Emulation of Polarization Mode Dispersion: Discrete-Time Models and Random Parameter Sampling

by

Ahmet Gökçen Mahmutoğlu

A Thesis Submitted to the
Graduate School of Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Electrical & Computer Engineering

Koç University

January, 2012

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master thesis by

Ahmet Gökçen Mahmutoğlu

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Reading Committee:

Assoc. Prof. Dr. Alper Tunga Erdoğan

Assoc. Prof. Dr. Alper Demir

Asst. Prof. Dr. Serdar Kozat

Asst. Prof. Dr. Emre Mengi

Asst. Prof. Dr. Şükrü Ekin Kocabaş

Date: _____

Abstract

We present continuous and discrete-time models of polarization mode dispersion (PMD) in single-mode optical fibers and random parameter sampling schemes for these models for accurate and low-complexity PMD emulation. Our discussion on continuous-time PMD emulation builds upon an earlier work, which uses model reduction techniques and Markov chain Monte Carlo (MCMC) methods. Here we develop an improved model parameter sampling procedure to generate continuous-time reduced complexity models making use of the proposed compensated MCMC method.

As the main contribution of this thesis, we present a discrete time PMD emulation scheme that is on a par with the emerging coherent receiver techniques based on digital signal processing algorithms. This scheme uses multiple-input multiple-output (MIMO) FIR filters that are lossless and therefore lend themselves as perfect candidates for emulation of fiber channels suffering from PMD. The concatenated composition of these filters resembles the continuous time lumped model of PMD channels and offers a flexible emulator structure in terms of computational complexity which constitutes the main bottleneck for real-time DSP applications. The parameter sampling problem for accurate PMD emulation considering the required statistical behavior of a PMD emulator is tackled using three different approaches, which are introduced in order of decreasing deviation from the desired statistics and increasing computational complexity.

ÖZETÇE

Optik tek kipli fiberler üzerindeki polarizasyon kipi saçılımı (PMD) öykünücülerinde kullanılmak üzere, zamanda devamlı ve ayrık modeller, ve bu modeller için rassal parametre örnekleme düzenekleri sunmaktayız. Zamanda devamlı PMD öykünmesi üzerine tartışmamızı, model indirgeme teknikleri ve Markov zinciri Monte Carlo (MCMC) yöntemleri kullanan geçmiş bir çalışmanın üzerine inşaa etmekte ve tarafımızdan önerilen dengelenmiş MCMC yöntemini kullanarak devamlı zamanlı, indirgenmiş karmaşıklıkta modeller yaratmak için iyileştirilmiş bir rassal parametre örnekleme yordamı geliştirmekteyiz.

Bu tezin ana katkısı olarak, yeni ortaya çıkmakta olan sayısal işaret işleme algoritmaları tabanlı faz uyumlu alıcı yöntemleriyle aynı düzlemde yer alan, zamanda ayrık bir PMD öykünmesi yordamı tanıtmaktayız. Bu yordam, kayıpsız ve bu nedenle PMD'den etkilenen kanallar için mükemmel adaylar olan çok-girişli çok-çıkışlı (MIMO) FIR filtreler kullanmaktadır. Bu filtreler, PMD kanallarının zamanda devamlı toplu modeli ile benzeşen art arda bağlı bir biçime sahiptirler ve gerçek zamanlı DSP uygulamalarında ana darboğazı oluşturan hesaplama karmaşılığı açısından esnek bir öykünücü yapısı sunmaktadırlar. Arzu edilen istatistiksel özellikleri göz önünde bulundurarak hassas bir PMD öykünücüsü için parametre örnekleme sorunu, üç farklı yöntem kullanarak çözülmüştür. Bu yöntemler azalan istatistiksel sapma ve artan hesaplama karmaşıklığı sırasıyla sunulmaktadırlar.

Acknowledgments

I am grateful to my advisors, Alper T. Erdoğan and Alper Demir, for their encouraging attitude and stimulating conversations that put me back on track when I was lost in the vast realm of trial and error. Without their knowledge and feedback this thesis would not have been possible.

I thank my friends, Berk İncekara, Işık Önay, Osman Dülek and Tunca Sekban for doing the same in my personal life for a very long time. My office-mates Pınar Karabulut and Ozan Yıldırım, who have been part of a comfortable and productive working environment, were also an unexpected turn of chance for me.

Last but not least, I express my deepest gratitude to my family who have been immensely supportive and incredibly patient with me.

Table of Contents

List of Figures

Nomenclature

CDF	cumulative distribution function
DGD	differential group delay
DSP	digital signal processing/processor
EM	expectation maximization
FIR	finite impulse response
i.i.d.	identically and independently distributed
LHS	left-hand side
LTI	linear time invariant
MCMC	Markov chain Monte Carlo
MIMO	multiple-input multiple-output
PDE	partial differential equation
PDF	probability density function
PMD	polarization mode dispersion
PSP	principal state of polarization
RHS	right-hand side
SVD	singular value decomposition
UPD	uniform parameter distribution

Chapter 1

Introduction

1.1. Motivation and Outline

Optical fiber links form the backbones of global and local telecommunication networks as they provide larger bandwidth compared to other media such as wireless and copper links. The growing demand for data and voice services has led to research for better spectral utilization of fiber links. With the use of single-frequency lasers and dispersion-shifted optical fibers, the transmission speed for a single channel in optical communication links has reached Gbit/s level. Currently, one of the biggest hurdles preventing these communication systems based on single-mode optical fibers from operating at even higher speeds is polarization mode dispersion (PMD).

PMD is a statistical phenomenon that occurs in single-mode optical fibers due to random imperfections breaking the circular symmetry of the fiber. PMD results in a random coupling and a speed difference between the two, normally degenerate orthogonal polarization modes of propagation. These effects give rise to pulse broadening and intersymbol interference (ISI) as well as mixing of two orthogonal polarization channels which limit the transmission speed while increasing the system outage probability. The causes of PMD, e.g. random flaws inflicted during the production process

or mechanical deformations on underground cables, and its properties is discussed in detail in optical communications literature [1, 2, 3, 4].

Since its first description [5, 6], PMD assumed a fundamental role in development of high-speed optical communication systems. Although recent advances in coherent receiver technologies brought some relief from optical distortions in optical fibers with the use of compensation algorithms [7, 8], PMD continues to be a major impediment at bit rates of 100 Gb/s and more [9, 10]. One of the important properties that needs to be considered during the design of optical communication systems operating on single-mode optical fibers is how much the system is affected by PMD. The statistical nature of PMD complicates this procedure immensely. Considering that different fibers will exhibit different PMD characteristics, that in themselves depend on the fiber's position and other environmental factors, the testing of such systems would have to involve repetitive steps with numerous different fibers of realistic lengths. The solution to this problem lies in the development of PMD emulators.

Communication channel emulators are critical devices for both the development and testing of the communication transceivers. Emulators mimic the physical behavior of the channel by distorting the transmitter output based on the mathematical model of the communication channel input/output characteristics. Therefore, the communication system developers and testers can use these devices to test the transmitter/receiver combinations for various link scenarios in the lab environment. This is a critical convenience, especially for the optical fiber communication systems, where the development and testing with real channels are not manageable for the reasons mentioned above.

For the case of PMD emulation, such a device must be able to capture the statistical properties of a PMD channel with the required accuracy for the particular communication system in question. Furthermore, they must be versatile enough to correctly emulate the behavior of older fibers installed in the existing optical links which exhibit relatively high differential group delays (DGD), and modern fibers that cause signal distortion mainly due to higher order effects [11].

Generally, PMD emulator design can follow two different routes. One of these methods results in devices that consist of optical parts such as concatenated birefringent plates and mechanical parts that allow the individual rotation of these plates together with both analog and digital electronic circuitry. The other method of PMD emulation employs analytical tools that approximate the desired statistics computationally.

The first group of PMD emulators with physical components is currently a commercially available technology and therefore widely used. Devices with concatenated polarization maintaining fibers combined with polarization scramblers can capture the DGD statistics as well as the autocorrelation function of PMD vectors to a certain degree [12]. The approximation of the autocorrelation function enables these devices to emulate not only the first order but also the higher order effects of PMD [13, 14]. PMD emulators using freely rotatable birefringent crystals [15] have the goal to uniformly distribute the polarization state on the Poincare sphere [2]. These emulators are only programmable for a single frequency and have limited control over higher order PMD. Devices that separate birefringent elements with freely rotatable thin waveplates, on the other hand, are designed as PMD sources that strive to create PMD effects with predetermined values [16]. Although PMD emulators constructed with long polarization maintaining fibers which have fiber twisters placed periodically on them or emulators that use ferroelectric crystals yield better results in terms of PMD statistics, their complexity constitutes a problem in their practical implementations [17].

Despite providing the desired statistical characteristics, the above discussed methods of PMD emulation suffer from limitations originating from their construction with movable optical and mechanical parts. This structure forces these devices not only to be complex and bulky in size but also their operation at higher transmission speeds is limited by the physical properties of the individual components.

The use of digital signal processing (DSP) based emulators would address these problems, namely hard-programmability and large size, regarding analog emulators.

The ability to construct such emulators is subject to the existence of effective DSP emulation algorithms capturing the statistical behavior of fiber channels. This is an area that is not sufficiently addressed in literature, and in this work, we are going to address the construction of such algorithms. Our treatment of this problem is presented in four parts.

In the remainder of this first chapter we will summarize the necessary mathematical groundwork for the description of PMD from a systems theoretical perspective. To this end, we first define two important quantities regarding PMD computations, the PMD vector and the PMD operator, and investigate the connection between them. At the end of this introductory chapter we discuss the statistical properties of the PMD vector.

The second chapter is dedicated to the presentation of different PMD models. The first part discusses the continuous-time lumped model of PMD with a high number of concatenated birefringent sections, which we call the "full model". In order to arrive at this final form of the model, the time-domain representation of a PMD channel is first given in the form of corresponding differential equations and the transition to a frequency-domain representation is then made using a modified trapezoidal discretization scheme.

The second PMD model we present has its origins in the full model. This second continuous-time model, which is our first step towards PMD emulation, relies on Pade approximation and Krylov subspace techniques for model reduction such that it has significantly fewer birefringent sections than the full model without losing the statistical properties of PMD channels.

The third PMD model introduces the main contribution of this work which is a discrete-time filter with low complexity that can be employed for PMD emulation on DSPs. Our approach in this regard will be based on the exploitation of the mathematical model corresponding to the PMD phenomenon. We will make use of the fact that the PMD channels are generalized paraunitary channels, which means they can be modeled by a lossless system with memory in combination with the frequency

independent loss factor. This fact enables the use of discrete time paraunitary filter structures. Consequently, the DSP based PMD emulation algorithm design boils down to coming up with an accurate and low complexity method to select parameters of the discrete time paraunitary filters.

In the third chapter we make a detour to discuss Markov chain Monte Carlo (MCMC) sampling methods before going into the subject of how the parameters of the models presented in Chapter 2 can be sampled such that the emerging model satisfies the statistical requirements. Here we will present a general modified MCMC sampling method, which we call "compensated MCMC", that can be employed for the solution of problems regarding random input sampling for complex models such as PMD emulators and investigate its properties. This chapter contains the argumentation why a new sampling scheme is necessary in the course of our work towards PMD emulators as well as other usage possibilities of the proposed algorithm with an example in numerical solutions of stochastic differential equations.

The final part of this work presents different parameter sampling algorithms (including compensated MCMC) for PMD emulators. The accuracy in terms of approximating the statistical properties of a real PMD channel is the main objective of our treatment. Another critical objective is the suitability for real time implementation. In fact, the frequency bandwidth to be simulated is relatively large in fiber communication systems and the corresponding DSP emulators need to work in a large sampling frequency regime. Therefore, the complexity of the algorithms will be a major concern in the design of parameter sampling schemes.

The first parameter sampling scheme we discuss is for the reduced complexity continuous time model – a variation of the random parameter sampling method presented in [18]. Here we use compensated MCMC instead of the standard MCMC algorithm and present simulation results that illustrate the improvement in the statistical accuracy of the models when their parameters are sampled with the proposed method. The second part of this chapter deals with the parameter sampling problem for the discrete-time paraunitary model. We propose three different methods for different

PMD channel scenarios and present these techniques in order of decreasing deviation from the desired statistics and increasing computational complexity. Finally, we conclude our treatment with a brief summary.

1.2. Background on PMD

This section covers the basics of PMD to the extent that it is necessary for further discussion. Interested readers should refer to [3] and the references therein for a detailed treatment of the subject.

1.2.1. The PMD Vector

In the classical description of polarization, a monochromatic plane wave propagating in the z-direction is represented by

$$\mathbf{E}(z,t) = \begin{pmatrix} E_x e^{j\phi_x} \\ E_y e^{j\phi_y} \end{pmatrix} e^{j(\omega t - kz)} \quad , \tag{1.1}$$

where E_x and E_y denote the electric fields components in x and y directions respectively. The 2×1 vector in (1.1) describes the polarization ellipse on the plane perpendicular to the propagation axis. After factoring out the common phase of its two components and normalizing their magnitudes, it can be written as

$$\mathbf{E} = E_0 \begin{pmatrix} \cos(\chi) \\ \sin(\chi) e^{j\phi} \end{pmatrix} \quad , \tag{1.2}$$

where $E_0 = \sqrt{E_x^2 + E_y^2}$, $\tan(\chi) = E_y/E_x$ and $\phi = \phi_y - \phi_x$. The above vector is called the Jones vector.

Another way of describing the polarization state of light is the four-dimensional Stokes vector, $\mathbf{S} = [S_0 \ S_1 \ S_2 \ S_3]^T$, which is connected to the Jones vector with the

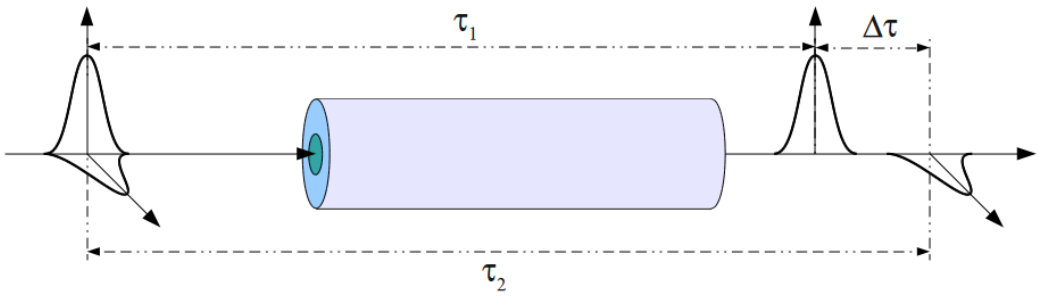


Figure 1.1: The speed difference between the two orthogonal modes of propagation on an optical fiber.

following equations:

$$\begin{aligned} S_0 &= \tfrac{1}{2}E_0^2 & S_1 &= \tfrac{1}{2}E_0^2\cos(2\chi) \\ S_2 &= \tfrac{1}{2}E_0^2\sin(2\chi)\cos(\phi) & S_3 &= \tfrac{1}{2}E_0^2\sin(2\chi)\sin(\phi) \end{aligned} \tag{1.3}$$

The reason for this redundancy in the description of the polarization states is that although the Jones vector can be derived directly from Maxwell's equations, it is not directly measurable, whereas the Stokes vector is. The Stokes vector is usually normalized by the intensity component S_0 which results in a three-dimensional unit norm vector. Every polarization state can be described by the pointing direction of this vector on the so-called Poincaré sphere. With the help of these formalisms we can characterize the effects of PMD.

Figure 1.1 illustrates how two orthogonal polarization modes on an optical fiber differ in their propagation speeds in the medium. It is intuitive to describe this phenomenon in terms of the eigenmodes of the fiber which remain unchanged in the course of their progress but experience different group velocities. However, this approach proves to be unfruitful because the eigenmodes change to first order with frequency. Therefore the computation of the group delay, which requires a frequency derivative, does not allow the separation of the delay term from its eigenmode. At this point a crucial observation can be made: according to Poole [6], two orthogonal polarization states at the output of a fiber can always be found that are stationary to

first order in frequency. Each of these states is called a principal states of polarization (PSP). The main object of interest in PMD related calculations, the PMD vector, can be defined using one of these states and the differential group delay between them [2].

Definition 1. *The PMD vector, $\boldsymbol{\tau}$, is the three dimensional vector which points in the direction of the slow PSP and has the length of the differential group delay between fast and slow PSPs.*

The PMD vector, $\boldsymbol{\tau}$, defines the infinitesimal precession rule which governs the behavior of the output polarization state, $\mathbf{s}$, with a variation in the frequency:

$$\frac{\mathrm{d}\mathbf{s}}{\mathrm{d}\omega} = \boldsymbol{\tau} \times \mathbf{s} \quad , \tag{1.4}$$

where "$\times$" denotes the usual cross product in three-dimensional space. We note that although the PMD vector is stationary to first order, this does not mean that it is frequency independent. Because the output PSP must be stationary to first order at every frequency, the corresponding input PSP has to change with frequency to account for the precession difference about the birefringent axis and this results in a frequency dependent PMD vector which is denoted with $\boldsymbol{\tau}(\omega)$. This frequency dependency is the source of the so-called higher order PMD effects.

Another way of connecting the two formalisms of polarization is the spin-vector calculus of polarization [2]. This treatment reshapes the equations (1.3) with the Pauli spin vector, $\vec{\sigma}$:

$$\mathbf{s} = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix} = \begin{pmatrix} \langle s|\sigma_1|s\rangle \\ \langle s|\sigma_2|s\rangle \\ \langle s|\sigma_3|s\rangle \end{pmatrix} = \langle s|\vec{\sigma}|s\rangle \quad , \tag{1.5}$$

where $\mathbf{s}$ is the normalized three-dimensional Stokes vector and $|s\rangle$ is the Jones vector. Here we use the bra-ket notation to conform with the existing literature as well as to clarify the distinction between vectors in Jones and Stokes spaces. The 2×2 spin

matrices in (1.5) are given by

$$\sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \; ; \; \sigma_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \; ; \; \sigma_3 = \begin{pmatrix} 0 & -j \\ j & 0 \end{pmatrix} \; . \tag{1.6}$$

Within this framework, the reciprocal conversion from a Stokes vector to a Jones vector can be written as the eigenvalue equation

$$\mathbf{s} \cdot \vec{\sigma} |s\rangle = |s\rangle \quad . \tag{1.7}$$

The dot product in the above equation is defined as the traceless Hermitian matrix,

$$\begin{aligned} \mathbf{s} \cdot \vec{\sigma} &= s_1\sigma_1 + s_2\sigma_2 + s_3\sigma_3 \\ &= \begin{pmatrix} s_1 & s_2 - js_3 \\ s_2 + js_3 & -s_1 \end{pmatrix} \quad . \end{aligned} \tag{1.8}$$

This matrix is very closely related to the so-called PMD operator which describes the transformation of an arbitrary input polarization state at the output of an optical fiber. Now we derive this connection which will also be useful in the systems theoretical treatment of PMD.

1.2.2. The PMD Operator

The input-output relation of polarization modes in an optical fiber can be described with the transformation equation

$$|t\rangle = e^{-j\phi_0}\mathbf{U}|s\rangle \quad , \tag{1.9}$$

where $\mathbf{U}$ is a 2×2 unitary matrix such that $\mathbf{U}\mathbf{U}^* = \mathbf{I}$. The frequency derivative of this equation consists of two delay terms

$$\begin{aligned} |t\rangle_\omega &= (-j\phi_0'\mathbf{U} + \mathbf{U}_\omega)e^{-j\phi_0}|s\rangle \\ &= -j(\phi_0' + j\mathbf{U}_\omega\mathbf{U}^*)|t\rangle. \end{aligned} \tag{1.10}$$

The first term in (1.10) is the common phase delay, τ_0 both polarization states undergo and the second factor, which is called the PMD operator, determines the differential group delay properties of the fiber. Since $\mathbf{U}$ is unitary we have $\mathbf{U}\mathbf{U}^* = \mathbf{I}$. Taking the frequency derivative of this equation results in $\mathbf{U}_\omega\mathbf{U}^* = -\mathbf{U}\mathbf{U}_\omega^*$ which implies that the matrix $j\mathbf{U}_\omega\mathbf{U}^*$ is Hermitian. Using the Taylor expansion of $\det(\mathbf{U}(\omega + \delta\omega))$ it can also be shown that this matrix is traceless.

$$\begin{aligned} \det(\mathbf{U}(\omega + \delta\omega)) &\approx \det(\mathbf{U}(\omega)) = 1 \\ &\approx \det(\mathbf{U} + \delta\omega\mathbf{U}_\omega) \\ &= \det(\mathbf{I} + \delta\omega\mathbf{U}_\omega\mathbf{U}^*)\det(\mathbf{U}) \\ &= 1 + \mathrm{tr}(\mathbf{U}_\omega\mathbf{U}^*)\delta\omega + O(\delta\omega^2) \end{aligned} \tag{1.11}$$

Therefore for this first order approximation to hold the trace of $\mathbf{U}_\omega\mathbf{U}^*$ must be zero.

