integrands.

The expression by Karagiannidis and Lioumpas in [1] is one of the most common tools to approximate the  $Q$ -function in the different problems of communication theory due to its tractability and accuracy compared to others. In particular, they approximate  $\text{erfc}(\cdot)$  by an inverse factorial series which is then truncated to a single term but the resulted expression is loose for small arguments. Therefore, they multiply it by a monotonically increasing function to tighten it there and, thus, to approximate accurately the  $Q$ -function for all  $x \geq 0$  as

$$Q(x) \approx a(1 - \exp(-cx)) \cdot \frac{\exp(-bx^2)}{x}, \quad (1)$$

where  $a = \frac{1}{1.135\sqrt{2\pi}}$ ,  $b = \frac{1}{2}$ , and  $c = \frac{1.98}{\sqrt{2}}$  originally, while [2] presents alternative coefficients for tailoring accuracy in different applications or transforming it into a bound.

Manuscript received September 3, 2021; revised November 29, 2021 and December 22, 2021; accepted December 23, 2021. Date of publication Month DD, 20YY; date of current version January 17, 2022. The associate editor coordinating the review of this letter and approving it for publication was A.-A. Boulogeorgos. (Corresponding author: Islam M. Tanash.)

The authors are with Tampere University, 33720 Tampere, Finland (e-mail: islam.tanash@tuni.fi; taneli.riihonen@tuni.fi).

Digital Object Identifier 10.1109/LCOMM.???????????

Inspired by the Karagiannidis–Lioumpas (KL) approximation, our first main contribution is to propose a new expression to approximate or bound the  $Q$ -function:

$$\tilde{Q}(x) \triangleq \underbrace{\frac{1 - \exp(-cx)}{x}}_{\triangleq g(x)} \cdot \underbrace{\sum_{n=1}^N a_n \exp(-b_n x^2)}_{\triangleq h(x)}, \quad (2)$$

which is referred to as the generalized KL (GKL) expression since it is reduced to the original KL expression in the special case of  $N = 1$  [1], [2] (but is novel herein for  $N > 1$ ). Conceptually, an approach analogous to that in [1] is used by first approximating the  $Q$ -function with the sum of exponentials  $h(x)$  as in [3, Eq. 8], which results in unbounded relative error and thus lower accuracy for the higher arguments; then it is multiplied by the term  $g(x)$  to bound the relative error with  $b_1 \triangleq \min\{b_n\}_{n=1}^N = \frac{1}{2}$ . This yields the accurate GKL expression in (2) that is limited to the domain  $x \geq 0$ , not so unlike most related approximations, but the relation  $Q(x) = 1 - Q(-x)$  extends it to  $x < 0$ .

As the second main contribution, we solve the research problem of optimizing the coefficients,  $\{(a_n, b_n)\}_{n=1}^N$  and  $c$ , in order to minimize the global or total absolute/relative error of the corresponding approximation. Furthermore, the coefficients are optimized in the minimax sense to derive tight lower and upper bounds too. We show that the GKL approximations/bounds together with the optimized coefficients achieve very high, increasing accuracy so that using not-so-large number of terms they can become virtually exact, i.e., the error may not be notable in many applications in communications systems' analysis. By these main contributions, we provide researchers with accuracy-controllable approximations/bounds in terms of several optimization criteria, from which they can choose one that best suits their needs in order to ease expression manipulations with extremely high accuracy.

Two types of complexity are of relevance herein, namely the analytical and the computational. The former, which refers to the difficulty of the analytical form of (2) and the tractability thereof in symbolic calculations for mathematical operations, is kept the same as for (1) while significantly increasing the accuracy. On the other hand, the latter can refer herein either to the difficulty and processing time of the proposed optimization methodology or to those in using the approximation. The offline complexity of coefficient optimization is hardly relevant since it is already implemented by us and the coefficients are released to public domain<sup>1</sup> so that there is no need for redoing it later, whereas the online complexity of using the GKL expression is directly proportional to the number of terms  $N$  used in the approximation. Hence, (2) with the optimized coefficients can reliably substitute the  $Q$ -function in derivations of almost exact closed-form expressions for

different performance measures with exactly the same analytical tractability as with the original KL approximation and with moderately increased computational complexity that is controllable with the choice of the number of terms.

The remainder of this paper is organized as follows. The next section presents our new approximations and bounds together with the optimization methodologies used for solving the sets of coefficients. The accuracy of the proposed approximations and bounds is validated in Section III by numerical results. After an overview of various applications of (2) in Section IV, the conclusion is given in Section V.

## II. NOVEL APPROXIMATIONS AND BOUNDS

This section finds the optimized coefficients,  $\{(a_n^*, b_n^*)\}_{n=1}^N$  and  $c^*$ , for the proposed GKL expression that offer variety to tailor accuracy for some specific application or to use bounds. For this reason, several optimization criteria are considered and each of them requires more or less different approach. The first two minimize maximum absolute and relative errors, whereas the remaining two minimize total absolute and relative errors.

### A. Minimax Approximations and Bounds

The GKL expression in (2) is optimized herein in the minimax sense by solving its corresponding coefficients as

$$\{(a_n^*, b_n^*)\}_{n=1}^N, c^* \triangleq \arg \min_{\{(a_n, b_n)\}_{n=1}^N, c} e_{\max}, \quad (3)$$

where

$$e_{\max} \triangleq \max_{x \geq 0} |e(x)|, \quad (4)$$

and the shorthand  $e \in \{d, r\}$  collectively represents both the absolute and relative error functions which are defined respectively as  $d(x) \triangleq \tilde{Q}(x) - Q(x)$  and  $r(x) \triangleq \frac{\tilde{Q}(x)}{Q(x)} - 1$ .

The minimax optimization results in uniform error functions that oscillate between local maximum and minimum values of equal magnitude and alternating signs as illustrated by the minimax approximations in Fig. 1(a). The absolute and relative error functions' derivatives vanish at these extrema points and are given respectively by  $d'(x) = \tilde{Q}'(x) - Q'(x)$  and  $r'(x) = (\tilde{Q}'(x)Q(x) - \tilde{Q}(x)Q'(x))/[Q(x)]^2$  where  $\tilde{Q}'(x) = -\frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)$  and  $\tilde{Q}(x) = -\frac{1}{x^2} \sum_{n=1}^N a_n ((2b_n x^2 + 1) \exp(cx) - 2b_n x^2 - cx - 1) \exp(-b_n x^2 - cx)$ .

We will shortly use the fact that the error functions converge to explicit values, which may be local extrema, at both ends of the non-negative real axis as follows:

$$\begin{aligned} d_0 &\triangleq \lim_{x \rightarrow 0} d(x) = c \sum_{n=1}^N a_n - \frac{1}{2}, & \lim_{x \rightarrow \infty} d(x) &= 0, \\ r_0 &\triangleq \lim_{x \rightarrow 0} r(x) = 2c \sum_{n=1}^N a_n - 1, \\ \lim_{x \rightarrow \infty} r(x) &= \begin{cases} \infty, & \text{if } b_1 < \frac{1}{2}, \\ \sqrt{2\pi} a_1 - 1, & \text{if } b_1 = \frac{1}{2}, \\ -1, & \text{if } b_1 > \frac{1}{2}, \end{cases} \end{aligned} \quad (5)$$

where  $a_1$  is the counterpart of  $b_1 \triangleq \min\{b_n\}_{n=1}^N$ .

It can be concluded from the above limit that global approximations and bounds exist in terms of the relative error if and only if  $b_1 = \frac{1}{2}$ , opposing to the absolute error function which is always bounded regardless of  $b_1$ 's value. Nevertheless, this study shows that the absolute and total errors can be reduced by allowing  $b_1 < \frac{1}{2}$ . Thus, we consider herein two variations of approximations w.r.t. absolute and total errors, namely, first variation with  $b_1 < \frac{1}{2}$  and second variation with  $b_1 = \frac{1}{2}$ .

1) *Approximations*: The optimized coefficients can be found by solving the following set of equations which describes the shape of the corresponding error function, for which  $x_k$  refers to the location of the error function's extrema and  $K$  refers to their number excluding the endpoints:

$$\begin{cases} f_0(\mathbf{v}) = e_0 + e_{\max} = 0 \\ f_k(\mathbf{v}) = e(x_k) + (-1)^k e_{\max} = 0, & \text{for } k = 1, 2, \dots, K, \\ f'_k(\mathbf{v}) = e'(x_k) = 0, & \text{for } k = 1, 2, \dots, K, \\ f_{K+1}(\mathbf{v}) = a_1 + \frac{r_{\max}-1}{\sqrt{2\pi}} = 0, & \text{only when } e = r. \end{cases} \quad (6)$$

Above,  $\mathbf{v}$  is a vector of the approximation's coefficients with  $e_{\max}$  which are to be optimized. More specifically,  $\mathbf{v} = [a_1, a_2, \dots, a_N, b_1, b_2, \dots, b_N, c, e_{\max}]$  with excluding  $b_1$  for the second variation of the absolute error and for the relative error since then  $b_1 = \frac{1}{2}$ . In addition,  $f_k(\mathbf{v})$  and  $f'_k(\mathbf{v})$  are two equations that report the error function's value and zero-derivative at each of the extrema points, and  $f_0(\mathbf{v})$  and  $f_{K+1}(\mathbf{v})$  result from evaluating the limits at both ends of the range  $[0, \infty)$  as in (5) to give one equation for the absolute error and two equations for the relative error that converges to  $-r_{\max}$  as  $x$  tends to infinity. For both error measures, the error function is assumed to start from  $e_0 = -e_{\max}$ .

When considering the absolute error,  $K = 2N + 1$  for the first variation and  $K = 2N$  for the second variation. A total of  $2K + 1$  equations including that at  $x = 0$  are formulated. On the other hand, for the bounded relative error, a total of  $2K + 2$  equations including those at the endpoint limits are formulated with  $K = 2N - 1$ . Generally, the number of equations for both error measures are equal to the number of unknowns, namely,  $\mathbf{v}$  and  $\{x_k\}_{k=1}^K$ . It is worth mentioning that the error function can also start from  $e_0 = 0$  to achieve continuity at the origin when extended to the negative values of  $x$  like for  $Q(x)$ , but at the expense of slightly less accuracy.

2) *Bounds*: Here we need to find the optimized sets of coefficients which, when substituted in (2), give uniform lower and upper bounds for which  $e(x) \leq 0$  and  $e(x) \geq 0$ , respectively. Error of a lower bound oscillates between zero and  $-e_{\max}$ , must have  $b_1 = \frac{1}{2}$  and start from  $e_0 = -e_{\max}$  for both error types. In addition, when it is optimized in terms of absolute error,  $K = 2N$  and its corresponding error function converges to zero as  $x$  tends to infinity, whereas when it is optimized in terms of relative error,  $K = 2N - 1$  and its error function converges to  $-r_{\max}$  as  $x$  tends to infinity.

On the other hand, the upper bound oscillates between zero and  $e_{\max}$  and must always start from  $e_0 = 0$  and converge to zero as  $x$  tends to infinity for both error types. In particular, for its optimization in terms of absolute error,  $K = 2N + 1$ , whereas  $K = 2N - 1$  for its optimization in terms of relative

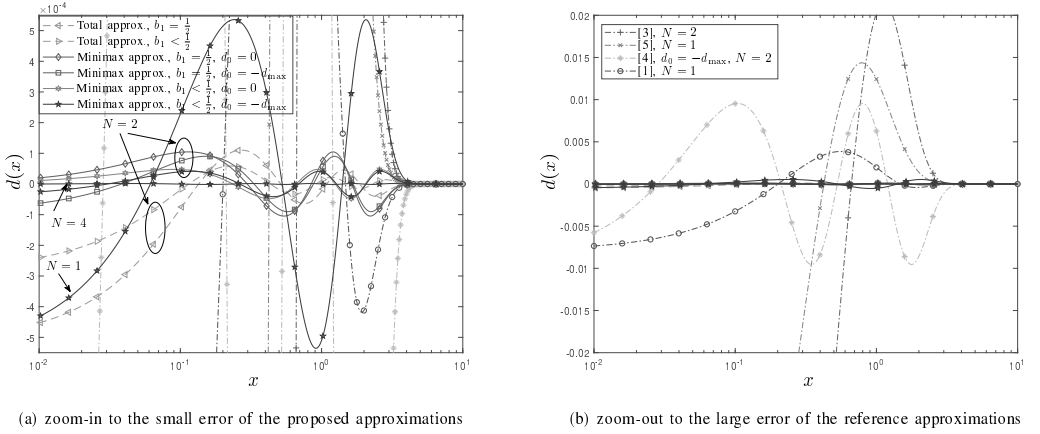


Fig. 1. Comparison between our approximations and the reference ones for  $N = 1$ ,  $N = 2$  and  $N = 4$  in terms of the absolute error.

error. Using the aforementioned description, the optimization problem can be easily formulated in the same way as in (6).

**3) Implementation of the Minimax Optimization and the Remez Exchange Algorithm:** The sets of equations formulated for each of the proposed approximations and bounds can be straightforwardly solved using any numerical tool. However, good initial guesses for the unknowns are required in order for their values to converge to the optimized ones. The initial guesses used herein are obtained heuristically and it was quite a challenge to get good ones for  $N > 5$ . We have solved this problem by proposing a variation of the Remez exchange algorithm for acquiring the optimized coefficients for  $N > 5$  and establishing the same uniform minimax error function but with  $K$  equations ( $\{f_k\}_{k=1}^K$ ) less than the approach introduced in Section II-A1. The absence of the derivative equations makes it less sensitive to the right choice of the initial guesses.

In particular, we construct a system of nonlinear equations describing the values of the extrema points of the corresponding error function, which alternate exactly  $L = K$  times for the absolute error and  $L = K + 1$  times for the relative error, as  $\mathbf{f}(\mathbf{v}) \triangleq [f_0(\mathbf{v}), f_1(\mathbf{v}), \dots, f_L(\mathbf{v})]^T$  for which  $f_l, l = 0, 1, \dots, L$  and  $\mathbf{v}$  are defined in (6), and  $\mathbf{f}$  and  $\mathbf{v}$  have equal lengths. We set up the Remez algorithm by initializing the locations of the  $K$  extrema while taking into consideration both endpoints, which might be local extrema.

Next, we start the first iteration by solving  $\mathbf{f}$  for  $\mathbf{v}$  using the iterative Newton–Raphson method whose iterations also require initial guesses for  $\mathbf{v}$  and are performed as

$$\mathbf{v}^{(t+1)} = \mathbf{v}^{(t)} - \left[ \mathbf{J}^{(t)}(\mathbf{v}^{(t)}) \right]^{-1} \mathbf{f}(\mathbf{v}^{(t)}), \quad (7)$$

where  $t$  is its counter and  $\mathbf{J}(\cdot)$  is the Jacobian matrix defined as  $\mathbf{J}(\mathbf{v}) = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial v_0} & \frac{\partial \mathbf{f}}{\partial v_1} & \dots & \frac{\partial \mathbf{f}}{\partial v_L} \end{bmatrix}$ . For the absolute error,  $\frac{\partial f_0}{\partial a_n} = c$ ,  $\frac{\partial f_0}{\partial b_n} = 0$ ,  $\frac{\partial f_k}{\partial c} = \sum_{n=1}^N a_n$ ,  $\frac{\partial f_k}{\partial a_n} = \frac{(1 - \exp(-c x_k))}{x_k} \exp(-b_n x_k^2)$ ,  $\frac{\partial f_k}{\partial b_n} = -a_n x_k (1 - \exp(-c x_k)) \exp(-b_n x_k^2)$ ,  $\frac{\partial f_k}{\partial c} = \exp(-c x_k) \sum_{n=1}^N a_n \exp(-b_n x_k^2)$ , whereas for the relative error, we multiply the above relations  $\frac{\partial f_0}{\partial a_n}$

and  $\frac{\partial f_0}{\partial c}$  by two and divide  $\frac{\partial f_k}{\partial a_n}$ ,  $\frac{\partial f_k}{\partial b_n}$  and  $\frac{\partial f_k}{\partial c}$  by  $Q(x_k)$ . Also, for the relative error only,  $\frac{\partial f_{K+1}}{\partial a_1} = 1$ ,  $\frac{\partial f_{K+1}}{\partial a_n} |_{n \neq 1} = \frac{\partial f_{K+1}}{\partial b_n} = \frac{\partial f_{K+1}}{\partial c} = 0$ , and  $\frac{\partial f_{K+1}}{\partial r_{\max}} = \frac{1}{\sqrt{2}\pi}$ . In addition,  $\frac{\partial f_0}{\partial e_{\max}} = 1$  and  $\frac{\partial f_k}{\partial e_{\max}} = (-1)^k$  for both error measures. The Newton–Raphson iterations are repeated until  $\Delta \mathbf{v} = \mathbf{v}^{(t+1)} - \mathbf{v}^{(t)}$  is less than a threshold value.

Then, we locate the new extrema of the resulting error function and use them for the following Remez iteration which we repeat until the difference between the old and new  $K$  extrema lies below a threshold value. Note that the Newton–Raphson method is implemented in every iteration of the Remez algorithm. Although the Remez algorithm still requires initial guesses for the unknowns like the approach in Section II-A1, it is much more robust against the accuracy of the initial guesses and converges very rapidly to the optimal solution. The optimized coefficients of minimax GKL approximations and bounds are solved herein up to  $N = 10$  for the two variations of the absolute error and for the relative error and released to public domain as a supplementary dataset.<sup>1</sup>

## B. Numerical Optimization in Terms of Total Error

The coefficients of the GKL expression can also be optimized in terms of the total integrated error as

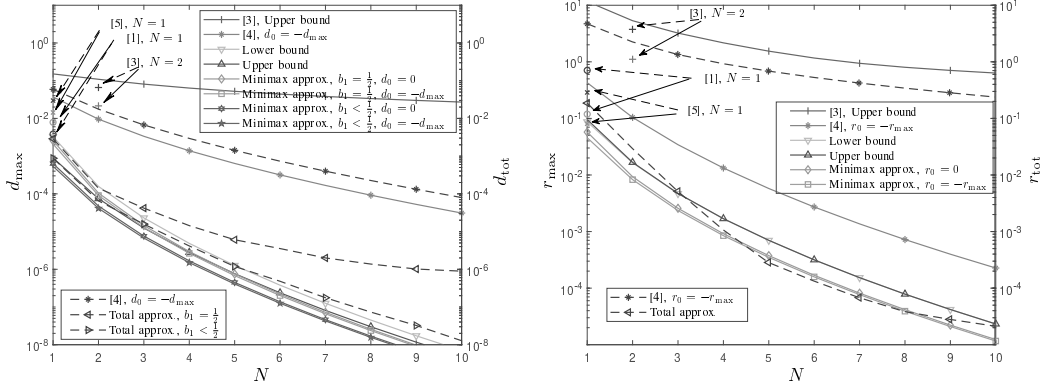
$$\{(a_n^*, b_n^*)\}_{n=1}^N, c^* \triangleq \arg \min_{\{(a_n, b_n)\}_{n=1}^N, c} e_{\text{tot}}, \quad (8)$$

where

$$e_{\text{tot}} \triangleq \int_0^R |e(x)| dx. \quad (9)$$

For  $e = d$ ,  $R \rightarrow \infty$  in order to obtain globally optimized approximations since  $d(x)$  converges to zero when  $x$  tends to infinity, whereas  $R$  is some constant for  $e = r$  which converges to a constant value when  $x$  tends to infinity for  $b_1 = \frac{1}{2}$ . We apply the quasi-Newton algorithm to perform the optimization

<sup>1</sup>Available at <https://doi.org/10.5281/zenodo.5806271> for download.



(a) absolute error of GKL approximations and bounds with reference cases

(b) relative error of GKL approximations and bounds with reference cases

Fig. 2. The proposed GKL approximations and bounds compared to existing ones from the literature including the original KL approximation.

herein. In particular, we used the `fminunc` command in Matlab with setting its 'Algorithm' to 'quasi-newton' in order to minimize the target function  $e_{\text{tot}}$ . The error function can also be forced to start from zero by adding the constraint  $\sum_{n=1}^N a_n = \frac{1}{2c}$ , which results from the limit at zero, and we then used `fmincon` command instead.

We start with heuristic initial guesses for the unknowns that converge eventually to the optimized values. In fact, we were able to use the minimax-optimized sets as mean values around which small random variance is introduced to work as initial guesses for their equivalent cases herein. Note that the `fminunc` command finds the local minimum of the target function. Therefore, we need to repeatedly run the local solver to locate a solution that has the lowest target function value. The optimized coefficients to the GKL approximations are also solved herein for the two variations of the absolute error and for the relative error with  $R = 10$  in terms of the total error.<sup>1</sup>

### III. NUMERICAL RESULTS

This section demonstrates how excellent the GKL approximations and bounds perform, were they achieve world-record low error levels as will be seen next. Figure 1 illustrates the absolute error functions resulted from applying our approximations and key existing ones. Obviously, the proposed approximations are extremely tight and even more interestingly, they substantially outperform all the reference cases for the whole non-negative real axis even with only one term as can be noted from the huge displacement in the corresponding curves in Fig. 1(b). The accuracy increases considerably further when increasing  $N$  as seen by the comparison between  $N = 1$ ,  $N = 2$  and  $N = 4$  for the minimax approximation.

Figure 2 plots the global error of the minimax approximations and bounds proposed in Section II-A together with the reference cases (solid lines and markers with solid arrows), in addition to the total error of the approximations proposed in Section II-B and the reference cases (dashed lines and markers with dashed arrows), both for  $N = 1, 2, \dots, 10$  and in terms of both error measures. With small  $N = 1, 2, 3$ , they

already significantly outperform the reference ones and their accuracy increases considerably by increasing  $N$ . Ultimately, the proposed GKL expression with optimized coefficients reaches extremely low levels in the order of  $10^{-9}$  and  $10^{-5}$  for absolute and relative errors, respectively, with  $N = 10$ . It should be noted that the proposed approximations and bounds in the special case of  $N = 1$  are the same as those in [2].

### IV. OVERVIEW OF APPLICATIONS

The applications of the original KL approximation and the newly proposed GKL approximation (2)—both have the same analytical complexity—are about the same and span different areas of communication theory. A popular application example would be evaluating the average symbol error probability for coherent detection, which results in linear combinations of the following integral with different integer values of  $P$ :

$$I_P(\bar{\gamma}) \triangleq \int_0^{\infty} Q^P(\sqrt{\gamma}) \phi_{\gamma}(\gamma) d\gamma, \quad (10)$$

where  $\phi_{\gamma}(\gamma)$  is the fading probability density function of the instantaneous signal-to-noise ratio  $\gamma$  with average  $\bar{\gamma}$ .

When assuming generalized  $\kappa - \mu$  distribution, (10) can be evaluated using [15, Eqs. 3.351.3, 3.462.1, and 8.445] after applying (2) to express tight approximations for the  $P$ th integer power of the Gaussian  $Q$ -function using the multinomial expansion. This yields

$$I_P(\bar{\gamma}) \approx \sum_{\tau=0}^{\infty} \frac{\mu^{\mu+2\tau}}{\exp(-\kappa\mu)} \frac{\kappa^{\tau}(1+\kappa)^{\mu+\tau}}{\Gamma(\mu+\tau)\tau!} \Psi_P(\bar{\gamma}) \quad (11)$$

and

$$I_P(\bar{\gamma}) \approx \frac{m^m}{\Gamma(m)} \Psi_P(\bar{\gamma}) \quad (12)$$

in the general case and in the special case of Nakagami- $m$  fading (that occurs at  $\kappa = 0$  and  $\mu = m$ ), respectively. The

convergent infinite series in (11) can be truncated to the desired accuracy. For the above expressions,

$$\begin{aligned} \Psi_P(\bar{\gamma}) = & \sum_{p_1+p_2+\dots+p_N=P} \sum_{j=0}^P \binom{P}{j} (-1)^j \frac{1}{\bar{\gamma}^\mu} G H \left[ C^{-A-1} \right. \\ & \times \Gamma(A+1) - C^{-A-\frac{3}{2}} \left[ \sqrt{C} \Gamma(A+1) {}_1F_1 \left( A+1; \frac{1}{2}; \frac{B^2}{4C} \right) \right. \\ & \left. \left. - B \Gamma \left( A + \frac{3}{2} \right) {}_1F_1 \left( A + \frac{3}{2}; \frac{3}{2}; \frac{B^2}{4C} \right) \right] \right], \quad (13) \end{aligned}$$

in which the first summation is taken over all combinations of non-negative integer indices  $p_1$  through  $p_N$  such that the sum of all  $p_n$  is  $P$ . The parameters are  $G = \binom{P}{p_1, p_2, \dots, p_N}$ ,  $H = a_1^{p_1} a_2^{p_2} \dots a_N^{p_N}$ ,  $A = \frac{2\mu+2\tau-2-P}{2} > -1$ ,  $B = c_j$ ,  $C = \Lambda + \frac{\mu(1+\kappa)}{\bar{\gamma}}$ , and  $\Lambda = b_1 p_1 + b_2 p_2 + \dots + b_N p_N$ ; moreover, the parameters  $A$  and  $C$  reduce to  $A = \frac{2m-2-P}{2} > -1$  and  $C = \Lambda + \frac{m}{\bar{\gamma}}$  for the special case of Nakagami- $m$  fading.

Let us then overview a few examples [16]–[20] from the wide range of applications available in the literature for which the proposed GKL expression is applicable as a substitute for the KL expression that was originally used in those publications. In particular, the GKL approximation/bound can be used to calculate the sampling bit error probability of binary phase shift keying [16], to approximate the phase noise probability density function in the system considered in [17], and to derive the coherent LoRa<sup>®</sup> symbol error rate under additive white Gaussian noise [18]. Beyond communications, it allows to approximate the distribution functions of particles experiencing compound subdiffusion [19] and to derive the predictive error of the probability of failure [20], for instance.

