APPROVAL SHEET

Title of Thesis: Advanced Monte Carlo Methods for Computation of Penalties Induced by Polarization Mode Dispersion in Optical Fiber Transmission Systems

Name of Candidate: Aurenice de Menezes Oliveira Lima
Doctor of Philosophy, 2005

Dissertation and Abstract Approved: _____

Professor Curtis R. Menyuk
Computer Science and Electrical Engineering

Date Approved: _____

Curriculum Vitae

Aurenice de Menezes Oliveira Lima
2922 7th Street North Apt. 2
Fargo, ND 58102 USA
Phone: 701-361-5627
e-mail: alima1@umbc.edu

Objective:
Tenure-track faculty position in Electrical Engineering or Mathematics

Birth:
Acajutiba, Bahia, Brazil, in March 22, 1971

Education:

Ph.D. **Electrical Engineering: Communications and Signal Processing**
University of Maryland Baltimore County, USA, August 2005
Dissertation: Advanced Monte Carlo Methods for Computation of Penalties Induced by Polarization Mode Dispersion in Optical Fiber Transmission Systems
Advisor: Dr. Curtis R. Menyuk

M.Sc. **Electrical Engineering: Electronics and Communications**
University of Campinas (Unicamp), Brazil, February 1998
Dissertation: Dimensioning of Common Channel Signaling Networks
Advisor: Dr. Ivanil S. Bonatti

B.Sc. **Electrical Engineering: Electronics and Communications**
Federal University of Bahia, Brazil, December 1995

Technical **Electronics**
Federal Technical School of Bahia, Brazil, September 1990

Employment:

Starting 8/2005 — *Adjunct faculty member*, **Minnesota State University, Moorhead**
Mathematics Department.

8/2000–8/2005 — *Research Assistant*, **University of Maryland, Baltimore County**
Research novel techniques for the mitigation of the limiting effects to the capacity of optical fiber transmission systems, and research advanced Monte Carlo methods for computation of penalties induced by polarization mode dispersion.

1/2005–5/2005 — *Adjunct faculty member*, **North Dakota State University**
Spring 2005: Instructor of the courses Electrical Engineering II (ECE 303), and Laboratory of Electrical Engineering II (ECE 306).

9/2003–12/2004 — *Visiting Scholar*, **North Dakota State University**
Conduct research in Electrical Engineering, collaborating with faculty at both North Dakota State University and University of Maryland Baltimore County.

4/1992–4/1995 — *Electronics Technician*, **Prologica Systems**
Salvador, Bahia, Brazil.
Maintenance of computers and computer peripheral devices.

1/1991–10/1991 — *Electronics Technician* (intern), **Telebahia**
(Former telecommunications agency of the state of Bahia),
Salvador, Bahia, Brazil.
Dimensioning of telephone networks, instalation and maintenance of telephone network equipments.

Awards and honors:

09/24/2003 – Venice Summer School on Polarization Mode Dispersion Award for paper that appeared in the proceedings of the 29^{th} European Conference on Optical Communication (ECOC 2003). The award was presented at ECOC 2003 in Italy.

03/01/1998 – Graduate Scholarship at Ph.D. degree level from the National Research Council (CNPq) of the Brazilian Ministry of Science and Technology for graduate education at the State University of Campinas, Campinas, Brazil. Period: 3/1998 to 8/1998.

03/01/1996 – Graduate Scholarship at M.Sc. degree level from CAPES of the Brazilian Ministry of Education for graduate education at the State University of Campinas, Campinas, Brazil. Period: 3/1996 to 2/1998.

Languages:

— English: Read, write, and speak fluently.
— Portuguese: Native speaker.
— Spanish: Read, communicate.

Professional societies:

— Lasers & Electro-Optics Society (LEOS) of the Institute of Electrical and Electronics Engineers (IEEE)
— IEEE Women in Engineering Society

Service:

Referee for *IEEE/OSA Journal of Lightwave Technology*, and *IEEE Photonics Technology Letters*.

Archival journal publications:

1) A. O. Lima, I. T. Lima, Jr., and C. R. Menyuk, "Error estimation in multicanonical Monte Carlo simulations with application to polarization mode dispersion emulators," to appear in *IEEE/OSA Journal of Lightwave Technology*, 2005.

2) A. O. Lima, C. R. Menyuk, and I. T. Lima, Jr., "Comparison of two biasing Monte Carlo methods for calculating outage probability in systems with multi-section PMD compensators," accepted subject to minor revisions by *IEEE Photonics Technology Letters*, 2005.

3) A. O. Lima, I. T. Lima, Jr., C. R. Menyuk, G. Biondini, and W. L. Kath, "Statistical analysis of the performance of PMD compensators using multiple importance

sampling," *IEEE Photonics Technology Letters*, Vol. 15, No. 12, pp. 1716–1718, December 2003.

4) A. O. Lima, I. T. Lima, Jr., C. R. Menyuk, and T. Adali, "Comparison of penalties resulting from first-order and all-order polarization mode dispersion in optical fiber transmission systems," *Optics Letters*, Vol. 28, No. 5, pp. 310–311, March 2003.

5) A. O. Lima, I. T. Lima, Jr., T. Adali and C. R. Menyuk, "A Novel Polarization Diversity Receiver for PMD Mitigation," *IEEE Photonics Technology Letters*, Vol. 14, No. 4, pp. 465–467, April 2002.

6) I. T. Lima, Jr., A. O. Lima, Y. Sun, H. Jiao, J. Zweck, C. R. Menyuk, and G. M. Carter, "Accurate receiver model for optical fiber communication systems with arbitrarily polarized noise," to appear in the March 2005 issue of *IEEE/OSA Journal of Lightwave Technology*.

7) I. T. Lima, Jr., A. O. Lima, G. Biondini, C. R. Menyuk, and W. L. Kath, "A comparative study of single-section polarization-mode dispersion compensators," *IEEE/OSA Journal of Lightwave Technology*. Vol. 22, No. 4, pp. 1023–1032, April 2004.

8) I. T. Lima, Jr., A. O. Lima, J. Zweck, and C. R. Menyuk, "Performance characterization of chirped return-to-zero modulation format using an accurate receiver model," *IEEE Photonics Technology Letters*, Vol. 15, No. 4, pp. 608–610, April 2003.

9) I. T. Lima, Jr., A. O. Lima, J. Zweck, and C. R. Menyuk, "Efficient computation of outage probabilities due to polarization effects in a WDM system using a reduced Stokes model and importance sampling," *IEEE Photonics Technology Letters*, Vol. 15, No. 1, pp. 45–47, January 2003.

10) Y. Sun, A. O. Lima, I. T. Lima, Jr., J. Zweck, L. Yan, C. R. Menyuk, and G. M. Carter, "Statistics of the system performance in scrambled recirculating loop with PDL and PDG," *IEEE Photonics Technology Letters*, Vol. 15, No. 8, pp. 1067–1069, August 2003.

11) Y. Sun, I. T. Lima, Jr., A. O. Lima, H. Jiao, J. Zweck, L. Yan, C. R. Menyuk, and G. M. Carter, "System variations due to partially polarized noise in a receiver," *IEEE Photonics Technology Letters*, Vol. 15, No. 11, pp. 1648–1650, November 2003.

12) J. Zweck, I. T. Lima, Jr., Y. Sun, A. O. Lima, C. R. Menyuk, and G. M. Carter, "Modeling Receivers in Optical Communications Systems with Polarization Effects," *OSA Optics and Photonics News*, Vol. 14, No. 11, pp. 30–35, November 2003.

Contributed papers at conferences:

1) A. O. Lima, I. T. Lima, Jr., C. R. Menyuk, and J. Zweck, "Performance evaluation of single-section and three-section PMD compensators using extended Monte Carlo methods," *in Proceedingd of Optical Fiber Communication Conference and Exposition and National Fiber Optic Engineers Conference (OFC/NFOEC) 2005*, Anaheim, California, USA, March 6–11, 2005, paper OME27.

2) A. O. Lima, I. T. Lima, Jr., J. Zweck, and C. R. Menyuk, "Efficient computation of PMD-induced penalties using Multicanonical Monte Carlo simulations," to appear *in Proceedings of the 29th European Conference on Optical Communication (ECOC) 2003*, Rimini, Italy, September 21–25, 2003, paper We364.

3) A. O. Lima, I. T. Lima, Jr., B. S. Marks, C. R. Menyuk, and W. L. Kath, "Performance analysis of single-section PMD compensators using multiple importance sampling," *in Proceedings of the Optical Fiber Communication Conference and Exposition (OFC) 2003*, Atlanta, Georgia, USA, March 23–28, 2003, paper ThA3.

4) A. O. Lima, I. T. Lima, Jr., T. Adali, and C. R. Menyuk, "Comparison of power penalties due to first- and all-order PMD distortions," *in Proceedings of the 28th European Conference on Optical Communication (ECOC) 2002*, Copenhagen, Denmark, September 8–12, 2002, paper 7.1.2.

5) A. O. Lima, I. T. Lima, Jr., T. Adali, and C. R. Menyuk, "Compensation of polarization mode dispersion in optical fiber transmission systems using a polarization diversity receiver," *Venice Summer School on Polarization Mode Dispersion (VSS) 2002*, Venice, Italy, June 24–26, 2002.

6) A. O. Lima, T. Adali, I. T. Lima, Jr., and C. R. Menyuk, "Polarization diversity and equalization for PMD mitigation," *in Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP) 2002*, Orlando, Florida, USA, May 13–17, 2002, Vol. III, pp. 2721–2724.

7) A. O. Lima, T. Adali, I. T. Lima, Jr., and C. R. Menyuk, "Polarization diversity receiver for PMD mitigation," *in Proceedings of the Optical Fiber Communication Conference and Exhibit (OFC) 2002*, Anaheim, California, USA, March 17–22, 2002, paper WI7.

8) A. O. Lima, I. T. Lima, Jr., T. Adali, and C. R. Menyuk, "PMD Mitigation Using Diversity Detection," *in Proceedings of the IEEE LEOS Summer Topical Meeting 2001*, Copper Mountain, Colorado, USA, July 30–1 August 2001, paper MD3.3.

9) A. O. Lima, I. S. Bonatti, A. K. Budri, and P. L. D. Peres, "Common Channel Signaling Networks: Implementation," *in Proceedings of the V Teletraffic Brazilian Symposium*, Recife-PE, Brazil, September 17-22, 1997.

10) I. T. Lima, Jr. and A. O. Lima, "Computation of the probability of power penalty and Q-penalty outages due to PMD," *in Proceedings of the LEOS Annual meeting 2003*, Tucson, Arizona, USA, October 26–30, 2003, paper TuQ1.

11) I. T. Lima, Jr., A. O. Lima, J. Zweck, and C. R. Menyuk, "An accurate formula for the Q-factor of a fiber transmission system with partially polarized noise," *in Proceedings of the Conference on Lasers and Electro Optics (CLEO) 2003*, Baltimore, Maryland, USA, June 1–6, 2003, paper CThJ2.

12) I. T. Lima, Jr., A. O. Lima, J. Zweck, and C. R. Menyuk, "Computation of the Q-factor in optical fiber systems using an accurate receiver model," *in Proceedings of the Optical Fiber Communication Conference and Exposition (OFC) 2003*, Atlanta, Georgia, U.S.A, March 23–28, 2003, paper MF81.

13) Y. Sun, A. O. Lima, I. T. Lima, Jr., L. Yan, J. Zweck, C. R. Menyuk, and G. M. Carter, "Accurate Q-factor distributions in optical transmission systems with polarization effects," *in Proceedings of the Optical Fiber Communication Conference and Exposition (OFC) 2003*, Atlanta, Georgia, USA, March 23–28, 2003, paper ThJ4.

14) I. T. Lima, Jr., A. O. Lima, J. Zweck, and C. R. Menyuk, "Computation of the penalty due to the polarization effects in a wavelength-division multiplexed system using a reduced Stokes model with a realistic receiver," *Venice Summer School on Polarization Mode Dispersion (VSS) 2002*, Venice, Italy, June 24–26, 2002.

15) I. T. Lima, Jr., A. O. Lima, Y. Sun, J. Zweck, B. S. Marks, G. M. Carter, and C. R. Menyuk, "Computation of the outage probability due to the polarization effects using importance sampling," *in Proceedings of the Optical Fiber Communication Conference and Exhibit (OFC) 2002*, Anaheim, California, USA, March 17–22, 2002, paper TuI7.