A Hermitian 2×2 matrix with zero trace has orthogonal eigenvectors and real eigenvalues that are symmetrical about the origin. If we denote these eigenvalue-eigenvector pairs by $(\pm\tau/2,\ |p_\pm\rangle)$, we can equivalently express (1.10) as

$$\begin{aligned} |p_\pm\rangle_\omega &= -j(\phi_0' + j\mathbf{U}_\omega\mathbf{U}^*)|p_\pm\rangle \\ &= -j\left(\phi_0' \pm \frac{\tau}{2}\right)|p_\pm\rangle \end{aligned} \tag{1.12}$$

Hence the total group delay for the two polarization components is $\tau_g = \tau_0 \pm \tau/2$.

Moreover, if we convert the frequency derivative to Stokes space we can see that,

$$\begin{aligned}\mathbf{p}_\omega &= \langle p_\omega|\vec{\sigma}|p_\omega\rangle \\ &= j\tau_g\langle p|\vec{\sigma}|p\rangle - j\tau_g\langle p|\vec{\sigma}|p\rangle = 0 \quad ,\end{aligned} \tag{1.13}$$

consequently the eigenvectors of $j\mathbf{U}_\omega\mathbf{U}^*$ are principal states of polarization with a differential group delay, τ, between them that is equal to the distance of the corresponding eigenvalues, i.e., the spacing of the matrix.

Lastly we investigate the connection between the PMD operator and the PMD vector. Using equations (1.10) and (1.12) we can obtain the useful relationship

$$j\mathbf{U}_\omega\mathbf{U}^* = \frac{1}{2}(\boldsymbol{\tau}\cdot\vec{\sigma}) \quad . \tag{1.14}$$

Furthermore, the precession rule in (1.4) can be derived if we consider how an arbitrary output polarization state changes with frequency.

$$\begin{aligned}\mathbf{s}_\omega &= \langle s_\omega|\vec{\sigma}|s\rangle + \langle s|\vec{\sigma}|s_\omega\rangle \\ &= j\langle s|\left(\tau_0 + \frac{1}{2}(\boldsymbol{\tau}\cdot\vec{\sigma})\right)\vec{\sigma}|s\rangle - j\langle s|\vec{\sigma}\left(\tau_0 + \frac{1}{2}(\boldsymbol{\tau}\cdot\vec{\sigma})\right)\vec{\sigma}|s\rangle \\ &= \frac{j}{2}\langle s|(\boldsymbol{\tau}\cdot\vec{\sigma})\vec{\sigma} - \vec{\sigma}(\boldsymbol{\tau}\cdot\vec{\sigma})|s\rangle \\ &= \langle s|\boldsymbol{\tau}\times\vec{\sigma}|s\rangle\end{aligned} \tag{1.15}$$

Equation (1.15) implies

$$\frac{\mathrm{d}\mathbf{t}}{\mathrm{d}\omega} = \boldsymbol{\tau}\times\mathbf{s} \quad . \tag{1.16}$$

1.2.3. PMD Vector Statistics

It can be shown theoretically as well as experimentally that the PMD vector, $\boldsymbol{\tau}$, is the result of a random walk process in three-dimensional space and it can be expressed as a variant of 3D Brownian motion. Therefore it has i.i.d. Gaussian vector components with zero-mean and variance σ^2 [19, 20].

As it will be discussed in Chapter 2, an optical fiber, subject to polarization mode dispersion, can be modeled as the concatenation of many statistically independent birefringent sections. As the number of sections grow, the accuracy of the model increases. High accuracy models can be constructed with one hundred to thousands of sections. Using such a model we can reveal some of the relevant statistics of the PMD vector in an experimental fashion and confirm their accordance with the existing theoretical results. Moreover, with the help of these experiments we can lay the groundwork for some of the statistical assumptions and approximations we will be using in the development of PMD emulators. During the course of this investigation, in order to determine the statistical properties of the fiber model, we will make use of a fiber simulation with 200 sections and 4.5×10^6 samples. The mean DGD of the model is set to be 10 psec.

Although their marginal distributions are known, there is no theoretical result about the joint distribution of the components of the PMD vector $\boldsymbol{\tau}(\omega)$ at different frequencies. The marginals of these components are Gaussian and independent at the same frequency. Figure 1.2 shows the contours of the joint distribution of the first and second components of $\boldsymbol{\tau}(\omega)$ at the center frequency, $\omega = 0$. The dotted lines are the contours of a jointly Gaussian distribution fitted to the simulation data. In this case, the covariance matrix is

$$C = \begin{bmatrix} 0.3943 & 0.0001 \\ 0.0001 & 0.3921 \end{bmatrix} \times 10^{-22}$$

which is in agreement with the expected result.

On the other hand, one expects the i^{th} component of $\boldsymbol{\tau}$ to be dependent on the i^{th} component at another frequency. The joint distribution is again unknown but there is evidence suggesting that it is not jointly Gaussian due to its frequency derivative not being Gaussian [19]. This odd behavior of random variables with Gaussian marginals not being jointly Gaussian can be seen in Figure 1.3. Although there is a clear deviation from the fitted jointly Gaussian distribution the similarity between two

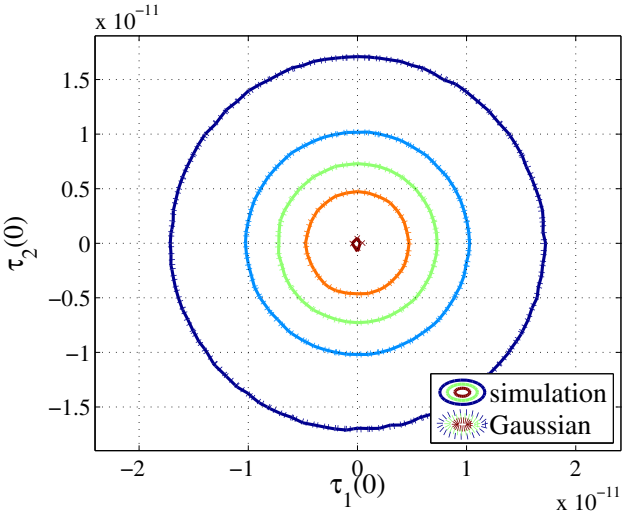


Figure 1.2: Contour plot of joint probability density function of first and second components of the PMD vectors at the frequency $f = 0$.

curves hints on a joint Gaussian approximation. The covariance matrix computed for the fitted Gaussian this time is

$$C = \begin{bmatrix} 0.3943 & 0.2951 \\ 0.2951 & 0.3939 \end{bmatrix} \times 10^{-22} \quad .$$

As a result of the Gaussianity of the PMD vector components the Euclidean norm of $\boldsymbol{\tau}$ has a Maxwellian distribution. A random variable $W = \sqrt{X^2 + Y^2 + Z^2}$ has a Maxwellian distribution if the random variables X, Y and Z are i.i.d. zero-mean Gaussians. The PDF of W is given by

$$f_W(w) = \sqrt{\frac{2}{\pi}} \frac{w^2}{\sigma^3} e^{-w^2/(2\sigma^2)} \quad , \tag{1.17}$$

where σ is the standard deviation of X, Y and Z. The mean of W is connected to σ by

$$E[W] = \sqrt{\frac{8}{\pi}} \sigma \tag{1.18}$$

Figure 1.4 shows the PDF and the complementary cumulative distribution function

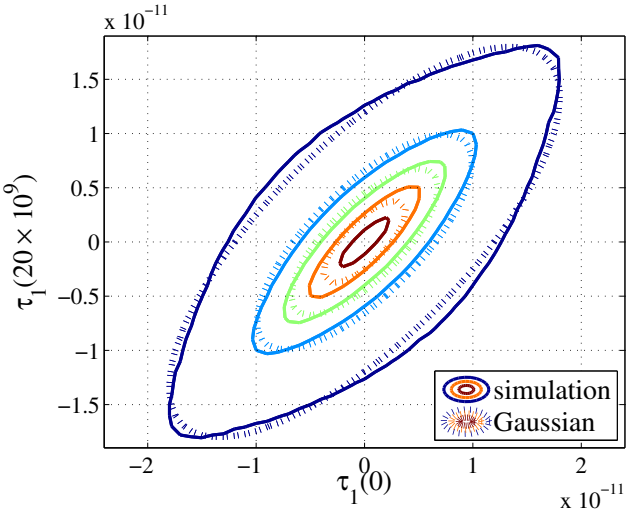


Figure 1.3: Contour plot of joint probability density function of first components of the PMD vectors at frequencies $f = 0$ and $f = 20GHz$.

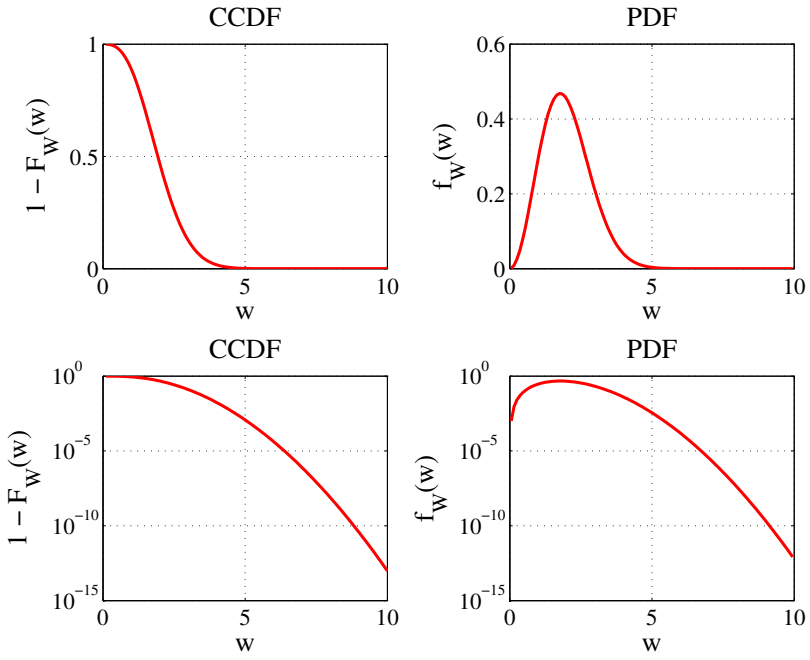


Figure 1.4: Probability density function (PDF) and complementary cumulative distribution function (CCDF) of Maxwellian random variable with mean 2. Both curves are plotted in linear and logarithmic scales.

(CCDF) of a Maxwellian distribution in linear as well as semi-logarithmic scale.

Chapter 2

Models of PMD

2.1. Introduction

In the most general case, one uses coupled nonlinear Schrödinger equations to devise a mathematical model of signal propagation in single mode optical fibers [21]. The first part of this chapter discusses the continuous-time lumped model of PMD with a high number of concatenated birefringent sections, which emerges from these equations with the terms for the effects of chromatic dispersion and fiber nonlinearities removed. In order to arrive at this final form, which we call the "full model", the time-domain representation of a PMD channel is first given with the corresponding differential equations and the transition to a frequency-domain representation is then made using a modified trapezoidal discretization scheme for the solution of these equations and converting them into the Laplace domain. The treatment in this section is based on the information in [18].

The second PMD model we present has its origins in the full model. In order to address the emulator complexity problem, caused by the higher number of birefringent sections in the full model, a model complexity reduction procedure was proposed in [18]. This reduced complexity continuous-time model, which is our first step towards PMD emulation, relies on Pade approximation and Krylov subspace techniques for

model reduction such that it has significantly fewer birefringent sections than the full model without losing the statistical properties of PMD channels. The results here serve as a brief summary for the techniques developed in [18] and are given without proof. Interested readers should refer to the original paper for details.

The third PMD model introduces the main contribution of this work which is a discrete-time filter with low complexity that can be employed for PMD emulation on DSPs. Our approach in this regard will be based on the exploitation of the mathematical model corresponding to the PMD phenomenon. We will make use of the fact that the PMD channels are generalized paraunitary channels, which means they can be modeled by a lossless system with memory in combination with the frequency independent loss factor. This fact enables the use of discrete time paraunitary filter structures. First we will introduce the general paraunitary FIR filter before turning our focus on the resulting structures concerning PMD computations. Lastly we will present the inherent DGD statistics of paraunitary filters.

2.2. Continuous-Time Lumped Lossless Model

A light wave propagating in a single-mode optical fiber carries two orthogonal polarization components. The complex envelops of the corresponding electric fields can be described with the following coupled partial differential equations [22, 23]:

$$\frac{\partial \mathbf{A}(z,t)}{\partial z} = \mathbf{D}\,\frac{\partial \mathbf{A}(z,t)}{\partial t} + \mathbf{R}(z)\;\mathbf{A}(z,t) \tag{2.1}$$

where t is time, z is the position along the propagation direction, with

$$\mathbf{A}(z,t) = \left[A_1(z,t)\;\; A_2(z,t)\right]^T \tag{2.2}$$

and

$$\mathbf{D} = \begin{bmatrix} 0 & 0 \\ 0 & -\Delta\beta \end{bmatrix} \quad . \tag{2.3}$$

The first term on the RHS of (2.1) expresses the time difference relationship of the two polarization components, and the second term accounts for the coupling between them. Because (2.1) represents a lossless two-dimensional system, $\mathbf{R}(z)$ in (2.1) should be a skew-Hermitian, 2×2 matrix, i.e., $\mathbf{R} = -\mathbf{R}^*$, where the superscript "$*$" denotes conjugate-transpose [22]. Furthermore this PDE represents a causal system because of its retarded frame of reference that moves with the group velocity of the faster polarization component. Hence the system described by (2.1) is a lossless, causal, distributed, continuous-time system.

In order to obtain the evolution equations with respect to spatial parameter z, we now define

$$\mathbf{B}(z,t) := \exp\left[-\int_0^z \mathbf{R}(u)\ \mathrm{d}u\right]\mathbf{A}(z,t) = \exp\left[-\mathbf{Q}(z)\right]\mathbf{A}(z,t) \quad , \tag{2.4}$$

where $\mathbf{Q}(z) = \int_0^z \mathbf{R}(u)\ \mathrm{d}u$. Note that the matrices $\mathbf{R}(z)$ and $\mathbf{Q}(z)$ commute since their product is Hermitian.

We differentiate (2.4) with respect to z

$$\frac{\partial \mathbf{B}(z,t)}{\partial z} = -\mathbf{R}(z)\ \mathbf{B}(z,t) + \exp\left[-\mathbf{Q}(z)\right]\frac{\partial \mathbf{A}(z,t)}{\partial z} \tag{2.5}$$

and substitute (2.1) in the above equation to obtain

$$\begin{aligned}
\frac{\partial \mathbf{B}(z,t)}{\partial z} &= -\mathbf{R}(z)\ \mathbf{B}(z,t) + \exp\left[-\mathbf{Q}(z)\right]\left[\mathbf{D}\,\frac{\partial \mathbf{A}(z,t)}{\partial t} + \mathbf{R}(z)\ \mathbf{A}(z,t)\right] \\
&= -\mathbf{R}(z)\ \mathbf{B}(z,t) + \\
&\exp\left[-\mathbf{Q}(z)\right]\left[\mathbf{D}\exp\left[\mathbf{Q}(z)\right]\frac{\partial \mathbf{B}(z,t)}{\partial t} + \mathbf{R}(z)\exp\left[\mathbf{Q}(z)\right]\mathbf{B}(z,t)\right] \\
&= \exp\left[-\mathbf{Q}(z)\right]\mathbf{D}\ \exp\left[\mathbf{Q}(z)\right]\frac{\partial \mathbf{B}(z,t)}{\partial t}
\end{aligned} \tag{2.6}$$

where we have used the commutativity of $\exp\left[-\mathbf{Q}(z)\right]$ with $\mathbf{R}(z)$. We note here that $\exp\left[\mathbf{Q}(z)\right] = \mathbf{U}(z)$ is a unitary matrix (i.e., $\mathbf{U}(z)^*\ \mathbf{U}(z) = \mathbf{I}$, which follows from the fact that $\mathbf{Q}(z)$ is skew-Hermitian) and its inverse is given by $\exp\left[-\mathbf{Q}(z)\right]$.

We can now discretize (2.6) on the z-axis and obtain a continuous-time lumped approximation of the original distributed system. Since the losslessness of the system is paramount to us, the discretization scheme we use must preserve this property. The modified trapezoidal scheme below satisfies this requirement if the granularity is fine enough for two consecutive values of $\mathbf{U}(z)$ to be approximately equal.

$$\frac{\mathbf{B}(z+\Delta z,t)-\mathbf{B}(z,t)}{\Delta z} = \frac{\partial}{\partial t}\mathbf{U}(z)^{*}\,\mathbf{D}\mathbf{U}(z)\left[\frac{\mathbf{B}(z+\Delta z,t)+\mathbf{B}(z,t)}{2}\right] \tag{2.7}$$

Equation (2.7) implies,

$$\begin{aligned}\mathbf{B}(z+\Delta z,t) = &\left[\mathbf{I}-\tfrac{\Delta z}{2}\frac{\partial}{\partial t}\mathbf{U}(z)^{*}\,\mathbf{D}\mathbf{U}(z)\right]^{-1}\\ &\left[\mathbf{I}+\tfrac{\Delta z}{2}\frac{\partial}{\partial t}\mathbf{U}(z)^{*}\,\mathbf{D}\mathbf{U}(z)\right]\mathbf{B}(z,t)\end{aligned} \tag{2.8}$$

where $\mathbf{I}$ is the 2×2 identity matrix. This is the description of a linear time-invariant LTI system that relates the individual building blocks (sections) of an optical fiber in a recursive manner. Its transfer function can be written in the Laplace domain with the complex frequency variable, s, as

$$\mathbf{H}^{\mathbf{B}}_{\Delta z}(z,s) = \left[\mathbf{I}-s\tfrac{\Delta z}{2}\mathbf{U}(z)^{*}\,\mathbf{D}\mathbf{U}(z)\right]^{-1}\left[\mathbf{I}+s\tfrac{\Delta z}{2}\mathbf{U}(z)^{*}\,\mathbf{D}\mathbf{U}(z)\right] \quad , \tag{2.9}$$

where

$$\mathbf{B}(z+\Delta z,s) = \mathbf{H}^{\mathbf{B}}_{\Delta z}(z,s)\;\mathbf{B}(z,s) \tag{2.10}$$

and $\mathbf{B}(z,s)$ is the Laplace transform of $\mathbf{B}(z,t)$ with respect to t.

Keeping in mind that $\mathbf{U}(z)$ is unitary, the transfer function above can be written as

$$\mathbf{H}^{\mathbf{B}}_{\Delta z}(z,s) = \mathbf{U}(z)^{*}\begin{bmatrix}1 & 0\\ 0 & \frac{1-s\frac{\Delta z}{2}\Delta\beta}{1+s\frac{\Delta z}{2}\Delta\beta}\end{bmatrix}\mathbf{U}(z) \quad . \tag{2.11}$$

It is easily verified that this transfer function is lossless since the entries of the 2×2 matrix in the middle are all-pass systems. At this point we can build a multi-section

system of length L by cascading one-section (not necessarily of equal length) transfer functions in the form of (2.11).

$$\mathbf{H^B}\left(L,s\right)=\prod_{k=0}^{N-1}\left\{\mathbf{U}\left(z_k\right)^*\begin{bmatrix}1 & 0\\ 0 & \frac{1-s\frac{\Delta z_k}{2}\Delta\beta}{1+s\frac{\Delta z_k}{2}\Delta\beta}\end{bmatrix}\mathbf{U}\left(z_k\right)\right\} \tag{2.12}$$

where

$$\mathbf{B}\left(L,s\right)=\mathbf{H^B}\left(L,s\right)\ \mathbf{B}\left(0,s\right) \tag{2.13}$$

and $\Delta z_k = z_{k+1} - z_k$, with $z_0 = 0$ and $z_N = L$. With these transfer functions, the expression for the Laplace transform of the complex envelopes of the two orthogonal polarization components, $\mathbf{A}\left(z,t\right)$, becomes

$$\mathbf{H^A}\left(L,s\right)=\mathbf{U}\left(L\right)\ \mathbf{H^B}\left(L,s\right)\ \mathbf{U}\left(0\right)^* \tag{2.14}$$

where

$$\mathbf{A}\left(L,s\right)=\mathbf{H^A}\left(L,s\right)\ \mathbf{A}\left(0,s\right)\quad . \tag{2.15}$$

We note that

$$\mathbf{H^B}\left(L,s=0\right)=\mathbf{I},\qquad \mathbf{H^A}\left(L,s=0\right)=\mathbf{U}\left(L\right)\ \mathbf{U}\left(0\right)^*\quad . \tag{2.16}$$

Since the two transfer functions $\mathbf{H^A}\left(L,s\right)$ and $\mathbf{H^B}\left(L,s\right)$ are linked with a frequency independent transformation, we will treat $\mathbf{H^B}\left(L,s\right)$ as the sole fiber transfer function and omit the superscript "$\mathbf{B}$" from now on.