Furthermore, the simplified series expansion of the original KL expression proposed in [21] can be applied likewise to (2) with the optimized coefficients, which results in

$$Q(x) \approx \sum_{l=1}^L \sum_{n=1}^N \frac{(-1)^{l+1} a_n c^l}{l!} \exp(-b_n x^2) x^{l-1}. \quad (14)$$

Since (14) can be used as a direct substitute for [21, Eq. 3], the proposed GKL approximations are also useful for the applications considered in [22]–[25] (and many others that cite [21]) and improve the accuracy of the analysis thereof.

## V. CONCLUSION

This Letter presented a new tractable expression for approximating the Gaussian  $Q$ -function together with multiple alternatives of coefficient sets for it<sup>1</sup> that are optimized to minimize either the global or total absolute/relative errors, from which the best suitable set is chosen for any application at hand. The extremely low error levels allow for their usage as highly reliable substitutions to the  $Q$ -function in order to derive virtually exact analytical expressions for different performance metrics in communication theory. Moreover, we extended the proposed expression to minimax bounds (with comparable accuracy to that of the approximations) that are useful when the worst/best case scenarios are of interest.

## REFERENCES

- [1] G. Karagiannidis and A. Lioumpas, “An improved approximation for the Gaussian  $Q$ -function,” *IEEE Commun. Lett.*, vol. 11, no. 8, pp. 644–646, Aug. 2007.
- [2] I. M. Tanash and T. Riihonen, “Improved coefficients for the Karagiannidis–Lioumpas approximations and bounds to the Gaussian  $Q$ -function,” *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1468–1471, May 2021.
- [3] M. Chiani *et al.*, “New exponential bounds and approximations for the computation of error probability in fading channels,” *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 840–845, Jul. 2003.
- [4] I. M. Tanash and T. Riihonen, “Global minimax approximations and bounds for the Gaussian  $Q$ -function by sums of exponentials,” *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6514–6524, Oct. 2020.
- [5] P. Börjesson and C. Sundberg, “Simple approximations of the error function  $Q(x)$  for communications applications,” *IEEE Trans. Commun.*, vol. 27, no. 3, pp. 639–643, Mar. 1979.
- [6] M. López-Bentéz and F. Casadevall, “Versatile, accurate, and analytically tractable approximation for the Gaussian  $Q$ -function,” *IEEE Trans. Commun.*, vol. 59, no. 4, pp. 917–922, Apr. 2011.
- [7] Q. Shi, “Novel approximation for the Gaussian  $Q$ -function and related applications,” in *Proc. 22nd IEEE PIMRC*, Sep. 2011, pp. 2030–2034.
- [8] G. Abreu, “Very simple tight bounds on the  $Q$ -function,” *IEEE Trans. Commun.*, vol. 60, no. 9, pp. 2415–2420, Sep. 2012.
- [9] O. Olabiyi and A. Annamalai, “Invertible exponential-type approximations for the Gaussian probability integral  $Q(x)$  with applications,” *IEEE Wireless Commun. Lett.*, vol. 1, no. 5, pp. 544–547, Oct. 2012.
- [10] D. Sadhwani *et al.*, “Tighter bounds on the Gaussian  $Q$ -function and its application in Nakagami- $m$  fading channel,” *IEEE Wireless Commun. Lett.*, vol. 6, no. 5, pp. 574–577, Oct. 2017.
- [11] J. Craig, “A new, simple and exact result for calculating the probability of error for two-dimensional signal constellations,” in *Proc. IEEE MILCOM*, vol. 2, Nov. 1991, pp. 571–575.
- [12] P. Lee, “Computation of the bit error rate of coherent  $M$ -ary PSK with Gray code bit mapping,” *IEEE Trans. Commun.*, vol. 34, no. 5, pp. 488–491, May. 1986.
- [13] M. Irshid and I. Salous, “Bit error probability for coherent  $M$ -ary PSK systems,” *IEEE Trans. Commun.*, vol. 39, no. 3, pp. 349–352, Mar. 1991.
- [14] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels*, 2nd ed. John Wiley and Sons, Inc., Jan. 2005.
- [15] I. Gradshteyn and I. Ryzhik, *Table of integrals, series, and products*, 7th ed. Elsevier/Academic Press, 2007.
- [16] W. M. Jang, “A simple performance approximation of general-order rectangular QAM with MRC in Nakagami- $m$  fading channels,” *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3457–3463, Sep. 2013.
- [17] S. Lin, Z. Wang, J. Xiong, Y. Fu, J. Jiang, Y. Wu, Y. Chen, C. Lu, and Y. Rao, “Rayleigh fading suppression in one-dimensional optical scatters,” *IEEE Access*, vol. 7, pp. 17 125–17 132, Jan. 2019.
- [18] O. Afisiadis, S. Li, J. Tapparel, A. Burg, and A. Balatsoukas-Stimming, “On the advantage of coherent LoRa detection in the presence of interference,” *IEEE Internet Things J.*, vol. 8, no. 14, pp. 11 581–11 593, Jul. 2021.
- [19] A. Shalchi and V. Arendt, “Distribution functions of energetic particles experiencing compound subdiffusion,” *Astrophys. J.*, vol. 890, no. 2, p. 147, Feb. 2020.
- [20] J. Wauters, I. Couckuyt, and J. Degroote, “A new surrogate-assisted single-loop reliability-based design optimization technique,” *Struct. Multidisc. Optim.*, vol. 63, no. 6, p. 2653–2671, Apr. 2021.
- [21] Y. Isukapalli and B. D. Rao, “An analytically tractable approximation for the Gaussian  $Q$ -function,” *IEEE Commun. Lett.*, vol. 12, no. 9, pp. 669–671, Sep. 2008.
- [22] —, “Packet error probability of a transmit beamforming system with imperfect feedback,” *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2298–2314, Apr. 2010.
- [23] M. Seyfi, S. Muhaidat, J. Liang, and M. Dianati, “Effect of feedback delay on the performance of cooperative networks with relay selection,” *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4161–4171, Dec. 2011.
- [24] A. Kumar, P. Thakur, S. Pandit, and G. Singh, “Threshold selection and cooperation in fading environment of cognitive radio network: Consequences on spectrum sensing and throughput,” *Int. J. Electron. Commun.*, vol. 117, p. 153101, Apr. 2020.
- [25] M. A. Al-Jarrah, E. Alsusa, A. Al-Dweik, and M.-S. Alouini, “Performance analysis of wireless mesh backhauling using intelligent reflecting surfaces,” *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3597–3610, Jun. 2021.



# PUBLICATION

6

**Tight logarithmic approximations and bounds for generic capacity  
integrals and their applications to statistical analysis of wireless  
systems**

I. M. Tanash and T. Riihonen

*IEEE Transactions on Communications*, in press

DOI: 10.1109/TCOMM.2022.3198435

**Publication reprinted with the permission of the copyright holders.**





# Tight Logarithmic Approximations and Bounds for Generic Capacity Integrals and Their Applications to Statistical Analysis of Wireless Systems

Islam M. Tanash  and Taneli Riihonen , *Senior Member, IEEE*

**Abstract**—We present tight yet tractable approximations and bounds for the ergodic capacity of any communication system in the form of a weighted sum of logarithmic functions, with the focus on the Nakagami and lognormal distributions that represent key building blocks for more complicated systems. A minimax optimization technique is developed to derive their coefficients resulting in uniform absolute or relative error. These approximations and bounds constitute a powerful tool for the statistical performance analysis as they enable the evaluation of the ergodic capacity of various communication systems that experience small-scale fading together with the lognormal shadowing effect and allow for simplifying the complicated integrals encountered when evaluating the ergodic capacity in different communication scenarios. Simple and tight closed-form solutions for the ergodic capacity of many classic and timely application examples are derived using the logarithmic approximations. The high accuracy of the proposed approximations is verified by numerical comparisons with existing approximations and with those obtained directly from numerical integration methods.

**Index Terms**—Ergodic capacity, minimax approximation, bounds, performance analysis, fading distributions.

## I. INTRODUCTION

**E**RGODIC capacity is an important measure for analyzing the performance of different communication systems [1]. It specifies the maximum transmission rate of reliable communication that can be achieved over time-varying channels. Specific formulations of ergodic capacity can be referred to as *capacity integrals* based on the way how they are found by calculating the expectation of instantaneous channel capacity using probability density functions (PDFs) that model fading. Establishing closed-form expressions for ergodic capacity is of great importance in communication theory since they enable us to gain scientific understanding of the behavior of communication systems and the effect of their parameters on the performance. In this area, our research work aims at facilitating the statistical performance analysis of wireless systems by developing novel mathematical tools that build upon the following general result in this article.

**Proposition 1:** For any wireless system with instantaneous capacity  $\mathcal{C} \triangleq \log_2(1 + \gamma_{\text{eff}})$  conditioned on fading states,

Manuscript received June 25, 2021; revised January 11, 2022, May 20, 2022; accepted May 21, 2022. This work was supported by the Academy of Finland under the grants 310991/326448, 315858, 341489, and 346622. The associate editor coordinating the review of this paper and approving it for publication was Sami Muhaidat. (*Corresponding author: Islam M. Tanash.*)  
The authors are with the Faculty of Information Technology and Communication Sciences, Tampere University, Korkeakoulunkatu 1, FI-33720 Tampere, Finland (e-mail: islam.tanash@tuni.fi; taneli.riihonen@tuni.fi).

Digital Object Identifier X

where  $\gamma_{\text{eff}} \triangleq 2^{\mathcal{C}} - 1$  denotes *effective* (not necessarily actual) signal-to-noise ratio (SNR) with average  $\bar{\gamma}_{\text{eff}} \triangleq \mathbb{E}[\gamma_{\text{eff}}]$ , the ergodic capacity can be approximated with arbitrary accuracy as

$$\bar{\mathcal{C}} \triangleq \mathbb{E}[\mathcal{C}] \approx \sum_{n=1}^N a_n \log_2(1 + b_n \bar{\gamma}_{\text{eff}}) \quad (1)$$

by choosing the coefficients  $\{(a_n, b_n)\}_{n=1}^N$  appropriately.

*Proof:* See Appendix A. ■

One will instantly notice that the generic approximation (1) is a weighted sum of the Shannon capacities of basic static additive white Gaussian noise (AWGN) channels. In other words, the greatness of Proposition 1 is that it proves that *any* system with fading channels is *in terms of capacity* equivalent to a system (cf. Fig. 1), wherein a scheduler employs randomly one of  $N + 1$  parallel static channels for the transmission of each data block:<sup>1</sup> Channel  $n$ ,  $n = 1, 2, \dots, N$ , having SNR of  $b_n \bar{\gamma}_{\text{eff}}$  is chosen with probability  $a_n$  and Channel 0 represents a completely blocked channel ( $b_0 = 0$ ), i.e., an outage event takes place with remaining probability  $a_0 = 1 - \sum_{n=1}^N a_n$ .

While Proposition 1 is powerful in proving the general existence of the approximation (1) for the ergodic capacity of any wireless system at large, it is not so applicable as an actual approximation for any specific system. This is because, firstly, the coefficients  $a_n$ ,  $n = 1, 2, \dots, N$ , are in the direct application computed from the PDF of  $\mathcal{C}$ , which is typically not derived explicitly in statistical performance analysis, and it may be cumbersome or even impossible to express. Secondly and more importantly, when choosing the coefficients from the Riemann sum according to the proof, the resulting approximations are inefficient, because a very large number of logarithmic terms are needed for adequate accuracy.

In this paper, we aim to evolve Proposition 1 into a useful, efficient tool in two ways. Firstly, we develop a systematic methodology to optimize coefficients  $\{(a_n, b_n)\}_{n=1}^N$  to approximate any communication system's ergodic capacity  $\bar{\mathcal{C}} = C(1/\bar{\gamma}_{\text{eff}}/\log_e(2))$  that can be expressed with the generic function  $C(\cdot)$  of some open or closed form. Secondly, we implement the presented optimization methodology to find  $\{(a_n, b_n)\}_{n=1}^N$  explicitly under Nakagami and lognormal fading (when  $C(\cdot)$  becomes  $C_m(\cdot)$ , the ‘Nakagami capacity integral’, or  $C_\sigma(\cdot)$ , the ‘lognormal capacity integral’) and show

<sup>1</sup>An alternative interpretation is a scheduler that employs the parallel channels sequentially for data blocks with relative durations  $a_n$ ,  $n = 0, 1, \dots, N$ .

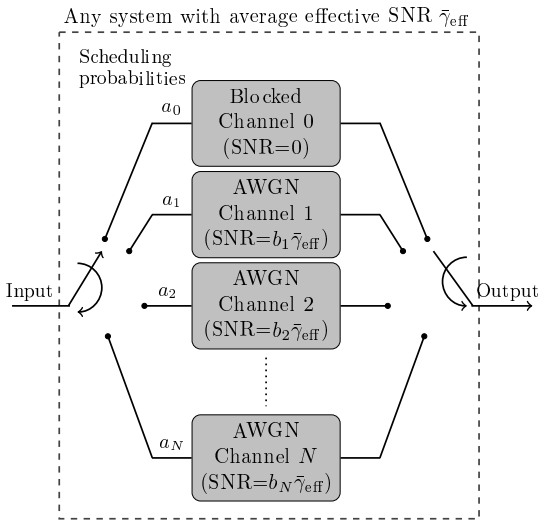


Fig. 1. Interpretation of the ergodic capacity of any communication system as a scheduler which randomly employs one of the parallel static channels when transmitting data blocks.

how to use them as building blocks for the capacity analysis of complex systems that manifest them in intermediate steps.

#### A. Related Works

Capacity integrals have been investigated extensively in the literature for countless transmission systems under various assumptions on transmitter and receiver channel knowledge and over different fading distributions [2]–[18]. In [2]–[12], the ergodic capacity over Rayleigh fading is evaluated for single-antenna systems and multi-antenna systems — namely, multiple-input single-output (MISO), single-input multiple-output (SIMO), and multiple-input multiple-output (MIMO) — for correlated or non-correlated channels and different combining techniques at the receiver. Moreover, the ergodic capacity for single-antenna and multi-antenna systems with non-correlated channels is evaluated over Nakagami fading in [13], [14] and over Rician fading in [15]–[17]. The ergodic capacity under  $\kappa - \mu$  fading is derived in [18].

Generally, the precise ergodic capacity expressions are difficult to express in analytical forms. This has motivated the work toward deriving approximations and bounds for capacity integrals [19]–[25]. They are also needed among many other purposes for optimal power allocation and network design. In particular, the authors in [19] present a lower bound for the capacity integral of MIMO Rayleigh channels with frequency-selective fading and/or channel correlation, together with an asymptotic approximation of the ergodic capacity over flat fading. Other asymptotic results are derived in [20] for specific multi-antenna scenarios with the channel knowledge at the receiver at first, and then at the transmitter as well.

In [21], more generic expressions for bounding the ergodic capacity are presented. In [22], two less accurate yet tractable

approximations that enable the development of analytical resource allocation strategies in Rayleigh MIMO systems are derived. The authors in [23] propose two simple yet accurate approximations for the ergodic capacity in the low-SNR region. Closed-form bounds for the ergodic capacity in dual-hop fixed-gain amplify and forward relay networks are proposed in [24] over Rayleigh fading channels, and in [25] over Nakagami fading channels.

In addition to the small-scale fading, the ergodic capacity is also investigated under the shadowing effect that is usually modeled by the lognormal distribution. The ergodic capacity of communication systems under lognormal fading channels does not admit a closed-form expression. Therefore, several approximations and bounds have been proposed to express it in terms of analytical functions [26]–[30]. The very first lower and upper bounds for evaluating the ergodic capacity over lognormal fading channels were presented in [26], resulting in simple yet loose bounds for lower values of SNR.

Other approximations were later developed in [27], [28] for single-input single-output (SISO) systems and the results were also generalized to approximate the capacity of diversity combining techniques with or without channel correlation, based on the fact that the sum of lognormal random variables can be well approximated by an equivalent lognormal one. In [29], a tight approximation for the lognormal capacity integral is presented and investigated for SISO and MIMO indoor ultra-wideband systems. The authors in [30] derive closed-form approximations for the capacity integral of various adaptive transmission schemes under lognormal distribution.

#### B. Contributions and Organization of the Paper

The unified fundamental tool, i.e., (1), contributed in this article enables the accurate evaluation of ergodic capacity in any communication system at large in the form of the weighted sum of logarithmic functions. It requires optimizing the corresponding coefficients so that they work as highly efficient replacements for those obtained from the numerical methods such as the Riemann sum in the proof of Proposition 1. Nevertheless, we also implement the proposed approach to offer novel logarithmic approximations and bounds with optimized coefficients specifically for the Nakagami and lognormal capacity integrals. Since these two integrals most frequently appear as building blocks for many more-complex communication systems, this often leads to logarithmic approximations and bounds in the same format of (1) for their capacity expressions. This avoids the need to formulate equivalent methodology and solve the coefficients specifically for every individual system despite the general tool facilitates that too.

We can summarize the contributions in this paper as follows.

- We propose a systematic methodology to optimize the approximations' coefficients and obtain the best logarithmic approximations in terms of the minimax absolute error for the capacity of any communication system. This requires redeveloping the related scheme that we previously presented in [31] for error probability analysis, which is inherently different from capacity analysis.

- We implement the optimization methodology on the Nakagami- $m$  channel (and over Rayleigh fading as a special case thereof) to derive minimax approximations for it. Especially, the approximations are valid for any value of  $m$ , opposing to the exact closed-form expression in [13, Eq. 23], which is valid only for its integer values.
- We show how the optimized approximation of the Nakagami capacity integral can be used as a building block to derive the capacity integral of many complicated communication systems [2]–[18], [32]–[37] and can even often lead to the same logarithmic form as an end result.
- Likewise, we find the optimized coefficients for the approximation of the lognormal capacity integral which enables the evaluation of the ergodic capacity for various communication systems that experience small-scale fading together with the lognormal shadowing effect, in the form of a sum of logarithmic terms. In particular, for a composite lognormal channel, we apply the sum of logarithms with its optimized coefficients to approximate the ergodic capacity over the small-scale fading channel first. The resulting integral has exactly the same form as the lognormal capacity integral, which we approximate again by the sum of logarithmic functions.
- We extend the proposed minimax method to find the optimized parameters of the logarithmic approximation in terms of the relative error. We also extend it to find new logarithmic lower and upper bounds with optimized parameters in terms of both error measures.

We validate the aforementioned contributions with an extensive set of application examples that demonstrate the wide range of applicability of the proposed approximations. We further illustrate their high accuracy by numerical comparisons with other existing approximations or those obtained by numerical integration methods. In fact, their accuracy is so high that they can be considered to be virtually exact in most applications while they allow deriving closed-form results in cases where exact analysis is considered to be impossible.

We organize the rest of this paper as follows. Section II introduces some needed background information to formulate and solve the research problem. The main contribution is presented in Section III, where we propose the new methodology to acquire tight logarithmic approximations and bounds for ergodic capacity at large. In Section IV, a wide range of applications are considered and their capacities are evaluated in terms of the proposed approximations. In Section V, the numerical results demonstrate the high accuracy of the proposed approximations compared to other existing and numerical ones. Finally, we conclude the paper in the last section.

## II. PRELIMINARIES

In this paper, we shall develop unified approximations and bounds in the format of (1) that apply for the ergodic capacity  $\bar{C} = C(1/\bar{\gamma}_{\text{eff}})/\log_e(2)$  of any communication system, where the *generic capacity function*  $C(x)$  can be of any mathematical form. In most communication systems' analysis,  $C(x)$  can be represented as a capacity integral that calculates the average of  $\mathcal{C} \triangleq \log_2(1 + \gamma_{\text{eff}})$  per the following definitions.

*Definition 1:* Given average effective SNR  $\bar{\gamma}_{\text{eff}}$  with  $G \triangleq \frac{2^{\bar{C}} - 1}{\bar{\gamma}_{\text{eff}}} = \frac{\gamma_{\text{eff}}}{\bar{\gamma}_{\text{eff}}}$ , whose PDF exists and is denoted by  $f_G(\cdot)$ , the ergodic capacity of the corresponding communication system is  $\bar{C} = C(1/\bar{\gamma}_{\text{eff}})/\log_e(2)$  [bit/s/Hz], where the *generic capacity integral* is defined as

$$C(x) \triangleq \int_0^\infty \log_e \left( 1 + \frac{t}{x} \right) f_G(t) dt. \quad (2)$$

One should note that the generic capacity function  $C(x)$  is not necessarily given by the above generic capacity integral when the presented tool is still applicable. Nevertheless, we shall focus on developing the approximations and applications for the following specific integrals, which originate from evaluating (2) for Nakagami (including Rayleigh) and lognormal fading channels. These integrals appear frequently as part of longer expressions or in intermediate calculation steps when analyzing the capacity of more complex wireless systems.

*Definition 2:* Given average SNR  $\bar{\gamma}$ , the ergodic capacity of a Nakagami- $m$  fading channel is  $\bar{C} = C_m(1/\bar{\gamma})/\log_e(2)$  [bit/s/Hz], where the *Nakagami capacity integral* is defined as

$$\begin{aligned} C_m(x) &\triangleq \int_0^\infty \frac{m^m}{\Gamma(m)} \log_e \left( 1 + \frac{t}{x} \right) t^{m-1} \exp(-mt) dt \\ &= \exp(mx) \sum_{k=0}^{m-1} \Gamma(-k, mx) (mx)^k, \end{aligned} \quad (3)$$

for  $x > 0$  [13, Eqs. 21 and 23] with  $\Gamma(\zeta, x) = \int_x^\infty t^{\zeta-1} \exp(-t) dt$  denoting the upper incomplete gamma function [38, Eq. 6.5.3] and  $m$  being the fading parameter; the latter expression is valid for integer values of  $m$  only.

Substituting  $m = 1$  in the above definition, we obtain the ergodic capacity of a Rayleigh fading channel as a special case as  $\bar{C} = C_1(1/\bar{\gamma})/\log_e(2)$  [bit/s/Hz], where the *Rayleigh capacity integral* is defined as

$$\begin{aligned} C_1(x) &= \int_0^\infty \log_e \left( 1 + \frac{t}{x} \right) \exp(-t) dt \\ &= \exp(x) E_1(x), \end{aligned} \quad (4)$$

for  $x > 0$  [2, Eqs. 4 and 5] with  $E_1(x) = \int_x^\infty \exp(-t)/t dt$  denoting the exponential integral [38, Eq. 5.1.1].

*Definition 3:* Given average SNR  $\bar{\gamma} = \exp(\eta + \frac{\sigma^2}{2})$ , in which  $\eta$  and  $\sigma$  are the mean and the standard deviation of the corresponding instantaneous SNR's natural logarithm, respectively, the ergodic capacity of a lognormal fading channel is  $\bar{C} = C_\sigma(1/\bar{\gamma})/\log_e(2)$  [bit/s/Hz], where the *lognormal capacity integral* is defined as

$$\begin{aligned} C_\sigma(x) &\triangleq \int_{-\infty}^\infty \frac{1}{\sqrt{\pi}} \log_e \left( 1 + \frac{1}{x} \exp \left( \sqrt{2} \sigma^2 t - \frac{\sigma^2}{2} \right) \right) \\ &\quad \times \exp(-t^2) dt, \end{aligned} \quad (5)$$

for  $x > 0$  [26, Eq. 29]; this integral does not admit a closed-form expression so its approximations are crucial to have.

The Rayleigh capacity integral in (4) admits a sandwich bound according to [38, Eq. 5.1.20] as

$$\frac{1}{2} \log_e \left( 1 + \frac{2}{x} \right) < C_1(x) < \log_e \left( 1 + \frac{1}{x} \right), \quad (6)$$

and any linear combination thereof could be used as an obvious, but loose, approximation for the ergodic capacity over a Rayleigh fading channel. Inspired by this fact and Proposition 1, we develop a family of *tractable* functions

$$\tilde{C}(x) \triangleq \sum_{n=1}^N a_n \log_e \left( 1 + \frac{b_n}{x} \right) \quad (7)$$

for  $x > 0$ , that offer *tight approximations and bounds* for  $C(x)$  as  $\tilde{C}(x)$ , for  $C_m(x)$  in (3) as  $\tilde{C}_m(x)$  and for  $C_\sigma(x)$  in (5) as  $\tilde{C}_\sigma(x)$  by proper parameter choice. They are directly related to Proposition 1 as  $\tilde{C}(1/\bar{\gamma}_{\text{eff}})/\log_e(2)$  results in the logarithmic approximation given in (1). Furthermore, it should be noted that all  $N!$  permutations of the parameter set  $\{(a_n, b_n)\}_{n=1}^N$  yield an equivalent function, although we always choose the canonical (sorted) representation with  $a_1 \leq a_2 \leq \dots \leq a_N$ .