16) T. Adali, W. Wang, A. O. Lima, "Electronic Equalization in Optical Fiber Communications," *in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2003*, Hong Kong, China, April 6-10, 2003, Vol.4, pp.497-500.

17) H. Jiao, I. T. Lima, Jr., A. O. Lima, Y. Sun, J. Zweck, L. Yan, C. R. Menyuk, and G. M. Carter, "Experimental validation of an accurate receiver model for systems with

unpolarized noise," *in Proceedings of the Conference on Lasers and Electro Optics (CLEO) 2003*, Baltimore, Maryland, USA, June 1–6, 2003, paper CThJ1.

18) J. Zweck, S. E. Minkoff, A. O. Lima, I. T. Lima, Jr., and C. R. Menyuk, "A comparative study of feedback controller sensitivity to all orders of PMD for a fixed DGD compensator," *in Proceedings of the Optical Fiber Communication Conference and Exposition (OFC) 2003*, Atlanta, Georgia, USA, March 23–28, 2003, paper ThY2.

19) W. Xi, T. Adali, A. O. Lima, W. Wang, J. Zweck, and C. R. Menyuk, "Electrical estimation of polarization mode dispersion parameters for compensation," *in Proceedings of the Optical Fiber Communication Conference and Exposition (OFC) 2003*, Atlanta, Georgia, USA, March 23–28, 2003, paper TuO5.

20) Y. Sun, I. T. Lima, Jr., A. O. Lima, H. Jiao, J. Zweck, L. Yan, C. R. Menyuk, and G. M. Carter, "Effects of partially polarized noise in a receiver," *in Proceedings of the Optical Fiber Communication Conference and Exposition (OFC) 2003*, Atlanta, Georgia, USA, March 23–28, 2003, paper MF82.

21) S. E. Minkoff, J. W. Zweck, A. O. Lima, I. T. Lima, Jr., and C. R. Menyuk, "Numerical Simulation and Analysis of Fiber Optic Compensators," *Society for Industrial and Applied Mathematics (SIAM) Annual Meeting*, Montreal, Canada, June 16-20, 2003.

Other publications:

1) A. O. Lima, I. T. Lima, Jr., T. Adali, and C. R. Menyuk, "Polarization pairing kills distortion," *EE TIMES online, In Focus: Communications*, Article ID 16504495, URL: http://www.eetimes.com, March 18, 2002.

2) A. O. Lima, "Dimensioning of Common Channel Signaling Networks," Master Thesis, State University of Campinas (Unicamp), Brazil, February 1998.

Abstract

Title of Dissertation: Advanced Monte Carlo Methods for Computation of Penalties Induced by Polarization Mode Dispersion in Optical Fiber Transmission Systems

Aurenice de Menezes Oliveira Lima, Doctor of Philosophy, 2005

Dissertation directed by: Professor Curtis R. Menyuk
Computer Science and Electrical Engineering

Polarization mode dispersion (PMD) is a major source of impairments in optical fiber communication systems. PMD broadens the optical pulses carrying the information and leads to inter-symbol interference. In long-haul transmission systems it is necessary to limit the penalty caused by polarization effects, so that the probability of exceeding a maximum specified penalty, such as 1 dB, will be small, typically 10^{-5} or less. This probability is referred to as the outage probability. Because of this stringent requirement, it has been very difficult to use either standard, unbiased Monte Carlo simulations or laboratory experiments to determine the outage probability of a system. Only outage probabilities larger than 10^{-4} can be efficiently computed with

standard, unbiased Monte Carlo simulations. A very large number of samples must be explored using standard Monte Carlo simulations in order to obtain an accurate estimate of rare events that lead to penalties of interest to system designers, which is computationally costly. To overcome this hurdle, advanced Monte Carlo methods such as importance sampling and multicanonical Monte Carlo (MMC) methods have recently been applied to compute PMD-induced penalties using a much smaller number of samples.

In this Ph.D. dissertation, I present work in which my colleagues and I investigated and applied importance sampling and MMC to accurately and efficiently compute penalties caused by PMD. These techniques allow low probability events to be efficiently computed by enabling one to concentrate the selection of samples on the most significant regions of the sample space. Even though these techniques are well-known in statistics, statistical physics, and some areas of communications, they have only recently been applied to optical fiber communication systems. Using these statistical techniques, we studied the performance of PMD compensators and compared the efficiency of these two advanced Monte Carlo methods to compute penalties of different types of compensated systems. We also used importance sampling to compare the penalty resulting from first-order and all-order PMD models, demonstrating the importance of accurately modeling PMD by including higher orders.

Since Monte Carlo methods are statistical, error estimates are essential to verify the accuracy of the results. MMC is a highly nonlinear, iterative method that generates correlated samples, so that standard error estimation techniques cannot be applied. A more sophisticated approach is needed. On the other hand, one can suc-

cessfully apply standard error estimation techniques and first-order error propagation to estimate errors in importance sampling simulations. In this Ph.D. dissertation, I also report the contribution that we made to estimate the statistical errors when using importance sampling and multicanonical Monte Carlo methods. We developed an efficient numerical method to estimate statistical errors when using MMC, which we refer to as the transition matrix method. We showed that this method is a variant of the bootstrap method.

Advanced Monte Carlo Methods for Computation of Penalties Induced by Polarization Mode Dispersion in Optical Fiber Transmission Systems

by

Aurenice de Menezes Oliveira Lima

Dissertation submitted to the Faculty of the Graduate School
of the University of Maryland in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2005

© Copyright by Aurenice de Menezes Oliveira Lima, 2005

Acknowledgements

I am grateful to Dr. Curtis Menyuk, my dissertation advisor, for his support and guidance throughout my graduate education at UMBC, and for allowing me to continue my work toward the Ph. D. degree away from the UMBC campus after I became a Ph. D. candidate. I am also grateful to Dr. Tulay Adalı and to Dr. Gary Carter for giving me the opportunity to collaborate with their research groups. I would like to thank Dr. John Zweck and Dr. Brian Marks for their contributions to my graduate education at UMBC.

I am thankful to all my colleagues and friends at UMBC, and to my colleagues who co-authored the publications that arose from this work: Dr. Ivan Lima Jr., Dr. William Kath, Dr. Gino Biondini, Dr. Li Yan, Lyn Randers, Dr. Ronald Holzlöhner, Oleg Sinkin, Jonathan Hu, Anshul Kalra, Dr. Yu Sun, Hua Jiao, Jiping Wen, Dr. Hai Xu, Dr. Wenze Xi, and Wei Wang. I am also thankful to my colleagues and friends of the Department of Electrical and Computer Engineering at North Dakota State University, and especially Dr. Daniel Ewert, for granting me access to his department's facilities during the last part of the research that I report in this dissertation.

I would like to express my immense gratitude to my parents José Aurinho and Eunice, and my sister Alvanice for their prayers, and for having supported me throughout the long and difficult journey of my education. I also thank my friends in Brazil for their support and friendship. I dedicate this dissertation to all of them.

I conclude by thanking God for his blessings, for giving me the strength to overcome the difficulties during this period of my life, and for his unconditional love.

TABLE OF CONTENTS

List of Tables ..... **vii**

List of Figures ..... **viii**

1 Introduction **1**

2 Polarization Mode Dispersion (PMD) **9**

2.1 Physical description of PMD 9

2.2 Polarization mode dispersion statistics 14

2.3 Modeling of polarization mode dispersion 17

3 Importance sampling applied to PMD **21**

3.1 Importance sampling to compute outage probability 21

3.2 Single-section PMD compensator and definition of penalty 24

3.3 Importance sampling that biases only the DGD 28

3.4 Importance sampling that biases the magnitude of first- and second-order PMD 36

3.5 Estimators of the mean and of the variance 42

3.6 Comparison of penalties resulting from first-order and all-order PMD . . . 45

3.7 Analysis of the performance of single-section PMD compensators using importance sampling 51

4 Multicanonical Monte Carlo method for PMD-induced penalty 60

4.1 The Multicanonical Monte Carlo method 60

4.2 MMC implementation to PMD emulators 61

4.2.1 Summary of the MMC algorithm 63

4.3 Correlations 66

4.4 MMC computation of PMD-induced penalty in uncompensated and single-section compensated systems 69

5 Estimation of Errors in MMC simulations 74

5.1 Why a new error estimation procedure ? 74

5.2 New error estimation procedure 77

5.2.1 The transition matrix method 78

5.2.2 Bootstrap method 81

5.2.3 Assessing the error in the MMC error estimation 83

5.3 Application and validation 86

6 Comparison of two biasing Monte Carlo methods for calculating outage probabilities in systems with multi-section PMD compensators 93

6.1 MMC and importance sampling to compute PMD-induced penalties 94

6.2 PMD Compensators 96

6.2.1 Three-section compensator 97

6.3 Simulation results and discussions 98

7 Conclusions 107

Bibliography 111

List of Tables

5.1 Selected data points from the curves shown in Fig. 5.3. The columns from left to right show: the normalized DGD value, the analytical probability density function, the estimated probability density function, the standard deviation computed using the transition matrix method, and the relative variation. 91

List of Figures

3.1 The pdf of the normalized DGD, $|\boldsymbol{\tau}| / \langle |\boldsymbol{\tau}| \rangle$, plotted on a logarithmic scale with 80 bins. The squares show the results of Monte Carlo simulations with importance sampling in which we only biased the DGD using 3×10^4 samples. The solid line shows the Maxwellian distribution with the same mean. 32

3.2 The joint pdf of the normalized $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$ with 25×25 bins. The solid lines show the results of Monte Carlo simulations with importance sampling in which we only biased the DGD using 3 biases with 6×10^4 samples in each bias. The dashed line shows the contour level corresponding to 10% relative variation in the results using importance sampling. The dotted lines show the results of standard Monte Carlo simulations using 10^8 samples. The dot-dashed line shows the contour level corresponding to 10% relative variation in the results using standard Monte Carlo simulations with 10^8 samples. The contours of the joint pdf from bottom to top of the plot, are at 3×10^{-n} , $n = 1, \cdots, 7$ and 10^{-m}, $m = 1, \cdots, 11$. 35

3.3 The pdf of the normalized $|\boldsymbol{\tau}_\omega|$ plotted on a logarithmic scale with 60 bins. The squares show the results of Monte Carlo simulations with importance sampling in which we biased both the first- and second-order PMD using 10^5 samples. The solid line shows the results of the theoretical distribution of the length of the frequency derivative of the polarization dispersion vector. 40

3.4 The joint pdf of the normalized $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$ with 25×25 bins. The solid lines show the results with importance sampling using 10 biases with 6×10^4 samples in each bias. The dashed line shows the contour level corresponding to 10% relative variation in the results using importance sampling. The dotted lines show the results of standard Monte Carlo simulations using 10^8 samples. The dot-dashed line shows the contour level corresponding to 10% relative variation in the results using standard Monte Carlo simulations with 10^8 samples. The contours of the joint pdf from the bottom to the top of the plot, are at 3×10^{-n} , $n = 1, \cdots, 7$ and 10^{-m}, $m = 1, \cdots, 11$. 41

3.5 Outage probability as a function of the eye-opening penalty margin for a 10 Gbit/s NRZ system with pulse edge rise and fall times of 30 ps and mean DGD, $\langle|\boldsymbol{\tau}|\rangle$, equal to 14 ps. The dashed line shows the outage probability considering only first-order PMD, while the solid line shows the results considering all-order (first- and higher-order) PMD distortion. The outage probability is the probability that the eye-opening penalty exceeds the value displayed on the horizontal axis. . 47

3.6 Same as Fig. 3.5, except that $\langle|\boldsymbol{\tau}|\rangle = 20$ ps and the pulse edge rise and fall times are 30 ps. 48

3.7 Same as Fig. 3.5, except that $\langle|\boldsymbol{\tau}|\rangle = 20$ ps and the pulse edge rise and fall times are 5 ps. 49

3.8 Conditional expectation of the second derivative of the DGD, $|\boldsymbol{\tau}|_{\omega\omega}$, given a value of the DGD, $|\boldsymbol{\tau}|$. 50