The expression in (2.12) serves as a realistic computational model of an optical fiber if the number of sections, N, is high and its parameters are sampled uniformly and independently. In this case it can be used for simulations in order to determine the characteristics of a real fiber. In fact this model has the same structure with the commonly used fiber model consisting of concatenated birefringent sections with random phase retardations and relative orientations among each other. This can be seen from the fact that the all-pass transfer functions in (2.12) correspond to first-

order Pade approximations of an ideal continuous-time delay $\exp(-s\tau)$ [24]:

$$\exp\left(-s\,\Delta z\,\Delta\beta\right) \approx \frac{1 - s\frac{\Delta z}{2}\Delta\beta}{1 + s\frac{\Delta z}{2}\Delta\beta} \tag{2.17}$$

where

$$\tau = \Delta z\,\Delta\beta \tag{2.18}$$

Another way to express the overall fiber transfer function is to use the orthogonal columns of the unitary matrix $\mathbf{U}(z_k)^* = \mathbf{U}_k^* = [\hat{\mathbf{u}}_k\ \mathbf{u}_k]$ which resembles the form of the models that will be the subject of further discussion on reduced complexity PMD emulation models:

$$\mathbf{H}_k(s) = \mathbf{I} - \frac{2s}{s+\alpha_k}\mathbf{u}_k\mathbf{u}_k^* = (\mathbf{I} - 2\mathbf{u}_k\mathbf{u}_k^*) + \frac{2\alpha_k}{s+\alpha_k}\mathbf{u}_k\mathbf{u}_k^* \tag{2.19}$$

with $\alpha_k = 2/\tau_k$. The equivalence of both forms follows from the fact that $\{\frac{\alpha_k - s}{\alpha_k + s}, \mathbf{u}_k\}$ and $\{1, \mathbf{u}_k\}$ are eigenvalue-eigenvector pairs of $\mathbf{H}_k$. The losslessness of this system can be verified by testing that $\mathbf{H}_k(s)^*\mathbf{H}_k(s) = \mathbf{I}$ for all real s and because these transfer functions are rational matrix functions, they represent lumped continuous-time systems [25]. In the course of the further discussion this computational model of an optical fiber will be called the "full model".

2.3. Reduced Complexity Continuous-Time Model

A single birefringent section can be represented in Laplace domain as a 2×2 transfer function given in equation (2.19). Here $\alpha_k = 2/\tau_k$ is a real scalar parameter determining the phase retardation of the section with τ_k as the differential group delay of that section. This transfer function can be converted into two-input, two-output state-space form as

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}x_k(t) &= -\alpha_k x_k(t) + \mathbf{u}_k^* \mathbf{i}_k(t) \\ \mathbf{y}_k(t) &= 2\alpha_k \mathbf{u}_k x_k(t) + (\mathbf{I} - 2\mathbf{u}_k\mathbf{u}_k^*)\mathbf{i}_k(t) \quad ,\end{aligned} \tag{2.20}$$

with x_k as the scalar state-space variable and 2×1 input and output vectors $\mathbf{i}_k$ and $\mathbf{y}_k$ respectively.

The concatenation of N such sections can be represented in frequency domain with a product of transfer functions of the form (2.19) and in state-space form as

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(t) &= \mathrm{A}\mathbf{x}(t) + \mathrm{B}\mathbf{i}(t) \\ \mathbf{y}(t) &= \mathrm{C}\mathbf{x}(t) + \mathrm{D}\mathbf{i}(t) \quad ,\end{aligned} \tag{2.21}$$

where $x = [x_0\ x_1 \ldots x_{N-1}]^T$ and A, B are $N \times N$ and $N \times 1$ matrices respectively.

At this point, model reduction techniques can be employed to obtain a reduced dimension, smaller system

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\tilde{\mathbf{x}}(t) &= \tilde{\mathrm{A}}\tilde{\mathbf{x}}(t) + \tilde{\mathrm{B}}\mathbf{i}(t) \\ \mathbf{y}(t) &= \tilde{\mathrm{C}}\tilde{\mathbf{x}}(\mathbf{t}) + \tilde{\mathrm{D}}\mathbf{i}(t) \quad .\end{aligned} \tag{2.22}$$

This new system can then be transferred back to the Laplace domain as the product of degree-one lossless rational matrix transfer functions to obtain,

$$\mathbf{H}_k(s) = \mathbf{I} - \frac{2\alpha_k}{\alpha_k - j\beta_k}\mathbf{u}_k\mathbf{u}_k^* + \frac{2\alpha_k}{s + \alpha_k + j\beta_k}\frac{\alpha_k + j\beta_k}{\alpha_k - j\beta_k}\mathbf{u}_k\mathbf{u}_k^* \,, \tag{2.23}$$

where $j = \sqrt{-1}$, $\mathbf{u}^*\mathbf{u} = 1$, and α_k and β_k are real scalars. Equation (2.23) differs from (2.19) in that its pole is not required to be real. These poles of the individual sections will assume different locations on the complex plane to approximate the behavior of

the full model. In [18] experiments were conducted with a full model of 200 section and several reduced models with dimensions ranging from two to sixteen. The result of a Monte Carlo simulation, in which the reduction algorithm was applied to 40000 random realizations of the full model, showed that the poles of the reduced models are complex and their approximate locations are distributed on an half-ellipse centered at the origin of the complex plane. This ellipse corresponds to the pole locations of the Pade approximation of an ideal delay, e^{-Ts}, where T is the total maximum DGD obtainable from the full model.

This fact enabled the authors [18] to generate random realizations of the reduced model directly by holding those pole locations fixed, and randomly sampling the unit vector $\mathbf{u}_k$ in (2.23). It was also observed that sampling the unit vectors uniformly and independently among sections of the reduced model, as it was done for the full model, does not produce the correct DGD statistics, which are expected to be frequency independent. The required dependency structure for the sampling of $\mathbf{u}_k$ was introduced implicitly by using Markov chain Monte Carlo (MCMC) methods, specifically the Metropolis-Hastings algorithm. This technique was shown to result in the correct mean and standard deviation values for the DGD with small deviations from the desired values.

Later in Section 4.2 it will be shown that these results can be improved by introducing a modification to the Metropolis-Hastings algorithm. This modification not only produces better approximations for the mean and the standard deviation of the DGD but also closely matches its Maxwellian PDF.

2.4. Discrete Time Lossless Model

2.4.1. Paraunitary FIR Filters

A matrix transfer function $\mathbf{H}(z)$ is said to be paraunitary if the following holds[26]:

$$\mathbf{H}^*(z^{-*})\mathbf{H}(z) = c^2\mathbf{I}_r \quad \forall z. \tag{2.24}$$

Here, $\mathbf{I}_r$ is the $r \times r$ identity matrix and c is a scalar constant.

If all the entries of $\mathbf{H}$ are stable transfer functions and equation (2.24) is satisfied with $c = 1$, then $\mathbf{H}$ is unitary on the unit circle and therefore corresponds to a lossless LTI system. Conversely, it can be shown that for rational transfer functions, the unitariness of $\mathbf{H}(e^{j\omega})$ for all ω implies (2.24) with $c = 1$. Hence a lossless system can be defined as a causal, stable paraunitary system [27]. Since FIR filters are inherently stable, from now on the terms "lossless" and "paraunitary" will be used interchangeably for causal systems.

A general $M \times M$ degree-one lossless transfer function can be written as [27],

$$\mathbf{H}_i(z) = [\mathbf{I} - \mathbf{v}_i\mathbf{v}_i^* + z^{-1}\mathbf{v}_i\mathbf{v}_i^*]\mathbf{H}_i(1) \quad . \tag{2.25}$$

Here, $\mathbf{v}_i$ is a complex $M \times 1$ vector of unit norm and $\mathbf{H}_i(1)$ is a unitary matrix. It can easily be verified that this transfer function is paraunitary. The extension to higher degree systems follows naturally with the concatenation of such degree-one blocks and a multiplication with a constant unitary matrix $\mathbf{R}$.

$$\mathbf{H}(z) = \mathbf{H}_1(z)\mathbf{H}_2(z)\ldots\mathbf{H}_N(z)\mathbf{R} \tag{2.26}$$

Note that every lossless transfer function can be expressed in this form and this factorization is not unique.

In PMD related calculations we will be dealing with 2×2 paraunitary transfer functions that act on 2×1 vector valued time series consisting of two orthogonal polarization directions of the transmitted signal. In this regard the DGD can be defined using the expression in (2.26). The distance between the fast and slow principal states of polarisation (PSP) can be expressed as the difference [3]

$$\tau(\omega) = |\Im\{\lambda_1[\mathbf{G}(e^{j\omega})] - \lambda_2[\mathbf{G}(e^{j\omega})]\}| \quad ,$$

where $\Im\{\lambda_i[\mathbf{G}(e^{j\omega})]\}$ denotes the imaginary part of the i$^{\text{th}}$ eigenvalue of $\mathbf{G}(e^{j\omega})$ and,

$$\mathbf{G}(e^{j\omega}) = \frac{\mathrm{d}\mathbf{H}(e^{j\omega})}{\mathrm{d}\omega}\mathbf{H}^*(e^{j\omega}). \tag{2.27}$$

If we now substitute equation (2.26) in (2.27) we get

$$\mathbf{G}(e^{j\omega}) = -j\sum_{i=1}^{N}\prod_{j=1}^{i}\mathbf{H}_{j-1}\mathbf{v}_i\mathbf{v}_i^*\mathbf{H}_{j-1}^* \quad , \tag{2.28}$$

with $\mathbf{H}_0 = \mathbf{I}_2$. This expression can be greatly simplified for $\omega = 0$ if we set $\mathbf{H}_k(1) = \mathbf{I}_2$ for all k. The DGD of the fiber, $\tau(\omega)$, evaluated at the center frequency can in this case be computed to be the difference between the two eigenvalues of the positive definite matrix $\mathbf{V}\mathbf{V}^*$ with $\mathbf{V} = [\mathbf{v}_1\mathbf{v}_2\ldots\mathbf{v}_N]$. Calculating the eigenvalues of this matrix results in a quadratic equation, and twice the discriminant of this equation gives us the difference between the two eigenvalues.

$$\tau(0) = \sqrt{(a-b)^2 + 4cc^*} \tag{2.29}$$

where,

$$j\mathbf{G}(1) = \begin{bmatrix} a & c \\ c^* & b \end{bmatrix}, \qquad \mathbf{v}_i = \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix}, \qquad \begin{array}{l} a = \sum_{i=1}^{N}|\alpha_i|^2 \\ b = \sum_{i=1}^{N}|\beta_i|^2 = N - a \\ c = \sum_{i=1}^{N}\alpha_i\beta_i^* \end{array} \quad . \tag{2.30}$$

2.4.2. DGD Statistics of Paraunitary Filters

As a reference point and motivation for further discussion, we investigate the statistical behavior of an ensemble of paraunitary filters when their parameters are sampled uniformly and independently. The structure of paraunitary FIR filters as concatenation of degree-one building blocks is similar to the continuous time lumped model of optical fibers discussed in Section 2.2 and their DGD distribution, therefore, resembles a real fiber. However, two important hurdles limit this resemblance: The mean DGD of filters tailored with this selection method is determined only by the

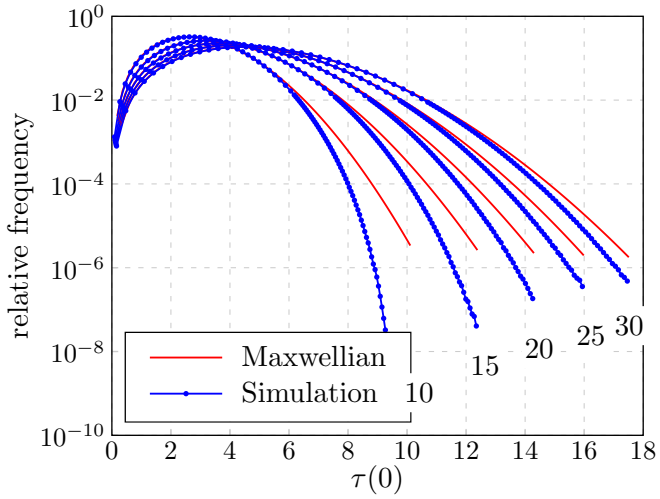


Figure 2.1: Probability density functions of the DGD for various number of sections (numbers at the end of the curves) for i.i.d. uniform filter parameters.

number of degree-one sections and is not adjustable which is a crucial feature for a PMD emulator. Furthermore, the probability density functions deviate heavily from the desired curves particularly in the tails which correspond to high DGD cases and are consequently of primary importance for system outage probability considerations.

The results for the DGD distribution are given in Figure 2.1 for the uniform i.i.d. parameter sampling method:

$$\alpha_i = \sqrt{X_i} e^{j2\pi Y_i} \qquad \beta_i = \sqrt{1 - |\alpha_i|^2} = \sqrt{1 - X_i} \quad ,$$

where X_i and Y_i are uniformly distributed in $[0, 1]$ and independent among themselves as well as among each other. The distribution of the DGD has been computed for different number of sections, $N = 10$ to $N = 30$ with 10^9 samples and compared with the expected Maxwellian distributions with the same mean as the generated data set. Although the PDF deviation becomes less prominent as N increases, the impact of a high number of sections on the computational complexity of the emulator renders this property unusable for our purposes.

Together with equation (2.29), (2.30) and the parameter selection method above

the expression for the DGD becomes

$$\begin{aligned}\tau &= \sqrt{(a-b)^2 + 4cc^*} \\ &= \sqrt{(2a-N)^2 + (2\Re\{c\})^2 + (2\Im\{c\})^2} \qquad (2.31)\\ &= \left[\left(2\sum_{i=1}^{N} X_i - N\right)^2 + \left(2\sum_{i=1}^{N}\sqrt{X_i - X_i^2}\cos(2\pi Y_i)\right)^2 + \left(2\sum_{i=1}^{N}\sqrt{X_i - X_i^2}\sin(2\pi Y_i)\right)^2\right]^{\frac{1}{2}}\end{aligned}$$

Remembering that the Maxwellian distribution can be expressed as the square root of sum of squares of three independent zero-mean Gaussian random variables with the same variance, the above expression can be used to verify that the distribution of τ indeed approximates a Maxwellian distribution. This is achieved invoking the central limit theorem on each of the three summations of length N. The first term, $2\sum_{i=1}^{N} X_i - N$, is zero-mean. Because X_i are independent the total random variable $2a - N$ has the sum of variances of X_i as its variance.

$$\mathrm{var}(2a - N) = 4N\mathrm{var}(X_i - \frac{1}{2}) = \frac{N}{3}$$

It is again straightforward to calculate the mean and the variance of the second and the third term of the sum. Because X_i are independent among each other and also independent of Y_i one can see that $\Re\{c\}$ is again zero mean because of the symmetry of the PDF of $\cos(2\pi Y_i)$ around zero. Furthermore the variance of $2\Re\{c\}$ can be computed as

$$\mathrm{var}(2\Re\{c\}) = N\mathrm{var}\left(2\sqrt{X_i - X_i^2}\right)\mathrm{var}(\cos(2\pi Y_i)) = 4N\left(\frac{1}{2} - \frac{1}{3}\right)\frac{1}{2} = \frac{N}{3} \quad .$$

The same applies for $2\Im\{c\}$ and hence the total random variable τ consists of three approximately normally distributed random variables with $\mu \approx 0$ and $\sigma^2 \approx N/3$. This enables us to compute the expected mean of τ as

$$\bar{\tau} \approx \sqrt{\frac{8}{\pi}}\sigma = \sqrt{\frac{8}{3\pi}N} \quad . \qquad (2.32)$$

These results are valid for the center frequency, $\omega = 0$ but can be extended to other frequencies. Further investigation reveals that this extension is justified for uniform i.i.d. parameter selection since there is no bias for any specific frequency point.

More in depth discussion of innate statistical properties of PMD emulators with i.i.d. uniform parameters including the derivation of the exact DGD distribution and examination of the deviation from the Maxwellian is outside of the scope of this work, but can be found in [28].

2.5. Conclusion

As the first step towards PMD emulation, we have presented three different models of an optical fiber suffering from PMD. The first one, the full model, serves as a mathematical abstraction of a real fiber. Its complicated structure due to its high number of concatenated birefringent sections enables this model to capture the statistical properties of PMD perfectly without any specially tailored parameter sampling scheme, while rendering its use as a PMD emulator impossible because of its high computational complexity.

The second model, the continuous-time reduced complexity model, springs from the necessity to reduce the model complexity of the full model. This is accomplished with a model order reduction scheme using Pade approximations and Krylov subspace techniques. The effect of this transformation on the model structure is that the poles of individual sections become complex-valued as opposed to equal-valued real poles of the model which in turn results in sections being of different lengths. For an ensemble of different reduced complexity models obtained by applying model reduction to a batch of full models, these poles are distributed approximately on a half ellipse on the LHS of the complex plane that represents the pole-locations of the Pade approximation of an ideal delay. This fact will be used as the basis for the parameter sampling method for this model to obtain the desired statistical properties.

The third model we presented constitutes the main contribution of this work.

The lossless nature of an optical fiber with PMD enables one to use discrete-time paraunitary FIR filters to represent such a system in a form that can be implemented on DSPs. These filters are built with two-input two-output sections of equal delays. When its parameters are sampled uniformly and independently, such a structure exhibits statistical properties close to those of the full model but suffers from an unadjustable mean DGD value that only depends on the number of sections and large deviations of the PDFs from the desired curves in their tail regions. Methods to remedy this discrepancy will be the main subject of the remainder of this work.

Chapter 3

Random Input Sampling for Complex Models Using Markov Chain Monte Carlo

3.1. Introduction

Most algorithms employed to sample from complicated probability distributions such as rejection sampling and importance sampling assume full knowledge of the target density [29]. Contrary to these approaches Markov Chain Monte Carlo (MCMC) methods can be used to sample from distributions for which the form of the density function is known, but the function value itself can only be evaluated up to a scalar constant. The versatility of MCMC algorithms make them powerful tools for complicated sampling problems. In this chapter we develop a general method for later implementation in the parameter sampling problem for PMD emulators.

MCMC algorithms devise a Markov chain on the sample space of a general vector random variable. In typical settings the probability density function of the distribution can only be evaluated up to a normalizing constant. The common Metropolis algorithm [30] starts with an initial state and generates samples of the random variable iteratively. At every step of the procedure a new state is proposed according to some proposal distribution. This proposal state is then accepted with a probability

determined by the ratio of the PDF values for the new state and the old state. Because the accept-reject rule only requires the evaluation of the ratio of the probability densities for the proposed and the old state, it is sufficient to know the target PDF up to a scalar constant. The sole restriction of the Metropolis algorithm is that the proposal density is symmetric and simple enough to sample directly.

One generalization of the Metropolis algorithm is the Metropolis-Hastings algorithm [31] which can employ non-symmetric proposal densities. To achieve this, the acceptance probability is modified to incorporate a ratio of the proposal density values.

MCMC methods are very general tools in regard to dealing with intractable probabilistic settings. This generality allows MCMC to be integrated into many practical problems in diverse fields like computational biology [32], statistical physics [33], random number generation [34, 35], artificial intelligence [36] and many more. A review for the applications of MCMC can be found in [37].

One of these broad applications deals with the model selection problem where one tries to choose a model among many competing models that is more likely to have generated the given probabilistic output data [38]. In this chapter we investigate a related problem in which the model that generates the given data is fixed, but accepts a random input with an unknown PDF. This situation arises naturally in the context of complex models which accept random inputs either as the data to be processed or as random model parameters. The realization of these models in practice can be in various ways such as lengthy and complicated computer routines, a set of involved mathematical equations or any kind of black box evaluation. Nevertheless they can be viewed as a mapping of random variables as illustrated in figure 3.1. If $h : \mathbb{R}^n \to \mathbb{R}^m$ is a multi-valued function of multiple variables with $\mathbf{Y} = \mathbf{h}(\mathbf{X})$ and its inverse h^{-1} does not exist or cannot be computed analytically then the question arises: How does one choose $\mathbf{X}$, for $\mathbf{Y}$ to have the desired probability density $f_{\mathbf{Y}d}$?

The answer to this question is not straightforward. First of all, generally h is very complicated and all we have at our disposal is some kind of routine that evaluates

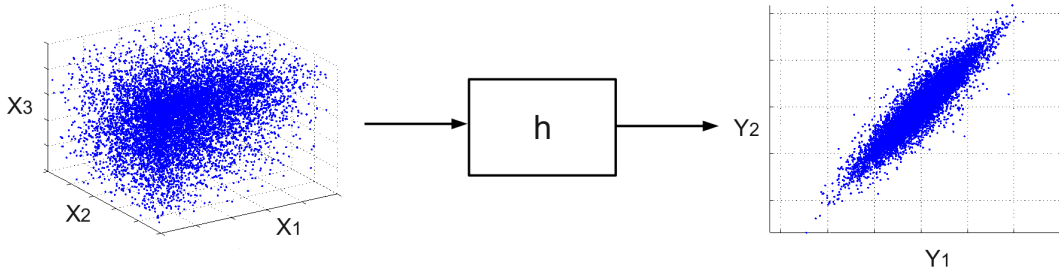


Figure 3.1: Graphical representation of a deterministic mapping with random input and output. The output of the complex system h, $\mathbf{Y} = h(\mathbf{X})$, is a random vector of dimension two while the input $\mathbf{X}$ is a random vector with three dimensions.

it for a given input. Therefore the unknown density of $\mathbf{X}$ can be computed through the inverse mapping only in special cases where h is known explicitly and $m = n$ with the Jacobian of h globally invertible. The use of standard sampling algorithms including MCMC to sample $\mathbf{X}$ is for this reason not possible. Additionally more than one input distribution can generate the desired output distribution creating an issue of non-uniqueness.

