

NONLINEAR SIGNAL PROCESSING FOR VISIBLE LIGHT COMMUNICATION

Ph.D. Thesis

by

RANGEET MITRA



**DISCIPLINE OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE**

JANUARY 2017

**NONLINEAR SIGNAL PROCESSING FOR VISIBLE
LIGHT COMMUNICATION**

A THESIS

*Submitted in partial fulfillment of the
requirements for the award of the degree*

of

DOCTOR OF PHILOSOPHY

by

RANGEET MITRA



DISCIPLINE OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

JANUARY 2017



INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled “**NONLINEAR SIGNAL PROCESSING FOR VISIBLE LIGHT COMMUNICATION**” in the partial fulfillment of the requirements for the award of the degree of DOCTOR OF PHILOSOPHY and submitted in the DISCIPLINE OF ELECTRICAL ENGINEERING, Indian Institute of Technology Indore, is an authentic record of my own work carried out during the time period from January 2014 to January 2017 under the supervision of Dr. Vimal Bhatia, Associate Professor, Indian Institute of Technology Indore, India.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Signature of the student with date

(Rangeet Mitra)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Signature of Thesis Supervisor with date

(Dr. Vimal Bhatia)

Rangeet Mitra has successfully given his Ph.D. Oral Examination held on

Signature of Thesis Supervisor

Convener, DPGC

Date:

Date:

Signature of PSPC Member:1

Signature of PSPC Member:2

Signature of External Examiner

Date:

Date:

Date:

ACKNOWLEDGEMENTS

I am immensely grateful to my PhD supervisor Dr. Vimal Bhatia for his invaluable guidance, encouragement and direction throughout this work. He has actively supported my research and has done everything humanly possible to support my research in IIT Indore. He has helped me in transforming myself into a better researcher, paper-writer and presenter. Further, I will be forever thankful to him for hiring me for a PhD position after my return from Turkey. I would also take this opportunity to thank my PSPC members Dr. Somnath Dey and Dr. Shaibal Mukherjee who have mentored my PhD thesis via interactions in yearly research progress seminars. Their suggestions and insights were really helpful in shaping my thesis and in forging directions of future work. Further, interactions with Dr. Murat Uysal and Mr. Farshad Miramirkhani acquainted me with the latest research in VLC and were quite useful in shaping my thesis towards something substantial.

I would thank Ministry of Human Resource Development (MHRD), Govt. of India and IIT Indore for providing stipend regularly and marginally hiking it when we were struggling economically. I would also take this opportunity to thank IIT Indore for partially funding my conference travels. These conferences helped me meet new people who have helped shape my PhD thesis. I will also thank the finance, administration, academic and R&D sections for all the support. The support from hostels, transport and mess has helped ease the research pressure significantly.

I would like to thank my math teachers Mr. Sanjay Misra and Mr. Syed Mohammad Omair of DPS Ranchi who have mentored my mathematics. Mr. Misra was the one who aroused my interest in mathematics and encouraged me to delve deeper into mathematics. Mr. Omair tutored me in Class 12 and made Calculus one of my strengths which has helped me a lot throughout my research.

Finally, I would like to thank my parents for their constant care and support. Interactions with Dr. Pankaj Sharma have been really insightful and morale-boosting. He always supported me when I was down and out, and encouraged me to hang in there and wait for my time to strike. Finally, I thank Mr. Nimesh Tiwary of Silver Springs, Phase-II for his

wonderful tea (and company) and two street dogs who have showered me with selfless affection.

RANGEET MITRA

ABSTRACT

Recently many research efforts have been directed towards communication in visible light spectrum (visible light communication (VLC)) keeping in mind bandwidth hungry future 5G and beyond systems. However, the performance of VLC systems is degraded by channel-impairments like inter-symbol interference (ISI) and inherent device impairments like light emitting diode (LED) nonlinearity which limits its overall throughput. To mitigate these impairments, and to maximize throughput, adaptive signal processing techniques are proposed in this work which do not need explicit knowledge of the VLC channel nonlinearity. In other words, online algorithms are proposed in this work, which adapts, learns and inverts the channel impairments. This adaptive learning can be useful in tracking scenarios when there is aging in the front-end devices/hardware. Additionally, there are savings in the manufacturing cost of the optical system hardware as these proposed techniques relax the minimum required tolerance in manufacturing front-end hardware. Finally, reliability of the overall VLC link increases as these signal processing techniques help in maintaining the overall signal to noise ratio in the presence of instantaneous outages. This thesis aims at developing signal processing techniques for boosting the throughput of the VLC system. In order to boost the achievable data rate in VLC system, the first work in this thesis proposes a Chebyshev polynomial based nonlinear pre-distortion technique for mitigating LED nonlinearity, and ISI of the VLC channel. Better bit error rate (BER) characteristics were found upon the use of Chebyshev pre-distortion as compared to popular linear normalized least mean squares (NLMS) based pre-distortion. The results obtained in this work may be useful for benchmarking with respect to a “perfect” scenario. However, the performance gains achieved in this work rely on perfect knowledge of the detected symbols at the receiver to be relayed to the transmitted symbols in the uplink which is not feasible in general. The second work focuses on open-loop VLC system in which the problem of improving the throughput of the VLC system is solved by open-loop post-distortion at the receiver. A novel sparse novelty criterion based kernel minimum symbol error rate (KMSER) post-distorter (or equalizer) based on reproducing kernel Hilbert space (RKHS) techniques is proposed in which savings in

computational complexity and superior BER characteristics are demonstrated. This work establishes the RKHS based post-distorter to be a better alternative for post-distortion as compared to Volterra post-distorters. The third work explores further techniques to bring the computational complexity of the sparse RKHS-based post-distorter down further by proposing fixed-budget based dictionary pruning criterion. It is demonstrated from this work that same BER performance is achieved by fixed-budget KMSE over impaired VLC channels with lower dictionary size as compared to novelty-criterion based KMSE technique. Exact mathematical insights on the transient performance of the dynamics of learning curves of fixed budget-KMSE are drawn in this work. The next work revisits the problem of post-distortion over VLC channels by exploring unsupervised techniques for post-distortion. In this work, the proposed multi-stage clustering based Hammerstein post-distorter is compared with the widely used modified cascaded-MMA (MCMMA)-based Volterra post-distorter. Superior convergence characteristics are obtained with the use of multi-stage Hammerstein based post-distortion. The steady-state mean squared error (MSE) characteristics of the multi-stage post-distorter is analyzed mathematically and many insights from the analyzed steady-state behavior are derived.

Finally, a multiple-input multiple-output (MIMO)-multi-user scenario is considered with non-orthogonal multiple access (NOMA) being the multiple-access technique. A special NOMA-scenario is considered where all users have correlated channel matrices (and hence similar channel conditions). A novel precoding technique and its corresponding power-allocation strategy is derived. This is the first effort in direction of NOMA in VLC being extended to arbitrary number of users with similar channel conditions. Based on these precoding techniques, expression for BER is derived for square-quadrature amplitude modulation (QAM) and insights are provided. Above all, the signal processing algorithms provided in this thesis are aimed at maximizing the overall throughput of the VLC system in presence of impairments for 5G and beyond communication systems.

Contents

ABSTRACT	i
LIST OF FIGURES	viii
LIST OF TABLES	xii
LIST OF ABBREVIATIONS	xiii
1 Introduction	1
1.1 Overview of visible light communication	2
1.2 Motivation, objective and significance	3
1.3 Thesis outline and contributions	5
2 Background	7
2.1 Overview of visible light communication	7
2.2 Components and applications of visible light communications	8
2.3 VLC channel impairments	10
2.3.1 LED nonlinearity impairments	10
2.3.2 Channel impairments	11
2.3.3 Performance measures	11
2.4 Review of classical LMS based adaptive filtering	14
2.5 Existing signal processing techniques for VLC	17
2.6 Potential of RKHS based adaptive filtering	19
2.7 Review of KLMS	20

2.8	Multiple-access for VLC	22
2.9	Conclusion	24
3	Chebyshev Polynomial based Adaptive Pre-distorter for VLC	27
3.1	Introduction	27
3.2	NLMS based pre-distortion	28
3.3	Chebyshev polynomial-expansion based pre-distortion in frequency-flat channels	30
3.4	Simulations	32
3.5	Summary	37
4	Adaptive Dictionary based Minimum Symbol Error Rate Post-distorter in VLC	39
4.1	Introduction	39
4.2	System Model	41
4.3	Review of KMSE based technique for nonlinear channel equalization . .	42
4.4	Sparsified-KMSE using novelty criterion based post-distortion for VLC channels	44
4.5	Theoretical analysis of KMSE	45
4.5.1	MSE transient dynamics for KMSE	46
4.5.2	Step-size range for convergence	48
4.5.3	Excess mean squared error analysis	49
4.5.4	Convergence analysis of KMSE	49
4.5.5	Does the analysis hold under sparsification of dictionary?	50
4.6	Simulations	52
4.7	Summary	57
5	Finite Dictionary Techniques for MSER Equalization in RKHS	59
5.1	Introduction	59
5.2	System model	60
5.3	Quantized kernel MSER	61

5.4	Fixed budget quantized kernel MSER	62
5.5	Expression for transient and steady-state behavior of QKMSER and FBQKMSER	64
5.5.1	Step-size range for convergence	67
5.5.2	Steady state misadjustment	68
5.6	Simulations	69
5.7	Summary	73
6	Unsupervised Multi-Stage Clustering based Hammerstein Post-distortion for VLC	75
6.1	Introduction	75
6.2	System model	77
6.3	Proposed normalized Hammerstein improved multi-stage clustering based post-distorter	78
6.3.1	Derivation of step-size range for convergence	80
6.3.2	Steady-state MSE vs step-size	83
6.4	Simulations	84
6.5	Summary	87
7	Precoded Chebyshev-NLMS based pre-distorter for nonlinear LED compen- sation in NOMA-VLC	89
7.1	Introduction	89
7.2	MIMO-NOMA system model	93
7.3	Review of QR-precoded NOMA	95
7.4	Proposed extension of Chebyshev pre-distortion to MIMO systems	96
7.5	Choice of precoding matrices	99
7.6	Power allocation strategy	100
7.6.1	Choice of $\lambda^{(u)}$ for each user	101
7.7	Analytical expression for BER of square M -QAM	102
7.8	Simulations	104

CONTENTS

7.8.1	Mixed constellation for two user scenario	106
7.8.2	<i>U</i> -user scenario	106
7.9	Summary	111
8	Conclusion and future work	113
8.1	Conclusion	113
8.2	Future work	114
	REFERENCES	116
	LIST OF PUBLICATIONS	129

List of Figures

2.1	Example of VLC deployment.	9
2.2	CIR of a VLC channel for “Open-Office” scenario.	12
2.3	CIR of a VLC channel for “office with cubicles” scenario.	12
2.4	Nonlinear LED transfer characteristics.	13
2.5	Application of adaptive filtering for system identification [26].	16
2.6	Application of adaptive filtering for system inversion [26].	16
2.7	RKHS adaptive filter shown as an equivalent RBF network.	21
2.8	A typical NOMA-VLC setup.	24
3.1	Block diagram of the proposed system model.	30
3.2	Performance comparison of the proposed approach with standard approaches in quadratic nonlinearity.	35
3.3	Performance comparison of the proposed approach with standard approaches in Rapp nonlinearity.	35
3.4	Performance comparison of the proposed approach with standard approaches in quadratic nonlinearity in precoded ISI environment for 16-QAM.	36
3.5	Performance comparison of the proposed approach with standard approaches in Rapp nonlinearity in precoded ISI environment for QPSK.	36
3.6	Performance comparison of the proposed approach with standard approaches in Rapp nonlinearity in precoded ISI environment for 4-PAM.	37
4.1	Block diagram of the used system model.	42

LIST OF FIGURES

4.2	BER and computational complexity comparison for 4.2(a) open office IEEE 802.15 PAN channel for 4-PAM. 4.2(b) office with cubicles IEEE 802.15 PAN channel for 4-PAM.	54
4.3	BER and computational complexity comparison for 4.3(a) open office IEEE 802.15 PAN channel for 16-QAM. 4.3(b) office with cubicles IEEE 802.15 PAN channel for 16-QAM.	55
4.4	4.4(a) Transient and steady-state MSE analysis of KMSER for “open office” channel for 16-QAM and 4.4(b) Transient and steady-state MSE analysis of KMSER for “office with cubicles” channel for 16-QAM. . . .	56
5.1	BER and computational complexity comparison of FBQKMSER with other sparsification approaches over IEEE 802.15 PAN ‘Office’ indoor VLC channel.	70
5.2	BER and computational complexity comparison of FBQKMSER with other sparsification approaches over IEEE 802.15 PAN ‘Office with cubicles’ indoor VLC channel.	70
5.3	Theoretical validation of MSE for FBQKMSER over IEEE 802.15 PAN ‘Office’ indoor VLC channel (for LED non-linearity).	71
5.4	Theoretical validation of MSE for FBQKMSER over IEEE 802.15 PAN ‘Office with cubicles’ indoor VLC channel (for LED non-linearity). . . .	72
6.1	Block diagram of the used system model.	78
6.2	MSE comparison for open office IEEE 802.15 PAN channel for 16-QAM.	85
6.3	MSE comparison for office with cubicles IEEE 802.15 PAN channel for 16-QAM.	85
6.4	BER comparison for open office over IEEE 802.15 PAN channel for 16-QAM.	86
6.5	BER comparison for office with cubicles over IEEE 802.15 PAN channel for 16-QAM.	86

6.6	Steady-state MSE vs step-size for open office over IEEE 802.15 PAN channel for 16-QAM.	86
6.7	Steady-state MSE vs step-size for office with cubicles over IEEE 802.15 PAN channel for 16-QAM.	87
7.1	Block diagram of the proposed system model.	95
7.2	2-user scenario with full CSI. One 4-QAM, another user 16-QAM. Chebyshev pre-distortion outperforms NLMS in SVD-precoded scenario which is designed keeping QoS requirements in consideration. Performance is shown for User-2 and User-1.	107
7.3	Convergence comparison for proposed SVD precoded Chebyshev pre-distorter and SVD precoded NLMS pre-distorter for User-2 at 25dB SNR in a two user scenario.	107
7.4	Convergence comparison for proposed SVD precoded Chebyshev pre-distorter and SVD precoded NLMS pre-distorter for User-1 at 40dB SNR in a two user scenario.	108
7.5	Sum-Rate plot for N -user scenario. After 3-users the sum-rate decays drastically. Specially, for the N -user scenario, all users are assigned 4-QAM.	108
7.6	Average BER plotted as a function of SNR for $U = 2$ for the proposed precoding technique. All users use 4-QAM modulation scheme.	109
7.7	Average BER plotted as a function of SNR for $U = 3$ for the proposed precoding technique. All users use 4-QAM modulation scheme.	109
7.8	Average BER plotted as a function of SNR for $U = 2,3$ for the proposed precoding technique. All users use 16-QAM modulation scheme in full CSI scenario.	110

List of Tables

7.1 Simulation Parameters 105

List of Abbreviations

AWGN additive white Gaussian noise.

BER bit error rate.

CIR channel impulse response.

CKLMS complex kernel least mean squares.

CMA constant-modulus algorithm.

CMMA cascaded-MMA.

CSI channel state information.

DFE decision feedback equalizer.

EMSE excess mean squared error.

FBQKMSER fixed budget quantized kernel MSER.

FEC forward error correction.

FIR finite impulse response.

FOV field of view.

HIMSC Hammerstein improved multi-stage clustering equalizer.

i.i.d independently identically distributed.

IMSC improved multi-stage clustering.

IoT internet of things.

IP internet protocol.

ISI inter-symbol interference.

KLMS kernel least mean squares.

KMSER kernel minimum symbol error rate.

LED light emitting diode.

LMS least mean squares.

LOS line of sight.

LUT lookup table.

MBER minimum bit error rate.

MCMMA modified cascaded-MMA.

MIMO multiple-input multiple-output.

MISO multiple-input single-output.

MMA multi-modulus algorithm.

MMSE minimum mean squared error.

MRC maximal ratio combining.

MSE mean squared error.

MSER minimum symbol error rate.

NAMBER normalized-adaptive MBER.

NCKLMS1 normalized CKLMS1.

NCKLMS2 normalized CKLMS2.

NLMS normalized least mean squares.

NOMA non-orthogonal multiple access.

p.d.f probability density function.

PAM pulse amplitude modulation.

PAPR peak to average power ratio.

PD photo-diode.

PLC power-line communication.

QAM quadrature amplitude modulation.

QKMSER quantized kernel MSER.

QoS quality of service.

QPSK quadrature phase shift keying.

RAM Random Access Memory.

RBF radial basis function.

RF radio frequency.

RKHS reproducing kernel Hilbert space.

RLS recursive least squares.

SER symbol error rate.

SIC successive interference canceller.

SINR signal to interference and noise ratio.

SISO single-input single-output.

SNR signal to noise ratio.

SVD singular value decomposition.

UE user equipment.

VLC visible light communication.

VNI visual network index.

List of Mathematical Symbols

$(\cdot)^T$	Transpose operator
$(\cdot)^{(u)}$	Quantity for the u^{th} user in a multi-user scenario
η	Step-size
$\hat{(\cdot)}$	Estimated value of the parameter (\cdot)
$\mathbb{E}[\cdot]$	Statistical expectation operator
\mathcal{D}_k	Dictionary of observations at time instant k
\mathcal{H}	Reproducing kernel Hilbert space
\mathcal{H}_σ	RKHS corresponding to kernel width σ
\mathcal{I}_k	Dictionary of error terms at time instant k
Ω_k	Adaptive post-distorter weight at time instant k
$\mathbf{x}(q)$	q^{th} component of the vector \mathbf{x}
\mathbf{x}_k	Vector of observations at time instant k
e_k	minimum mean squared error (MMSE) error term
I_k	minimum symbol error rate (MSER) error term
k	Time index
$r_k^{(i)}$	i^{th} pre-distorter weight

List of Abbreviations

- s_k Transmitted symbol at time instant k
- T_n Chebyshev polynomial of first kind of order n
- x_k Scalar at time instant k

Chapter 1

Introduction

The next generation of communication system would witness a paradigm shift in basic principles that govern the communication-link design in order to cater to the ever increasing demand for bandwidth in order to accommodate large number of users [1]. The estimated internet protocol (IP) traffic can be approximated to exceed 500 exabytes (1 exabyte= 10^{18} bytes) by 2020 [1]; which shows the rate at which the demand for bandwidth is simply growing with leaps and bounds. For example in applications like internet of things (IoT), there could be innumerable bandwidth-hungry devices within a single attocell which needs to be serviced at the same time with different levels of quality of service (QoS). In such scenarios, the visual network index (VNI) report published in 2014 indicates that an incremental change over 4G systems would not be able to meet future bandwidth requirements [1]. Thus, it is certain that well-established techniques for communication in the 4G standard will be outdated and considerable research efforts must be channelized in the direction of providing for high-speed communications for higher number of users. This need has initiated many proposals for increasing the available spectrum like moving the spectrum for communications towards mm-wave regime, nm-wave regime (VLC) and massive-MIMO.

One of the proposed techniques for meeting increasing demand for bandwidth is VLC [2]. In VLC, the LED are used for dual purposes of illumination and as a wireless transmitter. The intensity of the LED is varied at a rate imperceptible to the human eye in

accordance with the input modulating signal. This serves the dual purposes of illumination and signal transmission. This mode of communication is also environment friendly and hence belongs to the category of green communications. VLC systems/links have been reported in the literature that achieve speeds of 3Gbps and hence promise to address the bandwidth requirement of 5G systems (particularly 5G indoor attocells).

1.1 Overview of visible light communication

VLC is a technique of communication in which transmission of data is achieved by modulating the LED at high rates without affecting the primary purpose of illumination [3]. The purpose of illumination is unaffected as the intensity of the LED is modulated at a rate above the flicker fusion threshold [3, 4]. LEDs are ubiquitous in our surroundings. If we look around, we have LED lamps for lighting our rooms/offices, traffic lights, headlights of cars, billboards etc. In future, LEDs will dominate the market for illumination purposes and will be widely prevalent due to their cost-effectiveness [3]. The integration of using LED intensity modulation based VLC has the potential to complement radio frequency (RF)-spectrum and converge towards development of smart cities [5]. There are several advantages offered by VLC over RF communications, which are enlisted:

- The 2.4 GHz industrial, scientific and medical (ISM) band is getting increasingly crowded as compared to the less occupied visible-light spectrum. Thus VLC encounters lesser interference and facilitates better link quality.
- As light does not penetrates through the walls, VLC bands have better spectrum reusability and prevents eavesdropping/jamming from neighboring users.
- The front end devices of VLC based systems do not suffer from phase noise and IQ imbalance as RF devices, and are also cheap and easily available.
- As wavelength in visible band is in the sub-micron range the variance of channel estimation algorithms are much lower in general. This also helps in accurate positioning of the receiver for better link performance.

Due to these advantages of VLC, it has been chosen as one of the dominant research-directions which has the potential to cater to the ever increasing bandwidth requirements of 5G and beyond systems in the near future [6].

1.2 Motivation, objective and significance

Motivation

In spite of the advantages of VLC, the overall throughput of the VLC based systems is significantly reduced by channel impairments like ISI, and LED nonlinearity, which lower the overall effective signal to noise ratio (SNR). In order to achieve the promised throughput for VLC systems, and to maintain the SNR in severely degraded channel conditions many signal processing techniques have been suggested recently. It was noted that the signal processing techniques for VLC are of two types: a) open-loop post-processing at receiver, and b) closed-loop processing between receiver and transmitter. By use of both types of processing techniques, significant gains in performance of VLC systems can be observed. However, they are mostly based on heuristics and are not propounded by an optimality criterion. For example, one can peruse a vast amount of literature on Volterra adaptive equalizer [7, 8], which is generally the preferred solution, is well known to be subject to local minima and model order dependent. Motivated by these challenges, this thesis aims at finding nonlinear and convex solutions to the nonlinear VLC-channel effects. The specific motivation for each chapter is elaborated in the individual chapters.

Objectives

The objectives of this thesis are as follows:

- To propose signal processing algorithms for nonlinear, ISI-impaired VLC systems that are motivated by an optimality criterion as opposed to existing heuristic based detectors.
- To validate the performance of the proposed signal processing algorithms with re-

spect to the achieved BER floor and MSE characteristics.

- To theoretically analyze both the transient and steady-state behavior of the learning metric for proposed algorithms where applicable.

With these goals in mind, this thesis provides various signal processing techniques to mitigate the undesirable channel effects in VLC-systems. Also theoretical analysis is presented in this thesis to validate these signal processing techniques against the existing literature. First a closed-loop adaptive Chebyshev pre-distorter is proposed which outperforms the classical linear NLMS-based pre-distorter, which is motivated by a optimal min-max approximation error guarantee. Next, it is found that the exact knowledge of detected symbols is too idealistic an assumption (however the “idealistic” assumption may be used for benchmarking or as a bound) and therefore post-distortion techniques are proposed using RKHS-techniques based on the MSER-criterion which are motivated from notions of convexity and hence are global solutions. Theoretical analysis of the learning curves (MSE) of the proposed RKHS-based techniques is performed and simulations are performed to validate the theory. Consequently, post-distortion is approached from an unsupervised setting using the multi-stage clustering based approach and the blind multi-stage clustering cost function is found to be intimately linked to the correntropy criterion. Finally, the problem of pre-distortion is re-visited from a multi-user perspective in which NOMA being used as a multiple-access technique in a MIMO-setting. A hybrid closed-loop Chebyshev precoding is proposed for the NOMA-VLC system and BER-performance of the precoding technique is analyzed theoretically for square-QAM modulation. The insights gained from the analysis of the proposed algorithms guide the user towards choice of suitable values for parameters for the proposed algorithms for an improved VLC link in general.

Significance

VLC systems which are targeted for many 5G and beyond systems [2, 9–11], have attracted a lot of attention recently. However, these VLC systems significantly fail to deliver the promised throughput due to channel impairments [7] and without proper signal

processing techniques. To counteract the channel impairments, latest research in this domain is motivated mostly by sub-optimal signal processing techniques. All the works in this thesis are motivated by an optimality criterion; hence they exhibit superior performance than existing signal processing techniques, and may be used as an integral part of proposed 5G and beyond systems.

1.3 Thesis outline and contributions

The main aim of this thesis is to propose signal processing techniques to counter VLC-channel impairments, and analyze their learning curves and other performance measures theoretically. The organization of the thesis may be outlined as follows:

- Chapter 2 introduces the reader to VLC based systems, and briefly reviews the concepts of pre-distortion and post-distortion, linear and nonlinear adaptive filtering, and multiple access techniques in VLC.
- Chapter 3 investigates a closed-loop Chebyshev polynomial based pre-distorter motivated by NLMS algorithm. The choice of Chebyshev polynomial is motivated by min-max approximation error optimality criterion and superior BER characteristics as compared to linear pre-distorter is observed via simulations. This chapter assumes perfect knowledge of detected symbols at the receiver to be relayed to the transmitter via a feedback path which may be too idealistic in practice.
- To circumvent this idealistic assumption, Chapter 4 explores post-distortion techniques based on the RKHS-based minimum symbol error rate criterion by open-loop processing done at the receiver. A new KMSER-based post-distortion technique is proposed. Theoretical analysis for the transient and steady-state MSE curves is carried out. Asymptotic analysis for the novelty criterion based sparsification is done, and it is observed that the analysis is conserved under sparsification.
- Chapter 5 suggests a better sparsification technique for post-distortion over VLC channels called the fixed-budget KMSER approach. Detailed performance analysis

is carried out, and the derived analysis is validated by simulations. It is found that fixed-budget based sparsification yields lower dictionary sizes, and hence facilitates computational simplicity as compared to the novelty criterion based sparsification technique.

- Chapter 6 approaches the task of post-distortion over VLC channels via multi-stage clustering based blind equalization paradigm. In addition, this chapter analytically proves that the multi-stage clustering could be linked to the correntropy criterion (which incorporates higher order statistics), and hence this technique gives better MSE and BER performance as compared to existing blind post-distorters.
- Chapter 7 looks at closed loop multi-user MIMO-VLC with NOMA as the multiple-access technique. A novel hybrid precoding technique is proposed using Chebyshev polynomials. A novel power-allocation strategy is derived for the precoding technique. Using this power allocation, explicit expression for the BER is derived for square-QAM. Simulations are carried out to validate the theoretically derived expressions for BER for the proposed NOMA-VLC system.

Finally, Chapter 8 presents conclusion and direction for future works followed by bibliography and list of publications.

Chapter 2

Background

In this chapter, the area of visible light communications is introduced to the reader, and an overview of key signal processing techniques in VLC is provided. The organization of this chapter is as follows: a) an overview of VLC based systems is given in Section-2.1, b) components and applications VLC based systems are provided in Section-2.2, c) learning metrics to evaluate various signal processing algorithms are reviewed in Section-2.3, d) classical linear adaptive signal processing techniques are reviewed in Section-2.4, d) existing nonlinear adaptive filtering techniques are reviewed in Section-2.5, e) the potential of RKHS based techniques is explored in Section-2.6, f) the kernel least mean squares (KLMS) algorithm is reviewed in Section-2.7, and g) a recently proposed multiple access technique for VLC called NOMA is reviewed in Section-2.8.

2.1 Overview of visible light communication

VLC is a technique of communication in which transmission of data is achieved by modulating the LED at high rates without affecting the primary purpose of illumination [3]. The purpose of illumination is unaffected as the intensity of the LED is modulated at a rate above the flicker fusion threshold [3, 4]. LEDs are ubiquitous in our surroundings. If we look around, we have LED lamps for lighting our rooms/offices, traffic lights, headlights of cars, billboards etc. In future, LEDs will dominate the market for illumination purposes and will be widely prevalent due to their cost-effectiveness [3]. The integration of using

LED intensity modulation based VLC has the potential to complement RF-spectrum and converge towards development of smart cities [5]. There are several advantages offered by VLC over RF communications, which are enlisted:

- The 2.4 GHz industrial, scientific and medical (ISM) band is getting increasingly crowded as compared to the less occupied visible-light spectrum. Thus VLC encounters lesser interference and facilitates better link quality.
- As light does not penetrate walls through the VLC bands have better spectrum reusability and prevents eavesdropping/jamming from neighboring users.
- The front end devices of VLC do not suffer from phase noise and IQ imbalance as RF devices, and are also cheap and easily available.
- As wavelength in visible band is in the sub-micron range the variance of channel estimation algorithms are much lower in general. This also helps in accurate positioning of the receiver for better link performance.

Due to these advantages of VLC, it has been chosen as one of the dominant research-directions which has the potential to cater to the ever increasing bandwidth requirements of 5G and beyond systems in the near future [6].

2.2 Components and applications of visible light communications

Visible light communication refers to communication over visible light spectrum (400-700nm) [12] which is principally used for illumination. Moreover, this communication is “green” as compared to RF communication systems (due to use of light as a communication medium), uses spectrum which is much less cluttered as compared to the RF spectrum and uses the same infrastructure as compared to the RF systems. The overall downlink communication system consists of an intensity modulated LED, an LED driver and a power-line communication (PLC) modem/dongle. In certain scenarios considered in this

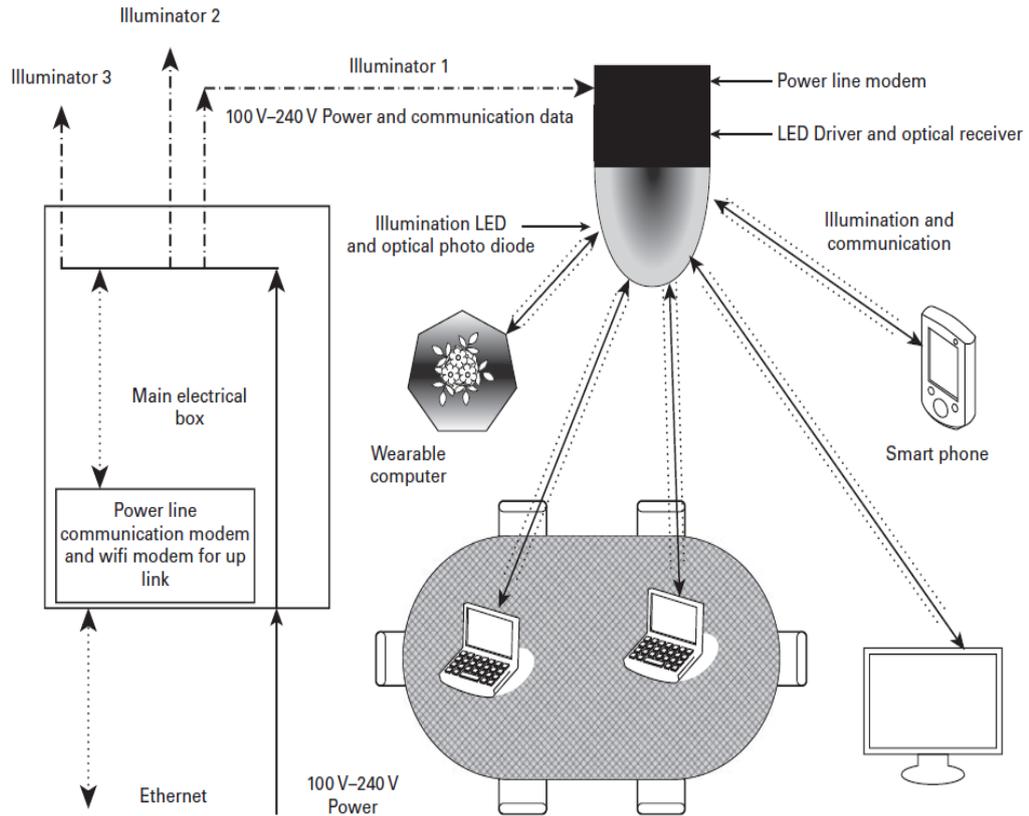


Figure 2.1: Example of VLC deployment.

work, when we assume a feedback uplink to transmitter to improve overall performance, the VLC-uplink consists of a photo-diode (used for converting light to an electrical signal), trans-impedance amplifier and a PLC modem. These are the components of a typical VLC communication system.

In future, there will be billions of bandwidth-hungry devices surrounding us as can be seen from the concept of IoT [13]. A typical VLC scenario is illustrated in Fig. 2.1 (adopted from [12]), in which one can see a plethora of devices connected to an LED which serves as a router. VLC would be an ideal technology that can act as an integral component of IoT as it is cheap, robust and can accommodate larger number of users as most of the visible spectrum is mostly unused/unoccupied as compared to the RF spectrum. This also has applications in inter-vehicular communications in smart cars by controlling traffic/communication between vehicles by using traffic lights, lights of cars etc. [12, 14, 15]. There are several other applications of VLC, including in medicine [16–18] where monitoring of patients in a hospital is done remotely by sensors and maintaining

records automatically. While RF spectrum is mostly occupied, this monitoring can be achieved by a VLC link that will not be as prone to interference and jamming.

Despite these desirable characteristics and applications of VLC links, the performance of VLC links are limited by the ISI and nonlinear characteristics of the LED. The following section provides information about these impairments in much greater detail.

2.3 VLC channel impairments

This section provides background on the impairments encountered in VLC links and emphasizes the need for signal processing techniques to mitigate these impairments. Also, performance metrics used in this work to assess performance of various signal processing algorithms (for mitigating these impairments) are explained. There are two major factors which impede the performance of a VLC link and prevent it from delivering the promised throughput for 5G systems: a) the LED characteristics, which presents saturation type nonlinearity causing distortion of the input signal, and b) the ISI/pulse spreading caused by different multi-path reflections from walls of the room. The following paragraphs elaborate on these performance-limiting impairments.

2.3.1 LED nonlinearity impairments

An LED's transfer characteristics are inherently nonlinear when one operates in the pulse-regime of LED or uses cheaply manufactured LEDs with lesser tolerance levels. In particular, white LED is modeled by saturation type nonlinearity in [19] called the Rapp model. Different values of parameters were derived in [19] to fit various nonlinearity types of white LEDs to the Rapp model which has been widely used previously in RF-communications to model power amplifiers. Mathematically, the Rapp model can be written as follows:

$$A(x) = \frac{x}{\left(1 + \left(\frac{x}{I_{sat}}\right)^{2p}\right)^{\frac{1}{2p}}} \quad , x > V_{th}$$
$$= 0 \quad , x < 0$$

where p is the parameter that defines the shape of the nonlinearity and I_{sat} indicates the LED saturation current. By invoking the well known Bussgang's theorem [20] it can be proven mathematically that the overall SNR falls considerably upon assuming a saturation nonlinearity to be acting on the input due to introduction of an uncorrelated additive component to the input signal. Techniques to mitigate LED nonlinearity are widely studied in the literature [7, 8, 21, 22], and it has been also widely established that additional signal processing techniques are required in order to maintain the BER below the forward error correction (FEC) limit.

2.3.2 Channel impairments

Early investigations on VLC assumed the channel to have a dominant line of sight component [23]. However, apart from the line of sight component, there exists multiple echoes/reflections of the transmission from the walls. In [24] these echoes were simulated using data-files of various LEDs by ray tracing technique and standardized channels were obtained. These channels were part of the IEEE 802.15 PAN VLC standard. From standardized channel models, significant delay spread causing severe ISI is observed. This is found to seriously hinder system throughput [24] without the use of advanced signal processing techniques.

A typical channel impulse response (CIR) is plotted in Fig. 2.2 for the "Open-Office" scenario and in Fig. 2.3 for the "office with cubicles" scenario. The LED nonlinearity characteristic is plotted in Fig. 2.4.