3.9 Probability density function of the eye-opening penalty for a system with a mean DGD of 30 ps and a single-section compensator. (i) Solid line: Results using importance sampling in which we only biased the DGD. (ii) Dashed line: Results using importance sampling in which we biased both the first- and second-order PMD. The confidence interval is shown with error bars. 54

3.10 Uncompensated system: The dotted lines show contour plots of the joint pdf of the normalized first- $|\boldsymbol{\tau}|$ and second-order PMD $|\boldsymbol{\tau}_\omega|$. The solid and the dashed lines show contour plots of the conditional expectation of the eye-opening penalty and the confidence interval of the contour plots, respectively. The contours of the joint pdf from the bottom to the top of the plot, are at 3×10^{-n} , $n = 1, \cdots, 7$ and 10^{-m}, $m = 1, \cdots, 11$. The curves of the conditional expectation of the eye-opening penalty in dB from the bottom to the top of the plot, are at $0.1, 0.2, 0.4, 0.6, 0.9, 1.2, 1.6, 2.2, 3.2$. 55

3.11 Same set of curves of Fig. 3.10 for a compensated system with a fixed-DGD compensator with constant DGD element equal to 2.5 $\langle|\boldsymbol{\tau}|\rangle$. The curves of the conditional expectation of the eye-opening penalty in dB from the bottom to the top of the plot, are at $0.1, 0.2, 0.3, 0.4$. 56

3.12 Same set of curves of Fig. 3.10 for a compensated system with a variable-DGD compensator with eye opening maximization. The curves of the conditional expectation of the eye-opening penalty in dB from the bottom to the top of the plot, are at $0.1, 0.2, 0.3, 0.4$. 57

3.13 Same set of curves of Fig. 3.10 for a compensated system with a variable-DGD compensator with minimized DGD after compensation at the central frequency of the channel. The solid lines show the contours of the conditional expectation of the eye-opening penalty in dB from the bottom to the top of the plot, are at $0.1, 0.2, 0.3, 0.4, 0.6, 0.9$. . 58

3.14 Outage probability as a function of the eye-opening penalty margin. The outage probability ($\hat{P}_o$) is the probability that the penalty exceeds the value displayed on the horizontal axis. (i) Dashed-dotted line: Uncompensated case; (ii) Dashed line: Variable-DGD compensator with the compensated DGD minimized at the central frequency of the channel; (iii) Solid line: Fixed-DGD compensator with DGD element equal to 2.5 $\langle|\boldsymbol{\tau}|\rangle$ and maximized eye opening; (iv) Solid-dotted line: Variable-DGD compensator with maximized eye opening. The error bars show the confidence interval for the curves that have at least one bin whose relative error ($\hat{\sigma}_{\hat{P}_o}/\hat{P}_o$) exceeds 10%. For those curves, we show the error bars for one out of three consecutive bins. 59

4.1 Correlation coefficients between bin i and bin j ($1 \leq j \leq 80$) for the 80-section emulator, where the bin i corresponds to $\mathrm{DGD}_i = 30$ ps ($1 \times$ mean DGD). The correlation coefficients are computed using 32 standard MMC simulations. Each standard MMC simulation consists of 30 MMC iterations with $8,000$ samples. 67

4.2 Correlation coefficients between bin i and bin j ($1 \leq j \leq 80$) for the 80-section emulator, where the bin i corresponds to $\mathrm{DGD}_i = 45$ ps ($1.5 \times$ mean DGD). The correlation coefficients are computed using 32 standard MMC simulations. Each standard MMC simulation consists of 30 MMC iterations with $8,000$ samples. 68

4.3 Correlation coefficients between bin i and bin j ($1 \leq j \leq 80$) for the 80-section emulator, where the bin i corresponds to $\text{DGD}_i = 75$ ps ($2.5 \times$ mean DGD). The correlation coefficients are computed using 32 standard MMC simulations. Each standard MMC simulation consists of 30 MMC iterations with $8,000$ samples. 69

4.4 Outage probability as a function of the eye-opening penalty. (i) Dotted line: Uncompensated system with a mean DGD of 30 ps. (ii) Dashed line and (iii) Open circles: Results for a variable-DGD compensator, obtained using MMC and IS, respectively, for a system with mean DGD of 30 ps. (iv) Solid line and (v) Squares: Results for an uncompensated system with mean DGD of 15 ps, obtained using MMC and IS, respectively. 70

4.5 Uncompensated system with a mean DGD of 15 ps. The dotted lines show the contour plots of the joint pdf of the normalized $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$, obtained using IS. The solid lines show the average eye-opening penalty given a value of $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$, obtained using MMC. The contours of joint pdf from the bottom to the top of the plot, are at 3×10^{-n} , $n = 1, \cdots, 7$ and 10^{-m}, $m = 1, \cdots, 11$. The penalty contours in dB from the left to the right of the plot, are at 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6. 71

4.6 Same set of curves of Fig. 4.5 for a compensated system with a variable-DGD compensator. The penalty contours in dB from the bottom to the top of the plot, are at 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6. 72

5.1 Relative variation ($\hat{\sigma}_{\hat{P}_{\mathrm{DGD}}}/\hat{P}_{\mathrm{DGD}}$) of the pdf of the normalized DGD, $|\boldsymbol{\tau}|/\langle|\boldsymbol{\tau}|\rangle$, for the 15-section PMD emulator using 14 MMC iterations with $4,000$ samples. The confidence interval is given by (5.5) when we compute an ensemble of standard deviations using bootstrap resampling for each of the 100 pseudo-transition matrices. 85

5.2 Relative variation ($\hat{\sigma}_{\hat{P}_{\mathrm{DGD}}}/\hat{P}_{\mathrm{DGD}}$) of the pdf of the normalized DGD, $|\boldsymbol{\tau}|/\langle|\boldsymbol{\tau}|\rangle$. (i) Circles: Transition matrix method based on a single standard MMC simulation for the 15-section PMD emulator; (ii) Solid: 10^3 standard MMC simulations for the 15-section emulator; (iii) Dashed: Confidence interval of the relative variation of the error estimated using the transition matrix method for the 15-section PMD emulator; (iv) Squares: Transition matrix method based on a single standard MMC simulation for the 80-section PMD emulator; (v) Dot-dashed: 10^3 standard MMC simulations for the 80-section PMD emulator. 87

5.3 The pdf of the normalized DGD, $|\boldsymbol{\tau}|/\langle|\boldsymbol{\tau}|\rangle$, for the 15-section PMD emulator using 14 MMC iterations with $4,000$ samples. (i) Diamonds: DGD pdf with error estimation using the transition matrix method, (ii) Dashed line: Maxwellian pdf, (iii) Solid line: Analytical pdf of the DGD for the 15-section PMD emulator. 89

5.4 The pdf of the normalized DGD, $|\boldsymbol{\tau}| / \langle |\boldsymbol{\tau}| \rangle$, for the 80-section PMD emulator using 30 MMC iterations with $8,000$ samples. (i) Diamonds: DGD pdf with error estimation using the transition matrix method, (ii) Dashed line: Maxwellian pdf, (iii) Solid line: Analytical pdf of the DGD for the 80-section PMD emulator. 90

6.1 Outage probability for a 1-dB penalty as function of the DGD element (τ_c) of the three-section compensator for a system with mean DGD of 30 ps. 100

6.2 Outage probability as a function of the eye-opening penalty for a system with mean DGD of 30 ps. (i) Dashed line (MMC) and triangles (IS): Uncompensated system. (ii) Dot-dashed line (MMC) and circles (IS): System with a single-section compensator. (iii) Solid line (MMC) and diamonds (IS): System with a three-section compensator. The error bars show the confidence interval for the MMC results. 101

6.3 Conditional expectation of the magnitude of the normalized second-order PMD, $|\boldsymbol{\tau}_\omega|$, given a value of the DGD of the transmission line, $|\boldsymbol{\tau}|$. Conditional expectation before (dashed) and after (solid) the three-section compensator. 103

6.4 Conditional expectation of the magnitude of the normalized third-order PMD, $|\boldsymbol{\tau}_{\omega\omega}|$, given a value of the DGD of the transmission line, $|\boldsymbol{\tau}|$. Conditional expectation before (dashed) and after (solid) the three-section compensator. 104

6.5 Conditional expectation of the magnitude of the normalized fourth-order PMD, $|\boldsymbol{\tau}_{\omega\omega\omega}|$, given a value of the DGD of the transmission line, $|\boldsymbol{\tau}|$. Conditional expectation before (dashed) and after (solid) the three-section compensator. 105

6.6 Three-section compensated system. The dotted lines are contour plots of the joint pdf of the normalized $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_{\omega}|$ from the bottom to the top of the plot, are at 3×10^{-n}, with $n = 1, \cdots, 7$ and 10^{-m}, with $m = 1, \cdots, 11$. The solid lines are contour plots of the conditional expectation of the eye-opening penalty in dB from the bottom to the top of the plot, are at 0.1, 0.2, 0.3, 0.4, 0.5, 0.6. 106

Chapter 1

Introduction

The rapid increase in the demand for bandwidth is driving most telecommunication operators toward the deployment of large capacity transmission systems. The data rates per channel in commercial systems have increased to the point that 10 Gbit/s systems have been on the market for many years, and systems with channel rates of 40 Gbit/s are beginning to be deployed. However, high speed optical fiber transmission systems face limitations imposed by physical properties of the transmission fiber. There are four principal impairments in optical fiber transmission: chromatic dispersion, nonlinearity, polarization effects, and amplified spontaneous emission noise. The polarization effects are polarization mode dispersion (PMD), polarization-dependent loss (PDL), and polarization-dependent gain (PDG). PMD is due to the randomly varying birefringence in optical fibers, PDL is caused by optical components such as directional couplers and isolators, while PDG is caused by polarization hole burning in optical amplifiers [1]. These effects can combine to produce signal impairments in single-channel as well as in wavelength-division-multiplexed systems [2]. Therefore, in long-haul transmission systems it is necessary to limit the penalty induced by po-

larization effects, requiring that the probability of exceeding a maximum specified penalty, such as 1 dB, be very small, typically 10^{-5} or less [3]. This probability is referred to as the outage probability. PMD-induced penalty is defined as the ratio between the back-to-back and the PMD-distorted eye opening. The back-to-back eye opening is computed when PMD is not included in the system. We defined the eye opening as the difference between the lowest mark and the highest space at the decision time in the received electrical noise-free signal. Because of this stringent requirement to penalty induced by polarization effects, it has been very difficult to use either standard, unbiased Monte Carlo simulations or laboratory experiments to determine the outage probability of a system. Only outage probabilities larger than 10^{-4} can be computed in practice by use of standard, unbiased Monte Carlo simulations, since a large number of samples must be obtained in order to accurately estimate the rare events that lead to penalties of interest to system designers. Rare events are almost always defined on the tails of probability density functions. They have small probability and occur infrequently in real applications or in simulations [4]. To overcome this hurdle, advanced Monte Carlo methods such as importance sampling and the multicanonical Monte Carlo (MMC) method have recently been applied to compute PMD-induced penalties using a much smaller number of samples. This dissertation investigates advanced Monte Carlo techniques that compute penalties induced by PMD in uncompensated and compensated systems. A major focus of this dissertation is the estimation of errors in the probability density function (pdf) of the penalties.

The use of Monte Carlo calculations pre-dates the electronic computer [5]; how-

ever, the name "Monte Carlo" is relatively recent. It was coined by Nicolas Metropolis in 1949 [6], but, under the older name of statistical sampling, the method has a history stretching back well into the last century, when numerical calculations were performed by hand [5]. As first envisaged, Monte Carlo was a method for estimating integrals that could not be solved otherwise — integrals in high-dimensional spaces [7]. The Monte Carlo method is a computer-based statistical sampling approach for solving numerical problems that arise in complex systems. The methodology was then further developed in the field of statistical physics during the early days of electronic computing (1945-55) [8]. The idea for constructing Markov-chain-based Monte Carlo algorithms was introduced in the 1950s [8]. This idea was later extended to handle increasingly complex physical systems [9]. In the 1980s, statisticians and computer scientists developed Monte-Carlo-based algorithms for a wide variety of integration and optimization tasks [10]. In the 1990s, the method began to play an important role in computational biology [11]. Over the past fifty years, researchers in diverse scientific fields have studied the Monte Carlo method and contributed to its development [12]. Today, a large number of scientists and engineers employ Monte Carlo techniques as an essential tool in their work.