The absolute and relative error functions  $d(x)$  and  $r(x)$ , respectively, as well as their first-order derivatives  $d'(x)$  and  $r'(x)$ , respectively, are needed in what follows. They are defined as

$$d(x) \triangleq \tilde{C}(x) - C(x), \quad (8)$$

$$r(x) \triangleq \frac{d(x)}{C(x)} = \frac{\tilde{C}(x)}{C(x)} - 1, \quad (9)$$

and their derivatives are given by

$$d'(x) = \tilde{C}'(x) - C'(x), \quad (10)$$

$$r'(x) = \frac{C(x)\tilde{C}'(x) - C'(x)\tilde{C}(x)}{[C(x)]^2}, \quad (11)$$

for which

$$\tilde{C}'(x) = - \sum_{n=1}^N \frac{a_n b_n}{(x + b_n)x}, \quad (12)$$

and generally, whenever  $C(x)$  is given by (2),

$$C'(x) = - \int_0^\infty \frac{t}{(t+x)x} f_G(t) dt. \quad (13)$$

For Nakagami- $m$  and lognormal fading, (13) becomes

$$C'_m(x) = -\frac{m}{x} + m C_m(x) + \left[ \frac{\exp(mx)}{x} \times \sum_{k=0}^{m-1} k (mx)^k \Gamma(-k, mx) \right] \quad (14)$$

and

$$C'_\sigma(x) = - \int_{-\infty}^\infty \frac{\exp(\sqrt{2\sigma^2}t - \frac{\sigma^2}{2} - t^2)}{\sqrt{\pi} \left( \exp(\sqrt{2\sigma^2}t - \frac{\sigma^2}{2}) + x \right) x} dt, \quad (15)$$

respectively.

### III. NEW LOGARITHMIC APPROXIMATIONS AND BOUNDS

Inspired by Proposition 1 and by the table-book bounds restated in (6) for the Rayleigh capacity integral defined in (4), we replace the generic capacity function  $C(x)$  as well as the generic, Nakagami and lognormal capacity integrals in (2), (3) and (5), respectively, by a *weighted sum of logarithmic functions* and design appropriate values for the corresponding

coefficients. A possible choice would be to use the numerical coefficients that result from applying the numerical integration rules. However, much higher accuracy can be achieved by optimizing these coefficients in the minimax sense to give the best logarithmic approximations and bounds as will be explained soon. To begin with, we can make two minor but useful observations.

*Remark 1:* An approximation for the exponential integral,  $E_1(x)$ , is directly derived from approximating (4) by (7) as

$$E_1(x) \approx \exp(-x) \sum_{n=1}^N a_n \log_e \left( 1 + \frac{b_n}{x} \right). \quad (16)$$

Thus, the following results are applicable also beyond ergodic capacity analysis and in other fields of science than communication engineering, where the exponential integral occurs.

*Remark 2:* As originally reported in [39], the numerical evaluation of the latter form of the Rayleigh capacity integral in (4) is subject to a severe stability issue. In particular with double-precision floating-point arithmetic,  $\exp(x)$  overflows and  $E_1(x)$  underflows whenever  $x \geq 740$  although their product,  $C_1(x)$ , is finite and of the magnitude of  $1/x$  as shown by [38, Eq. 5.1.19]:  $1/(x+1) < C_1(x) < 1/x$  for all  $x > 0$ . On the other hand, all approximations and bounds according to (7) avoid this stability issue completely.

#### A. Approximations from Numerical Integration

As already mentioned, a possible choice for the parameters of (7) can be acquired by applying the Riemann sum method. However, slightly better parameter choice is achieved by applying the various quadrature numerical integration methods which are more direct and efficient to be used than the Riemann sum method. Therefore, the numerical coefficients can be easily found as given in the following three lemmas, for which the common proof given underneath holds for all, and where  $\{t_n\}_{n=1}^N$  are the nodes and  $\{w_n\}_{n=1}^N$  are the quadrature weights of the corresponding numerical integration rule [40].

*Lemma 1:* The generic capacity integral can be numerically approximated by (7) with its numerical coefficients given as

$$\{(a_n, b_n)\}_{n=1}^N = \left\{ \left( w_n f_G(t_n), t_n \right) \right\}_{n=1}^N. \quad (17)$$

*Lemma 2:* The Nakagami capacity integral can be numerically approximated as (7) with its numerical coefficients given as

$$\{(a_n, b_n)\}_{n=1}^N = \left\{ \left( w_n \frac{m^m}{\Gamma(m)} t_n^{m-1} \exp(-m t_n), t_n \right) \right\}_{n=1}^N. \quad (18)$$

*Lemma 3:* The lognormal capacity integral can be numerically approximated as (7) with its numerical coefficients given as

$$\{(a_n, b_n)\}_{n=1}^N = \left\{ \left( \frac{w_n}{\sqrt{\pi}} \exp(-t_n^2), \exp(\sqrt{2\sigma^2} t_n - \frac{\sigma^2}{2}) \right) \right\}_{n=1}^N. \quad (19)$$

*Proof:* Starting from the capacity integral expressions in Section II, we implement the quadrature numerical integration techniques, which approximate any integral of the form  $\int_u^v f(t) dt$  as a finite sum of the form  $\sum_{n=1}^N w_n f(t_n)$  for which  $f(t)$  is given in (2) for Lemma 1, in (3) for Lemma 2 and in (5) for Lemma 3. This yields the same logarithmic sum as in (7) with the numerical coefficients stated in the lemmas to approximate the respective generic, Nakagami and lognormal capacity integrals. ■

In particular, the capacity integral is an improper convergent integral that can be approximated directly by applying the Gauss–Laguerre or Gauss–Hermite quadrature rules or by considering a large yet finite integration interval with Newton–Cotes methods [38]. Another alternative way would be to use transformation of variables to limit the integration interval and thus enable the application of various other integration techniques. Nevertheless, the numerical approximations have relatively low accuracy in terms of global error and need a large number of logarithmic terms in order to achieve adequate accuracy. Therefore, we only consider the commonly used Gauss–Laguerre and Gauss–Hermite quadrature rules in the analysis of the proposed approximations in this paper.

### B. Minimax Approximations

The adopted weighted sum of logarithmic functions in (7) can be optimized to establish best minimax approximations and bounds for the generic capacity function as well as the generic, Nakagami and lognormal capacity integrals. In particular, the *best* approximation or bound refers to the member of the function family (7) that is the tightest of them all for given  $N$  and always occur with optimal set of coefficients  $\{(a_n^*, b_n^*)\}_{n=1}^N$  that minimizes the maximum error and is expressed as the solution to the following *minimax optimization problem*:

$$\{(a_n^*, b_n^*)\}_{n=1}^N = \arg \min_{\{(a_n, b_n)\}_{n=1}^N} e_{\max} \quad (20)$$

where  $e \in \{d, r\}$  represents both the absolute and relative errors collectively in what follows, and  $e_{\max}$  is the *maximum* error, which is defined as

$$e_{\max} \triangleq \sup \{|e(x)| : x > 0\} = \max \{|e_0|, |e_1|, \dots, |e_L|, |e_\infty|\}. \quad (21)$$

The latter expression comes from Fermat’s theorem, where  $e_l = e(x_l)$ ,  $l = 1, 2, \dots, L$ , are the error values at the stationary points  $x_l$ ,  $l = 1, 2, \dots, L$ , at which  $e'(x_l) = 0$ .

In the following proposition, we describe the expected shape of the solution to the minimax optimization problem in (20) that gives the best approximation or bound.

*Proposition 2:* The unique best logarithmic approximation or bound of the function family (7) with degree  $D$  for the capacity integral occurs when the corresponding error function  $e(x)$  alternates  $D$  times between  $D + 1$  extrema points of the same value of error and alternating signs. Its extreme points are found at the roots of its derivatives or asymptotically at the endpoints of its open domain.

*Proof:* According to the theorem in [41], the proposed approximation defined in (7) with  $\{(a_n^*, b_n^*)\}_{n=1}^N$  is the best minimax approximation to  $C(x)$  (including  $C_m(x)$  and  $C_\sigma(x)$ ), if and only if  $d(x)$  or  $r(x)$  defined respectively in (8) and (9), alternate  $D$  times. Moreover, the uniqueness of the solution,  $\{(a_n^*, b_n^*)\}_{n=1}^N$ , is guaranteed since the set of functions  $\{\log_e \left(1 + \frac{b_n}{x}\right), n = 1, 2, \dots, N\}$  used in the approximation in (7) satisfies the Haar condition on  $(0, \infty)$  with a null set  $\{\infty\}$ , since for every set of  $N$  distinct points  $\{x_n\}_{n=1}^N, x > 0$ , the determinant of the  $N \times N$  matrix, whose  $(i, j)$ th entry is  $\log_e \left(1 + \frac{b_i}{x_j}\right)$ , is nonzero [42]. This condition is essential to establish a unique best Chebyshev approximation [43, Theorem 1]. ■

After characterizing the shape of the minimax error function, we need to find the solution which gives such an error function. This is achieved by formulating a set of nonlinear equations and solving them as explained next.

1) *Optimization in Terms of Absolute Error:* When considering the absolute error, the best logarithmic approximation for the ergodic capacity can be found by optimizing its corresponding parameters according to (20), which implies that we seek to minimize the maximum/global error. This problem can be solved by formulating a set of nonlinear equations that describe the best absolute error function which is proved to be uniform with all its extrema points alternating in sign with the same value of error per Proposition 2.

*Corollary 1:* The best approximation in terms of the absolute error is found as the solution to the following set of equations:

$$\begin{cases} d'(x_l) = 0, & \text{for } l = 1, 2, \dots, L, \\ d(x_l) = (-1)^l d_{\max}, & \text{for } l = 1, 2, \dots, L, \\ d_0 = \lim_{x \rightarrow 0} d(x) = d_{\max}, \\ \sum_{n=1}^N a_n = 1, \end{cases} \quad (22)$$

where  $L = 2N - 1$ .

The equation  $\sum_{n=1}^N a_n = 1$  in (22) is actually a condition that is necessary to construct a bounded error function from the left, otherwise  $d_0 = \pm\infty$ . In particular, the first extrema point occurs asymptotically at zero, i.e., we choose  $x_0$  to be a very small value near zero and assign  $d_0 = d(x_0) = d_{\max}$ . Thus, when meeting the condition,  $x_0$  contributes only with a single equation that expresses the error value at that point, opposing to the other extrema points which contribute with two equations; one expresses its value and the other expresses the zero derivative of the error function at the corresponding stationary point. The absolute error is also bounded from the right, i.e.,  $d_\infty = \lim_{x \rightarrow \infty} d(x) = 0$ . Therefore, with including the imposed condition, a total of  $4N$  equations are constructed and their number is equal to the number of unknowns, namely,  $\{(a_n^*, b_n^*)\}_{n=1}^N, \{x_l\}_{l=1}^L$  and  $d_{\max}$ .

It should be noted that  $\tilde{C}(x)$  has a degree  $D = 2N$  at the optimized set of coefficients  $\{(a_n^*, b_n^*)\}_{n=1}^N$ . However, the imposed condition  $\sum_{n=1}^N a_n = 1$  decreases its degrees of freedom by one to be  $D = 2N - 1$ . Therefore,  $\tilde{C}(x)$  with  $\{(a_n^*, b_n^*)\}_{n=1}^N$  is the best Chebyshev approximation that alternates exactly  $2N - 1$  times between local maximum and minimum values of equal magnitude according to Proposition 2. This confirms exactly with the proposed approach in

(22) which alternates  $2N - 1$  times and results in a total of  $2N$  extrema points including  $x_0$ .

2) *Optimization In Terms of Relative Error:* Similar to optimizing the approximation's parameters in terms of the absolute error, the best approximation in terms of the relative error is derived by solving the minimax optimization problem in (20) through formulating a set of nonlinear equations describing the uniform minimax relative error function.

*Corollary 2:* The best approximation in terms of the relative error is found by the solution to the following set of equations:

$$\begin{cases} r'(x_l) = 0, & \text{for } l = 1, 2, \dots, L, \\ r(x_l) = (-1)^{l+1} r_{\max}, & \text{for } l = 1, 2, \dots, L, \\ r_0 = \lim_{x \rightarrow 0} r(x) = -r_{\max}, \\ \sum_{n=1}^N a_n b_n = -r_{\max} + 1. \end{cases} \quad (23)$$

In a similar way as for the absolute error, the extrema point  $x_0$  is set to be a very small value near zero and it only contributes with a single equation ( $r_0 = r(x_0) = -r_{\max}$ ). On the other hand, the relative error converges to a constant value when  $x$  tends to infinity opposing to  $d(x)$  which converges to zero, i.e.,  $r_\infty = \lim_{x \rightarrow \infty} r(x) = \sum_{n=1}^N a_n b_n - 1$  and we assign  $r_\infty = -r_{\max}$ , which results in the last equation in (23). A solution to this system of equations yields the required optimized parameters  $\{(a_n^*, b_n^*)\}_{n=1}^N$  that define the best approximation. Since no condition is imposed herein,  $D = 2N$  and, hence,  $r(x)$  alternates  $2N$  times as seen in Fig. 4(a) which confirms with the proposed approach in (23).

3) *Lower and Upper Bounds:* The proposed minimax optimization method for the logarithmic approximation in (7) can also be extended to give upper and lower bounds in terms of both absolute and relative errors. They are additionally constrained in (21) by  $e(x) \leq 0$  or  $e(x) \geq 0$  when solving for the best lower or upper bound, respectively. Therefore, we construct the lower bound by shifting down the corresponding error function in such a way as to make it oscillate between zero and  $-e_{\max}$  with  $2N$  extrema and  $e_0 = 0$  for the absolute error, and  $2N + 1$  extrema and  $e_0 = -r_{\max}$  for the relative error. With these properties of the corresponding error function, the optimization problem can be easily formulated in the same manner as in (22) and (23) for both error measures.

Similarly, using the shifting approach, the error function is forced to oscillate between zero and  $e_{\max}$  for the upper bound, resulting in an error function with  $2N - 1$  extrema and  $e_0 = d_{\max}$  for the absolute error and with  $2N + 1$  extrema and  $e_0 = 0$  for the relative error. It should be noted that for the upper bound in terms of absolute error, an extra equation  $\sum_{n=1}^N a_n b_n - 1 = 0$  is added to the system of equations in order to get an equal number of equations and unknowns. In addition, for the absolute error,  $d_\infty$  is never counted as an extremum since it converges to zero when  $x$  tends to infinity, whereas for the relative error,  $r_\infty$  is counted as an extremum since it converges to a constant value when  $x$  tends to infinity.

### C. Proof by Construction

We prove the existence of the proposed solution to (20) by construction. While the set of equations in (22) and (23)

can be directly formulated and solved for any communication system in order to find the optimized sets of coefficients  $\{(a_n^*, b_n^*)\}_{n=1}^N$  for the novel minimax approximations in (7), we have implemented the proposed methodology to find the optimized coefficients in terms of the absolute error for the Nakagami and lognormal capacity integrals which are to be used as building blocks in the capacity analysis of the more complicated systems as will be seen shortly. These coefficients are calculated by constructing (22) through substituting (7) with (3) for Nakagami capacity integral, or (5) for lognormal capacity integral, in (8) together with substituting (12) with (14) for Nakagami capacity integral, or (15) for lognormal capacity integral, in (10). Each formulated system of equations is then numerically solved using the `fsolve` command in Matlab with an equal number of equations and unknowns after using good initial guesses for the unknowns.

The coefficients  $\{(a_n^*, b_n^*)\}_{n=1}^N$  are calculated for up to  $N = 10$  or when the order of accuracy is  $10^{-9}$ , and are released to public domain in a supplementary digital file.<sup>2</sup> Likewise, we prove the existence of the solutions to (23) and the bounds by finding them for two example cases in Section V. Together with the released data sets, we also provide a basic Matlab code that implements solving (22) to calculate the optimized coefficients of (7) for any communication system in terms of the absolute error.<sup>2</sup>

Despite the simplicity of implementing this numerical approach, the challenge is to find heuristic initial guesses for the unknowns:  $\{(a_n, b_n)\}_{n=1}^N$ ,  $\{x_l\}_{l=1}^L$  and  $e_{\max}$ . In this work, we have used iteratively random values for the lower values of  $N$  and then used curve fitting techniques to draw some relationships that indicate their successive values for higher values of  $N$ . We followed this procedure to find the initial guesses for one certain value for both  $m$  and  $\sigma$  and found the optimized values of the corresponding unknowns which are then used as initial guesses for the same optimization problem but with shifted values of  $m$  and  $\sigma$  with small steps; the new optimized values are then used for the next shifted values and so on. It is worth mentioning that the numerical coefficients of  $\{(a_n, b_n)\}_{n=1}^N$  in Lemma 1 can be a very good choice as initial guesses too, especially for the lower values of  $N$ , to converge to the optimized values or at least to work as mean values around which small random variance is introduced.

## IV. APPLICATIONS OF THE PROPOSED APPROXIMATIONS AND BOUNDS

As discussed earlier, one can straightforwardly apply the methodology of Section III-B in order to solve the coefficients of (1) for any communication system. Alternatively, one can use  $\tilde{C}_m(x)$  and  $\tilde{C}_\sigma(x)$  as building blocks to derive the ergodic capacity whenever possible. Particularly in this section, we mainly focus on the second approach for which we study its important role in simplifying the complicated integrals encountered when evaluating the ergodic capacity in different communication scenarios.

A frequently seen integral in the intermediate steps when analyzing the performance of many wireless communication

<sup>2</sup>Available at <https://doi.org/10.5281/zenodo.6641977> for download.

TABLE I  
VALUES OF  $\Phi_j$ ,  $m_j$  and  $\theta_j$  FOR THE ERGODIC CAPACITIES OF SISO  
SYSTEMS UNDER DIFFERENT FADING DISTRIBUTIONS.

Fading	$\Phi_j$	$m_j$	$\theta_j$
Rice	$\frac{K^j}{j!} \exp(-K)$	$j + 1$	$\frac{1+K}{1+j}$
Nakagami- $q$ (Hoyt)	$\frac{(2j)!(1+q^2)^{-2j-1}(1-q^2)^{2j}}{j!2^{2j-1}}$	$2j + 1$	$\frac{(1+q^2)^2}{4(2j+1)q^2}$
$\eta - \mu$	$\frac{\sqrt{\pi}\Gamma(2\mu+2j)2^{-2\mu-2j+1}h^{-\mu-2j}H^{2j}}{\Gamma(\mu+j+\frac{1}{2})\Gamma(\mu)\Gamma(j+1)}$	$2\mu + 2j$	$\frac{2\mu h}{2\mu+2j}$
$\kappa - \mu$	$\frac{(\mu\kappa)^j}{j!\exp(\mu\kappa)}$	$\mu + j$	$\frac{\mu(1+\kappa)}{(\mu+j)}$

\* notes:  $K$  is the Rician factor,  $h = (2 + \eta^{-1} + \eta)/4$  and  $H = (\eta^{-1} - \eta)/4$  for Format 1 of the  $\eta - \mu$  distribution and  $h = \frac{1}{(1-\eta^2)}$  and  $H = \eta/(1-\eta^2)$  for Format 2

systems with respect to their ergodic capacity [2]–[18], [32]–[37] has a similar form to that of the Nakagami capacity integral as

$$\begin{aligned} I_{m,\phi}(x) &\triangleq \int_0^\infty \log_e(1+xt) t^{m-1} \exp(-\phi t) dt \\ &= \phi^{-m} \Gamma(m) C_m(\phi/(xm)) \\ &\approx \phi^{-m} \Gamma(m) \tilde{C}_m(\phi/(xm)). \end{aligned} \quad (24)$$

Above, the second line has been written in terms of (3) and the third line is correspondingly approximated in terms of (7).

In particular,  $\tilde{C}_m(x)$  can be used to directly approximate the ergodic capacity of a Nakagami- $m$  channel as  $\bar{C} \approx \tilde{C}_m(1/\bar{\gamma})/\log_e(2)$  including Rayleigh fading as a special case with  $m = 1$ , and  $\tilde{C}_\sigma(x)$  can be used to directly approximate the ergodic capacity of a lognormal channel as  $\bar{C} \approx \tilde{C}_\sigma(1/\bar{\gamma})/\log_e(2)$ . Next, we illustrate the use of  $\tilde{C}_m(x)$  to evaluate the ergodic capacity in different communications systems under small-scale fading, and then the use of  $\tilde{C}_\sigma(x)$  to approximate the ergodic capacity when the lognormal shadowing is introduced to the system. One can also use  $\tilde{C}_m(x)$  to evaluate the ergodic capacity of the more complicated systems that encounter a similar integral as  $I_{m,\phi}(x)$  in (24) and do not eventually result in the logarithmic expression (1); such a case is illustrated in Section IV-D.

#### A. Ergodic Capacity Under Small-Scale Fading

In addition to the Nakagami- $m$  distribution (and Rayleigh distribution thereof),  $\tilde{C}_m$  is used to approximate the capacity integral of the single-antenna systems over the more complicated distributions as

$$\bar{C} \approx \frac{1}{\log_e(2)} \sum_{j=0}^{\infty} \Phi_j \tilde{C}_m \left( \frac{\theta_j}{\bar{\gamma}} \right) \approx \sum_{n=1}^N a_n \log_2(1 + b_n \bar{\gamma}), \quad (25)$$

where  $\Phi_j$ ,  $j = 0, 1, \dots$ , are constants. Table I lists the values of  $\Phi_j$ ,  $m_j$  and  $\theta_j$  for the ergodic capacities of SISO systems under different fading distributions. It should be mentioned that the infinite series in (25) results from expanding the modified Bessel function of the first kind as a power series (which is included in the PDF of many of the fading distributions) [38, Eq. 9.6.12], and it can be truncated up to several terms that are adequate to obtain the required accuracy. The double-summation logarithmic terms (when including the approximation sum) can be rearranged into a single summation, yielding the same logarithmic approximation as in (1)

Table II lists closed-form expressions for the ergodic capacity of various point-to-point multi-antenna systems in terms of  $\tilde{C}_m(x)$ , where they usually encounter similar integrals as  $I_{m,\phi}(x)$  in (24). In particular, we consider two diversity combining techniques for SIMO, namely, maximum ratio combining (MRC) and selection combining (SC) at the receiver (RX). We also consider some MISO schemes including beamforming (BF) or distributed MISO systems with channel distribution information (CDIT) at the transmitter (TX), in addition to space time block codes (STBCs). Finally, some combined transmit–receive diversity and spatial multiplexing schemes are considered for MIMO channels.

#### B. Ergodic Capacity under Small-Scale Fading Channels with Lognormal Shadowing

Another side of novelty is that this tool enables the evaluation of the ergodic capacity for different communication systems in the presence of shadowing and results in the same logarithmic approximation as in (1). In particular, the capacity integral of a composite fading channel with  $\gamma_{\text{eff}} = \psi s$ , where  $\psi$  and  $s$  are two independent random variables representing the respective small-scale and lognormal fading, is calculated by averaging the small-scale distributed SNR over the conditional density of the lognormal-distributed conditional SNR, i.e., the average SNR of the small-scale fading is lognormally distributed, thus

$$\bar{C} = \mathbb{E}_{\gamma_{\text{eff}}}[\log_2(1 + \gamma_{\text{eff}})] = \mathbb{E}_s[\mathbb{E}_{\psi|\gamma_{\text{eff}}}\{\log_2(1 + \gamma_{\text{eff}})\}]. \quad (26)$$

The inner expectation which refers to the small-scale fading can be directly evaluated in terms of  $\tilde{C}_m(x)$  and it results in a similar expression as in (5) when considering the outer expectation which refers to the shadowing effect, for which we apply  $\tilde{C}_\sigma(x)$ .

Next, we calculate the ergodic capacity for some single-antenna and multi-antenna systems, for which we use  $\{(a_{n_1,m}, b_{n_1,m})\}_{n_1=1}^{N_1}$  to refer to the optimized coefficients of  $\tilde{C}_m$  of the Nakagami capacity integral in (3). The ergodic capacity of a Nakagami–lognormal composite fading channel can be approximated as a function of the lognormal average SNR ( $\bar{\gamma}_s$ ) as  $\bar{C} \approx \sum_{n_1=1}^{N_1} a_{n_1,m} \tilde{C}_\sigma(1/(b_{n_1,m} \bar{\gamma}_s))/\log_e(2)$  including Rayleigh fading as a special case with  $m = 1$ .

Moreover, it is calculated using Table I for the more complicated small-scale distributions with lognormal shadowing as

$$\begin{aligned} \bar{C} &\approx \frac{1}{\log_e(2)} \sum_{j=0}^{\infty} \sum_{n_1=1}^{N_1} a_{n_1,m_j} \Phi_j \tilde{C}_\sigma \left( \frac{\theta_j}{b_{n_1,m_j} \bar{\gamma}_s} \right) \\ &\approx \sum_{n=1}^N a_n \log_2(1 + b_n \bar{\gamma}_s) = \frac{1}{\log_e(2)} \tilde{C} \left( \frac{1}{\bar{\gamma}_s} \right), \end{aligned} \quad (27)$$

where the latter form occurs after applying  $\tilde{C}(x)$  twice and rearranging the triple summation into a single one with truncating the outer summation to sufficient number of terms.

In the same way as above, the ergodic capacity of some multi-antenna systems under small-scale fading and lognormal shadowing can also be approximated using  $\tilde{C}_m(x)$  and  $\tilde{C}_\sigma(x)$ . In particular, the ergodic capacity of MIMO spatial

TABLE II  
THE ERGODIC CAPACITY OF SOME MULTI-ANTENNA SYSTEMS IN TERMS OF  $\tilde{C}_m(x)$ .