2.3.3 Performance measures

The following performance measures, to compare the proposed algorithm with existing algorithms, are used in this work:

MSE

MSE is defined as the variance of the deviation of an estimate from its desired value. MSE gives a measure of the reliability of an estimation technique. A low value of MSE for an

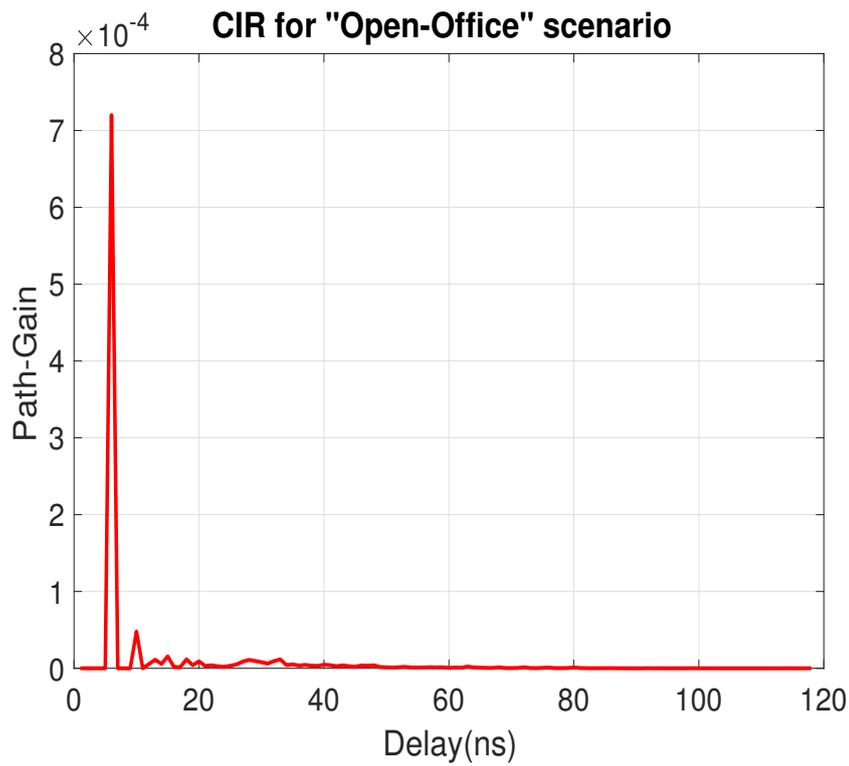


Figure 2.2: CIR of a VLC channel for “Open-Office” scenario.

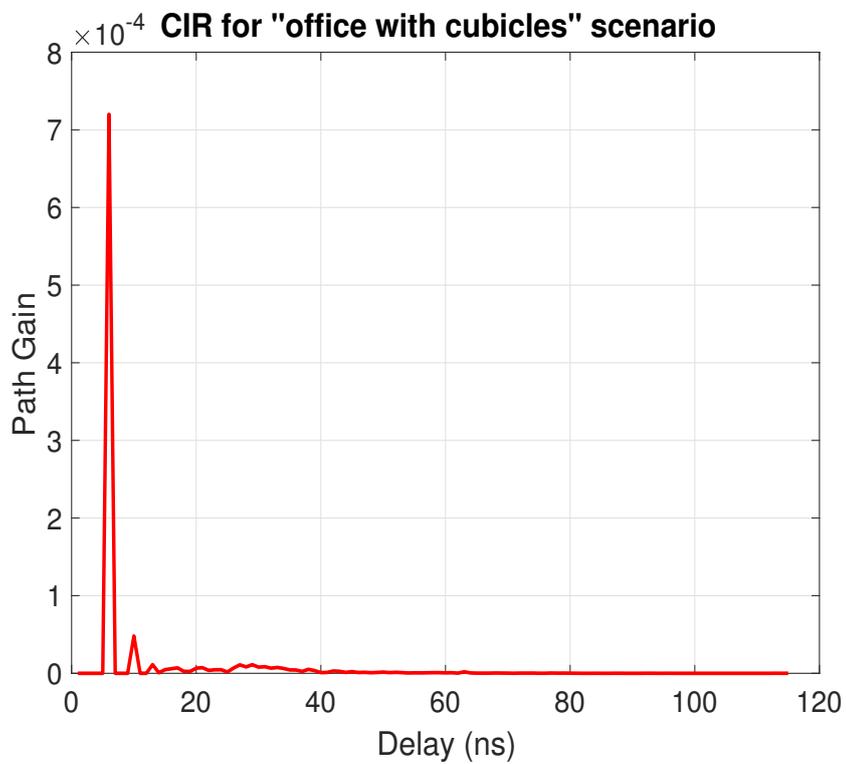


Figure 2.3: CIR of a VLC channel for “office with cubicles” scenario.

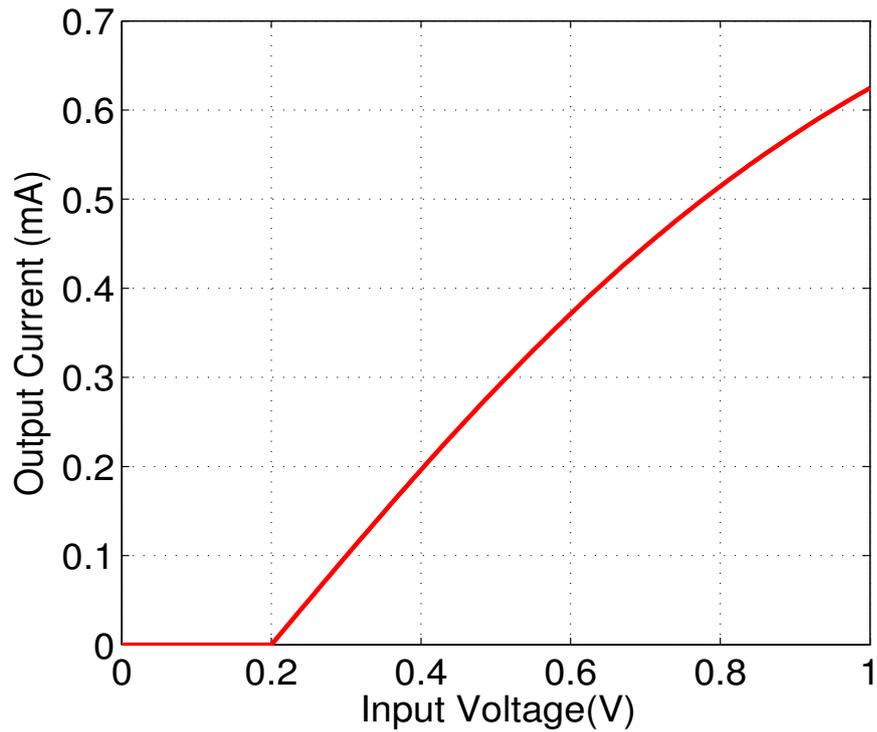


Figure 2.4: Nonlinear LED transfer characteristics.

estimator is always desired in order to ensure accurate parameter estimation. The MSE of an estimate value \hat{x} of a parameter x , is given by the following equation:

$$MSE(x) = \mathbb{E}[(x - \hat{x})^2] \quad (2.1)$$

where $\mathbb{E}[\cdot]$ denotes the statistical expectation operator. MSE gives a measure of variance for an unbiased estimator, and is a sufficient statistic in additive white Gaussian noise (AWGN) scenarios.

BER

In order to recover the transmitted bits in the presence of channel nonlinearity and ISI, another important optimization criterion for online learning is BER. The BER of the overall system is the probability of a received bit being flipped with respect to a transmitted bit

obtained via Monte-Carlo simulations. The BER is given by the following equation:

$$\text{BER} = \lim_{N_1 \rightarrow \infty} \frac{k_1}{N_1} \quad (2.2)$$

where k_1 is number of received bits that are in error, and N_1 is the total number of transmitted bits. BER is an important optimality criterion and is an independent cost function in its own right.

SER

Similar to BER, the symbol error rate (SER) is defined as the number of received symbols in error. Symbols are defined to be groups of q_1 -bits which are mapped to a 2^{q_1} -ary number. The SER can be defined as:

$$\text{SER} = \lim_{N_2 \rightarrow \infty} \frac{k_2}{N_2} \quad (2.3)$$

where k_2 is number of received symbols that are in error, and N_2 is the total number of transmitted symbols. Alternatively, SER can also be shown to be related to BER as follows:

$$\text{SER} = 1 - (1 - \text{BER})^{q_1} \quad (2.4)$$

Optimization of the minimum bit error rate (MBER)/MSER cost function in general yields faster convergence due to incorporation of order statistics and hence is a better learning criterion particularly over nonlinear channels.

2.4 Review of classical LMS based adaptive filtering

A classical linear adaptive filter [25], in its simplest form estimates a weighted sum of input \mathbf{x}_k , with the adaptive filter weights Ω_k to form the output $y_k = \Omega_k^T \mathbf{x}_k$, where $(\cdot)^T$ denotes the transpose operator. Next, the following cost function, J_{LMS} , is formulated

which is equivalent to the quadratic loss function:

$$J_{\text{LMS}} = \mathbb{E}[(s_{k-D} - y_k)^2] \quad (2.5)$$

where $\mathbb{E}[\cdot]$ denotes the statistical expectation operator, s_k denotes the transmitted symbols, and D is the overall lag of the channel and the equalizer. Since, it is difficult to evaluate the expectation, an instantaneous approximation of the cost function is assumed, and since J_{LMS} is convex in Ω_k , convergence to the global minima by a stochastic gradient learning algorithm is achieved, by adapting the weights recursively in the negative direction of the instantaneous gradient. The adaptation equation for the least mean squares (LMS) algorithm can hence be written as follows:

$$\Omega_{k+1} = \Omega_k + \eta e_k \mathbf{x}_k \quad (2.6)$$

where $e_k = (s_{k-D} - y_k)$ denotes the error term of the LMS algorithm, and η denotes the step-size.

A drawback of LMS algorithm is its sensitivity to the scaling of observations which could prevent its convergence. To counter this, the NLMS algorithm [25] scales the observations by the square of its Euclidean norm and hence prevents instability in convergence due to high variance of the observation vector. The NLMS algorithm uses a time varying step-size as follows:

$$\Omega_{k+1} = \Omega_k + \eta_k e_k \mathbf{x}_k \quad (2.7)$$

where $\eta_k = \frac{\eta}{\|\mathbf{x}_k\|_2^2}$.

The LMS algorithm and its variants have reasonable performance, are simple to implement (as their growth of computational complexity scales linearly with the dimensionality of regressors), and find widespread applicability in many areas like channel estimation, equalization, noise cancellation, and antenna beam-forming. Two examples of applications of the classical adaptive filtering in system identification and inverse channel mod-

eling for a given plant/system [26], are illustrated in Fig. 2.5, and Fig. 2.6 respectively.

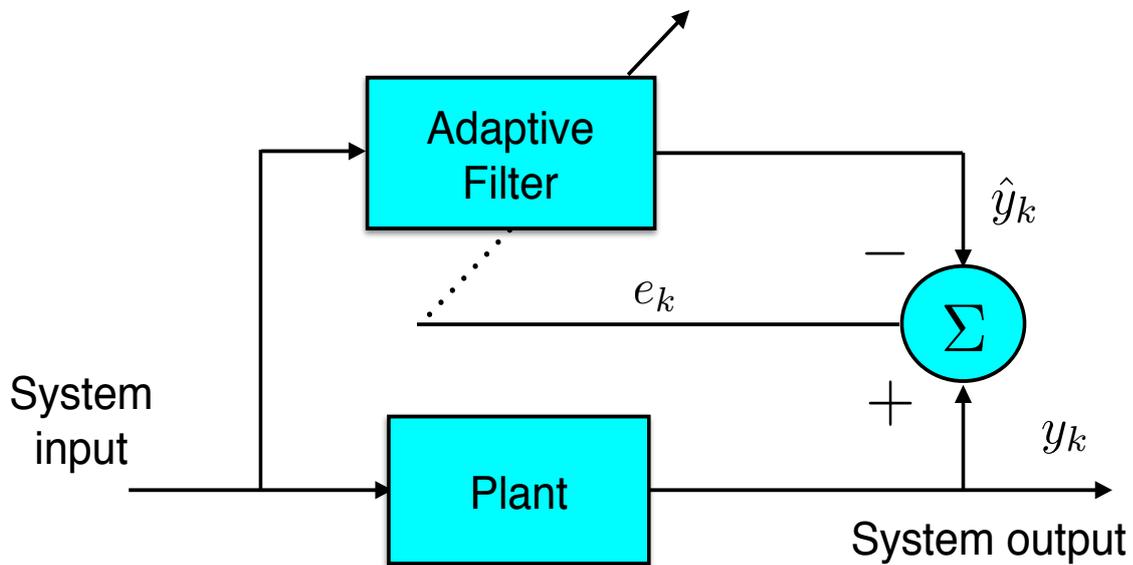


Figure 2.5: Application of adaptive filtering for system identification [26].

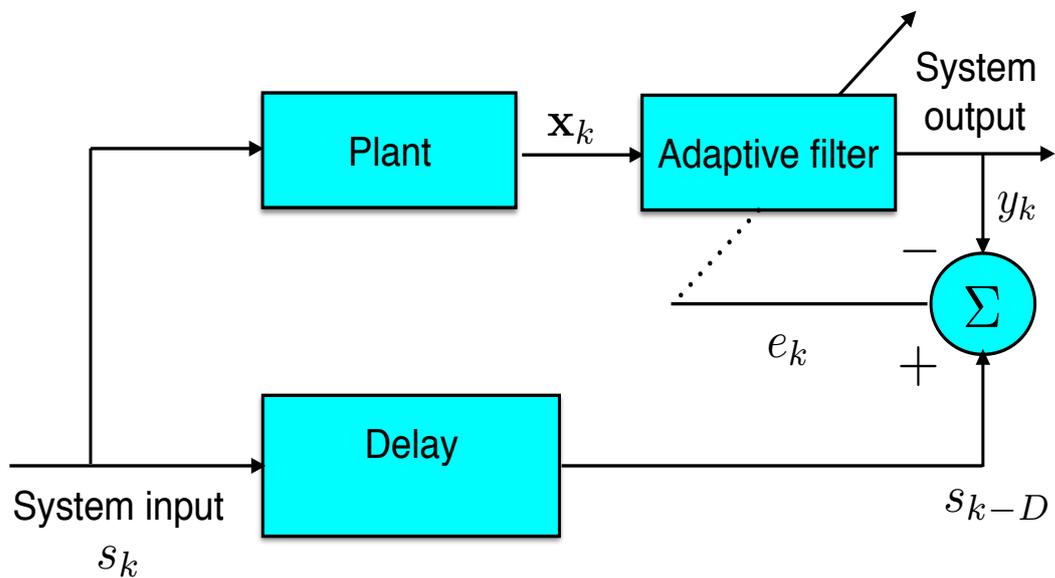


Figure 2.6: Application of adaptive filtering for system inversion [26].

However, the linear adaptive filtering techniques lose their validity, and are sub-optimal in scenarios when there is an inherently nonlinear system/plant, and therefore the observations are not linearly separable. In such scenarios, nonlinear adaptive filters need to be considered as candidate signal processing techniques.

2.5 Existing signal processing techniques for VLC

In VLC based systems, complex nonlinear signal processing techniques are required to achieve the FEC limit for BER of 3.8×10^{-3} due to the VLC channel impairments like ISI and LED nonlinearity. The existing signal processing methods for mitigating the impairments in VLC based systems can be classified into two categories: a) closed loop pre-distortion at the transmitter [27], and b) open-loop post-distortion at the receiver [21, 22, 28]. Linear closed loop pre-distortion uses a linear weight as a pre-distorter, which is learnt by NLMS algorithm using the detected symbols at the receiver as feedback to the transmitter. However, assumption of perfect feedback of labels is too idealistic an assumption, and hence open loop post-distortion has been explored in the literature. Among existing techniques for post-distortion, Volterra, and Hammerstein polynomial based adaptive filters have been explored lately, which rely on a truncated Taylor's series approximation of a function. However, their performance is severely affected by the modeling error incurred due to abrupt truncation of the Volterra series. As an example, let $g : \mathbb{R}^n \rightarrow \mathbb{R}$ be any differentiable function. Then, by the Taylor's series expansion, we arrive at the following approximation for $g(\mathbf{x})$ till second order terms, where $\mathbf{x} \in \mathbb{R}^n$:

$$g(\mathbf{x}) = g(\mathbf{x}_0) + \nabla g(\mathbf{x}_0)^T (\mathbf{x} - \mathbf{x}_0) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_0)^T \mathbf{J}(\mathbf{x}_0) (\mathbf{x} - \mathbf{x}_0) + \text{higher order terms} \quad (2.8)$$

where \mathbf{x}_0 is an arbitrary expansion point, $\nabla g(\mathbf{x}_0)$ denotes the gradient operator, and $\mathbf{J}(\mathbf{x}_0)$ denotes the Hessian matrix evaluated at the expansion point \mathbf{x}_0 . Assuming the second order expansion of $g(\mathbf{x})$, we have:

$$\hat{g}(\mathbf{x}) = g(\mathbf{x}_0) + \nabla g(\mathbf{x}_0)^T (\mathbf{x} - \mathbf{x}_0) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_0)^T \mathbf{J}(\mathbf{x}_0) (\mathbf{x} - \mathbf{x}_0) \quad (2.9)$$

It is observed from (2.9), that the final estimate of $g(\cdot)$, \hat{g} , is a linearly weighted combination of first order terms $[x_1, x_2 \dots x_n]$, and the second order terms $[x_1^2, x_1 x_2, x_1 x_3 \dots x_n^2]$. These terms are cumulatively called the second order Volterra kernel. Let the regressors

corresponding to the second order Volterra kernel be denoted as:

$$\phi_{\text{VOLTERRA}}(\mathbf{x}_k) = [x_1, x_2 \dots x_n, x_1^2, x_1 x_2, x_1 x_3 \dots x_n^2] \quad (2.10)$$

and the output as:

$$y_k = \Omega_k^T \phi_{\text{VOLTERRA}}(\mathbf{x}_k) \quad (2.11)$$

Existing generic Volterra filters (the Volterra filters can handle memory as well and hence are better than simple Taylor's series approximation which is presented here as an illustrative special case) optimize the following cost function given below:

$$J_{\text{VOLTERRA}} = \mathbb{E}[(g(\mathbf{x}_k) - y_k)^2] \quad (2.12)$$

where, the weights Ω_k are adapted by taking instantaneous gradient of J_{VOLTERRA} as follows:

$$\Omega_{k+1} = \Omega_k + \eta e_k \phi_{\text{VOLTERRA}}(\mathbf{x}_k) \quad (2.13)$$

where $e_k = g(\mathbf{x}_k) - y_k$. It should be noted that truncation till the second order Volterra kernel introduces modeling error in estimation of $g(\cdot)$. Further, the performance of polynomial expansion based approaches like Volterra filtering are susceptible to local minima, and is also sensitive to the choice of model order. Additionally, these approaches exhibit slow convergence, and have high computational requirement which prevent their deployment in practical VLC based systems.

Therefore, signal processing techniques need to be developed that have the following desirable properties: a) convexity, b) representation of a wide variety of functions without modeling errors, and c) which can be readily sparsified for computational tractability.

2.6 Potential of RKHS based adaptive filtering

In order to circumvent the computational complexity, and incurred modeling error by classical nonlinear signal processing techniques like Volterra filtering, online RKHS based learning techniques have been proposed in the literature which rely on the representer theorem in RKHS. In order to understand the theory behind RKHS, one can introduce the notion of evaluation function $\mathcal{F} : \mathcal{H} \rightarrow \mathbb{C}$ corresponding to a functional $f(\cdot)$, such that $\mathcal{F}_x(f) = f(x)$ (\mathcal{H} denotes the Hilbert space of functionals $f(\cdot)$). A Hilbert space is defined to be an RKHS if the evaluation functionals are bounded. Therefore, by Reisz representation theorem [29], there exists a $k_x \in \mathcal{H}$ such that:

$$\mathcal{F}_x(f) = f(x) = \langle k_x, x \rangle_{\mathcal{H}} \quad (2.14)$$

Since k_x is a function in RKHS in its own right, by reproducing property, one can write:

$$k_x(x) = \langle k_t, k_x \rangle_{\mathcal{H}} = K(t, x) \quad (2.15)$$

where $K(t, x)$ is the reproducing kernel defined on \mathcal{H} . It can be easily proven that all valid reproducing kernels are symmetric and positive semi-definite functions defined over $\mathbb{C}^n \times \mathbb{C}^n$.

An important result over RKHS is the representer theorem. Considering the loss function, $J(f)$, given below:

$$J(f) = l(f) + \lambda \|f\|_{\mathcal{H}}^2 \quad (2.16)$$

where $l(f)$ is a loss function of f and $\|\cdot\|_{\mathcal{H}}$ is the norm defined on RKHS \mathcal{H} . From the representer theorem,

$$\hat{f}(x) = \arg \min_f J(f) = \sum_{\forall i} \alpha_i K(x_i, x) \quad (2.17)$$

Thus, inference over a high (or even infinite) dimensional RKHS \mathcal{H} becomes equivalent

to minimizing over \mathbb{C}^n by use of kernels $K(\cdot, \cdot)$. Thus, the minimizer of a regularized non-linear loss-function can be written as a weighted sum of kernel functions which indicate inner-product in RKHS.

From (2.17), it has been found that the original RKHS based approaches have a temporally increasing computational and storage requirement, which is an undesirable feature of these techniques. Hence, sparsification of the learning techniques based on many criteria like novelty criterion, surprise criterion, fixed-budget criterion, have been proposed in the literature. These techniques selectively learn from the incoming observations using a dictionary consisting of “different” observations and error terms (the notion of “difference” being given by the respective sparsification criterion). Thus, these sparsification techniques curtail the temporally expanding requirement of the kernel adaptive filters without sacrificing performance. The RKHS based techniques can be viewed as an radial basis function (RBF) structure as shown in Fig. 2.7 [30], assuming a dictionary of observations, $\{\mathcal{D}_k^{(i)}\}_{i=1}^{|\mathcal{D}_k|}$, and a corresponding dictionary of error terms \mathcal{J}_k defined similarly. As opposed to classical post-distortion techniques like Volterra filtering, the RKHS based approaches are convex (and hence do not converge to local minima), allow for computational simplicity via sparsification, and have no modeling error (by the representer theorem), and hence is promising for post-distortion for standardized VLC channels.

2.7 Review of KLMS

In this section we provide a review of an online RKHS based adaptive filtering technique called KLMS [31]. This paradigm refers to online stochastic minimization of the quadratic loss function in (2.16).

By the representer theorem given in [32], unique representation for any arbitrary non-linearity exists in RKHS. Therefore, to recover the input symbols, the nonlinear equation as given in (6.1) is inverted by mapping a received vector of observation \mathbf{x}_k to a linearly separable high dimensional RKHS \mathcal{H}_σ corresponding to a single kernel width σ by a feature map $\phi(\cdot)$, where $\phi : \mathbb{C}^n \rightarrow \mathcal{H}_\sigma$ corresponding to the kernel width σ . Consequent

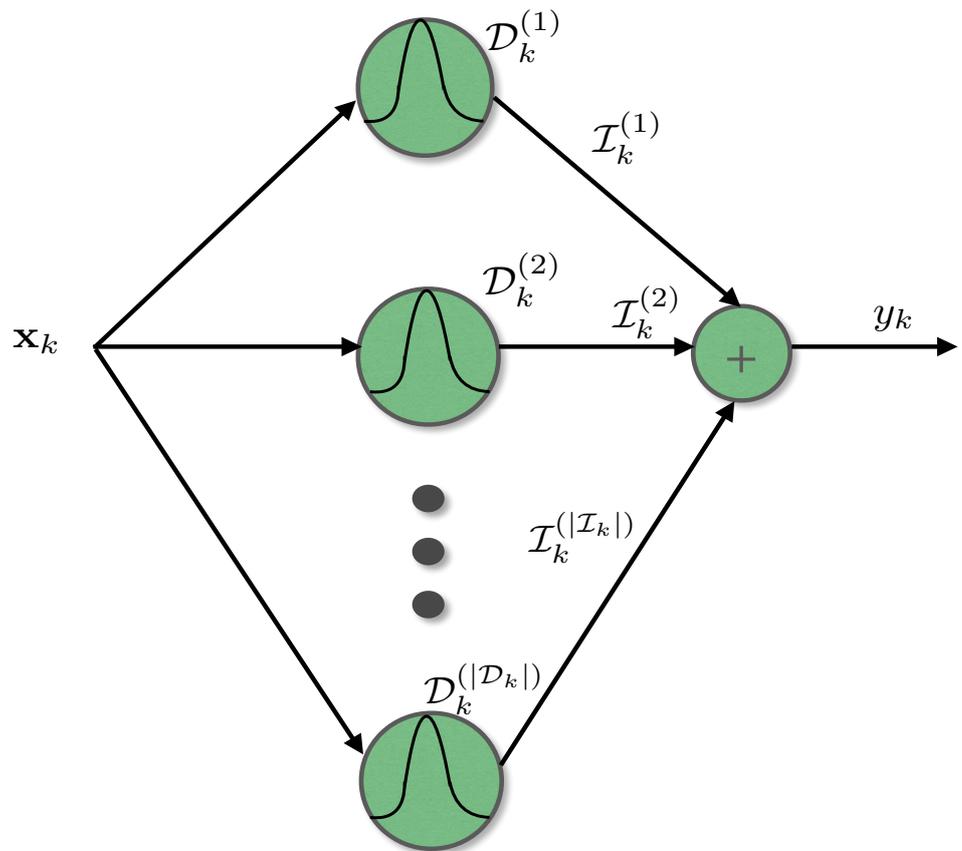


Figure 2.7: RKHS adaptive filter shown as an equivalent RBF network.

to this mapping, the following cost function is formed for the KLMS based post-distorter:

$$J_{\text{KLMS}}(k) = \mathbb{E}[(s_{k-D} - \Omega_k^T \phi(\mathbf{x}_k))^2] \quad (2.18)$$

where Ω_k is the post-distorter weight in RKHS, s_{k-D} is the delayed desired symbols (D being the overall lag of the channel and the equalizer), and $\mathbb{E}[\cdot]$ denotes the statistical expectation operator. The Ω_k is optimized using a stochastic gradient based approach. Let us denote $y_k = \Omega_k^T \phi(\mathbf{x}_k)$. Then the classical adaptation equation is given as follows:

$$\Omega_{k+1} = \Omega_k + \eta e_k \phi(\mathbf{x}_k) \quad (2.19)$$

where η is the step size, and $e_k = s_{k-D} - y_k$ denotes the error term for the KLMS algorithm. However, from (2.19), we can write as follows:

$$y_k = \eta \sum_{i=1}^{k-1} e_i \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_k) \rangle_{\mathcal{H}_\sigma} \quad (2.20)$$

where $\langle \cdot, \cdot \rangle_{\mathcal{H}_\sigma}$ denotes the inner product in RKHS. By the kernel trick, explicit knowledge of $\phi(\cdot)$ is not essential to evaluate this inner product. Instead it can be written as:

$$\begin{aligned} \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_k) \rangle_{\mathcal{H}_\sigma} &= \kappa_{\sigma, \mathbb{C}^n}(\mathbf{x}_i, \mathbf{x}_k) = \\ &= \exp\left(-\frac{\sum_{\forall q} (\mathbf{x}_i(q) - \mathbf{x}_k(q)^*)^2}{\sigma^2}\right) \end{aligned} \quad (2.21)$$

where $\mathbf{x}^{(q)}$ denotes the q^{th} component of the vector, \mathbf{x} , and $(\cdot)^*$ denotes complex conjugation.

2.8 Multiple-access for VLC

In classical multiple access techniques, users share an orthogonal set of time-frequency resources, which are allocated to them; but their overall throughput is less than the upper bound for the achievable capacity. However, with the increase in demand for spectrum for upcoming 5G based deployments, it is possible to overlap many users on the same

time-frequency resource by superposition coding. The same time-frequency resource is shared by multiple users with different power-levels allocated to them (which is called power-domain multiplexing or classical NOMA) [33, 34]. At each user equipment, there is an ordered successive interference canceller (SIC) which recovers the users' signals in the order decided by the signals' allocated power level. As an example, consider two users UE1 and UE2 with channel gains $h^{(1)}$, and $h^{(2)}$, symbols $s^{(1)}$ and $s^{(2)}$, and the corresponding channel outputs $y^{(1)}$, and $y^{(2)}$. We have the following equation relating these parameters at the k^{th} time instant assuming a quasi-static single-input single-output (SISO) channel model [33]:

$$\begin{aligned} y_k^{(1)} &= h^{(1)}(\sqrt{P^{(1)}}s_k^{(1)} + \sqrt{P^{(2)}}s_k^{(2)}) + n_k^{(1)} \\ y_k^{(2)} &= h^{(2)}(\sqrt{P^{(1)}}s_k^{(1)} + \sqrt{P^{(2)}}s_k^{(2)}) + n_k^{(2)} \end{aligned} \quad (2.22)$$

where $n_k^{(1)}$ and $n_k^{(2)}$ are the noise terms at each user equipment (UE) with variance σ_n^2 . If $h^{(1)} > h^{(2)}$, $P^{(1)} < P^{(2)}$. Hence, after one-tap equalization, $s_k^{(2)}$ is recovered from $y_k^{(2)}$ at UE2 with the following signal to interference and noise ratio (SINR) (denoted by $\Gamma^{(u)}$ for the u^{th} user):

$$\Gamma^{(2)} = \frac{|h^{(2)}|^2 P^{(2)}}{|h^{(2)}|^2 P^{(1)} + \sigma_n^2} \quad (2.23)$$

At UE1, however, detection proceeds in two stages of SIC. First, the $s_k^{(2)}$ is detected from $y_k^{(1)}$ treating the contribution of $s_k^{(2)}$ to $y_k^{(1)}$ as noise. Next, the contribution of the detected $s_k^{(2)}$ (denoted by $\hat{s}_k^{(2)}$) is subtracted from $y_k^{(1)}$, and is detected with the following SINR:

$$\Gamma^{(1)} = \frac{P^{(1)}|h^{(1)}|^2}{\sigma_n^2} \quad (2.24)$$

Although initially proposed for RF based systems, the NOMA has been recently found to be useful in VLC. A typical NOMA setup for a VLC based system is given in Fig. 2.8.

Classical NOMA assumes diverse channel conditions at each UE and hence facilitates

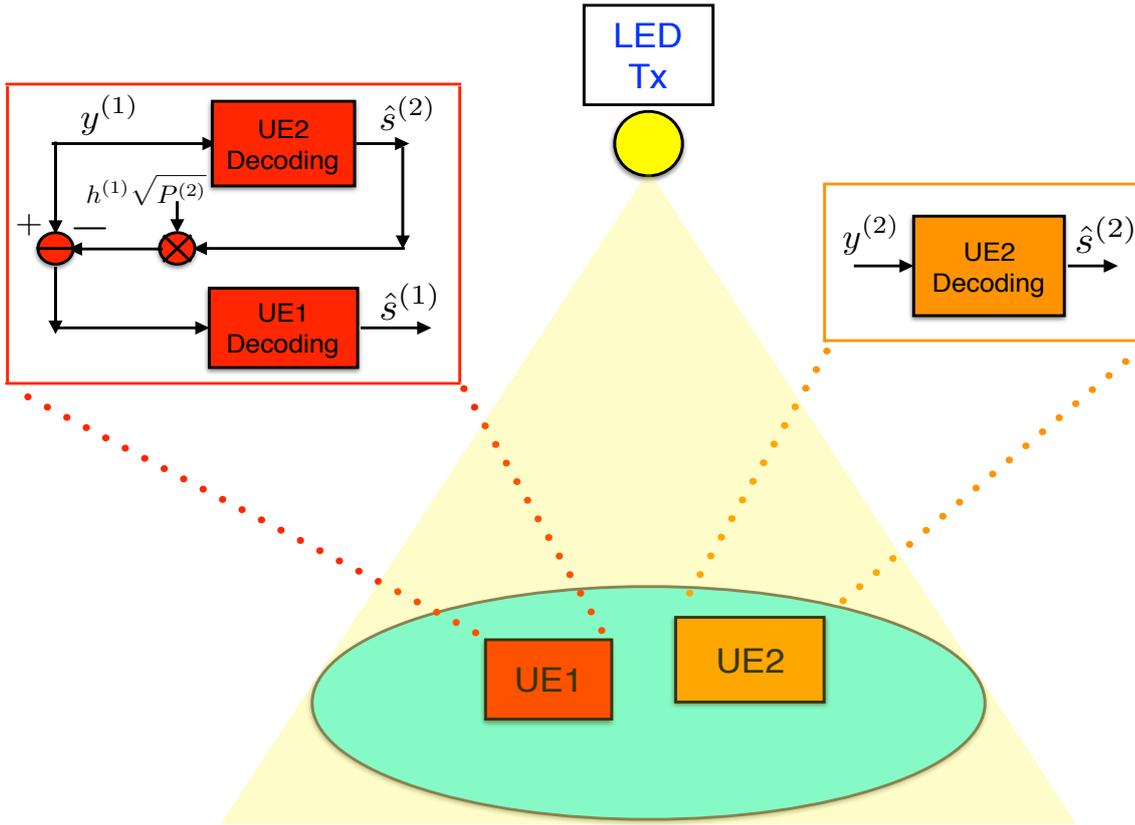


Figure 2.8: A typical NOMA-VLC setup.

diverse power allocation coefficients for each user. However, in practice, all users in a typical attocell in a VLC network may not experience similar channels, but have differing data-rate requirements [35]. Existing precoding techniques do not address this scenario in general, and this problem is addressed only for the two-user scenario [35]. Hence, precoding and power-allocation techniques need to be developed that extends the work in [35] for a multiple user scenario, when many users experience similar/correlated channels, which can happen in practical VLC scenarios when some users are located close to each other.

2.9 Conclusion

In this chapter, a brief overview of VLC based systems was provided to aid the understanding of the contents of the thesis. Next, the major impairments for VLC based communication techniques, and existing signal processing techniques to mitigate those im-

pairments were reviewed. Next, the potential of RKHS based learning was discussed. Finally, multiple-access in VLC was taken up. An upcoming multiple-access technique called NOMA was discussed, and potential challenges in VLC based systems for NOMA were raised.

Chapter 3

Chebyshev Polynomial based Adaptive Pre-distorter for VLC

3.1 Introduction

In VLC, the device-nonlinearity of an LED degrades the overall performance of a communication link [36]. To improve the overall link-performance in such scenarios, several pre-distortion schemes have been suggested to counter the nonlinear characteristics of LED. The simplest pre-distortion scheme is to use a lookup table (LUT) which stores the input-output pairs for the nonlinearity of the LED and pre-distorts the transmitted signal by assigning the nearest member in the LUT. This type of receiver is simple to implement; however the LED characteristics are prone to change due to factors like temperature variation and aging [27]. To mitigate this, an adaptive NLMS based pre-distorter that learns a scaling factor for pre-distortion, and tracks changes in LED characteristics was proposed in [27].

However, in order to counter-effect LED nonlinear characteristics, one must “learn” a nonlinear transformation on the input as opposed to a linear hypothesis proposed in [27] in order to incorporate order statistics. The reason is that the inverse of a nonlinearity should be a nonlinearity in general. Therefore, a Chebyshev polynomial expansion of the input is suggested in this chapter for pre-distorter design using a familiar stochastic-gradient

based approach.

There has been some ongoing research on joint ISI and LED nonlinearity mitigation using techniques like Volterra adaptive filters [22, 28] and Hammerstein filters [21]. However, the Volterra based adaptive filters are computationally complex. The indoor ISI channel of a room is mostly static which can be obtained from ray-tracing techniques [37]. Hence, at the receiver, the data received after photo-diode (which is assumed to be linear) can be post-processed by incorporating the pseudo-inverse of the convolutional matrix corresponding to the ISI channel in the feedback loop of [27] as a zero-forcing solution to mitigate ISI. Using the post-processed data the system model in [27] can be used to mitigate the LED distortion (using Chebyshev polynomial regression) so that the transmitted symbols can be recovered accurately. To evaluate robustness of the proposed approach, comparisons have been made against the NLMS based pre-distortion approach and post-distortion based approaches like Volterra [28] and Hammerstein filters [21] for a variety of modulation schemes. There have also been works that use an adaptive mean square criterion [38] for equalization which is suboptimal for nonlinear scenarios.

This chapter is organized as follows: Section-3.2 reviews NLMS based pre-distortion, the proposed approach is given in Section-3.3, Section-3.4 describes the simulation results and Section-3.5 summarizes this chapter.