The most common use of Monte Carlo simulations in optical fiber communications is when we need to compute the distribution of a random variable y, where $y = f(x)$. Here, x represents a state of the system, for example a fiber realization. In this case $f : S \to \mathbf{R}$ is a deterministic function from the state space S of the system to the real numbers, such as the mapping from fiber realizations to eye-opening penalties. Often, there is no analytical formula for f, and so we cannot easily obtain an analytical

formula for the moments or the pdf of y. In fact, we not always need to compute the distribution of y, but only its moments.

The principle behind Monte Carlo simulation is that the behavior of a statistical quantity, such as a moment of y, can be assessed by the empirical process of drawing many random samples x from the state space S, computing the values of $y = f(x)$ and then computing the statistics of y. The basic Monte Carlo procedure is as follows [13]:

1. Use pseudo-random numbers produced by a random number generation algorithm to generate the random samples x.

2. Calculate $y = f(x)$.

3. Calculate the estimator $\hat{\theta}$ of the desired statistical quantity of the random variable y.

4. Repeat steps 2 and 3 N times, where N is the number of trials.

5. Construct a relative frequency distribution (histogram) of the resulting $\hat{\theta}_N$ values, which is the Monte Carlo estimate of the desired statistical quantity.

In a Monte Carlo simulation, the generated random samples depend on the sequence of pseudo-random numbers that are generated during the simulation. Numbers obtained by a formula that imitate the values of a random variable uniformly distributed between 0 and 1 are called pseudo-random numbers [14]. With a second, different sequence of pseudo-random numbers the simulation will not give identical results but will yield different values for $\hat{\theta}$ which agree with those obtained from the

first sequence to within some statistical error [15]. In the field of optical fiber communications, standard Monte Carlo simulations have been applied to study outage probability due to PMD [16] among other applications. Although this approach has proved useful, the error in standard Monte Carlo simulations is given approximately by $N_I^{-1/2}$, where N_I is the number of hits in a given bin of the histogram [7]. Therefore, this technique is not efficient in the estimation of the probability of rare events, such as outage probabilities on the order of 10^{-5}, since it would be necessary to use at least 10^6 standard Monte Carlo samples.

In this Ph.D. dissertation, I describe the work in which my colleagues and I used the advanced Monte Carlo techniques of importance sampling [17] and multicanonical Monte Carlo [18] to accurately and efficiently estimate penalties induced by PMD. These techniques allow low probability events to be efficiently computed by concentrating samples in the most significant regions of the sample space. For example, in the application of Monte Carlo techniques to compute the pdf of differential group delay (DGD) for a PMD emulator, the most significant regions are those where the rare events or large DGD values are located. Using these methods, we evaluated the performance of PMD compensators and compared the efficiency of the two methods to compute penalties in different types of compensated systems [19], [20]–[25]. We also used importance sampling to compare the penalty resulting from first-order and all-order PMD models, demonstrating that simple theoretical models that take into account only first-order PMD overestimate the penalty when the DGD is large compared to the mean DGD [26], [27].

Importance sampling [17] is a well-known statistical technique that uses *a priori*

knowledge of which regions in the sample space are most significant to concentrate samples in those regions. This technique allows one to compute almost arbitrarily small probabilities at a fraction of the computational cost of standard, unbiased Monte Carlo simulations. Monte Carlo simulations with importance sampling have been used to bias first-order PMD [28] and to bias both first- and second-order PMD [29] to compute the probability of large PMD-induced penalties, which are rare. The results obtained from this method rely on the correlation between the penalty and the first- and second-order PMD. When both first- and second-order PMD are biased, the entire first- and second-order PMD plane is statistically resolved, even though only a portion of this plane determines the penalty, as we showed in [23].

The multicanonical Monte Carlo method was proposed by Berg and Neuhaus [18] and was recently applied to compute the pdf of the DGD for a PMD emulator [30]. In the work presented in this dissertation, we applied the MMC simulation technique to directly compute the probability of large penalties resulting from all orders of PMD in uncompensated and compensated systems. In contrast to importance sampling and most other biasing Monte Carlo methods, MMC does not require prior knowledge of which rare events contribute significantly to the large penalty values in the tail of the pdf. MMC is an iterative method which in each iteration produces a biased random walk that automatically searches the state space for important rare events. MMC also has the advantage that it does not require one to combine different biased pdfs of the quantity of interest in order to obtain the entire pdf, as is the case with importance sampling.

As mentioned earlier in this chapter, Monte Carlo methods rely on statistics since

they use random numbers, so that statistical error estimates are essential to assess the accuracy of the results. The goal of any scheme for biasing Monte Carlo simulations is to reduce the variance of the desired statistical quantity of the random variable y [10], [17]. MMC uses a set of systematic procedures to reduce the variance, which are highly nonlinear as well as iterative and have the effect of inducing a complex web of correlations from sample to sample in each iteration and between iterations. These in turn induce bin-to-bin correlations in the histograms of the pdfs of the quantities of interest. Thus, calculating the error is significantly more difficult than in standard Monte Carlo simulations. Due to the correlations, one cannot apply to MMC the standard error analysis [31] that is traditionally used for simulations with uncorrelated samples. For the same reason, one cannot determine the contribution of the variance from each iteration using standard error propagation methods as is possible with importance sampling [24]. In this Ph.D. dissertation, I report the contribution that we made to estimate the statistical errors when using importance sampling and multicanonical Monte Carlo methods. We applied standard error estimation techniques and first-order error propagation to estimate errors in importance sampling simulations. For MMC, we developed an efficient numerical method to estimate statistical errors, which we refer to as the transition matrix method. We showed that this method is a variant of the well-known bootstrap method [32].

The remainder of this Ph.D. dissertation is organized as follows: In Chapter 2, I describe the physical characteristics of PMD and its associated statistical quantities, and I also describe how we model PMD. In Chapter 3, I describe how to implement the technique that uses Monte Carlo simulations with importance sampling to

compute the probability density function of the eye-opening penalty and the outage probability due to PMD for a single-channel optical fiber transmission system. Using importance sampling, we were able to compare penalties from a first-order PMD model with the penalties from an all-order PMD model. I also show how to analyze the performance of PMD single-section compensators using importance sampling. In Chapter 4, I describe how to implement the multicanonical Monte Carlo method to compute penalties from polarization mode dispersion in optical fiber communication systems. In Chapter 5, I present the transition matrix method that we developed to estimate errors in MMC simulations. In Chapter 6, I describe a comparative study of the two techniques of MMC and importance sampling that were introduced in Chapters 3 and 4 to compute penalties due to PMD in systems with multi-section PMD compensators. Finally, in Chapter 7, I present the conclusions of this Ph.D. dissertation.

Chapter 2

Polarization Mode Dispersion (PMD)

In this chapter, I give a physical and a statistical description of PMD. Since PMD varies with temperature, stress, and other conditions, a statistical method is required to estimate outage probabilities. In addition, I show how my colleagues and I modeled PMD in the systems that we studied.

2.1 Physical description of PMD

Polarization mode dispersion in optical fibers is due to deviations from circular symmetry in the core and cladding. These deviations give rise to two distinct polarization modes with distinct phase and group velocities. Although telecommunications fibers are often called "single-mode," there are two polarized HE_{11} modes [33] even in an ideal, circularly-symmetric fiber. Therefore, light propagation in single-mode fibers is governed by two orthogonally polarized fundamental modes (or eigenmodes), which are degenerate in the case of ideal fibers. The asymmetry in actual optical fibers breaks the degeneracy of the HE_{11} modes, resulting in birefringence, which is defined as the difference in the phase and group velocities of the two modes [33]. In opti-

cal fibers, the birefringence is nearly linear, and one may refer to the two orthogonal eigenmodes of polarization as the x- and y-polarizations. The strength of birefringence B may then be defined as

$$B = \frac{|\beta_x - \beta_y|}{k_0} = |n_x - n_y|\,, \tag{2.1}$$

where n_x and n_y are the modal refractive indices for the two orthogonally polarized eigenmodes, β_x and β_y are the corresponding wavenumbers, and $k_0 = 2\pi/\lambda$, where λ is the wavelength of light. For a given value of B, the two modes exchange their powers in a periodic fashion as they propagate inside a fiber section with constant birefringence. This period is called the beat length L_B [34],

$$L_B = \frac{2\pi}{|\beta_x - \beta_y|} = \frac{\lambda}{B}. \tag{2.2}$$

The beat length is the propagation distance over which a 2π phase difference accumulates between the two modes, or equivalently the polarization states vary through a full cycle.

The axis along which the modal refractive index is smaller is called the fast axis because the phase velocity is larger for light propagating in that direction. For the same reason, the axis with the larger modal refractive index is called the slow axis. In standard optical fibers, the birefringence changes randomly along the fiber due to vibrations and variations of the temperature on a time scale that varies from milliseconds to hours [35]. The fiber can be modeled as a sequence of birefringence sections [35], where the amount of birefringence remain constant in each section but the orientation of the principal axes changes randomly from section to section along

the fiber length. As a result, light launched into the fiber with a fixed state of polarization changes its polarization in a random fashion along the fiber.

Considering a single section of constant birefringence, if an input pulse excites both polarization eigenmodes of the fiber section, the two polarization eigenmodes drift apart along the fiber section because they travel at different group velocities. The pulse becomes broader at the output because the two eigenmodes undergo different delays in the fiber section. In the frequency domain, the state of polarization at the output of the fiber section varies with frequency for a fixed input polarization. When displayed on the Poincaré sphere, the polarization at the output moves on a circle on the surface of the sphere as the optical frequency is varied. The evolution of the state of polarization along the fiber section is the result of the increase in the frequency-dependent phase difference between orthogonal components of the electrical field in the two eigenmodes of polarization. In the presence of extensive mode coupling, *i.e.*, considering many fiber sections, light in different frequencies will couple differently into the fast and the slow axes of birefringence along the fiber sections, which will undergo different propagation delays. In this case, the states of polarization evolve randomly over the surface of the Poincaré sphere as the light propagates along the fiber until the surface of the sphere is uniformly covered. This phenomenon is referred as polarization mode dispersion.

Polarization mode dispersion is characterized in Stokes space by the polarization dispersion vector $\boldsymbol{\tau}(\omega)$. The polarization dispersion vector is defined as [34]

$$\frac{\mathrm{d}\mathbf{s}}{\mathrm{d}\omega} = \boldsymbol{\tau} \times \mathbf{s}, \tag{2.3}$$

where $\mathbf{s}$ is a three-dimensional unit Stokes vector representing the output polarization state of light. The direction of the polarization dispersion vector determines the direction of the two orthogonal principal states of polarizations (PSP), $\pm\boldsymbol{\tau}/|\boldsymbol{\tau}|$. The differential group delay (DGD) between the PSPs is equal to the length of the polarization dispersion vector, $|\boldsymbol{\tau}| = \tau$. When the polarization dispersion vector varies slowly over the bandwidth of the optical signal, one may neglect the higher order terms of the Taylor expansion of $\boldsymbol{\tau}(\omega)$ at the central frequency of the channel ω_c, so that $\boldsymbol{\tau}(\omega) \approx \boldsymbol{\tau}(\omega_c)$. In this regime, the penalty is said to be dominated by distortion due to first-order PMD. In this approximation, the signal splits into two PSPs, which propagate at different group velocities, causing intersymbol interference. Light launched in a single PSP does not experience first-order PMD distortion. Hence, in the frequency domain, a PSP is defined as that input polarization for which the output state of polarization is independent of frequency to first order, *i.e.*, over a small frequency range.

In the first-order approximation, the PSPs and the DGD do not vary with frequency. The correlation bandwidth [34] $\Delta\nu_{\mathrm{PMD}}$ over which the DGD can be considered constant is inversely proportional to the expected value of the DGD with a constant of proportionality close to 1/2. For pulses with a small bandwidth compared to $\Delta\nu_{\mathrm{PMD}}$, which means that the expected DGD is small compared to the duration of the optical pulse, the first-order model is practically useful. In most sections of this dissertation, however, we use a PMD model that takes into account the frequency variation of the DGD and of the PSPs. The random variations with wavelength of both the DGD and PSPs are the origin of the higher orders of PMD [36]. In

Section 2.3 we describe the PMD simulation model that we use [37].