Communication system	Fading	$\mathcal{C} \cdot \log_e(2)$
<b>Receiver spatial diversity (SIMO)</b> with optimal rate adaptation to channel fading with constant transmit power	Rayleigh	MRC at RX [10]: $\tilde{C}_{N_r} \left( \frac{1}{N_r \bar{\gamma}} \right)$ SC at RX [10]: $N_r \sum_{i=0}^{N_r-1} \frac{(-1)^i}{i+1} \binom{N_r-1}{i} \tilde{C}_i \left( \frac{i+1}{\bar{\gamma}} \right)$
	Nakagami- $m$	MRC at receiver [13]: $\tilde{C}_{N_r \times m} \left( \frac{1}{N_r \bar{\gamma}} \right)$
<b>Transmitter spatial diversity (MISO)</b>	Rayleigh	STBC for uncorrelated channels: $\tilde{C}_{N_t} \left( \frac{1}{N_t \bar{\gamma}} \right)$ Distributed MISO system [6, Eq. 4]: $\sum_{i=1}^M \sum_{n=1}^{K_i} a_{in} \bar{\gamma}_i^n \tilde{C}_n \left( \frac{1}{n \bar{\gamma}_i} \right)$
	Nakagami- $m$	STBC for uncorrelated channels: $R C_{m \times N_t} \left( \frac{1}{m N_t \bar{\gamma}} \right)$
	Rice	STBC for uncorrelated channels: $R \sum_{i=0}^{\infty} \frac{(N_t K)^i}{\Gamma(i+1)} \exp(-N_t K) \tilde{C}_{N_t+i} \left( \frac{1}{(N_t+i) \bar{\gamma}} \right)$ Optimum BF with CDIT at TX [17]: $\exp\left(-\frac{m \bar{\gamma}}{\sigma_V^2}\right) \sum_{i=0}^{\infty} \frac{m \bar{\gamma}^i}{i! \sigma_V^2} \tilde{C}_{i+1} \left( \frac{1}{(i+1) \sigma_V \rho} \right)  _{V=V_\theta}$
<b>Combined transmit-receive diversity (MIMO)</b>	Rayleigh	Maximum ratio transmission with MRC at RX [12]: $\sum_{k=1}^m \sum_{l=n-m}^{(n+m-2k)k} a_{k,l} \tilde{C}_{l+1} \left( \frac{k}{(l+1) \bar{\gamma}} \right)$ STBC for uncorrelated channels [3], [4]: $R \tilde{C}_{N_r \times N_t} \left( \frac{1}{N_r N_t \bar{\gamma}} \right)$ STBC for correlated channels [5]: $R \sum_{i=1}^g \sum_{j=1}^{v_g} K_{i,j} \tilde{C}_j \left( \frac{1}{j a \lambda_i} \right)$
	Nakagami- $m$	STBC for uncorrelated channels [3], [4]: $R C_{m \times N_r \times N_t} \left( \frac{1}{m N_r N_t \bar{\gamma}} \right)$
	Rice	STBC for uncorrelated channels [3], [4]: $R \sum_{i=0}^{\infty} \frac{(N_r N_t K)^i}{\Gamma(i+1)} \exp(-N_r N_t K) \tilde{C}_{N_r \times N_t+i} \left( \frac{1}{(N_r N_t+i) \bar{\gamma}} \right)$
<b>Spatial multiplexing (MIMO)</b>	Rayleigh	i.i.d. channels [8], [9]: $\sum_{z=0}^{\alpha-1} \sum_{j=0}^{2j} \frac{(-1)^j (2j)! (\beta - \alpha + i)!}{2^{2z-1} j! i! (\beta - \alpha + j)!} \binom{2z-2j}{z-j} \binom{2\beta-2\alpha+2j}{2j-i} \tilde{C}_{\beta-\alpha+i+1} \left( \frac{N_t}{(\beta - \alpha + i + 1) \rho} \right)$ Correlated channels without CSI at TX: [11, Eq. 25] with $\{\Psi_1(k)\}_{i,j} = \left( \frac{1}{\phi_j} \right)^{i-1} \Gamma(t-i+1) \tilde{C}_{t-i+1} \left( \frac{1}{\rho(t-i+1)\phi_j} \right)$ , if $i = k$ Correlated channels with partial CSI at TX: [11, Eqs. 27 and 28] with $\{\Psi_{2B}(k)\}_{i,j} = \left( \frac{1}{\phi_j} \right)^{s-i+1} \Gamma(s-i+1) \tilde{C}_{s-i+1} \left( \frac{1}{\rho(s-i+1)\phi_j} \right)$ , if $i = k$

\*notes:  $N_t$  is the number of transmit antennas,  $N_r$  is the number of receive antennas,  $\alpha = \min\{N_t, N_r\}$ ,  $\beta = \max\{N_t, N_r\}$ ,  $\rho$  is the transmit SNR,  $a_{k,l}$ ,  $M$  and  $K_i$  are defined in [6],  $a_{k,l}$  is derived in [12],  $R$  is the code rate of the STBC,  $K$  is the Rician factor,  $K_{i,j}$ ,  $a$ ,  $g$  and  $\lambda_i$  are defined in [5],  $m$ ,  $\sigma_V$ ,  $\sigma_V$  and  $V_\theta$  are defined in [17].

multiplexing over Rayleigh fading channels with lognormal shadowing [32], [33] is calculated as

$$\begin{aligned} \bar{C} &\approx \frac{1}{\log_e(2)} \sum_{z=0}^{\alpha-1} \sum_{j=0}^{2j} \sum_{i=0}^{2j} \sum_{n_1=1}^{N_1} a_{n_1, \beta - \alpha + i + 1} \\ &\times \frac{(-1)^i (2j)! (\beta - \alpha + i)!}{2^{2z-1} j! i! (\beta - \alpha + j)!} \binom{2z-2j}{z-j} \binom{2\beta-2\alpha+2j}{2j-i} \\ &\times \tilde{C}_\sigma \left( \frac{N_t}{(\beta - \alpha + i + 1) \rho b_{n_1, \beta - \alpha + i + 1} \bar{\gamma}_s} \right). \end{aligned} \quad (28)$$

Moreover, the ergodic capacity of cooperative spatial multiplexing systems with Rayleigh fading and lognormal shadowing [34] is calculated as

$$\begin{aligned} \bar{C} &\approx \frac{1}{\log_e(2)} \sum_{k=1}^{\varrho} \sum_{n_1}^{N_1} \frac{a_{n_1, N_r - \varrho + 1}}{2} \\ &\times \tilde{C}_\sigma \left( \frac{1}{(N_r - \varrho + 1) \rho_0 \Omega_{RD,k} b_{n_1, N_r - \varrho + 1}} \right), \end{aligned} \quad (29)$$

where  $\Omega_{RD,k}$  is the channel mean power for the link from the  $k$ th relay to the destination,  $\varrho$  is the number of relays and  $\rho_0$  is the average SNR per symbol.

### C. Ergodic Capacity in Recent Research Directions

After the above wide range of fundamental applications for the proposed approximations/bounds, let us proceed to

illustrate their applicability and usefulness in timely wireless systems with specific applications from the recent literature.<sup>3</sup>

In particular, the ergodic capacity (2) of downlink non-orthogonal multiple access (NOMA) system over the  $\alpha - \mu$  fading distribution [44] does not admit a similar integral as  $I_{m,\phi}(x)$  in (24) as intermediate step and, thus, we cannot use  $\tilde{C}_m(x)$  to calculate its ergodic capacity. For that, we implement the first proposed approach which means directly approximating the ergodic capacity (2) by (7). We have used the openly released Matlab code<sup>2</sup> which we have modified to make it comply with the studied system in order to find the optimized coefficients for  $\alpha = \mu = N = 2$  with two users,  $L = 2$ , ( $U_1$  and  $U_2$ ) in terms of the absolute error to approximate the ergodic capacity for both users respectively as

$$\begin{aligned} \bar{C}_{U_1} &\approx \frac{1}{\log(2)} \left[ \sum_{n=1}^2 a_n \log_2(1 + b_n \bar{\gamma}) \right. \\ &\quad \left. - \sum_{n=1}^2 a_n \log_2(1 + b_n \beta_2 \bar{\gamma}) \right], \end{aligned} \quad (30)$$

with  $\{(a_n, b_n)\}_{n=1}^2 = \{(0.336, 0.172), (0.664, 0.835)\}$ , and

$$\bar{C}_{U_2} \approx \frac{1}{\log(2)} \sum_{n=1}^2 a_n \log_2(1 + b_n \beta_2 \bar{\gamma}), \quad (31)$$

<sup>3</sup>We had to use some notations and symbols herein which are the same as in the original publications to preserve comparability, due to which some unavoidable overloading exists in this subsection compared to the rest of the article.



with  $\{(a_n, b_n)\}_{n=1}^2 = \{(0.409, 0.610), (0.591, 1.887)\}$ . The parameter  $\beta_l, l = 1, 2, \dots, L$  is the power allocation coefficient. In particular,  $\{(a_n, b_n)\}_{n=1}^N$  can be calculated for the logarithmic approximation of  $\tilde{C}_l = C_l(1/(\beta_l \bar{\gamma})) / \log_e(2)$  in [44, Eq. 46] by formulating (22) through substituting (7) and (2) to (8) together with substituting (12) and (13) to (10). The PDF  $f_G(t)$  in (2) corresponds herein to  $f_\gamma(\frac{\gamma}{\bar{\gamma}})$  in [44, Eq. 8]. These equations are then solved using the `f_solve` command in Matlab. The openly released code<sup>2</sup> can be used after modification to find the optimized coefficients for any values of  $\alpha, \mu$  and  $L$ .

On the other hand, we can derive the ergodic capacity in terms of  $\tilde{C}_m(x)$ , if the system encounters similar integral as  $I_{m,\phi}(x)$  in (24). For example, the ergodic capacity for a system with coordinated multipoint reception for mm-wave uplink with blockages and Nakagami- $m$  fading [35] can be calculated as

$$\bar{c} \approx \sum_{n=1}^N \sum_{i=1}^n \sum_{k=1}^{m_i} \frac{k^k}{\log(2)} q_n \Lambda_{n,i,k} \tilde{C}_k \left( \frac{N}{k \bar{\gamma}_i} \right), \quad (32)$$

where  $m_i$  is the Nakagami parameter of the  $i$ th link,  $N$  is the number of base stations,  $q_n$  is defined in [35, Eq. 8] and  $\Lambda_{n,i,k}$  is recursively obtained using [35, Eq. 9 and Eq. 10].

Likewise, the ergodic capacity is calculated for a mm-wave downlink NOMA system over fluctuating two-ray channels under general power allocation in [36] as

$$\begin{aligned} \bar{c} \approx & \frac{1}{\log(2)} \left[ \sum_{j_p=0}^{\infty} H_p \tilde{C}_{j_p+1} \left( \frac{1}{2\sigma_p^2(j_p+1) a Q_p \bar{\gamma}} \right) \right. \\ & + \sum_{j_q=0}^{\infty} H_p \tilde{C}_{j_q+1} \left( \frac{1}{2\sigma_q^2(j_q+1) Q_q \bar{\gamma}} \right) \\ & \left. - \sum_{j_q=0}^{\infty} H_q \tilde{C}_{j_q+1} \left( \frac{1}{2\sigma_q^2(j_q+1) a Q_q \bar{\gamma}} \right) \right], \quad (33) \end{aligned}$$

where its parameters are defined in [36]. In addition, the ergodic capacity for reflecting intelligent surface-assisted SISO system with correlated channels [37] can be approximated as

$$\bar{c} \approx \frac{1}{\log(2)} \tilde{C}_{k_a} \left( \frac{1}{k_a w_a \rho_0} \right), \quad (34)$$

where  $k_a$  and  $w_a$  are defined in [37, Eqs. 5 and 6], respectively.

#### D. Ergodic Capacity of Dual-Hop Fixed-Gain Relay Networks in Nakagami- $m$ Fading

This subsection gives an application example on the use of  $\tilde{C}_m$  in the intermediate steps when analyzing the capacity of the more complex wireless systems without necessarily resulting in the same logarithmic expression as in (1). In particular, we study the performance of a dual-hop fixed-

gain relay network under Nakagami- $m$  fading [25]. Its ergodic capacity can be approximated as

$$\begin{aligned} \bar{c} \approx & \frac{1}{\log_e(2)} \exp\left(\frac{m_1}{\bar{\gamma}_1}\right) \frac{m_1^{m_1}}{\Gamma(m_1)} \sum_{n_1=1}^{N_1} a_{n_1, m_2} \\ & \times \left[ \sum_{j=0}^{m_1-1} \sum_{n_2=1}^{N_2} a_{n_2, j+1} \binom{m_1-1}{j} m_1^{-j-1} j! \bar{\gamma}_1^{-m_1+j+1} \right. \\ & \times (-1)^{m_1-j-1} \log_e \left( 1 + \frac{(j+1) \bar{\gamma}_1 \bar{\gamma}_2 b_{n_1, m_2} b_{n_2, j+1}}{m_1 \varkappa} \right) \\ & - \sum_{i=0}^{\infty} \sum_{j=0}^{m_1-1} \binom{m_1-1}{j} \frac{m_1^i}{i!} \frac{(-1)^{m_1+j+i-1}}{j+i+1} \frac{1}{\bar{\gamma}_1^{m_1+i}} \\ & \times \left[ \left( 1 - \left( \frac{-\varkappa}{\bar{\gamma}_2 b_{n_1, m_2}} \right)^{j+i+1} \right) \log_e \left( 1 + \frac{\bar{\gamma}_2 b_{n_1, m_2}}{\varkappa} \right) \right. \\ & \left. \left. + \sum_{q=1}^{j+i+1} \frac{(-1)^q \varkappa^{q-1}}{(j+i-q+2) (\bar{\gamma}_2 b_{n_1, m_2})^{q-1}} \right] \right] \\ & - \frac{1}{\log_e(2)} \tilde{C}_{m_2} \left( \frac{\varkappa}{\bar{\gamma}_2} \right). \quad (35) \end{aligned}$$

where  $\varkappa$  is a constant defined in [45, Eq. 16],  $\bar{\gamma}_1$  and  $\bar{\gamma}_2$  are the statistical averages of the instantaneous SNRs  $\gamma_1$  and  $\gamma_2$  of the first and second hop, respectively, whose fading parameters are  $m_1$  and  $m_2$ . This expression is valid for any value of  $m_1$  opposing to [25] which is valid only for integer values of  $m_1$ . It is worth mentioning that the same expression in (35) is also obtained when evaluating the ergodic capacity under Rician, Nakagami- $q$  (Hoyt),  $\eta - \mu$  and  $\kappa - \mu$  distributions without performing individual analysis for each fading distribution. The detailed derivation of (35) is available in Appendix B.

#### E. Tractability Comparison

In this section, we illustrate the mathematical tractability of the proposed approximations and bounds and the insightful observations gained from using them for calculating the ergodic capacity of the different communication systems. For that, we consider some of the previous example applications and compare the novel analytical expressions derived herein with the corresponding expressions in the literature.

In Sections IV-A and IV-B, the capacity of the single-antenna and multi-antenna systems under small-scale fading or when combined with lognormal shadowing is evaluated using the proposed tool into the elegant simple logarithmic form in (1) which is unified for all these systems. On the contrary, it is evaluated in the literature as different complicated expressions that are unique always to the specific system under study so that a complete study and analysis are required for each system independently and using different mathematical steps.

In particular, the ergodic capacity is written in terms of the exponential integral and the incomplete gamma function in [15] for Rician fading, and in terms of the Meijer  $G$ -function in [18] for  $\kappa - \mu$  fading. In [26]–[29], the ergodic capacity is written in terms of the Gaussian  $Q$ -function or the multiplication of the complementary error function by the exponential function. On the other hand, to the best of our knowledge,

there is no available ergodic capacity analysis for the SISO system in the literature on the composite lognormal fading models and only asymptotic analysis is available in [46]. Thus, the proposed tool renders new analytical solutions that were previously deemed unattainable.

Table II evaluates the ergodic capacity of different multi-antenna system models and various fading distributions as a summation of logarithmic functions, whereas the corresponding references write them using complex functions such as the exponential integral and incomplete gamma functions. Moreover, the ergodic capacity of the multi-antenna systems under combined fading in (28) and (29) is written respectively in terms of the exponential function together with exponential integral functions in [32] and in terms of the incomplete gamma function together with the power function in [34].

The impressive advantage in terms of analytical complexity is best seen in the timely applications of Section IV-C. In particular, the ergodic capacity of the downlink NOMA over the  $\alpha - \mu$  fading in [44] and of the NOMA-based mm-wave communications in [36] are written respectively in terms of the complicated Fox  $H$ -function and the Meijer  $G$ -function which are themselves unsolvable integrals, whereas they are written in (30), (31) and (33) in the unified logarithmic form.

The importance and elegance of the proposed tool are demonstrated by its ability to provide direct or even visual insights into the system's performance opposing to expressions that comprise special functions. For example, from Table II, we can immediately see that the ergodic capacity improves with increasing  $N_r$  and  $\bar{\gamma}$  for the SIMO systems, whereas it improves with increasing  $N_t$ ,  $\bar{\gamma}$ ,  $m$  under Nakagami- $m$  fading,  $K$  under Rician fading and  $R$  of the STBC, all for the MISO systems. On the other hand, none of these observations can be concluded from the corresponding complicated expressions in the literature.

## V. NUMERICAL RESULTS AND DISCUSSION

This section demonstrates the accuracy of the proposed approximations and bounds while the actual behavior of the corresponding systems has already been analyzed extensively in the references. In particular, we compare them with previously derived ones, in addition to the numerical approximations obtained by applying Gauss–Laguerre and Gauss–Hermite quadrature rules for the Nakagami and the lognormal capacity integrals, respectively. Furthermore, we validate and compare some of the application examples presented in Section IV with those obtained from the numerical and existing approximations and bounds.

Let us begin with plotting the global absolute error,  $d_{\max}$ , for the Nakagami capacity integral in Fig. 2(a) for different values of  $m$ , using our approximations and the numerical approximation resulting from applying the Gauss–Laguerre quadrature rule. It is clearly realized from the figure that our minimax coefficients result in much more accurate logarithmic approximations in terms of the global error than those resulting from numerical integration. Moreover, as the number of terms increases, the accuracy increases substantially, especially for higher values of  $m$ . We further verify the accuracy of the

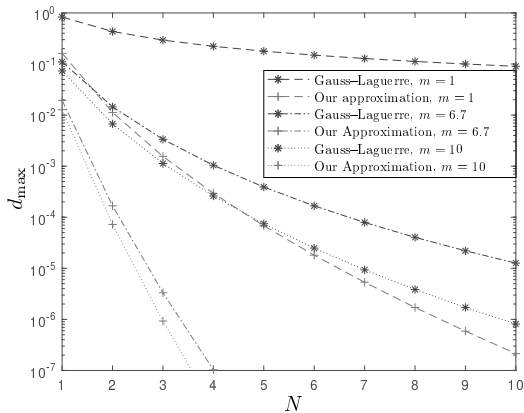
proposed approximation by comparing its absolute error with that of the Gauss–Laguerre approximation for the whole considered range of the argument in Fig. 2(b). Obviously, our optimized coefficients not only have the least global error, but they also achieve higher accuracy for most of the considered range of the argument for the different values of  $m$ .

Moreover, the same comparisons are made for the lognormal capacity integral for different values of  $\sigma$ . Our approximations are compared with those obtained using the Gauss–Hermite quadrature rule which has the same logarithmic form, in addition to the existing approximations which encounter very complicated functions such as the complementary Gaussian error function and the trigonometric functions [26]–[29]. The proposed approximations mostly outperform all the other ones in terms of the global error as depicted in Fig. 3(a). They also have comparable or even better accuracy than those with the very complex form over the whole considered range of the argument as seen in Fig. 3(b) despite their significantly simpler form.

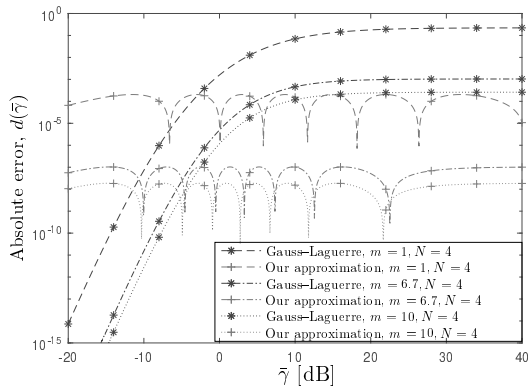
The minimax optimization method is not only used for constructing the approximations in terms of the absolute error but also for the approximations in terms of relative error as explained in Section III-B2, and for the lower and upper bounds in terms of both error measures as explained in Section III-B3. The approximation for the special case of Rayleigh capacity integral is optimized in terms of the relative error for  $N = 3$  and the corresponding relative error function is plotted in Fig 4(a), whereas in Fig. 4(b), we plot the uniform absolute error functions resulting from the optimized upper and lower bounds of the Nakagami capacity integral for  $m = 3$  and  $N = 3$ . As expected, the resulting error functions oscillate uniformly and achieve high accuracy. We can conclude from Figs. 2, 3 and 4 that the proposed approximations with the optimized coefficients achieve significant improvement in accuracy by several orders of magnitude when compared to the numerical and existing approximation. The absolute and relative errors are so small that they are virtually exact with the actual capacity measures.

Next, we numerically investigate some of the applications of the proposed approximations which are included in Section IV. In Fig. 5, the ergodic capacity for Rician fading channel with lognormal shadowing is studied and its absolute error is plotted for different values of the Rician factor using three approaches, namely, (i) Gauss–Laguerre and Gauss–Hermite rules respectively, (ii) using (3) for the small-scale stage and then Gauss–Hermite rule for the shadowing stage and finally (iii) using (27) with the necessary coefficients from Table I. We can observe that approach (iii) results in a tighter approximation than that of approach (i) which has exactly the same analytical form. It also has the same accuracy as that of approach (ii).

Figure 6 illustrates the error resulting from applying our approximation to evaluate the ergodic capacity in  $2 \times 2$  MIMO network over shadowed-Rayleigh channel as in (28), and compares it with the theoretical results presented in [32], [33]. Our optimized coefficients yield significantly higher accuracy than those of [33], having exactly the same logarithmic form and number of terms. Despite the simplicity of our approximation's



(a)

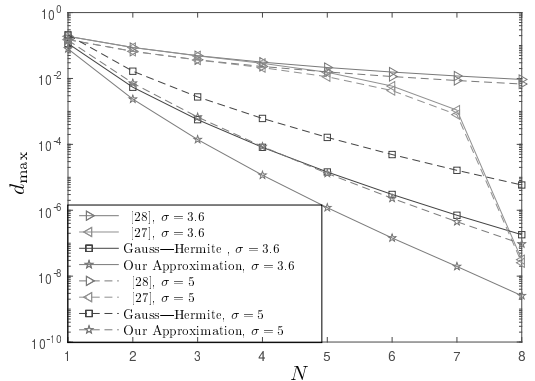


(b)

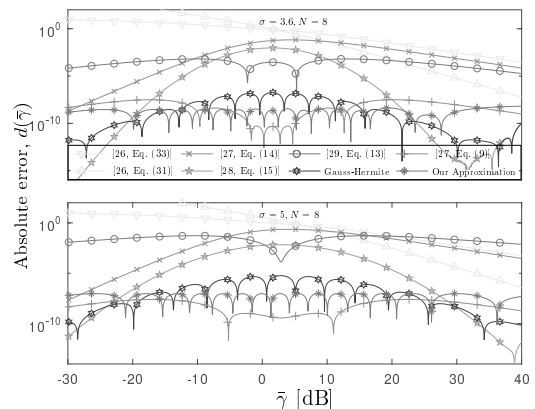
Fig. 2. (a) Comparison between our approximations and those obtained using Gauss-Laguerre for the Nakagami capacity integral with different values of  $m$  in terms of global absolute error. (b) Same as (a) but in terms of the absolute error function over the whole considered range of the argument with  $N = 4$ .

analytical form compared to that of [32], it achieves higher accuracy over a wide range of the argument. The ergodic capacity of both users, (30) and (31), is plotted in Fig. 7 along with the exact capacity derived in [44] for different selections of power allocation coefficients. The figure shows virtually exact match between the logarithmic approximation and the exact results with only two logarithmic terms ( $N = 2$ ).

In Fig. 8, the absolute error of the ergodic capacity in a dual-hop cooperative system is plotted as a function of the average SNR of each hop, where we considered  $\bar{\gamma}_1 = \bar{\gamma}_2$ . It is clear from the figure that the ergodic capacity resulting from applying our approximation is extremely accurate. In particular, the mathematical form of the ergodic capacity in (35) is not only much more tractable than that in [24] and [25], but also its accuracy outperforms [24] for the lower and moderate values, when considering Rayleigh fading channels,



(a)



(b)

Fig. 3. (a) Comparison between our approximations and those obtained using Gauss-Hermite and the existing approximations for the lognormal capacity integral with different values of  $\sigma$  in terms of global absolute error. (b) Same as (a) but in terms of the absolute error function over the whole considered range of the argument with  $N = 8$ .

and outperforms [25] over the whole range of the argument when considering Nakagami- $m$  fading channels, with less error by three orders of magnitude.