3.2 NLMS based pre-distortion

NLMS based pre-distortion has been one of the seminal techniques suggested for mitigating the effect of LED nonlinearity [27]. In NLMS based pre-distortion, the pre-distorter is chosen as a linear scaling factor (considering the Bussgang decomposition of the LED nonlinearity), and this pre-distorter is adaptively learnt by the NLMS algorithm with the detected symbols at the receiver being relayed to the transmitter in order to form the error signal.

In the considered system model, let x_k denote the sequence of input symbols indexed

over time k . The $A(\cdot)$ in [39] is modeled by the following “quadratic” equation:

$$A(x) = b_0 + b_1(x - 0.5) + b_2(x - 0.5)^2 \quad (3.1)$$

b_0 , b_1 and b_2 are the polynomial coefficients which are parameterized by $b_0 = \zeta$, $b_1 = 1$, $b_2 = -4\zeta + 2$, where ζ is a parameter to control the severity of nonlinearity. Another kind of nonlinearity considered in the simulations is taken from [36], which is called the “Rapp” nonlinearity and is widely used to model power-amplifiers in RF communications.

The work in [27] assumes a scaling factor r_k , which is multiplied with the input symbol sequence x_k . Consequently, it passes through the LED nonlinearity $A(\cdot)$, and an independently identically distributed (i.i.d) AWGN n_k is added after passing through the channel so as to form an estimated signal \hat{l}_k .

Several online stochastic gradient based online adaptive filtering techniques exist in the literature [40] for parameter estimation (the parameter in this chapter would indicate the inverse of the LED nonlinearity). Chief among such techniques is the LMS algorithm which optimizes the stochastic approximation of the squared error, and is widely known for its robustness and simple implementation. However, it suffers from the problem of gradient noise amplification upon encountering observations towards the higher end of its dynamic range. To counter this phenomenon, the observations are divided by their norm and then used for stochastic-gradient adaptation. This adaptive technique prevents high values of the observations from amplifying the gradient noise. This adaptive algorithm is called the NLMS algorithm. Mathematically, it has also been motivated as optimization of the MSE cost function under assumption of a temporal smoothness constraint in [41], and also as a rank one approximation of the Newton-update as given in [40].

Due to the above desirable characteristics of the NLMS algorithms, in [27] the linear pre-distorter r_k is updated using the following NLMS-based algorithm:

$$r_{k+1} = r_k + \eta_k e_k x_k \quad (3.2)$$

where, $\eta_k = \frac{\eta}{\sum_{\forall k} |x_k|^2}$ is the step-size for the NLMS algorithm [40] at the k^{th} iteration given

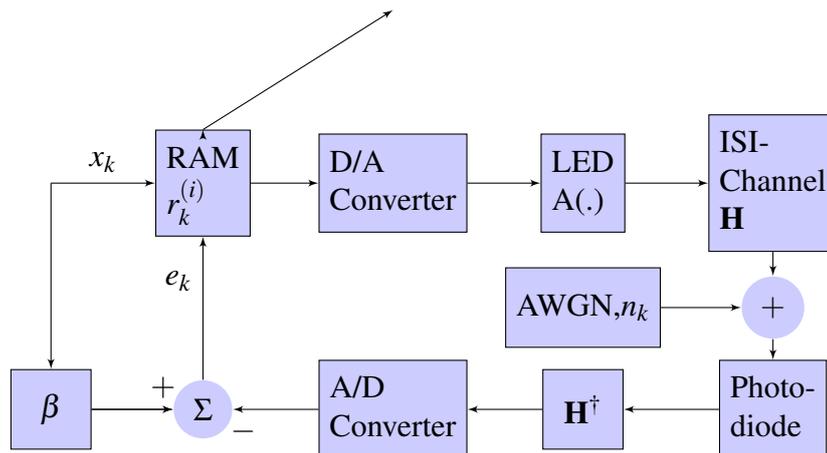


Figure 3.1: Block diagram of the proposed system model.

in [40, 42] (η being a small positive number), $e_k = \beta x_k - \hat{l}_k$ (β being a biasing constant), where \hat{l}_k is the output given by $\hat{l}_k = A(r_k x_k) + n_k$. β is a constant which is the gain term of LED.

3.3 Chebyshev polynomial-expansion based pre-distortion in frequency-flat channels

Orthogonal polynomials [43] are defined to be a set of orthogonal basis functions with respect to an inner product. In other words, given a closed subset of the real line $[a, b]$, two orthogonal polynomials $f_m(x)$ and $f_n(x)$ must satisfy the following condition,

$$\int_a^b f_m(x) f_n(x) dx = 0 \quad (3.3)$$

Orthogonal polynomials find numerous applications in approximation theory according to which an arbitrary function in a Sobolev space could be expressed as a linear combination of orthogonal basis polynomials. There exists many kinds of orthogonal polynomials (e.g. Hermite, Laguerre) among which the Chebyshev polynomials are widely preferable due to its min-max error approximation property in the interval $[-1, 1]$. They are given by the

following recurrence relation:

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \quad (3.4)$$

where T_n denotes the Chebyshev polynomial of first kind of order n . These functions can also be viewed as solutions to the following differential equation:

$$(1 - x^2) \frac{d^2y}{dx^2} - x \frac{dy}{dx} + \alpha^2 y = 0 \quad (3.5)$$

where α is an arbitrary constant. Due to min-max approximation error optimality of the Chebyshev expansion, this Chapter explores Chebyshev polynomial based pre-distortion. The basic blocks for Chebyshev pre-distortion are described below and illustrated in Fig. 3.1. The noise process is modeled by AWGN which comes after passing through the channel \mathbf{H} (considered according to the system model in [44], where \mathbf{H} denotes the convolutional matrix corresponding to the CIR). Please note that there could be a possibility of an amplifier after the photodiode in Fig. 3.1. However, as this is a closed loop control system scaling factors are compensated by the closed loop adaptive pre-distorter system. Hence, this block can be omitted without loss of generality. To correct an LED nonlinearity $A(\cdot)$, it is desirable that the characteristics of the pre-distorter be a nonlinear function of the input x_k , as opposed to $r_k x_k$ (discussed in previous the section). A polynomial pre-distorter is proposed which would make $\hat{l}_k = A(\sum_{\forall i} r_k^{(i)} T_i(x_k)) + \mathbf{H}^\dagger n_k$, where $T_i(x)$ is a Chebyshev polynomial of i^{th} order, and $r_k^{(i)}$ is the i^{th} pre-distorter weight at time instant k . This is a nonlinear transformation on input signals which is written as a sum of orthonormal basis polynomials in interval $[-1, 1]$. This expansion is better suited for approximating the inverse of the restriction of the LED nonlinearity in the interval $[-1, 1]$ as compared to approximating by just a scaling factor in [27], since the inverse of the nonlinear characteristic is better modeled by a nonlinear function.

Chebyshev polynomials have been found in the control literature for nonlinear system identification and nonlinear parameter learning in general [45]. Chebyshev polynomials have desirable properties like: a) orthogonality, b) minimizing the min-max error from

the desired polynomial, and c) having decaying coefficients (eigenvalues for orthonormal polynomials) for smooth functions; hence it can ascertain the optimal polynomial order to avoid overfitting. Hence, the Chebyshev polynomials are good choice for approximating smooth functions within the closed interval $[-1, 1]$. Although the nonlinearity may not be contained in $[-1, 1]$ one can consider its restriction over the closed interval by proper biasing of LED such that the signal and its nonlinear transformation are within $[-1, 1]$. This is implementation-wise practical as in a real-life scenario one deals with signals with a finite dynamic range which can be scaled to the interval $[-1, 1]$.

To perform Chebyshev NLMS based pre-distortion, the cost-function to be minimized, $J = \min_{r_k^{(i)}} \mathbb{E}[(\beta x_k - \hat{l}_k)]^2$, is the mean squared deviation from the detected signal \hat{l}_k and transmitted symbol x_k . The coefficients $r_k^{(i)}$ are updated by taking derivative of J with respect to $r_k^{(i)}$ (which is practically implemented as a Random Access Memory (RAM)) via a similar stochastic gradient NLMS approach as:

$$r_{k+1}^{(i)} = r_k^{(i)} + \eta_k e_k T_i(x_k) \quad (3.6)$$

The first four Chebyshev polynomials which was used in this chapter are given as $T_0(x) = 1, T_1(x) = x, T_2(x) = 2x^2 - 1, T_3(x) = 4x^3 - 3x$.

Please note that for coefficient update and comparison both the algorithms in [27] and the proposed approach uses the NLMS algorithm, however with different inputs. The former approach passes the inputs directly to the NLMS filter while the latter uses orthogonal polynomials. Hence, the covariance matrix of the data on the latter case will be the identity matrix, which will culminate in a larger Eigen-value spread of the data and hence faster convergence is achieved [41].

3.4 Simulations

In this section, simulations are presented to validate the proposed adaptive pre-distortion approach against existing pre-distortion/post-distortion techniques for VLC. First, the proposed algorithm is compared to the existing pre-distortion algorithms given in [27]

as shown in the Fig. 3.2. The proposed algorithms are compared against: a) no pre-distortion, b) lookup table with $\zeta = 0.582$, c) lookup table with $\zeta = 0.582 + 0.025$, d) NLMS algorithm in [27] with $\zeta = 0.582$, e) NLMS algorithm in [27] with $\zeta_1 = 0.582 + 0.025$, f) proposed polynomial expansion based algorithm with $\zeta = 0.582$, and g) proposed polynomial expansion based algorithm with $\zeta_1 = 0.582 + 0.025$. It is observed that the proposed algorithm has better performance as compared to alternative approaches compared in [27] for various values of ζ, ζ_1 . A gain of 1 decade of SER at an SNR of 14dB is found for the proposed algorithm as compared to the NLMS based approach in [27] for $\zeta = 0.582, \zeta_1 = 0.582 + 0.025$ for 16-QAM. Thus it can be inferred from simulations that the polynomial expansion based approach outperforms NLMS based approach in [27] as a pre-distortion technique for compensating the nonlinear characteristics of the LED.

Similar advantages are observed in the case of LED characteristic given in [36] for the proposed algorithm vis-a-vis the same approaches described in the literature. SER gain of a decade is shown in Fig. 3.3. The modulation scheme in this figure was quadrature phase shift keying (QPSK).

To further validate the robustness of the proposed algorithm, ISI VLC channels are considered. The CIR for an indoor channel \mathbf{h} can be estimated easily as given in [37, 46–48] by ray-tracing techniques. Assuming that the impulse response is known, consider the data block (in time) to be multiplied by the convolutional matrix \mathbf{H} corresponding to \mathbf{h} , where \mathbf{H} is full-rank. If the data-block is post-processed in time by the pseudo-inverse of \mathbf{H} , the channel reduces to a frequency-flat channel and the block diagram in Fig. 3.1 will be applicable. This post-processing is necessary as ISI introduces memory in the system and makes the transmitted symbols temporally dependent. Thus, upon post-processing by a left-inverse of \mathbf{H} , the temporal whiteness of symbols is restored and hence the Chebyshev polynomial expansion may be used to learn the LED nonlinearity in the feedback loop. The CIR considered in this scenario has 169 taps as Configuration C of [49]. In Figs. 3.4 and 3.5, better SER improvements is observed for the proposed Chebyshev polynomial expansion based nonlinearity mitigation approach as compared with NLMS based nonlinearity mitigation approach in the ISI scenario in [49] using the

system model in Fig. 3.1.

In Fig. 3.6 the performance of 4-pulse amplitude modulation (PAM) for the proposed approach, NLMS based approach and Volterra adaptive filtering in [28] are compared. The same 169 taps ISI channel as [49] was used as in previous simulations so as to model an indoor channel. Same temporal post-coding, by the pseudo-inverse of the convolutional matrix, was used in case of the proposed approach and NLMS based approaches. For Volterra equalizer, the post processing was performed at the receiver using the LMS algorithm [28]. It is observed in Fig. 3.6 that by using the proposed approach, better performance is achieved as compared to NLMS based adaptive pre-distorter and Volterra based adaptive post-distorter (at high SNR). Please note that the system model in Fig. 3.1 is used only for Chebyshev regression and NLMS pre-distorter; and the respective system models for other algorithms are used as given in their corresponding seminal works. In other words, for the Chebyshev and Hammerstein filters, the system models as given in [21, 28] are used. The purpose of this simulation is to compare how the proposed pre-distortion technique compare against existing post-processing techniques which are also valid ways of optimizing the bit error rate. In Fig. 3.6 the general Hammerstein based approach in [21] is compared with the proposed approach using the recursive least squares (RLS) based approach proposed in their work. It is found that the proposed algorithm delivers better SER performance as compared to these approaches.

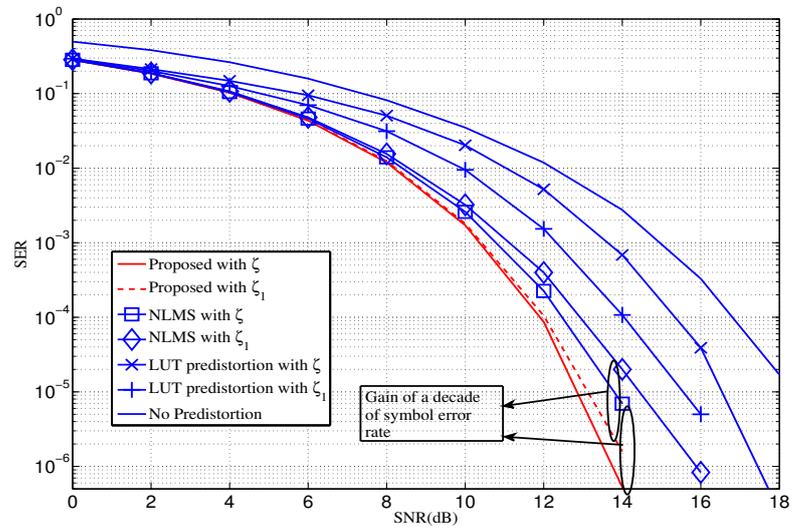


Figure 3.2: Performance comparison of the proposed approach with standard approaches in quadratic nonlinearity.

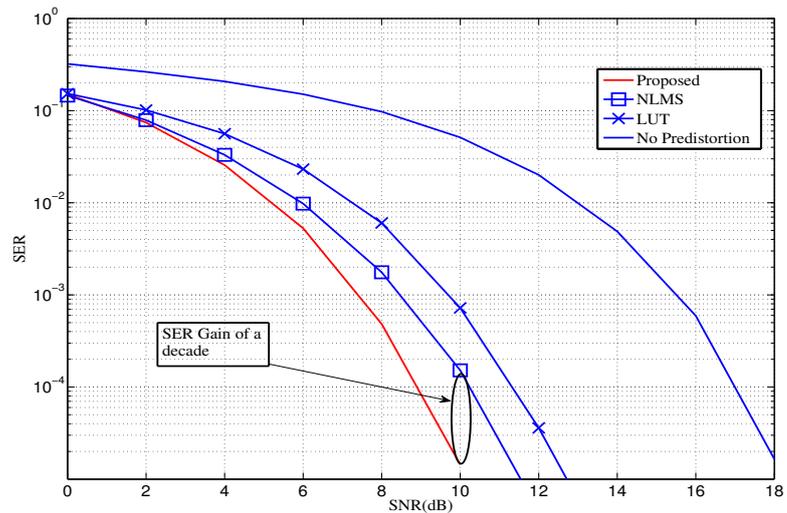


Figure 3.3: Performance comparison of the proposed approach with standard approaches in Rapp nonlinearity.

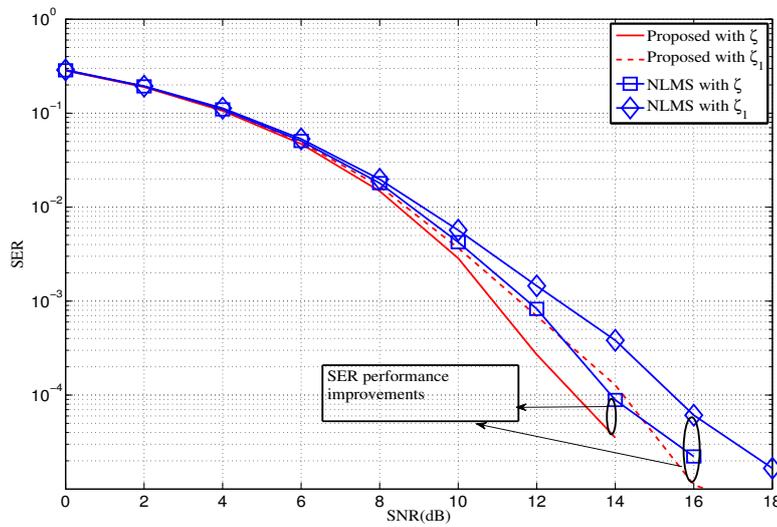


Figure 3.4: Performance comparison of the proposed approach with standard approaches in quadratic nonlinearity in precoded ISI environment for 16-QAM.

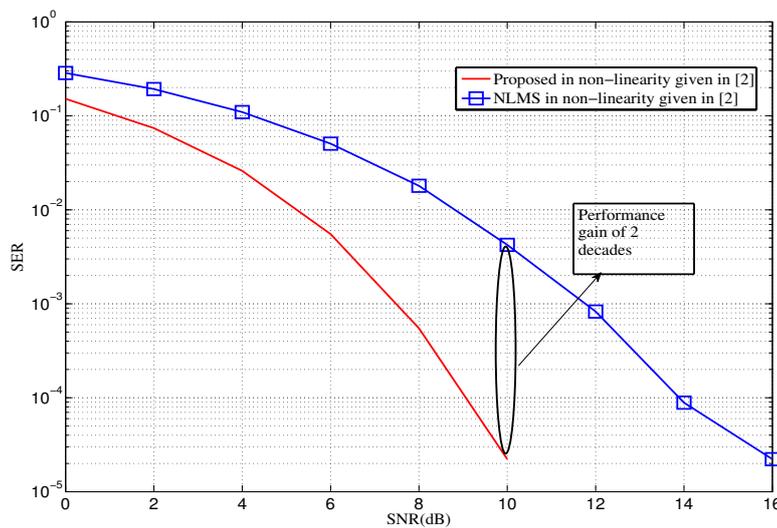


Figure 3.5: Performance comparison of the proposed approach with standard approaches in Rapp nonlinearity in precoded ISI environment for QPSK.

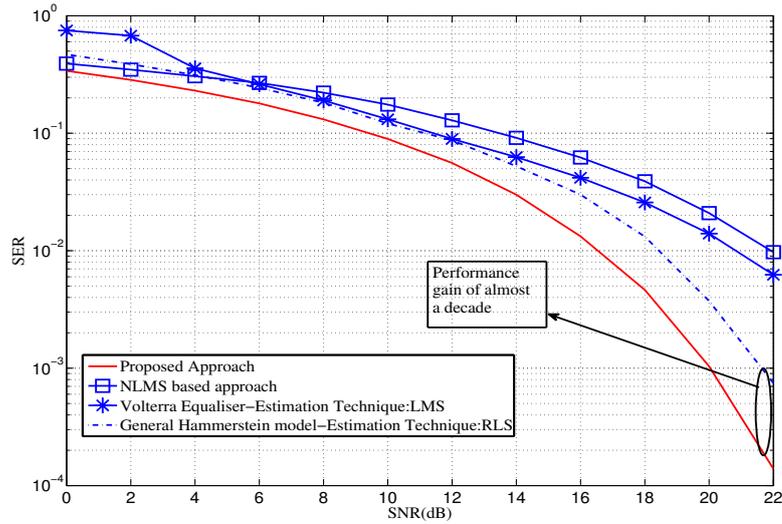


Figure 3.6: Performance comparison of the proposed approach with standard approaches in Rapp nonlinearity in precoded ISI environment for 4-PAM.

3.5 Summary

In this chapter, a typical closed-loop VLC system is considered and is found to be impaired by LED nonlinearity and ISI. To mitigate these impairments, a technique for transmit-side mitigation of LED-nonlinearity (also called pre-distortion) is contributed. The nonlinear adaptive pre-distortion algorithm is more effective in correcting LED nonlinearities in a VLC setting as compared to other existing linear pre-distortion algorithms. Instead of the scaling factor-based pre-distortion proposed in the literature, this chapter suggests the use of orthogonal polynomials to mitigate the nonlinear LED characteristics. In this chapter, normalized Chebyshev regression based pre-distortion is suggested to learn the pre-distorter weights due to min-max approximation properties of Chebyshev polynomials in the interval $[-1,1]$. The polynomial coefficients of the Chebyshev pre-distorter were derived by a stochastic gradient based NLMS algorithm which was found to be more suited for correcting nonlinearity as compared to a linear adaptive pre-distorter as given in [27]. Additionally, the system model was extended to model ISI scenarios by incorporation of pseudo-inverse of the circulant channel matrix over the closed loop system model proposed in this chapter. Simulations were carried over frequency-flat/ISI scenarios, and the proposed Chebyshev polynomial based pre-distortion algorithm was

compared with a) Volterra post-distortion, b) Hammerstein post-distortion, and c) simple NLMS pre-distortion [27]. As observed from simulations, considerable SER performance improvement is achieved using the proposed Chebyshev polynomial based pre-distortion algorithm as compared to existing linear pre-distortion based approaches.

Chapter 4

Adaptive Dictionary based Minimum Symbol Error Rate Post-distorter in VLC

4.1 Introduction

VLC communication systems promise bandwidths of 4-5 orders of magnitude higher than 4G systems, and 2-3 orders of magnitude higher than modern optical fiber systems (assuming there is no electrical to optical conversion issue) with speeds upto 3-Gb/s being reported [50]. However, their performance is limited by impairments like LED nonlinearity and ISI due to the propagation channel, as highlighted in Chapter 1. LED nonlinearity particularly affects the performance of VLC systems when multilevel constellations like PAM and QAM with high peak to average power ratio (PAPR) are used. One can argue that LED always has a linear working region, and lifetime of a modern LED is 10 years due to which the characteristics may remain more or less invariant (which can be resolved by manual calibration). However, the strength of post-distortion is apparent in the following scenarios: (a) high power/pulse mode of LED, in which nonlinearity is higher and the life time of LED is shorter, b) robustness on the tolerance of manufacturing, since different production bins give difference grades of LED, and post-processing technique

would be useful to save the cost in manufacturing, and maintaining SNR under dynamic environment in presence of interference, power instability, etc.

For a VLC system, the CIR of an indoor environment is typically long which requires complex receiver architecture based on Volterra-decision feedback equalizer (DFE) [28]. Many VLC systems have been demonstrated successfully in the literature which mitigate these impairments by pre-distortion techniques, and post-distortion techniques or an (iterative) combination of both. Post-distortion, and pre-distortion (as discussed in previous chapter) techniques generally go hand-in-hand [22] and are both valid solutions to mitigate the overall VLC channel impairments. Post-distortion, as an independent task, aims to mitigate the overall nonlinearity of the cascade of LED nonlinearity and ISI channel by a long computationally demanding solution like Volterra-DFE [7, 8, 28]. This formulation, though attractive, is based on heuristics like choice of optimal filter order and may suffer from convergence to local minima if the filter parameters are not chosen properly. Moreover, the overall throughput can be affected if the Volterra-DFE takes a large number of iterations to converge.

In parallel, recently there has been a growing interest in RKHS based techniques [31, 51, 52] for equalization and channel estimation of nonlinear systems. These classes of algorithms are generally adaptive and find convex solutions to nonlinear optimization problems [30]. Several classes of linear adaptive algorithms like LMS algorithm, and RLS algorithm have been absorbed within the framework of RKHS techniques with reasonable computational complexity. Recently, the MSER based equalization [53, 54] (a better paradigm than MMSE paradigm as it incorporates higher order statistics) has also been adopted into the RKHS framework by the name the kernel minimum symbol error rate equalizer [55]. Both KLMS [31] and KMSE based approaches [55], give good BER performance; however they require infinite storage and polynomial computational complexity, thereby calling for dictionary sparsification techniques [56].

In order to lower the computational complexity and make the KMSE viable for VLC channels, a sparsification technique is proposed in Section 4.4. From the literature survey, the use of RKHS techniques has not been explored for equalization of VLC

channels. Simulations reveal that the proposed sparsified-KMSER has equivalent/better performance in most scenarios as compared to Volterra-DFE, with lower computations required as compared to the Volterra-DFE. As an additional novelty, theoretical expressions dictating MSE dynamics for KMSER are derived and validated by simulations. This analysis provides control over the desired MSE floor, and the convergence rate of KMSER by varying the step-size in a manner dictated by the mathematically derived expressions.

The following terminology is used in this chapter: scalar at time k is represented by $(\cdot)_k$ and vector of past M samples at time k is represented by boldface with subscript k such as \mathbf{x}_k (which are elements of \mathbb{C}^M , M being the row dimension of the vector, and \mathbb{C} denotes the field of complex numbers) and italicised variables like \mathcal{D}_k and \mathcal{I}_k denote an online dictionary of observation and error terms respectively, at the time instance k . Real part of a complex quantity is denoted by the superscript $(\cdot)^R$ and $\Re\{\cdot\}$, and the imaginary part is denoted by $(\cdot)^I$ and $\Im\{\cdot\}$.

This chapter is organized as follows: Section-4.2 provides the system model, Section-4.3 reviews KMSER, Section-4.4 proposes the sparsification of KMSER for VLC channels via novelty criterion, Section-4.5 derives the equations that dictate the MSE dynamics of KMSER, Section-4.6 provides simulations to validate the proposed approach against the Volterra DFE, and Section-4.7 summarizes this chapter.

4.2 System Model

In this section, the system model considered in this chapter is described. Let s_k denote the input constellation (with DC bias so as to be placed in the forward-bias of LED) at the k^{th} time instant. The transmit symbol can be complex valued, where the real, and imaginary parts of the transmitted symbols would be modulated by orthogonal pulse-shapes at the baseband prior to transmission. It is passed through an finite impulse response (FIR) filter $\{h_i\}_{i=0}^{L-1}$, of order $L - 1$. The received symbol at k^{th} time instant, x_k , is given as follows:

$$x_k = \sum_{i=0}^{L-1} h_i A(s_{k-i}) + n_k \quad (4.1)$$

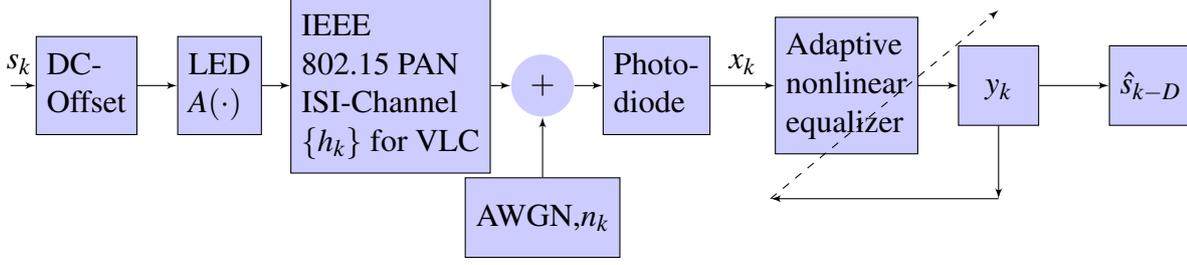


Figure 4.1: Block diagram of the used system model.

where $A(\cdot)$ is a Rapp LED nonlinearity [19, 57] which had been used to model a white LED, and n_k is i.i.d AWGN with variance σ_n^2 . The block diagram of the system model is given in Fig. 4.1. The channel consists of a linear FIR filter $\mathbf{h} = [h_0, h_1, \dots, h_{L-1}]$ (which models the VLC indoor channel from standard channel models [58, 59] obtained by techniques given in [24] and surveyed in [60]), and a Rapp nonlinearity $A(\cdot)$ which models the LED characteristics [23, 61]. The equalizer considered in this chapter is a nonlinear adaptive equalizer with observation \mathbf{x}_k as input. \hat{s}_{k-D} is the equalizer estimate delayed by D samples, where D is the cumulative lag of the equalizer.

4.3 Review of KMSER based technique for nonlinear channel equalization

In this section, a recent adaptive algorithm for nonlinear channel equalization based on KMSER cost function in RKHS [55] is reviewed.

In the KMSER channel equalizer, a nonlinear implicit feature map, $\Phi(\mathbf{x}_k)$, is applied on the observation at k^{th} instant (denoted by \mathbf{x}_k). KMSER is the application of the kernel trick to the normalized-adaptive MBER (NAMBER) recursion [54] which is written as follows for nonlinear channel equalization,

$$\Omega_k = \Omega_{k-1} - \eta I_k \frac{\Phi(\mathbf{x}_k)^*}{\langle \Phi(\mathbf{x}_k), \Phi(\mathbf{x}_k) \rangle_{\mathcal{H}} + \tau} \quad (4.2)$$

where $\Phi(\cdot)$ is an implicit feature map from $\mathbb{C}^M \rightarrow \mathcal{H}$, $*$ denotes complex conjugation, η is the step-size, τ is a small positive number, $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ denotes the inner product in RKHS

\mathcal{H} , and Ω_k is the implicit parameter weight in RKHS at time k . $I_k = \tanh(\beta(y_k^R - s_{k-D}^R + 1)) + \tanh(\beta(y_k^R - s_{k-D}^R - 1)) + j(\tanh(\beta(y_k^I - s_{k-D}^I + 1)) + \tanh(\beta(y_k^I - s_{k-D}^I - 1)))$ where β is a constant high enough to approximate the signum function. Thus the output at k^{th} instant y_k can be written as

$$y_k = -\eta \sum_{i=1}^{k-1} I_i \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_k) \rangle_{\mathcal{H}} \quad (4.3)$$

By applying the kernel trick, the adaptation at the k^{th} instant would then become [31],

$$y_k = -\eta \sum_{i=1}^{k-1} I_i \kappa_{\gamma, \mathbb{C}^d}(\mathbf{x}_i, \mathbf{x}_k) \quad (4.4)$$

where,

$$\kappa_{\gamma, \mathbb{C}^d}(\mathbf{x}_i, \mathbf{x}_{k+1}) = \exp \left(- \sum_{\forall q} (\mathbf{x}_i(q) - \mathbf{x}_{k+1}(q)^*)^2 \gamma \right) \quad (4.5)$$

where $\mathbf{x}(q)$ denotes the q^{th} entry of the vector \mathbf{x} , I_i is the corresponding error term at i^{th} instant which is defined as a soft function of the MSER constraint and γ is the kernel parameter. This is the basic KMSE algorithm as introduced in [55]. This algorithm was validated over some benchmark channels in the literature and was found to outperform other RKHS based equalizers, thus being a strong candidate to be implemented over practical scenarios like VLC channel equalization. Please note that while using a Gaussian kernel, the explicit knowledge of $\Phi(\cdot)$ is not required as the overall adaptation is expressed as the inner product in (4.5). From [32] it is known that if γ is chosen properly, any arbitrary nonlinearity can be modeled by this inner product in RKHS by Representation theorem [32]. Another advantage of these approaches is that knowledge of the nonlinearity, that is to be compensated for, is not required to be known a-priori for processing at the receiver. Additionally, the receiver is not constrained by the structure of nonlinearity, as assumed by Volterra-DFE (viz. maximum order of Volterra series) and the KMSE based approach also leads to a convex optimization problem in RKHS [55]. Despite these desirable properties, the computational complexity of KMSE grows with time which is its major drawback and makes the system unrealizable in practical scenarios. In the

next section, a modification to KMSER equalizer is proposed via *novelty criterion* [56] so as to further curtail its computational complexity and make it suitable for practical VLC channel equalization.

4.4 Sparsified-KMSER using novelty criterion based post-distortion for VLC channels

The computational complexity of Volterra-DFE (which is a popular post-distortion algorithm for VLC systems) increases with an increase in memory of the first order and second order terms. Also, truncation till second-order coefficients in the Volterra-series representation introduces modeling error (as the representation for the nonlinearity is being approximated by a second order Taylor's series [62]) in the symbol estimates. As opposed to this, the RKHS based techniques have ability to represent any nonlinearity exactly (due to closedness of Hilbert spaces and the Representation theorem [32]). Due to this, the modeling error is significantly reduced by using RKHS based techniques. However, from (4.3), it is observed that the number of computations required for KMSER (or any online RKHS technique like KLMS) grows undesirably as $O(k)$. With the huge influx of observations (data) for high data-rate 5G systems, an infinitely growing dictionary (i.e a set of observations based on which the equalizer is estimated in RKHS) cannot be maintained. Hence, in order to make the RKHS based techniques practical, a selectively growing dictionary based on pruning/sparsification is proposed.

Here, an RKHS based sparsified post-distorter is proposed, which: a) does not suffer from the approximation error (due to Taylor's series truncation) and the computational complexity of the Volterra-DFE, and b) does not require infinite storage and computational requirements as required for KLMS and KMSER. Therefore, a dictionary sparsification technique based on novelty criterion as has been suggested in [30] for KLMS is proposed for KMSER criterion for VLC channels. The notion of a dictionary $\mathcal{D}_k = \{\mathcal{D}_k^{(i)}\}_{i=1}^{|\mathcal{D}_k|}$ denotes a collection of observations indexed by (i) at the k^{th} instant which is maintained and updated in accordance with the incoming observations. The output, y_k , is calcu-

lated according to the dictionaries \mathcal{D}_k (a collection of significantly different observations which represent the data) and \mathcal{I}_k (a collection of corresponding I_k which have been selectively added to the dictionary indexed by (i)). If the incoming observation \mathbf{x}_k satisfies $\min_i \|\mathcal{D}_k^{(i)} - \mathbf{x}_k\| \leq \tau_1$ (i.e. a similar entry to the current observation exists in the dictionary), then the new observation is rejected and the dictionary remains unchanged. If not, then if $|I_{k+1}| > \tau_2$ (i.e. if the MSER constraint is violated significantly), the \mathbf{x}_k is added to $\mathcal{D}_k^{(i)}$, otherwise the \mathbf{x}_k is rejected. This algorithm is described in detail in Algorithm 1.

From the analysis provided in Section 4.5.5, it is proven asymptotically that, the MSE

Algorithm 1 Sparsified-KMSER

- 1: Initialize constants τ_1 and τ_2 . Dictionary $\mathcal{D}_1 = \{\mathbf{x}_1\}$, $\mathcal{I}_1 = I_1$, η , kernel-width γ .
 - 2: **while** $k \leq 10000$ (maximum number of iterations as given in [28]) **do**
 - 3: $y_k = -\eta \sum_{i=1}^{|\mathcal{D}_k^{(i)}|} \mathcal{I}_k^{(i)} \kappa_{\gamma, \mathbb{C}^d}(\mathcal{D}_k^{(i)}, \mathbf{x}_k)$
 - 4: $I_k = \tanh(\beta(y_k^R - s_{k-D}^R + 1)) + \tanh(\beta(y_k^R - s_{k-D}^R - 1)) + j(\tanh(\beta(y_k^I - s_{k-D}^I + 1)) + \tanh(\beta(y_k^I - s_{k-D}^I - 1)))$
 - 5: **if** $\min_i \|\mathcal{D}_k^{(i)} - \mathbf{x}_k\| \geq \tau_1$ and $|I_k| > \tau_2$ **then**
 - 6: $\mathcal{D}_{k+1} := \mathcal{D}_k \cup (\mathbf{x}_k)$
 - 7: $\mathcal{I}_{k+1} := \mathcal{I}_k \cup (I_k)$
 - 8: **end if**
 - 9: **end while**
-

behavior of the proposed sparsified-KMSER is similar to that of KMSER if $\tau_1 = 0.1 \sqrt{\frac{1}{2\gamma}}$ and $\tau_2 = \sigma_e$ are chosen by rules given in [30], with σ_e^2 denoting the targeted steady-state MSE floor, while providing a practical and low complexity adaptive solution. Thus, the proposed sparsified KMSER based post-distorter is robust to nonlinearity, requires far less storage and computations, thereby making it a practical solution for VLC communication.