The PMD coefficient D_{PMD} of an optical fiber is related to the average DGD $\langle\tau\rangle$ by [34]

$$D_{\mathrm{PMD}} = \sqrt{\frac{3\pi}{8L}}\,\langle\tau\rangle, \tag{2.4}$$

where L is the fiber length. Typical values for D_{PMD} are in the range from 0.1 to 1.0 ps/km$^{1/2}$ [38]. Because of its $L^{1/2}$ dependence, PMD-induced pulse broadening is relatively small compared to broadening due to chromatic dispersion. However, chromatic dispersion is a deterministic effect that can be compensated in a deterministic way, while PMD is a random effect. Since $\langle\tau\rangle \approx D_{\mathrm{PMD}} \cdot L^{1/2}$, PMD can become a limiting factor for optical fiber communication systems designed to operate over long distances. In this dissertation, we focus on systems whose PMD is large enough to produce significant waveform distortion. In these systems, the mean of the accumulated DGD of the transmission line exceeds 10% of the bit period. These systems include 10 Gbit/s long-haul terrestrial communications systems that use high PMD (> 1 ps/km$^{1/2}$) optical fibers that were deployed before 1990 [39]. For systems in which the mean DGD exceeds 10% of the bit period, there is a need for PMD compensation that takes into account the effects of second and higher orders of PMD impairments. In the following chapters of this dissertation, I address the issue of PMD compensation for first and higher orders of PMD.

2.2 Polarization mode dispersion statistics

Polarization mode dispersion is a stochastic phenomenon that changes with wavelength and time. Therefore, one cannot predict the impairment of the system due to PMD at any particular wavelength and time, but must instead rely upon a statistical description. Since the probability density functions for PMD in most cases have asymptotic tails that extend to unacceptable large impairments, the system cannot be designed to handle the worst case PMD impairment and must instead be designed for a specific outage probability [40]. For similar reasons, the goal of PMD compensation cannot be to eliminate the impairment, but rather to reduce the PMD outage probability. To understand and predict system outage probabilities, to design PMD compensators, and to accurate model PMD in systems, one must understand the statistics of the phenomena associated with PMD. In this section, the statistical characteristics of PMD that we investigate include the probability densities of first-order PMD, $\boldsymbol{\tau}$, second-order PMD, $\boldsymbol{\tau}_\omega$, and the probability densities of their magnitudes. Probability densities can be obtained from an analytical model developed by Foschini and Poole [36], [41]. This model, like most models used to obtain analytical results for PMD statistics, assumes the ideal, perfectly random, fiber birefringence. It is important to mention that real systems are not perfectly random and real densities presumably do not exhibit the asymptotic tails extending to infinity that are found in densities obtained analytically [40].

The derivation of the PMD probabilities density functions (pdfs) is based on the known characteristic function of the six-dimensional (6-D) vector $(\boldsymbol{\tau}, \boldsymbol{\tau}_\omega)$, which is

defined to be the 6-D Fourier transform of the probability density function $p_{\boldsymbol{\tau}',\boldsymbol{\tau}'_\omega}$ [36]

$$F\left\{p_{\boldsymbol{\tau}',\boldsymbol{\tau}'_\omega}\right\}(\mathbf{z},\boldsymbol{\Psi}) = \int_{\Re^6} \mathbf{e}^{i(\mathbf{z}\cdot\boldsymbol{\tau}+\boldsymbol{\Psi}\cdot\boldsymbol{\tau}_\omega)} p_{\boldsymbol{\tau}',\boldsymbol{\tau}'_\omega}(\mathbf{u},\mathbf{v})\, du_x\, du_y\, du_z\, dv_x\, dv_y\, dv_z. \tag{2.5}$$

The normalized vector, $(\boldsymbol{\tau}', \boldsymbol{\tau}'_\omega)$, is obtained from the 6-D vector $(\boldsymbol{\tau}, \boldsymbol{\tau}_\omega)$ according to

$$(\boldsymbol{\tau}', \boldsymbol{\tau}'_\omega) = \left(\sqrt{\frac{8}{\pi}}\frac{\boldsymbol{\tau}}{\langle|\boldsymbol{\tau}|\rangle},\ \frac{8}{\pi}\frac{\boldsymbol{\tau}_\omega}{\langle|\boldsymbol{\tau}|\rangle^2}\right). \tag{2.6}$$

We may write $F\left\{p_{\boldsymbol{\tau}',\boldsymbol{\tau}'_\omega}\right\}$ explicitly as

$$F\left\{p_{\boldsymbol{\tau}',\boldsymbol{\tau}'_\omega}\right\} = \operatorname{sech}|\boldsymbol{\Psi}| \exp\left\{-\frac{1}{2}\left[\frac{|\mathbf{z}|^2 \tanh|\boldsymbol{\Psi}|}{\boldsymbol{\Psi}} + \frac{(\mathbf{z}\cdot\boldsymbol{\Psi})^2}{|\boldsymbol{\Psi}|^2}\left(1 - \frac{\tanh|\boldsymbol{\Psi}|}{|\boldsymbol{\Psi}|}\right)\right]\right\}, \tag{2.7}$$

where F indicates the 6-D Fourier transform of a function of $(\boldsymbol{\tau}', \boldsymbol{\tau}'_\omega)$ into a function of its conjugate vector $(\mathbf{z}, \boldsymbol{\Psi})$, and the vectors $\mathbf{z}$ and $\boldsymbol{\Psi}$ represent 3-D transform variables corresponding to the components of $\boldsymbol{\tau}'$ and $\boldsymbol{\tau}'_\omega$, respectively. The normalization allows the results obtained from (2.7) to be independent of the mean DGD. For simplicity, in the pdfs that we present in this section, we use $\langle\tau\rangle$ instead of $\langle|\boldsymbol{\tau}|\rangle$ to represent the mean DGD.

We can use (2.7) to obtain a variety of probability density functions and moments [36], [40], [41]. For instance, the characteristic function of a single variable can be obtained from the six-dimensional function by setting the uninteresting conjugate variables to zero. In this way, the Gaussian characteristic function for a first-order component, τ_x for example, is found from (2.7) by setting z_y, z_z, and $\boldsymbol{\Psi}$ to zero. An inverse Fourier transform can be used to provide the desired density, which is also Gaussian [36]. By setting $\boldsymbol{\Psi}$ to zero and performing an inverse Fourier transform with

respect to $\mathbf{z}$, we obtain the 3-D Gaussian density function of $\boldsymbol{\tau}$. Letting $\tau_i = \tau_x$, τ_y, or τ_z, we find

$$p_{\tau_i} = \frac{2}{\pi \langle\tau\rangle} \exp[-\left(2\tau_i / \langle\tau\rangle\right)^2 / \pi], \tag{2.8}$$

and the pdf of the DGD $|\boldsymbol{\tau}| = \tau$, which is Maxwellian is

$$p_\tau = \frac{8}{\pi^2 \langle\tau\rangle} \left(\frac{2\tau}{\langle\tau\rangle}\right)^2 \exp[-\left(2\tau / \langle\tau\rangle\right)^2 / \pi], \tag{2.9}$$

where $\tau \geq 0$.

Setting $\mathbf{z} = 0$ in (2.7), one obtains the pdf of $\boldsymbol{\tau}_\omega$. In addition, setting $\Psi_{j\neq i} = 0$, followed by a 1-D inverse Fourier transform, we find that the probability density of a component, $\tau_{\omega,i}$, is given by

$$p_{\tau_{\omega,i}} = \frac{4}{\pi \langle\tau\rangle^2} \text{sech}\left(\frac{4\tau_{\omega,i}}{\langle\tau\rangle^2}\right). \tag{2.10}$$

Conversion to polar coordinates after one obtains the pdf of $\boldsymbol{\tau}_\omega$ shows that the probability density of the magnitude $|\boldsymbol{\tau}_\omega|$ is given by [40], [42]

$$p_{|\boldsymbol{\tau}_\omega|} = \frac{32\,|\boldsymbol{\tau}_\omega|}{\pi \langle\tau\rangle^4} \tanh\left(\frac{4\,|\boldsymbol{\tau}_\omega|}{\langle\tau\rangle^2}\right) \text{sech}\left(\frac{4\,|\boldsymbol{\tau}_\omega|}{\langle\tau\rangle^2}\right). \tag{2.11}$$

While these analytical expressions are useful, it is not possible to obtain exact analytical expressions for the densities after compensation or for penalties either before or after compensation. Simulations must be used to obtain these other densities. The tails of both the analytical and simulated distributions are the regions of greatest importance because it is the large values of PMD that lead to system outages. Assuming that ρ is a random variable representing the penalty or any other quantity for which we described the probability density function above, the outage probability

with respect to ρ is the probability of exceeding some value of ρ, which is typically the threshold for unacceptable impairment. The estimation of system outage probabilities requires the computation of the complement of the cumulative density function (cdfc), where $\mathrm{cdfc}(\rho) = \int_{\rho}^{\infty} p(\rho')\,\mathrm{d}\rho'$, and $p(\rho)$ is the corresponding pdf.

It is very difficult to obtain the tails of the distributions using standard, unbiased Monte Carlo simulations. The advanced Monte Carlo techniques of importance sampling and multicanonical Monte Carlo can be used to efficiently estimate rare penalties induced by PMD. The techniques of importance sampling and multicanonical Monte Carlo and their applications to PMD are presented in the following chapters of this dissertation.

2.3 Modeling of polarization mode dispersion

Perturbations that cause loss of circular symmetry in the core of optical fibers lead to birefringence and hence to PMD. Here, I describe the PMD simulation model that my colleagues and I used [37]. This model is based on the representation of the fiber link by a frequency-dependent transfer function, expressed as a 2×2 matrix, $\bar{\mathsf{T}}(\omega)$, referred to as the Jones matrix. This transfer function takes into account all orders of PMD. Since we focused our investigation on the penalty caused by PMD, we did not take into account polarization-dependent loss, chromatic dispersion, or fiber nonlinearity in this model. In this case, the Jones matrix is unitary, and the complex envelope of the electrical field vector at the end of the fiber link, $\mathbf{E}_{\mathrm{out}}(\omega)$, may be written as

$$\mathbf{E}_{\mathrm{out}}(\omega) = \bar{\mathsf{T}}(\omega)\,\mathbf{E}_{\mathrm{in}}(\omega), \tag{2.12}$$

where $\mathbf{E}_{\text{in}}(\omega)$ is the input field vector in the Jones space [43]. Using the coarse step model of a fiber [37] in which we assume that the fiber passes ergodically through all possible orientations of the birefringence and that each section is long compared to the fiber correlation length [37], the Jones transfer matrix $\bar{\mathsf{T}}(\omega)$ of an optical fiber that consists of N linearly birefringent sections may be written as [37]

$$\bar{\mathsf{T}}(\omega) = \prod_{n=1}^{N} \bar{\mathsf{T}}_n(\omega), \tag{2.13}$$

where

$$\bar{\mathsf{T}}_n(\omega) = \bar{\mathsf{P}}(\omega)\,\bar{\mathsf{S}}_n \tag{2.14}$$

is the transmission matrix of the n-th fiber section. The matrix

$$\bar{\mathsf{S}}_n = \begin{pmatrix} \cos(\xi_n/2)\exp[i(\psi_n+\phi_n)/2] & i\sin(\xi_n/2)\exp[i(\psi_n-\phi_n)/2] \\ i\sin(\xi_n/2)\exp[-i(\psi_n-\phi_n)/2] & \cos(\xi_n/2)\exp[-i(\psi_n+\phi_n)/2] \end{pmatrix} \tag{2.15}$$

is a Jones matrix that corresponds to a uniformly distributed rotation on the Poincaré sphere, while

$$\bar{\mathsf{P}}(\omega) = \begin{pmatrix} \exp(-i\omega\tau_s/2) & 0 \\ 0 & \exp(i\omega\tau_s/2) \end{pmatrix}, \tag{2.16}$$

models the frequency-dependent phase rotation of the light through a birefringent section. The parameter τ_s is the DGD in a single fiber section. The angles ξ_n, ψ_n, and ϕ_n will be independent at each n and from each other. The angles ψ_n and ϕ_n are uniformly distributed between 0 and 2π, while the random variables $\cos\xi_n$ are uniformly distributed between -1 and 1. The Müller matrix [43] S_n that is equivalent to the Jones matrix $\bar{\mathsf{S}}_n$ in (2.15) is comprised of elementary rotations around two of the three orthogonal axes [43] of the Poincaré sphere,

$$\mathsf{S}_n = \mathsf{R}_x(\psi_n)\,\mathsf{R}_y(\xi_n)\,\mathsf{R}_x(\phi_n), \tag{2.17}$$

which produces a uniform rotation on the Poincaré sphere [37] provided that the angles ψ_n, ξ_n, ϕ_n are distributed as described above. In (2.17), $\mathsf{R}_x(\psi_n)$ is a rotation around the x-axis, and $\mathsf{R}_y(\xi_n)$ is a rotation around the y-axis. Likewise, the Müller matrix $\mathsf{P}(\omega)$ equivalent to the Jones matrix $\bar{\mathsf{P}}(\omega)$ in (2.16) is comprised of an elementary rotation around the x-axis,

$$\mathsf{P}(\omega) \;=\; \mathsf{R}_x(-\omega\tau_s)\,. \tag{2.18}$$

Since the Müller matrix of a section is equal to $\mathsf{P}(\omega)\mathsf{S}_n$, the polarization dispersion vector of a single section is given by [44]

$$\boldsymbol{\tau}_s \;= -\tau_s\,\hat{\mathbf{a}}_x, \tag{2.19}$$

where $\hat{\mathbf{a}}_x$ is a unit vector along the x-axis.