In order to address this problem we first review some Markov chain and MCMC theory then we provide a detailed description of the problem at hand and a toy example to demonstrate the concept before proceeding to develop a solution. Finally we conclude our discussion with a numerical example of a stochastic differential equation (SDE) and demonstrate how our method can be used to sample from the space of solution paths to this equation.

3.2. Background

In this section we present some elementary Markov chain and MCMC theory which will be required for later discussion.

Definition 2. *A sequence of indexed random variables* $\mathbf{X}_i$, $i \in \mathbb{N}$ *is called a Markov*

chain if the following property holds for every measurable set $\mathcal{A} \subseteq \mathbb{R}^n$

$$\Pr(\mathbf{X}_{i+1} \in \mathcal{A} | \mathbf{X}_i, \mathbf{X}_{i-1}, \ldots, \mathbf{X}_0) = \Pr(\mathbf{X}_{i+1} \in \mathcal{A} | \mathbf{X}_i) \ . \tag{3.1}$$

$T(\mathbf{x}, \mathcal{A}) = \Pr(\mathbf{X}_{i+1} \in \mathcal{A} | \mathbf{X}_i = \mathbf{x})$ *is called its transition kernel with the transition density* $\tau(\mathbf{x}, \mathbf{x}')$*, where*

$$T(\mathbf{x}, \mathcal{A}) = \int_{\mathcal{A}} \tau(\mathbf{x}, \mathbf{x}') \mathrm{d}\mathbf{x}', \qquad \mathbf{x} \in \mathbb{R}^n \tag{3.2}$$

Definition 3. $f(\mathbf{x})$ *is called a stationary distribution of the Markov chain if it satisfies*

$$f(\mathbf{x}') = \int_{\mathbb{R}^n} f(\mathbf{x}) \tau(\mathbf{x}, \mathbf{x}') \mathrm{d}\mathbf{x} \ . \tag{3.3}$$

Lemma 1. *A sufficient condition for* $f(\mathbf{x})$ *to be a stationary distribution of the Markov chain* $\mathbf{X}_i$ *is the detailed balance equation:*

$$f(\mathbf{x}) \tau(\mathbf{x}, \mathbf{x}') = f(\mathbf{x}') \tau(\mathbf{x}', \mathbf{x}), \qquad \mathbf{x}, \mathbf{x}' \in \mathbb{R}^n \ . \tag{3.4}$$

Proof. Integrating both sides of equation (3.4) and using definition 3,

$$\begin{aligned} \int_{\mathbb{R}^n} f(\mathbf{x}) \tau(\mathbf{x}, \mathbf{x}') \mathrm{d}\mathbf{x} &= \int_{\mathbb{R}^n} f(\mathbf{x}') \tau(\mathbf{x}', \mathbf{x}) \mathrm{d}\mathbf{x} \\ &= f(\mathbf{x}') \int_{\mathbb{R}^n} \tau(\mathbf{x}', \mathbf{x}) \mathrm{d}\mathbf{x} \\ &= f(\mathbf{x}') \end{aligned} \tag{3.5}$$

□

Note that lemma 1 gives a sufficient condition, and the necessary condition is much looser [39]. Furthermore under certain conditions the stationary distribution is unique [40].

Above definitions and lemma 1 are sufficient to describe the Metropolis-Hastings

(and the Metropolis algorithm as its special case) in a formal way. Given a target distribution $f(\mathbf{x})$ for the random vector $\mathbf{X}$ the strategy of the algorithm is to construct a Markov chain on the state space of interest, $\mathbb{R}^n$, and choose a transition kernel such that the Markov chain has $f(\mathbf{x})$ as its stationary distribution. This is accomplished in two stages. At the first stage the procedure takes a random step in the state space according to some proposal density $p(\mathbf{x}, \mathbf{x}')$ which describes the probability of moving from the state, $\mathbf{X}_i = \mathbf{x}$, to the next one, $\mathbf{X}_{i+1} = \mathbf{x}'$. Most common choice for p uses a form of increment on $\mathbf{x}$ such that $\mathbf{x}' = \mathbf{x} + \Delta\mathbf{x}$. Commonly used densities for the random increment $\Delta\mathbf{x}$ are tractable ones like the uniform and the Gaussian density. At the second stage of the algorithm a decision is made whether the chain will advance to $\mathbf{x}'$ as its next state or stay at $\mathbf{x}$. The decision mechanism uses the ratio $\frac{f(\mathbf{x}')}{f(\mathbf{x})}$ in the decision rule

$$\alpha(\mathbf{x}, \mathbf{x}') = \min\left(1, \frac{f(\mathbf{x}')}{f(\mathbf{x})}\frac{p(\mathbf{x}', \mathbf{x})}{p(\mathbf{x}, \mathbf{x}')}\right) \tag{3.6}$$

which gives the acceptance probability of the proposed move. After evaluating the accept-reject ratio, a random number u is sampled according to a standard uniform distribution and the move is accepted if $u \leq \alpha(\mathbf{x}, \mathbf{x}')$. If the proposed state is not accepted the Markov chain remains in its previous state. Theorem 1 shows that the distribution of the samples taken in this way indeed converges to $f(\mathbf{x})$.

Theorem 1. *The transition kernel of the Metropolis-Hastings algorithm satisfies the detailed balance condition and $f(\mathbf{x})$ is the stationary distribution of the resulting Markov chain.*

Proof. The transition kernel $T(\mathbf{x}, \mathcal{A})$ can be written as the sum of two probabilities: The probability of an accepted step to a point $\mathbf{x}'$ in $\mathcal{A}$ and the probability of a rejection while the point $\mathbf{x}$ lies in $\mathcal{A}$.

$$T(\mathbf{x}, \mathcal{A}) = \int_{\mathcal{A}} p(\mathbf{x}, \mathbf{x}')\alpha(\mathbf{x}, \mathbf{x}')\mathrm{d}\mathbf{x}' + 1_{\{\mathbf{x}\in\mathcal{A}\}} \int_{\Omega} p(\mathbf{x}, \mathbf{x}')(1 - \alpha(\mathbf{x}, \mathbf{x}'))\mathrm{d}\mathbf{x}'$$

Hence the transition density is given by

$$\tau(\mathbf{x}, \mathbf{x}') = p(\mathbf{x}, \mathbf{x}')\alpha(\mathbf{x}, \mathbf{x}') + \delta_{\mathbf{x}}(\mathbf{x}')r(\mathbf{x}) \ ,$$

where $\delta_{\mathbf{x}}(\mathbf{x}')$ is the point mass at $\mathbf{x}$ and $r(\mathbf{x}) = 1 - \int_{\Omega} \alpha(\mathbf{x}, \mathbf{x}')p(\mathbf{x}, \mathbf{x}')\mathrm{d}\mathbf{x}'$ is the probability that the chain does not leave its current position $\mathbf{x}$.

Lemma 1 gives us a way of checking whether this transition kernel has the desired PDF as its stationary distribution. If we now check if equation (3.4) is satisfied we find for the first summand of the transition density

$$\begin{aligned} f(\mathbf{x})p(\mathbf{x}, \mathbf{x}')\alpha(\mathbf{x}, \mathbf{x}') &= f(\mathbf{x})p(\mathbf{x}, \mathbf{x}')\min\left(1, \frac{f(\mathbf{x}')}{f(\mathbf{x})}\frac{p(\mathbf{x}', \mathbf{x})}{p(\mathbf{x}, \mathbf{x}')}\right) \\ &= \min\left(f(\mathbf{x})p(\mathbf{x}, \mathbf{x}'), f(\mathbf{x}')p(\mathbf{x}', \mathbf{x})\right) \\ &= \min\left(\frac{f(\mathbf{x})p(\mathbf{x}, \mathbf{x}')}{f(\mathbf{x}')p(\mathbf{x}', \mathbf{x})}, 1\right) f(\mathbf{x}')p(\mathbf{x}', \mathbf{x}) \\ &= f(\mathbf{x}')p(\mathbf{x}', \mathbf{x})\alpha(\mathbf{x}', \mathbf{x}) \quad . \end{aligned} \tag{3.7}$$

Finally for the second summand the requirement is trivially satisfied

$$\delta_{\mathbf{x}}(\mathbf{x}')r(\mathbf{x}) = \delta_{\mathbf{x}'}(\mathbf{x})r(\mathbf{x}') \quad ,$$

and this completes the proof. □

Further discussion of Markov chain and MCMC theory is outside the scope of this work but excellent material on this subject can be found in [41] and [42].

3.3. An Illustrative Example

Now let us recap the problem described in the introduction. Suppose we are given a general many-to-one, non-isometric map $h : \mathbb{R}^n \to \mathbb{R}^m$ which maps a random vector $\mathbf{X}$ to another random vector $\mathbf{Y} = h(\mathbf{X})$ where $\mathbf{X} \in \mathbb{R}^n$, $\mathbf{Y} \in \mathbb{R}^m$ and let $f_{\mathbf{Y}d}$ be the desired probability density of $\mathbf{Y}$. Given $f_{\mathbf{Y}d}$ how must $f_{\mathbf{X}}$ be chosen such that the

transformed variable $\mathbf{Y} = h(\mathbf{X})$ has the desired PDF?

Given this setting one might be tempted to construct a Markov chain in the space of the input variables to sample $\mathbf{X}$ while evaluating the accept-reject rule probabilities of the Metropolis-Hastings algorithm in the space of the random output vector $\mathbf{Y}$.

As the following toy-example illustrates, this method does not result in a Markov chain in the space of input variables with the desired stationary distribution $f_{\mathbf{Y}d}$ of the output variables.

Figure 3.2 describes a discrete state space consisting of three states $\mathcal{A} = \{X_1, X_2, X_3\}$. The arrows represent a many-to-one function $h : \mathcal{A} \to \mathcal{B}$ with $h(X_1) = Y_1$ and $h(X_2) = h(X_3) = Y_2$ and $Y_1, Y_2 \in \mathcal{B}$. It is of no importance if $\mathcal{B}$ is discrete or continuous but the range of h is discrete for obvious reasons. The desired distribution of Y is $f_{Yd}(Y_1) = 0.9$ and $f_{Yd}(Y_2) = 0.1$. The results about Markov chains and the Metropolis-Hastings algorithm given in section 3.2 can easily be adopted to general finite state spaces and to this specific example.

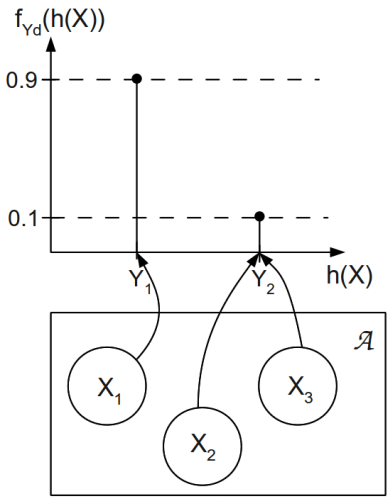


Figure 3.2: A toy-example to illustrate the problem of mapping the state variables X to another random variable Y with the desired probability distribution f_{Yd}.

Now suppose that we are running the Metropolis algorithm on $\mathcal{A}$ with a symmetric proposal distribution P. For the current state X_i a new state is proposed according to the rule

$$\Pr(X_j|X_i) = P(X_i, X_j) = \begin{cases} \frac{1}{2} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} .$$

Together with the accept-reject rule of the Metropolis algorithm, this results in a Markov chain with the transition probabilities $T(X_i, X_j) = P(X_i, X_j)\min(1, \frac{f_{Yd}(h(X_j))}{f_{Yd}(h(X_i))})$ for $i \neq j$ and the transition probability matrix

$$\mathbf{T} = \begin{pmatrix} \frac{16}{18} & \frac{1}{18} & \frac{1}{18} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \tag{3.8}$$

The left eigenvector of this matrix that corresponds to the eigenvalue 1 gives us the stationary distribution of the chain, which is $(\frac{9}{11}, \frac{1}{11}, \frac{1}{11})$. It can be easily seen that this distribution does not provide the desired stationary distribution on the range of h. In fact this distribution corresponds to a function h' which maps X_3 to a different value $h'(X_3)$ which has the same probability as $h'(X_2)$. This behaviour can also be observed with the more general Metropolis-Hastings algorithm by choosing a non-symmetric proposal distribution and applying the corresponding accept-reject rule.

This toy-example illustrates clearly that to address the problem of creating the target probability density $f_{\mathbf{Y}d}$ we have to take the properties of the mapping h into account and modify the Metropolis-Hastings algorithm accordingly.

3.4. Modification of MCMC with a Probing Term

The reason that the above example fails to converge to the desired stationary distribution f_{Yd} lies within the properties of the general mapping h. First h is not one-to-one and hence the probability of a state Y_i appearing in the chain on $\mathcal{B}$ depends on the probability of all the states X_j on the space of inputs for which $h(X_j) = Y_i$ holds. Additionally for the continuous case, even if h was one-to-one it would not necessarily be an isometry so that volumes are distorted under the mapping creating a similar effect on the stationary distribution. In this section we develop a method to overcome the shortcomings of MCMC sampling for the problem described in the previous section.

In this context for the general case we first implement a probing procedure for the mapping h by using the output distribution that results when the input parameters are sampled uniformly and independently. Then we show that a modification of the

target density with this uniform output density can be used in the space of parameters for the accept-reject rule in MCMC to achieve the desired density $f_{\mathbf{Y}d}$ on the range of h.

Theorem 2. *Let* $\mathbf{U}$ *be a uniform random vector on the probability space* $(\Omega,\ \mathcal{F},\ F_{\mathbf{U}})$ *where* Ω *is a bounded subset of* $\mathbb{R}^n$ *such that* $f_{\mathbf{U}}(\mathbf{u}_1) = f_{\mathbf{U}}(\mathbf{u}_2)$ *for all* $\mathbf{u}_1, \mathbf{u}_2 \in \Omega$ *with* $f_{\mathbf{U}}$ *as the probability density function of the cumulative distribution function* $F_{\mathbf{U}}$*. And let* $h : \mathbb{R}^n \to \mathbb{R}^m$ *be a mapping satisfying the required regularity conditions such that* $(\Omega',\ \mathcal{F}')$ *with* $\Omega' = h(\Omega)$ *is the induced sample space by* h *and the associated* σ*-algebra. Then a random variable* $\mathbf{X} \in \mathbb{R}^n$ *constructs another random variable* $\mathbf{Y} = h(\mathbf{X}),\ \mathbf{Y} \in \mathbb{R}^m$ *with the desired probability density* $f_{\mathbf{Y}d}$ *if* $\mathbf{X}$ *has the unnormalized probability density*

$$f_{\mathbf{X}}(\mathbf{x}) \propto \frac{f_{\mathbf{Y}d}(h(\mathbf{x}))}{f_{\mathbf{Q}}(h(\mathbf{x}))}$$

where $f_{\mathbf{Q}}$ *is the probability density of the transformed random variable* $\mathbf{Q} = h(\mathbf{U})$*.*

Proof. Consider the bounded sample space $\Omega \subseteq \mathbb{R}^n$, $\Omega = [\alpha_1,\ \beta_1] \times [\alpha_2,\ \beta_2] \times \cdots \times [\alpha_n,\ \beta_n]$ in which we assume $f_{\mathbf{Q}}$ is strictly positive. For the cumulative distribution function of $\mathbf{Q}$ we get

$$F_{\mathbf{Q}}(\mathbf{q}) = \Pr(\mathbf{Q} \leq \mathbf{q}) = \Pr(\{\mathbf{u} : \mathbf{u} \in \Omega,\ h(\mathbf{u}) \leq \mathbf{q}\}) \quad . \tag{3.9}$$

which can be written with the indicator function as

$$F_{\mathbf{Q}}(\mathbf{q}) = \int_{\Omega} 1_{\{\mathbf{u}:h(\mathbf{u})\leq\mathbf{q}\}} f_{\mathbf{U}}(\mathbf{u}) \mathrm{d}\mathbf{u} \tag{3.10}$$

Note that the indicator function in equation (3.10) can be expressed with the components of the random vector $\mathbf{q}$ and the function h as

$$1_{\{\mathbf{u}:h(\mathbf{u})\leq\mathbf{q}\}} = s(q_1 - h_1(\mathbf{u}))s(q_2 - h_2(\mathbf{u})) \ldots s(q_m - h_m(\mathbf{u})) \tag{3.11}$$

where s is the unit step function.

The PDF of $\mathbf{Q}$ is given by $f_{\mathbf{Q}}(\mathbf{q}) = \frac{\partial^m F_{\mathbf{Q}}(\mathbf{q})}{\partial q_1 \partial q_2 \ldots \partial q_m}$. Using generalized functions and equation (3.11) we can write this expression as

$$\begin{aligned} f_{\mathbf{Q}}(\mathbf{q}) &= \int_{\Omega} \delta(q_1 - h_1(\mathbf{u}))\delta(q_2 - h_2(\mathbf{u})) \ldots \delta(q_m - h_m(\mathbf{u})) f_{\mathbf{U}}(\mathbf{u}) \mathrm{d}\mathbf{u} \\ &\propto \int_{\alpha_1}^{\beta_1} \int_{\alpha_2}^{\beta_2} \ldots \int_{\alpha_n}^{\beta_n} \delta(q_1 - h_1(\mathbf{u}))\delta(q_2 - h_2(\mathbf{u})) \ldots \delta(q_m - h_m(\mathbf{u})) \mathrm{d}\mathbf{u} \end{aligned} \quad (3.12)$$

If we now set the distribution of the input random variable $\mathbf{X}$ proportional to the ratio of the desired distribution of $\mathbf{Y}$ and the distribution of $\mathbf{Q}$,

$$f_{\mathbf{X}}(\mathbf{x}) \propto \frac{f_{\mathbf{Y}d}(h(\mathbf{x}))}{f_{\mathbf{Q}}(h(\mathbf{x}))} \quad (3.13)$$

we have for the cumulative distribution function of $\mathbf{Y} = h(\mathbf{X})$

$$\begin{aligned} F_{\mathbf{Y}}(\mathbf{y}) &= \Pr(\mathbf{Y} \leq \mathbf{y}) \\ &= \Pr(\{\mathbf{x} : \mathbf{x} \in \Omega,\ h(\mathbf{x}) \leq \mathbf{y}\}) \\ &= \int_{\Omega} 1_{\{\mathbf{x}: h(\mathbf{x}) \leq \mathbf{y}\}} f_{\mathbf{X}}(\mathbf{x}) \mathrm{d}\mathbf{x} \\ &= \int_{\Omega} 1_{\{\mathbf{x}: h(\mathbf{x}) \leq \mathbf{y}\}} \frac{f_{\mathbf{Y}d}(h(\mathbf{x}))}{f_{\mathbf{Q}}(h(\mathbf{x}))} \mathrm{d}\mathbf{x} \quad . \end{aligned} \quad (3.14)$$

Finally we have for the output probability density,

$$
\begin{aligned}
f_{\mathbf{Y}}(\mathbf{y}) &= \frac{\partial^m}{\partial y_1 \partial y_2 \dots \partial y_m} F_{\mathbf{Y}}(\mathbf{y}) \\
&\propto \int_{\alpha_1}^{\beta_1} \int_{\alpha_2}^{\beta_2} \dots \int_{\alpha_n}^{\beta_n} \delta(y_1 - h_1(\mathbf{x}))\delta(y_2 - h_2(\mathbf{x})) \dots \delta(y_m - h_m(\mathbf{x})) \frac{f_{\mathbf{Y}d}(h(\mathbf{x}))}{f_{\mathbf{Q}}(h(\mathbf{x}))} \mathrm{d}\mathbf{x} \\
&\propto \frac{f_{\mathbf{Y}d}(\mathbf{y})}{f_{\mathbf{Q}}(\mathbf{y})} \underbrace{\int_{\Omega} \delta(\mathbf{y} - h(\mathbf{x})) \mathrm{d}\mathbf{x}}_{f_{\mathbf{Q}}(\mathbf{y})} \\
&\propto f_{\mathbf{Y}d}(\mathbf{y}) \qquad (3.15)
\end{aligned}
$$

Since both are normalized probability densities $f_{\mathbf{Y}} = f_{\mathbf{Y}d}$ holds and the distribution of the image of samples on $\mathbb{R}^n$ will be equal to the desired distribution on $\mathbb{R}^m$. □

Theorem 2 shows that we can find the distribution of a random variable $\mathbf{X}$ that gives us the desired density $f_{\mathbf{Y}d}$ through the mapping h provided that we know the uniform input distribution $f_{\mathbf{Q}}$. This can be accomplished by modifying the Metropolis-Hastings accept-reject rule in (3.6) as

$$
\alpha(\mathbf{x}, \mathbf{x}') = \min\left(1, \frac{f_{\mathbf{Y}d}(h(\mathbf{x}'))}{f_{\mathbf{Y}d}(h(\mathbf{x}))} \frac{p(\mathbf{x}', \mathbf{x})}{p(\mathbf{x}, \mathbf{x}')} \frac{f_{\mathbf{Q}}(h(\mathbf{x}))}{f_{\mathbf{Q}}(h(\mathbf{x}'))}\right) \quad . \qquad (3.16)
$$

Note that Theorem 2 assumes a bounded support for the uniformly sampled random vector $\mathbf{Q}$, with $\Omega = [\alpha_1,\ \beta_1] \times [\alpha_2,\ \beta_2] \times \dots \times [\alpha_n,\ \beta_n]$. This assumption implies that the support of the input vector $\mathbf{X}$ is equal to or a subset of Ω. Therefore in case $\mathbf{X}$ has unbounded support, this technique will sample a truncated version of the input random vector. Nevertheless practical difficulties caused by this fact can be overcome with an adjustment of Ω which theoretically can be chosen arbitrarily large.