4.5 Theoretical analysis of KMSER

In this section, various properties of KMSER are demonstrated by theoretically analyzing its MSE dynamics. This analysis is necessary, as step-size can be varied (within the range specified by the following MSE analysis), so as to achieve a specified MSE floor within a number of iterations.

First, the theoretically derived transient dynamics of the KMSER is provided. Next,

the KLMS is analyzed in the derived framework (of analysis), so that the proposed approach can be compared with KLMS. Consequently, a step-size range for convergence of KMSE is provided which helps us in the choice of step-size. Then the steady-state MSE of KMSE is theoretically compared with that of the KLMS, and it is proven that KMSE has lower excess misadjustment as compared to KLMS for a given step-size. Next, a step-size range is derived, within which the KMSE converges faster as compared to KLMS. Finally, it is shown that the analysis for transient and steady state MSE of KMSE asymptotically holds under the assumption of an online sparsified dictionary and closely matches the MSE curves obtained from simulations.

4.5.1 MSE transient dynamics for KMSE

In this section, from analysis, the MSE behavior of the KMSE is predicted. This analysis is necessary to theoretically predict the converged MSE floor and iterations required for convergence as the step-size η is varied.

Let Ω_k denote the implicit parameter in RKHS which is being estimated using KMSE criterion. Let $\tilde{\Omega}_k$ be the deviation of Ω_k from the optimal parameter Ω^o , such that $s_{k-D} = \langle \Omega^o, \Phi(\mathbf{x}_k) \rangle_{\mathcal{H}} + n_k$ (existence of Ω^o stems from the Representation theorem of Mercer kernels). Also, let the a-priori deviation of the implicit parameters in RKHS be denoted as $\tilde{\Omega}_k^a = \Omega_k - \Omega^o$ and a-posteriori deviation be denoted as $\tilde{\Omega}_k^p = \Omega_{k+1} - \Omega^o$. Then from [63], $\tilde{\Omega}_k^p$ and $\tilde{\Omega}_k^a$ are related by the following equation,

$$\tilde{\Omega}_k^p = \tilde{\Omega}_k^a - \eta I_k \langle \Phi(\mathbf{x}_k), \cdot \rangle_{\mathcal{H}} \quad (4.6)$$

Let, at the k^{th} instant, $\tilde{y}_k = \langle \tilde{\Omega}_k^a, \Phi(\mathbf{x}_k) \rangle_{\mathcal{H}}$ and $\tilde{y}_{k+1} = \langle \tilde{\Omega}_k^p, \Phi(\mathbf{x}_k) \rangle_{\mathcal{H}}$. The $\tilde{y}_k = y_k - y^o$ is the stochastic deviation from the fixed point over expectation. As the trace of the kernel Gram is “1” for the Gaussian kernel, the following equation is inferred,

$$\tilde{y}_{k+1} = \tilde{y}_k - \eta I_k \quad (4.7)$$

$$\implies \mathbb{E}[|\tilde{y}_{k+1}|^2] = \mathbb{E}[|\tilde{y}_k|^2] + \mathbb{E}[\eta^2 |I_k|^2 - 2\eta \Re\{I_k^* \tilde{y}_k\}] \quad (4.8)$$

as,

$$\mathbb{E}[I_k^* \tilde{y}_k + \tilde{y}_k^* I_k] = \mathbb{E}[2\Re\{I_k^* \tilde{y}_k\}] \quad (4.9)$$

Since, $I_k = \tanh(\beta(y_k^R - s_{k-D}^R + 1)) + \tanh(\beta(y_k^R - s_{k-D}^R - 1)) + j(\tanh(\beta(y_k^I - s_{k-D}^I + 1)) + \tanh(\beta(y_k^I - s_{k-D}^I - 1)))$, the deviation of the signal point s_{k-D} from the signal point y_k can be written as $y_k - s_{k-D} = \tilde{y}_k + n_k$. The Taylor series approximation for I_k around “1” upto third order exponents for $\beta = 1$ (a high enough value for approximating a signum function [54]) is found to be,

$$I_k^R \approx -0.29(\tilde{y}_k^R + n_k^R)^3 + 0.80(\tilde{y}_k^R + n_k^R)^2 + 0.33(\tilde{y}_k^R + n_k^R) + 0.11 \quad (4.10)$$

This Taylor series approximation is valid since a smooth adaptation is assumed. Assuming small \tilde{y}_k and n_k , $\mathbb{E}[\tilde{y}_k] \rightarrow 0$ (unbiasedness), neglecting terms of 3rd order assuming Gaussianity of \tilde{y}_k and n_k in RKHS (therefore odd moments are zero) and assuming \tilde{y}_k and n_k to be independent.

$$\mathbb{E}[I_k^R \tilde{y}_k^R] \approx \mathbb{E}[0.33(\tilde{y}_k^R + n_k^R) + 0.116] \tilde{y}_k^R \approx 0.33 \mathbb{E}[|\tilde{y}_k^R|^2]$$

$$|\hat{I}_k^R|^2 \approx 0.29(\tilde{y}_k^R + n_k^R)^2 + 0.077(\tilde{y}_k^R + n_k^R) + 0.0135 \approx 0.29(\tilde{y}_k^R + n_k^R)^2 + 0.077(\tilde{y}_k^R + n_k^R) \quad (4.11)$$

$$\implies \mathbb{E}[|\hat{I}_k^R|^2] \approx 0.29[\mathbb{E}[\tilde{y}_k^R{}^2] + \sigma_n^2]$$

Without loss of generality, assuming similar results for imaginary part, (4.8) can be written as,

$$\mathbb{E}[|\tilde{y}_{k+1}|^2] \approx (1 - 0.66\eta + 0.29\eta^2)\mathbb{E}[|\tilde{y}_k|^2] + \delta_1 \quad (4.12)$$

where $\delta_1 = 0.29\eta^2\sigma_n^2$.

For validation and comparison of the proposed approach against KLMS, the KLMS algorithm can also be analyzed in the similar way as the proposed KMSER algorithm. For KLMS, the adaptation equation is as given in [31], in which instead of I_k , the er-

ror term, $e_k = y_k - s_{k-D} = \tilde{y}_k + n_k$ (which follows from adaptive filter theory as $e_k = \langle \tilde{\mathbf{Q}}_k, \Phi(\mathbf{x}_k) \rangle + n_k$),

$$\tilde{y}_{k+1} = \tilde{y}_k - \eta(\tilde{y}_k + n_k)\kappa_{\gamma, \mathbb{C}^d}(\mathbf{x}_k, \mathbf{x}_k) \quad (4.13)$$

Therefore, after squaring both sides and then taking expectation, the adaptation in (4.13) can be written approximately as,

$$\mathbb{E}[|\tilde{y}_{k+1}|^2] = \mathbb{E}[|\tilde{y}_k|^2](1 - 2\eta + \eta^2) + \delta_2 \quad (4.14)$$

where $\delta_2 = \eta^2 \sigma_n^2$.

Thus, having derived a unified framework for analysis of MSE-dynamics for KMSER in (4.12) and KLMS in (4.14), several desirable features of the KMSER as compared to KLMS, are highlighted in the following sections.

4.5.2 Step-size range for convergence

In this section, bounds are derived for step-size for which KMSER converges. The adaptation equation for KMSER (4.8) converges iff,

$$0 < 1 - 0.66\eta + 0.29\eta^2 < 1 \implies 0 < \eta < 2.27 \quad (4.15)$$

Comparing with KLMS, which converges iff $0 < \eta < \frac{2}{\text{Trace}[G_\phi]} \implies 0 < \eta < 2$ (as given in [30] and can be found by putting $(1 - 2\eta + \eta^2) < 1$ from (4.14)), where G_ϕ is the kernel Gram matrix. Therefore, it is observed that the proposed algorithm converges over a wider step-size range as compared to KLMS, which is a desirable property of the proposed algorithm.

4.5.3 Excess mean squared error analysis

In this section, the excess mean squared error (EMSE) for KMSER is derived and then compared with KLMS. In (4.12), assuming convergence for KMSER, $\mathbb{E}[|\tilde{y}_{k+1}|^2] \approx \mathbb{E}[|\tilde{y}_k|^2]$

$$\mathbb{E}[|\tilde{y}_{k+1}|^2] = \frac{0.29\eta^2\sigma_n^2}{0.66\eta - 0.29\eta^2} \quad (4.16)$$

At low enough step-size, for the KMSER, the steady state excess misadjustment is given by the following equation,

$$\mathbb{E}[|\tilde{y}_\infty|^2] \approx 0.4394\eta\sigma_n^2 \quad (4.17)$$

The approximation in (4.17) is under the assumption of a small step-size.

Using the same assumption for KLMS, $\mathbb{E}[|\tilde{y}_{k+1}|^2] \approx \mathbb{E}[|\tilde{y}_k|^2]$ in (4.14), implies the steady state excess misadjustment for KLMS would be,

$$\mathbb{E}[|\tilde{y}_\infty|^2] = \frac{\eta\sigma_n^2}{2-\eta} \approx 0.5\eta\sigma_n^2 \quad (4.18)$$

The approximation in (4.18) is also under assumption of small step-size. This expression is same as derived in [31]. This shows that the methodology for the analysis used in this chapter justifies the results in literature [31]. Also, from (4.17) and (4.18), it is observed that EMSE of KMSER is lower than that of KLMS which is a desirable property of the proposed algorithm.

4.5.4 Convergence analysis of KMSER

In this section, the convergence rate of the KMSER is analyzed and compared with convergence rate of KLMS algorithm. Let there be an arbitrary adaptation equation as given below,

$$t_{k+1} = \alpha t_k + \delta \quad (4.19)$$

Then, at a particular instant k , the value of t_k is given by,

$$t_k = \delta \frac{1 - \alpha^{k-1}}{1 - \alpha} \quad (4.20)$$

The steady state value, t^o , of t_k is

$$t^o = \frac{\delta}{1 - \alpha} \quad (4.21)$$

assuming $0 < |\alpha| < 1$ for convergence [64]. Without loss of generality, let us model t_k as $\mathbb{E}[|\tilde{y}_k|^2]$. Let us now model the proposed KMSER algorithm by this recursion. From (4.12), $\alpha_1 = 1 - 0.66\eta + 0.29\eta^2$ for the proposed algorithm. By the same methodology for KLMS, from (4.14), $\alpha_2 = 1 - 2\eta + \eta^2$. Hence, for the proposed algorithm to converge faster than KLMS, the following condition from (4.20) is required, assuming convergence to the same steady state value,

$$1 - \alpha_1^{k-1} < 1 - \alpha_2^{k-1} \quad (4.22)$$

Substituting the above values of α_1 and α_2 , the following condition is found,

$$0 < \eta < 1.88 \quad (4.23)$$

Hence, under these ranges of step-sizes for KMSER, faster convergence is achieved in case of KMSER as compared to KLMS algorithm.

4.5.5 Does the analysis hold under sparsification of dictionary?

In line with the above analysis provided above for KMSER with no sparsification, a proof is now provided that the dynamical equations derived in this chapter for transient dynamics are valid upon sparsification of the dictionary \mathcal{D}_k . In other words, it is proven that the sparsification does not affect the results derived in this chapter (which are based on dynamical equation (4.12)) if the τ_1 and τ_2 are chosen according to $\tau_1 = 0.1\sqrt{\frac{1}{2\gamma}}$ and $\tau_2 = \sigma_e$ as given in [30]. It can be assumed that the set of input observations $\{\mathbf{x}_k\}$ belongs to a compact subset of \mathbb{C}^M (as it is a closed and bounded subset of \mathbb{C}^M). By Cover's theorem,

a dictionary exists $\mathcal{D}_k = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_P\}$ with P centers whose union of disjoint Euclidean neighborhoods cover the compact input domain [30]. Again, if the a-posteriori and a-priori deviation Ω_k^p and Ω_k^a is calculated, two scenarios arise as considered in Algorithm 1:

Scenario 1: \mathbf{x}_k is added at the k^{th} instant to dictionary

In this scenario, $\mathcal{D}_k (= \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_P\}) \neq \mathcal{D}_{k+1} (= \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_P, \mathbf{x}_k\})$

$$\Omega_k^a = -\eta \sum_{i=1}^P \mathcal{J}^{(i)} \langle \Phi(\mathbf{c}_i), \cdot \rangle_{\mathcal{H}} \quad (4.24)$$

$$\Omega_k^p = -\eta \sum_{i=1}^P \mathcal{J}^{(i)} \langle \Phi(\mathbf{c}_i), \cdot \rangle_{\mathcal{H}} - \eta I_k \langle \Phi(\mathbf{x}_k), \cdot \rangle_{\mathcal{H}} \quad (4.25)$$

Taking inner product on both sides (4.24) and (4.25) by $\Phi(\mathbf{x}_k)$ in RKHS \mathcal{H} (similar to (4.7)),

$$\tilde{y}_{k+1} = \tilde{y}_k - \eta I_k \langle \Phi(\mathbf{x}_k), \Phi(\mathbf{x}_k) \rangle_{\mathcal{H}} \quad (4.26)$$

where, $\tilde{y}_k = \langle \tilde{\Omega}_k^a, \Phi(\mathbf{x}_k) \rangle_{\mathcal{H}}$ and $\tilde{y}_{k+1} = \langle \tilde{\Omega}_k^p, \Phi(\mathbf{x}_k) \rangle_{\mathcal{H}}$. For the Gaussian kernel $\langle \Phi(\mathbf{x}_k), \Phi(\mathbf{x}_k) \rangle_{\mathcal{H}} = 1$. Therefore, \tilde{y}_{k+1} can be written as,

$$\tilde{y}_{k+1} = \tilde{y}_k - \eta I_k \quad (4.27)$$

which is same as (4.7).

Scenario 2: Center similar to \mathbf{x}_k exists in the dictionary or $I_k < \tau_2$

In this scenario, $\mathcal{D}_k = \mathcal{D}_{k+1}$. Then, the \tilde{y}_{k+1} can be written as:

$$\tilde{y}_{k+1} = \tilde{y}_k \quad (4.28)$$

Thus, from (4.27) and (4.28), probabilistically, \tilde{y}_{k+1} can be written as follows:

$$\tilde{y}_{k+1} = \tilde{y}_k - p\eta I_k \quad (4.29)$$

where p and $1 - p$ are probabilities of the event of addition of an element to the dictionary \mathcal{D}_k and the event of keeping the dictionary same respectively. In the initial phase of adaptation, $p \rightarrow 1$ as more and more samples are being added to the dictionary and (4.29) approximates (4.7). In the convergence phase, $p \rightarrow 0$ and $1 - p \rightarrow 1$. That would imply $\mathbb{E}[|\tilde{y}_{k+1}|^2] = \mathbb{E}[|\tilde{y}_k|^2]$. This implies that in the initial transient phase the MSE behavior of sparsified-KMSER is conserved with respect to KMSER and the growth of the adaptive dictionary is turned off upon achieving convergence. Therefore, transient equations developed based on (4.7) are approximately conserved upon sparsification. Thus it can be concluded that the analysis presented in this chapter is valid upon sparsification of the dictionary for the proposed post-distorter.

4.6 Simulations

The proposed sparsified-KMSER is compared against sparsified-KLMS, Volterra-DFE (with 45 linear taps and 25×25 second order taps [8]), and linear DFE with 45 linear taps. The kernel width γ of KLMS and KMSER is determined by Silverman's rule [65], the τ_1 is chosen as 10^{-2} and τ_2 is chosen as 0.3. From Figs. (4.2(a)), (4.2(b)), it is observed that the proposed sparsified-KMSER outperforms the BER performance of Volterra-DFE in the high-SNR regime while maintaining lower computational cost as compared to sparsified-KLMS and Volterra-DFE over "open office" and "office with cubicles channels" of IEEE 802.15 standard respectively, for 4-PAM. For 4-PAM, the sparsified-KMSER has the least computational complexity in all scenarios with equivalent BER performance as compared to sparsified-KLMS and Volterra-DFE. An ensemble of 10000 samples is considered over 500 Monte-Carlo trials in all the BER plots.

In Figs. (4.3(a)), (4.3(b)), it is observed that the proposed sparsified-KMSER actually outperforms the Volterra-DFE in "open-office" and "office with cubicles" scenarios

(in terms of BER) in the low SNR regime with much lower computational complexity in case of 16-QAM. However, as the SNR is increased the RKHS based post-distorters and the Volterra-DFE exhibit similar behavior for 16-QAM modulation scheme. Among the Volterra-DFE, the proposed sparsified-KMSER and sparsified-KLMS, the sparsified-KMSER is found to be computationally more efficient in case of 16-QAM as compared to sparsified-KLMS. Sparsified-KMSER outperforms Volterra-DFE in case of 16-QAM, both computationally, as well as in terms of BER-performance.

Finally, the transient and steady state MSE behavior of the KMSER in Figs. (4.4(a)), (4.4(b)) given by (4.12) and (4.17), are compared over “open-office” and “office with cubicles” scenario for 16-QAM constellation. The transient and the steady state behavior is plotted over a range of step-sizes from $\eta \in [0.25, 0.4]$ and it is found that the simulations closely match the theoretically derived MSE derived curves, thus providing us full control over the step-size range so as to achieve a given MSE floor at a desired convergence rate. Also, the theoretically derived MSE characteristics via Taylor series approximation is also compared with the MSE characteristics obtained assuming perfect knowledge of I_k . The expectations involving I_k -terms in (4.8) (like $\mathbb{E}[\Re\{I_k^* \tilde{y}_k\}]$ and $\mathbb{E}[|I_k|^2]$) were evaluated using numerical techniques as suggested in [54] under Gaussian distribution. It is observed that in all scenarios, they almost overlap, which highlights the validity and applicability of Taylor series approximation of I_k .

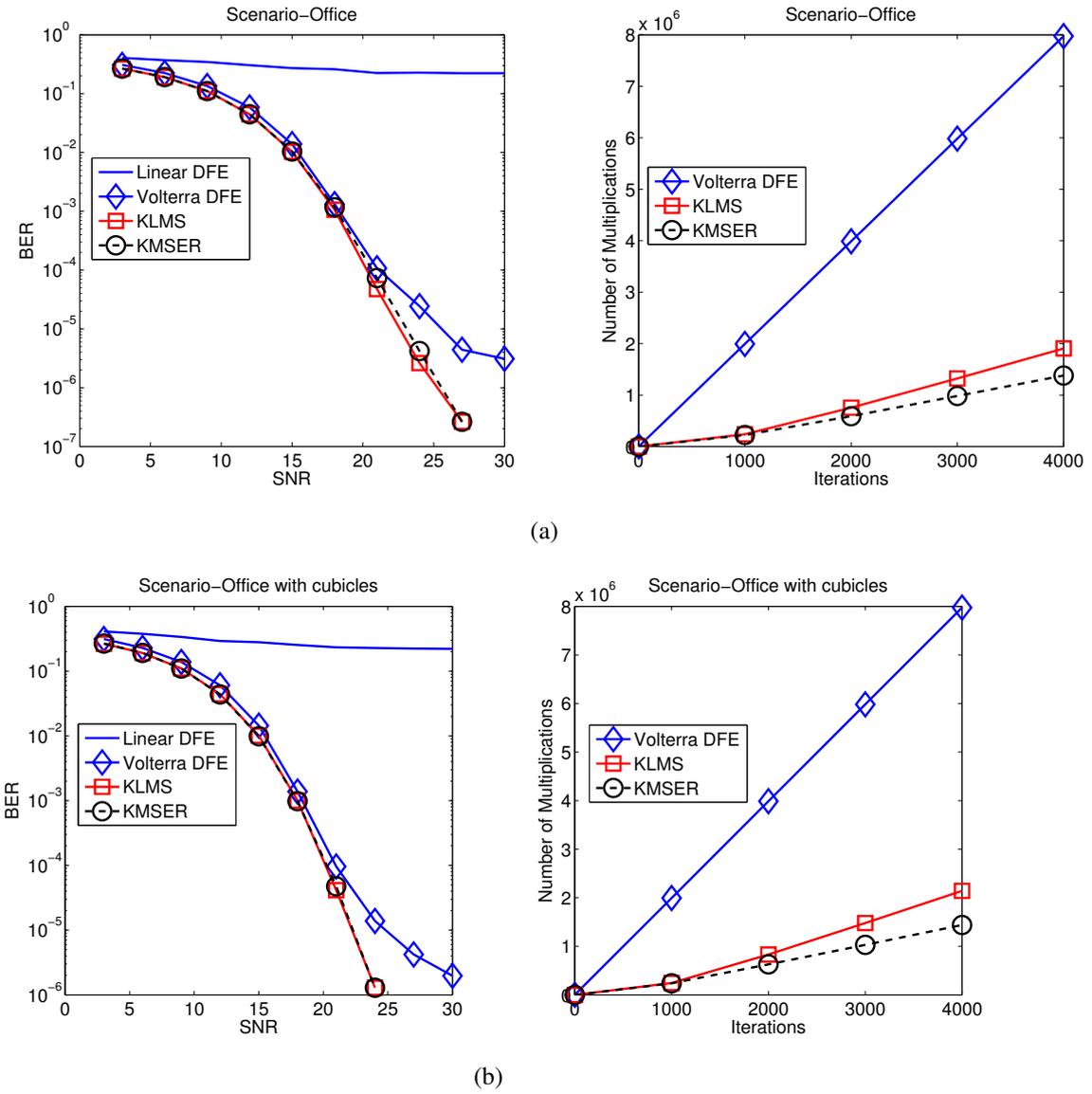


Figure 4.2: BER and computational complexity comparison for 4.2(a) open office IEEE 802.15 PAN channel for 4-PAM. 4.2(b) office with cubicles IEEE 802.15 PAN channel for 4-PAM.

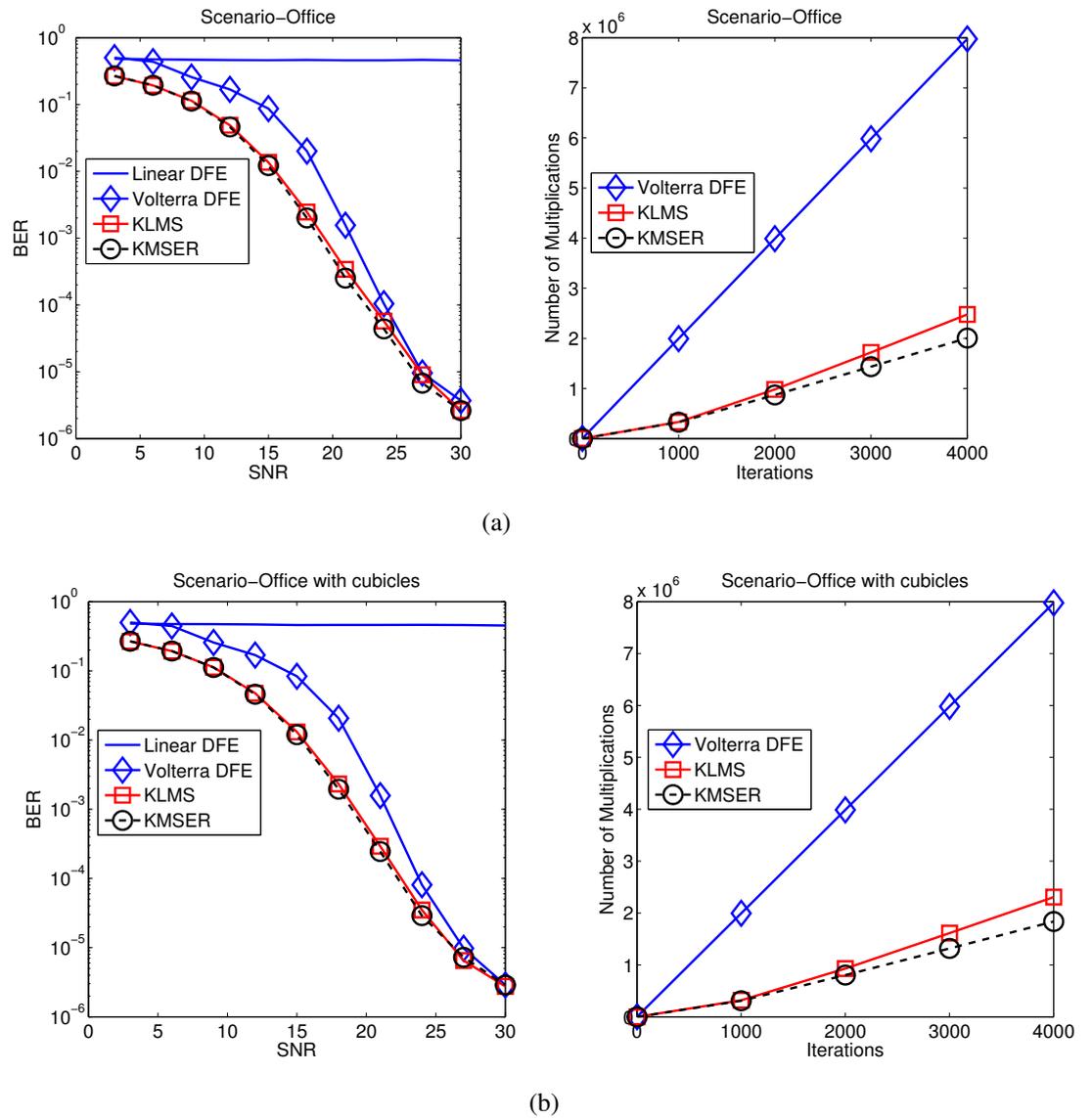


Figure 4.3: BER and computational complexity comparison for 4.3(a) open office IEEE 802.15 PAN channel for 16-QAM. 4.3(b) office with cubicles IEEE 802.15 PAN channel for 16-QAM.

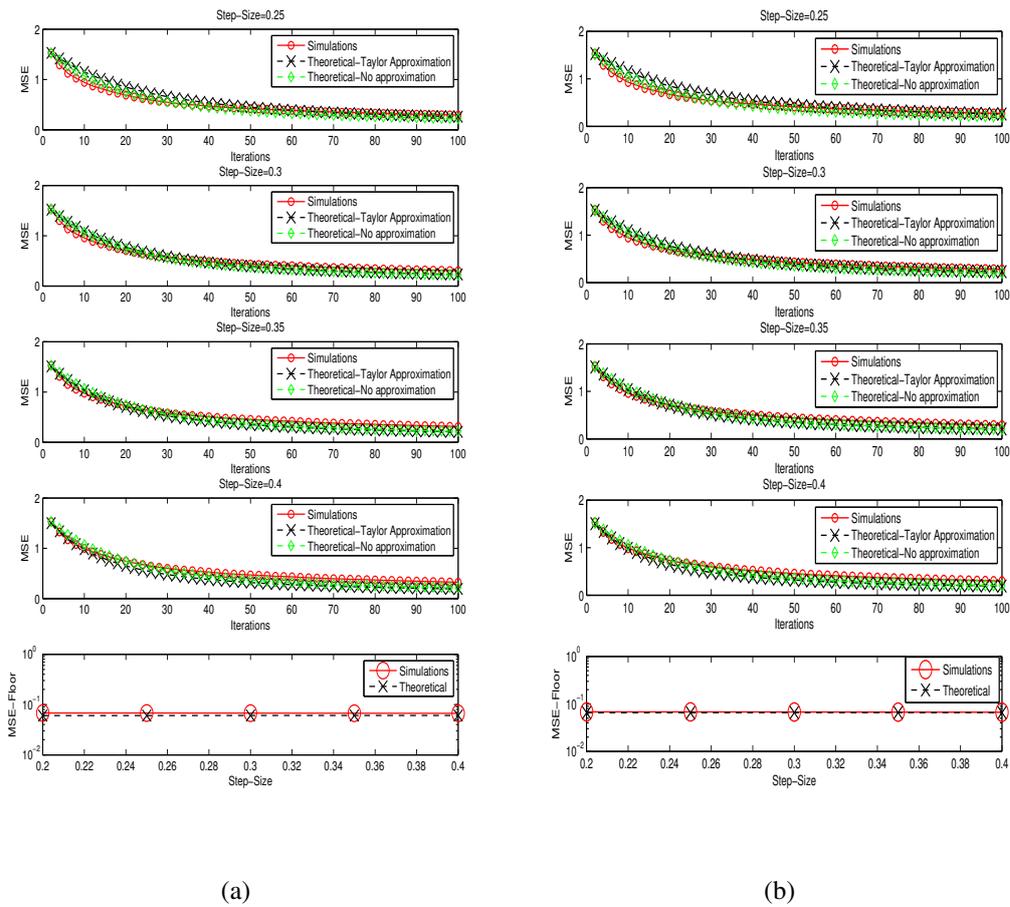


Figure 4.4: 4.4(a) Transient and steady-state MSE analysis of KMSER for “open office” channel for 16-QAM and 4.4(b) Transient and steady-state MSE analysis of KMSER for “office with cubicles” channel for 16-QAM.

4.7 Summary

In this chapter, a severely impaired open-loop VLC system with LED nonlinearity and ISI scenario is considered. Signal processing in this scenario is more difficult as relaxed assumptions of knowledge of channel at the receiver and estimated symbols at the transmitter do not exist. To deal with such nonlinear channel impairments in these scenarios, a novel RKHS based post-distortion technique is proposed to mitigate the LED-nonlinearity at the receiver. First, existing Volterra based post-distorters were surveyed and were found to suffer from three drawbacks: a) convergence to local minima, b) high computational complexity, and c) modeling error due to truncation of Volterra series. As opposed to Volterra-DFE, sparsified KMSER-NC was proposed which has the following advantages: a) problem is convex in RKHS; hence there is guarantee of global optima, b) the approach is computationally simple due to sparsification by novelty criterion, and c) no modeling error occurs due to guarantee of exact representation by the representer theorem. The convergence of sparsified-KMSER is analyzed theoretically, and the simulated MSE behavior is found to be closely matching the theoretically derived dynamical equations for MSE. It is also found asymptotically, that the MSE characteristics of the KMSER are approximately preserved upon sparsification with proper choice of spread parameters and error-thresholds. Additionally, an approximate dynamical equation for convergence was derived and validated with different choice of step-sizes. Superior BER characteristics were observed in case of KMSER as compared to KLMS and Volterra-DFE which indicates that the proposed sparsified-KMSER based post-distortion is a viable solution for equalization of nonlinear indoor IEEE 802.15 PAN VLC channels in the presence of nonlinear LED characteristics.

Chapter 5

Finite Dictionary Techniques for MSER Equalization in RKHS

5.1 Introduction

In this chapter, the problem of post-distortion over IEEE 802.15 PAN VLC channels is once again taken up. It was seen in the last chapter that RKHS techniques deliver equivalent performance as compared to Volterra-filters and have significantly lower computational complexity. The computational complexity of RKHS techniques depends on the sparsification criterion used in order to learn the dictionary. In the previous chapter, the novelty criterion as a sparsification mechanism for learning the dictionary. This chapter explores the use of the quantized criterion [66], and the fixed-budget criterion [67] in conjunction with the KMSER-equalization for learning the dictionary. The fixed budget criterion introduces a mechanism to prune “unimportant” entries in the dictionary (which does not exist among other dictionary-learning techniques) thereby ensuring faster convergence and lower dictionary size.

In this chapter, the quantized kernel MSER (QKMSER) and fixed budget quantized kernel MSER (FBQKMSER) are proposed and analyzed which requires lower dictionary-size as compared to other sparsification approaches without significant loss of performance. Following are the contributions in this chapter:

- Proposing finite dictionary variants of KMSE called QKMSE and FBQKMSE.
- Derive corresponding exact theoretical expression for transient and steady-state MSE within a unified framework.
- Finding conditions (bounds on η) for convergence for QKMSE and FBQKMSE.

Simulation carried over IEEE 802.15 PAN VLC channels, indicate that: i) the MSE performance of QKMSE and FBQKMSE is not affected significantly while working with a finite dictionary, ii) the FBQKMSE outperforms all algorithms in terms of convergence metric characteristics in considered simulation conditions, and iii) the theoretically derived and simulated MSE dynamical curves are almost overlapping, which establishes the validity of the derived dynamical equations for QKMSE and FBQKMSE.

The following terminology is used in this chapter: scalar at time k is represented by $(\cdot)_k$ and vector of past M samples at time k is represented by boldface with subscript k such as \mathbf{x}_k (which are elements of \mathbb{C}^M , M being the row dimension of the vector, and \mathbb{C} denotes the field of complex numbers) and italicised variables like \mathcal{D}_k and \mathcal{I}_k denote an online dictionary of observation and error terms respectively, at the time instance k . Real part of a complex quantity is denoted by the superscript $(\cdot)^R$ and $\Re\{\cdot\}$, and the imaginary part is denoted by $(\cdot)^I$ and $\Im\{\cdot\}$.

This chapter is organized as follows: Section-5.2 provides the system model assumed in this chapter that will be used in following sections. The QKMSE and FBQKMSE are proposed in Section-5.3 and Section-5.4 respectively. Mathematical analysis of QKMSE and FBQKMSE is done in Section-5.5. To validate the proposed algorithm against existing approaches, simulations are provided in Section-5.6. To summarize the contributions of this chapter, the conclusions are drawn in Section-5.7.

5.2 System model

In this section, the system model of the considered VLC system is described. Let s_k denote the input constellation at the k^{th} time instant. It is passed through an FIR filter

$\{h_i\}_{i=1}^L$ (given by IEEE 802.15 PAN VLC standard) denoting channel coefficients, where L is the tap length. The received symbol at k^{th} time instant, x_k , is given as follows:

$$x_k = \sum_{i=0}^{L-1} h_i A(s_{k-i}) + n_k \quad (5.1)$$

$A(\cdot)$, as mentioned previously, denotes the-transmit side Rapp non-linearity, n_k is i.i.d AWGN with variance σ_n^2 and D denotes the equalizer delay. The channel consists of a linear FIR filter $\mathbf{h} = [h_0, h_1, \dots, h_{L-1}]$. The equalizer considered in this chapter is a non-linear adaptive equalizer with symbols s_k and observation \mathbf{x}_k as input. \hat{s}_{k-D} is the equalizer estimate delayed by D samples, where D is the cumulative lag of the equalizer.

5.3 Quantized kernel MSER

To avoid the unbounded growth of the kernel dictionary, it is selectively grown depending on a similarity criterion for incoming regressors for the MSER based algorithm in [68]. Along the lines of the work reviewed in [66, 67, 69], the quantized kernel MSER is proposed. Let us assume a dictionary $\mathcal{D}_k = \{\mathcal{I}_k^{(j)}, \mathbf{x}_k^{(j)}\}_{j=1}^{|\mathcal{D}_k|}$, where $|\mathcal{D}_k|$ denotes cardinality of the dictionary. Please note the terminology here; \mathcal{I}_k is the set of all I_k which are present at instant k in the dictionary. $\mathcal{I}_k^{(j)}$ denotes the j^{th} element of the set \mathcal{I}_k , which is a measure of deviation from the MSER constraint [54]. Using above terminology, quantized kernel MSER algorithm is proposed (as described in Algorithm 2). This algorithm (as described in Algorithm 2) initializes by taking a step-size of η , kernel bandwidth σ and an initial dictionary of size 1 consisting of the first observation and innovation. Then, for every k^{th} iteration, the output y_k is calculated and the innovation I_k for KMSER is calculated using y_k . Then the current observation \mathbf{x}_k , is compared using Euclidean distance $\|\cdot\|_2$ to all members in the dynamic dictionary of observations \mathcal{D}_k . If the Euclidean distance to the closest member of the dictionary is less than threshold ϵ , then the MSER innovation $\mathcal{I}_k^{j^*}$ in \mathcal{D}_k is updated. Otherwise, the new observation \mathbf{x}_k is added to the dictionary. Proceeding with such selective addition of data, the unbounded growth of the dictionary is partly curtailed, while maintaining a reasonable level of performance.