The formulation of (2.14) is consistent with the one in [44], where the random mode coupling in the n-th section occurs prior to the birefringent element of that section. We set τ_s equal to

$$\tau_s = \sqrt{\frac{3\pi}{8N}}\,\langle\tau\rangle, \tag{2.20}$$

where $\langle\tau\rangle = \langle\tau^{(N)}\rangle$ is the mean DGD of the fiber with N sections [45]. Throughout this dissertation, we emulate an optical fiber with eighty birefringent sections $N = 80$ unless otherwise stated. In [3], Lima *et al.* showed that $N = 80$ is sufficient to obtain a Maxwellian distribution of the DGD in the outage probability range up to 10^{-6}.

The pdf of the DGD τ, for a PMD emulator with polarization controllers, which implements the coarse step method [37], and N concatenated birefringent sections

each with a DGD of τ_s, is given by [46]

$$p\left(\tau, N\right) \;=\; \frac{\tau}{2\,{\tau_s}^2\,(N-2)!}\sum_{j=0}^{k}(-1)^j \left(\begin{array}{c} N \\ j \end{array}\right)(Nm-j)^{(N-2)}\,, \tag{2.21}$$

where $m = 0.5\,(1-\tau/N\tau_s)$ and k is the integer part of Nm. The expression given in (2.21) is valid only for $N > 1$. In the case $N = 1$ the pdf is a delta function $p(\tau, 1) = \delta(\tau - \tau_s)$. The pdf $p\,(\tau, N)$ is identically zero outside the interval $0 < \tau < N\tau_s$. This interval is reasonable since N concatenated vectors of length τ_s cannot be longer than $N\tau_s$. In the limit $N \to \infty$, the pdf is the Maxwellian pdf [36], [46], where the mean DGD is equal to $\tau_s\sqrt{8N/3\pi}$.

Chapter 3

Importance sampling applied to PMD

In this chapter, I discuss the application of the biasing Monte Carlo techniques of importance sampling to compute outage probabilities in optical fiber transmission systems. I show how to implement importance sampling in which one biases just the DGD and importance sampling in which one biases both the first- and second-order PMD. Then, I show how my colleagues and I estimate the mean and the variance of the penalty given a value of first- and second-order PMD. Finally, I apply importance sampling to compare penalties resulting from first-order and all-order PMD and to analyze the performance of single-section PMD compensators. I also describe the single-section PMD compensator that we investigated.

3.1 Importance sampling to compute outage probability

As discussed in Chapter 1, optical fiber transmission systems are vulnerable to problems arising from a variety of fiber polarization effects, including PMD. PMD has become a serious barrier to the deployment of high bit rate (10 Gbit/s and higher)

transmission systems. One of the goals in designing these systems is to minimize the probability of channel outage due to polarization effects. The penalty due to PMD is defined as the ratio between the eye opening when PMD is not included in the system and the PMD-distorted eye opening. The eye opening is defined as the difference between the lowest mark and the highest space. System designers commonly allocate a prescribed penalty margin to polarization effects, such as 1 dB, with a fixed probability that the margin will be exceeded, such as 10^{-5}, corresponding to approximately five minutes per year. When this margin is exceeded, an outage is said to occur. Thus, we define outage probability as the probability of the penalty exceeding this allowed value. When using standard, unbiased Monte Carlo simulations, a very large number of samples must be explored in order to obtain an accurate estimate of the rare events that lead to unacceptable penalties, such as large DGD values. In this chapter, I focus on the use of a biased Monte Carlo technique called importance sampling to accurately estimate the penalty induced by polarization mode dispersion. This technique allows low probability events to be efficiently computed using a far smaller number of samples than would be required with standard, unbiased Monte Carlo simulations.

The goal of any biased Monte Carlo simulation, including simulations based on importance sampling, is to cause the events that contribute to the statistical quantities of interest to occur more frequently and thus to reduce the relative variation in the numerical estimate of those quantities with a fixed number of samples [17]. In this dissertation, the quantities of interest are the eye-opening penalties and their probability density functions both before and after compensation.

In Section 3.2, I explain how we implement the single-section compensator and how we determine the eye-opening penalty. In Section 3.3, I describe the implementation of importance sampling to the length of the polarization dispersion vector, the DGD. I show that biasing the DGD is sufficient to accurately calculate the uncompensated penalties and their pdfs, but it is not sufficient to accurately calculate the compensated penalties and their pdfs. To study single-section compensated systems where second-order PMD also plays a role it is necessary to use an importance sampling method capable of biasing both the DGD and the length of the frequency derivative of the polarization dispersion vector. In Section 3.4, I show how the probability of rare events with large first- and second-order PMD can be efficiently computed using Monte Carlo simulation with importance sampling applied to both the DGD and length of the second-order PMD. In Section 3.5, I show how to compute the relative variation in the results that one obtains using Monte Carlo simulations with importance sampling. In Sections 3.6 and 3.7, I show the use of the technique of importance sampling applied to PMD. In Section 3.6, I show results when we compare the eye-opening penalty from a first-order PMD model with the penalty from an all-order PMD model. Finally, in Section 3.7, I show the simulation results of PMD-induced penalty in the plane of first- and second-order PMD and the outage probability for uncompensated and single-section compensated systems.

3.2 Single-section PMD compensator and definition of penalty

In this section, I explain how my colleagues and I implemented the single-section compensator and the receiver model that is used throughout this dissertation. The increased understanding of PMD and its system impairments, together with a quest for higher transmission bandwidths, has motivated considerable effort to mitigate the effects of PMD, based on different compensation schemes [3], [16], [20]. One of the primary objectives has been to enable system upgrades from 2.5 Gbit/s to 10 Gbit/s or from 10 Gbit/s to 40 Gbit/s on old, embedded, high-PMD fibers. PMD compensation techniques must reduce the impact of first-order PMD and should reduce higher-order PMD effects or at least not increase the higher orders of PMD. The techniques should also be able to rapidly track changes in PMD, including changes both in the DGD and the PSPs. Other desired characteristics of PMD mitigation techniques are low cost and small size to minimize the impact on existing system architectures. In addition, mitigation techniques should have a small number of feedback parameters to control [40].

In this chapter, I will describe a PMD compensator with an arbitrarily rotatable polarization controller and a single DGD element, which can be fixed [47] or variable [48]. The adjustable DGD element or birefringent element is used to minimize the impact of the fiber PMD and the polarization controller is used to adjust the direction of the polarization dispersion vector of the compensator. The expression for the polarization dispersion vector after compensation, which is equivalent to the one

in [49], is given by

$$\boldsymbol{\tau}_{\mathrm{tot}}(\omega) = \boldsymbol{\tau}_c + \mathsf{T}_c(\omega)\mathsf{R}_{\mathrm{pc}}\boldsymbol{\tau}_f(\omega), \tag{3.1}$$

where $\boldsymbol{\tau}_c$ is the polarization dispersion vector of the compensator, $\boldsymbol{\tau}_f(\omega)$ is the polarization dispersion vector of the transmission fiber, R_{pc} is the polarization transformation in Stokes space that is produced by the polarization controller of the compensator, and $\mathsf{T}_c(\omega)$ is the polarization transformation in Stokes space that is produced by the DGD element of the compensator, which is similar to (2.18). We model the polarization transformation R_{pc} as

$$\mathsf{R}_{\mathrm{pc}} = \mathsf{R}_x(\phi_{\mathrm{pc}})\mathsf{R}_y(\psi_{\mathrm{pc}})\mathsf{R}_x(-\phi_{\mathrm{pc}}). \tag{3.2}$$

I note that the two parameters of the polarization controller's angles in (3.2) are the only free parameters that a compensator with a fixed DGD element possesses, while the value of the DGD element of a variable DGD compensator is an extra free parameter that must be adjusted during the operation. In (3.2), the parameter ϕ_{pc} is the angle that determines the axis of polarization rotation in the y-z plane of the Poincaré sphere, while the parameter ψ_{pc} is the angle of rotation around that axis of polarization rotation. An appropriate selection of these two angles will transform an arbitrary input Stokes vector into a given output Stokes vector. While most electronic polarization controllers have two or more parameters to adjust that are different from ϕ_{pc} and ψ_{pc}, it is possible to configure them to operate in accordance to the transformation matrix R_{pc} in (3.2) [50].

In all the work reported in this chapter, we used the eye opening as the feedback parameter for the optimization algorithm unless otherwise stated. We defined the

eye opening as the difference between the lowest mark and the highest space at the decision time in the received electrical noise-free signal. The eye-opening penalty is defined as the ratio between the back-to-back and the PMD-distorted eye opening. The back-to-back eye opening is computed when PMD is not included in the system. Since PMD causes pulse spreading in amplitude-shift keyed modulation formats, the isolated marks and spaces are the ones that suffer the highest penalty [51]. To define the decision time, we recovered the clock using an algorithm based on one described by Trischitta and Varma [52]. We simulated the 16-bit string "0100100101101101." This bit string has isolated marks and spaces, in addition to other combinations of marks and spaces. In most of other simulations in this dissertation we use pseudorandom binary sequence pattern. The receiver model consists of an Gaussian optical filter with full width at half maximum (FWHM) of 60 GHz, a square-law photodetector, and a fifth-order electrical Bessel filter with a 3 dB bandwidth of 8.6 GHz. To determine the decision time after the electronic receiver, we delayed the bit stream by half a bit slot and subtracted it from the original stream, which is then squared. As a result a strong tone is produced at 10 GHz. The decision time is set equal to the time at which the phase of this tone is equal to $\pi/2$.

The goal of our study is to determine the performance limit of the compensators. In order to do that, we search for the angles ϕ_{pc} and ψ_{pc} of the polarization controller for which the eye opening is largest. In this case, the eye opening is our compensated feedback parameter. We therefore show the global optimum of the compensated feedback parameter for each fiber realization. To obtain the optimum, we start with 5 evenly spaced initial values for each of the angles ϕ_{pc} and ψ_{pc} in the polarization

transformation matrix R_{pc}, which results in 25 different initial values. If the DGD of the compensator is adjustable, we start the optimization with the DGD of the compensator equal to the DGD of the fiber. We then apply the conjugate gradient algorithm [53] to each of these 25 initial polarization transformations. To ensure that this procedure yields the global optimum, we studied the convergence as the number of initial polarization transformations is increased. We examined 10^4 fiber realizations spread throughout our phase space, and we never found more than 12 local optima in the cases that we examined. We missed the global optimum in three of these cases because several optima were closely clustered, but the penalty difference was small. We therefore concluded that 25 initial polarization transformations were sufficient to obtain the global optimum with sufficient accuracy for our purposes. We observed that the use of the eye opening as the objective function for the conjugate gradient algorithm produces multiple optimum values when both the DGD and the length of the frequency derivative of the polarization dispersion vector are very large.