Furthermore Theorem 2 gives us only the unnormalized PDF which is sufficient to sample $\mathbf{X}$ with MCMC. But the above method can be used irrespective of the specific sampling method once this density is normalized. Hence we obtain a general method to control the input of complex systems with prescribed random outputs.

In practical applications one will not always be able to compute $f_{\mathbf{Q}}$ analytically.

In these situations $f_{\mathbf{Q}}$ will have to be substituted with an approximation $\hat{f}_{\mathbf{Q}}$. For this purpose one can use various density estimation schemes available. For large data sets nonparametric schemes like kernel density estimators and nearest neighbour methods [43] can be used. For other settings Bayesian schemes like the EM algorithm [44] can be employed for inference.

3.5. An Application: Stochastic Differential Equations

In this section we demonstrate an example for our algorithm on stochastic differential equations. In this case the model is given by a differential equation driven by random noise and the input random variable takes the form of the solution to this equation.

Consider the one dimensional Itō stochastic differential equation

$$\begin{aligned} &\mathrm{d}X_t = b(X_t, t)\mathrm{d}t + a(X_t, t)\mathrm{d}W_t \,, \qquad 0 \le t \le T \\ &X_0 = c \end{aligned} \tag{3.17}$$

where $a, b : \mathbb{R} \times [0, T] \to \mathbb{R}$ are measurable functions and W_t is the Wiener process.

A numerical treatment of this equation can be done by discretization using the simple Euler scheme.

$$X_{i+1} = X_i + b(X_i, t_i)\Delta t + a(X_i, t_i)\Delta W_i \,, \qquad 0 = t_0 \le t_1 \le \cdots \le t_N = T \tag{3.18}$$

We now set $\Delta X_i = X_i - X_{i-1}$ and define the random vectors $\Delta\mathbf{X} = (\Delta X_1, \Delta X_2, \ldots, \Delta X_N)^T$ and $\mathbf{Y} = h(\Delta\mathbf{X}) = (Y_1, Y_2, \ldots, Y_N)^T$ with

$$Y_i = \frac{\Delta X_i - b(X_{i-1}, t_{i-1})\Delta t}{a(X_{i-1}, t_{i-1})} = \Delta W_{i-1} \tag{3.19}$$

Note that $\Delta\mathbf{X}$ together with X_0 completely determines the sample path. Hence if

we can sample $\Delta\mathbf{X}$ such that it satisfies equation (3.18), that means we can generate a solution path to the stochastic differential equation. The distribution of $\Delta\mathbf{X}$ is unknown but we know that $\mathbf{Y}$ is a Gaussian random vector with i.i.d. zero mean components with variance Δt. Using this fact we can employ the modified MCMC algorithm to sample solution paths.

For this general class of stochastic differential equations we can obtain the PDF of the output vector when the input random variables are sampled uniformly. First we derive the expression of the joint output distribution for the uniformly sampled input variables. Let $\Delta\mathbf{U} \in \mathbb{R}^N$ be a random vector with i.i.d. components distributed uniformly in $[-\rho, \rho]$ and $\mathbf{Q} = h(\Delta\mathbf{U}) \in \mathbb{R}^N$ another random vector with the joint PDF $f_{\mathbf{Q}}$. We can express $f_{\mathbf{Q}}$ as the product of conditional PDFs as follows.

$$f_{\mathbf{Q}}(\mathbf{q}) = f_{\mathbf{Q}}(q_1, q_2, \ldots, q_N) = f(q_1)f(q_2|q_1)\ldots f(q_N|q_{N-1}, q_{N-2}, \ldots, q_1) \tag{3.20}$$

It can easily be seen from equation (3.19) that each of these conditional PDFs are uniform in a range determined by the previous values of q_i. Particularly since

$$Q_i = \frac{\Delta U_i - b(U_{i-1}, t_{i-1})\Delta t}{a(U_{i-1}, t_{i-1})}\,, \qquad U_k = U_{k-1} + \Delta U_k$$

we have

$$\begin{aligned} f(q_i|q_{i-1}, q_{i-2}, \ldots, q_1) &= f(q_i|u_{i-1}, u_{i-2}, \ldots, u_1, u_0) \\ &\propto |a_{i-1}| \left[s\Big(q_i - \frac{-\rho - b_{i-1}}{|a_{i-1}|}\Big) - s\Big(q_i - \frac{\rho - b_{i-1}}{|a_{i-1}|}\Big) \right] \end{aligned} \tag{3.21}$$

where $a_k = a(u_k, t_k)$, $b_k = b(u_k, t_k)$ and s is the step function. Now we can write the joint density function.

$$f_{\mathbf{Q}}(q_1, q_2, \ldots, q_N) \propto \begin{cases} \prod_{i=0}^{N-1} |a_i| & \text{if } q_i \in \left[\frac{-\rho - b_{i-1}}{|a_{i-1}|}, \frac{\rho - b_{i-1}}{|a_{i-1}|}\right] \forall i \in \{1, 2, \ldots, N\} \\ 0 & \text{o.w.} \end{cases} \tag{3.22}$$

As discussed in the previous section, the restriction of ΔU_i in $[-\rho, \rho]$ is a practical necessity and does not create any problems in real world applications since ρ can be chosen arbitrarily large, and the points where $f_{\mathbf{Q}}$ is zero can be viewed as proposals of impossible states and rejected immediately.

Combining equations (3.16) and (3.22) the whole accept-reject probability of the MCMC algorithm can be written as

$$\alpha(\Delta\mathbf{x}, \Delta\mathbf{x}') = \min\left(1, \frac{f_{\mathbf{Y}d}(h(\Delta\mathbf{x}'))}{f_{\mathbf{Y}d}(h(\Delta\mathbf{x}))} \frac{f_{\mathbf{Q}}(h(\Delta\mathbf{x}))}{f_{\mathbf{Q}}(h(\Delta\mathbf{x}'))}\right) \tag{3.23}$$

with $f_{\mathbf{Y}d}(\mathbf{y}) \propto e^{-\mathbf{y}^T\mathbf{y}/(2\Delta t)}$.

As a numerical example for the above procedure consider the linear stochastic differential equation

$$\begin{aligned} &\mathrm{d}X_t = \mu X_t \mathrm{d}t + \sigma X_t \mathrm{d}W_t \,, \qquad 0 \le t \le 1 \\ &X_0 = 1 \end{aligned} \tag{3.24}$$

where μ and σ are scalar constants. This equation describes the geometric Brownian motion [45] which finds applications in mathematical finance, particularly in the Black-Scholes model of financial markets [46]. This is a good example for demonstration purposes because one can obtain its solution analytically. A stochastic process satisfying equation (3.24) will have the form

$$X_t = X_0 e^{(\hat{\mu}t + \sigma W_t)} \tag{3.25}$$

where $\hat{\mu} = \mu - \frac{\sigma^2}{2}$. It's PDF has a lognormal distribution,

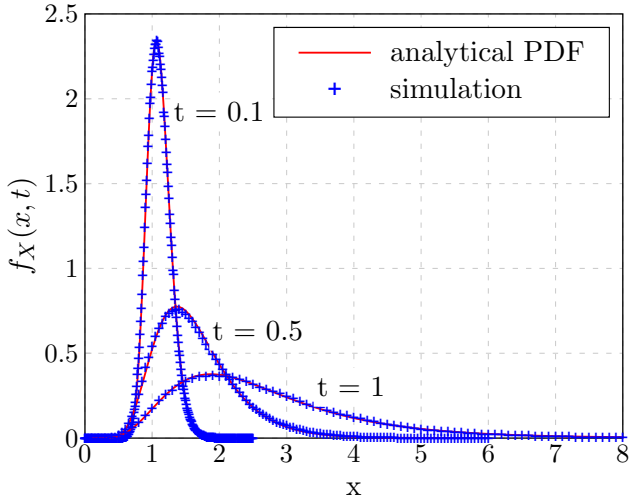


Figure 3.3: Analytical probability density functions of X_t at t = 0.1, t= 0.5 and t=1 compared with the empirical PDFs of the simulation data.

$$f_{X_t}(x,t) = \frac{1}{\sigma x\sqrt{2\pi t}} e^{-(\ln x - \ln X_0) - \hat{\mu}t)^2/(2\sigma^2 t)} \tag{3.26}$$

and the autocorrelation of X_t is given by

$$R(s,t) = e^{\mu(s+t)}(e^{\sigma^2 \min(s,t)} - 1) \quad . \tag{3.27}$$

Figures 3.3 and 3.4 display the results of a simulation with the modified MCMC algorithm. The scalar constants in equation (3.24) were chosen as $\mu = 1$ and $\sigma = 0.5$. The time axis was divided in one hundred equal length intervals with $\Delta t = 0.01$. Initially 5.1×10^6 samples were generated and the first 10^5 samples were discarded as the burn-in length. For this setting ρ was chosen to be 2 and a uniform distribution in $[-0.2, 0.2]$ was used as the proposal distribution for $\Delta\mathbf{X}$. Figure 3.3 shows three analytical PDFs at different time points compared with the empirical PDFs obtained from the simulation data and figure 3.4 shows the normalized autocorrelation with one time point held fixed and the second one varied between 0 and 1. These graphical results verify that the sample paths built using our algorithm converge to the desired stationary distribution and hence satisfy the given stochastic differential equation.

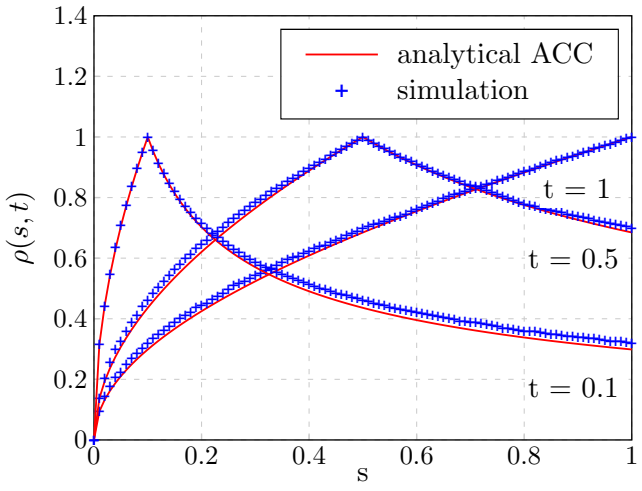


Figure 3.4: Normalized autocorrelation of X_t at three different time points compared with simulation data.

One noteworthy property of numerical solutions using the modified MCMC algorithm is that all points of a sample path get sampled in parallel as opposed to classical iterative methods such as the Euler-Maruyama scheme [47]. These methods usually begin with the initial value X_0, and sample later points of the solution path with an iterative update rule given by the difference equation (3.18). For this reason dealing with more complicated settings like stochastic boundary value problems of the form

$$\begin{aligned} &\mathrm{d}X_t = b(X_t,t)\mathrm{d}t + a(X_t,t)\mathrm{d}W_t\ , \qquad 0 \le t \le T \\ &h(X_0, X_T) = 0 \end{aligned} \tag{3.28}$$

becomes troublesome because the points of a sample path are not independent of its future values. On the other hand the incorporation of boundary conditions to the modified MCMC algorithm is straightforward since the points of the proposed sample paths are obtained simultaneously with independent increments.

3.6. Conclusion

We have presented a solution to the problem of input variable sampling for complex stochastic models with prescribed output distribution. This approach is based on a modification to the Metropolis-Hastings algorithm with an additional expression which can be viewed as a probing term for the model of interest. Our algorithm is easy to implement, benefits from the extensive literature on MCMC and hence we believe that it can be adapted to a variety of applications. We have demonstrated one such application on general stochastic differential equations viewing them from the perspective of stochastic input-output models enabling us to apply our algorithm to obtain solution paths. In the next chapter we will implement the compensated MCMC algorithm to sample input parameters for continuous and discrete-time models in order to generate the desired output statistics.

Although this treatment is based on MCMC, the approach taken to tackle the input variable sampling problem does not require any specific sampling method to be used. The algorithm presented here can be implemented equally well with other sampling methods once the output distribution for uniformly sampled input variables is worked out and therefore offers a fresh approach for dealing with general stochastic models.

Chapter 4

Parameter Sampling Methods for Accurate PMD Emulation

4.1. Introduction

In Chapter 2 we have presented three different computational models of PMD, two of which were of interest regarding PMD emulation, namely the reduced complexity continuous-time model and the discrete-time paraunitary model. Although it was mentioned several times that the uniform i.i.d. parameter sampling method for these models does not produce correct PMD statistics, no remedy was offered for this problem. In this chapter we investigate solution methods for the input parameter sampling problem of PMD models.

We begin our treatment with the continuous-time reduced complexity model and take up the problem where it was left off in [18]. Here, the authors fixed the pole locations of the reduced model on a half ellipse and sampled the remaining parameters of the model using MCMC. We show that this method results in suboptimal DGD statistics and offer a solution which greatly improves the DGD PDFs of the model by using compensated MCMC instead of the standard Metropolis-Hastings algorithm.

In the second part of this chapter we focus on the discrete time model and provide

a more in depth treatment of the parameter sampling problem. The discussion in Section 2.4.2 suggests that a different parameter sampling scheme must be implemented in order to achieve accurate DGD statistics which is only the first step in the construction of a PMD emulator since higher order statistics must be considered as well. To this end, in this section we propose three different parameter sampling methods for discrete time paraunitary FIR filters which can be used for more accurate PMD emulation. From such an emulator we expect a Maxwellian DGD distribution with an adjustable mean value and frequency independent behavior over the whole frequency range of interest. In order to capture higher order effects, we also require a good approximation of the frequency autocorrelation of the PMD vector [48]. With respect to these performance criteria, the three random parameter sampling schemes we propose can be listed and classified as follows:

- **Cascading of two-section blocks**: First order statistics matching.
- **Compensated MCMC**: Higher order statistics matching.
- **Greedy paraunitary approximation algorithm**: Transfer function matching.

There is a direct link between this classification and the computational complexity of the sampling schemes: The least "general" method is also the fastest.

4.2. Parameter Sampling for Reduced Complexity Continuous-Time Models using Compensated MCMC

In order to create the required dependency structure among the random model parameters $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N\}$, the standard Metropolis-Hastings algorithm starts with an initial value for $\mathbf{u}_k, k = 1 \ldots N$, and takes a random step in a subset of $\mathbb{R}^{3N}$ by randomly incrementing the components of the complex 2×1 unit vectors using a symmetric proposal density. The new set of parameters, $\{\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2, \ldots, \tilde{\mathbf{u}}_N\}$, construct

another model with different DGD values $\tilde{\tau}(\omega)$. This step is then accepted with the probability,

$$\alpha = \min\left(1, \prod_{i=1}^{K} \frac{f_M(\tilde{\tau}(\omega_i), \bar{\tau})}{f_M(\tau(\omega_i), \bar{\tau})}\right) \quad , \tag{4.1}$$

where $f_M(\bullet, \bar{\tau})$ denotes the Maxwellian PDF with mean $\bar{\tau}$ and ω_i is the i$^{\text{th}}$ frequency point. Note that in (4.1) the term with the proposal densities was left out because of the symmetry of the proposal distribution. If the proposed step is rejected, the algorithm stays at the same values of $\mathbf{u}_k$, and sets the new sample of model parameters equal to the previous one. In (4.1) the frequency axis is discretized in such a way that the frequency points are well separated from each other to be approximately independent. Therefore their joint DGD distribution can be written as the product of marginal PDFs.

The distribution of the samples generated in this way is expected to converge to the desired stationary distribution f_M. It was argued in the previous section that the stationary distribution of the above Markov chain deviates from the desired distribution because of the mapping between the random variables ($\mathbf{u}_k$'s) lying in the space of the random walk, and the random variable (DGD) the accept-reject rule in (4.1) is based on. In order to compensate for this discrepancy, we use the modified accept-reject rule in (4.16). The new acceptance probability of the compensated algorithm reads

$$\alpha_c = \min\left(1, \prod_{i=1}^{K} \frac{f_M(\tilde{\tau}(\omega_i), \bar{\tau})}{f_M(\tau(\omega_i), \bar{\tau})} \frac{f_U(\tau(\omega_i))}{f_U(\tilde{\tau}(\omega_i))}\right) \quad . \tag{4.2}$$

Here, f_U denotes the probability density function of the resulting DGD values when the model parameters $\mathbf{u}_k$ are sampled uniformly and independently.

We now present simulation results comparing the standard method [18] in (4.1) and the proposed compensated method in (4.2) to sample the parameters of the reduced model in (2.23).

In order to avoid the issues related to convergence, the exact same simulation

parameters are used for the compensated algorithm as they were used in the example given in [18]. A reduced model with 16 sections was considered to emulate an optical fiber with a mean-DGD of $\bar{\tau} = 10$ ps over a 40 GHz bandwidth. The random walk on $\mathbf{u}_k$ is achieved by random increments to the components of the unit vectors.

$$\mathbf{u}_k = \begin{pmatrix} \sqrt{\gamma_k} e^{j\theta_k} \\ \sqrt{1-\gamma_k} e^{j\phi_k} \end{pmatrix}, \quad \tilde{\mathbf{u}}_k = \begin{pmatrix} \sqrt{\tilde{\gamma}_k} e^{j\tilde{\theta}_k} \\ \sqrt{1-\tilde{\gamma}_k} e^{j\tilde{\phi}_k} \end{pmatrix}, \quad \begin{aligned} \tilde{\gamma}_k &= \gamma_k + \delta\gamma_k \\ \tilde{\theta}_k &= \theta_k + \delta\theta_k \\ \tilde{\phi}_k &= \phi_k + \delta\phi_k \end{aligned}$$

where $\delta\gamma_k$ is a uniform random variable in $[-0.1, 0.1]$ and $\delta\theta_k$ and $\delta\phi_k$ are uniformly distributed in $[-\frac{\pi}{20}, \frac{\pi}{20}]$.

The probability density function $f_U(\tau)$ for i.i.d. uniform model parameters is rather complicated and can be found in [49]. Note that the sections of the reduced model do not have the same poles and therefore the DGD of each section is different. The DGD of one section is given by

$$\mathrm{DGD}_k(\omega) = \frac{2\alpha_k}{\alpha_k^2 + (\omega + \beta_k)^2} \quad . \tag{4.3}$$

For the simulations, K in (4.1) and (4.2) was chosen to be 3. The frequencies used to compute the acceptance probabilities α and α_c are 0, 10 and 20 GHz. Initially 2.01×10^6 samples were generated and the first 10^4 were discarded as the burn-in length.

Figure 4.1 shows the mean DGD and the standard deviation over the whole frequency range. As observed, the compensated MCMC technique proposed performs considerably better than the standard MCMC algorithm. Figure 4.2 displays the DGD distribution on the corner frequency, 20 GHz, in semi-logarithmic scale. Here, one can see that although the mean and the standard deviation results of the uncompensated algorithm are not very far from the expected values, the shape of the DGD probability density function does not resemble the Maxwellian. Simulation results on the other two frequencies used to compute the accept-reject rule, 0 and 10 GHz, exhibit a similar behavior. Finally, Figure 4.3 displays the log-scale DGD distribution

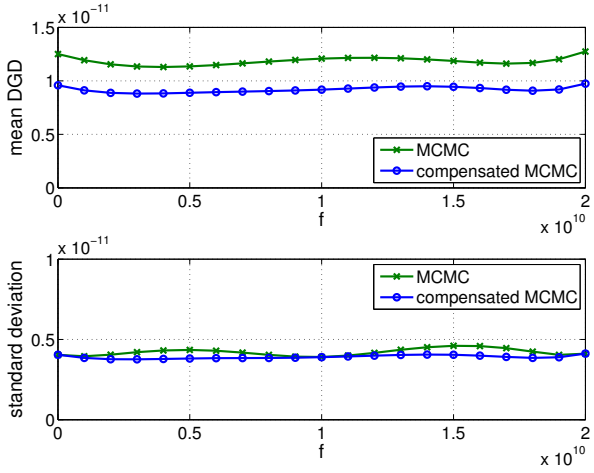


Figure 4.1: Mean DGD and standard deviation values for both algorithms over the whole frequency range.

at an intermediate frequency, 15 GHz, which is not one of the three frequencies used in (4.1) and (4.2). As expected the DGD distribution of the output of the compensated MCMC algorithm compares similarly to the uncompensated MCMC much like at other frequencies. Above results confirm the predicted discrepancy in the DGD distribution of the randomized reduced model when the samples are generated with the standard MCMC as well as the correction that can be achieved using the proposed compensated MCMC technique.