Algorithm 2 Quantized Kernel MSER (QKMSER)

```

1: Initialise step-size  $\eta$ , kernel width  $\gamma = \sqrt{\frac{1}{2\sigma^2}}$  and quantization threshold  $\varepsilon > 0$ , and
   initial dictionary  $\mathcal{D}_0 = \{\delta_0, x_0\}$ .
2: while  $|\mathcal{D}_k| \geq 1$  do
3:    $y_k = \eta \sum_{j=1}^{|\mathcal{D}_{k-1}|} \kappa_{\gamma, \mathbb{C}^d \times \mathbb{C}^d}(\mathcal{D}_{k-1}^{(j)}, \mathbf{x}_k) I_{k-1}^{(j)}$ 
4:    $I_k = \tanh(\beta(\Re(y_k) - \Re(s_{k-D}) + 1)) + \tanh(\beta(\Re(y_k) - \Re(s_{k-D}) - 1)) +$ 
      $\sqrt{-1}(\tanh(\beta(\Im(y_k) - \Im(s_{k-D}) + 1)) +$ 
5:      $\tanh(\beta(\Im(y_k) - \Im(s_{k-D}) - 1)))$ 
6:    $j^* = \arg \min_{1 \leq j \leq |\mathcal{D}_{k-1}|} \|\mathbf{x}_k - \mathcal{D}_{k-1}^{(j)}\|$ 
7:   if  $\|\mathbf{x}_k - \mathcal{D}_{k-1}^{j^*}\| \leq \varepsilon$  then
8:      $\mathcal{D}_k = \mathcal{D}_{k-1}$ 
9:      $\mathcal{I}_k^{(j^*)} = \mathcal{I}_{k-1}^{(j^*)} + \eta I_k$ 
10:  else
11:     $\mathcal{D}_k = \mathcal{D}_{k-1} \cup \mathbf{x}_k, \mathcal{I}_k = \mathcal{I}_{k-1} \cup I_k$ 
12:  end if
13: end while
    
```

Though this algorithm curtails the infinite memory-storage requirement of KMSER by selectively growing the dictionary based on Euclidean distance to the nearest entry in the dictionary, it gives no technique of shrinking the given dictionary by rejecting unimportant samples. The algorithm given in the next section introduces a technique for shrinking the dictionary based on significance values for kernel MSER.

5.4 Fixed budget quantized kernel MSER

This technique gives a method to prune the size of the available dictionary by techniques which are based on online estimation of a term called ‘‘significance’’. Let ζ be a forgetting factor such that $0 < \zeta < 1$. Significance is estimated in the following manner depending on whether a new center is added, merged or pruned. In case a center is added, the significance, $E_k^{(j)}$, for j^{th} entry of the dictionary at k^{th} time instant is updated as follows [67]:

$$E_k^{(j)} = \zeta E_{k-1}^{(j)} + |\mathcal{I}_k^{|\mathcal{D}_{k-1}|+1}| \kappa_{\gamma, \mathbb{C}^d \times \mathbb{C}^d}(\mathcal{D}_{k-1}^{(j)}, \mathcal{D}_k^{(|\mathcal{D}_{k-1}|+1)}), \quad (5.2)$$

$$\forall 1 \leq j \leq |\mathcal{D}_{k-1}|$$

(i) In case of merging:

$$E_k^{(j \neq j^*)} = \zeta E_{k-1}^{(j \neq j^*)} + |\mathcal{S}_k^{(j \neq j^*)}| \kappa_{\gamma, \mathbb{C}^d \times \mathbb{C}^d}(\mathcal{D}_{k-1}^{(j \neq j^*)}, \mathcal{D}_k^{(j^*)}) \quad (5.3)$$

where $\lambda_k^{(j)}$ is a variable which is updated as:

$$\lambda_k^{(j)} = \zeta \lambda_{k-1}^{(j)} \quad (5.4)$$

and,

$$E_k^{(j^*)} = \frac{|\mathcal{S}_k^{(j)} + \eta I_k|}{|\mathcal{S}_k^{(j)}|} \mathcal{S}_k^{(j)} \zeta E_{k-1}^{(j^*)} + |\mathcal{S}_k^{(j)} + \eta I_k| \kappa_{\gamma, \mathbb{C}^d \times \mathbb{C}^d}(\mathcal{D}_{k-1}^{(j^*)}, \mathcal{D}_k^{(j^*)}) \quad (5.5)$$

(ii) In case of deletion/pruning of the M^{th} dictionary entry,

$$E_k^{(j)} = E_{k-1}^{(j)} - |\mathcal{S}_k^{(j)}| \lambda_{k-1}^{(j)} \kappa_{\gamma, \mathbb{C}^d \times \mathbb{C}^d}(\mathcal{D}_{k-1}^{(j)}, \mathcal{D}_{k-1}^{(M)}) \quad (5.6)$$

$$\lambda_k^{(j)} = \zeta \lambda_{k-1}^{(j)} + 1 \quad (5.7)$$

Using these online estimates of significance, the proposed FBQKMSER is given in Algorithm 3. This algorithm begins with a quantization threshold ε , kernel-width σ an initial dictionary \mathcal{D}_0 with initial values of the innovation and first observation \mathbf{x}_0 . Then for every k^{th} iteration the output at the k^{th} instant y_k is calculated using the innovation I_k at the k^{th} iteration. Now this algorithm branches into three possibilities: a) Updation: If the current observation \mathbf{x}_k is close enough in the Euclidean norm $\|\cdot\|_2$ to any of the elements of the dictionary \mathcal{D}_k at time instant k , the corresponding innovation in \mathcal{D}_k is updated, b) Addition: If \mathbf{x}_k is significantly “different” (in the sense of $\|\cdot\|_2$) as compared to all elements in the dictionary, then the innovation I_k and \mathbf{x}_k are appended to the dictionary, and c) Pruning: The entry with minimum significance at each iteration E_k^j is deleted. Thus

Algorithm 3 Fixed Budget Quantized Kernel MSER (FBQKMSER)

```

1: Initialise step-size  $\eta$ , kernel width  $\gamma = \sqrt{\frac{1}{2\sigma^2}}$  and quantization threshold  $\varepsilon > 0$ , and
   initial dictionary  $\mathcal{D}_0 = \{\delta_0, \mathbf{x}_0\}$ .
2: while  $|\mathcal{D}_k| \geq 1$  do
3:    $y_k = \eta \sum_{j=1}^{|\mathcal{D}_{k-1}|} \mathbf{k}_{\gamma, \mathbb{C}^d \times \mathbb{C}^d}(\mathcal{D}_{k-1}^{(j)}, \mathbf{x}_k) I_{k-1}^{(j)}$ 
4:    $I_k = \tanh(\beta(\Re(y_k) - \Re(s_{k-D}) + 1)) + \tanh(\beta(\Re(y_k) - \Re(s_{k-D}) - 1))$ 
5:    $+ \sqrt{-1}(\tanh(\beta(\Im(y_k) - \Im(s_{k-D}) + 1)) + \tanh(\beta(\Im(y_k) - \Im(s_{k-D}) - 1)))$ 
6:    $j^* = \arg \min_{1 \leq j \leq |\mathcal{D}_{k-1}|} \|\mathbf{x}_k - \mathcal{D}_{k-1}^{(j)}\|$ 
7:   if  $\|\mathbf{x}_k - \mathcal{D}_{k-1}^{(j^*)}\| \leq \varepsilon$  then
8:      $\mathcal{D}_k = \mathcal{D}_{k-1}$ 
9:      $\mathcal{I}_k^{(j^*)} = \mathcal{I}_{k-1}^{(j^*)} + \eta I_k$ 
10:    Update significance  $\{E_k^{(j)}\} \forall j$  as per eq. (5.3) and eq. (5.5).
11:   else
12:      $\mathcal{D}_k = \mathcal{D}_{k-1} \cup \mathbf{x}_k, \mathcal{I}_k = \mathcal{I}_{k-1} \cup I_k$ 
13:     Update significance  $\{E_k^{(j)}\} \forall j$  as per eq. (5.3) and eq. (5.5).
14:     Also update significance for newly added tuple as per eq. (5.2).
15:   end if
16:    $\mathcal{D}'_{k-1} = \{(\mathcal{I}_k^{(j)}, \mathbf{x}_k^{(j)}) \in \mathcal{D}_{k-1} \forall j : E_k^{(j)} = \min\{E_k\}\}$ 
17:    $\mathcal{D}_k = \mathcal{D}_{k-1} - \mathcal{D}'_{k-1}$ 
18:   Update significance  $\{E_k^{(j)}\} \forall j$  as per eq. (5.6).
19: end while

```

this provides a means of deleting unimportant entries in the dictionary that helps in better learning and tracking in non-stationary environments as observed from simulation in Section-5.6.

In the next section, analysis of the performance of the proposed QKMSER and FBQKMSER is presented. Recursive dynamical equations parameterized by the step-size η that dictate the MSE evolution with respect to iterations, are derived. Also expressions for steady state MSE and conditions for convergence of the proposed algorithms are derived theoretically.

5.5 Expression for transient and steady-state behavior of QKMSER and FBQKMSER

In this section, in order to analyze the MSE dynamics of the proposed approaches, the equations for the transient behavior of the QKLMS and FBQKMSER are developed. First, the state transition probabilities are calculated and consequently a generic probabilistic

adaptation equation for QKMSER and FBQKMSER is defined.

First, measure of the interval, given by $\mu_1(n)$, that a new entry is not added in the dictionary. The mathematical expression for $\mu_1(n)$ is given by:

$$\mu_1(k) = \nu(\mathbb{E}[\kappa(\mathbf{x}, \mathbf{x}_k)] > \varepsilon) \quad (5.8)$$

where $\nu(\cdot)$ denotes measure of the interval. This can be written as follows using result from [67]:

$$\mu_1(k) \approx \frac{1}{|\mathcal{D}_k|} \sum_{j=1}^{|\mathcal{D}_k|} \int_{-\infty}^{\infty} \kappa(\mathbf{x}, \mathbf{x}_k) \kappa(\mathbf{x}, \mathbf{x}_j) d\mathbf{x} \quad (5.9)$$

as ε is assumed to be small. This is equivalent to (for the widely-used Gaussian kernel):

$$\mu_1(k) \approx \frac{1}{|\mathcal{D}_k|} \sum_{j=1}^{|\mathcal{D}_k|} \kappa(\mathbf{x}_k, \mathbf{x}_j) \quad (5.10)$$

Measure μ_2 of the dictionary-size being incremented is given by, $\mu_2(k) \approx 1 - \mu_1(k)$. To compute measure of interval such that the significance drops below a threshold for a dictionary to get decremented, $p_3(n)$, one needs to find probability of the following event:

$$\mathbb{E} \left[|I_k| \exp \left(\frac{-\|\mathbf{x}_j - \mathbf{x}_k\|^2}{2\sigma^2} \right) \right] < \varepsilon' \quad (5.11)$$

where $\varepsilon' \rightarrow 0$. Hence $p_3(n)$ is given by,

$$\begin{aligned} \mu_3(k) &\approx \mathbb{E}[|I_k|] \left[1 - \int \dots \int_{-\infty}^{\infty} \exp \left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|^2}{2\sigma^2} \right) \right. \\ &\quad \left. \frac{1}{|\mathcal{D}_k|} \sum_{j=1}^{|\mathcal{D}_k|} \kappa(\mathbf{x} - \mathbf{x}_j) d\mathbf{x} \right] = \\ &\mathbb{E}[|I_k|] \left\{ 1 - \left[\frac{1}{|\mathcal{D}_k|} \sum_{j=1}^{|\mathcal{D}_k|} \exp \left(\frac{-\|\mathbf{x}_j - \mathbf{x}_k\|^2}{2\sigma^2} \right) \right] \right\} \end{aligned} \quad (5.12)$$

assuming the random variable I_k and proximity measure $\exp(\frac{-\|\mathbf{x}_j - \mathbf{x}_k\|^2}{2\sigma^2})$ to be independent of each other. $\mathbb{E}[|I_k|]$ can be estimated online to calculate theoretical $\mu_3(k)$. Hence the transformed transition probabilities would be given by, $p_i(k) = \frac{\mu_i(k)}{\sum_{i=1}^3 \mu_i(k)}$. The assumption

5.5. EXPRESSION FOR TRANSIENT AND STEADY-STATE BEHAVIOR OF QKMSER AND FBQKMSER

used in the following derivation are: a) I_k is i.i.d, and b) $0 < \eta \ll 1$. By using the probabilities $\{p_i(k)\}$, one can evaluate the transient behavior of QKMSER and FBQKMSER. Dynamical equations for the a-priori and a-posteriori deviation of the output y_k from the fixed point y^o can be written probabilistically as \tilde{y}_k :

$$\begin{aligned} \tilde{y}_{k+1} = & \tilde{y}_k - \eta p_1(k) \left(\frac{|\mathcal{D}_k| - 1}{|\mathcal{D}_k| + 1} \right) I_k \langle \phi(\mathbf{x}_{j^*}), \phi(\mathbf{x}_k) \rangle \\ & - \eta p_2(k) \left(\frac{1}{|\mathcal{D}_k| + 1} \right) \mathcal{I}_k^{(|\mathcal{D}_k|+1)} \langle \phi(\mathbf{x}_{|\mathcal{D}_k|+1}), \phi(\mathbf{x}_k) \rangle \\ & + \eta p_3(k) \left(\frac{1}{|\mathcal{D}_k| + 1} \right) \mathcal{I}_k^{(M)} \langle \phi(\mathbf{x}_M), \phi(\mathbf{x}_k) \rangle \end{aligned} \quad (5.13)$$

Squaring both sides and taking expectation,

$$\begin{aligned} \mathbb{E}(|\tilde{y}_{k+1}|^2) = & \mathbb{E}(|\tilde{y}_k|^2) + \eta^2 \left(\frac{|\mathcal{D}_k| - 1}{|\mathcal{D}_k| + 1} \right)^2 p_1^2(k) \mathbb{E}[|I_k|^2] \\ & \eta^2 p_2^2(k) \left(\frac{1}{|\mathcal{D}_k| + 1} \right)^2 \mathbb{E}[|I_k|^2] + \eta^2 p_3^2(k) \left(\frac{1}{|\mathcal{D}_k| + 1} \right)^2 \mathbb{E}[|I_k|^2] - \\ & 2\eta \mathbb{E}[\tilde{y}_k I_k] \left\{ \left(\frac{|\mathcal{D}_k| - 1}{|\mathcal{D}_k| + 1} \right) p_1(k) + \frac{1}{|\mathcal{D}_k| + 1} p_2(k) - \frac{1}{|\mathcal{D}_k| + 1} p_3(k) \right\} \end{aligned} \quad (5.14)$$

From Taylor's series approximation of I_k around 1, $\mathbb{E}(|I_k|^2) \approx 0.29(\mathbb{E}[|\tilde{y}_k|^2] + \sigma_n^2)$ and $\mathbb{E}[\tilde{y}_k I_k] \approx 0.33\mathbb{E}[|\tilde{y}_k|^2]$. Let $C_1 = \left(\frac{|\mathcal{D}_k| - 1}{|\mathcal{D}_k| + 1} \right)^2 p_1^2(k) + p_2^2(k) \left(\frac{1}{|\mathcal{D}_k| + 1} \right)^2 + p_3^2(k) \left(\frac{1}{|\mathcal{D}_k| + 1} \right)^2$.

Using these values, the following *exact* equation for transient behavior of FBQKMSER is derived,

$$\begin{aligned} \mathbb{E}(|\tilde{y}_{k+1}|^2) = & \left[1 - 0.66\eta \left\{ \left(\frac{|\mathcal{D}_k| - 1}{|\mathcal{D}_k| + 1} \right) p_1(k) \right. \right. \\ & \left. \left. + \frac{1}{|\mathcal{D}_k| + 1} p_2(k) - \frac{1}{|\mathcal{D}_k| + 1} p_3(k) \right\} + 0.29C_1 \eta^2 \right] \mathbb{E}(|\tilde{y}_k|^2) \\ & + 0.29\eta^2 C_1 \sigma_n^2 \end{aligned} \quad (5.15)$$

Using the assumption $0 < \eta \ll 1$ the approximate dynamical equation for the transient

behavior for the FBQKMSER is given by:

$$\begin{aligned} \mathbb{E}(|\tilde{y}_{k+1}|^2) = & \left[1 - 0.66\eta \left\{ \left(\frac{|\mathcal{D}_k| - 1}{|\mathcal{D}_k| + 1} \right) p_1(k) \right. \right. \\ & \left. \left. + \frac{1}{|\mathcal{D}_k| + 1} p_2(k) - \frac{1}{|\mathcal{D}_k| + 1} p_3(k) \right\} \right] \mathbb{E}(|\tilde{y}_k|^2) + 0.29\eta^2 C_1 \sigma_n^2 \end{aligned} \quad (5.16)$$

In the same framework putting $p_3 = 0$ gives the *exact* transient behavior for QKMSER as under:

$$\begin{aligned} \mathbb{E}(|\tilde{y}_{k+1}|^2) = & \left[1 - 0.66\eta \left\{ \left(\frac{|\mathcal{D}_k| - 1}{|\mathcal{D}_k|} \right) p_1(k) \right. \right. \\ & \left. \left. + \frac{1}{|\mathcal{D}_k|} p_2(k) \right\} + 0.29\eta^2 \left\{ \left(\frac{|\mathcal{D}_k| - 1}{|\mathcal{D}_k|} \right)^2 p_1^2(k) + \right. \right. \\ & \left. \left. p_2^2(k) \left(\frac{1}{|\mathcal{D}_k|} \right)^2 \right\} \right] \mathbb{E}(|\tilde{y}_k|^2) + 0.29\eta^2 \left\{ \left(\frac{|\mathcal{D}_k| - 1}{|\mathcal{D}_k|} \right)^2 p_1^2(k) + \right. \\ & \left. \left(\frac{1}{|\mathcal{D}_k|} \right)^2 p_2^2(k) \right\} \sigma_n^2 \end{aligned} \quad (5.17)$$

Under approximation of small step size, the approximate dynamical equation for the transient behavior for the QKMSER is given by:

$$\begin{aligned} \mathbb{E}(|\tilde{y}_{k+1}|^2) = & \left[1 - 0.66\eta \left\{ \left(\frac{|\mathcal{D}_k| - 1}{|\mathcal{D}_k|} \right) p_1(k) \right. \right. \\ & \left. \left. + \frac{1}{|\mathcal{D}_k|} p_2(k) \right\} \right] \mathbb{E}(|\tilde{y}_k|^2) \\ & + 0.29\eta^2 \left\{ \left(\frac{|\mathcal{D}_k| - 1}{|\mathcal{D}_k|} \right)^2 p_1^2(k) + p_2^2(k) \left(\frac{1}{|\mathcal{D}_k|} \right)^2 \right\} \sigma_n^2 \end{aligned} \quad (5.18)$$

These expressions which were derived for transient behavior of the proposed algorithms imply the following performance metric properties:

5.5.1 Step-size range for convergence

In this section, step-size range for both QKMSER and FBQKMSER is derived by the analysis which is provided below:

FBQKMSER

From eq. (5.15), let there be a constant C_2 , which is given as follows:

$$C_2 = \left(\frac{|\mathcal{D}_k| - 1}{|\mathcal{D}_k| + 1} \right) p_1(k) + \left(\frac{1}{|\mathcal{D}_k| + 1} \right) p_2(k) - \left(\frac{1}{|\mathcal{D}_k| + 1} \right) p_3(k) \quad (5.19)$$

From eq. (5.15), the adaptation converges iff,

$$|(1 - 0.66\eta C_2 + 0.29C_1\eta^2)| < 1 \quad (5.20)$$

Hence, the step-size range for convergence is given by:

$$0 < \eta < \frac{C_2}{C_1} 2.2 \quad (5.21)$$

QKMSER

Let there be two constants C_3 and C_4 as follows:

$$C_3 = \left(\frac{|\mathcal{D}_k| - 1}{|\mathcal{D}_k|} \right) p_1(k) + \frac{1}{|\mathcal{D}_k|} p_2(k) \quad (5.22)$$

$$C_4 = \left(\frac{|\mathcal{D}_k| - 1}{|\mathcal{D}_k|} \right)^2 p_1^2(k) + p_2^2(k) \frac{1}{|\mathcal{D}_k|} \quad (5.23)$$

From eq. (5.18), the adaptation converges iff,

$$|(1 - 0.66\eta C_3 + 0.29C_4\eta^2)| < 1 \implies 0 < \eta < 2.2 \frac{C_3}{C_4} \quad (5.24)$$

5.5.2 Steady state misadjustment

The steady-state misadjustment for FBQKMSER and QKMSER is mathematically given as follows:

FBQKMSER

From eq. (5.15), the steady-state value is given by:

$$\mathbb{E}(|\tilde{y}_\infty|^2) = \frac{0.29}{0.66} \eta \sigma_n^2 \frac{C_1}{C_2} \quad (5.25)$$

This equation governs the steady state misadjustment of FBQKMSER as a function of step-size η . Note that $\frac{C_3}{C_4}, \frac{C_1}{C_2} \rightarrow 1$ at convergence with $\frac{C_1}{C_2} < \frac{C_4}{C_3}$.

QKMSER

From eq. (5.18), the adaptation converges to a steady-state value under the assumption of small step-size given by,

$$\mathbb{E}(|\tilde{y}_\infty|^2) = \frac{0.29}{0.66} \eta \sigma_n^2 \frac{C_4}{C_3} \quad (5.26)$$

This equation governs the steady state misadjustment of QKMSER as a function of step-size η .

5.6 Simulations

In this section, simulations are provided that will validate the proposed approaches against the existing literature. The kernel parameter $\gamma = \sqrt{\frac{1}{2\sigma^2}}$ used throughout for all the algorithms is chosen by the well known Silverman's rule [65, 70, 71]. An ensemble of 200 Monte-Carlo simulations has been chosen for all the simulations as further averaging did not change the variance of the output-simulation parameters significantly. The modulation scheme was assumed as QPSK. In Fig. 5.1 and Fig. 5.2, the proposed FBQKMSER is compared with existing sparsification procedures like KLMS with novelty criterion (KLMS-NC) and KMSER with novelty criterion (KMSER-NC) as proposed in [72] for post-distortion in VLC channels [24]. Two IEEE PAN 802.15 standardized VLC channels were considered, namely, for the ‘‘office’’ scenario and for the ‘‘office with cubicles’’ scenario. BER performance was monitored for various sparsification algorithms considered

in this chapter whilst monitoring dictionary size/computational complexity trade-off.

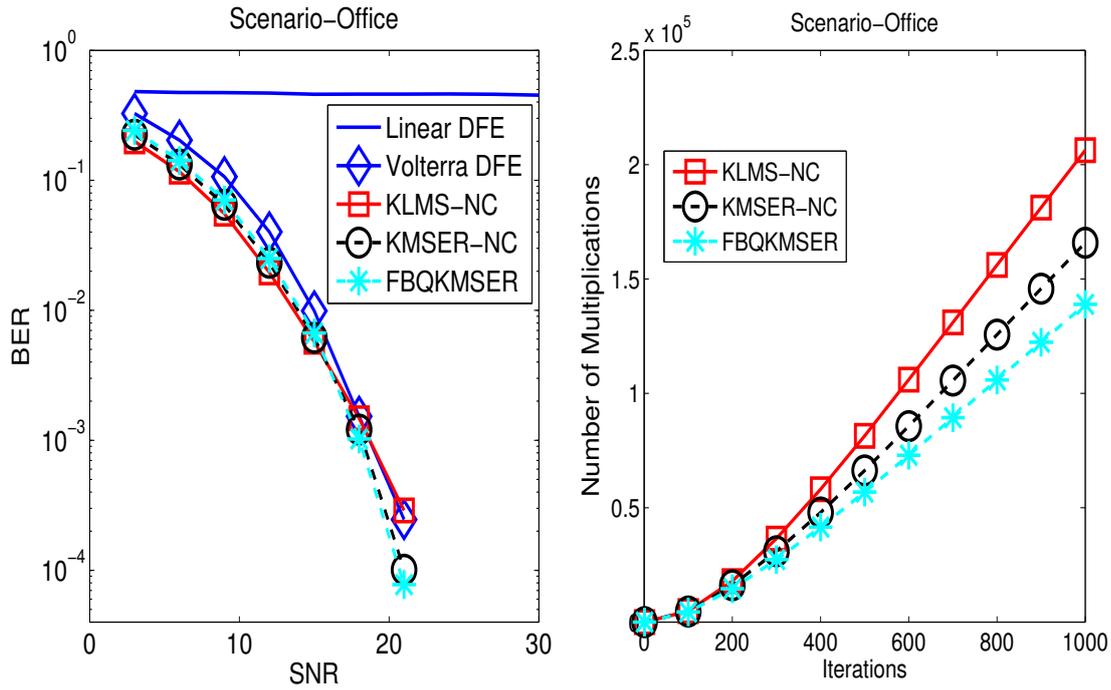


Figure 5.1: BER and computational complexity comparison of FBQKMSER with other sparsification approaches over IEEE 802.15 PAN 'Office' indoor VLC channel.

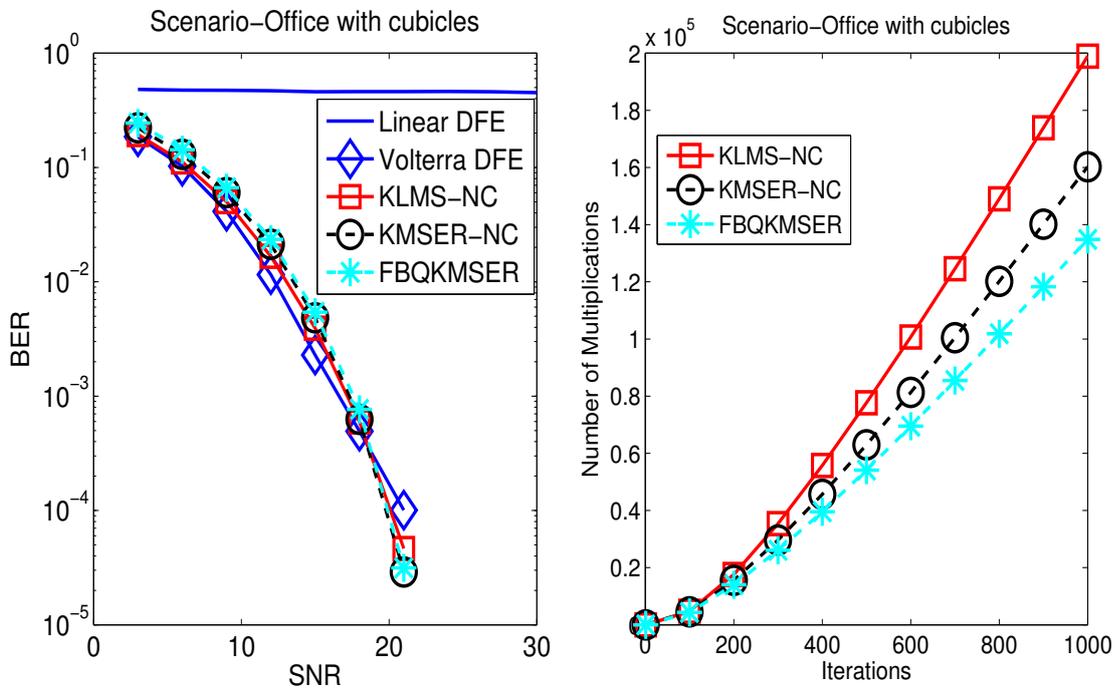


Figure 5.2: BER and computational complexity comparison of FBQKMSER with other sparsification approaches over IEEE 802.15 PAN 'Office with cubicles' indoor VLC channel.

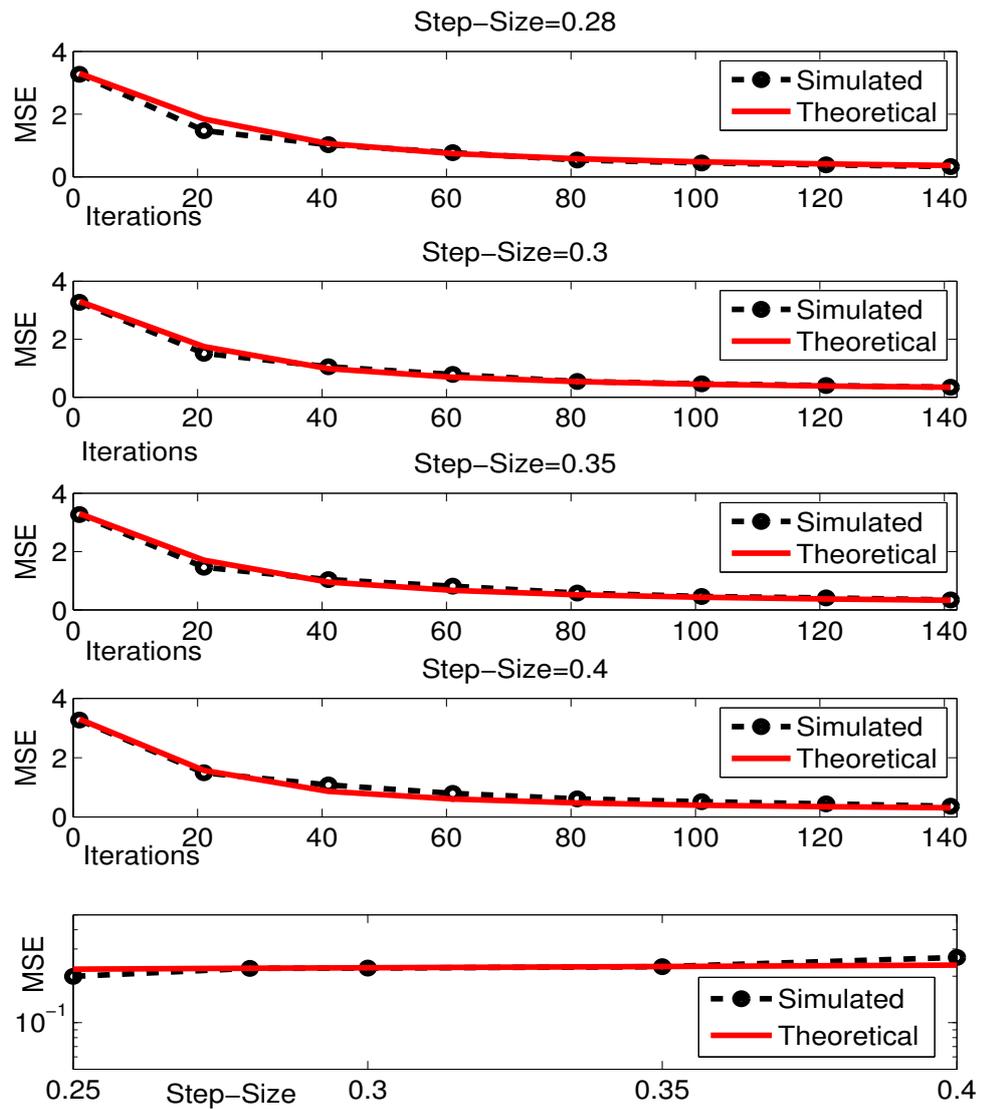


Figure 5.3: Theoretical validation of MSE for FBQKMSER over IEEE 802.15 PAN 'Office' indoor VLC channel (for LED non-linearity).

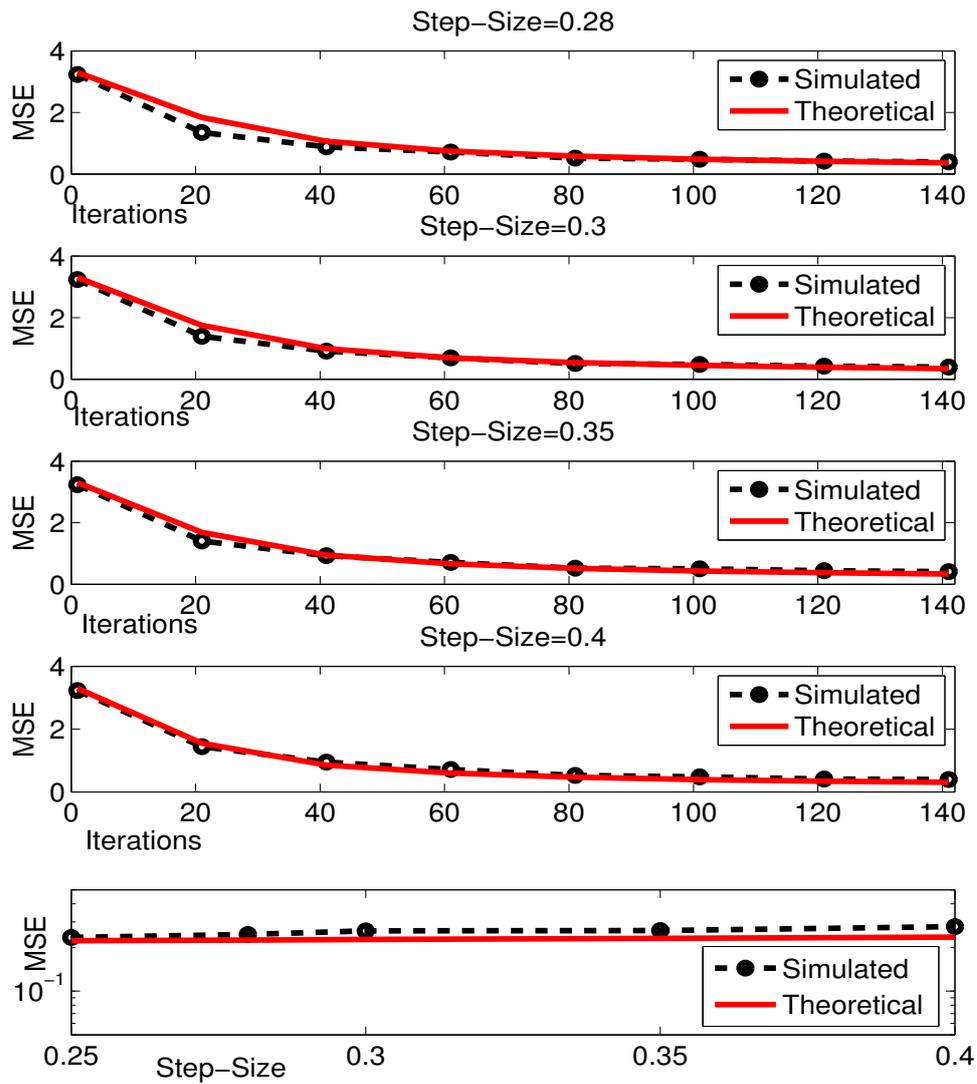


Figure 5.4: Theoretical validation of MSE for FBQMSER over IEEE 802.15 PAN 'Office with cubicles' indoor VLC channel (for LED non-linearity).

It can be observed from Fig. 5.1 and Fig. 5.2, that the proposed FBQKMSEr maintains equivalent performance as compared to KMSEr-NC and KLMS-NC (proposed in [72]) with much lesser number of computations. This validates FBQKMSEr as a better and more applicable sparsification mechanism as it facilitates for computational simplicity without sacrificing performance.

Additionally, in Fig. 5.3 and Fig. 5.4, the transient behavior for MSE dynamics for FBQKMSEr is compared against the theoretical MSE behavior of FBQKMSEr derived from (5.15) for various step-size values. It is observed that the theoretical and simulated curves almost overlap with each other for both the considered VLC channels. This validates the derived dynamical equation in (5.15) in a practical VLC scenario.

5.7 Summary

In this chapter the problem of mitigating device impairments over VLC channels like LED nonlinearity and ISI was re-visited by providing techniques that facilitate for further lowering computational complexity as compared to previous approaches. A unifying framework of analysis was given for sparse-KMSEr filtering techniques. In this chapter, two new sparse RKHS adaptive post-distortion techniques inspired by MSER criterion have been proposed to mitigate the LED nonlinearity at the receiver, namely: a) QKMSEr, and b) FBQKMSEr over VLC channels. In particular, FBQKMSEr relies on a measure of “significance” of an observation to grow/prune the dictionary which gives us a mechanism for rejecting the redundant observations, thereby facilitating for lower-dictionary sizes as compared to other sparsification criteria like novelty criterion/QKMSEr. Theoretical formulae have been derived to analyze the transient and steady-state convergence of these approaches (QKMSEr and FBQKMSEr), along with KMSEr and complex kernel least mean squares (CKLMS) within a unified framework. Ranges for step-size η have been found in this chapter in which the proposed algorithms converge. This makes the analysis presented in this chapter, an important theoretical contribution in the analysis of non-linear sparse kernel adaptive equalization over VLC channels. It is also found in this

chapter that the FBQKMSER converges to a lower dictionary size as QKMSER/KMSER-NC while delivering similar BER performance which confirms its viability as an efficient sparsification mechanism as compared to quantized/ NC-based approaches.