The performance of the compensator depends on how the DGD and the effects of the first- and higher-order frequency derivatives of the polarization dispersion vector of the transmission fiber interact with the DGD element of the compensator to produce a residual polarization dispersion vector and on how the signal couples with the residual principal states of polarization over the spectrum of the channel. Therefore, the operation of single-section PMD compensators is a compromise between reducing the DGD and setting one principal state of polarization after compensation that is approximately co-polarized with the signal. An expression for the pulse spreading due to PMD as a function of the polarization dispersion vector of the transmission

fiber and the polarization state over the spectrum of the signal was given in [54].

3.3 Importance sampling that biases only the DGD

Importance sampling is a well-known technique in statistics that makes efficient use of Monte Carlo simulations to compute the probability of rare events. The rare events that lead to outage due to PMD are highly correlated with the large DGD values. These values are large compared to the average or expected value of the DGD. When we use importance sampling to bias the DGD, we are taking advantage of the large correlation that exists between the PMD-induced penalty and the DGD. We note that first- and higher-order frequency derivatives of the polarization dispersion vector are included in the simulations, although this approach does not produce larger values of first- and higher-order frequency derivatives of the polarization dispersion vector than the moderately large values that are naturally obtained when one biases the DGD.

My colleagues and I observed that the use of a single biased distribution with importance sampling is insufficient to resolve the histogram of the penalty produced by polarization effects in the configuration space. Thus, it is necessary to combine the statistical results from several different biased distributions, since each distribution resolves a different region in the sample space. This approach is referred to as multiple importance sampling.

To apply the multiple importance sampling technique, we first recall that P_I, the probability of an event defined by the indicator function $I(\mathbf{x})$, can be estimated as [55]

$$\hat{P}_I = \sum_{j=1}^{J} \frac{1}{M_j} \sum_{i=1}^{M_j} I(\mathbf{x}_{j,i}) w_j(\mathbf{x}_{j,i}) L_j(\mathbf{x}_{j,i}), \tag{3.3}$$

where

$$L_j(\mathbf{x}_{j,i}) = \frac{p(\mathbf{x}_{j,i})}{p_j^*(\mathbf{x}_{j,i})} \tag{3.4}$$

is the likelihood ratio of the i-th sample $\mathbf{x}_{j,i}$ drawn from the j-th biasing distribution and where M_j is the number of samples drawn from the j-th biasing distribution $p_j^*(\mathbf{x})$. The term $p(\mathbf{x})$ is the pdf of the unbiased distribution, and J is the number of different biasing distributions. The weights $w_j(\mathbf{x})$ allow one to combine different biasing distributions and are defined in (3.5) below. In this application, the random vector $\mathbf{x}$ corresponds to the random realization of the fiber PMD, which is determined by the random mode coupling between the birefringent sections, and the unbiased pdf $p(\mathbf{x})$ is assumed to be known *a priori*. The indicator function $I(\mathbf{x})$ in (3.3) is chosen to compute the probability of having an eye-opening penalty within a given range, such as a bin in a histogram. Thus, the indicator function $I(\mathbf{x})$ is defined as 1 inside the desired penalty range and 0 otherwise.

An efficient technique to combine the samples from multiple biased distributions is the balanced heuristic method [55], which my colleagues and I used for the work reported in this dissertation. The balanced heuristic weight assigned to the sample $\mathbf{x}$ drawn from the j-th distribution is given by

$$w_j(\mathbf{x}) = \frac{M_j L_j^{-1}(\mathbf{x})}{\sum_{k=1}^{J} M_k L_k^{-1}(\mathbf{x})}. \tag{3.5}$$

The idea behind the balanced heuristic method is that samples are weighted according to the likelihood that each biased distribution produces samples in that region; distributions that are more likely to put samples there are weighted more heavily. The computation of the balanced heuristic weight for any given sample requires that

the likelihood ratio of all the biased distributions be evaluated for that sample. In other words, the likelihood ratio of all samples in all J biased distributions have to be evaluated for the i-th sample drawn from the j-th distribution, even though this sample was obtained using only the biased pdf of the j-th distribution.

The polarization dispersion vector after n fiber sections is determined by the same concatenation rule showed in (3.1),

$$\boldsymbol{\tau}^{(n)} = \boldsymbol{\tau}_n + \mathsf{T}_n \boldsymbol{\tau}^{(n-1)}, \tag{3.6}$$

where T_n is the equivalent Müller matrix of the n-th section in (2.14). Biondini, *et al.* [28] demonstrated that the appropriate parameters to bias are the angles θ_n between the polarization dispersion vector of the first $(n-1)$ sections $\boldsymbol{\tau}^{(n-1)}$ and the polarization dispersion vector of the n-th section $\boldsymbol{\tau}_n$ at the center frequency of the channel, such that $\cos\theta_n$ is biased towards one, thereby increasing the probability that the polarization dispersion vector at that frequency will lengthen after that section. In other words, $\boldsymbol{\tau}_n$ is biased towards a direction $\hat{\mathbf{b}}$ that is equal to the direction of $\boldsymbol{\tau}^{(n-1)}$. The angles θ_n are directly determined by the realization of the random mode coupling between the birefringent sections. Thus, the values of $\cos\theta_n$ play the role of the components of the random vector $\mathbf{x}$ in (3.3). In standard Monte Carlo simulations, in which the PMD is modeled using the coarse step method [37], the cosines of each of the angles θ_n are uniformly distributed in the interval $[-1, 1]$. One can note that an unbiased importance sampling simulation, for which $L(\mathbf{x}) = 1$, is exactly the same as a standard Monte Carlo simulation. However, in biased importance sampling one fixes a biasing parameter α and selects the $\cos\theta_n$ from a biased pdf, $p^*_\alpha(\cos\theta)$. In the

work reported in this dissertation, we choose $\cos\theta_n$ from the pdf [56]

$$p_\alpha^*(\cos\theta) = \frac{\alpha}{1-e^{-2\alpha}} e^{-\alpha(1-\cos\theta)}, \tag{3.7}$$

which biases $\cos\theta$ towards 1 when α is positive, and corresponds to standard Monte Carlo simulations in the limit $\alpha = 0$. The likelihood ratio for each value of $\cos\theta$ that is obtained from the biasing pdf in (3.7) is given by

$$L_\alpha(\cos\theta) = \frac{1-e^{-2\alpha}}{2\alpha} e^{\alpha(1-\cos\theta)}. \tag{3.8}$$

Since the values of $\cos\theta_n$ are independent random variables, the likelihood ratio for a biased realization of the fiber PMD is equal to the product of the likelihood ratios for each of its biased angles,

$$L_j(\mathbf{x}_{j,i}) = \prod_{n=1}^{N} L_{\alpha_j}(\cos\theta_{j,i,n}), \tag{3.9}$$

where α_j is the amount of bias used in the j-th distribution, and $\cos\theta_{j,i,n}$ is $\cos\theta$ for the n-th section of the i-th sample obtained from the j-th distribution.

To efficiently apply importance sampling, we must determine the value of the biasing parameter α that enables us to statistically resolve the histogram of the eye-opening penalty over a range of large eye-opening penalty values whose probability is on the order of a given target probability P_α, such as $P_\alpha = 10^{-5}$. Intuitively, we anticipate that the required value of α is the one for which the target probability P_α is equal to the likelihood ratio of the biased realization of the fiber PMD evaluated at the expected value of the random variable $\cos\theta$ with biasing pdf $p_\alpha^*(\cos\theta)$. That is, the bias parameter α satisfies the equation

$$P_\alpha = \left[L_\alpha\left(\langle\cos\theta\rangle\right)\right]^N = \left[L_\alpha\left(\frac{1+e^{-2\alpha}}{1-e^{-2\alpha}} - \frac{1}{\alpha}\right)\right]^N, \tag{3.10}$$

where N is the number of fiber sections, and $\langle\cdot\rangle$ is the expectation operator. Our motivation for this heuristic comes from (3.3) and from the observation that, over a given range of penalties, the biased samples statistically resolve the histogram of the eye-opening penalty when the indicator function for this range has the value 1 for a large proportion of the biased samples. In Fig. 3.1, I show the results in which we

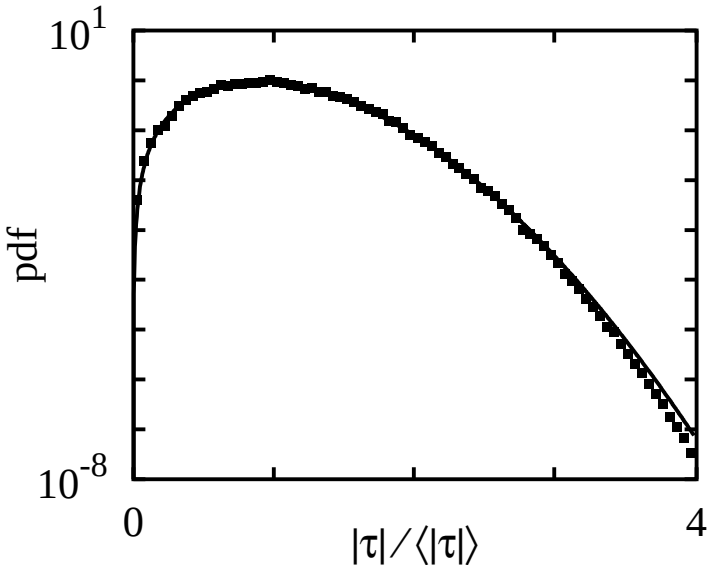

Figure 3.1: The pdf of the normalized DGD, $|\boldsymbol{\tau}|/\langle|\boldsymbol{\tau}|\rangle$, plotted on a logarithmic scale with 80 bins. The squares show the results of Monte Carlo simulations with importance sampling in which we only biased the DGD using 3×10^4 samples. The solid line shows the Maxwellian distribution with the same mean.

only biased the DGD, choosing $\alpha = 0$, which produces unbiased samples, together with $\alpha = 0.5$, and $\alpha = 1.0$. The target probabilities of the biased simulations are $P_{0.5} = 3.9 \times 10^{-2}$ and $P_{1.0} = 5.4 \times 10^{-6}$. We used 10^4 samples for each of the three biases, except as noted. We point out that as we increase the number of samples with bias parameter α, the size of the interval about P_α for which the histogram of the

eye-opening penalty is well resolved increases.

In order to bias the angle θ_n between the polarization dispersion vector of the first $(n-1)$ sections $\boldsymbol{\tau}^{(n-1)}$ and the polarization dispersion vector of the n-th section $\boldsymbol{\tau}_n$, the polarization rotation matrix S_n in (2.17) has to be modified to properly account for the bias. In practice, we bias the direction of the polarization dispersion vector of the previous $(n-1)$ sections $\boldsymbol{\tau}^{(n-1)}$ toward $\boldsymbol{\tau}_n$ because the polarization dispersion vector of the previous sections $\boldsymbol{\tau}^{(n-1)}$ is the vector rotated by the matrix T_n, as shown in (3.6). The polarization dispersion vector of any of the sections of the transmission fiber modeled by (2.14) is given by $\boldsymbol{\tau}_s = -\tau_s\,\hat{\mathbf{a}}_x$, as in (2.19), which is independent of S_n. Therefore, the matrix S_n of the n-th section must bias the polarization dispersion vector of the previous sections toward the vector $-\hat{\mathbf{a}}_x$. We obtain this bias by replacing the first two random rotations in (2.17) by the combination of one random rotation with two deterministic rotations around the y- and the z-axes. The first rotation $\mathsf{R}_y(\chi_n)$ eliminates the z-component of the polarization dispersion vector of the previous $(n-1)$ sections $\boldsymbol{\tau}^{(n-1)}$, which is accomplished by choosing $\chi_n = \arctan(\tau_z/\tau_x)$. The second rotation $\mathsf{R}_z(\zeta_n)$ eliminates the y-component of $\mathsf{R}_y(\chi_n)\,\boldsymbol{\tau}^{(n-1)}$, where ζ_n is chosen like χ_n, with the additional constraint that the resultant vector $\mathsf{R}_z(\zeta_n)\,\mathsf{R}_y(\chi_n)\,\boldsymbol{\tau}^{(n-1)}$ should be in the $-\hat{\mathbf{a}}_x$ direction. Then, we chose a random angle θ_n from the biasing pdf in (3.7) to rotate $\mathsf{R}_z(\zeta_n)\,\mathsf{R}_y(\chi_n)\,\boldsymbol{\tau}^{(n-1)}$ around the z-axis. This rotation can be combined with the previous deterministic rotation around the z-axis by ζ_n to produce a single rotation. Finally, we add a uniformly distributed random rotation ψ_n around the x-axis to obtain the polarization rotation

matrix for the DGD bias $\mathsf{S}_n^{(1)}$, which becomes

$$\mathsf{S}_n^{(1)} \;=\; \mathsf{R}_x\left(\psi_n\right)\mathsf{R}_z\left(\theta_n+\zeta_n\right)\mathsf{R}_y\left(\chi_n\right). \tag{3.11}$$

A uniform rotation like $\bar{\mathsf{S}}_n$ in (2.15) could in principle be added at the end of the fiber so that the direction of $\boldsymbol{\tau}^{(N)}$ is uniformly distributed on the Poincaré sphere. However, the receiver model used for the results in this dissertation has no polarization dependence; so, this final rotation is unnecessary.