## VI. CONCLUSIONS

This paper presented an accurate and efficient tool for facilitating statistical performance analysis in different wireless communication systems in terms of ergodic capacity. A novel systematic methodology was also developed in order to optimize its accuracy in the minimax sense. This tool was applied to a wide range of fundamental and recent applications, including single-antenna and multi-antenna systems under small-scale fading and with or without lognormal shadowing in order to derive tractable closed-form expressions for the ergodic capacity. We validated the tightness of the proposed

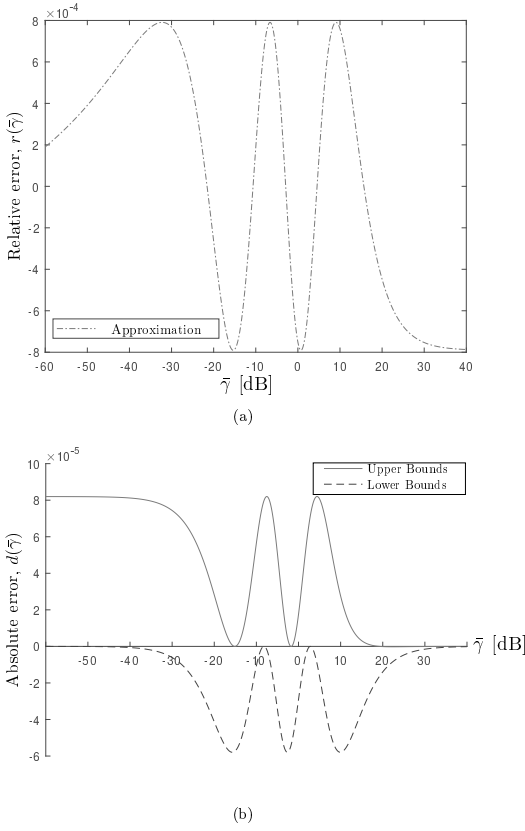


Fig. 4. (a) The optimized approximation in terms of relative error for Rayleigh capacity integral with  $N = 3$ . (b) The optimized upper and lower bounds in terms of the absolute error for Nakagami capacity integral with  $m = 3$  and  $N = 3$ .

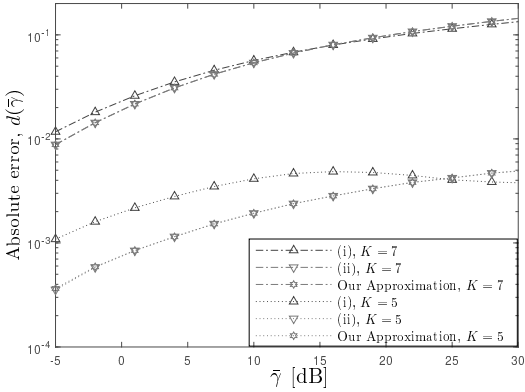


Fig. 5. Comparison of the absolute error function of the proposed approximations and the numerical ones for shadowed-Rician network with  $N_1 = 4$ ,  $N_2 = 6$ ,  $\sigma = 4$ , and different values of  $K$ .

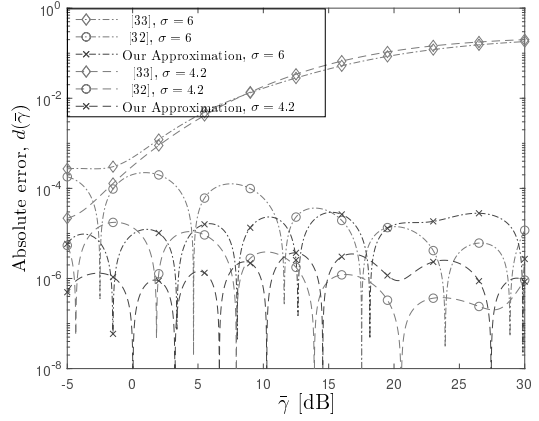


Fig. 6. Comparison of the absolute error function of the proposed approximations and the existing ones for  $2 \times 2$  MIMO network with  $N_1 = 7$ ,  $N_2 = 5$ ,  $\rho = 1$  dB and different levels of shadowing.

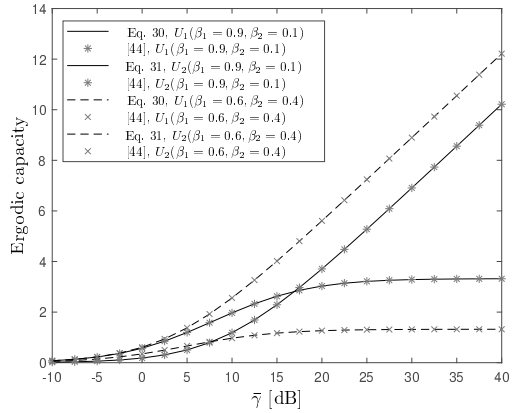


Fig. 7. Ergodic capacity for two NOMA users with  $\alpha = \mu = 2$  and for different values of  $\beta_2$ .

tool by numerical comparisons with existing and numerical ones, in which our tool showed significant improvement in the accuracy by several orders of magnitude.

#### APPENDIX A PROOF OF PROPOSITION 1

Denoting that the PDF of instantaneous capacity  $\mathcal{C}$  is given by  $f_{\mathcal{C}}(c)$ , the ergodic capacity is calculated as

$$\begin{aligned} \mathbb{E}[\mathcal{C}] &= \int_0^{\infty} c f_{\mathcal{C}}(c) dc \\ &= \int_0^{\infty} \underbrace{\frac{\bar{\gamma}_{\text{eff}} f_{\mathcal{C}}(\log_2(1 + \bar{\gamma}_{\text{eff}} g))}{\log_e(2) (1 + \bar{\gamma}_{\text{eff}} g)}}_{\triangleq f_{\mathcal{C}}(g)} \log_2(1 + \bar{\gamma}_{\text{eff}} g) dg, \end{aligned} \quad (36)$$

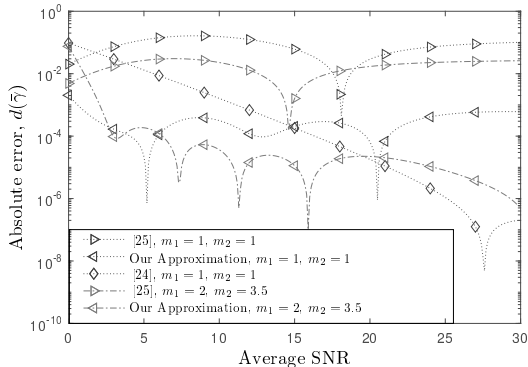


Fig. 8. The absolute error of the ergodic capacity in a dual-hop cooperative system with  $N_1 = N_2 = 4$ .

where the second expression is obtained by changing the integration variable to  $g \triangleq (2^c - 1)/\bar{\gamma}_{\text{eff}}$ . Next, we implement the Riemann sum method to approximate the above integral by truncating it and dividing the integration interval into  $N$  partitions, each of length  $\delta$ . Therefore, the ergodic capacity can be approximated by a finite sum of logarithmic functions according to (1) by choosing

$$a_n \triangleq \frac{\delta \bar{\gamma}_{\text{eff}} f_C(\log_2(1 + \bar{\gamma}_{\text{eff}} n \delta))}{\log_e(2) (1 + \bar{\gamma}_{\text{eff}} n \delta)} \quad \text{and} \quad b_n \triangleq n \delta, \quad (37)$$

while arbitrary accuracy can be achieved when  $\delta \rightarrow 0$  and  $N \rightarrow \infty$ . Furthermore, by applying appropriately the left, intermediate or the right rule for each partition, one can always guarantee that  $\sum_{n=1}^N a_n \leq 1$ , since each  $a_n$  represents a part of the total probability mass of random variable  $G \triangleq (2^c - 1)/\bar{\gamma}_{\text{eff}}$ , whose PDF is denoted by  $f_G(g)$ .

In the general case without making specific assumptions about the distribution of  $\mathcal{C}$ , coefficients  $a_n$ ,  $n = 1, 2, \dots, N$ , will depend on  $\bar{\gamma}_{\text{eff}}$ , which would make (1) an inconvenient approximation for the statistical analysis of specific systems. However, we can express  $f_C(c)$  in terms of  $f_G(g)$  as

$$f_C(c) = \frac{2^c \log_e(2)}{\bar{\gamma}_{\text{eff}}} f_G\left(\frac{2^c - 1}{\bar{\gamma}_{\text{eff}}}\right), \quad (38)$$

which results, by substitution into (37), in  $a_n = \delta f_G(n \delta)$ . Thus, whenever  $f_G(g)$  is independent of  $\bar{\gamma}_{\text{eff}}$ ,  $a_n$  becomes independent of  $\bar{\gamma}_{\text{eff}}$  too, and the same approximation (i.e., the same coefficients) can be conveniently applied with any value of  $\bar{\gamma}_{\text{eff}}$ . This condition is not very restrictive in practice, and it is satisfied in the applications discussed in this article.

## APPENDIX B

### DERIVATION OF (35) FOR DUAL-HOP FIXED-GAIN RELAY NETWORKS UNDER NAKAGAMI FADING

From [24], the end-to-end SNR herein is  $\gamma_e \triangleq \frac{\gamma_1 \gamma_2}{\varkappa + \gamma_2}$  and the ergodic capacity is calculated as

$$\bar{c} = \underbrace{\frac{1}{\log_e(2)} \mathbb{E} \left[ \log_e \left( 1 + \frac{(1 + \gamma_1) \gamma_2}{\varkappa} \right) \right]}_A - \underbrace{\frac{1}{\log_e(2)} \mathbb{E} \left[ \log_e \left( 1 + \frac{\gamma_2}{\varkappa} \right) \right]}_B. \quad (39)$$

We will consider Nakagami- $m$  fading channels. Part  $B$  of (39) can be directly approximated using our logarithmic approximation with the optimized parameters  $\{(a_{n_1, m_2}, b_{n_1, m_2})\}_{n_1=1}^{N_1}$  as

$$B = \frac{1}{\log_e(2)} \tilde{C}_{m_2} \left( \frac{\varkappa}{\bar{\gamma}_2} \right), \quad (40)$$

whereas part  $A$  is evaluated as

$$\begin{aligned} A &= \frac{1}{\log_e(2)} \int_0^\infty \left( \frac{m_1}{\bar{\gamma}_1} \right)^{m_1} \frac{\gamma_1^{m_1-1}}{\Gamma(m_1)} \exp\left(-m_1 \frac{\gamma_1}{\bar{\gamma}_1}\right) \\ &\quad \times \int_0^\infty \log_e \left( 1 + \frac{(1 + \gamma_1) \gamma_2}{\varkappa} \right) \left( \frac{m_2}{\bar{\gamma}_2} \right)^{m_2} \frac{\gamma_2^{m_2-1}}{\Gamma(m_2)} \\ &\quad \times \exp\left(-m_2 \frac{\gamma_2}{\bar{\gamma}_2}\right) d\gamma_2 d\gamma_1. \end{aligned} \quad (41)$$

We approximate the inner integral which is of the form  $I_{m_2, \frac{m_2}{\bar{\gamma}_2}} \left( \frac{1+\gamma_1}{\varkappa} \right)$  using (24). Therefore, (41) becomes

$$\begin{aligned} A &= \frac{1}{\log_e(2)} \sum_{n_1=1}^{N_1} a_{n_1, m_2} \int_0^\infty \log_e \left( 1 + \frac{(1 + \gamma_1) \bar{\gamma}_2 b_{n_1, m_2}}{\varkappa} \right) \\ &\quad \times \left( \frac{m_1}{\bar{\gamma}_1} \right)^{m_1} \frac{\gamma_1^{m_1-1}}{\Gamma(m_1)} \exp\left(-m_1 \frac{\gamma_1}{\bar{\gamma}_1}\right) d\gamma_1. \end{aligned} \quad (42)$$

Using change of variables  $z = \frac{1+\gamma_1}{\bar{\gamma}_1}$ , we obtain

$$\begin{aligned} A &= \frac{1}{\log_e(2)} \frac{m_1^{m_1}}{\Gamma(m_1)} \sum_{n_1=1}^{N_1} a_{n_1, m_2} \\ &\quad \times \left[ \int_0^\infty P_1(z) dz - \int_0^{1/\bar{\gamma}_1} P_1(z) dz \right], \end{aligned} \quad (43)$$

where

$$\begin{aligned} P_1(z) &= \log_e \left( 1 + \frac{\bar{\gamma}_1 \bar{\gamma}_2 b_{n_1, m_2} z}{\varkappa} \right) \bar{\gamma}_1^{-m_1+1} \\ &\quad \times (\bar{\gamma}_1 z - 1)^{m_1-1} \exp\left(-m_1 z + \frac{m_1}{\bar{\gamma}_1}\right). \end{aligned} \quad (44)$$

Next, we expand  $(\bar{\gamma}_1 z - 1)^{m_1-1}$  using the binomial theorem, and approximate the resulting expression

which contains  $I_{j+1, m_1} \left( \frac{\tilde{\gamma}_1 \tilde{\gamma}_2 b_{n_1, m_2}}{\varkappa} \right)$  using (24) with  $\{(a_{n_2, j+1}, b_{n_2, j+1})\}_{n_2=1}^{N_2}$  to evaluate  $A_1 = \int_0^\infty P_1(z) dz$  as

$$A_1 = \exp \left( \frac{m_1}{\tilde{\gamma}_1} \right) \sum_{j=0}^{m_1-1} \sum_{n_2=1}^{N_2} a_{n_2, j+1} \binom{m_1-1}{j} \\ \times m_1^{-j-1} j! \tilde{\gamma}_1^{-m_1+j+1} (-1)^{m_1-j-1} \\ \times \log_e \left( 1 + \frac{(j+1)\tilde{\gamma}_1 \tilde{\gamma}_2 b_{n_1, m_2} b_{n_2, j+1}}{m_1 \varkappa} \right), \quad (45)$$

whereas for  $A_2 = \int_0^{1/\tilde{\gamma}_1} P_1(z) dz$ , we apply [47, Eqs. 1.110, 1.211.1, 2.729] as

$$A_2 = \exp \left( \frac{m_1}{\tilde{\gamma}_1} \right) \sum_{i=0}^\infty \sum_{j=0}^{m_1-1} \binom{m_1-1}{j} \frac{m_1^i}{i!} (-1)^{m_1+j+i-1} \\ \tilde{\gamma}_1^{-m_1+j+1} \int_0^{1/\tilde{\gamma}_1} z^{j+i} \log_e \left( 1 + \frac{\tilde{\gamma}_1 \tilde{\gamma}_2 b_{n_1, m_2} z}{\varkappa} \right) dz \\ = \exp \left( \frac{m_1}{\tilde{\gamma}_1} \right) \sum_{i=0}^\infty \sum_{j=0}^{m_1-1} \binom{m_1-1}{j} \frac{(m_1)^i}{i!} \frac{(-1)^{m_1+j+i-1}}{j+i+1} \\ \frac{1}{\tilde{\gamma}_1^{m_1+i}} \left[ \left( 1 - \left( \frac{-\varkappa}{\tilde{\gamma}_2 b_{n_1, m_2}} \right)^{j+i+1} \right) \log_e \left( 1 + \frac{\tilde{\gamma}_2 b_{n_1, m_2}}{\varkappa} \right) \right. \\ \left. + \sum_{q=1}^{j+i+1} \frac{(-1)^q \varkappa^q q^{-1}}{(j+i-q+2) (\tilde{\gamma}_2 b_{n_1, m_2})^{q-1}} \right]. \quad (46)$$

We substitute (45) and (46) back into (43) which then are substituted together with (40) into (39) to obtain a closed-form approximation for the ergodic capacity in a dual-hop fixed-gain relay networks over Nakagami- $m$  fading channels according to (35).

## REFERENCES

- [1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [2] W. C. Y. Lee, "Estimate of channel capacity in Rayleigh fading environment," *IEEE Transactions on Vehicular Technology*, vol. 39, no. 3, pp. 187–189, Aug. 1990.
- [3] H. Zhang and T. Gulliver, "Capacity and error probability analysis for space-time block codes over fading channels," in *Proc. IEEE Pacific Rim Conference on Communications Computers and Signal Processing*, vol. 1, Aug. 2003, pp. 102–105.
- [4] H. Zhang and T. A. Gulliver, "Closed form capacity expressions for space time block codes over fading channels," in *Proc. International Symposium on Information Theory*, Jun. 2004, p. 411.
- [5] L. Musavian, M. Nakhai, and A. Aghvami, "Capacity of space time block codes with adaptive transmission in correlated Rayleigh fading channels," in *Proc. IEEE Vehicular Technology Conference*, vol. 3, May 2006, pp. 1511–1515.
- [6] D. Castanheira and A. Gameiro, "Distributed MISO system capacity over Rayleigh flat fading channels," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Sep. 2008, pp. 1–5.
- [7] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, Jun. 2003.
- [8] I. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–596, Sep. 1999.
- [9] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering, and keyhole," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2636–2647, Oct. 2003.
- [10] M.-S. Alouini and A. J. Goldsmith, "Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 4, pp. 1165–1181, Jul. 1999.
- [11] M. Kang and M.-S. Alouini, "Capacity of correlated MIMO Rayleigh channels," *IEEE Transactions on Wireless Communications*, vol. 5, no. 1, pp. 143–155, Jan. 2006.
- [12] A. Maaref and S. Aissa, "Closed-form expressions for the outage and ergodic Shannon capacity of MIMO MRC systems," *IEEE Transactions on Communications*, vol. 53, no. 7, pp. 1092–1095, Jul. 2005.
- [13] M.-S. Alouini and A. Goldsmith, "Capacity of Nakagami multipath fading channels," in *Proc. IEEE Vehicular Technology Conference*, vol. 1, May 1997, pp. 358–362.
- [14] G. Fraidenraich, O. Leveque, and J. M. Cioffi, "On the MIMO channel capacity for the Nakagami- $m$  channel," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3752–3757, Aug. 2008.
- [15] I.-S. Koh and T. Hwang, "Simple expression of ergodic capacity for Rician fading channel," *IEICE Transactions on Communications*, vol. 93-B, no. 6, pp. 1594–1596, Jul. 2010.
- [16] M. Kang and M.-S. Alouini, "Capacity of MIMO Rician channels," *IEEE Transactions on Wireless Communications*, vol. 5, no. 1, pp. 112–122, Jan. 2006.
- [17] D. E. Kontaxis, G. V. Tsoulos, and S. Karaboyas, "Ergodic capacity optimization for single-stream beamforming transmission in MISO Rician fading channels," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 2, pp. 628–641, Feb. 2013.
- [18] C. García-Corrales, F. J. Cañete Corripio, and J. Paris, "Capacity of  $\kappa - \mu$  shadowed fading channels," *International Journal of Antennas and Propagation*, vol. 2014, pp. 1–8, Jul. 2014.
- [19] O. Oyman, R. U. Nabar, H. Bolcskei, and A. J. Paulraj, "Tight lower bounds on the ergodic capacity of Rayleigh fading MIMO channels," in *Proc. Global Telecommunications Conference*, vol. 2, Nov. 2002, pp. 1172–1176.
- [20] E. Gauthier, A. Yongacoglu, and J.-Y. Chouinard, "Capacity of multiple antenna systems in Rayleigh fading channels," in *Proc. Canadian Conference on Electrical and Computer Engineering*, vol. 1, May 2000, pp. 275–279.
- [21] G. Alex, "Rayleigh fading multi-antenna channels," *EURASIP Journal on Advances in Signal Processing*, vol. 2002, no. 3, p. 316–329, Mar. 2002.
- [22] M. Dohler and H. Aghvami, "On the approximation of MIMO capacity," *IEEE Transactions on Wireless Communications*, vol. 4, no. 1, pp. 30–34, Jan. 2005.
- [23] B. Banerjee, A. Abu Al Haija, C. Tellambura, and H. A. Suraweera, "Simple and accurate low SNR ergodic capacity approximations," *IEEE Communications Letters*, vol. 22, no. 2, pp. 356–359, Feb. 2018.
- [24] O. Waqar, M. Ghogho, and D. McLernon, "Tight bounds for ergodic capacity of dual-hop fixed-gain relay networks under Rayleigh fading," *IEEE Communications Letters*, vol. 15, no. 4, pp. 413–415, Apr. 2011.
- [25] D. B. da Costa and S. Aissa, "Capacity analysis of cooperative systems with relay selection in Nakagami- $m$  fading," *IEEE Communications Letters*, vol. 13, no. 9, pp. 637–639, Sep. 2009.
- [26] M.-S. Alouini and A. J. Goldsmith, "Area spectral efficiency of cellular mobile radio systems," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 4, pp. 1047–1066, Jul. 1999.
- [27] A. Laourine, A. Stephenne, and S. Affes, "Estimating the ergodic capacity of log-normal channels," *IEEE Communications Letters*, vol. 11, no. 7, pp. 568–570, Jul. 2007.
- [28] A. Laourine, A. Stephenne, and S. Affes, "On the capacity of log-normal fading channels," *IEEE Transactions on Communications*, vol. 57, no. 6, pp. 1603–1607, Jun. 2009.
- [29] F. Heliot, X. Chu, R. Hoshyar, and R. Tafazolli, "A tight closed-form approximation of the log-normal fading channel capacity," *IEEE Transactions on Wireless Communications*, vol. 8, no. 6, pp. 2842–2847, Jun. 2009.
- [30] G. Pan, E. Ekici, and Q. Feng, "Capacity analysis of log-normal channels under various adaptive transmission schemes," *IEEE Communications Letters*, vol. 16, no. 3, pp. 346–348, Mar. 2012.
- [31] I. M. Tanash and T. Riihonen, "Global minimax approximations and bounds for the Gaussian  $Q$ -function by sums of exponentials," *IEEE Transactions on Communications*, vol. 68, no. 10, pp. 6514–6524, Oct. 2020.
- [32] L. Yang, "On the capacity of MIMO Rayleigh fading channels with log-normal shadowing," in *Proc. Congress on Image and Signal Processing*, vol. 5, May 2008, pp. 479–482.

- [33] D. Shen, A. Lu, Y. Cui, F. Kuang, X. Zhang, K. Wu, and J. Yao, "On the channel capacity of MIMO Rayleigh-lognormal fading channel," in *Proc. International Conference on Microwave and Millimeter Wave Technology*, May 2010, pp. 156–159.
- [34] T. Q. Duong and H.-J. Zepernick, "On the ergodic capacity of cooperative spatial multiplexing systems in composite channels," in *Proc. IEEE Radio and Wireless Symposium*, Jan. 2009, pp. 175–178.
- [35] B. Maham and P. Popovski, "Capacity analysis of coordinated multipoint reception for mmwave uplink with blockages," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 16 299–16 303, Dec. 2020.
- [36] Y. Tian, G. Pan, and M.-S. Alouini, "On NOMA-based mmwave communications," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 15 398–15 411, Dec. 2020.
- [37] T. Van Chien, A. K. Papazafeiropoulos, L. T. Tu, R. Chopra, S. Chatzinothas, and B. Ottersten, "Outage probability analysis of IRS-assisted systems under spatially correlated channels," *IEEE Wireless Communications Letters*, vol. 10, no. 8, pp. 1815–1819, Aug. 2021.
- [38] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications, 1972.
- [39] T. Riihonen, S. Werner, and R. Wichman, "Comments on 'Simple formulas for SIMO and MISO ergodic capacities,'" *Electronics Letters*, vol. 48, no. 2, pp. 127–127, Jan. 2012.
- [40] P. Davis and P. Rabinowitz, *Methods of Numerical Integration*, 2nd ed. Academic Press, 1984.
- [41] C. B. Dunham, "Chebyshev approximation by logarithmic families," *Zeitschrift Angewandte Mathematik und Mechanik*, vol. 53, no. 5, pp. 352–353, Jan. 1973.
- [42] E. W. Cheney, *Introduction to Approximation Theory*. New York McGraw-Hill, 1966.
- [43] C. Dunham, "Families satisfying the Haar condition," *Journal of Approximation Theory*, vol. 12, pp. 291–298, Nov. 1974.
- [44] A. Alqahtani, E. Alsusa, A. Al-Dweik, and M. Al-Jarrah, "Performance analysis for downlink NOMA over  $\alpha - \mu$  generalized fading channels," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 7, pp. 6814–6825, Jul. 2021.
- [45] M. Hasna and M.-S. Alouini, "A performance study of dual-hop transmissions with fixed gain relays," *IEEE Transactions on Wireless Communications*, vol. 3, no. 6, pp. 1963–1968, Nov. 2004.
- [46] I. Ansari and M.-S. Alouini, "Asymptotic ergodic capacity analysis of composite lognormal shadowed channels," in *Proc. IEEE Vehicular Technology Conference*, Jul. 2015.
- [47] I. Gradshteyn and I. Ryzhik, *Table of integrals, series, and products*, 7th ed. Elsevier/Academic Press, 2007.



**Islam M. Tanash** received the B.Sc. and M.Sc. degrees in electrical engineering from Jordan University of Science and Technology (JUST), Irbid, Jordan in 2014 and 2016, respectively. She is currently a PhD student and doctoral researcher at the Faculty of Information Technology and Communication Sciences, Tampere University, Finland. She was selected among the best 200 young researchers to participate in the Seventh Heidelberg Laureate Forum (HLF) in 2019. She received the Huawei Best PhD Student Paper Award in 2020. Her research

interests include the areas of communications theory, wireless networks, and wireless systems security.



**Taneli Riihonen** (S'06–M'14–SM'22) received the D.Sc. degree in electrical engineering (with distinction) from Aalto University, Helsinki, Finland, in August 2014. He is currently an Associate Professor (tenure track) at the Faculty of Information Technology and Communication Sciences, Tampere University, Finland. He held various research positions at Aalto University School of Electrical Engineering from September 2005 through December 2017. He was a Visiting Associate Research Scientist and an Adjunct Assistant Professor at Columbia University in the City of New York, USA, from November 2014 through December 2015. He has been nominated twelve times as an Exemplary/Top Reviewer/Editor of various IEEE journals and is serving as an Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS since June 2022. He has previously served as an Editor for IEEE COMMUNICATIONS LETTERS and IEEE WIRELESS COMMUNICATIONS LETTERS. He received the Finnish technical sector's award for the best doctoral dissertation of the year and the EURASIP Best PhD Thesis Award 2017. His research activity is focused on physical-layer OFDM(A), multiantenna, relaying and full-duplex wireless techniques with current interest in the evolution of beyond 5G systems.