Chapter 6

Unsupervised Multi-Stage Clustering based Hammerstein Post-distortion for VLC

6.1 Introduction

In this chapter, the problem of post-distortion over impaired VLC channels is revisited once again. The open loop VLC system model is considered with a nonlinear post-distorter/equalizer at the receiver. There are three basic techniques of training an equalizer/learning the equalizer coefficients: a) supervised based training, b) unsupervised based training and c) semi-supervised training. Supervised training assumes knowledge of training data/labels to be available. On the other hand, unsupervised-based training does not assume any knowledge of training data. Thus it saves precious bandwidth by reducing the overhead required for training symbols and hence enhances the achievable data rate. Semi-supervised based training assumes partial knowledge of labels. This chapter focuses on the problem of unsupervised learning of the post-distorter coefficients (without knowledge of training symbols). While there has been a growing literature on various supervised post-distortion techniques based on Volterra filtering and other nonlinear approaches [28, 72], most literature on *unsupervised* post-distortion for VLC still relies on

multi-modulus algorithm (MMA) based approaches like cascaded-MMA (CMMA) [73] and the MCMMA Volterra filtering [7, 8, 74].

In parallel, there is another independent paradigm that has existed in the domain of blind equalization for the past two decades called “multi-stage clustering” based approach [75], which was found to outperform traditional paradigms like constant-modulus algorithm (CMA), MMA and many others as found in [76, 77]. More recently, a normalized phase-splitting variant of the multi-stage clustering paradigm was proposed in [78], which significantly accelerates the convergence of the multi-stage clustering based approach. However, all the algorithms derived till date for the improved multi-stage clustering (IMSC) paradigm are designed for linear channels and are unsuitable for LED nonlinearity affected scenarios; and thus calls for complex receiver architectures.

In this chapter, the scope of unsupervised learning based post-distortion is explored with relevance to VLC channels. For blind post-distortion over VLC channels affected by LED nonlinearity, an unsupervised normalized Hammerstein post-distortion structure for the improved multi-stage clustering based approach is proposed. This approach is termed as normalized Hammerstein improved multi-stage clustering equalizer (HIMSC). Various properties of the normalized HIMSC is highlighted which reinforces its suitability for VLC system as compared to the existing *blind* MCMMA-Volterra based approach. Additionally, a bound is derived for step-size over which the normalized HIMSC converges using classical adaptive filter theory based techniques. Further, an explicit expression for steady-state MSE floor as a function of step-size is derived in this chapter for normalized HIMSC. Simulation carried over standard IEEE 802.15 PAN VLC channels indicate that indeed normalized HIMSC delivers better MSE and BER performance as compared to the IMSC or the MCMMA-Volterra algorithm with much lower computational complexity. Furthermore, the derived expression for steady-state MSE is validated via simulations carried over IEEE 802.15 PAN VLC channels.

In this chapter, the following terminology is used: scalar at time k is represented by $(\cdot)_k$ and vector at time k is represented by boldface with subscript k such as \mathbf{x}_k (which are elements of \mathbb{C}^K , K being the row dimension of the vector, and \mathbb{C} denotes the field

of complex numbers). Real part of a complex quantity is denoted by the superscript $(\cdot)^R$ and $\Re\{\cdot\}$, and the imaginary part is denoted by $(\cdot)^I$ and $\Im\{\cdot\}$. The transpose operation is denoted by $(\cdot)^T$.

This chapter is organized as follows: Section-6.2 provides the system model provided used in this work, Section-6.3 gives the proposed unsupervised Hammerstein post-distorter, Section-6.4 presents the simulations to validate the proposed approach, Section-6.5 summarizes this chapter.

6.2 System model

In this section, the system model considered in this chapter is described. Let s_k denote the input constellation (with DC bias so as to be placed in the forward-bias of LED) at the k^{th} time instant. It is passed through an FIR filter $\{h_i\}_{i=0}^{L-1}$, of order $L - 1$. This channel is modeled by the convolution of the lowpass impulse response of LED (which is modeled by a lowpass filter with cutoff frequency 20MHz) given in [79] and the IEEE 802.15 PAN VLC channel. It is to be noted that the lowpass response of the LED models its finite modulation bandwidth which greatly affects the system throughput. The received symbol at k^{th} time instant, x_k , is given as follows:

$$x_k = \sum_{i=0}^{L-1} h_i f(s_{k-i}) + n_k \quad (6.1)$$

where $f(\cdot)$ is a Rapp LED nonlinearity [19] which had been used to model a white LED, and n_k is i.i.d (independent identically distributed) AWGN with variance σ_n^2 . The factor p controls the level of nonlinearity. For example, $p = 0.5$ has been considered to be a significant nonlinear characteristic in [57] and therefore the same has been considered throughout this chapter. The block diagram of the system model is given in Fig. 6.1. The channel consists of a linear FIR filter $\mathbf{h} = [h_0, h_1, \dots, h_{L-1}]$ (which models the IEEE 802.15 PAN VLC indoor channel [24, 58]). The equalizer considered in this chapter is a nonlinear adaptive equalizer with observation \mathbf{x}_k as input. \hat{s}_{k-D} is the equalizer estimate delayed by D samples, where D is the cumulative lag of the equalizer, such that $K = 2D + 1$. The

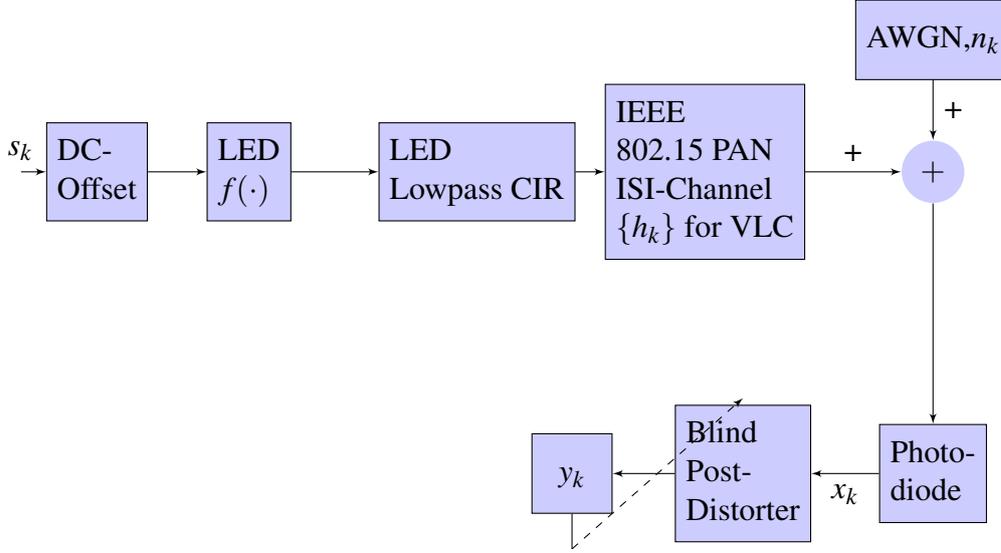


Figure 6.1: Block diagram of the used system model.

output of the blind post-distorter is denoted by y_k .

6.3 Proposed normalized Hammerstein improved multi-stage clustering based post-distorter

In this section, the normalized HIMSC is formulated for unsupervised post-distortion in nonlinear LED affected VLC IEEE 802.15 channels. The following transformation on the regressor \mathbf{x}_k is defined as follows as per the Hammerstein expansion,

$$\begin{aligned} \phi(\mathbf{x}_k) = & [x_k, x_{k-1}, \dots, x_{k-2D}, \\ & x_k^2, x_{k-1}^2, \dots, x_{k-2D}^2, \dots, x_k^M, x_{k-1}^M, \dots, x_{k-2D}^M] \end{aligned} \quad (6.2)$$

where M is the order of the Hammerstein expansion. Let the equalizer weight at the k^{th} instant be denoted by $\Omega_k = \Omega_k^R + j\Omega_k^I$ which is of same dimension as $\phi(\cdot)$. Let us define,

$$y_k^R = \Re(\Omega_k^T \phi(\mathbf{x}_k)), \quad \text{and} \quad y_k^I = \Im(\Omega_k^T \phi(\mathbf{x}_k)) \quad (6.3)$$

Using these expressions, the following cost function for the real and imaginary parts is formed:

$$J_{\text{HIMSC}}^R = \frac{1}{Q} \sum_{j=1}^Q \exp\left(-\frac{(y_k^R - \mu_j^R)^2}{\rho}\right) \quad (6.4)$$

$$J_{\text{HIMSC}}^I = \frac{1}{Q} \sum_{j=1}^Q \exp\left(-\frac{(y_k^I - \mu_j^I)^2}{\rho}\right)$$

where $\{\mu_j^R\}_{j=1}^Q$ and $\{\mu_j^I\}_{j=1}^Q$ are the possible real and imaginary parts of the transmitted constellation, ρ is the spread parameter, and Q is the number of moduli used for channel equalization in a given stage of clustering. The cost-function in (6.4) is optimized by using a hierarchical clustering approach. For example, for 16-QAM, first initial convergence is achieved using classification of the constellation in the four-quadrants, i.e., μ_j^R and μ_j^I will take values $\{+2, -2\}$. The second stage acts like a soft-decision-directed phase, and then μ_j^R and μ_j^I can be allowed to take values $\{+1, -1, +3, -3\}$. Some characteristics of the proposed normalized HIMSC cost function are: a) The multi-stage clustering cost function is actually the a-posteriori probability density function (p.d.f) of y_k parameterized by Ω_k (assuming equally likely transmit-constellation symbols) [75], and b) The multi-stage clustering criterion can be viewed as the sum of correntropy between the filtered output and the points of the transmit constellation. Hence, the proposed normalized HIMSC cost function is a better similarity measure as compared to deviation from fixed moduli as in CMMA or MCMMA due to incorporation of higher order statistics of the error term. As mathematically proven in the literature [80], correntropy converges to a lower MSE as compared to a simple squared deviation criterion due to incorporation of higher order statistics. Hence, intuitively, the second order fixed modulus based classical Busgang cost functions like MMA are not sufficient statistics for nonlinear channel equalization.

The cost function in (6.4) is optimized by a stochastic gradient algorithm with a nor-

malized descent direction which is updated as:

$$\Omega_{k+1}^R = \Omega_k^R + \mu e_{\text{HIMSC}}^R \frac{\phi(\mathbf{x}_k)}{\|\phi(\mathbf{x}_k)\|_2^2} \quad (6.5)$$

$$\Omega_{k+1}^I = \Omega_k^I + \mu e_{\text{HIMSC}}^I \frac{\phi(\mathbf{x}_k)}{\|\phi(\mathbf{x}_k)\|_2^2} \quad (6.6)$$

where,

$$e_{\text{HIMSC}}^R = \sum_{j=1}^Q \exp\left(-\frac{(y_k^R - \mu_j^R)^2}{\rho}\right) (y_k^R - \mu_j^R) \quad (6.7)$$

$$e_{\text{HIMSC}}^I = \sum_{j=1}^Q \exp\left(-\frac{(y_k^I - \mu_j^I)^2}{\rho}\right) (y_k^I - \mu_j^I)$$

which as found in [78] (which has been derived for the linear channel case and can be extended to the VLC channels as well by considering the regressors as $\phi(\mathbf{x}_k)$) is the solution to the following optimization problems for the real and imaginary parts of the phase splitting equalizer:

$$\begin{aligned} & \underset{\Omega_k^R, \Omega_k^I}{\text{minimize}} && \frac{1}{2} \|\Omega_{k+1}^R - \Omega_k^R\|_2^2 + \frac{1}{2} \|\Omega_{k+1}^I - \Omega_k^I\|_2^2 \\ & \text{subject to} && J_{\text{HIMSC}}^R = (1 - \varepsilon), J_{\text{HIMSC}}^I = (1 - \varepsilon) \\ & && \varepsilon \rightarrow 0 \end{aligned}$$

6.3.1 Derivation of step-size range for convergence

In this section, an upper bound on the step-size is derived so as to ensure convergence of the proposed normalized HIMSC. Let us denote the a-priori deviation from optimal weight $\Omega^o = \Omega^{R(o)} + j\Omega^{I(o)}$ as $\tilde{\Omega}_k = \Omega_k - \Omega^o$ and a-posteriori deviation from optimal weights $\tilde{\Omega}_{k+1} = \Omega_{k+1} - \Omega^o$. Additionally, let us denote the real and imaginary modulus as μ_*^R and μ_*^I , such that $\mu_*^R = \Omega^{R(o)T} \phi(\mathbf{x}_k)$ and $\mu_*^I = \Omega^{I(o)T} \phi(\mathbf{x}_k)$. Further, it is noted that

J_{HIMSC} can be approximated as follows:

$$\begin{aligned} J_{\text{HIMSC}}^R &\approx \hat{J}_{\text{HIMSC}}^R = \frac{1}{Q} \exp\left(-\frac{(y_k^R - \mu_*^R)^2}{\rho}\right) \\ J_{\text{HIMSC}}^I &\approx \hat{J}_{\text{HIMSC}}^I = \frac{1}{Q} \exp\left(-\frac{(y_k^I - \mu_*^I)^2}{\rho}\right) \end{aligned} \quad (6.8)$$

where $(\hat{\cdot})$ operator denotes the approximation of the HIMSC cost-function to the nearest constellation point. In other words, the multi-modal mixture in (6.4) is being approximated by the most dominant component. Let the output of the adaptive filter be denoted as $\tilde{y}_k = \tilde{\Omega}_k^T \phi(\mathbf{x}_k)$ and $\tilde{y}_{k+1} = \tilde{\Omega}_{k+1}^T \phi(\mathbf{x}_k)$. From (6.5), one can write as follows;

$$\tilde{y}_{k+1} = \tilde{y}_k - \mu e_{\text{HIMSC}} \quad (6.9)$$

Squaring and taking expectation on both sides, one arrives at:

$$\begin{aligned} \mathbb{E}[|\tilde{y}_{k+1}|^2] &\approx \mathbb{E}[|\tilde{y}_k|^2] + \mu^2 \mathbb{E}[|\hat{e}_{\text{HIMSC}}|^2] - \\ &2\mu \mathbb{E}[(\hat{e}_{\text{HIMSC}}^R \tilde{y}_k^R + \hat{e}_{\text{HIMSC}}^I \tilde{y}_k^I)] \end{aligned} \quad (6.10)$$

and,

$$\begin{aligned} \hat{e}_{\text{HIMSC}}^R &= \frac{1}{Q} \exp\left(-\frac{(y_k^R - \mu_*^R)^2}{\rho}\right) (y_k^R - \mu_*^R), \\ \hat{e}_{\text{HIMSC}}^I &= \frac{1}{Q} \exp\left(-\frac{(y_k^I - \mu_*^I)^2}{\rho}\right) (y_k^I - \mu_*^I), \end{aligned} \quad (6.11)$$

Consequently, one arrives at the following step-size range for convergence from (6.10) (i.e. to ensure that $\mathbb{E}[|\tilde{y}_{k+1}|^2] \leq \mathbb{E}[|\tilde{y}_k|^2]$),

$$0 \leq \mu \leq \frac{2\mathbb{E}[\hat{e}_{\text{HIMSC}}^R \tilde{y}_k^R + \hat{e}_{\text{HIMSC}}^I \tilde{y}_k^I]}{\mathbb{E}[|\hat{e}_{\text{HIMSC}}|^2]} \quad (6.12)$$

Let $\rho = 2\sigma^2$, $\mathbb{E}[|\tilde{y}_\infty|^2] = S$, and $f(\tilde{y}_k + n_k) = \hat{e}_{\text{HIMSC}}$. Noting that $f(x) = \exp(-\frac{x^2}{2\sigma^2})x$, one can arrive at the following expressions under the assumption of Gaussian noise with

6.3. PROPOSED NORMALIZED HAMMERSTEIN IMPROVED
MULTI-STAGE CLUSTERING BASED POST-DISTORTER

variance σ_n^2 as per integrals given in [80]:

$$\begin{aligned}\mathbb{E}[|\hat{e}_{\text{HIMSC}}|^2] &= \mathbb{E}[f(\tilde{y}_k + n_k)^2] = \frac{2\sigma^3(S + \sigma_n^2)}{(2S + 2\sigma_n^2 + \sigma^2)^{\frac{3}{2}}} \\ \mathbb{E}[\hat{e}_{\text{HIMSC}}^R \tilde{y}_k^R + \hat{e}_{\text{HIMSC}}^I \tilde{y}_k^I] &= \frac{2\sigma^3 S}{(S + \sigma_n^2 + \sigma^2)^{\frac{3}{2}}}\end{aligned}\quad (6.13)$$

Under the assumption of high signal to noise ratio, one gets the following bound for the step-size range,

$$0 \leq \mu \leq 2 \left[\frac{2S + \sigma^2}{S + \sigma^2} \right]^{\frac{3}{2}} \quad (6.14)$$

The following special cases of the expression derived in (6.14) can be derived based on value-ranges of S and σ^2 :

- For $S \gg \sigma^2$, (6.14) can be approximated as $0 \leq \mu \leq 2 \times 2^{\frac{3}{2}}$. This is intuitive as our bound can be relaxed to achieve a higher targeted S .
- For $S \ll \sigma^2$, (6.14) can be approximated as, $0 \leq \mu \leq 2$. Thus, one has a more strict bound to achieve a lower targeted S .
- For $S \approx \sigma^2$, (6.14) can be approximated as, $0 \leq \mu \leq 2 \times 1.5^{\frac{3}{2}}$.

Thus one can conclude from above three points that the step-size range required for convergence becomes more stringent as one reduces the targeted excess mean square error floor S .

For adaptive equalizers, steady-state misadjustment is an important learning metric which, in turn, is influenced by choice of step-size. In the next section, the impact of varying the step-size on the steady-state misadjustment of the proposed normalized HIMSC is studied.

6.3.2 Steady-state MSE vs step-size

Assuming convergence in (6.10), one arrives at:

$$\mu^2 \frac{S + \sigma_n^2}{(2S + 2\sigma_n^2 + \sigma^2)^{\frac{3}{2}}} = 2\mu \frac{S}{(S + \sigma_n^2 + \sigma^2)^{\frac{3}{2}}} \quad (6.15)$$

Assuming small S and σ_n^2 at high signal to noise ratio (SNR), $S + \sigma_n^2 \ll \sigma^2$ and $\sigma_n^2 \ll S$ (6.15) can be approximated as:

$$\frac{\mu}{2} \left(1 + \frac{S + \sigma_n^2}{\sigma^2}\right) \left(1 - \frac{2S + 2\sigma_n^2}{\sigma^2}\right) = \left(1 - \frac{2}{3} \frac{\sigma_n^2}{S + \sigma_n^2}\right) \quad (6.16)$$

Simplifying further and neglecting fourth order error terms, one arrives at:

$$\frac{\mu}{2} \left(1 - \frac{S + \sigma_n^2}{\sigma^2}\right) \approx 1 - \frac{2}{3} \frac{\sigma_n^2}{S} \quad (6.17)$$

Simplifying and assuming $S^2 \rightarrow 0$:

$$S \approx \frac{4\sigma^2\sigma_n^2}{\left(1 - \frac{\mu}{2}\right)6\sigma^2 + 3\mu\sigma_n^2} \quad (6.18)$$

Now, an interesting property of the proposed approach is demonstrated. In (6.18), if $\sigma_n^2 \ll \sigma^2$, (6.18) can be reduced as:

$$S \approx \frac{4\sigma_n^2}{6 - 3\mu} \quad (6.19)$$

It is interesting to note that at high SNR, the steady-state MSE of the proposed normalized HIMSC is independent of the spread parameter σ^2 . This means that small tweaks in parameter values of σ^2 does not affect the performance of the proposed normalized HIMSC at high SNR. One can also infer from (6.18), that as μ is increased, the steady state mean squared error also increases, which is intuitive from classical adaptive filtering theory [40].

6.4 Simulations

In this section, simulations are presented to validate the normalized HIMSC with respect to MCMMA based Volterra post-distorter. The Hammerstein expansion was chosen with $D = 4$ and corresponding step-size were chosen so that they learn the LED characteristics at the same rate of the MSE learning curve. For normalized HIMSC, $\mu = 0.25$ was used; while for Volterra-MCMMA, the $\mu = 0.0003$ was chosen. The rms delay spread of the overall channel was calculated to be 19.94ns and 18.75ns respectively for “office” and “office with cubicles” scenarios respectively. The transmit bit-rate was assumed to be 400Mbps.

For Volterra-MCMMA, 45-linear taps and 25-second order taps as used in [8] was chosen. It is observed in Fig. 6.2 and Fig. 6.3 that both in the open office and office with cubicles IEEE 802.15 PAN VLC channels, the normalized HIMSC outperforms Volterra-MCMMA and converges to a lower MSE. The MSE curves are compared at 33dB signal SNR. Also from Figs. 6.4 and 6.5, it can be concluded that at high SNR, IMSC and normalized HIMSC demonstrate superior BER performance as compared to Volterra-MCMMA. This indicates that amongst unsupervised approaches, the normalized HIMSC is a superior paradigm to learn the post-distorter from the observations itself without the need of training symbols.

Furthermore, the steady-state MSE values obtained via simulations for various step-sizes are validated by the theoretical expression derived in (6.18). It could be observed from Fig. 6.6 and Fig. 6.7 that the theoretically derived expression in (6.18) closely matches the simulated MSE floors in the high SNR regime. Also, it is observed that at high SNR (as the SNR is changed from 40dB to 36dB) there is not much change in simulated value of steady-state MSE floor as a function of step-size which further validates the theory for steady-state MSE derived in this chapter. In other words, at high SNR, the same σ produces similar MSE characteristics.

The proposed normalized HIMSC is much computationally simpler as compared to the blind MCMMA based Volterra equalizer in [8]. The unsupervised algorithm in [8],

needs $O(8(45 + 25^2) + 2)$ operations as there are 45 linear taps and 25 second order taps in [8] and computational complexity of a simple Widrow-Hopf rule is $O(8G + 2)$ real operations [40], where G is the dimensionality of the regressors. Whereas for comparison, for simulating the proposed normalized HIMSC, the use of $D = 4$ and $M = 5$ leads to $O(10(45) + 2)$ real operations (that is, $O(10G + 2)$ for normalized Widrow-Hopf rule, G in our case is 45). This indicates that the proposed normalized HIMSC is computationally less complex as compared to the Volterra-MCMMMA in [8] and provides computational savings by a factor of 11.86 coupled with superior BER performance shown in Fig. 6.4 and Fig. 6.5.

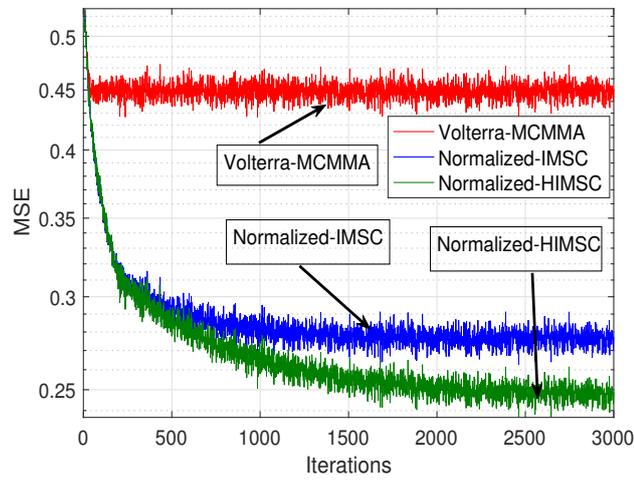


Figure 6.2: MSE comparison for open office IEEE 802.15 PAN channel for 16-QAM.

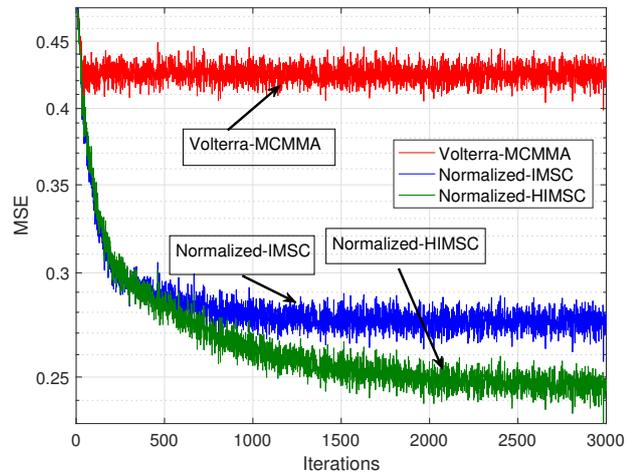


Figure 6.3: MSE comparison for office with cubicles IEEE 802.15 PAN channel for 16-QAM.

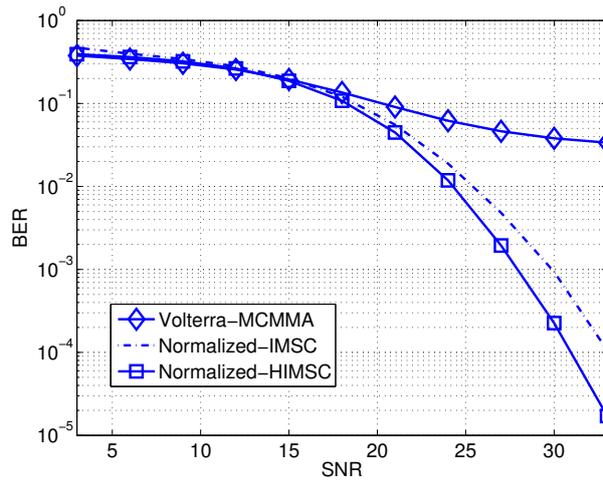


Figure 6.4: BER comparison for open office over IEEE 802.15 PAN channel for 16-QAM.

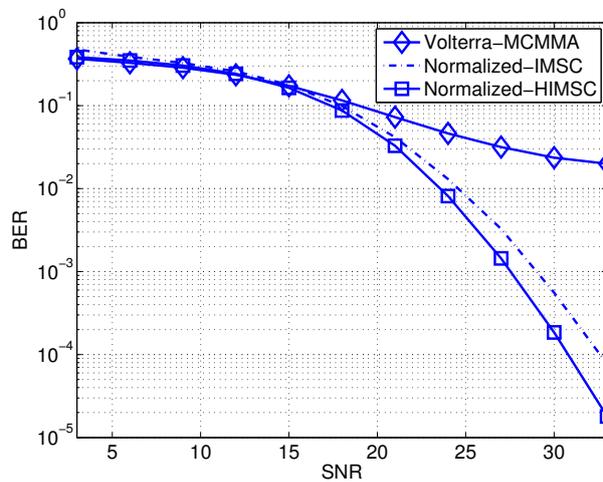


Figure 6.5: BER comparison for office with cubicles over IEEE 802.15 PAN channel for 16-QAM.

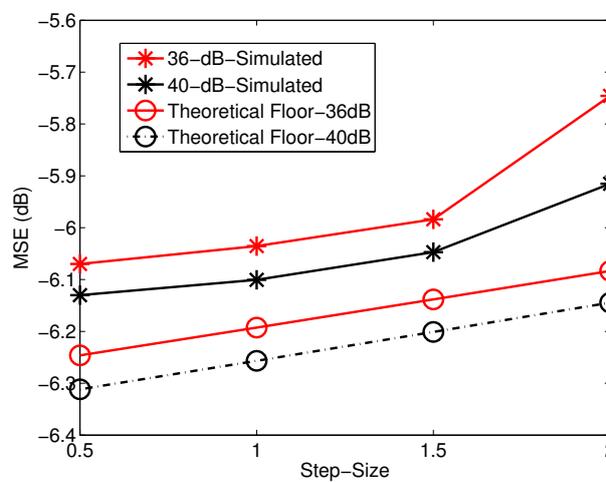


Figure 6.6: Steady-state MSE vs step-size for open office over IEEE 802.15 PAN channel for 16-QAM.

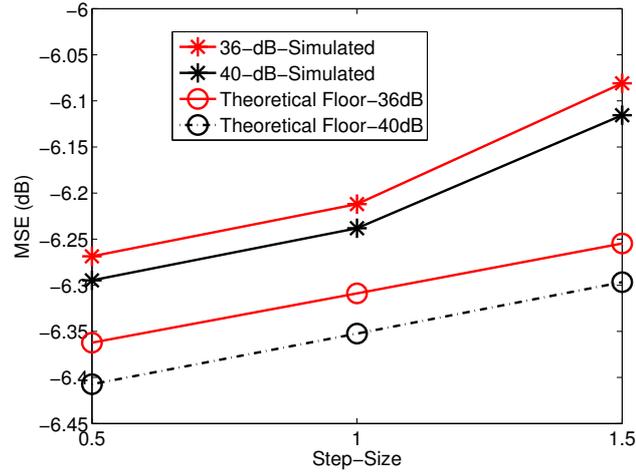


Figure 6.7: Steady-state MSE vs step-size for office with cubicles over IEEE 802.15 PAN channel for 16-QAM.

6.5 Summary

In this chapter, the problem of post-distortion over dispersive VLC channels was re-considered. However, the focus of this chapter was limited to unsupervised approaches as opposed to supervised approaches in previous chapters. Additionally, the effect of finite modulation bandwidth of the LED was considered in the system model. A novel unsupervised nonlinear post-distorter based on multi-stage clustering paradigm called normalized HIMSC was proposed and validated against existing Volterra-MCMMA over IEEE 802.15 PAN VLC channels. A suitable step-size range was theoretically derived to guarantee convergence of the proposed adaptive post-distorter. Superior MSE curves and BER performance as compared to Volterra-MCMMA was observed with lower computational complexity, thus validating its suitability as a blind post-distorter for VLC. Furthermore, theoretical expressions for steady-state MSE was derived analytically and validated by simulations over IEEE 802.15 PAN VLC channels, which reinforces correctness of the theory derived in this chapter. These results also indicate that the multi-stage clustering based nonlinear filtering paradigm can be a preferred alternative as compared to MMA/CMA and other blind post-distorters for VLC.

Chapter 7

Precoded Chebyshev-NLMS based pre-distorter for nonlinear LED compensation in NOMA-VLC

7.1 Introduction

To meet the ever increasing demand of users in VLC for 5G systems by 2020, a novel multiple access scheme called NOMA has attracted much attention recently [81, 82]. In this multiple access scheme, instead of allocation of orthogonal dimensions to each user (be it code, frequency, time), the signals of users are overlapped (superimposed) over the same dimension with power diversity, and the signals of users are separated out at each UE by a SIC. Recently, NOMA has been found to be viable for adoption into VLC mainly due to the following reasons [33]: a) NOMA can handle small number of users which typically happens in a Li-Fi attocell (a small femtocell) [83], b) channel is generally dominated by a line of sight (LOS) path which facilitates for accurate channel estimation, and c) adjusting tuning angles and field of view (FOV) gives additional degrees of freedom for multiplexing multi-user signals which can induce differential channel gains facilitating for power diversity.

Among existing major techniques for NOMA, the work in [34] suggest choices of

precoding matrices which rely on the channel conditions to be different. For example, the work in maximal ratio combining (MRC) post-processing [34] would work well if the channel matrices are well conditioned/ the columns of the channel matrices of all users (with all users having different QoS requirements) are all independent/dissimilar to ensure power-diversity. However, if the channel matrices are aligned/similar, the MRC based processing would render the same effective channel conditions at each UE thereby culminating in a failure of the NOMA system. While the work in [84] is related to the problem addressed in this chapter, however it assumes a single antenna for each UE in its formulation. In MIMO-VLC, however, one has an array of photodiodes at each UE thus restricting the validity of the approach in [84] (which essentially decomposes the MIMO detection problem into several spaced multiple-input single-output (MISO) problems and losing diversity gain as given in [84]). In fact, the work in [84] is the “one-antenna per UE”-analog of the problem handled in this chapter. Moreover, the impact of device impairments like nonlinearity is not considered in a MIMO-VLC scenario in a NOMA setting. The precoding technique in [35] holds even when channel conditions for UEs are similar, however is applicable to only two-user scenario (additionally it promises a diversity gain for one of the users) and leaves its possible extension to arbitrary number of users as an open problem. The works in [34, 35] propose a new paradigm for NOMA power allocation called “cognitive-radio inspired power allocation”. In this paradigm, the user with a better channel condition is considered as a secondary-user and the user with worse channel condition is treated as a primary-user. When a new user wants to access the link, it is served opportunistically under the condition that the existing users’ QoS requirements are maintained. This policy applies in many 5G scenarios like IoT where users have diverse QoS requirements and not necessarily diverse channel conditions [35]. To accommodate multiplicity of users in such scenarios, it is crucial to have diverse channel conditions which can be either in-built [34], or achieved by precoding [35] (though this precoding, as mentioned before, was done for only two users). In this chapter, techniques for cognitive-radio inspired NOMA (which arises typically in 5G scenarios like internet of things [35]) are explored wherein channels of all users exhibit significant correlation

coupled with device-impairments.

Despite employing NOMA-VLC, the capacity of VLC systems is limited by inherent LED nonlinearity [39], which can be mitigated by using pre-distortion or post-distortion techniques. Among pre-distortion techniques, the simplest technique would be to maintain a lookup table of the estimated nonlinearity. However, such static lookup tables lose their practicability due to varying LED characteristics [27] (due to device aging); hence there is a need for adaptive pre-distorters. Such adaptive pre-distorters have been suggested in the literature which are learnt using NLMS algorithm [27], and Chebyshev regression using NLMS approach [85]. This chapter proposes a modified Chebyshev-NLMS based pre-distortion with hybrid eigen-decomposition based precoding in a MIMO-VLC scenario for IoT applications and demonstrates the superiority of Chebyshev regression based pre-distortion approach.

Additionally, to the best of authors' knowledge, all studies so far in a VLC-NOMA system [33, 86, 87] do not consider LED nonlinearity in a MIMO scenario for the design of a pre-distorter. In view of the existing NOMA literature in VLC, the major contributions of this chapter are summarized as follows:

- A new hybrid precoding technique for NOMA-VLC channels is suggested in a closed loop adaptive Chebyshev pre-distorter, based on singular value decomposition (SVD), in IoT applications. The proposed approach works even when the left and the right eigenvectors of all the available channel matrices for all users are correlated. In other words, if the channels have correlated CIR, one cannot design different precoding vectors for users by the framework of analysis given in [34]. For two users, a QR decomposition technique based approach in [35] is able to deliver varying grades of service even in scenarios with similar channel conditions. However, the precoding technique presented in this chapter holds for arbitrary number of users which is a challenge raised in [35].
- A novel power allocation technique is found for the proposed precoding scheme based on QoS requirements of various users. While finding the optimal power allocation coefficients, the SVD based precoding technique given in this chapter is

considered in a scenario when the channel matrices of all the users have similar CIR with each user being opportunistically served whilst maintaining the QoS requirements of existing users. In such conditions as given in [35], the users have diverse QoS requirements as opposed to diverse channel conditions. Simulations indicate that the proposed power allocation technique performs well for square M -QAM modulation schemes in the above mentioned scenario for varying number of users.

- In order to validate the proposed NOMA precoding scheme via simulations, generalized formulae for BER vs SNR for varying number of users is derived and validated by simulations for square M -QAM. The simulations indicate that the theoretically derived BER formulae indeed match the simulated BER curves for varying number of users, which further validates the framework of analysis presented in this chapter.

In this chapter, the following terminology is adopted: scalars at time k are denoted by subscript k such as x_k , vectors (which are tuples of scalars), are denoted as $\{x_k\}$ and matrices are denoted by capital boldface such as \mathbf{H} . Transpose of matrices/vectors are denoted by $(\cdot)^T$. Additionally, inverse of transpose of a matrix is denoted by $(\cdot)^{-T}$ and the pseudo-inverse is denoted by $(\cdot)^\dagger$. Sets are denoted by $\langle \cdot \rangle$ in this chapter.