4.3. Parameter Sampling for Paraunitary FIR Filters

The following section introduces the aforementioned sampling schemes and demonstrate their performance emulating an optical communication link with a bandwidth of 40 GHz and a mean DGD of 0.4 symbol period. All of the experiments use a discrete time filter consisting of 20 birefringent sections. Furthermore we assume that the continuous time system was sampled four times over its minimum rate (4×40 GHz) and therefore set the mean DGD of the emulators to 1.6 sampling periods.

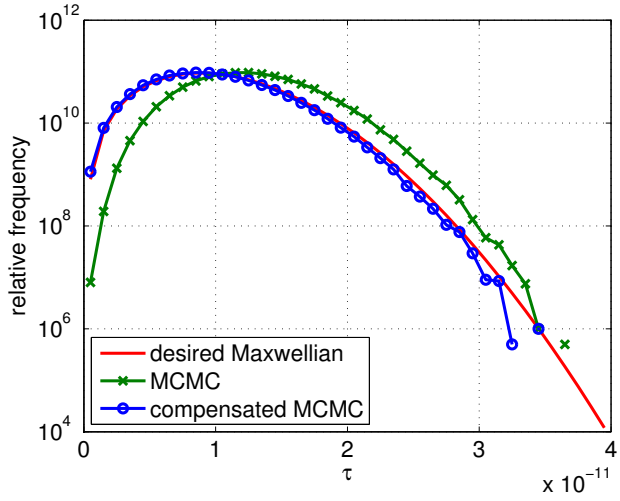


Figure 4.2: Probability density function of DGD at 20 GHz for both algorithms compared with the desired Maxwellian PDF.

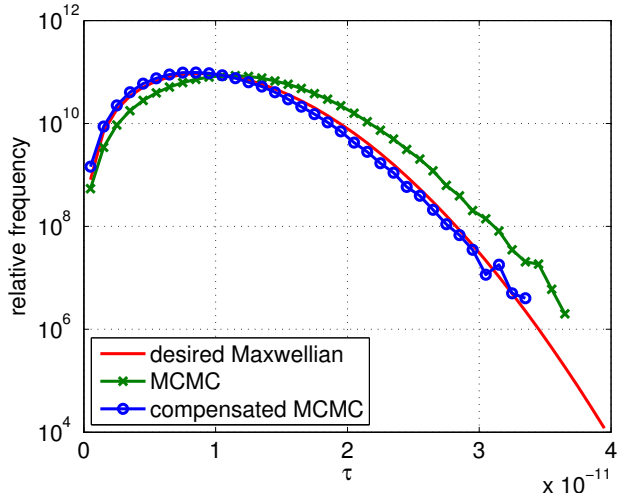


Figure 4.3: Probability density function of DGD at 15 GHz for both algorithms compared with the desired Maxwellian PDF.

4.3.1. Cascading Method

In order to sample the filter parameters in such a way that the mean DGD becomes adjustable and frequency independent, one can extend the degree-one paraunitary building blocks of the filter to two-section blocks that have dependent $\mathbf{v}_i$. We call this technique "cascading" and analyze its properties below.

Constructing a Paraunitary FIR Filter with Maxwellian DGD Distribution at the Center Frequency

The difference between the eigenvalues of a Hermitian matrix,

$$\mathbf{G} = \begin{bmatrix} A & C \\ C^* & B \end{bmatrix} \quad , \tag{4.4}$$

has the form: $\tau = \sqrt{(A-B)^2 + 4CC^*}$. Therefore an ensemble of 2×2 Hermitian matrices will have a Maxwellian spacing distribution if the individual components in (4.4) are distributed as follows:

$$\begin{aligned} A \sim \mathcal{N}(\tfrac{N}{2}, \sigma) \quad & \Re\{C\} \sim \mathcal{N}(0, \sigma) \\ B = N - A \quad & \Im\{C\} \sim \mathcal{N}(0, \sigma) \end{aligned} \tag{4.5}$$

Here $\mathcal{N}(\mu, \sigma)$ denotes the Gaussian distribution with mean μ and standard deviation σ. Considering equation 1.18, the resulting Maxwellian distribution will have the following mean value:

$$\bar{\tau} = 2\sqrt{\frac{8}{\pi}}\sigma \quad . \tag{4.6}$$

At this stage one can ask if unit norm vectors $\mathbf{v}_i$ can be found such that $\mathbf{G} = \sum_{i=1}^{N} \mathbf{v}_i \mathbf{v}_i^*$. If one can find such vectors they can be used to construct a paraunitary FIR filter so that it will have a Maxwellian DGD at the center frequency by construction. This is indeed the case if $\mathrm{tr}(\mathbf{G}) \in \mathbb{Z}$ and $\mathrm{tr}(\mathbf{G}) \geq \mathrm{rank}(\mathbf{G})$ [50].

Since $N = \mathrm{tr}(\mathbf{G})$ can be restraint to positive integers, this condition can always

be satisfied. Being the sum of rank 1 projection matrices, $\mathbf{v}_i\mathbf{v}_i^*$, puts one additional constraint on $\mathbf{G}$: $\mathbf{G}$ must be positive definite. This enforces following inequalities on the parameters in (4.5).

$$A > 0, \quad B > 0, \quad AB > |C|^2 \tag{4.7}$$

Because of these constraints, the distributions of A and C in (4.5) must have finite support and hence cannot be actual Gaussians. In section 4.3.1 we will show that adjusting the standard deviation of the distributions one can eliminate this discrepancy for all practical purposes.

The case for $N = 2$

Since $\mathbf{G}$ has real eigenvalues and orthogonal eigenvectors if it is constructed as in (4.4) and (4.5) with $N = 2$, it can be easily verified that $\mathbf{G}$ can be partitioned in the following way:

$$\mathbf{G} = \mathbf{v}_1\mathbf{v}_1^* + \mathbf{v}_2\mathbf{v}_2^*, \qquad \mathbf{v}_1^*\mathbf{v}_1 = \mathbf{v}_2^*\mathbf{v}_2 = 1$$

with,

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}}\left(\sqrt{\lambda_1}\mathbf{e}_1 + \sqrt{\lambda_2}\mathbf{e}_2\right) \quad \mathbf{v}_2 = \frac{1}{\sqrt{2}}\left(\sqrt{\lambda_1}\mathbf{e}_1 - \sqrt{\lambda_2}\mathbf{e}_2\right) \quad . \tag{4.8}$$

Where $\mathbf{e}_i$ are the unit norm orthogonal eigenvectors of $\mathbf{G}$ with λ_i as the corresponding eigenvalues. In order to get a marginally uniform distribution for magnitude squares of the components of $\mathbf{v}_i$, the eigenvectors $\mathbf{e}_i$ can be multiplied with a phase term $e^{j\phi_i}$, where ϕ_i is selected uniformly at random in $[0, 2\pi]$.

The resulting DGD density of a simulation with this selection method for $\mathbf{v}_i$ is displayed in Figure 4.4. The sample size is 10^7 and $\bar{\tau}$ was taken to be 0.4.

Because of the positive definiteness constraint given in equation (4.7), the above two-section unit can only reach DGD values up to a limit. This shortcoming of the second degree FIR filter can be overcome with the extension of the same idea to higher degree filters. The first possibility that comes to mind is to cascade M second degree

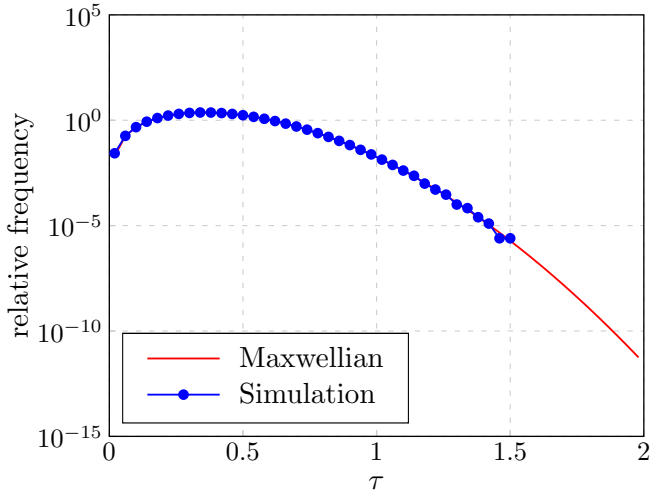


Figure 4.4: The PDF of the DGD of the second degree FIR filter compared against a true Maxwellian with the same mean.

filters and build a more general FIR filter with 2M sections.

Extension to Higher Degree Filters

Using the fact that the sum of two independent Gaussian random variables is again a Gaussian random variable, we can extend the above discussed second order FIR filter to a more general system of even order.

$$\mathbf{G} = \mathbf{G}_1 + \mathbf{G}_2 + \cdots + \mathbf{G}_M$$

$\mathbf{G}$ has a Maxwellian spacing distribution if the summands $\mathbf{G}_i$ are constructed independently according to (4.4) and (4.5). In this case, the unit norm vectors $\mathbf{v}_i$ can be selected with the same scheme as (4.8).

$$\mathbf{G} = \sum_{k=1}^{M} \begin{bmatrix} A_k & C_k \\ C_k^* & B_k \end{bmatrix} = \sum_{k_1,k_2=1}^{M} \mathbf{v}_{k_1}\mathbf{v}_{k_1}^* + \mathbf{v}_{k_2}\mathbf{v}_{k_2}^* \tag{4.9}$$

$$\begin{aligned} A_k &\sim \mathcal{N}(1,\sigma) & \Re\{C_k\} &\sim \mathcal{N}(0,\sigma) \\ B_k &= N - a_k & \Im\{C_k\} &\sim \mathcal{N}(0,\sigma) \end{aligned} \tag{4.10}$$

This filter of order 2M will have the following mean DGD:

$$\bar{\tau} = 2\sqrt{\frac{8M}{\pi}}\sigma \quad . \tag{4.11}$$

Constraints on Model Parameters

As discussed earlier, the selection of the standard deviation σ in (4.5) is not arbitrary. The fact that $\mathbf{G}$ must be positive definite restrains σ with an upper bound determined by the minimum probability value we want to match in the DGD distribution. For a fixed value of σ the probability that $\mathbf{G}$ in (4.4) is not positive definite can be computed from (4.7).

The last constraint in (4.7) takes precedence over the other two since A and B can never be negative simultaneously and $|C|^2$ is always positive. The maximum value $AB = A(N-A)$ can get is $\frac{N^2}{4}$. Hence in the case that $|C| \geq \frac{N}{2}$, $\mathbf{G}$ is never positive definite. Similarly for $A \notin (0, N)$ $\mathbf{G}$ cannot be positive definite. This restricts the range of τ to $(0, N)$ as it is expected from an FIR filter built with N delay elements.

With the conditions above we can calculate the probability that $\mathbf{G}$ is positive definite with given N and σ.

$$\begin{aligned} \mathrm{P}(\mathbf{G} \succ 0) &= P(\{A = a \,|\, a \in (0, N)\} \cap \{|C| = c \,|\, c^2 \leq a(N-a)\}) \\ &= \int_0^{\frac{N^2}{4}} P(|C|^2 \leq x \,|\, X = x)\, \mathrm{d}x \\ &= \int_0^{\frac{N^2}{4}} P(|C|^2 \leq x) f_X(x)\, \mathrm{d}x \end{aligned} \tag{4.12}$$

Where $X = AB$. The last line of the above equation holds because X and C are independent.

The PDF of X is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(N^2/4-x)}{2\sigma^2}} \frac{1}{\sqrt{\frac{N^2}{4}-x}}$$

The random variable $|C|$ is Reighley distributed since $|C| = \sqrt{\Re\{C\}^2 + \Im\{C\}^2}$ and its CDF is given by

$$F_{|C|}(c) = 1 - e^{-\frac{c^2}{2\sigma^2}}$$

Hence

$$\begin{aligned} \mathrm{P}(\mathbf{G} \succ 0) &= \int_0^{\frac{N^2}{4}} \left(1 - e^{-\frac{x}{2\sigma^2}}\right) \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(N^2/4-x)}{2\sigma^2}} \frac{1}{\sqrt{\frac{N^2}{4}-x}} \,\mathrm{d}x \\ &= \frac{-N}{\sqrt{2\pi}\sigma} e^{\frac{-N^2}{8\sigma^2}} + \frac{1}{\sqrt{2\pi}\sigma} \int_0^{\frac{N^2}{4}} e^{\frac{-(N^2/4-x)}{2\sigma^2}} \frac{1}{\sqrt{\frac{N^2}{4}-x}} \,\mathrm{d}x \end{aligned}$$

After the variable transformation $u^2 = N^2/4 - x$ we arrive at the final result.

$$\mathrm{P}(\mathbf{G} \succ 0) = \mathrm{erf}\left(\frac{N}{2\sqrt{2}\sigma}\right) - \frac{N}{\sqrt{2\pi}\sigma} e^{\frac{-N^2}{8\sigma^2}}$$

Where erf(•) is the error function.

Note that this expression is equal to the probability that a Maxwellian random variable with the same mean as in equation (4.6) is smaller than N. In other words, the positive definiteness of $\mathbf{G}$ ensures that the resulting DGD is at most N which is the maximum delay of a discrete time filter with N sections.

This result can be used as a model order selection tool for a given minimum probability we want to match, p, and a mean DGD, $\bar{\tau}$.

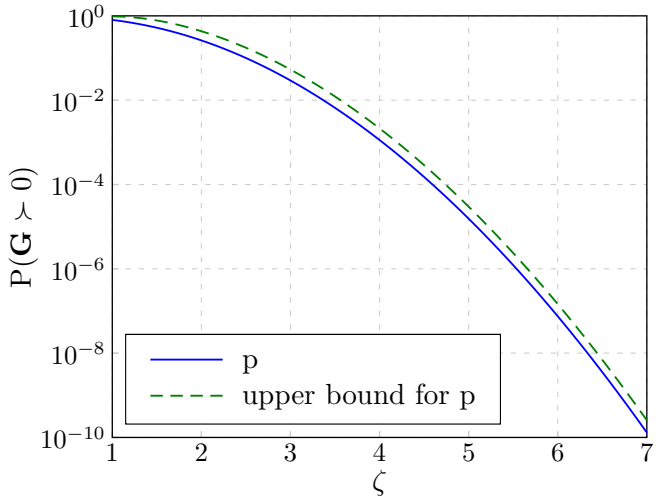


Figure 4.5: The value of $\mathrm{P}(\mathbf{G} \succ 0)$ and its upper bound.

$$\begin{aligned} p &= 1 - \mathrm{P}(\mathbf{G} \succ 0) \\ &= \operatorname{erfc}\left(\frac{N}{2\sqrt{2}\sigma}\right) + \frac{N}{\sqrt{2\pi}\sigma} e^{\frac{-N^2}{8\sigma^2}} \\ &= 2\left[\mathrm{Q}\left(\frac{N}{2\sigma}\right) + \frac{N}{\sqrt{2\pi}2\sigma} e^{-\frac{(N/2\sigma)^2}{2}}\right] \end{aligned} \tag{4.13}$$

Furthermore we can use the well known upper bound for the Q-function

$$\mathrm{Q}(\zeta) < \frac{\zeta}{\sqrt{2\pi}} e^{\frac{-\zeta^2}{2}}, \quad \zeta \geq 1$$

to get a more useful expression. If we set $\zeta = \frac{N}{2\sigma}$ in the above expression we get

$$p < 4\frac{N}{2\sigma\sqrt{2\pi}} e^{-\frac{(N/2\sigma)^2}{2}} \tag{4.14}$$

For a predetermined p, we can calculate the value of σ. For example for $N = 2$ and $\mathrm{p} = 10^{-6}$ the maximum value of σ is approximately 0.18. This equals to a maximum mean DGD of 0.5745.

The fact that σ has an upper bound limits the cascading method in terms of the

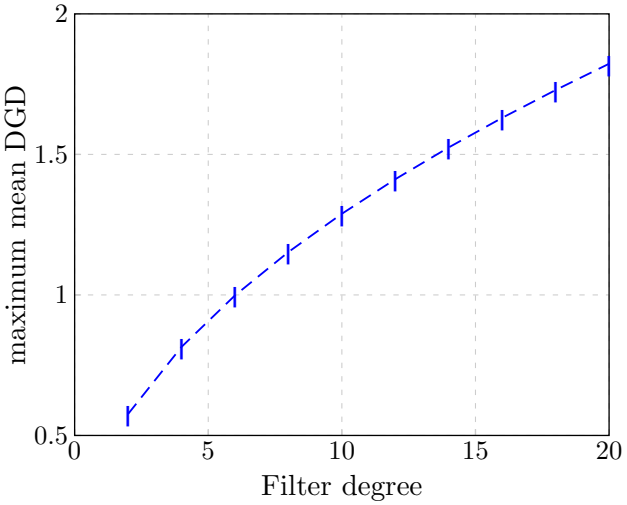


Figure 4.6: Maximum achievable mean DGD of the cascade method. Notice that the cascading method can only produce filters with even filter degrees.

maximum achievable mean DGD $\bar{\tau}_{max}(p)$ for a given threshold probability p in (4.13) and the filter length N. This value can be computed for varying filter degrees via (4.6) and (4.11). Figure 4.6 shows the maximum achievable DGD values for $p = 10^{-6}$.

The second restriction on σ originates from the discrete time nature of the problem at hand. The above calculations of τ, particularly (4.6), are normalized in terms of the sampling period T_S.

$$\tau = \frac{\mathrm{DGD}}{T_S}$$

Because the sampling period must be less than or equal to the symbol period T_{Sym} we have

$$\frac{\overline{\mathrm{DGD}}}{T_S} = 2\sqrt{\frac{8}{\pi}}\sigma \geq \frac{\overline{\mathrm{DGD}}}{T_{Sym}}$$

and

$$\sigma \geq \frac{\overline{\mathrm{DGD}}}{2T_{Sym}}\sqrt{\frac{\pi}{8}} \quad . \tag{4.15}$$

Where $\overline{\mathrm{DGD}}$ is the mean DGD of the system in seconds.

As an example a communication channel with a 40 GHz bandwidth has a symbol period of 25 psecs when binary signalling is used. A typical mean DGD value is 10 psecs. If we evaluate (4.15) with these values we get approximately $\sigma \geq 0.125$. With the upper bound above we have $0.125 \leq \sigma \leq 0.18$.

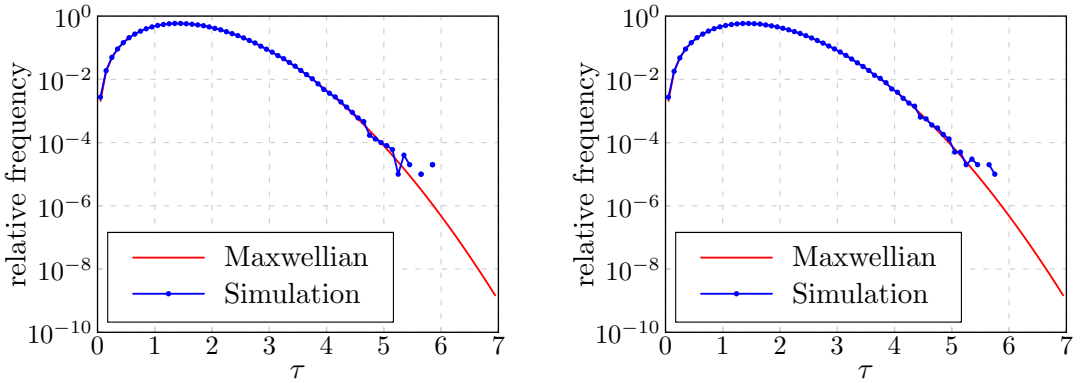
Although with the continuous time system sampled at Nyquist rate gives us a viable range for σ this situation changes immediately if we oversample the signal. An oversampling rate of 2 is sufficient to render the second order FIR filter unusable. The treatment of a general communication link is possible with the extension of the second order FIR filter to higher orders.