# PUBLICATION

7

**Ergodic capacity analysis of RIS-aided systems with spatially correlated channels**

I. M. Tanash and T. Riihonen

In *Proc. IEEE International Conference on Communications (ICC)*, May 2022,  
pp. 3293–3298

DOI: 10.1109/ICC45855.2022.9839244

**Publication reprinted with the permission of the copyright holders.**



# Ergodic Capacity Analysis of RIS-Aided Systems with Spatially Correlated Channels

Islam M. Tanash and Taneli Riihonen

Faculty of Information Technology and Communication Sciences, Tampere University, Finland

e-mail: {islam.tanash, taneli.riihonen}@tuni.fi

**Abstract**—This paper investigates the ergodic capacity of reflecting intelligent surface (RIS)-aided single-input single-output communication systems with spatially correlated Rayleigh-fading channels. The ergodic capacity for such systems does not admit an exact closed-form expression. Therefore, we consider two alternative fading distributions to approximate the systems' statistical characterization to enable the derivation of closed-form expressions for the ergodic capacity. We further simplify the ergodic capacity by proposing novel and unified approximations in the form of a weighted sum of logarithmic functions with optimized coefficients. We validate the effectiveness and the high accuracy of the adopted schemes and the proposed approximations through numerical results. Performance analysis to study the impact of several system parameters on the ergodic capacity is also conducted. Deploying an RIS to the communication system can significantly increase the ergodic capacity which increases even further with increasing the number of reflecting elements equipped on the RIS, and this effect is best seen when the direct path is weak.

## I. INTRODUCTION

The reconfigurable intelligent surfaces (RISs) have emerged recently as a promising technology to intelligently control the wireless environment and significantly improve the spectral and energy efficiency thereof. More specifically, an RIS is a large metasurface that consists of low-cost, sub-wavelength-sized passive reflecting elements (REs) that serve as nearly isotropic scatterers and can be adjusted via a microcontroller to collaboratively steer the incident electromagnetic signals into the desired direction.

This topic has attracted many research efforts from both the academia and industry to extensively study the design [1], [2], optimization [3]–[5] and the potential applications [6], [7] of the RIS-aided systems. A considerable amount of theoretical studies has also been conducted in the literature to analyze their different performance measures [8]–[13]. The majority of the existing research assumes independent and identically distributed (i.i.d.) fading channels [8]–[12]. However, independence does not represent a realistic assumption as has been shown by Björnson and Sanguinetti in [14], who also introduce a more realistic spatially correlated Rayleigh fading system model, which will be adopted for our current work too.

Moreover, limited analytical assessment of the ergodic capacity, which refers to the upper rate at which information can be reliably transmitted over a time varying channel, has been reported in the literature for the RIS-aided systems. In fact, as far as we are aware of, the ergodic capacity in single-input single-output (SISO) RIS-aided systems has been investigated

only for the case of i.i.d. fading channels in [9]–[11], leaving the case with spatial correlation unstudied yet. Motivated by these facts, we present herein a more realistic performance study in terms of the ergodic capacity for a SISO system model with direct path and correlated Rayleigh fading channels.

In particular, since the exact statistical characterization of the end-to-end equivalent channel of the SISO system with direct path and spatially correlated channels is unknown, we adopt two distributions to approximate it, namely the non-central chi-square and the Gamma distribution. This leads us toward deriving closed-form expressions, which are exact in respect to the adopted distribution, for the ergodic capacity. Nevertheless, in order to simplify the ergodic capacity even further and provide additional engineering insight into it, we present tractable and tight approximations for the ergodic capacity in the form of a weighted sum of logarithmic functions with optimally choosing the corresponding coefficients.

We verify the presented expressions by Monte Carlo simulations that illustrate an excellent match between the adopted distributions and the exact channel statistics. The high accuracy of the proposed logarithmic approximation, whose accuracy increases with increasing number of terms in the summation, is also confirmed. Numerical results show that the ergodic capacity depends on multiple factors, namely, the transmitted power, number of REs equipped on the RIS and the large-scale coefficient of the direct path, where their increase result in improving the ergodic capacity significantly.

## II. SYSTEM AND CHANNEL MODELS

The system under study is illustrated in Fig. 1 and it consists of a single-antenna source (S), a single-antenna destination (D) and a two-dimensional RIS equipped with  $M = M_H \times M_V$  REs, where  $M_H$  is the number of REs per row and  $M_V$  is the number of REs per column. Each RE have an area of  $\Lambda = d_H \times d_V$ , where  $d_H$  is its horizontal width and  $d_V$  is its vertical height. The destination can overhear the signal from the RIS, as well as through the direct path.

The received signal at the destination can be written as

$$y = As + w, \quad (1)$$

for which the channel response of the RIS-aided system is

$$A = \mathbf{g}^T \Theta \mathbf{h} + u, \quad (2)$$

where  $s$  is the transmitted signal with transmitted power  $E_s = E[|s|^2]$ ,  $w \sim \mathcal{N}_{\mathbb{C}}(0, N_0)$  is the additive white Gaussian

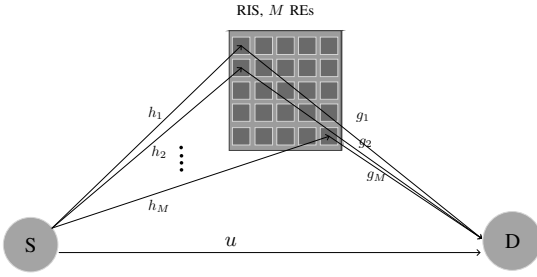


Fig. 1. SISO wireless system with one RIS. The S-RIS and RIS-D paths consist of multiple propagation paths through the  $M$  REs.

noise with zero mean and variance  $N_0 = \mathbb{E}[|u|^2]$  that follows circularly symmetric Gaussian distribution, and  $\Theta$  is the diagonal phase-shift matrix  $\Theta = \text{diag}(e^{j\theta_1}, \dots, e^{j\theta_M})$ . On the other hand, the channel vectors of the links S-RIS and RIS-D, which are assumed to be independent, are denoted respectively as  $\mathbf{h} = [h_1, \dots, h_M]^T \in \mathbb{C}^M$  and  $\mathbf{g} = [g_1, \dots, g_M]^T \in \mathbb{C}^M$  while  $u = |u|e^{j\angle u} \in \mathbb{C}$  is the fading coefficient of the direct S-D link which is independent of the indirect RIS links.

We adopt the spatially correlated Rayleigh fading model proposed in [14]. Accordingly, we characterize the fading distribution of the encountered links as  $u \sim \mathcal{N}_{\mathbb{C}}(0, \Omega_u)$ ,  $\mathbf{h} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \Lambda\mu_h\mathbf{R})$ , and  $\mathbf{g} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \Lambda\mu_g\mathbf{R})$ , where  $\Omega_u$ ,  $\mu_h$  and  $\mu_g$  are the large-scale fading coefficients (path strength) of S-D, S-RIS and RIS-D, respectively, whereas  $\mathbf{R} \in \mathbb{C}^{M \times M}$  is the spatial correlation matrix, which is assumed to be the same for both RIS links, and its elements are calculated according to [14, Eq. 10] as

$$\mathbf{R}_{n,m} = \text{sinc}\left(\frac{2\|\mathbf{a}_n - \mathbf{a}_m\|}{\lambda}\right), \quad n, m = 1, \dots, M, \quad (3)$$

where  $\lambda$  is the wavelength, and  $\mathbf{a}_n$  and  $\mathbf{a}_m$  are the respective locations of the  $n$ th and  $m$ th elements w.r.t. the origin.

Assuming perfect CSI at the RIS, we consider optimal phase configuration by choosing  $\theta_i = \angle u - (\angle h_i + \angle g_i)$ ,  $i = 1, \dots, M$ . Therefore,

$$|A| = \sum_{i=1}^M |h_i g_i| + |u|. \quad (4)$$

Consequently, the instantaneous end-to-end signal-to-noise ratio (SNR) at the receiver is defined as

$$\rho = \rho_0 |A|^2, \quad (5)$$

where  $\rho_0 = E_s/N_0$  denotes the transmit SNR.

### III. ERGODIC CAPACITY ANALYSIS

In this section, we derive analytical expressions for the ergodic capacity of the system under study. The ergodic capacity [bit/s/Hz] is defined as

$$\bar{C} \triangleq \mathbb{E}[\log_2(1 + \rho)] = \int_0^\infty \log_2(1 + x) f_\rho(x) dx, \quad (6)$$

where  $f_\rho(\cdot)$  is the probability density function (PDF) of the end-to-end SNR. Since the exact distribution of  $\rho$  given by (4) and (5) is intractable, it may not be possible to derive the exact ergodic capacity. Therefore, we shall first approximate  $f_\rho(\cdot)$  by either the non-central chi-square or the Gamma distribution.

#### A. Statistical Channel Characterization Based on Non-Central Chi-Square Distribution

According to the central limit theorem (CLT), the sum of weakly correlated random variables converges toward a normal random variable. Therefore, the channel response defined in (4) will be nearly normally distributed and, consequently, the PDF of the end-to-end SNR in (5) can be approximated with a non-central chi-square random variable with one degree-of-freedom as

$$f_\rho(x) \simeq \frac{1}{2\rho_0\kappa^2} \left(\frac{x}{\rho_0\eta}\right)^{-\frac{1}{4}} \exp\left(-\frac{x + \eta\rho_0}{2\rho_0\kappa^2}\right) I_{-\frac{1}{2}}\left(\frac{\sqrt{\eta x}}{\sqrt{\rho_0\kappa^2}}\right), \quad (7)$$

for which  $\eta = (\mathbb{E}[|A|])^2$ ,  $\kappa^2 = \text{Var}[|A|]$  and  $I_\nu(\cdot)$  is the modified Bessel function of the first kind [15, Eq. 8.406]. We calculate  $\mathbb{E}[|A|]$  using the linearity property together with the independency assumption of  $\mathbf{h}$  and  $\mathbf{g}$  as  $\mathbb{E}[|A|] = \mathbb{E}[\sum_{i=1}^M |h_i g_i|] + \mathbb{E}[|u|] = \sum_{n=1}^M \mathbb{E}[|h_{n,i}|] \mathbb{E}[|g_{n,i}|] + \mathbb{E}[|u|]$ , where  $\mathbb{E}[|u|] = \frac{\sqrt{\pi\Omega_u}}{2}$ . Therefore,

$$\eta = \left(\frac{M\pi\Lambda}{4} \sqrt{\mu_h\mu_g} + \frac{\sqrt{\pi\Omega_u}}{2}\right)^2. \quad (8)$$

On the other hand, we calculate  $\kappa^2 = \text{Var}[|A|]$  as

$$\kappa^2 = \sum_{i=1}^M \sum_{j=1}^M \frac{\Psi\Lambda^2\mu_h\mu_g\mathbf{R}_{i,j}}{4} [\Psi\mathbf{R}_{i,j} + \pi] + \frac{\Psi\Omega_u}{2}, \quad (9)$$

where  $\Psi = \left(\frac{4-\pi}{2}\right)$ ; see Appendix A for the derivation of (9).

Let us next approximate the ergodic capacity in (6) by substituting the exact  $f_\rho(\cdot)$  by its non-central chi-square-distributed approximation in (7) as

$$\bar{C}_{\chi^2} = \int_0^\infty \frac{1}{2\kappa^2} \log_2(1 + \rho_0 t) \left(\frac{t}{\eta}\right)^{-\frac{1}{4}} \times \exp\left(-\frac{t + \eta}{2\kappa^2}\right) I_{-\frac{1}{2}}\left(\frac{\sqrt{\eta t}}{\kappa^2}\right) dt. \quad (10)$$

This integral will result in the same analytical expression as [16, Eq. 21] that is rewritten in (11) at the top of the next page, with substituting  $\eta$  and  $\kappa^2$  by the novel expressions (8) and (9), respectively. In (11),  $H[\cdot, \dots, \cdot]$  is the multivariable Fox  $H$ -function [17, Eq. 8.3.1].

#### B. Statistical Channel Characterization Based on Gamma Distribution

Since the PDF of the end-to-end SNR can be approximated with a non-central chi-square random variable with one degree-of-freedom in Section III-A, it will look similar to the Gaussian PDF with a single maximum, and its tails extend to

$$\bar{C}_{\chi^2} = \frac{\pi}{2\sqrt{\kappa}} \ln 2 \left( \frac{\kappa^3}{8\eta\rho_0^3} \right)^{-\frac{1}{4}} \exp\left(-\frac{\eta}{2\kappa^2}\right) H_{1,0:1,1,0:1,2}^{0,1:1,1,0:1,2} \left[ \begin{matrix} (\frac{1}{4}; 1, 1) \\ - \end{matrix} \middle| \begin{matrix} (\frac{1}{4}, 1) \\ (-\frac{1}{4}, 1), (\frac{1}{4}, 1), (\frac{1}{4}, 1) \end{matrix} \middle| \begin{matrix} (1, 1), (1, 1) \\ (1, 1), (0, 1) \end{matrix} \middle| \frac{\eta}{2\kappa^2}, 2\rho_0\kappa^2 \right] \quad (11)$$

$$\begin{aligned} \text{Var}[|A|^2] &= \overbrace{\sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^M \sum_{m=1}^M \frac{\Psi^4 \Lambda^4 \mu_h^2 \mu_g^2}{16} \hat{R}^2 + \frac{\Psi^3 \Lambda^4 \mu_h^2 \mu_g^2 \pi}{16} \hat{R} \tilde{R} + \frac{\Psi^2 \Lambda^4 \mu_h^2 \mu_g^2 \pi^2}{32} \left( \hat{R} + \frac{\tilde{R}^2}{2} \right) + \frac{\Psi \Lambda^4 \mu_h^2 \mu_g^2 \pi^3}{64} \tilde{R} + \frac{\Lambda^4 \mu_h^2 \mu_g^2 \pi^4}{256}}^{P_2} \\ &+ \overbrace{2\sqrt{\pi\Omega_u} \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^M \left( \frac{\pi\Lambda}{4} \right)^3 (\mu_h \mu_g)^{\frac{3}{2}} + \Psi \frac{\pi^2 \Lambda^3 (\mu_h \mu_g)^{\frac{3}{2}}}{16} \tilde{R} + \Psi^2 \frac{\Lambda^3 (\mu_h \mu_g)^{\frac{3}{2}} \pi}{16} \tilde{R}^2}^{P_3} \\ &+ 6\Omega_u \sum_{i=1}^M \sum_{j=1}^M \left[ \frac{\Psi^2 \Lambda^2 \mu_h \mu_g \mathbf{R}_{i,j}^2}{4} + \frac{\Psi \pi \Lambda^2 \mu_h \mu_g \mathbf{R}_{i,j}}{4} + \frac{\pi^2 \Lambda^2 \mu_h \mu_g}{16} \right] + M\pi\Lambda\sqrt{\mu_h \mu_g} \beta^{\frac{3}{2}} \Gamma\left(\frac{5}{2}\right) \\ &- \left( \sum_{i=1}^M \sum_{j=1}^M \frac{\Psi \Lambda^2 \mu_h \mu_g \mathbf{R}_{i,j}}{4} [\Psi \mathbf{R}_{i,j} + \pi] + \frac{\pi^2 \Lambda^2 M^2 \mu_h \mu_g}{16} + \frac{M\pi\Lambda\sqrt{\mu_h \mu_g} \pi \Omega_u}{4} + \Omega_u \right)^2 \quad (15) \end{aligned}$$

\* note:  $\hat{R} = [\mathbf{R}_{i,j} \mathbf{R}_{k,m} + \mathbf{R}_{i,k} \mathbf{R}_{j,m} + \mathbf{R}_{i,m} \mathbf{R}_{j,k}]$ ,  $\tilde{R} = [\mathbf{R}_{i,j} + \mathbf{R}_{k,m} + \mathbf{R}_{i,k} + \mathbf{R}_{j,m} + \mathbf{R}_{i,m} + \mathbf{R}_{j,k}]$ ,  $\tilde{R} = [\mathbf{R}_{j,k} + \mathbf{R}_{i,k} + \mathbf{R}_{i,j}]$

infinity from the right side but is truncated to zero from the left side. Hence, this PDF can be tightly approximated by the first term of a Laguerre series expansion as stated in [18] as

$$f_\rho(x) \simeq \frac{x^{\alpha-1}}{(\rho_0\beta)^\alpha \Gamma(\alpha)} \exp\left(-\frac{x}{\rho_0\beta}\right), \quad (12)$$

where

$$\alpha = \frac{(\mathbb{E}[|A|^2])^2}{\text{Var}[|A|^2]} \quad \text{and} \quad \beta = \frac{\text{Var}[|A|^2]}{\mathbb{E}[|A|^2]}. \quad (13)$$

The mean of  $|A|^2$  is calculated as

$$\begin{aligned} \mathbb{E}[|A|^2] &= \sum_{i=1}^M \sum_{j=1}^M \Psi \frac{\Lambda^2 \mu_h \mu_g \mathbf{R}_{i,j}}{4} [\Psi \mathbf{R}_{i,j} + \pi] \\ &+ \frac{\pi^2 \Lambda^2 M^2 \mu_h \mu_g}{16} + \frac{M\pi\Lambda\sqrt{\mu_h \mu_g} \pi \Omega_u}{4} + \Omega_u, \quad (14) \end{aligned}$$

whereas the variance of  $|A|^2$  is given by (15) at the top of this page. The detailed derivations of (14) and (15) are available in Appendices B and C, respectively.

Let us next approximate the ergodic capacity in (6) by substituting the exact  $f_\rho(\cdot)$  by its Gamma-distributed approximation in (12) as

$$\bar{C}_\Gamma = \int_0^\infty \frac{1}{\beta^\alpha \Gamma(\alpha)} \log_2(1 + \rho_0 t) t^{\alpha-1} \exp\left(-\frac{t}{\beta}\right) dt. \quad (16)$$

The integral can be evaluated using [19, Eq. 78] as

$$\bar{C}_\Gamma = \frac{\exp\left(\frac{1}{\rho_0\beta}\right)}{\log(2)} \sum_{j=1}^{\alpha} \Gamma\left(-\alpha + j, \frac{1}{\rho_0\beta}\right) (\rho_0\beta)^{j-\alpha}. \quad (17)$$

However, (17) is only valid for integer values of  $\alpha$ , i.e., (16) has no closed-form expression for non-integer values of  $\alpha$ .

### C. Unified Logarithmic Approximations for Ergodic Capacity

It can be observed from (11) and (17) that the resulted ergodic capacity is very complex and it is almost impossible to get insightful observations from these expressions. In addition, (16) does not even admit a closed-form expression for the non-integer values of  $\alpha$ . Therefore, we introduce novel and tractable approximations for  $\bar{C}_{\chi^2}$  and  $\bar{C}_\Gamma$  in the form of a weighted sum of logarithmic functions as

$$\tilde{C}(\rho_0) \triangleq \sum_{n=1}^N a_n \log_2(1 + b_n \rho_0), \quad (18)$$

for which the research problem is to choose appropriate values for the coefficients  $\{(a_n, b_n)\}_{n=1}^N$ . The proposed approximation (18) is unified for both of the adopted distributions and it reveals that the ergodic capacity increases with  $\rho_0$ .

We acquire  $\{(a_n, b_n)\}_{n=1}^N$  using the minimax optimization principle according to the Remez exchange algorithm, which we slightly modify to make it comply with the nonlinearity that occurs from the logarithmic approximation. In particular, the Remez algorithm is an iterative method that can be used to derive the minimax approximation, which results in a uniform error function with equalized extrema of alternating signs [20]. In this paper, we use the sum of logarithms in (18) as the approximating function to obtain the best unique approximations for (10) and (16) in terms of the relative error. The resulting relative error function is defined as  $r(\rho_0) \triangleq \frac{\tilde{C}(\rho_0)}{C^*(\rho_0)} - 1$  with  $\star \in \{\chi^2, \Gamma\}$ , and it should have  $2N + 1$  extrema on  $[0, \infty)$ .

We start the Remez algorithm by constructing a system of  $2N + 1$  nonlinear equations  $f_k(\mathbf{v}) \triangleq r(x_k) + (-1)^k r_{\max} = 0$ ,  $k = 0, 1, \dots, 2N$ , describing the values of extrema points as

$$\mathbf{f}(\mathbf{v}) \triangleq [f_0(\mathbf{v}), f_1(\mathbf{v}), \dots, f_{2N}(\mathbf{v})]^T = \mathbf{0}, \quad (19)$$

where  $x_k$  is the location of the  $k$ th extremum of the error function,  $r_{\max}$  is the value of the maximum error at the uniform extrema points, and  $\mathbf{v} = [a_1, a_2, \dots, a_N, b_1, b_2, \dots, b_N, r_{\max}]^T$  is a vector of the unknowns. The first extrema point always occurs asymptotically at zero, i.e., we choose  $x_0$  to be a very small fixed value near zero, whereas the last extrema point always occurs asymptotically at infinity, i.e., we choose  $x_{2N}$  to be a very large fixed value; the other  $x_k$  are variables that are found through the outer Remez iterations. Next, we initialize these remaining locations of the extrema points and start the first Remez iteration by solving  $\mathbf{f}$  for  $\mathbf{v}$ . We then find the locations of the new extrema points of the resulting error function and use them for the following Remez iteration which we repeat until the difference between the locations of the old and new extrema is smaller than a predefined threshold value.

In each outer iteration of the Remez algorithm, we solve  $\mathbf{f}$  for  $\mathbf{v}$  using the Newton–Raphson method since the logarithmic approximation results in a nonlinear type of equations. The Newton–Raphson method is also iterative and requires initial guesses for the vector of unknowns  $\mathbf{v}$ . Its iterations are referred to as the inner iterations to differentiate them from the outer ones of the Remez algorithm, and are performed as

$$\mathbf{v}^{(\tau+1)} = \mathbf{v}^{(\tau)} - \left[ \mathbf{J}^{(\tau)}(\mathbf{v}^{(\tau)}) \right]^{-1} \mathbf{f}(\mathbf{v}^{(\tau)}), \quad (20)$$

where  $\tau$  is its inner-iteration counter and  $\mathbf{J}(\cdot)$  is the Jacobian matrix defined as  $\mathbf{J}(\mathbf{v}) = \left[ \frac{\partial \mathbf{f}}{\partial v_0}, \frac{\partial \mathbf{f}}{\partial v_1}, \dots, \frac{\partial \mathbf{f}}{\partial v_{2N}} \right]$ , for which  $\frac{\partial f_k}{\partial a_n} = \frac{\log_2(1+b_n x_k)}{C_*(x_k)}$ ,  $\frac{\partial f_k}{\partial b_n} = \frac{a_n x_k}{\log(2)(1+b_n x_k)C_*(x_k)}$ ,  $\frac{\partial f_k}{\partial r_{\max}} = (-1)^k$ . The Newton–Raphson iterations are repeated until  $\Delta \mathbf{v} = \mathbf{v}^{(\tau+1)} - \mathbf{v}^{(\tau)}$  is less than a threshold value.

#### IV. NUMERICAL RESULTS AND DISCUSSIONS

This section gives insight into the performance of the considered system in terms of the ergodic capacity, where it studies the effect of several parameters on the ergodic capacity. In addition, it verifies the accuracy of the adopted non-central chi-square and Gamma approximations by means of Monte Carlo simulations. In particular, we consider herein a network setup with a carrier frequency of 3 GHz,  $d_H = d_V = \frac{\lambda}{4}$  and  $\mu_h = \mu_g = -45$  dB since the RIS is usually placed in an elevated place from the ground and thus has less path losses.

In Fig. 2, we examine the accuracy of the non-central chi-square and Gamma statistical models in (7) and (12), respectively. As can be seen, both approximations are well corroborated with the true PDF with or without direct path and for different values of  $M$ . The right shifting of the PDF which occurs upon increasing the number of REs equipped on the RIS, indicates an increase in the system power gain.

The ergodic capacity of the considered RIS-aided system is evaluated using (11) for the non-central chi-square distribution and using (16) for the Gamma distribution and the corresponding results are plotted in Fig. 3, where they coincide well with the simulated ergodic capacity for different  $M$  values. Moreover, the approximation proposed in (18) is plotted for both distributions after finding the optimized coefficients<sup>1</sup>

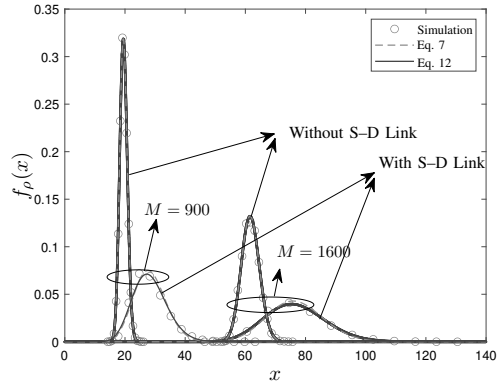


Fig. 2. The true PDF of the end-to-end SNR compared to approximated non-central chi-square and the Gamma PDFs with  $\rho_0 = 110$  dB and  $\Omega_u = -120$  dB.

using the Remez algorithm. Both approximations match well with the exact expressions with only four logarithmic terms ( $N = 4$ ). The figure also depicts the impact of imposing an RIS to the communication system, where the ergodic capacity shows much better performance when compared to the scenario where communication is achieved only through the direct path. Moreover, we can see that the ergodic capacity increases with increasing the transmitted power and better performance is achieved with increasing the number of REs i.e., less transmitted power is required to achieve a certain level of the ergodic capacity. For example, for ergodic capacity of 1.5 bit/s/Hz, an increment by 300 REs will decrease the required transmitted power by approximately 6.4 dB.