This chapter is organized as follows: Section-7.2 discusses the MIMO-NOMA system model, and existing QR precoded linear NOMA is reviewed in Section-7.3. Section-7.4, presents proof of the feasibility of the proposed precoded Chebyshev-pre-distortion for nonlinear LED affected VLC-NOMA, and Section-7.5 suggests the choice of suitable precoding matrix. Section-7.6 presents the power allocation strategy for the given system. Expression for BER performance for square M -QAM is derived in Section-7.7. The simulation results are described in Section-7.8. Finally, the chapter is summarized in Section-7.9.

7.2 MIMO-NOMA system model

In this section, the MIMO-NOMA system model considered in this chapter is presented and the terminology followed throughout the chapter is introduced. The input vector at time instant k is denoted as, $\mathbf{x}_k = \{x_k\}_{k=zM_T+1}^{(z+1)M_T}$, where z is an arbitrary integer that denotes the sample duration, and M_T is the number of LEDs at the transmitter. The symbol $\mathbf{x}'_k = \{x'_k\}$ indicates precoding of $\{x_k\}$ by a matrix \mathbf{P} . In case different precoding matrices are used for each user (which has been done previously in [88, 89], although in a different setting), the precoding matrix for the u^{th} user is denoted as $\mathbf{P}^{(u)}$. The proposed system model is given in Fig. 7.1. This input vector is actually a non-orthogonal superposition of many users' signals (called a mixed-constellation in this chapter), with suitable power allocation. Mathematically, this can be written as:

$$\{x_k\} = \sum_{u=1}^U \sqrt{P^{(u)}} \{s_k^{(u)}\} \quad (7.1)$$

where U is the number of users and u is the u^{th} user index variable. $P^{(u)}$ is the power allocated to each user and $\{s_k^{(u)}\}$ is the u^{th} user's constellation at k^{th} time instant. If $\{s_k^{(u)}\}$ is a bipolar constellation, then it should be offset by a DC value so that the entire constellation lies in the forward bias regime of the LED. Without loss of generality, to impose a constant power constraint, it is assumed that $\sum_{u=1}^U P^{(u)} = 1$. This is followed by pre-distorter mapping $T(\cdot)$. The considered channel matrix for the u^{th} user is denoted by $\mathbf{H}^{(u)} \in \mathbb{R}^{M_R \times M_T}$ (where M_R is the number of receive photodiodes in the photodiode array for each user) and consequently the i.i.d AWGN is added, and the superposition of signals is broadcasted from the transmitter LED array. For the u^{th} user, the received signal vector $\mathbf{y}_k^{(u)} = \{y_k^{(u)}\}$ can be written as:

$$\{y_k^{(u)}\} = \mathbf{H}^{(u)} A(T(\{x'_k\})) + \{n_k\} \quad (7.2)$$

where $\mathbf{n}_k = \{n_k\}$ denotes AWGN with zero mean and covariance matrix $\sigma_n^2 \mathbf{I}$ (\mathbf{I} being the identity matrix). $T(\cdot)$ denotes the adaptive pre-distorter transformation which can be

learnt by popular techniques like NLMS [27] or Chebyshev regression based NLMS [85]. At each UE, $\{s_k^{(u)}\}$ is then recovered by SIC. The LED nonlinearity $A(\cdot)$ in [19] is a Rapp LED nonlinearity.

In the case of imperfect channel state information (CSI) (both transmit and receive) considered in this chapter, one can express the following measure of deviation in terms of actual elements of channel matrix $\mathbf{H} = [\mathbf{h}_g]$, and $\hat{\mathbf{h}}_g$, which is the g^{th} column of channel matrix affected by estimation error:

$$\|\mathbf{h}_g - \hat{\mathbf{h}}_g\|_2 \leq \gamma \quad (7.3)$$

where γ -neighborhood is assumed to be Gaussian distributed with variance σ_γ^2 as given in [90]. $\|\cdot\|_2$ denotes the Euclidean norm.

In this paragraph, every considered block of the system model in Fig. 7.1 is described. In the first block the mixed-constellation $\{x_k\}$ is taken, precoded by the matrix $\mathbf{P}^{(u)}$ corresponding to User u , and passed through the pre-distorter $T(\cdot) = \sum_{\forall i} r_k^{(i)} T_i(\cdot)$, where T_i denotes the i^{th} Chebyshev polynomial and $r_k^{(i)}$ denotes the pre-distorter weights. This precoded transmission is passed to the transmitter LED array, where the nonlinearity $A(\cdot)$ is (implicitly) applied. Then the AWGN at the k^{th} instant, n_k , is added. At the receiver, for each user u , the precoded transmission affected by nonlinearity is received by the u^{th} photodiode array. Then, according to the QoS of the user, an estimate of the mixed signal $\{\hat{l}_k^{(u)}\}$ for each user u is recovered by multiplying with $(\mathbf{H}^{(u)} \mathbf{P}^{(u)})^\dagger$ to form an estimate of the mixed constellation $\{x_k\}$ at each UE, given by $\{\hat{l}_k^{(u)}\}$. After this, the SIC is performed at each UE to recover the user's symbols. It is to be noted that in the simulations, errors induced at each layer of SIC are taken into account. The \hat{l}_k , which is chosen among the $\{\hat{l}_k^{(u)}\}$ from UEs with the highest QoS and the input signal is used in the feedback loop (which can be an RF uplink [13, 91] or PLC-based [23, 92]) to adjust the pre-distorter coefficients $r_k^{(i)}$ using the NLMS or the Chebyshev-NLMS algorithms.

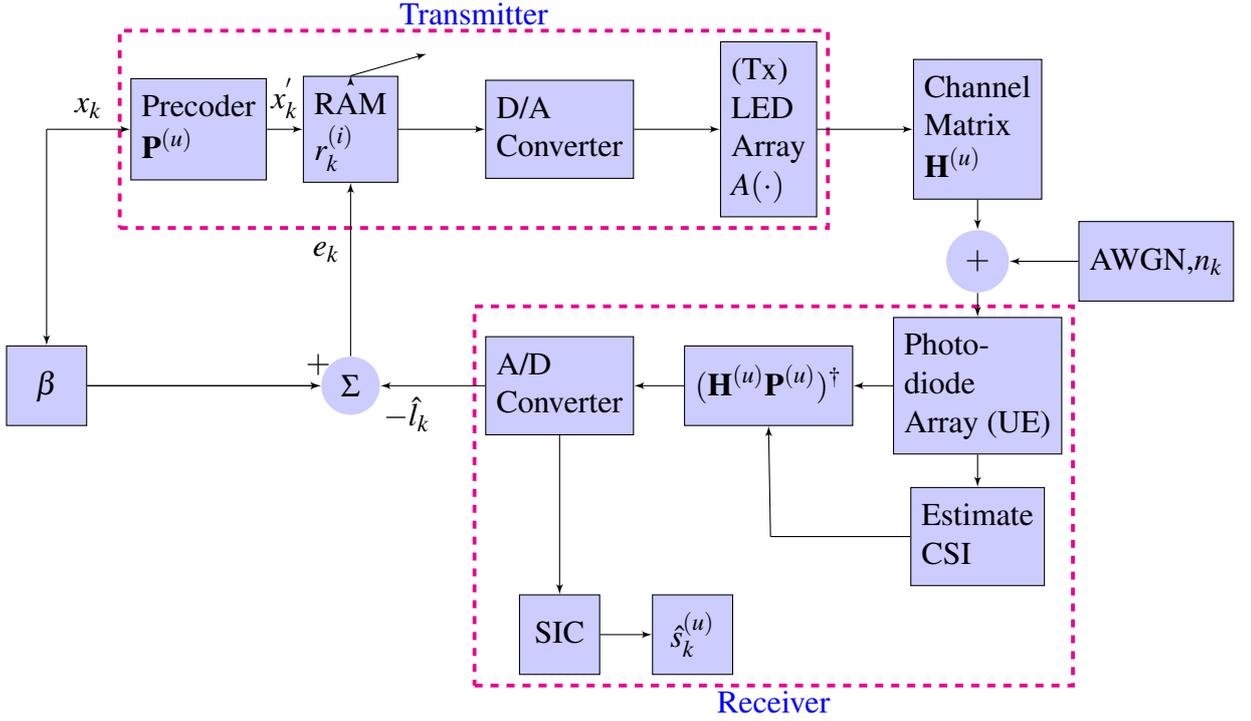


Figure 7.1: Block diagram of the proposed system model.

7.3 Review of QR-precoded NOMA

In this section, a technique for NOMA is reviewed which considers NOMA for a special two-user scenario, even when channels of the users exhibit significant correlation. This problem is generalized in this chapter to arbitrary number of users and hence it will be insightful to review the work given in [35] prior to the proposed algorithm. In the work in [35], a two user scenario is considered which experiences similar channels, $\mathbf{H}^{(u)}$, with $u \in \{1, 2\}$ but with varying QoS requirements.

$$\{x_k\} = \sum_{u=1}^2 \sqrt{P^{(u)}} \{s_k^{(u)}\} \quad (7.4)$$

The idea in [35] is to select a user (say User 2), whose experience one would like to improve selectively, and consider the QR decomposition of the transpose of the corresponding channel such that $\mathbf{H}^{(2)T} = \mathbf{Q}^{(2)}\mathbf{R}^{(2)}$. Here, $\mathbf{Q}^{(2)} \in \mathbb{R}^{M_R \times M_R}$ is a unitary matrix and $\mathbf{R}^{(2)} \in \mathbb{R}^{M_R \times M_T}$ is an upper triangular matrix obtained by QR-decomposition. Consequently, the broadcast \mathbf{x}_k is precoded by $\mathbf{P}^{(u)} = \mathbf{Q}^{(2)}$, $u = 1, 2$ to form \mathbf{x}'_k . In other words,

$\mathbf{x}'_k = \mathbf{Q}^{(2)}\mathbf{x}$. Then User 2's signals are recovered by successive interference cancellation and User 2 experiences a diversity gain.

On the other hand, User 1's signal is recovered by the zero forcing solution upon \mathbf{x}'_k by the matrix $(\mathbf{H}^{(1)}\mathbf{Q}^{(2)})^\dagger$. This step actually induces different channel condition at UE with lesser QoS due to multiplication by the Wishart matrix corresponding to $(\mathbf{H}^{(1)}\mathbf{Q}^{(2)})^\dagger$. In other words, the multiplication of $(\mathbf{H}^{(1)}\mathbf{Q}^{(2)})^\dagger$ the equivalent channel condition at User 1 is degraded as compared to that of User 2, to the extent controlled by the power allocation, which is done so that User 1's QoS are met while improving the performance of User 2. This is achieved by minimizing the probability of rate-outage of User 1 as given in [35].

7.4 Proposed extension of Chebyshev pre-distortion to MIMO systems

For a single UE scenario, the Chebyshev polynomial based pre-distortion was proposed in Chapter 2 for SISO systems. However, this chapter considers a multi-user MIMO VLC scenario with NOMA as the multiple-access technique. In this chapter, it is assumed that all UEs are equipped with a photodiode array and hence the received signal for all users are vectors. In order to learn the pre-distorter weights in a MIMO scenario, the modified cost function, $J_{\text{MIMO}} = \min_{r_k^{(i)}} \mathbb{E}[(\|\{\beta x_k\} - \{\hat{l}_k\}\|_2^2)]$, involving Euclidean norm, is minimized with respect to the weights $\{r_k^{(i)}\}$. J_{MIMO} could also be written as:

$$J_{\text{MIMO}} = \sum_{\forall x_k \in \{x_k\}, \hat{l}_k \in \{\hat{l}_k\}} (\beta x_k - \hat{l}_k)^2 \quad (7.5)$$

To clarify the terminology in the previous equation, a double summation is not denoted in above equation. A single summation, after element-wise subtraction and squaring of elements of the vectors, is denoted. As the Euclidean norm is convex in the Chebyshev coefficient weights, stochastic gradient descent based adaptation is guaranteed to converge to the global optimum of the cost function.

The estimate \hat{l}_k is derived from a feedback uplink from the receiver (which can be

achieved by time-division duplexing, frequency division duplexing or by a separate RF uplink [13, 91]). Each of the U UEs sends an estimate $\hat{l}_k^{(u)}$ among which \hat{l}_k is chosen from the link with the highest QoS as it will have the maximum SINR. The pre-distorter coefficients $r_k^{(i)}$ are updated by taking gradient of J_{MIMO} with respect to $r_k^{(i)}$ via a similar stochastic gradient NLMS based approach as:

$$r_{k+1}^{(i)} = r_k^{(i)} + \sum_{\forall x_k \in \{x_k\}, \hat{l}_k \in \{\hat{l}_k\}} \frac{\eta}{\sum_{\forall i} T_i(x_k)^2} e_k T_i(x_k) \quad (7.6)$$

for all, x_k in $\{x_k\}$. $r_k^{(i)}$ denotes the i^{th} pre-distorter weight at k^{th} time instant and $e_k = \beta x_k - \hat{l}_k$ for each x_k in $\{x_k\}$.

The Chebyshev weight-estimation inverts nonlinearity of the LED. However, to incorporate power-diversity in the framework, precoding techniques are required. This raises important questions regarding the overall system like: a) how does Chebyshev pre-distortion perform when precoding is employed? Are the signals detected successfully?; b) if so, then at what signal to noise ratio (this is important as this would influence the rate which in turn would affect the power-allocation technique)? These questions are answered in the next paragraph using a modified Busgang theorem. In further sections, more elaboration on suitable precoding matrices and derive the power allocation strategy for the considered NOMA-MIMO scenario over VLC channels, is provided.

In this paragraph, the feasibility of a hybrid Chebyshev pre-distorter that is coupled with precoding is proven to extend its suitability to NOMA channels. The Chebyshev pre-distortion in [85] is modified by adding a precoder $\mathbf{P}^{(u)}$ at the transmitter for each user u (in addition to the Chebyshev pre-distorter which mitigates the LED nonlinearity), and the mixed constellation $\hat{l}_k^{(u)}$ at the receiver $\{x_k\}$ is recovered by multiplying by $(\mathbf{H}^{(u)}\mathbf{P}^{(u)})^\dagger$. In the next paragraph, it is shown that the recovery of symbols is made possible by using this pre-distorter in the presence of LED nonlinearity. The output $\hat{l}_k^{(u)}$ can be written as

follows:

$$\{\hat{l}_k^{(u)}\} = (\mathbf{H}^{(u)}\mathbf{P}^{(u)})^\dagger \quad (7.7)$$

$$[\mathbf{H}^{(u)}A(\{\sum_{\forall i} r_k^{(i)} T_i(\mathbf{P}^{(u)}\{x_k\})\}) + \{n_k\}]$$

From extension of Bussgang's theorem [93], one can write

$$\{\hat{l}_k\}^{(u)} = \{\alpha x_k\} + (\mathbf{H}^{(u)}\mathbf{P}^{(u)})^\dagger \mathbf{H}^{(u)}\{\delta\} + (\mathbf{H}^{(u)}\mathbf{P}^{(u)})^\dagger \{n_k\} \quad (7.8)$$

where δ is zero mean uncorrelated noise sequence with variance σ_δ^2 and α is a scaling correlation factor between $A(T(\cdot))$ with its argument. $\mathbf{H}^{(u)\dagger}$ is the pseudo inverse of a matrix $\mathbf{H}^{(u)}$ and is assumed to be a left-inverse of \mathbf{H} . Hence, $\hat{l}_k^{(u)}$ lies in the same subspace of x_k . Hence any precoding as done in the case of the linear scenario in [35], works for the nonlinear closed loop system model as well.

In order to study the impact of nonlinearity on the overall detection performance, let us define the following terms: $\mathbf{Z}^{(u)} = (\mathbf{H}^{(u)}\mathbf{P}^{(u)})^\dagger$ and $\mathbf{W}^{(u)} = (\mathbf{H}^{(u)}\mathbf{P}^{(u)})^\dagger \mathbf{H}^{(u)}$. Thus, from (7.8) the u^{th} user is decoded with the following SINR (denoted by $\Gamma^{(u)}$),

$$\Gamma^{(u)} = \frac{\alpha^2 P^{(u)}}{\left(\alpha^2 \sum_{\forall b>u} P^{(b)} + \text{Tr}(\mathbf{W}^{(u)T} \mathbf{W}^{(u)}) \sigma_\delta^2 + \text{Tr}(\mathbf{Z}^{(u)T} \mathbf{Z}^{(u)}) \sigma_v^2 \right)} \quad (7.9)$$

, where $\text{Tr}(\cdot)$ denotes the trace. Upon convergence of Chebyshev coefficients to their optimal values (ensured by NLMS algorithm), $\alpha \rightarrow 1$ and $\sigma_\delta^2 \rightarrow 0$. Then, the SINR for the u^{th} user is as follows:

$$\Gamma^{(u)} = \frac{P^{(u)}}{\sum_{\forall b>u} P^{(b)} + \text{Tr}(\mathbf{Z}^{(u)T} \mathbf{Z}^{(u)}) \sigma_v^2} \quad (7.10)$$

Please note that in the above expression, the transmit constellation of each user is assumed to be normalized to unit power, and b is the index variable denoting the layer of SIC.

7.5 Choice of precoding matrices

In this section, the technique of choosing precoding matrices is described in detail to induce a power-diversity over the considered multi-user VLC scenario. In (7.10), the SINR experienced by the u^{th} user depends on the pseudo-inverse of the overall channel matrix $\mathbf{H}^{(u)}\mathbf{P}^{(u)}$. It can be noted that the following salient points to motivate the proposed precoding technique:

- The channel matrix is ill-conditioned in VLC-MIMO systems [94, 95]. This makes some of the eigenvalues of the channel-matrix very small as compared to other eigenvalues of the channel matrix.
- Traditional beamforming as in [88] may not work as some of the channel matrices have correlated eigenvectors, thereby reducing the available degrees of freedom to accommodate more users (a phenomenon which is more pronounced in VLC due to the ill conditioned channel matrix).

The ill-conditioned channel matrix is actually a blessing in disguise, as this helps us in inserting many levels of QoS for each user by the precoding technique described below. To design a precoding matrix $\mathbf{P}^{(u)}$, let us consider the SVD of the u^{th} users' channel matrix $\mathbf{H}^{(u)}$ as follows:

$$\mathbf{H}^{(u)} = \mathbf{U}^{(u)}\mathbf{\Sigma}^{(u)}\mathbf{V}^{(u)T} \quad (7.11)$$

where

$$\mathbf{\Sigma}^{(u)} = \text{diag}(\sigma^{(1,(u))}, \sigma^{(2,(u))}, \dots, \sigma^{(c^{(u)},{(u)})}, 0, 0, 0\dots) \quad (7.12)$$

$c^{(u)}$ denotes the rank of the u^{th} user's matrix. The matrices $\mathbf{U}^{(u)}$ and $\mathbf{V}^{(u)}$ consists of eigenvectors of $\mathbf{H}^{(u)}\mathbf{H}^{(u)T}$ and $\mathbf{H}^{(u)T}\mathbf{H}^{(u)}$. Thus the precoding matrix $\mathbf{P}^{(u)}$ for each user

would be given by:

$$\mathbf{P}^{(u)} = \mathbf{V}^{(u)} \boldsymbol{\Sigma}^{(u)\lambda^{(u)}-1} \quad (7.13)$$

where $\lambda^{(u)}$ is the exponent assigned to u^{th} user and

$$\begin{aligned} \boldsymbol{\Sigma}^{(u)\lambda^{(u)}-1} = \text{diag}(\sigma^{(1,(u))\lambda^{(u)}-1}, \sigma^{(2,(u))\lambda^{(u)}-1}, \dots, \\ \sigma^{(c^{(u)},(u))\lambda^{(u)}-1}, 0, 0, 0\dots) \end{aligned}$$

Upon such precoding, each user would experience a virtual ‘‘parallel’’ channel given by the following equation (assuming $\alpha \rightarrow 1$ and $\sigma_{\delta}^2 \rightarrow 0$):

$$\{\hat{l}_k\}^{(u)} \approx \{\alpha x_k\} + (\mathbf{U}^{(u)} \boldsymbol{\Sigma}^{(u)\lambda^{(u)}})^{\dagger} \{n_k\} \quad (7.14)$$

Thus each user would experience a channel $\mathbf{U}^{(u)} \boldsymbol{\Sigma}^{(u)\lambda^{(u)}}$ with a distinctive trace and hence inducing diverse channel conditions with proper choice of $\lambda^{(u)}$. Further, as in [35], one may want to improve the condition number of one of the users’ (say u_1) channel with a higher QoS by QR factorization technique as done in [35], i.e. one may assign $\mathbf{P}^{(u)} = \mathbf{V}^{(u)} \boldsymbol{\Sigma}^{(u)\lambda^{(u)}-1} \mathbf{Q}^{(u_1)}$, where $\mathbf{U}^{(u_1)} \boldsymbol{\Sigma}^{(u_1)\lambda^{(u_1)}} = \mathbf{R}^{(u_1)T} \mathbf{Q}^{(u_1)T}$. However, this QR factorization technique is not a pressing necessity as the generalized power diversity is achieved by varying $\lambda^{(u)}$. The exact technique of finding $\lambda^{(u)}$ for the u^{th} user according to varying levels of user-QoS is given in the next section.

7.6 Power allocation strategy

In this section, the power allocation strategy that needs to fulfill a given QoS for a given user, is derived. QoS for the u^{th} user is typically designated by a reliable transmission rate $R^{(u)}$. This can be written mathematically as:

$$\log_2(1 + \Gamma^{(u)}) \geq R^{(u)} \quad (7.15)$$

Let us define $\varepsilon^{(u)} = 2^{R^{(u)}} - 1$. Thus,

$$\frac{P^{(u)}}{\sum_{b>u} P^{(b)} + \text{Tr}(\mathbf{Z}^{(u)T} \mathbf{Z}^{(u)}) \sigma_v^2} \geq \varepsilon^{(u)} \quad (7.16)$$

Rearranging terms one gets,

$$P^{(u)} \geq \varepsilon^{(u)} \sum_{b>u} P^{(b)} + \varepsilon^{(u)} \text{Tr}(\mathbf{Z}^{(u)T} \mathbf{Z}^{(u)}) \sigma_v^2 \quad (7.17)$$

adding $\sum_{b>u} P^{(b)}$ both sides,

$$\begin{aligned} 1 &\geq P^{(u)} + \sum_{b>u} P^{(b)} \geq (1 + \varepsilon^{(u)}) \sum_{b>u} P^{(b)} + \\ &\quad \varepsilon^{(u)} \text{Tr}(\mathbf{Z}^{(u)T} \mathbf{Z}^{(u)}) \sigma_v^2 \\ \implies \sum_{b>u} P^{(b)} &\leq \frac{1 - \varepsilon^{(u)} \text{Tr}(\mathbf{Z}^{(u)T} \mathbf{Z}^{(u)}) \sigma_v^2}{1 + \varepsilon^{(u)}} \end{aligned} \quad (7.18)$$

Assuming $1 - P^{(u)} \geq 0$, one arrives at the following power allocation for each user:

$$P^{(u)} = \min \left(1, \varepsilon^{(u)} \frac{(1 + \text{Tr}(\mathbf{Z}^{(u)T} \mathbf{Z}^{(u)}) \sigma_v^2)}{1 + \varepsilon^{(u)}} \right) \quad (7.19)$$

It is to be noted that this power allocation strategy is derived by considering different users' differing QoS requirements with each user having a specific $\mathbf{Z}^{(u)}$ depending on its requirement $\varepsilon^{(u)}$. Thus this power-allocation technique is a better and a more generalized power allocation technique as compared to the gain ratio power allocation in [33] for the considered MIMO NOMA-VLC channel considering correlated channels.

7.6.1 Choice of $\lambda^{(u)}$ for each user

In NOMA it is very important to induce power diversity, which is provided by precoding. One can simplify the expression for the residual noise power from (7.14) for each user, the following expression is arrived at:

$$\text{Tr}(\mathbf{Z}^{(u)T} \mathbf{Z}^{(u)}) \sigma_v^2 = \text{Tr}(\mathbf{\Sigma}^{(u)-\lambda^{(u)T}} \mathbf{\Sigma}^{(u)-\lambda^{(u)}}) \sigma_v^2 \quad (7.20)$$

For power diversity, i.e. for different channel-quality at various UE,

$$\begin{aligned}
 & \log \frac{|\mathbf{Z}^{(u)T} \mathbf{Z}^{(u)}| \sigma_v^2}{|\mathbf{Z}^{(u+1)T} \mathbf{Z}^{(u+1)}| \sigma_v^2} \geq 0 & (7.21) \\
 \implies & \frac{1}{\lambda^{(u)}} \log \left(\sum_{g=1}^{c^{(u+1)}} |\sigma^{(g,(u+1))}|^2 \right) \\
 & \geq \frac{1}{\lambda^{(u+1)}} \log \left(\sum_{g=1}^{c^{(u)}} |\sigma^{(g,(u))}|^2 \right) \\
 \implies & \lambda^{(u+1)} \geq \lambda^{(u)} \frac{\log \left(\sum_{g=1}^{c^{(u)}} |\sigma^{(g,(u))}|^2 \right)}{\log \left(\sum_{g=1}^{c^{(u+1)}} |\sigma^{(g,(u+1))}|^2 \right)}
 \end{aligned}$$

Hence this gives us a recursive rule for choosing $\lambda^{(u)}$ for each user by initializing $\lambda^{(1)} = 1$. The lower-bound for $\lambda^{(u+1)}$ in (7.21) is considered so as to improve the overall condition number of the channel matrix.

7.7 Analytical expression for BER of square M -QAM

In this section, an analytical expression for bit error rate of M -QAM is derived for the proposed precoding algorithm. This analytical expression is necessary to predict the system performance without computationally intensive Monte-Carlo simulations and for overall calibration of the wireless link for link-optimization [96]. The case of square M -QAM is chosen in this chapter because of its widespread use in the VLC literature [50, 97, 98].

Let us denote a square M -QAM constellation with amplitude and phase set being given by $\langle A_n \rangle_{n=1}^{\sqrt{M}}, \langle \phi_m \rangle_{m=1}^{\sqrt{M}}$ such that M -QAM constellation is expressed as $\langle A_n \exp(j\phi_m) \rangle_{n,m=1}^{\sqrt{M}}$. Hence, at each layer, one can have the following recursion for the

respective set of moduli $\mu_{b,m,n}$ for \hat{l}_k (using laws of vector addition):

$$\begin{aligned} \mu_{1,(m,n)} &= \sqrt{P^{(1)}|A_m|^2 + P^{(2)}|A_n|^2 + 2\sqrt{P^{(1)}P^{(2)}}A_m^*A_n \cos(\phi_n)} \\ \mu_{b+1,(m,n)} &= \sqrt{P^{(b)}|A_m|^2 + |\mu_{b,(m,n)}|^2 + 2\sqrt{P^{(b)}}\mu_{b,(m,n)}A_m^* \cos(\phi_n)} \\ \forall m, n &= 1, 2, \dots, \sqrt{M} \end{aligned} \quad (7.22)$$

Finally the probability of bit-error $P_{\sqrt{M}}$ for an equivalent \sqrt{M} -PAM modulation scheme could be written as (i.e. the BER of square M -QAM modulation can be assumed as two independent \sqrt{M} -PAM modulation [99]):

$$\begin{aligned} P_{\sqrt{M}} &= 2 \left(\frac{\sqrt{M}-1}{\sqrt{M}} \right) \sum_{\forall b} \sum_{\forall m} \sum_{\forall n} \\ &Q \left(\sqrt{\frac{|\mu_{b+1,(m,n)} - \mu_{b,(m,n)}|^2}{\text{Tr}[\mathbf{Z}^{(u)T} \mathbf{Z}^{(u)} \boldsymbol{\sigma}_v^2]}} \right) \end{aligned} \quad (7.23)$$

Finally, the final expression for square M -QAM can be given by the following equation by considering the real and imaginary parts of the M -QAM constellation as independent:

$$P_M = 1 - (1 - P_{\sqrt{M}})^2 \quad (7.24)$$

Thus (7.24) gives us an expression for probability of error for square M -QAM for the proposed SVD based precoding technique in a NOMA-VLC scenario. For the special case of 4-QAM, one arrives at the following expression for BER

$$\begin{aligned} P_{\text{QAM}_{\sqrt{M}=2}} &= \sum_{\forall b} \sum_{m=1}^{\sqrt{M}} \sum_{n=1}^{\sqrt{M}} \\ &Q \left(\sqrt{\frac{|\mu_{b+1,(m,n)} - \mu_{b,(m,n)}|^2}{\text{Tr}[\mathbf{Z}^{(u)T} \mathbf{Z}^{(u)} \boldsymbol{\sigma}_v^2]}} \right) \end{aligned} \quad (7.25)$$

and,

$$P_{\text{QAM}_{\sqrt{M}=2}} = 1 - (1 - P_{\text{QAM}_{\sqrt{M}=2}})^2 \approx 2P_{\text{QAM}_{\sqrt{M}=2}} \quad (7.26)$$

where the relation between μ_{b+1} and μ_b can be written as follows (considering the amplitude set to be ± 1 and phase set to be $\pm \frac{\pi}{4}$):

$$\mu_{b+1} = \sqrt{P^{(b)} + |\mu_b|^2 + \sqrt{2P^{(b)}}\mu_b} \quad (7.27)$$

It could be readily seen that the above equation is the special case of (7.23) assuming the amplitude set to be having only the moduli $\{\pm 1\}$ and phase $\pm \frac{\pi}{4}$ as encountered for 4-QAM modulation.

7.8 Simulations

A typical channel matrix was generated mathematically as given in [33], for a room of size $5\text{m} \times 5\text{m} \times 3\text{m}$, with refractive index of lens 1.5, height of LED-2.25m, area of the photo-diode (PD) 1cm^2 , spacing between transmitter LED array and receiver photodiode array 0.4m, (X,Y) coordinates of PDs corresponding to User 1's photodiode array are (0.2,0.2) and (0.2,-0.2). (X,Y) coordinates of User 2's PDs in its photodiode array are (-0.2,0.2) and (-0.2,-0.2). FOV of the photodetectors is kept fixed at 60 degrees. A 2×2 MIMO channel was considered for each user. All users were chosen close to each other such that all experience almost similar channel conditions. For *power allocation* for each user, (7.19) was used in all simulations. These simulation parameters are tabulated in Table 7.1. For all the simulations 10^6 symbols were considered with ensemble of 3000 Monte-Carlo runs.

The VLC quasi-static [44] deterministic channel model used for simulating the chan-

nel matrix is given by the following equation [33, 95]:

$$\mathbf{H}^{(u)} = [h_{ij}]^{(u)} = \frac{A_{e,i}}{d_{ij}^2 \sin^2 \Psi} R(\phi_{ij}) \cos \theta_{ij}, \quad 0 < \phi_{ij} < \Psi \quad (7.28)$$

$$= 0, \quad \text{otherwise}$$

$A_{e,i}$ denotes the area of the i^{th} photodetector. d_{ij} is the distance between i^{th} transmitter LED and j^{th} photodetector for each UE, ϕ_{ij} is the perpendicular angle of j^{th} LED, θ_{ij} is the angle between i^{th} transmit LED in transmitter-array and j^{th} photodetector in a photodetector array with the receiver axis. Ψ denotes the FOV for each photodetector. $R(\phi_{ij})$ denotes the Lambertian radiant intensity which can be written as follows:

$$R(\phi_{ij}) = \frac{(\kappa + 1) \cos^\kappa(\phi_{ij})}{2\pi} \quad (7.29)$$

κ is the order of Lambertian emission given as follows:

$$\kappa = -\frac{\ln 2}{\ln(\cos(\phi_{\frac{1}{2}}))} \quad (7.30)$$

Typically, as given in [94, 95], these channel-matrices are inherently ill-conditioned and exhibit correlatedness when users are located near to each other.

Table 7.1: Simulation Parameters

Room-Size	5m × 5m × 3m
LED-Height	2.25m
Photodetector-Area	1cm ²
LED array spacing	0.4m
LED Emission Half-Angle $\phi_{\frac{1}{2}}$	70 degrees
η	0.00022
User 1's PD array coordinates	(0.2, 0.2, -3), (0.2, -0.2, -3)
User 2's PD array coordinates	(-0.2, 0.2, -3), (-0.2, -0.2, -3)
User 3's PD array coordinates	(-0.7, 0.7, -3), (-0.7, -0.7, -3)
User 4's PD array coordinates	(0.7, 0.7, -3), (0.7, -0.7, -3)
FOV, Ψ	60 degrees

Using the given simulation setup, the simulations that validates the proposed SVD-precoded

Chebyshev-NLMS based pre-distorter are described. Two scenarios are considered: a) the two user mixed constellation scenario in which the Chebyshev pre-distortion is compared with NLMS in conjunction with the proposed precoding to highlight gains received from Chebyshev pre-distortion as compared to NLMS pre-distortion, and b) the U -user scenario in which sum-rate and the average BER is studied by varying the number of users.

7.8.1 Mixed constellation for two user scenario

In all the simulations considered in this sub-section for the two-user mixed-constellation scenario, User 1 is assumed to use a 16-QAM constellation while User 2 is assumed to have 4-QAM modulation. In Fig. 7.2, performance of the NLMS and the proposed SVD-precoded Chebyshev NLMS based approach is compared in full-CSI scenario (which means that $\sigma_\gamma^2 = 0$ or in other words, perfect knowledge of channel at the transmitter and the receiver). It is found that the proposed SVD precoded Chebyshev-NLMS pre-distorter exhibits superior BER vs SNR characteristics as compared to NLMS pre-distortion. Additionally, from Fig. 7.3 and Fig. 7.4, it is clear that the proposed SVD precoded Chebyshev-NLMS algorithm has superior convergence characteristics with respect to NLMS pre-distortion in a MIMO VLC-NOMA scenario in full CSI scenario. Performance gains of several decades of BER are observed for the proposed SVD-precoded Chebyshev-NLMS over simple NLMS. These simulations indicate that SVD-precoded Chebyshev pre-distortion is a better solution for NOMA-VLC system as compared to NLMS algorithm.

7.8.2 U -user scenario

Using the proposed power allocation scheme derived in (7.19), simulations are performed by varying the number of users as monitoring the trend of sum-rate and the average BER for the proposed SVD-precoded Chebyshev-NLMS for 4-QAM. From Fig. 7.5, it is observed that initially the sum-rate increases as the number of users is increased from one to two. As the number of users is increased to three, there is a slight dip in the sum-rate, although the sum rate obtained in this case is still greater than the scenario for $U = 1$.

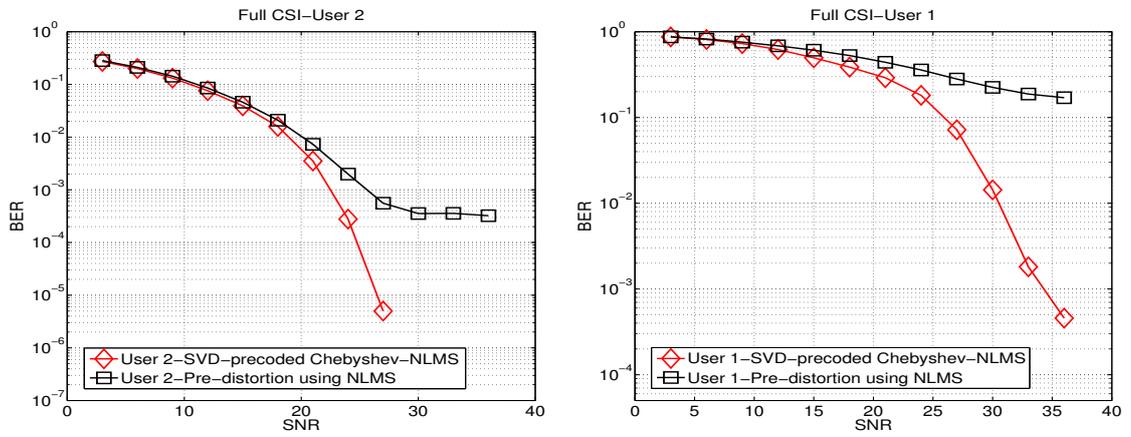


Figure 7.2: 2-user scenario with full CSI. One 4-QAM, another user 16-QAM. Chebyshev pre-distortion outperforms NLMS in SVD-precoded scenario which is designed keeping QoS requirements in consideration. Performance is shown for User-2 and User-1.