For the example communication link above with 40 GHz bandwidth and an oversampling rate of 2 ($T_S = \frac{T_{Sym}}{2}$) the use of the cascading method allows a viable standard deviation if M is large enough: $\frac{0.25}{\sqrt{M}} \leq \sigma \leq 0.18$. With the use of a sixth order FIR filter ($M = 3$) one can obtain a mean DGD of 10 psecs if σ is selected to be 0.1447.

Frequency Behavior of the Cascading Method

The analysis given above only discusses the properties of the cascade method at the center frequency but the reason this method was chosen among other alternatives to partition a positive definite matrix as sum of idempotents is its uniform frequency behavior.

Figure 4.7 illustrates this behavior in terms of the PDF of the DGD. These graphics display the DGD distribution of a filter with 20 sections and a mean DGD value of 1.6 first at the center frequency, $\omega = 0$, and then at the corner frequency, $\omega = \frac{\pi}{4}$. The solid curves in the graphics are the expected Maxwellians with mean 1.6. Figure 4.7 shows complete agreement of the model output with the expected values at the center as well as the corner frequency. This behavior is typical for the cascade method with low enough mean DGD values however higher order statistics do not display a similar behavior. Figure 4.8 shows that although the mean of the DGD distribution is constant over the whole frequency range, the correlation structure deviates from



(a) PDF at $\omega = 0$ (b) PDF at $\omega = \frac{\pi}{4}$

Figure 4.7: PDF of the DGD at the center frequency ($\omega = 0$) and the corner frequency ($\omega = \frac{\pi}{4}$). Red curves represent the Maxwellian with mean 1.6.

the desired curve. Therefore the cascade method, in its current form, remains only as a tool for matching the first order statistics of a real PMD channel.

4.3.2. Compensated MCMC Method

Now let us investigate how MCMC can be employed to generate samples of fiber models using paraunitary FIR filters. A discrete time fiber model with N concatenated degree-one sections can be viewed as a complex mapping from the sample space of filter parameters to the space of PMD vector values. This mapping accepts a set of 2×1 complex valued unit norm vectors, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N\}$, which have a total of $2N$ real scalar parameters as input and produces a frequency dependent PMD vector, $\boldsymbol{\tau}(\omega)$, at the output. If we discretize the frequency axis such that we force the statistical properties of the PMD vectors, $\{\boldsymbol{\tau}(\omega_1), \boldsymbol{\tau}(\omega_2), \ldots \boldsymbol{\tau}(\omega_M)\}$ at a set of frequencies, ω_1 through ω_M, in the frequency range of interest, we can expect the model to behave similarly at intermediate frequencies. Consequently, we obtain a mapping from $\mathbb{R}^{2N}$ to $\mathbb{R}^M$. Therefore, the problem of sampling the input parameters, such that the output statistics exhibit the desired behavior, can be described as follows.

Suppose we are given a general many-to-one, non-isometric map $h : \mathbb{R}^n \to \mathbb{R}^m$

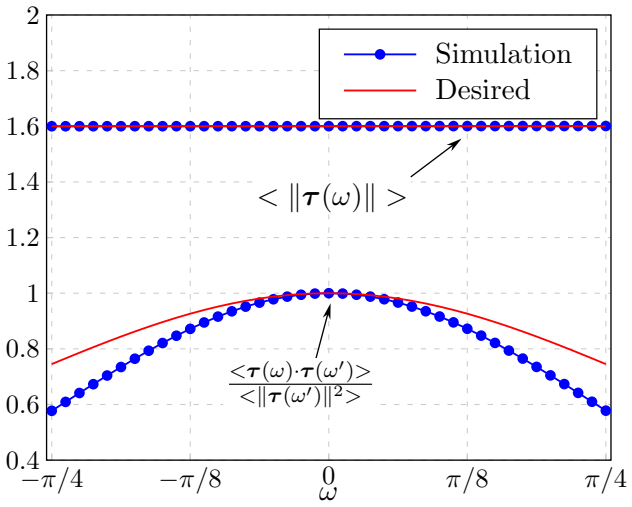


Figure 4.8: Mean and normalized autocorrelation curves of the cascade method compared with the expected values.

which maps a random vector $\mathbf{X}$ to another random vector $\mathbf{Y} = h(\mathbf{X})$ where $\mathbf{X} \in \mathbb{R}^n$, $\mathbf{Y} \in \mathbb{R}^m$ and let $f_{\mathbf{Y}d}$ be the desired probability density of $\mathbf{Y}$. Given $f_{\mathbf{Y}d}$ how must $f_{\mathbf{X}}$ be chosen such that the transformed variable $\mathbf{Y} = h(\mathbf{X})$ has the desired PDF?

The answer to this question was given in Chapter 3 under the framework of the compensated MCMC algorithm which modifies the accept-reject rule in the standard Metropolis-Hastings algorithm as

$$\alpha_c(\mathbf{x}, \mathbf{x}') = \min\left(1, \frac{f_{\mathbf{Y}d}(h(\mathbf{x}'))}{f_{\mathbf{Y}d}(h(\mathbf{x}))} \frac{p(\mathbf{x}', \mathbf{x})}{p(\mathbf{x}, \mathbf{x}')} \frac{f_U(h(\mathbf{x}))}{f_U(h(\mathbf{x}'))}\right) \quad , \tag{4.16}$$

where $f_U(\bullet)$ is the distribution of the output random vectors with uniform i.i.d. input parameters. This distribution will be called uniform parameter distribution and abbreviated as UPD from now on.

Although the general setting for the discrete time model parameter sampling problem is the same as the continuous time case, the particulars assume a more complicated form. For the continuous time model we only consider the first order statistics of the DGD and build the accept reject rule of the MCMC algorithm accordingly, whereas

in the discrete time case we strive to match higher order statistics of a true PMD channel. In order to achieve this goal, we use the output PMD vector of the model as the target random variable. Due to the complicated nature of the PMD vector, we make use of simplifying assumptions and approximations about its probability distributions concerning the full model as well as the paraunitary FIR filter.

Target Density for Compensated MCMC

As the discussion in Section 1.2.3 suggests, we expect an ensemble of PMD vectors to satisfy the higher order statistical requirements of a true PMD channel when the individual components have a jointly Gaussian distribution. Since we assume that only the same components at different frequency points are dependent on each other, we can express the joint probability density of a collection of PMD vectors as follows:

$$f_T(\mathbf{T}, \bar{\tau}) = K(\mathbf{\Sigma}) \exp\left(-\frac{1}{2}\mathrm{tr}(\mathbf{T}\mathbf{\Sigma}^{-1}\mathbf{T}^T)\right) \quad . \tag{4.17}$$

Here, $\mathbf{T}$ is the $3 \times k$ matrix of PMD vectors, $\mathbf{T} = [\boldsymbol{\tau}(\omega_1)\, \boldsymbol{\tau}(\omega_2)\, \ldots\, \boldsymbol{\tau}(\omega_k)]$, $\mathbf{\Sigma}$ is the $k \times k$ covariance matrix and $K(\mathbf{\Sigma})$ is the normalizing constant of the PDF.

The covariance matrix depends only on one parameter which is the desired mean DGD, $\bar{\tau}$, and can be computed with the expression describing the expected value of the inner product of two PMD vectors at different frequency points ω and ω' [13]:

$$\langle \boldsymbol{\tau}(\omega) \cdot \boldsymbol{\tau}(\omega') \rangle = \frac{3}{\Delta\omega^2}\left(1 - \exp\left(-\frac{\Delta\omega^2 \langle \bar{\tau}^2 \rangle}{3}\right)\right) \quad , \tag{4.18}$$

where $\Delta\omega = |\omega - \omega'|$ and $\langle \bar{\tau}^2 \rangle = \frac{3\pi}{8}\bar{\tau}^2$.

Using (4.18) we can fill in the entries of $\mathbf{\Sigma}$ with

$$\mathbf{\Sigma}_{ij} = \frac{1}{3}\langle \boldsymbol{\tau}(\omega_i) \cdot \boldsymbol{\tau}(\omega_j) \rangle \quad ,$$

since different components of PMD vectors are independent. Because the covariance between two components depends on their distance on the frequency axis, this matrix has a Toeplitz structure.

Uniform Parameter Distribution of the FIR Filter

The evaluation of the accept-reject rule in the compensated MCMC algorithm requires the knowledge of the distribution of the PMD vectors when the model parameters are sampled uniformly and independently. Although this distribution resembles the PMD vector distribution of the full model, its exact form deviates from a joint Gaussian much more than the PDF of the full model does because it has significantly fewer birefringent sections. Indeed, the effort to approximate this PDF with a joint Gaussian or even a Gaussian mixture model results in inaccurate PMD vector statistics in terms of mean DGD and covariance structure. Moreover, upon close inspection one can observe that despite being uncorrelated, different PMD vector components at different frequency points exhibit a tail dependency which cannot be captured with a jointly Gaussian distribution. In order to overcome these difficulties we model the uniform parameter distribution as a copula vine [51].

Copulas are multivariate functions that are employed to describe dependency structures of random variables [52, 53]. For our purposes, without going into their theoretical description, copulas can be viewed as linkage functions that express the joint PDFs of random variables as the product of their marginals and their dependency structure. For the case of two dependent random variables, X and Y with marginal CDFs F_X and F_Y, one can write

$$f_{XY}(x,y) = f_X(x) f_Y(y) c_{XY}(F_X(x), F_Y(y)) \quad , \tag{4.19}$$

where f_X and f_Y are marginal PDFs of X and Y respectively and f_{XY} is their joint PDF. The non-negative bivariate function $c_{XY} : [0,1] \times [0,1] \to \mathbb{R}_+$ is the copula density of the two random variables. According to Sklar's theorem [54], under some regularity conditions, such a function always exists.

Equation (4.19) becomes especially useful if the marginals are known. This is indeed the case in the UPD. The PMD vector at the i$^{\text{th}}$ frequency point, $\boldsymbol{\tau}(\omega_i) = [\tau_1(\omega_i), \tau_2(\omega_i), \tau_3(\omega_i)]^T$, has independent components that are distributed according to

the uniform sum distribution. A random variable X that is the sum of N independent uniform random variables has the PDF

$$f_U(x,N) = \frac{1}{2\,(N-1)!}\sum_{k=0}^{N}(-1)^k\binom{N}{k}(x-k)^{N-1}\operatorname{sgn}(x-k) \quad . \tag{4.20}$$

Its CDF is given by

$$F_U(x,N) = \frac{1}{N!}\sum_{k=0}^{\lfloor x\rfloor}(-1)^k\binom{N}{k}(x-k)^N \quad . \tag{4.21}$$

The subscript U denotes that this distribution is the univariate margin of the UPD. As one can easily guess, in this case N is the number of sections of the paraunitary FIR filter.

For the construction of the copula density, the multivariate t-copula has proven to be useful [55]. In fact the set of same PMD vector components (e.g. the first component at all the frequency points etc.) follows a multivariate t-copula distribution with uniform sum marginals almost exactly. The remaining dependency among cross-components is modeled using bivariate t-copulas and arranging them into a D-vine [56] in order to exploit the stationarity property of the joint PMD vector distribution. The parameters of the multivariate as well as the bivariate copulas can be estimated using standard maximum likelihood algorithms.

Figure 4.9 illustrates the strategy for building trees of pair copulas for three frequency points. The three-variate t-copula, $c_{123}(F_U(\tau_m(\omega_1)), F_U(\tau_m(\omega_2)), F_U(\tau_m(\omega_3)))$, is the joint density of the m^{th} PMD vector components. The dependency among cross-components, $(\tau_m(\omega_i), \tau_n(\omega_j))$, $m \neq n$, $i \neq j$, is accounted for with the pair copulas c_1 and c_2, and 1 is the independence copula that connects the components of a single PMD vector. The symmetry in Figure 4.9 is caused by a simplifying assumption we make in order to obtain a more tractable UPD. Note that the pair copulas connecting the cross-components are in fact conditional PDFs that not only depend on the conditioning variables but also operate on the transformed forms of their arguments. Here

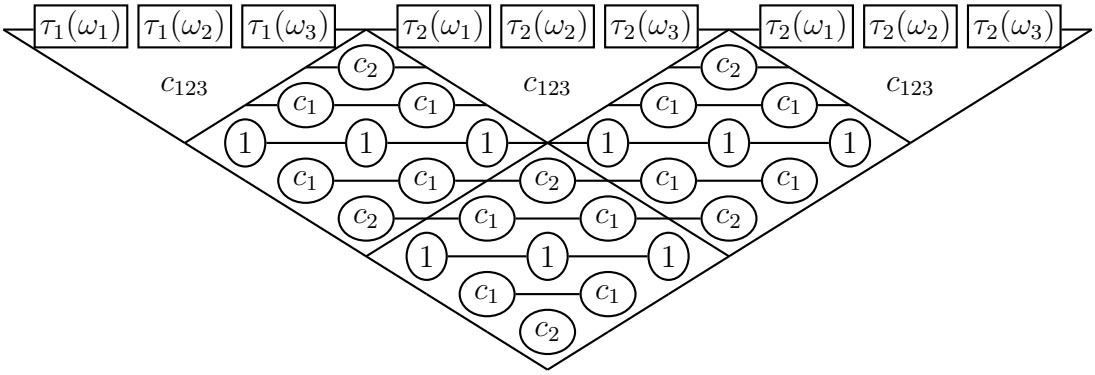


Figure 4.9: Copula vine structure for the uniform parameter distribution.

we make the assumption that these copulas are sufficiently flat so that we can describe the relationship among cross-components solely based on their frequency separation. At the end, we obtain the whole joint copula of all nine variables by multiplying all the components in Figure 4.9.

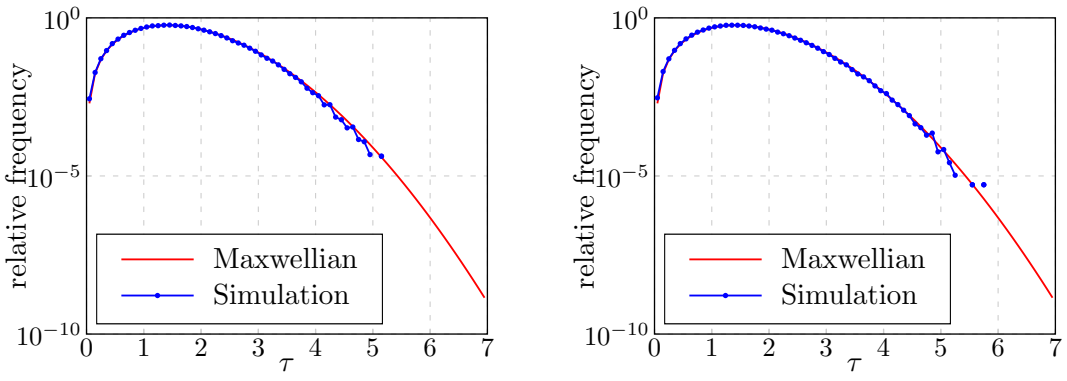
Performance of Compensated MCMC

Using the results from previous discussion, we can construct an accept-reject rule in the compensated MCMC algorithm such that the statistics of the ensemble of output PMD vectors will approximate the desired values. To this end we can write

$$\alpha_c = \min\left(1, \frac{f_T(\tilde{\mathbf{T}}, \bar{\tau})}{f_T(\mathbf{T}, \bar{\tau})} \frac{f_U(\mathbf{T}, N)}{f_U(\tilde{\mathbf{T}}, N)}\right) \quad , \tag{4.22}$$

The statistics of the compensated MCMC algorithm output is illustrated in Figures 4.10 and 4.11. The simulation was ran with 2×10^6 samples and the first 10^4 samples were discarded as the burn-in phase. The number of frequencies in the accept-reject rule, k, was chosen to be 3 ($-\pi/4$, 0 and $\pi/4$). It can be observed that this algorithm performs much better in terms of the autocorrelation function than the cascade algorithm at the cost of a small deviation in the mean value.

The limiting factor on the accuracy of the compensated MCMC algorithm is how



(a) PDF at $\omega = 0$ (b) PDF at $\omega = \frac{\pi}{4}$

Figure 4.10: PDF of the DGD at the center frequency ($\omega = 0$) and the corner frequency ($\omega = \frac{\pi}{4}$). Red curves represent the Maxwellian with mean 1.6.

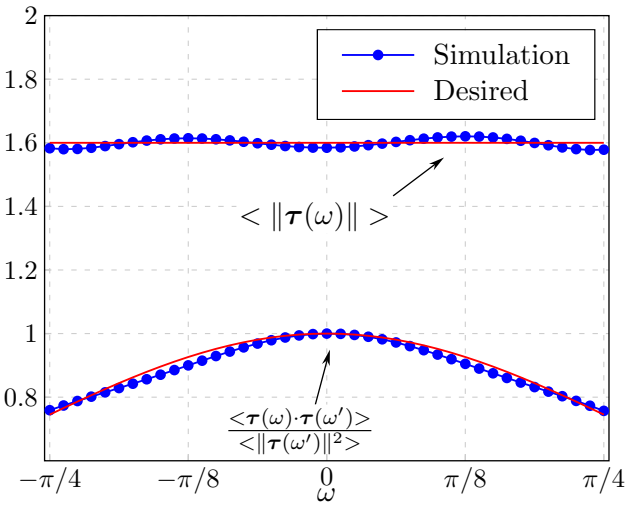


Figure 4.11: Mean and normalized autocorrelation curves of the compensated MCMC method compared with the expected values.

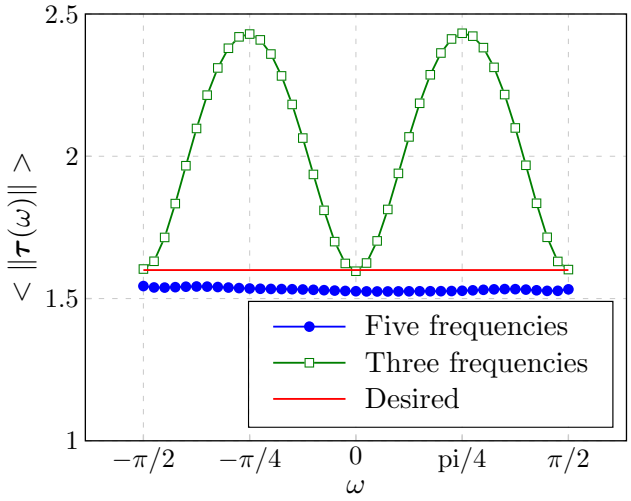


Figure 4.12: The mean DGD values for a twice oversampled system with three and five frequency points in the accept-reject rule of the compensated MCMC algorithm.

well the UDP can be modeled and approximated. The assumptions about the UDP and the pair copulas describing it are a source of increasing inaccuracy as the number of frequencies in the accept-reject rule grows. On the other hand, holding the number of frequencies fixed while increasing their distance results in non-uniform frequency behavior of the output. Figure 4.12 illustrates these two properties together by displaying the mean DGD values for simulations using three and five frequencies over the frequency range $[-\pi/2, \pi/2]$, i.e. for a system that is oversampled at twice its minimum sampling rate. The simulation with three frequency points in the accept-reject rule matches the desired mean DGD value at these frequencies but exhibits large deviations at intermediate points while the simulation with five frequencies results in a relatively constant mean DGD value smaller than the desired one.

4.3.3. Greedy Approximation Algorithm

At the heart of the last method we consider for the discrete time PMD emulation parameter sampling problem lies a greedy iterative transfer function approximation algorithm [57]. This algorithm takes a general matrix transfer function as input and

tries to approximate it with a paraunitary FIR filter. This is achieved by iteratively optimizing each section of the filter. Figure 4.13 shows the result of one sample run of the algorithm with 200 iterations. Although only the absolute value of the 2×2 transfer function is displayed, the same behavior of complete overlap of the curves can also be observed with the phases.

The iterative approximation algorithm method first defines a distance measure in the space of discrete time transfer functions as a mean-squared weighted Frobenius norm,

$$\xi \triangleq \frac{1}{2\pi} \int_0^{2\pi} W(\omega) \left\| \mathbf{D}(e^{j\omega}) - \mathbf{H}(e^{j\omega}) \right\|_F^2 \mathrm{d}\omega \quad , \tag{4.23}$$

where, for our purposes, $\mathbf{D}(e^{j\omega})$ is the transfer function of a real fiber constructed with a high number of birefringent sections (full model) and $\mathbf{H}(e^{j\omega})$ is the transfer function of the paraunitary FIR filter. $W(\omega)$ is a weighting function which is set equal to identity in the frequency range of interest and zero otherwise. This distance is then iteratively minimized by handling the unitary matrix $\mathbf{R}$ and each degree-one section $\mathbf{H}_i$ in equation (2.26) separately.