The impact of the number of REs equipped on the RIS to the ergodic capacity for different  $\Omega_u$  and  $\rho_0$  values is investigated further in Fig. 4, which illustrates the significant increase in the ergodic capacity when increasing  $M_H = M_V$  and  $M$  thereof, e.g., for  $\Omega_u = -110$  dB and  $\rho_0 = 115$  dB, as  $M_H = M_V$  changes from 15 to 25, the ergodic capacity improves by approximately 75%. The figure also confirms the significant improvement of the ergodic capacity with increasing the transmitted power, e.g., for  $M_H = M_V = 30$ , as  $\rho_0$  shifts from 105 to 110 dB, the ergodic capacity improves by approximately 71%. Moreover, as the strength of the direct path increases, the ergodic capacity increases considerably, e.g., for  $M_H = M_V = 15$  and  $\rho_0 = 110$  dB, as  $\Omega_u$  increases from  $-110$  to  $-90$  dB, the ergodic capacity improves by approximately 214%.

We further study the effect of utilizing an RIS to assist the communication between S and D to the ergodic capacity and how it behaves with changing the strength of the direct path in Fig. 5. In particular, the RIS increases the ergodic capacity considerably, especially for lower values of  $\Omega_u$  where the RIS contributes more to the communication process, e.g., for  $\Omega_u = -130$  dB, the ergodic capacity increases by 2.9, 4.4 and 6.0 bit/s/Hz for  $\rho_0 = 105, 110$  and 115 dB, respectively,

<sup>1</sup> Available at <https://doi.org/10.5281/zenodo.6087447> for download.

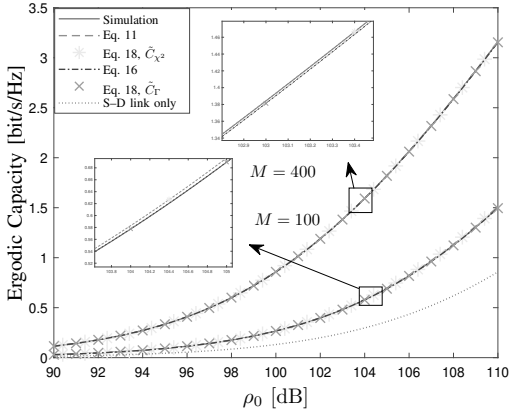


Fig. 3. The ergodic capacity in terms of the transmitted power with  $\Omega_u = -110$  dB and for different values of  $M$ .

when adding an RIS with 900 REs. On the other hand, the RIS contributes less to the communication process for the higher values of  $\Omega_u$ , where the direct path has a strong effect. In general, the ergodic capacity increases with increasing  $\Omega_u$ .

## V. CONCLUSION

This paper studied the ergodic capacity of a SISO communication system with correlated Rayleigh-fading channels and direct path between the source and destination. Specifically, it presented two approximating schemes for the PDF of the end-to-end SNR and thus derived closed-form expressions for the corresponding ergodic capacity. Furthermore, it presented more simplified, but very accurate, logarithmic approximations to the ergodic capacity for both schemes. Numerical simulations verified the performed statistical analysis and confirmed the high accuracy of the proposed approximations. The conducted analysis revealed that ergodic capacity improves with the transmitted SNR, the number of REs, and the strength of the direct path. The effect of adding an RIS to the communication system or increasing the number of its REs, is best seen for the lower values of the direct path strength.

## APPENDIX

### A. Derivation of (9)

We calculate the variance of  $|A|$  which is defined in (4) as

$$\text{Var} [|A|] = \text{Var} \left[ \sum_{i=1}^M |h_i g_i| \right] + \text{Var} [|u|]. \quad (21)$$

By utilizing the definition of the variance, we get

$$\text{Var} \left[ \sum_{i=1}^M |h_i g_i| \right] = \text{E} \left[ \left( \sum_{i=1}^M |h_i g_i| \right)^2 \right] - \left( \text{E} \left[ \sum_{i=1}^M |h_i g_i| \right] \right)^2, \quad (22)$$

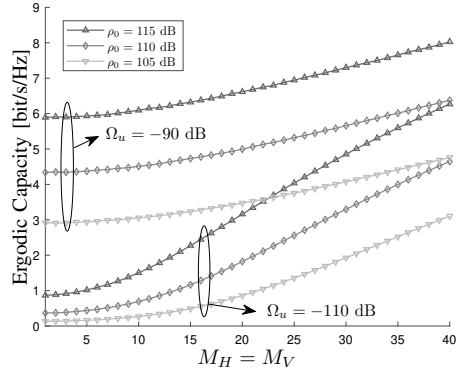


Fig. 4. Impact of the number of REs per dimension on the ergodic capacity with different  $\rho_0$  and  $\Omega_u$  values.

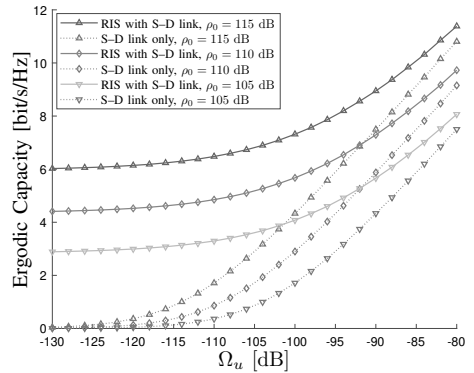


Fig. 5. Impact of the strength of the direct path on the ergodic capacity with  $M = 900$ .

where

$$\text{E} \left[ \left( \sum_{i=1}^M |h_i g_i| \right)^2 \right] = \sum_{i=1}^M \sum_{j=1}^M \text{E} [|h_i h_j|] \text{E} [|g_i g_j|]. \quad (23)$$

The channel coefficient  $|h_i|$  is Rayleigh-distributed by which its mean and variance are given as

$$\text{E} [|h_i|] = \sqrt{\frac{\pi \Lambda \mu_h}{4}} \quad \text{and} \quad \text{Var} [|h_i|] = \left( \frac{4 - \pi}{2} \right) \frac{\Lambda \mu_h}{2}, \quad (24)$$

respectively; the corresponding fact applies also for the channel coefficient  $|g_i|$  in terms of  $\mu_g$ .

Now using [21, Eq. 10], we get

$$\begin{aligned} \mathbb{E} \left[ \left( \sum_{i=1}^M |h_i g_i| \right)^2 \right] &= \sum_{i=1}^M \sum_{j=1}^M \left[ \left( \frac{4-\pi}{2} \right) \frac{\Lambda \mu_h \mathbf{R}_{i,j}}{2} + \frac{\pi \Lambda \mu_h}{4} \right] \\ &\quad \times \left[ \left( \frac{4-\pi}{2} \right) \frac{\Lambda \mu_g \mathbf{R}_{i,j}}{2} + \frac{\pi \Lambda \mu_g}{4} \right] \quad (25) \\ &= \sum_{i=1}^M \sum_{j=1}^M \left[ \left( \frac{4-\pi}{2} \right)^2 \frac{\Lambda^2 \mu_h \mu_g \mathbf{R}_{i,j}}{4} \right. \\ &\quad \left. + \left( \frac{4-\pi}{2} \right) \frac{\pi \Lambda^2 \mu_h \mu_g \mathbf{R}_{i,j}}{4} + \frac{\pi^2 \Lambda^2 \mu_h \mu_g}{16} \right]. \end{aligned}$$

On the other hand, from (8) we find

$$\mathbb{E} \left[ \sum_{i=1}^M |h_i g_i| \right] = \frac{M \pi \Lambda \sqrt{\mu_h \mu_g}}{4}. \quad (26)$$

By substituting (25) and (26) back into (22), which is then substituted together with  $\text{Var}[|u|] = \left(\frac{4-\pi}{4}\right)\Omega_u$  into (21), we then complete the derivation of (9).

### B. Derivation of (14)

We calculate the mean of  $|A|^2$ , where  $|A|$  is defined in (4), as

$$\begin{aligned} \mathbb{E}[|A|^2] &= \mathbb{E} \left[ \left( \sum_{i=1}^M |h_i g_i| \right)^2 \right] + 2 \mathbb{E}[|u|] \mathbb{E} \left[ \sum_{i=1}^M |h_i g_i| \right] \\ &\quad + \mathbb{E}[|u|^2]. \quad (27) \end{aligned}$$

By substituting (25) and (26) in (27) together with using  $\mathbb{E}[|u|^\zeta] = \Omega_u^{\frac{\zeta}{2}} \Gamma\left(1 + \frac{\zeta}{2}\right)$ , we complete the derivation of (14).

### C. Derivation of (15)

We calculate the variance of  $|A|^2$ , where  $|A|$  is defined in (4), as

$$\begin{aligned} \text{Var}[|A|^2] &= \mathbb{E} \left[ \underbrace{\left( \sum_{i=1}^M |h_i g_i| + |u| \right)^4}_{P_1} \right] \\ &\quad - \mathbb{E} \left[ \left( \sum_{i=1}^M |h_i g_i| + |u| \right)^2 \right]^2, \quad (28) \end{aligned}$$

where the first term is given by

$$\begin{aligned} P_1 &= \mathbb{E} \left[ \left( \sum_{i=1}^M |h_i g_i| \right)^4 \right] + 4 \mathbb{E}[|u|] \mathbb{E} \left[ \left( \sum_{i=1}^M |h_i g_i| \right)^3 \right] \\ &\quad + 6 \mathbb{E}[|u|^2] \mathbb{E} \left[ \left( \sum_{i=1}^M |h_i g_i| \right)^2 \right] + 4 \mathbb{E}[|u|^3] \mathbb{E} \left[ \sum_{i=1}^M |h_i g_i| \right] \\ &\quad + \mathbb{E}[|u|^4], \quad (29) \end{aligned}$$

for which  $\mathbb{E} \left[ \left( \sum_{i=1}^M |h_i g_i| \right)^4 \right]$  and  $\mathbb{E} \left[ \left( \sum_{i=1}^M |h_i g_i| \right)^3 \right]$  are calculated in a similar way to (25) with using [22, Eq. 1.1] for the former and [21, Eq. 14] for the latter. This result in evaluating both terms respectively as  $P_2$  and  $P_3$  in (15). Moreover, by trivially solving the remaining terms and then substituting the resulting  $P_1$  term with (27) into (28), we then complete the derivation of (15).

## REFERENCES

- [1] H. Yang, X. Cao, F. Yang, J. Gao, S. Xu, M. Li, X. Chen, Y. Zhao, Y. Zheng, and L. Sijia, "A programmable metasurface with dynamic polarization, scattering and focusing control," *Sci. Rep.*, vol. 6, Oct. 2016.
- [2] N. Kaina, M. Dupre, G. Lerosey, and M. Fink, "Shaping complex microwave fields in reverberating media with binary tunable metasurfaces," *Sci. Rep.*, vol. 4, p. 6693, Oct. 2014.
- [3] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wirel. Commun.*, vol. 18, no. 8, pp. 4157–4170, Jun. 2019.
- [4] Q. Wu and R. Zhang, "Beamforming optimization for intelligent reflecting surface with discrete phase shifts," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, May 2019, pp. 7830–7833.
- [5] C. Pan, H. Ren, K. Wang, W. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Trans. Wirel. Commun.*, vol. 19, no. 8, pp. 5218–5233, Aug. 2020.
- [6] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, "Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 14960–14973, Oct. 2020.
- [7] A. Almohamad, A. M. Tahir, A. Al-Kababji, H. M. Furqan, T. Khattab, M. O. Hasna, and H. Arslan, "Smart and secure wireless communications via reflecting intelligent surfaces: A short survey," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1442–1456, Sep. 2020.
- [8] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, Aug. 2019.
- [9] D. Kudathanthirige, D. Gunasinghe, and G. Amarasuriya, "Performance analysis of intelligent reflective surfaces for wireless communication," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2020.
- [10] A. A. Boulogeorgos and A. Alexiou, "Performance analysis of reconfigurable intelligent surface-assisted wireless systems and comparison with relaying," *IEEE Access*, vol. 8, pp. 94463–94483, May 2020.
- [11] A. M. Salhab and M. H. Samuh, "Accurate performance analysis of reconfigurable intelligent surfaces over Rician fading channels," *IEEE Wirel. Commun. Lett.*, vol. 10, no. 5, pp. 1051–1055, 2021.
- [12] I. M. Tanash and T. Riihonen, "Link performance of multiple reconfigurable intelligent surfaces and direct path in general fading," in *Proc. Int. Conf. Signal Process. Commun. Syst.*, Dec. 2021, pp. 1–6.
- [13] Q.-U.-A. Nadeem, H. Alwazani, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Intelligent reflecting surface-assisted multi-user MISO communication: Channel estimation and beamforming design," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 661–680, May 2020.
- [14] E. Björnson and L. Sanguinetti, "Rayleigh fading modeling and channel hardening for reconfigurable intelligent surfaces," *IEEE Wirel. Commun. Letters*, vol. 10, no. 4, pp. 830–834, Apr. 2021.
- [15] I. Gradshteyn and I. Ryzhik, *Table of integrals, series, and products*, 7th ed. Elsevier/Academic Press, 2007.
- [16] Y. Zhang, J. Zhang, M. D. Renzo, H. Xiao, and B. Ai, "Performance analysis of RIS-aided systems with practical phase shift and amplitude response," *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4501–4511, May 2021.
- [17] A. Prudnikov, Y. Brychkov, and O. Marichev, *Integrals and Series. Volume 3: More Special Functions*. CRC Press, Oct. 1990, vol. 3.
- [18] S. Primak, *Stochastic Methods and Their Applications to Communications: Stochastic Differential Equations Approach*. Wiley, 2004.
- [19] M.-S. Alouini and A. J. Goldsmith, "Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques," *IEEE Trans. Veh. Technol.*, vol. 48, no. 4, pp. 1165–1181, Jul. 1999.
- [20] I. M. Tanash and T. Riihonen, "Remez exchange algorithm for approximating powers of the Q-function by exponential sums," in *Proc. IEEE Veh. Technol. Conf.*, Apr. 2021.
- [21] D. Tavela, *Mean and Variance of the Product of Random Variables*, Apr. 2019.
- [22] P. Janssen and P. Stoica, "On the expectation of the product of four matrix-valued Gaussian random variables," *IEEE Trans. Automat. Contr.*, vol. 33, pp. 867–870, Oct. 1988.



# PUBLICATION

8

**Link performance of multiple reconfigurable intelligent surfaces and  
direct path in general fading**

I. M. Tanash and T. Riihonen

In *Proc. International Conference on Signal Processing and Communication Systems  
(ICSPCS)*, Dec. 2021, pp. 1–6

DOI: 10.1109/ICSPCS53099.2021.9660225

**Publication reprinted with the permission of the copyright holders.**



# Link Performance of Multiple Reconfigurable Intelligent Surfaces and Direct Path in General Fading

Islam M. Tanash and Taneli Riihonen

Faculty of Information Technology and Communication Sciences, Tampere University, Finland

e-mail: {islam.tanash, taneli.riihonen}@tuni.fi

**Abstract**—We analyze the performance of a single-input single-output wireless link that is aided by multiple reconfigurable intelligent surfaces (RISs) — in terms of outage probability, average symbol error probability and ergodic capacity, for which we derive analytical expressions in closed form. In particular, we consider a realistic system model, where the direct path may not be blocked and for which channels corresponding to different RISs are assumed to be independent but not identical and follow the generic  $\kappa$ - $\mu$  fading distribution, which can be reduced to a number of fading scenarios (namely Rayleigh, Rice, Nakagami- $m$ , and one-sided Gaussian). This enables the evaluation of the system performance when adopting any combination of these special cases or the generic  $\kappa$ - $\mu$  distribution for both hops of the multiple distributed RISs. The direct path is modeled by Rayleigh fading assuming no line-of-sight between the source and the destination. We verify the accuracy of the adopted approach by means of Monte Carlo simulations and conduct a performance analysis that demonstrates the significant improvement in the system performance due to the usage of the RISs. Especially, we show that increasing the number of reflecting elements equipped on the RISs and placing the RISs closer to either communication endpoints improve the performance considerably.

## I. INTRODUCTION

The reconfigurable intelligent surface (RIS) is a promising emerging technology for future wireless communication networks since it gives more control over the wireless environment for the aim of improving the quality-of-service and spectrum efficiency. It consists of a large surface that has low-cost passive reflecting elements (REs) that can be adapted by a microcontroller to collaboratively reflect the incident electromagnetic signals into the desired direction.

Most of the research work conducted on this topic focuses on the design [1], [2], optimization [3]–[5] and potential applications [6]–[8] of RIS-aided systems. Specifically, in [1], a digitally controlled metasurface, whose units can be adapted independently, is designed to dynamically manipulate the electromagnetic waves and, thus, achieve more versatility; whereas in [2], a tunable metasurface is designed to work as a spatial microwave modulator with energy feedback.

Prior works have also investigated optimizing the performance of RIS-aided wireless systems: In [3], the authors solve a non-convex optimization problem to maximize their system's energy efficiency; and in [4], the discrete phase shifts together with the transmit beamforming of a multi-antenna base station are optimized to minimize transmission power. In addition,

the authors in [5], who adopt RISs at the edge of cells to enhance the downlink transmission for cell-edge users, aim toward maximizing the weighted sum rate of all users by optimizing the transmitter's active precoding matrices together with the REs' phase shifts.

The applications of RISs span the different areas of wireless communications, where it is adopted in [6] to support the communication in unmanned aerial vehicle-assisted wireless systems and in [7] to assist the data transmission from a base station to a single-antenna receiver in an RIS-assisted millimeter wave system. The RIS technology can also be adopted in wireless networks to enhance the physical layer security as explained in [8]. On the other hand, the theoretical study of RIS-aided wireless networks still in its early stage, where limited number of research works have been established to analyze the performance of these systems due to the difficulty in evaluating the statistical characterization of the end-to-end signal-to-noise ratio (SNR). Therefore, several approximations, bounds or asymptotic analysis have been developed to analyze the RIS-aided systems [9], [10].

Noticeable efforts have been made on studying the generic single-input single-output (SISO) system model without direct path, where the central limit theorem (CLT) is used to derive bounds or approximations for the different performance measures for Rayleigh distribution in [11], [12]. A different approximating approach is used in [13], [14] to achieve high accuracy regardless of the number of REs at the RIS. The SISO system with Rician fading and direct path between the source (S) and destination (D) is studied in [15], for which the statistical characterization of the end-to-end SNR is not evaluated and thus the symbol error rate is not derived either.

A more generic SISO system with multiple RISs is investigated in [16], [17] and different approaches are used to approximate the channel statistics. All the fading channels associated with different RISs are assumed to be independent and identically distributed (i.i.d.). However, this does not represent a realistic assumption since the RISs may be distributed over a wide geographical area. Therefore, different RISs are expected to experience non-identical channels of the same or different fading distribution. On the other hand, for each RIS, the channels encountered by REs can be assumed to be i.i.d. since they are placed on the same surface, i.e., the REs of a single RIS are located very close to each other.

Motivated by the fact that the literature only considers the case where the same fading model is assumed for both hops (S-RIS and RIS-D) among all the distributed RISs and with i.i.d. channels, we present herein a more realistic performance study of a generic SISO system model with multiple RISs and direct path with independently but non-identically distributed (i.n.i.d.) fading channels across the distributed RISs which are geographically far apart from each other, and thus each RIS may also experience different fading distribution. Therefore, we choose to evaluate the system's performance over the generic  $\kappa$ - $\mu$  distribution which can be reduced to a number of the most used fading scenarios, namely, Rayleigh, Rice, Nakagami- $m$  and one-sided Gaussian distribution. This allows us not only to consider the same double fading channels for all the distributed RISs, but also to consider different combinations of the special cases or the generic distribution for the S-RIS and RIS-D links of the different RISs.

In particular, we implement the Laguerre series method [18] to approximate the statistical characterization of the end-to-end equivalent channel of the SISO system with multiple RISs and direct path. Closed-form expressions for the outage probability and ergodic capacity are presented as well as a novel expression for the average symbol error probability (ASEP) is derived. Our work presents generalized results that are valid for any number of RISs equipped with arbitrary numbers of REs. It is also valid for any combination of the fading distributions covered by the  $\kappa$ - $\mu$  distribution and with or without direct path, where the latter represents a special case of the former when the direct channel gain is set to zero.

## II. SYSTEM AND CHANNEL MODELS

The system under study is illustrated in Fig. 1 and it consists of a single-antenna source (S),  $N$  RISs, where the  $n$ th one (RIS $_n$ ) is equipped with  $M_n$  REs, and a single-antenna destination (D). The destination can overhear the signal from all the distributed RISs as well as through the direct path. It is worth mentioning that the considered system model includes the special case of an obstructed direct path between S and D, for which the channel coefficient  $u$  below in (2) equals zero.

### A. Signal Models

The received signal at the destination can be written as

$$y = A s + w, \quad (1)$$

for which the combined channel response is

$$A = \sum_{n=1}^N A_n + u, \quad (2)$$

where the channel response of the  $n$ th RIS is

$$A_n = \sum_{i=1}^{M_n} h_{n,i} g_{n,i} r_{n,i}, \quad (3)$$

and  $s$  is the transmitted signal,  $h_{n,i}$ ,  $g_{n,i}$  and  $u$  are the fading coefficients of S-RIS $_n$ , RIS $_n$ -D and S-D links, respectively, while the additive white Gaussian noise is denoted by  $w$  in (1)

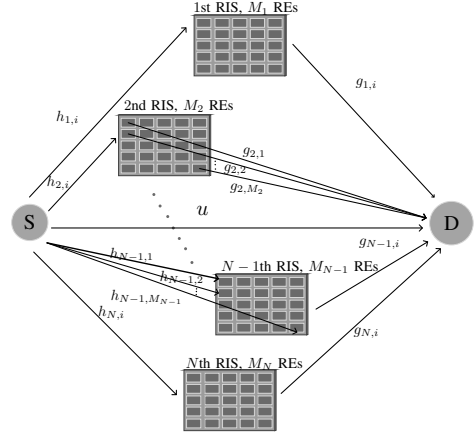


Fig. 1. A SISO wireless system with  $N$  RISs. Each S-RIS $_n$  and RIS $_n$ -D path consists of multiple propagation paths through the  $M_n$  REs. For simplicity, we illustrate the multipath components via two RISs only.

with zero mean and variance  $N_0 = E[|w|^2]$ . The instantaneous end-to-end SNR is defined as  $\rho = E_s |A|^2 / N_0 = \rho_0 |A|^2$  with  $E_s = E[|s|^2]$  being the transmitted power and  $\rho_0 = E_s / N_0$  denoting the transmit SNR. In addition,  $r_{n,i} = \exp(j\theta_{n,i})$  is the response of the  $i$ th RE in the  $n$ th RIS for which its magnitude is assumed to be equal to one and its phase shift is optimized to maximize the SNR at the receiver by choosing  $\theta_{n,i} = \angle u - (\angle h_{n,i} + \angle g_{n,i})$ , assuming ideal global channel state information and centralized coordination.

### B. Fading Models

The flat fading coefficients  $h_{n,i}$ ,  $g_{n,i}$  and  $u$  are assumed to be statistically independent, identical per RIS, and slowly varying. On the other hand, they are not identical for the different RISs, which are geographically separated far apart from each other. The average gains of their envelopes are defined respectively as  $\sigma_{h_n}^2 = E[|h_{n,i}|^2] = \nu_0 (\frac{d_0}{d_{h_n}})^{\eta_{h_n}}$ ,  $\sigma_{g_n}^2 = E[|g_{n,i}|^2] = \nu_0 (\frac{d_0}{d_{g_n}})^{\eta_{g_n}}$  and  $\sigma_u^2 = E[|u|^2] = \nu_0 (\frac{d_0}{d_u})^{\eta_u}$ , where  $\nu_0$  is the reference path loss at the reference distance  $d_0$ , and  $d_j$  and  $\eta_j$ ,  $j \in \{h_n, g_n, u\}$  denote respectively the distance and path loss exponent of the corresponding link. We let  $|h_{n,i}|$  and  $|g_{n,i}|$  follow generalized  $\kappa$ - $\mu$  fading distribution, for which  $\kappa$  is the ratio between the total power of the dominant components and the total power of the scattered waves, and  $\mu$  is the number of multipath clusters [19]. Assuming there is no line-of-sight (LoS) in the direct path, the S-D link can be modeled by Rayleigh fading.

The  $\kappa$ - $\mu$  distribution encloses most of common small-scale fading models as special cases that are obtained by controlling the values of its fading parameters. In particular, for Rayleigh ( $\kappa = 0, \mu = 1$ ), Nakagami- $m$  ( $\kappa = 0, \mu = m$ ), Rice ( $\kappa = K, \mu = 1$ ) and one-sided Gaussian ( $\kappa = 0, \mu = 0.5$ ), where  $m$  and  $K$  refer respectively to the shape parameter of the

Nakagami- $m$  distribution and to the Rician factor. Therefore, in addition to the generic  $\kappa$ - $\mu$  distribution, we can consider the same or combination of the special-case distributions for both links of the  $N$  distributed RISs by assigning the corresponding values to  $\kappa_{h_n}$  and  $\mu_{h_n}$  of the S-RIS $_n$  hop and to  $\kappa_{g_n}$  and  $\mu_{g_n}$  of the RIS $_n$ -D hop.