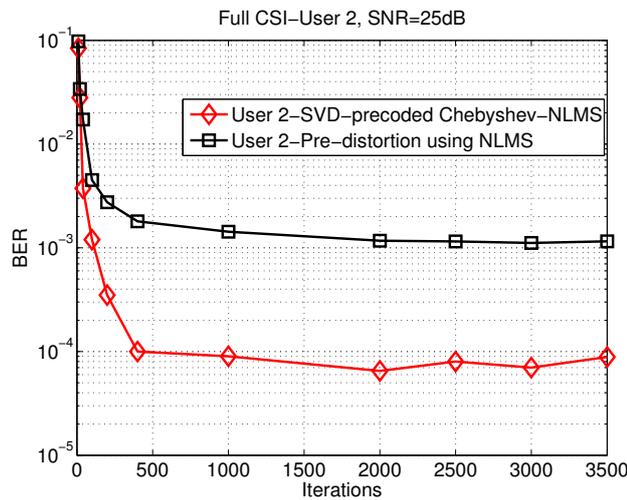


Figure 7.3: Convergence comparison for proposed SVD pre-coded Chebyshev pre-distorter and SVD pre-coded NLMS pre-distorter for User-2 at 25dB SNR in a two user scenario.

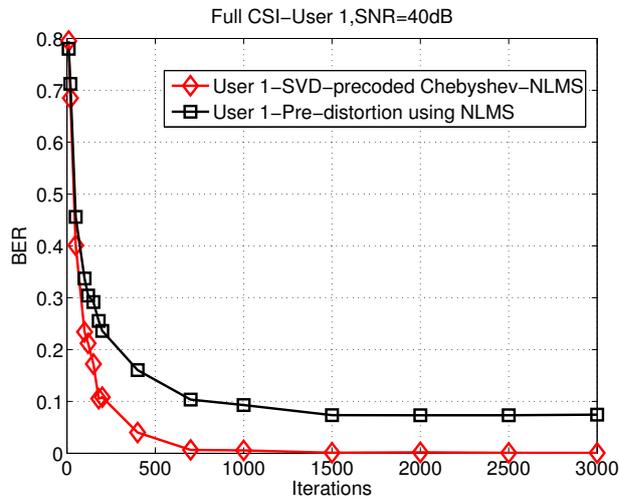


Figure 7.4: Convergence comparison for proposed SVD precoded Chebyshev pre-distorter and SVD precoded NLMS pre-distorter for User-1 at 40dB SNR in a two user scenario.

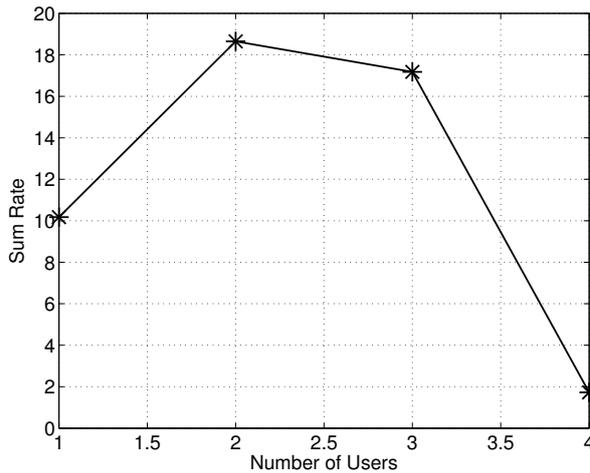


Figure 7.5: Sum-Rate plot for N -user scenario. After 3-users the sum-rate decays drastically. Specially, for the N -user scenario, all users are assigned 4-QAM.

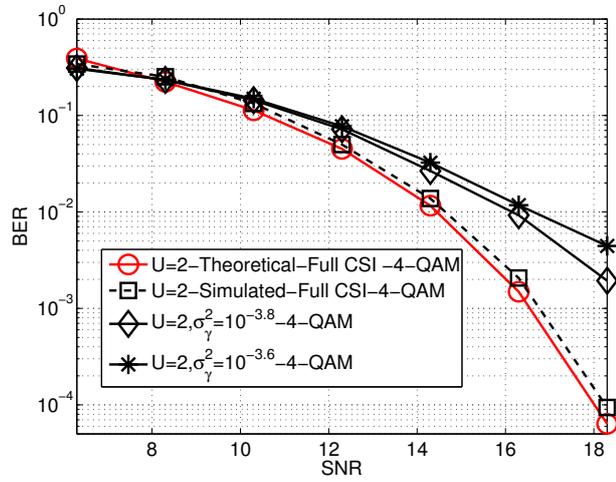


Figure 7.6: Average BER plotted as a function of SNR for $U = 2$ for the proposed pre-coding technique. All users use 4-QAM modulation scheme.

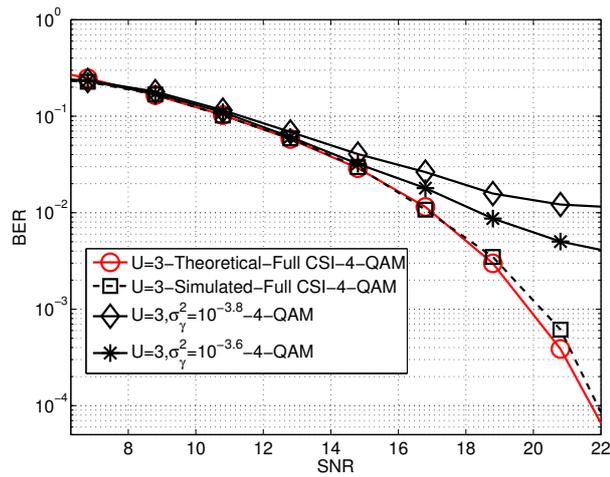


Figure 7.7: Average BER plotted as a function of SNR for $U = 3$ for the proposed pre-coding technique. All users use 4-QAM modulation scheme.

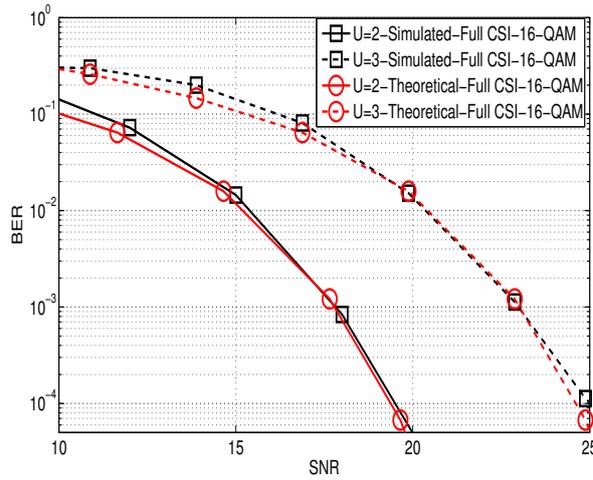


Figure 7.8: Average BER plotted as a function of SNR for $U = 2, 3$ for the proposed precoding technique. All users use 16-QAM modulation scheme in full CSI scenario.

As one more user is added ($U = 4$), the sum rate decreases and drastically degrades the overall system performance. It is assumed that all users are served according to their QoS.

Additionally, the average BER vs SNR is plotted for $U = 2, 3$ in Fig. 7.6 and Fig. 7.7. It can be observed that till 3 users, the overall BER performance of the system is quite acceptable. However, as the number of users is increased to $U = 4$, the BER becomes totally unacceptable (and hence is not plotted). This further reinforces the observation in sum-rate plots that the number of users that can be accommodated is at maximum three in the considered scenario. Moreover, the derived expression for BER in (7.23) (denoted as “theoretical”) approximates the simulated BER curve (derived for $\sigma_\gamma^2 = 0$ and perfect knowledge of the channel scenario denoted as full CSI), which confirms validity of the derived theoretical expressions. Also, a marked loss of BER performance is observed upon small values in $\sigma_\gamma^2 = 10^{-3.8}, 10^{-3.6}$ [90] which can be attributed to ill-conditioned channel matrices which one typically encounters in VLC scenarios [95].

Furthermore, for $U = 2, 3$, the performance of the proposed SVD-precoded Chebyshev NLMS based technique is simulated for 16-QAM, and the BER vs SNR is plotted in Fig. 7.8 and compared to the theoretically derived formula in (7.23) for the full CSI scenario ($\sigma_\gamma^2 = 0$). It is observed that the theoretical BER curves match the simulated BER curves very closely for varying number of users $U = 2, 3$. This validates the generalized framework of analysis for higher order QAM modulation schemes proposed in this

chapter.

7.9 Summary

In this chapter, the problem of pre-distortion was taken up in a multi-user scenario. For multiple-access, NOMA was considered as it is one of the major emerging multiple-access techniques for 4G/5G systems. A novel generalized hybrid precoding technique is proposed for IoT scenarios for arbitrary number of users with differing QoS requirements. This precoding technique is combined with Chebyshev pre-distortion for applicability in LED nonlinearity impaired scenarios. Performance gains were found with respect to NLMS pre-distortion in perfect CSI as well as imperfect CSI scenarios. A novel power-allocation algorithm is also proposed for the precoding algorithm derived in this chapter. The proposed precoding also works in situations in which there are ill-conditioned channel matrices and number of degrees of freedom in the channel matrix are limited to accommodate arbitrary number of users. This work could have applications in VLC as an integral part of IoT and Li-fi devices where the applications of NOMA are more prevalent as one converges towards 5G technologies.

Chapter 8

Conclusion and future work

This chapter provides conclusions and major insights gained from this thesis, and further outlines directions of future work.

8.1 Conclusion

In this thesis, novel signal processing techniques for mitigating LED nonlinearity and ISI (which are two major performance limiting factors in VLC), were proposed and validated. These techniques are required for delivering the promised throughput of VLC systems targeted for 5G systems. These techniques can be classified into two categories: a) pre-distortion techniques, and b) post-distortion (equalization) techniques.

In Chapter 3, in order to mitigate LED-nonlinearity, pre-distortion techniques were considered and a novel Chebyshev polynomial based pre-distorter was proposed in this work. This proposed choice of orthogonal polynomial based NLMS, instead of classical linear-NLMS based pre-distorter resulted in better BER performance. These desirable characteristics were justified by the whiteness and min-max error approximation of Chebyshev regression, which facilitate for better convergence characteristics. However, the system model in that work used some idealistic assumptions like a perfect uplink between transmitter and receiver, which would not hold good in practical scenarios.

To look into some practical VLC links with ISI, the problem of post-distortion was subsequently considered. Supervised post-distortion techniques for mitigating ISI and

LED-nonlinearity were proposed using the RKHS methods, which guarantee convex and universal representation for nonlinear optimization problems. Novel MSER techniques for dictionary sparsification were proposed using novelty-criterion (refer Chapter 4), and fixed budget criterion (in Chapter 5) to facilitate for computational simplicity. Mathematical analysis is carried for the MSE dynamics of the proposed post-distorter, both in the transient and the steady-state regime.

In addition, in Chapter 6, an unsupervised Hammerstein-based post-distorter based on a multi-stage clustering based paradigm is proposed which demonstrates faster convergence and better BER performance as compared to existing blind post-distorters based on MCMMA-criterion. The steady-state MSE behavior was mathematically analyzed and validated via simulations. Additionally, it was found to be computationally simpler as compared to existing unsupervised implementations of post-distorters available in the literature.

Finally, in Chapter 7, the system model given in Chapter-3 was extended to multi-user MIMO scenario. For multiple access, NOMA was considered as it is one of the emerging multiple-access techniques for 5G systems, particularly for VLC. Novel precoding and power-allocation techniques were derived which hold even when users do not encounter dissimilar channels. Theoretical expressions for BER for square M -QAM were derived and validated via simulations. A power allocation technique is derived for ill-conditioned channels in order to opportunistically satisfy the individual rate constraint based QoS. This happens to be one of the seminal investigations in NOMA-VLC in a MIMO-setting.

8.2 Future work

VLC is an emerging area of active research and all the algorithms developed in this work are novel approaches to improve throughput of the VLC link. The proposed algorithms and novel techniques proposed in this thesis leads to several open problems that could be explored.

Firstly, another direction of research could be deployment of a novel sparsification

criterion apart from novelty criterion/fixed budget criterion. Also, exact analysis MSE-dynamics of all the post-distorters in this work without Taylor-series approximations is an open problem.

Secondly, in NOMA-VLC, it will be interesting to investigate the performance of the proposed pre-distorter in ISI-MIMO fading channels, which may facilitate for better throughput of the overall VLC-link.

Lastly, a practical implementation of all algorithms proposed in this thesis needs to be made and fixed point analysis of all algorithms needs to be done in order to study the behavior of the proposed approaches under finite-precision arithmetic, and real time operations.

References

- [1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, “What will 5G be?” *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.
- [2] H. Haas, L. Yin, Y. Wang, and C. Chen, “What is LiFi?” *Journal of Lightwave Technology*, vol. 34, no. 6, pp. 1533–1544, 2016.
- [3] A. Jovicic, J. Li, and T. Richardson, “Visible light communication: opportunities, challenges and the path to market,” *IEEE Communications Magazine*, vol. 51, no. 12, pp. 26–32, 2013.
- [4] J. Roufs and F. Blommaert, “Temporal impulse and step responses of the human eye obtained psychophysically by means of a drift-correcting perturbation technique,” *Vision Research*, vol. 21, no. 8, pp. 1203–1221, 1981.
- [5] S. Ayub, S. Kariyawasam, M. Honary, and B. Honary, “A practical approach of VLC architecture for smart city,” in *Antennas and propagation conference (LAPC), 2013 Loughborough*. IEEE, 2013, pp. 106–111.
- [6] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, “Five disruptive technology directions for 5G,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, 2014.
- [7] Y. Wang, L. Tao, X. Huang, J. Shi, and N. Chi, “8-Gb/s RGBY LED-based WDM VLC system employing high-order CAP modulation and hybrid post equalizer,” *IEEE Photonics Journal*, vol. 7, no. 6, pp. 1–7, 2015.

- [8] ———, “Enhanced performance of a high-speed WDM CAP64 VLC system employing Volterra series-based nonlinear equalizer,” *IEEE Photonics Journal*, vol. 7, no. 3, pp. 1–7, 2015.
- [9] S. Zvanovec, P. Chvojka, P. A. Haigh, and Z. Ghassemlooy, “Visible light communications towards 5G,” *Radioengineering*, vol. 24, no. 1, pp. 1–9, 2015.
- [10] D. Karunatilaka, F. Zafar, V. Kalavally, and R. Parthiban, “LED based indoor visible light communications: State of the art,” *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1649–1678, 2015.
- [11] R. Zhang, J. Wang, Z. Wang, Z. Xu, C. Zhao, and L. Hanzo, “Visible light communications in heterogeneous networks: Paving the way for user-centric design,” *IEEE Wireless Communications*, vol. 22, no. 2, pp. 8–16, 2015.
- [12] S. Arnon, *Visible Light Communication*, 1st ed. Cambridge University Press, 2015.
- [13] H. Burchardt, N. Serafimovski, D. Tsonev, S. Videv, and H. Haas, “VLC: Beyond point-to-point communication,” *IEEE Communications Magazine*, vol. 52, no. 7, pp. 98–105, 2014.
- [14] I. Takai, T. Harada, M. Andoh, K. Yasutomi, K. Kagawa, and S. Kawahito, “Optical vehicle-to-vehicle communication system using LED transmitter and camera receiver,” *IEEE Photonics Journal*, vol. 6, no. 5, pp. 1–14, 2014.
- [15] S.-H. Yu, O. Shih, H.-M. Tsai, N. Wisitpongphan, and R. D. Roberts, “Smart automotive lighting for vehicle safety,” *IEEE Communications Magazine*, vol. 51, no. 12, pp. 50–59, 2013.
- [16] Z. Wang, C. Yu, W.-D. Zhong, J. Chen, and W. Chen, “Performance of a novel LED lamp arrangement to reduce SNR fluctuation for multi-user visible light communication systems,” *Optics Express*, vol. 20, no. 4, pp. 4564–4573, 2012.
- [17] R. Murai, T. Sakai, H. Kawano, Y. Matsukawa, Y. Kitano, Y. Honda, and K. C. Campbell, “A novel visible light communication system for enhanced control

- of autonomous delivery robots in a hospital,” in *System Integration (SII), 2012 IEEE/SICE International Symposium on*. IEEE, 2012, pp. 510–516.
- [18] Y.-F. Liu, Y. C. Chang, C.-W. Chow, and C.-H. Yeh, “Equalization and pre-distorted schemes for increasing data rate in in-door visible light communication system,” in *National Fiber Optic Engineers Conference*. Optical Society of America, 2011, p. JWA083.
- [19] H. Elgala, R. Mesleh, and H. Haas, “An LED model for intensity-modulated optical communication systems,” *IEEE Photonics Technology Letters*, vol. 22, no. 11, pp. 835–837, 2010.
- [20] R. Price, “A useful theorem for nonlinear devices having Gaussian inputs,” *IRE Transactions on Information Theory*, vol. 4, no. 2, pp. 69–72, 1958.
- [21] H. Qian, S. Yao, S. Cai, and T. Zhou, “Adaptive postdistortion for nonlinear LEDs in visible light communications,” *IEEE Photonics Journal*, vol. 6, no. 4, pp. 1–8, 2014.
- [22] K. Ying, Z. Yu, R. J. Baxley, H. Qian, G.-K. Chang, and G. T. Zhou, “Nonlinear distortion mitigation in visible light communications,” *IEEE Wireless Communications*, vol. 22, no. 2, pp. 36–45, 2015.
- [23] H. Elgala, R. Mesleh, and H. Haas, “Indoor optical wireless communication: potential and state-of-the-art,” *IEEE Communications Magazine*, vol. 49, no. 9, pp. 56–62, 2011.
- [24] F. Miramirkhani and M. Uysal, “Channel modeling and characterization for visible light communications,” *IEEE Photonics Journal*, vol. 7, no. 6, pp. 1–16, 2015.
- [25] S. S. Haykin, *Adaptive filter theory*. Pearson Education India, 2007.
- [26] V. Bhatia, “Distribution dependent adaptive learning,” Ph.D. dissertation, IDCOR, The University of Edinburgh, 2005.

- [27] J. K. Kim, K. Hyun, and S. K. Park, "Adaptive predistorter using NLMS algorithm for nonlinear compensation in visible-light communication system," *Electronics Letters*, vol. 50, no. 20, pp. 1457–1459, 2014.
- [28] G. Stepniak, J. Siuzdak, and P. Zwierko, "Compensation of a VLC phosphorescent white LED nonlinearity by means of Volterra DFE," *IEEE Photonics Technology Letters*, vol. 25, no. 16, pp. 1597–1600, 2013.
- [29] E. Kreyszig, *Introductory functional analysis with applications*, 1st ed. Wiley New York, 1989, vol. 81.
- [30] W. Liu, J. C. Principe, and S. Haykin, *Kernel adaptive filtering: a comprehensive introduction*, 1st ed. John Wiley & Sons, 2011, vol. 57.
- [31] W. Liu, P. P. Pokharel, and J. C. Principe, "The kernel least-mean-square algorithm," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 543–554, 2008.
- [32] B. Schölkopf, R. Herbrich, and A. J. Smola, "A generalized Representer theorem," in *Computational Learning Theory*. Springer, 2001, pp. 416–426.
- [33] H. Marshoud, V. M. Kapinas, G. K. Karagiannidis, and S. Muhaidat, "Non-orthogonal multiple access for visible light communications," *IEEE Photonics Technology Letters*, vol. 28, no. 1, pp. 51–54, Jan 2016.
- [34] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 537–552, 2016.
- [35] Z. Ding, L. Dai, and H. V. Poor, "MIMO-NOMA design for small packet transmission in the internet of things," *IEEE Access*, vol. 4, pp. 1393–1405, 2016.
- [36] H. Elgala, R. Mesleh, and H. Haas, "Predistortion in optical wireless transmission using OFDM," in *Hybrid Intelligent Systems, 2009. HIS'09. Ninth International Conference on*, vol. 2. IEEE, 2009, pp. 184–189.

- [37] M. Noshad and M. Brandt-Pearce, "Hadamard coded modulation for visible light communications," *arXiv preprint arXiv:1406.2897*, 2014.
- [38] Y. Yi, C. Li, and K. Lee, "Adaptive MMSE equalizer for optical multipath dispersion in indoor visible light communication," *IETE Journal of Research*, vol. 58, no. 5, pp. 347–355, 2012.
- [39] T. Lee, "The nonlinearity of double-heterostructure LED's for optical communications," *Proceedings of the IEEE*, vol. 65, no. 9, pp. 1408–1410, 1977.
- [40] A. H. Sayed, *Fundamentals of adaptive filtering*, 1st ed. John Wiley & Sons, 2003.
- [41] S. S. Haykin, *Adaptive filter theory*, 4th ed. Pearson Education India, 2008.
- [42] B. Farhang-Boroujeny, *Adaptive filters: theory and applications*, 2nd ed. John Wiley & Sons, 2013.
- [43] M. Abramowitz, I. A. Stegun *et al.*, "Handbook of mathematical functions," *Applied mathematics series*, vol. 55, p. 62, 1966.
- [44] Z. Ghassemlooy, W. Popoola, and S. Rajbhandari, *Optical wireless communications: system and channel modelling with Matlab®*, 1st ed. CRC Press, 2012.
- [45] N. Nguyen, J. Burken, and A. Ishihara, "Least-squares adaptive control using Chebyshev orthogonal polynomials," in *AIAA Infotech@ Aerospace Conference, AIAA-2011-1402*, 2011.
- [46] O. González, S. Rodríguez, R. Pérez-Jiménez, B. R. Mendoza, and A. Ayala, "Error analysis of the simulated impulse response on indoor wireless optical channels using a Monte Carlo-based ray-tracing algorithm," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 124–130, 2005.
- [47] K. Lee, H. Park, and J. R. Barry, "Indoor channel characteristics for visible light communications," *IEEE Communications Letters*, vol. 15, no. 2, pp. 217–219, 2011.

- [48] A. H. Azhar, T. Tran, and D. O'Brien, "A gigabit/s indoor wireless transmission using MIMO-OFDM visible-light communications," *IEEE Photonics Technology Letters*, vol. 25, no. 2, pp. 171–174, 2013.
- [49] E. Sarbazi, M. Uysal, M. Abdallah, and K. Qaraqe, "Indoor channel modelling and characterization for visible light communications," in *Transparent Optical Networks (ICTON), 2014 16th International Conference on*. IEEE, 2014, pp. 1–4.
- [50] D. Tsonev, H. Chun, S. Rajbhandari, J. J. McKendry, S. Videv, E. Gu, M. Haji, S. Watson, A. E. Kelly, G. Faulkner *et al.*, "A 3-Gb/s single-LED OFDM-based wireless VLC link using a gallium nitride," *IEEE Photonics Technology Letters*, vol. 26, no. 7, pp. 637–640, 2014.
- [51] Y. Engel, S. Mannor, and R. Meir, "The kernel recursive least-squares algorithm," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2275–2285, 2004.
- [52] W. Liu and J. Príncipe, "Kernel affine projection algorithms," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. 1, pp. 1–12, 2008.
- [53] S. Chen, S. Tan, L. Xu, and L. Hanzo, "Adaptive minimum error-rate filtering design: A review," *Signal Processing*, vol. 88, no. 7, pp. 1671–1697, 2008.
- [54] M. Gong, F. Chen, H. Yu, Z. Lu, and L. Hu, "Normalized adaptive channel equalizer based on minimal symbol-error-rate," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1374–1383, 2013.
- [55] R. Mitra and V. Bhatia, "A kernel based technique for MSER equalisation for non-linear channels," in *Signal Processing Conference (EUSIPCO), 23rd European*, 2015, pp. 1446–1450.
- [56] P. Bouboulis and S. Theodoridis, "Extension of Wirtinger's calculus to reproducing kernel Hilbert spaces and the complex kernel LMS," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 964–978, 2011.

- [57] X. Zhang, S. Dimitrov, S. Sinanović, and H. Haas, “Optimal power allocation in spatial modulation OFDM for visible light communications,” in *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*. IEEE, 2012, pp. 1–5.
- [58] “TG7r1, LiFi reference channel models,” *IEEE 802.15 Std., Rev. 1*, July 2015.
- [59] M. Uysal and F. Miramirkhani, “Channel modeling for visible light communications,” *IEEE P802.15 Working Group for Wireless Personal Area Networks (WPANs)*, no. IEEE 15-15-0352-02-007a, 2015.
- [60] V. Jungnickel, M. Uysal, N. Serafimovski, T. Baykas, D. O’Brien, E. Ciaramella, Z. Ghassemlooy, R. Green, H. Haas, P. A. Haigh, V. P. G. Jimenez, F. Miramirkhani, M. Wolf, and S. Zvanovec, “A European view on the next generation optical wireless communication standard,” in *Standards for Communications and Networking (CSCN), 2015 IEEE Conference on*, Oct 2015, pp. 106–111.
- [61] R. Mitra and V. Bhatia, “Chebyshev polynomial-based adaptive predistorter for nonlinear LED compensation in VLC,” *IEEE Photonics Technology Letters*, vol. 28, no. 10, pp. 1053–1056, 2016.
- [62] P. S. Diniz, *Adaptive filtering*, 3rd ed. Springer, 1997.
- [63] S. Zhao, B. Chen, and J. C. Principe, “Kernel adaptive filtering with maximum correntropy criterion,” in *Neural Networks (IJCNN), The 2011 International Joint Conference on*. IEEE, 2011, pp. 2012–2017.
- [64] T. W. Gamelin and R. E. Greene, *Introduction to topology*. Courier Corporation, 1983.
- [65] B. W. Silverman, *Density estimation for statistics and data analysis*, 1st ed. CRC press, 1986, vol. 26.
- [66] B. Chen, S. Zhao, P. Zhu, and J. C. Principe, “Quantized kernel least mean square algorithm,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 1, pp. 22–32, 2012.

- [67] S. Zhao, B. Chen, P. Zhu, and J. C. Príncipe, “Fixed budget quantized kernel least-mean-square algorithm,” *Signal Processing*, vol. 93, no. 9, pp. 2759–2770, 2013.
- [68] R. Mitra and V. Bhatia, “A kernel based technique for MSER equalisation for non-linear channels,” in *EUSIPCO 2015*, 2015, pp. 1461–1465.
- [69] B. Chen, S. Zhao, P. Zhu, and J. C. Principe, “Quantized kernel recursive least squares algorithm,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 9, pp. 1484–1491, 2013.
- [70] B. Chen, J. Liang, N. Zheng, and J. C. Príncipe, “Kernel least mean square with adaptive kernel size,” *Neurocomputing*, vol. 191, pp. 95–106, 2016.
- [71] V. Bhatia and B. Mulgrew, “Non-parametric likelihood based channel estimator for Gaussian mixture noise,” *Signal Processing*, vol. 87, no. 11, pp. 2569–2586, 2007.
- [72] R. Mitra and V. Bhatia, “Adaptive sparse dictionary based kernel minimum symbol error rate post-distortion for nonlinear LEDs in visible light communications,” *IEEE Photonics Journal*, vol. 8, no. 4, pp. 472–484, 2016.
- [73] L. Tao, Y. Wang, Y. Gao, A. P. T. Lau, N. Chi, and C. Lu, “Experimental demonstration of 10 Gb/s multi-level carrier-less amplitude and phase modulation for short range optical communication systems,” *Optics Express*, vol. 21, no. 5, pp. 6459–6465, 2013.
- [74] Y. Wang, L. Tao, Y. Wang, and N. Chi, “High speed WDM VLC system based on multi-band CAP64 with weighted pre-equalization and modified CMMMA based post-equalization,” *IEEE Communications Letters*, vol. 18, no. 10, pp. 1719–1722, 2014.
- [75] S. Chen, S. McLaughlin, P. M. Grant, and B. Mulgrew, “Multi-stage blind clustering equaliser,” *IEEE Transactions on Communications*, vol. 43, no. 234, pp. 701–705, 1995.

- [76] R. Mitra, S. Singh, and A. Mishra, "Improved multi-stage clustering-based blind equalisation," *IET Communications*, vol. 5, no. 9, pp. 1255–1261, 2011.
- [77] R. Mitra and V. Bhatia, "Improved multi-stage clustering-based blind equalisation in interference-limited CDMA environments," *Electronics Letters*, vol. 50, no. 14, pp. 1027–1029, 2014.
- [78] —, "Normalised multi-stage clustering equaliser for underwater acoustic channels," in *Sensor Signal Processing for Defence (SSPD), 2015*. IEEE, 2015, pp. 1–5.
- [79] J. Grubor, O. Jamett, J. Walewski, S. Randel, and K.-D. Langer, "High-speed wireless indoor communication via visible light," *ITG-Fachbericht-Breitbandversorgung in Deutschland-Vielfalt für alle?*, 2007.
- [80] B. Chen, L. Xing, J. Liang, N. Zheng, and J. C. Principe, "Steady-state mean-square error analysis for adaptive filtering under the maximum correntropy criterion," *IEEE Signal Processing Letters*, vol. 21, no. 7, pp. 880–884, 2014.
- [81] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th*. IEEE, 2013, pp. 1–5.
- [82] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 74–81, 2015.
- [83] H. Haas, L. Yin, Y. Wang, and C. Chen, "What is LiFi?" *Journal of Lightwave Technology*, vol. 34, no. 6, pp. 1533–1544, March 2016.
- [84] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis, "A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems," *IEEE Transactions on Signal Processing*, vol. 64, no. 1, pp. 76–88, 2016.

- [85] R. Mitra and V. Bhatia, “Chebyshev polynomial-based adaptive predistorter for non-linear LED compensation in VLC,” *IEEE Photonics Technology Letters*, vol. 28, no. 10, pp. 1053–1056, May 2016.
- [86] L. Yin, X. Wu, and H. Haas, “On the performance of non-orthogonal multiple access in visible light communication,” in *Personal, Indoor, and Mobile Radio Communications (PIMRC), 2015 IEEE 26th Annual International Symposium on*, Aug 2015, pp. 1354–1359.
- [87] R. C. Kizilirmak, C. R. Rowell, and M. Uysal, “Non-orthogonal multiple access (NOMA) for indoor visible light communications,” in *Optical Wireless Communications (IWOW), 2015 4th International Workshop on*, Sept 2015, pp. 98–101.
- [88] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. ElKashlan, H. V. Poor *et al.*, “Application of non-orthogonal multiple access in LTE and 5G networks,” *arXiv preprint arXiv:1511.08610*, 2015.
- [89] A. Benjebbour, A. Li, Y. Kishiyama, H. Jiang, and T. Nakamura, “System-level performance of downlink NOMA combined with SU-MIMO for future LTE enhancements,” in *2014 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2014, pp. 706–710.
- [90] Z. Yang, Z. Ding, P. Fan, and G. K. Karagiannidis, “On the performance of non-orthogonal multiple access systems with partial channel information,” *IEEE Transactions on Communications*, vol. 64, no. 2, pp. 654–667, 2016.
- [91] D. C. O’Brien, L. Zeng, H. Le-Minh, G. Faulkner, J. W. Walewski, and S. Randel, “Visible light communications: Challenges and possibilities,” in *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*. IEEE, 2008, pp. 1–5.
- [92] H. Ma, L. Lampe, and S. Hranilovic, “Integration of indoor visible light and power line communication systems,” in *Power Line Communications and Its Applications (ISPLC), 2013 17th IEEE International Symposium on*. IEEE, 2013, pp. 291–296.

- [93] W. Zhang, "A general framework for transmission with transceiver distortion and some applications," *IEEE Transactions on Communications*, vol. 60, no. 2, pp. 384–399, 2012.
- [94] T. Fath and H. Haas, "Performance comparison of MIMO techniques for optical wireless communications in indoor environments," *IEEE Transactions on Communications*, vol. 61, no. 2, pp. 733–742, 2013.
- [95] Q. Wang, Z. Wang, and L. Dai, "Multiuser MIMO-OFDM for visible light communications," *IEEE Photonics Journal*, vol. 7, no. 6, pp. 1–11, 2015.
- [96] D. Dixit and P. Sahu, "Performance analysis of rectangular QAM with SC receiver over Nakagami-fading channels," *IEEE Communications Letters*, vol. 18, no. 7, pp. 1262–1265, 2014.
- [97] D. Lee, K. Choi, K.-D. Kim, and Y. Park, "Visible light wireless communications based on predistorted OFDM," *Optics Communications*, vol. 285, no. 7, pp. 1767–1770, 2012.
- [98] C. Chow, C. Yeh, Y. Liu, P. Huang, and Y. Liu, "Adaptive scheme for maintaining the performance of the in-home white-LED visible light wireless communications using OFDM," *Optics Communications*, vol. 292, pp. 49–52, 2013.
- [99] J. G. Proakis and M. Salehi, *Fundamentals of communication systems*, 3rd ed. Pearson Education India, 2007.

List of publications

Publications from thesis

Journal papers:

1. R. Mitra, and V. Bhatia, "Improved multi-stage clustering-based blind equalisation in interference-limited CDMA environments," *Electronics Letters*, vol. 50, no. 14, pp. 1027-1029, 2014.
2. R. Mitra, and V. Bhatia, "Chebyshev Polynomial-Based Adaptive Predistorter for Nonlinear LED Compensation in VLC," *IEEE Photonics Technology Letters*, vol. 28, no. 10, pp. 1053-1056, 2016.
3. R. Mitra, and V. Bhatia, "Adaptive Sparse Dictionary-Based Kernel Minimum Symbol Error Rate Post-Distortion for Nonlinear LEDs in Visible Light Communications," *IEEE Photonics Journal*, vol. 8, no. 4, pp. 1-13, 2016.
4. R. Mitra, and V. Bhatia, "Finite Dictionary Techniques for MSER Equalization in RKHS," *Signal Image Video Processing (SIVP)*, Springer, vol. 11, no. 5, pp. 849-856, 2017.
5. R. Mitra, and V. Bhatia, "Unsupervised Multi-Stage Clustering based Hammerstein post-distortion for VLC," *IEEE Photonics Journal*, vol. 9, no. 1, pp. 1-10, 2017.
6. R. Mitra, and V. Bhatia, "Low complexity post-distorter for visible light communications," *IEEE Communication Letters*, 2017 (Accepted).
7. R. Mitra, and V. Bhatia, "Kernel-based parallel multi-user detector for massive-MIMO," *Elsevier Computers and Electrical Engineering (special issue on "Recent Trends in Signal Processing for 5G Technologies, 2017")*, (Accepted).

Conference papers:

1. R. Mitra, and V. Bhatia, "Improved Multi-stage Clustering Based Blind Equalisation in Distributed Environments," Information Technology (ICIT), 2014 International Conference on. IEEE, India, 2014.
2. R. Mitra, and V. Bhatia, "The diffusion-KLMS algorithm," Information Technology (ICIT), 2014 International Conference on. IEEE, India, 2014.
3. R. Mitra, and V. Bhatia, "A kernel based technique for MSER equalisation for non-linear channels," Signal Processing Conference (EUSIPCO), 2015 23rd European. IEEE, France, 2015.
4. R. Mitra, and V. Bhatia, "Normalised multi-stage clustering equaliser for underwater acoustic channels," Sensor Signal Processing for Defence (SSPD), 2015. IEEE, UK, 2015.
5. R. Mitra, N. Kumar, and V. Bhatia, "Blind MBER based selection diversity combining in amplify and forward relay protocol in Rayleigh fading environments," Signal Processing and Communication (ICSC), 2015 International Conference on. IEEE, India, 2015.
6. R. Mitra, N. Kumar, and V. Bhatia, "A hybrid MBER-IMSC based blind equalisation scheme in amplify and forward relay protocol in Rayleigh fading environments," 2015 IEEE Bombay Section Symposium (IBSS). IEEE, India, 2015.