The optimization of $\mathbf{R}$ boils down to maximizing $\Re\{\mathrm{tr}(\mathbf{R}^*\mathbf{A})\}$, with

$$\mathbf{A} = \frac{1}{2\pi} \int_0^{2\pi} W(\omega) \mathbf{V}^*(e^{j\omega}) \mathbf{D}(e^{j\omega}) \mathrm{d}\omega \tag{4.24}$$

and $\mathbf{V}(e^{j\omega}) = \prod_{i=N}^{1} \mathbf{H}_i(e^{j\omega})$. This expression can be optimized elegantly by the closest unitary matrix to $\mathbf{A}$ which in turn can be computed via the SVD of $\mathbf{A}$ [58]. Similarly the optimization for individual $\mathbf{v}_i$ in equation (2.25) is achieved by minimizing the quadratic form $\mathbf{v}_i^*(\mathbf{B} + \mathbf{B}^*)\mathbf{v}_i = \mathbf{v}_i^* \mathbf{Q} \mathbf{v}_i$ with

$$\mathbf{B} = \frac{1}{2\pi} \int_0^{2\pi} W(\omega)(1 - e^{-j\omega}) \mathcal{R}_i(e^{j\omega}) \mathbf{R} \mathbf{D}^*(e^{j\omega}) \mathcal{L}_i(e^{j\omega}) \mathrm{d}\omega \quad . \tag{4.25}$$

Here, $\mathcal{R}_i$ and $\mathcal{L}_i$ denote the right and left factors of $\mathbf{V}$ respectively, such that

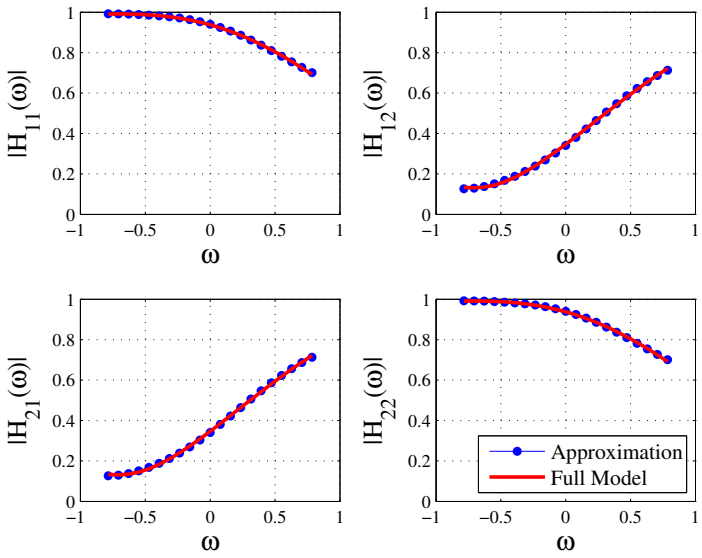


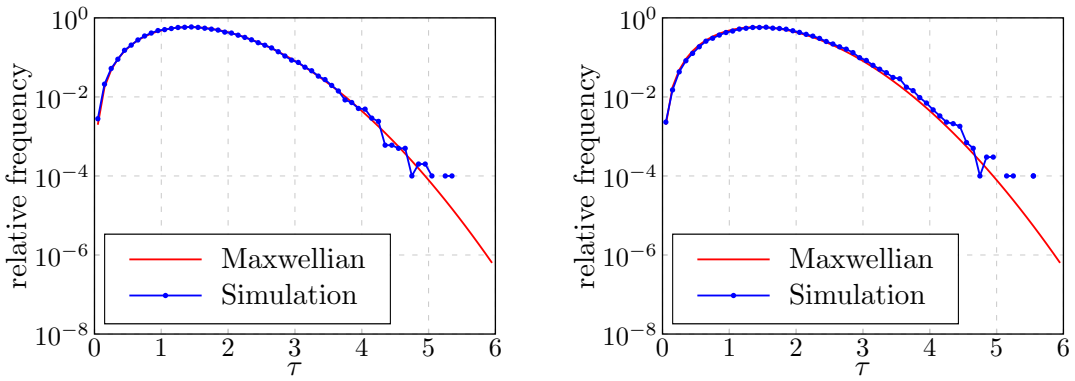
Figure 4.13: The result of the greedy approximation on one sample transfer function. Continuous lines represent the original transfer function and the dots represent the approximation.

$\mathbf{V}(e^{j\omega}) = \mathcal{L}_i(e^{j\omega})\mathbf{H}_i(e^{j\omega})\mathcal{R}_i(e^{j\omega})$. The minimization of this expression is achieved by setting $\mathbf{v}_i$ equal to the corresponding normalized eigenvector of the smallest eigenvalue of $\mathbf{B}$. It is shown in [57] that with every such iteration the error term in (4.23) is reduced. Hence, the algorithm moves forward by optimizing each section and the paraunitary matrix iteratively.

The statistical behavior of this method is demonstrated in Figures 4.14 and 4.15. It is obvious that greedy approximation method outperforms the previous two strategies at the cost of increased computational complexity.

4.4. Conclusion

Building upon an earlier work on PMD emulation and compensation which uses model reduction techniques and Markov chain Monte Carlo methods, we have presented an improved sampling procedure to generate low complexity continuous-time



(a) PDF at $\omega = 0$ (b) PDF at $\omega = \frac{\pi}{4}$

Figure 4.14: PDF of the DGD at the center frequency ($\omega = 0$) and the corner frequency ($\omega = \frac{\pi}{4}$). Red curves represent the Maxwellian with mean 1.6.

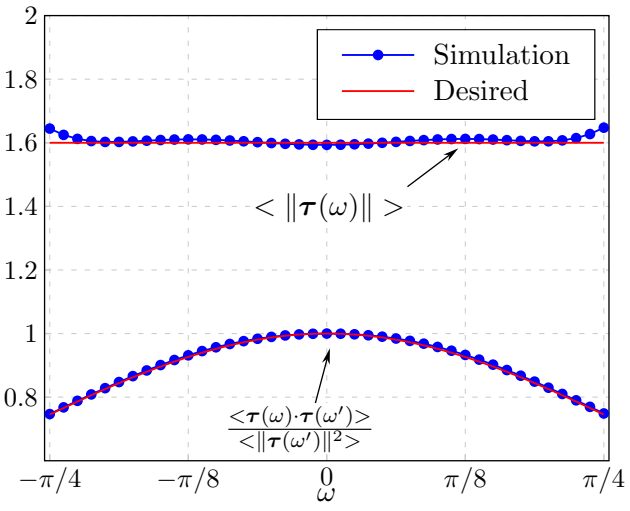


Figure 4.15: Mean and normalized autocorrelation curves of the greedy approximation method compared with the expected values.

models of an optical fiber subject to PMD. Introducing a modification to the standard Metropolis-Hastings algorithm, we have resolved a discrepancy of the sampling method in [18] and obtained better results in terms of approximating the desired DGD statistics. The models obtained with our technique preserve losslessness and offer a close approximation to PMD statistics while being computationally efficient. Therefore they can be used for efficient emulation of PMD in optical fibers.

For the emulation of PMD channels on discrete time devices we have presented a paraunitary FIR filter structure. This structure constitutes a good candidate for PMD emulation not only because of its losslessness property but also for its cascaded nature that enables one to adjust the complexity of the filter. Using three different approaches we have addressed the question of how the parameters of an ensemble of such filters have to be chosen for them to capture the statistical behavior of a real PMD channel. Using theoretical and simulation results we have shown that the cascading method can be employed for systems that suffer mainly from effects that can be described by the first order statistics of PMD. For higher order effects one can use the compensated MCMC algorithm which provides a good approximation for the mean and the autocorrelation values of the PMD vector. The final approach approximates the transfer function of an optical fiber in the frequency range of interest and hence provides the best results in terms of desired statistics. The choice of appropriate method depends on the trade-off between computational cost and the statistical accuracy.

Chapter 5

Summary and Conclusion

We have presented continuous and discrete-time models for PMD emulation and random parameter sampling schemes for these models with the goal of accurately emulating PMD in terms of desired statistical properties of a real single mode optical fiber. As an introduction to our treatment, in Chapter 1 we have discussed why accurate PMD emulation is necessary as well as the fundamentals of the mathematical description of PMD and its statistical behavior.

Chapter 2 was dedicated to the presentation of different PMD models relevant to our discussion. First of these models was the continuous-time lumped system with a high number of birefringent sections which serves as a realistic, high complexity representation of an optical fiber. This model was employed to determine the desired statistical properties of a PMD emulator as well as the accuracy of the approximations used to capture the statistical nature of PMD channels. Based on the full model, we continued our discussion with the reduced complexity continuous time model. This model is obtained by applying Pade approximation and Krylov subspace techniques to reduce the number of birefringent sections in a full model. Investigating its Laplace domain representation, we have made an important observation about the distribution of the pole locations of the reduced complexity model, which was our first step towards the random parameter sampling problem for PMD emulators. The last PMD model

we presented was designed to be implemented on DSPs in order to be used with the emerging coherent receivers technologies. This model was based on discrete-time paraunitary filters which inherently satisfy the losslessness condition of a PMD channel. The statistical properties of these models was also discussed from which we concluded that a specifically tailored random parameter sampling scheme was needed in order to be able to produce viable PMD emulators.

Before continuing with the solution of this parameter sampling problem, we made a detour to introduce an extension of the standard MCMC method for random input selection tasks for general complex models. In this chapter we developed a sampling scheme based on a modification to the Metropolis-Hastings algorithm with an additional expression which can be viewed as a probing term for the model of interest. The proposed algorithm is easy to implement, benefits from the extensive literature on MCMC and hence we believe that it can be adapted to a variety of applications. We have demonstrated one such application on general stochastic differential equations viewing them from the perspective of stochastic input-output models enabling us to apply our algorithm to obtain solution paths.

The compensated MCMC algorithm found further use during the course of our discussion in Chapter 4. Using the proposed sampling method, we have resolved a discrepancy that was present in earlier treatments of the continuous-time reduced complexity model and obtained better results in terms of approximating the desired DGD statistics. For the emulation of PMD channels on discrete time devices we have used three different approaches and addressed the question of how the parameters of an ensemble of such filters have to be chosen in order to capture the statistical behavior of a real PMD channel. Using theoretical and simulation results we have shown that the cascading method can be employed for communication systems which suffer mainly from first order effects of PMD. For higher order effects, one can use the compensated MCMC algorithm which provides a good approximation for the mean and the autocorrelation values of the PMD vector. The final approach approximates the transfer function of an optical fiber in the frequency range of interest and hence

provides the best results in terms of desired statistics. The choice of appropriate sampling method for a particular communication system depends on the requirements of statistical accuracy and computational complexity.

The methods presented in this thesis can be used to construct versatile and efficient PMD emulators as parts of continuous as well as discrete-time communication systems. The discrete-time models are of particular interest because they can be employed in testing of systems operating with coherent receivers built on digital signal processing algorithms. In parallel to the development of these devices, our algorithms can be extended to incorporate the evolution of PMD effects in time such that PMD channels can be emulated with models which not only offer a batch of independent optical fiber realisations but also are able to capture the gradual changes a single optical fiber undergoes. This task still stands as an open problem and can constitute the main objective of further work on this subject.

Bibliography

[1] C. Poole and J. Nagel, "Polarization effects in lightwave systems," *Optical Fiber Telecommunications*, vol. 3, pp. 114–161, 1997.

[2] J. Gordon and H. Kogelnik, "Pmd fundamentals: Polarization mode dispersion in optical fibers," *Proceedings of the National Academy of Sciences*, vol. 97, no. 9, p. 4541, 2000.

[3] J. Damask, *Polarization optics in telecommunications*. Springer Verlag, 2005, vol. 101.

[4] M. Karlsson, "Polarization mode dispersion induced pulse broadening in optical fibers," *Optics letters*, vol. 23, no. 9, pp. 688–690, 1998.

[5] S. Rashleigh and R. Ulrich, "Polarization mode dispersion in single-mode fibers," *Optics Letters*, vol. 3, no. 2, pp. 60–62, 1978.

[6] C. Poole and R. Wagner, "Phenomenological approach to polarisation dispersion in long single-mode fibres," *Electronics Letters*, vol. 22, no. 19, pp. 1029–1030, 1986.

[7] A. Erdogan, A. Demir, and T. Oktem, "Automatic pmd compensation by unsupervised polarization diversity combining coherent receivers," *Journal of Lightwave Technology*, vol. 26, no. 13, pp. 1823–1834, 2008.

[8] T. Oktem, A. Erdogan, and A. Demir, "Adaptive receiver structures for fiber communication systems employing polarization-division multiplexing," *Lightwave Technology, Journal of*, vol. 27, no. 23, pp. 5394–5404, 2009.

[9] S. Savory, "Digital filters for coherent optical receivers," *Optics Express*, vol. 16, no. 2, pp. 804–817, 2008.

[10] B. Spinnler, "Equalizer design and complexity for digital coherent receivers," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 16, no. 5, pp. 1180–1192, 2010.

[11] J. Faure, B. Lavigne, C. Bresson, O. Bertran-Pardo, A. Colomer, and R. Cantó, "40g and 100g deployment on 10g infrastructure: market overview and trends, coherent versus conventional technology," in *Optical Fiber Communication Conference*. Optical Society of America, 2010.

[12] I. Lima, R. Khosravani, P. Ebrahimi, E. Ibragimov, C. Menyuk, and A. Willner, "Comparison of polarization mode dispersion emulators," *Journal of Lightwave Technology*, vol. 19, no. 12, p. 1872, 2001.

[13] M. Karlsson and J. Brentel, "Autocorrelation function of the polarization-mode dispersion vector," *Optics Letters*, vol. 24, no. 14, pp. 939–941, 1999.

[14] M. Shtaif and A. Mecozzi, "Study of the frequency autocorrelation of the differential group delay in fibers with polarization mode dispersion," *Optics Letters*, vol. 25, no. 10, pp. 707–709, 2000.

[15] J. Damask, "A programmable polarization-mode dispersion emulator for systematic testing of 10 gb/s pmd compensators," in *Optical Fiber Communication Conference, 2000*, vol. 3. IEEE, 2000, pp. 28–30.

[16] J. Damask, G. Simer, K. Rochford, and P. Myers, "Demonstration of a programmable pmd source," *Photonics Technology Letters, IEEE*, vol. 15, no. 2, pp. 296–298, 2003.

[17] R. Noe, D. Sandel, M. Yoshida-Dierolf, S. Hinz, V. Mirvoda, C. Glingener, E. Gottwald, C. Scheerer, G. Fischer, T. Weyrauch *et al.*, "Polarization mode dispersion compensation at 10, 20, and 40 gb/s with various optical equalizers," *Journal of lightwave Technology*, vol. 17, no. 9, p. 1602, 1999.

[18] A. Demir and A. Erdogan, "Emulation and inversion of polarization mode dispersion: A lossless system and reduced-order modeling perspective," *J. Lightw. Technol*, vol. 26, no. 17, pp. 3071–3089, 2008.

[19] G. Foschini and C. Poole, "Statistical theory of polarization dispersion in single mode fibers," *Lightwave Technology, Journal of*, vol. 9, no. 11, pp. 1439–1456, 1991.

[20] F. Curti, B. Daino, G. De Marchis, and F. Matera, "Statistical treatment of the evolution of the principal states of polarization in single-mode fibers," *Lightwave Technology, Journal of*, vol. 8, no. 8, pp. 1162–1166, 1990.

[21] G. Agrawal, "Nonlinear fiber optics," *Nonlinear Science at the Dawn of the 21st Century*, pp. 195–211, 2000.

[22] R. Mearz, *Integrated optics: design and modeling.* Artech House, 1995.

[23] S. Savory and F. Payne, "Pulse propagation in fibers with polarization-mode dispersion," *Lightwave Technology, Journal of*, vol. 19, no. 3, pp. 350–357, 2001.

[24] G. Baker and P. Graves-Morris, *Padé approximants.* Cambridge Univ Pr, 1996, vol. 59.

[25] M. Wohlers, *Lumped and distributed passive networks: A generalized and advanced viewpoint*. Academic press, 1969.

[26] P. Vaidyanathan and Z. Doganata, "The role of lossless systems in modern digital signal processing: A tutorial," *Education, IEEE Transactions on*, vol. 32, no. 3, pp. 181–197, 1989.

[27] P. Vaidyanathan, *Multirate systems and filter banks*. Pearson Education India, 1993.

[28] M. Karlsson, "Probability density functions of the differential group delay in optical fiber communication systems," *Lightwave Technology, Journal of*, vol. 19, no. 3, pp. 324–331, 2001.

[29] C. Robert and G. Casella, *Monte Carlo statistical methods*. Springer Verlag, 2004.

[30] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller *et al.*, "Equation of state calculations by fast computing machines," *The journal of chemical physics*, vol. 21, no. 6, p. 1087, 1953.

[31] W. Hastings, "Monte carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[32] J. Huelsenbeck and F. Ronquist, "Mrbayes: Bayesian inference of phylogenetic trees," *Bioinformatics*, vol. 17, no. 8, pp. 754–755, 2001.

[33] D. Heermann, "Computer simulation methods: in theoretical physics," 1986.

[34] B. Ripley and E. Corporation, *Stochastic simulation*. Wiley Online Library, 1987, vol. 21.

[35] L. Devroye, "Non-uniform random variate generation," 1986.

[36] C. Andrieu, N. De Freitas, A. Doucet, and M. Jordan, "An introduction to mcmc for machine learning," *Machine learning*, vol. 50, no. 1, pp. 5–43, 2003.

[37] R. Neal and U. of Toronto. Department of Computer Science, *Probabilistic inference using Markov chain Monte Carlo methods*. Citeseer, 1993.

[38] P. Dellaportas, J. Forster, and I. Ntzoufras, "On bayesian model and variable selection using mcmc," *Statistics and Computing*, vol. 12, no. 1, pp. 27–36, 2002.

[39] A. Mira and C. Geyer, "On non-reversible markov chains," *Monte Carlo methods*, vol. 26, p. 95, 2000.

[40] K. Athreya, H. Doss, and J. Sethuraman, "On the convergence of the markov chain simulation method," *The Annals of Statistics*, vol. 24, no. 1, pp. 69–100, 1996.

[41] L. Tierney, "Markov chains for exploring posterior distributions," *the Annals of Statistics*, vol. 22, no. 4, pp. 1701–1728, 1994.

[42] ——, "A note on metropolis-hastings kernels for general state spaces," *Annals of Applied Probability*, pp. 1–9, 1998.

[43] C. Bishop and SpringerLink (Service en ligne), *Pattern recognition and machine learning*. Springer New York, 2006, vol. 4.

[44] R. Redner and H. Walker, "Mixture densities, maximum likelihood and the em algorithm," *SIAM review*, pp. 195–239, 1984.

[45] B. Øksendal, *Stochastic differential equations: an introduction with applications*. Springer Verlag, 2003.

[46] F. Black and M. Scholes, "The pricing of options and corporate liabilities," *The journal of political economy*, pp. 637–654, 1973.

[47] P. Kloeden, E. Platen, and H. Schurz, "Stochastic differential equations," *Numerical Solution of SDE Through Computer Experiments*, pp. 63–90, 1994.

[48] R. Khosravani, I. Lima Jr, P. Ebrahimi, E. Ibragimov, A. Willner, and C. Menyuk, "Time and frequency domain characteristics of polarization-mode dispersion emulators," *Photonics Technology Letters, IEEE*, vol. 13, no. 2, pp. 127–129, 2001.

[49] C. Antonelli and A. Mecozzi, "Statistics of the dgd in pmd emulators," *Photonics Technology Letters, IEEE*, vol. 16, no. 8, pp. 1840–1842, 2004.

[50] R. Hartwig and M. Putcha, "When is a matrix a sum of idempotents?" *Linear and Multilinear Algebra*, vol. 26, no. 4, pp. 279–286, 1990.

[51] T. Bedford and R. Cooke, "Vines–a new graphical model for dependent random variables," *The Annals of Statistics*, vol. 30, no. 4, pp. 1031–1068, 2002.

[52] R. Nelsen, *An introduction to copulas*. Springer Verlag, 2006.

[53] E. Frees and E. Valdez, "Understanding relationships using copulas," *North American actuarial journal*, vol. 2, no. 1, 1998.

[54] A. Sklar, "Fonctions de répartition à n dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris*, vol. 8, no. 1, p. 11, 1959.

[55] S. Demarta and A. McNeil, "The t copula and related copulas," *International Statistical Review*, vol. 73, no. 1, pp. 111–129, 2005.

[56] K. Aas, C. Czado, A. Frigessi, and H. Bakken, "Pair-copula constructions of multiple dependence," *Insurance: Mathematics and Economics*, vol. 44, no. 2, pp. 182–198, 2009.

[57] A. Tkacenko and P. Vaidyanathan, "Iterative greedy algorithm for solving the fir paraunitary approximation problem," *Signal Processing, IEEE Transactions on*, vol. 54, no. 1, pp. 146–160, 2006.

[58] R. Horn and C. Johnson, *Matrix analysis*. Cambridge Univ Pr, 1990.

Vita

Ahmet Gökçen Mahmutoğlu was born in Istanbul, Turkey, on 6 November 1984. Following his graduation from Sankt Georg High School in 2003, he started his electrical and electronics engineering education in Technical University of Vienna. He received his bachelor of science degree in 2008 and started working as an instrumentation engineer in the private sector. He is currently continuing his education at Koc University, Istanbul as a research and teaching assistant.