Toward evaluating the performance measures of the considered system, we need to derive the probability density function (PDF) of the end-to-end SNR for the system under study. We achieve that by first deriving the PDF and the cumulative distribution function (CDF) of the combined channel response defined in (2). It is obvious that the channel response of the  $n$ th RIS defined in (3) is a sum of  $M_n$  identical double  $\kappa$ - $\mu$  random variables, which all are continuous, independent and defined over the positive real axis. Therefore, their sum converges toward a normal random variable according to the central limit theorem. As a result, the combined channel response, which is a sum of the  $N$  resulted normal variables plus a single Rayleigh random variable will also be nearly normally distributed and its PDF will look similar to the Gaussian PDF with a single maximum, and its tails extend to infinity from the right side but is truncated to zero from the left side.

The PDF of the combined nearly-Gaussian channel response can be further tightly approximated by the first term of a Laguerre series expansion as stated in [18] as

$$f_{|A|}(x) \simeq \frac{x^\alpha}{\beta^{\alpha+1} \Gamma(\alpha+1)} \exp\left(-\frac{x}{\beta}\right), \quad (4)$$

where

$$\alpha = \frac{(\mathbb{E}[|A|])^2}{\text{Var}[|A|]} - 1, \quad (5)$$

$$\beta = \frac{\text{Var}[|A|]}{\mathbb{E}[|A|]}. \quad (6)$$

The corresponding CDF can be derived [13, Appendix A] as

$$F_{|A|}(x) \simeq \frac{\gamma(\alpha+1, x/\beta)}{\Gamma(\alpha+1)}, \quad (7)$$

where  $\gamma(\cdot, \cdot)$  denotes the lower incomplete Gamma function.

The mean of  $|A|$  is calculated using its linearity property together with the independency assumption as  $\mathbb{E}[|A|] = \sum_{n=1}^N \mathbb{E}[|A_n|] + \mathbb{E}[|u|] = \sum_{n=1}^N M_n \mathbb{E}[|h_{n,i}|] \mathbb{E}[|g_{n,i}|] + \mathbb{E}[|u|]$  for which the expectation of a  $\kappa$ - $\mu$  distributed fading coefficient is given in [20, Eq. 3] and the  $c$ th moment of a Rayleigh-distributed fading coefficient is  $\mathbb{E}[|u|^c] = \sigma_u^c \Gamma(1 + \frac{c}{2})$ . Therefore,

$$\begin{aligned} \mathbb{E}[|A|] &= \sum_{n=1}^N M_n \frac{\sigma_{h_n} \Gamma(\mu_{h_n} + \frac{1}{2}) \exp(-\kappa_{h_n} \mu_{h_n})}{\Gamma(\mu_{h_n}) ((1 + \kappa_{h_n}) \mu_{h_n})^{\frac{1}{2}}} \\ &\times \frac{\sigma_{g_n} \Gamma(\mu_{g_n} + \frac{1}{2}) \exp(-\kappa_{g_n} \mu_{g_n})}{\Gamma(\mu_{g_n}) ((1 + \kappa_{g_n}) \mu_{g_n})^{\frac{1}{2}}} \\ &\times {}_1F_1\left(\mu_{h_n} + \frac{1}{2}; \mu_{h_n}; \kappa_{h_n} \mu_{h_n}\right) \\ &\times {}_1F_1\left(\mu_{g_n} + \frac{1}{2}; \mu_{g_n}; \kappa_{g_n} \mu_{g_n}\right) + \sqrt{\frac{\pi \sigma_u^2}{4}}, \quad (8) \end{aligned}$$

where  ${}_1F_1(\cdot; \cdot; \cdot)$  is the confluent hypergeometric function of the first kind [21, Eq. 9.210.1].

Likewise, the variance of  $|A|$  is calculated as  $\text{Var}[|A|] = \sum_{n=1}^N \text{Var}[|A_n|] + \text{Var}[|u|]$ , where

$$\begin{aligned} \text{Var}[|A_n|] &= M_n \text{Var}[|h_{n,i} g_{n,i}|] \\ &= M_n (\mathbb{E}[|h_{n,i}|^2] \mathbb{E}[|g_{n,i}|^2] - \mathbb{E}[|h_{n,i}|]^2 \mathbb{E}[|g_{n,i}|]^2) \end{aligned} \quad (9)$$

and  $\text{Var}[|u|] = \mathbb{E}[|u|^2] - (\mathbb{E}[|u|])^2$ , which leads us to evaluating it as shown in (10) at the top of the next page.

Finally, we can derive the PDF of the end-to-end SNR by taking the derivative of the CDF of  $\rho$  that is defined as

$$F_\rho(x) = \Pr(\rho \leq x) = F_{|A|}\left(\sqrt{\frac{x}{\rho_0}}\right). \quad (11)$$

Therefore,

$$f_\rho(x) \simeq \frac{1}{2\beta^{\alpha+1} \Gamma(\alpha+1)} \rho_0^{-\frac{\alpha+1}{2}} x^{\frac{\alpha-1}{2}} \exp\left(-\sqrt{\frac{x}{\beta^2 \rho_0}}\right). \quad (12)$$

### III. PERFORMANCE ANALYSIS

The performance of the considered system is studied in this section in terms of three central performance metrics, namely outage probability, ASEP and ergodic capacity.

The outage probability that is defined as the probability that the end-to-end instantaneous SNR falls below a predefined threshold value,  $\rho_{th}$ , is given directly [13, Eq. 31] by

$$P_O = F_\rho(\rho_{th}) \simeq \frac{\gamma\left(\alpha+1, \frac{1}{\beta} \sqrt{\frac{\rho_{th}}{\rho_0}}\right)}{\Gamma(\alpha+1)}. \quad (13)$$

The average symbol error probability (ASEP) under fading for coherent detection is obtained in most cases by evaluating

$$\bar{P}_E = \int_0^\infty \Omega\left(Q\left(\sqrt{\zeta x}\right)\right) f_\rho(x) dx, \quad (14)$$

where  $\Omega(\cdot)$  is some polynomial of the  $Q$ -function that corresponds to the conditional error probability, e.g.,

$$\begin{aligned} \Omega\left(Q\left(\sqrt{\zeta x}\right)\right) &= 4 \left(\frac{\sqrt{\mathcal{M}}-1}{\sqrt{\mathcal{M}}}\right) Q\left(\sqrt{\zeta x}\right) \\ &- 4 \left(\frac{\sqrt{\mathcal{M}}-1}{\sqrt{\mathcal{M}}}\right)^2 Q^2\left(\sqrt{\zeta x}\right) \quad (15) \end{aligned}$$

for square  $\mathcal{M}$ -quadrature amplitude modulation ( $\mathcal{M}$ -QAM) [22], whereas the constant  $\zeta = \frac{3}{\mathcal{M}-1}$ . We can derive a closed-form expression for (14) by substituting the exponential approximation proposed in [23] into the above integral as

$$\bar{P}_E = \sum_{r=1}^R a_r \int_0^\infty \exp(-b_r \zeta x) f_\rho(x) dx, \quad (16)$$

where  $\{(a_r, b_r)\}_{r=1}^R$  is some set of coefficients from [24]. The above expression is presented with an equality because there is practically no approximation error in the present application despite its being an approximation in the strict sense.

$$\begin{aligned} \text{Var}[|A|] = & \sum_{n=1}^N M_n \left( \sigma_{h_n}^2 \sigma_{g_n}^2 - \frac{\sigma_{h_n}^2 \Gamma^2(\mu_{h_n} + \frac{1}{2}) \exp(-2\kappa_{h_n} \mu_{h_n})}{\Gamma^2(\mu_{h_n})(1 + \kappa_{h_n}) \mu_{h_n}} \frac{\sigma_{g_n}^2 \Gamma^2(\mu_{g_n} + \frac{1}{2}) \exp(-2\kappa_{g_n} \mu_{g_n})}{\Gamma^2(\mu_{g_n})(1 + \kappa_{g_n}) \mu_{g_n}} \right. \\ & \left. \times {}_1F_1^2\left(\mu_{h_n} + \frac{1}{2}; \mu_{h_n}; \kappa_{h_n} \mu_{h_n}\right) {}_1F_1^2\left(\mu_{g_n} + \frac{1}{2}; \mu_{g_n}; \kappa_{g_n} \mu_{g_n}\right) \right) + \frac{4 - \pi}{4} \sigma_u^2 \end{aligned} \quad (10)$$

$$\begin{aligned} \bar{C} \simeq & \frac{1}{\ln(2) \Gamma(\alpha + 1)} \left( \frac{\Gamma(\alpha - 1) {}_2F_3\left(1, 1; 2, 1 - \frac{\alpha}{2}, \frac{3}{2} - \frac{\alpha}{2}; -\frac{1}{4\beta^2 \rho_0}\right)}{\beta^2 \rho_0} + \frac{\pi \beta^{-\alpha-2} \rho_0^{-\frac{\alpha}{2}-1} \csc\left(\frac{\pi\alpha}{2}\right) {}_1F_2\left(\frac{\alpha}{2} + 1; \frac{3}{2}, \frac{\alpha}{2} + 2; -\frac{1}{4\beta^2 \rho_0}\right)}{\alpha + 2} \right. \\ & + \frac{\pi \beta^{-\alpha-1} \rho_0^{-\frac{\alpha}{2}-\frac{1}{2}} \sec\left(\frac{\pi\alpha}{2}\right) {}_1F_2\left(\frac{\alpha}{2} + \frac{1}{2}; \frac{1}{2}, \frac{\alpha}{2} + \frac{3}{2}; -\frac{1}{4\beta^2 \rho_0}\right)}{\alpha + 1} - 2\alpha^2 \Gamma(\alpha - 1) \ln\left(\frac{1}{\beta\sqrt{\rho_0}}\right) + 2\alpha \Gamma(\alpha - 1) \ln\left(\frac{1}{\beta\sqrt{\rho_0}}\right) \\ & \left. + 2(\alpha - 1) \alpha \Gamma(\alpha - 1) \psi^{(0)}(\alpha + 1) \right) \end{aligned} \quad (18)$$

By substituting (12) in (16) and using [21, Eq. 3.462.1], we obtain

$$\begin{aligned} \bar{P}_E = & \frac{1}{2\beta^{\alpha+1} \Gamma(\alpha + 1)} \sum_{r=1}^R a_r (\rho_0 \zeta b_r)^{-\frac{\alpha+1}{2}} \left( \Gamma\left(\frac{\alpha + 1}{2}\right) \right. \\ & \times {}_1F_1\left(\frac{\alpha + 1}{2}, \frac{1}{2}, \frac{1}{4\beta^2 \rho_0 \zeta b_r}\right) - \left(\beta^2 \rho_0 \zeta b_r\right)^{-\frac{1}{2}} \\ & \left. \times \Gamma\left(\frac{\alpha}{2} + 1\right) {}_1F_1\left(\frac{\alpha}{2} + 1, \frac{3}{2}, \frac{1}{4\beta^2 \rho_0 \zeta b_r}\right) \right), \end{aligned} \quad (17)$$

for which  $\alpha$  and  $\beta$  are defined respectively in (5) and (6).

The ergodic capacity of the considered system has the same analytical form as [14, Eq. 11] that is rewritten in (18) with substituting novel expressions of  $\alpha$  and  $\beta$ , which are calculated herein using the mean and variance of the combined channel response in (8) and (10), respectively. The  $\psi^{(0)}(\cdot)$  in (18) is the 0th polygamma function and  $\csc(\cdot)$  is the cosecant function.

#### IV. NUMERICAL RESULTS AND DISCUSSIONS

This section gives insight into the performance of the considered system in terms of the outage probability, ASEP and ergodic capacity. In addition, it verifies the accuracy of the adopted Laguerre series approximation by means of Monte Carlo simulations. We assume five different RISs ( $N = 5$ ) whose number of REs is given as  $\{M_n\}_{n=1}^N = \{14, 26, 16, 24, 20\}$  or  $\{M_n\}_{n=1}^N = \{28, 52, 32, 48, 40\}$ . Also,  $M$  refers to the total number of REs in all the  $N$  distributed RISs, i.e.,  $M = \sum_{n=1}^N M_n$ . Thus,  $M = 100$  and  $M = 200$  for the two considered cases. For calculating the average gains  $\sigma_{h_n}^2, \sigma_{g_n}^2, \sigma_u^2$ , we set  $d_0 = 1$  m,  $\iota_0 = -30$  dB,  $\eta_{h_n} = 2.4, \eta_{g_n} = 2.3$  for all  $n = 1, 2, \dots, 5$  and  $\eta_u = 3$ . The RISs are assumed to be distributed between S and D which are located in the  $x$ -axis and separated by a distance  $d_u = 100$  m. The location of each RIS is given in the Cartesian coordinate system as  $(d_{x_n}, d_{y_n})$  and the total dis-

tances of the links are calculated as  $d_{h_n} = \sqrt{d_{x_n}^2 + d_{y_n}^2}$  and  $d_{g_n} = \sqrt{(d_u - d_{x_n})^2 + d_{y_n}^2}$ .

Unless otherwise stated, we consider the location setting  $D = [(25, 50), (40, 30), (55, 10), (82, -20), (95, -40)]$  m and S-RIS <sub>$n$</sub> -D paths' distributions with

$\kappa_{h_1} = 0, \mu_{h_1} = 1, \kappa_{g_1} = 0, \mu_{g_1} = 1$  (double Rayleigh),  
 $\kappa_{h_2} = 0, \mu_{h_2} = 3, \kappa_{g_2} = 0, \mu_{g_2} = 2$  (double Nakagami),  
 $\kappa_{h_3} = 2, \mu_{h_3} = 1, \kappa_{g_3} = 2, \mu_{g_3} = 1$  (double Rician),  
 $\kappa_{h_4} = 1, \mu_{h_4} = 2, \kappa_{g_4} = 1, \mu_{g_4} = 2$  (double  $\kappa - \mu$ ), and  
 $\kappa_{h_5} = 2.5, \mu_{h_5} = 1, \kappa_{g_5} = 0, \mu_{g_5} = 3.3$  (Rician-Nakagami).

The accuracy of the first-term Laguerre approximation (4) for the end-to-end channel's PDF of the considered system model with and without direct path between S and D is tested and illustrated in Fig. 2. It can be noted that the used approximation is very tight for both communication scenarios (with or without direct path) and for any combination of the fading distributions, where we verified its accuracy over two fading scenarios; all links experience Rician fading or each RIS experiences different fading distribution using the setting specified above. The high accuracy is maintained for low and high numbers of the RISs' REs. The communication scenario, where only a S-D link exist, is also presented for comparison and it shows that imposing the RISs in the system increases its power gain which increases even further by increasing  $M$  as can be depicted from the right-shifting of the PDF.

Figure 3 depicts the impact of using RISs to assist the communication between S and D and enhance the different performance metrics. In particular, the outage probability, ASEP and the ergodic capacity, whose analytical values coincide well with the true measures, show much better performance when compared to the scenario where communication is achieved only through the direct path. In addition, the impact of increasing the number of REs equipped on the distributed RISs is clearly noted where as  $M$  increases, the outage probability and ASEP decrease and the ergodic capacity increases, indicating improved performance, i.e., less transmitted power is required

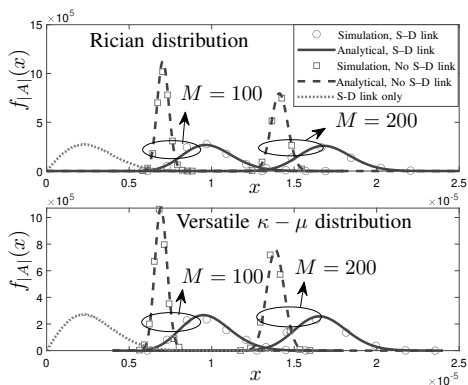


Fig. 2. The PDF of the end-to-end channel with and without S-D link for  $N = 5$  of two different RISs systems.

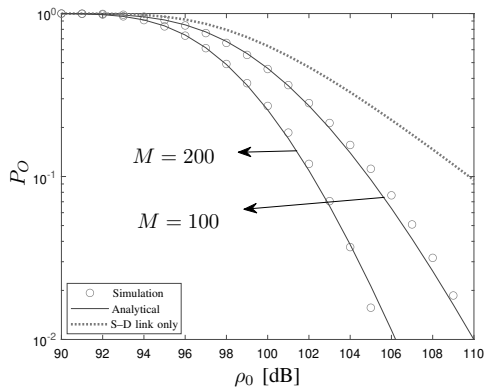
to achieve a certain level of the considered measure.

The effect of increasing  $M$  on the different orders of the considered  $M$ -QAM scheme in Fig. 3(b) is nearly the same, e.g., for ASEP of 10%, an increment by 100 REs will decrease the required transmitted power by approximately 2.2 dB for both schemes. Moreover, it can be noted from Fig. 3(a) and (b), that as  $M$  increases, the rate of change in the slope of the outage probability and the ASEP increases which indicates higher diversity gain.

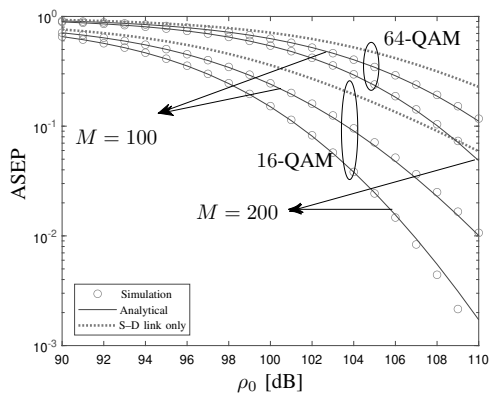
Finally, we demonstrate the impact of the locations of the  $N$  distributed RISs to the system's performance. To give a better insight into it, we test the  $x$ -position and the  $y$ -position separately, while keeping the other dimension's position constant. In particular, in Fig. 4(a), we choose three different location settings for the five distributed RISs as indicated by the three different marker symbols in the smaller subfigure to represent the different possibilities of movements along the  $x$ -axis. The corresponding ASEP is calculated and plotted. We conclude from the figure that as the  $x$ -position of the RISs is nearer to either S or D, better performance is achieved. On the other hand, placing the RISs near the half-way between S and D results in worse performance since the path losses for both hops are maximized. Similarly, the  $y$ -placement of the RISs is also tested in Fig. 4(b) and shows better performance when the RISs are placed vertically closer to S and D, where the path losses are less and thus they contribute more efficiently to the communication process.

## V. CONCLUSION

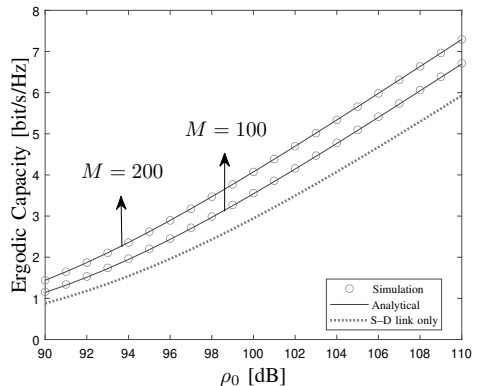
This paper studied the performance of a generalized system setup, namely, a SISO communication system with multiple RISs and direct path between the source and the destination over the generic  $\kappa$ - $\mu$  fading channels. Specifically, it presented tight expressions for the corresponding outage probability, average symbol error probability and ergodic capacity. The considered fading distribution includes most of the widely used fading models. This validates the use of all the derived



(a) Outage probability,  $\rho_{th} = 10$  dB



(b) Average symbol error probability



(c) Ergodic capacity

Fig. 3. The outage probability, average symbol error probability and ergodic capacity for different values of  $M$ , i.e., the total number of REs.

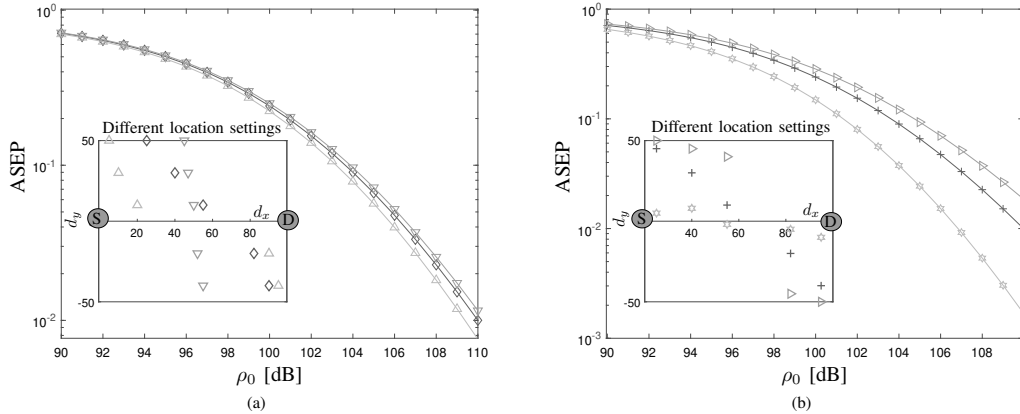


Fig. 4. Impact of the  $x$ -position in (a) and the  $y$ -position in (b) of the  $N$  distributed RISs to the ASEP, while keeping the other dimension's position constant.

expressions for these special cases. The numerical results verified the performed statistical analysis and confirmed the high accuracy of the derived performance measures. Moreover, we showed that increasing the number of reflecting elements equipped on the RISs and placing them closer either to the source or destination, improve the system's performance significantly and increase its diversity gain.

#### REFERENCES

- [1] H. Yang *et al.*, "A programmable metasurface with dynamic polarization, scattering and focusing control," *Sci. Rep.*, vol. 6, Oct. 2016.
- [2] N. Kaina, M. Dupré, G. Lerosey, and M. Fink, "Shaping complex microwave fields in reverberating media with binary tunable metasurfaces," *Sci. Rep.*, vol. 4, no. 6693, Oct. 2014.
- [3] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wirel. Commun.*, vol. 18, no. 8, pp. 4157–4170, Jun. 2019.
- [4] Q. Wu and R. Zhang, "Beamforming optimization for intelligent reflecting surface with discrete phase shifts," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, May 2019, pp. 7830–7833.
- [5] C. Pan, H. Ren, K. Wang, W. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Trans. Wirel. Commun.*, Jun. 2020.
- [6] M. Al-Jarrah, A. Al-Dweik, E. Alsusa, Y. Idrissi, and M.-S. Alouini, "IRS-assisted UAV communications with imperfect phase compensation," *IEEE Trans. Wirel. Commun.*, 2021.
- [7] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, "Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 14960–14973, Oct. 2020.
- [8] A. Almoahamad, A. M. Tahir, A. Al-Kababji, H. M. Furqan, T. Khattab, M. O. Hasna, and H. Arslan, "Smart and secure wireless communications via reflecting intelligent surfaces: A short survey," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1442–1456, Sep. 2020.
- [9] V. C. Thirumavalavan and T. S. Jayaraman, "BER analysis of reconfigurable intelligent surface assisted downlink power domain NOMA system," in *Proc. Int. Conf. on Commun. Syst. Netw.*, Mar. 2020.
- [10] M. Badiu and J. P. Coon, "Communication through a large reflecting surface with phase errors," *IEEE Wirel. Commun. Lett.*, vol. 9, no. 2, pp. 184–188, Feb. 2020.
- [11] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, Aug. 2019.
- [12] D. Kudathanthirige, D. Gunasinghe, and G. Amarasinghe, "Performance analysis of intelligent reflective surfaces for wireless communication," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2020.
- [13] A.-A. A. Boulogeorgos and A. Alexiou, "Performance analysis of reconfigurable intelligent surface-assisted wireless systems and comparison with relaying," *IEEE Access*, vol. 8, pp. 94463–94483, May 2020.
- [14] A. Salhab and M. Samuh, "Accurate performance analysis of reconfigurable intelligent surfaces over Rician fading channels," *IEEE Wirel. Commun. Lett.*, vol. 10, no. 5, pp. 1051–1055, May 2021.
- [15] Q. Tao, J. Wang, and C. Zhong, "Performance analysis of intelligent reflecting surface aided communication systems," *IEEE Commun. Lett.*, vol. 24, no. 11, pp. 2464–2468, Nov. 2020.
- [16] D. L. Galappaththige, D. Kudathanthirige, and G. Amarasinghe, "Performance analysis of distributed intelligent reflective surface aided communications," in *Proc. IEEE Glob. Commun. Conf.*, Dec. 2020.
- [17] L. Yang, Y. Yang, D. B. da Costa, and I. Trigui, "Outage probability and capacity scaling law of multiple RIS-aided networks," *IEEE Wirel. Commun. Letters*, vol. 10, no. 2, pp. 256–260, Feb. 2021.
- [18] S. Primak, *Stochastic Methods and Their Applications to Communications: Stochastic Differential Equations Approach*. Wiley, 2004.
- [19] M. D. Yacoub, "The  $\kappa$ - $\mu$  distribution and the  $\eta$ - $\mu$  distribution," *IEEE Antennas Propag. Mag.*, vol. 49, no. 1, pp. 68–81, Feb. 2007.
- [20] N. Bhargava and Y. J. Chun, "On the product of two  $\kappa$ - $\mu$  random variables and its application to double and composite fading channels," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 4, pp. 2457–2470, Apr. 2018.
- [21] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. Elsevier/Academic Press, 2007.
- [22] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels*, 2nd ed. John Wiley and Sons, Inc., Jan. 2005.
- [23] I. M. Tanash and T. Riihonen, "Global minimax approximations and bounds for the Gaussian  $Q$ -function by sums of exponentials," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6514–6524, Oct. 2020.
- [24] I. M. Tanash and T. Riihonen, "Coefficients for global minimax approximations and bounds for the Gaussian  $Q$ -function by sums of exponentials," Jul. 2020. [Online]. Available: <https://doi.org/10.5281/zenodo.4112978>