In Fig. 3.1, I show the pdf of the normalized DGD, $\tau/\langle\tau\rangle$, of a fiber with 80 birefringent sections and 30 ps of mean DGD $\langle\tau\rangle$, where $\tau = |\boldsymbol{\tau}|$ and the DGD is normalized with respect to $\langle\tau\rangle$. Hence, these results are independent of $\langle\tau\rangle$. The unbiased probability of obtaining normalized DGD values outside the domain $[0, 4]$ that I show in Fig. 3.1 is less than 10^{-8}. This curve was obtained with only 10^4 samples from Monte Carlo simulations for each of the three biasing distributions: $\alpha = 0$, which produces unbiased samples, $\alpha = 0.5$, and $\alpha = 1.0$. The results of the biasing distributions were combined using the balanced heuristic method previously described. The largest relative variation over the domain $[0.3, 4]$ is 8%. The relative variation is the ratio between the standard deviation of the probability and the probability value. In Section 3.5, I describe in detail how we compute the relative variation for the results shown in this chapter. We observed an excellent agreement between the numerically calculated pdf of the DGD obtained with importance sampling and the Maxwellian pdf with the same mean. The slight deviation in the tail between the numerically calculated pdf and the Maxwellian distribution occurs because we use 80 sections rather than a larger number [57].

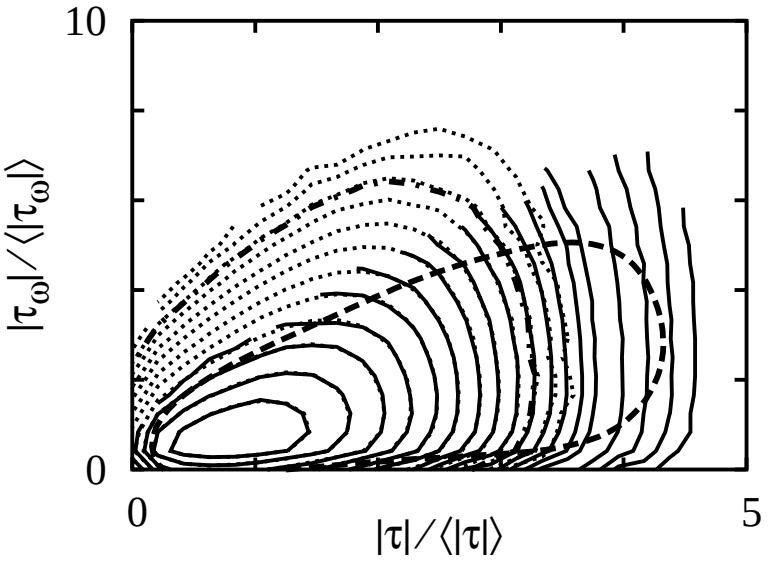

Figure 3.2: The joint pdf of the normalized $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$ with 25×25 bins. The solid lines show the results of Monte Carlo simulations with importance sampling in which we only biased the DGD using 3 biases with 6×10^4 samples in each bias. The dashed line shows the contour level corresponding to 10% relative variation in the results using importance sampling. The dotted lines show the results of standard Monte Carlo simulations using 10^8 samples. The dot-dashed line shows the contour level corresponding to 10% relative variation in the results using standard Monte Carlo simulations with 10^8 samples. The contours of the joint pdf from bottom to top of the plot, are at 3×10^{-n} , $n = 1, \cdots, 7$ and 10^{-m}, $m = 1, \cdots, 11$.

In Fig. 3.2, we compare the joint pdf of the DGD $|\boldsymbol{\tau}|$ and the length of the frequency derivative of the polarization dispersion vector $|\boldsymbol{\tau}_\omega|$ that is obtained with the implementation of the importance sampling with DGD bias to the joint pdf obtained with standard Monte Carlo simulations with 10^8 samples. We had the same configuration as in Fig. 3.1, except that we used 6×10^4 samples per bias. We observed that the length of the frequency derivative of the polarization dispersion vectors that are statistically correlated to the DGD that we bias are correctly accounted for. How-

ever, this implementation is not efficient in obtaining samples with large lengths of the frequency derivative of the polarization dispersion vector associated with moderately small values of DGD. Hence, the use of DGD bias is limited to systems where the DGD is the dominant source of penalties, which is the case in uncompensated systems and in systems with limited PMD compensation. In Fig. 3.9 of Section 3.7, I show that biasing only the DGD is not sufficient to accurately compute penalties in a single-section PMD-compensated system, where the compensator has a variable DGD-section in which the residual DGD of the system at the central frequency of the channel is canceled after compensation.

3.4 Importance sampling that biases the magnitude of first- and second-order PMD

As I will show in Section 3.7, biasing the DGD is sufficient to accurately calculate the uncompensated penalties and their pdfs, but it is not sufficient to accurately calculate the compensated penalties and their pdfs. To study compensated systems where second-order PMD also plays a role it is necessary to use an importance sampling method capable of biasing both the DGD and the length of the frequency derivative of the polarization dispersion version. Monte Carlo simulations with multiple importance sampling in which first- and second-order PMD is biased uses multiply-biased simulations that generate a sufficient number of combinations of first- and second-order PMD to cover the statistically significant regions of the first- and second-order PMD plane.

The derivative of the polarization dispersion vector with respect to the angular frequency ω after n fiber sections $\boldsymbol{\tau}_{\omega}^{(n)}$ is determined by the concatenation rule [49],

$$\boldsymbol{\tau}_{\omega}^{(n)} = \boldsymbol{\tau}_{n_\omega} + \boldsymbol{\tau}_n \times \left(\mathsf{T}_n \boldsymbol{\tau}^{(n-1)}\right) + \mathsf{T}_n \boldsymbol{\tau}_{\omega}^{(n-1)}, \tag{3.12}$$

where T_n is the Müller matrix that is equivalent to the Jones matrix in (2.14), and $\boldsymbol{\tau}_{n_\omega}$ is the derivative of the polarization dispersion vector of the n-th section with respect to the angular frequency. Note that $|\boldsymbol{\tau}_{n_\omega}| = 0$ in this problem, since each section has a constant—frequency independent—polarization dispersion vector $\boldsymbol{\tau}_n$.

In order to obtain large values of the frequency derivative of the polarization dispersion vector $\boldsymbol{\tau}_\omega$ with a relatively small number of Monte Carlo simulations, Fogal, *et al.* [29] demonstrated that it is necessary to bias the polarization dispersion vector of the n-th section in a direction $\hat{\mathbf{b}}$ that is different from the direction used in the DGD bias that was described in Section 3.3. The biasing direction $\hat{\mathbf{b}}$ is located in a plane that contains the vectors $\boldsymbol{\tau}^{(n-1)}$ and $\boldsymbol{\tau}^{(n-1)} \times \boldsymbol{\tau}_{\omega}^{(n-1)}$, and this direction is chosen so that the angle between the biasing direction $\hat{\mathbf{b}}$ and the polarization dispersion vector of the previous sections $\boldsymbol{\tau}^{(n-1)}$ varies linearly along the fiber sections from $\beta_1 = 0$ in the first fiber section to $\beta_N = \beta$ in the last fiber section, where the values of β and α in (3.7) determine the region in the plane formed by the DGD $|\boldsymbol{\tau}|$ and the length of the frequency derivative of the polarization dispersion vector $|\boldsymbol{\tau}_\omega|$ that one wants to statistically resolve. Specifically [29], my colleagues and I chose

$$\beta_n = \frac{n}{N}\beta, \tag{3.13}$$

where $0 \leq \beta \leq \pi$. Note that the choice $\beta = 0$ produces only DGD bias. However, the parameter α completely determines the target probability (3.10), since the pa-

rameter β simply selects a region of equal probability in the parameter space. For the simulation results presented in this dissertation, in which we biased both the DGD and the length of the frequency derivative of the polarization dispersion vector, we chose the following values of (α, β) to bias the distributions with 10^4 samples each: $(0, 0)$, which produces unbiased samples, $(0.5, 0)$, $(0.5, \pi/3)$, $(0.5, 2\pi/3)$, $(0.5, \pi)$, $(1, 0)$, $(1, \pi/3)$, $(1, 2\pi/3)$, $(1, \pi)$, and $(0.7, 0)$.

In order to implement the bias for both the DGD and the length of the frequency derivative of the polarization dispersion vector, the polarization rotation matrix S_n in (2.17) has to be modified in a way that is analogous to the derivation of $\mathsf{S}_n^{(1)}$ in Section 3.3. The goal is to choose a set of rotations so that the vector $\mathrm{T}_n^{-1}\boldsymbol{\tau}_n$ ends up at angle θ with the biasing direction $\hat{\mathbf{b}}$ in the three-dimensional Stokes space. The first step is similar to the one described in Section 3.3, where two deterministic rotations are obtained to rotate the polarization dispersion vector of the previous $(n-1)$ sections $\boldsymbol{\tau}^{(n-1)}$ to the $-\hat{\mathbf{a}}_x$ direction, $\mathsf{R}_z(\zeta_n)\,\mathsf{R}_y(\chi_n)\,\boldsymbol{\tau}^{(n-1)}$. Then, a rotation around the x-axis $\mathsf{R}_x(\nu_n)$ eliminates the y-component of $\mathsf{R}_z(\zeta_n)\,\mathsf{R}_y(\chi_n)\,\boldsymbol{\tau}_\omega^{(n-1)}$, while leaving the z-component positive. The next step is to apply a deterministic rotation $\mathsf{R}_z(\beta_n)$ so that $[\mathsf{R}_z(\beta_n)\,\mathsf{R}_x(\nu_n)\,\mathsf{R}_z(\zeta_n)\,\mathsf{R}_y(\chi_n)]^{-1}\boldsymbol{\tau}_n$ is parallel to the biasing direction $\hat{\mathbf{b}}$, where β_n is determined by (3.13). Then, a uniformly distributed rotation around the x-axis $\mathsf{R}_x(\psi_n)$ is added to produce the correct statistical randomization of the polarization rotation matrix of first- and second-order bias $\mathsf{S}_n^{(2)}$. Finally, a biased rotation around the z-axis, $\mathsf{R}_z(\theta_n)$, is applied to obtain an appropriate bias for both the DGD and the length of the frequency derivative of the polarization dispersion vector, where $\cos\theta_n$ is obtained from the pdf in (3.7). The matrix $\mathsf{S}_n^{(2)}$ in this case

becomes

$$\mathsf{S}_n^{(2)} \;=\; \mathsf{R}_z\left(\theta_n\right)\mathsf{R}_x\left(\psi_n\right)\mathsf{R}_z\left(\beta_n\right)\mathsf{R}_x\left(\nu_n\right)\mathsf{R}_z\left(\zeta_n\right)\mathsf{R}_y\left(\chi_n\right), \tag{3.14}$$

where ψ_n is a random variable whose pdf is uniformly distributed between 0 and 2π as in (3.11). Note that none of the angles χ_n, ζ_n, ν_n, and β_n in (3.14) are random, and that these rotations are not unique; it is possible to produce a bias for both the DGD and the length of the frequency derivative of the polarization dispersion vector using a different set of elementary rotations. A uniform rotation like $\bar{\mathsf{S}}_n$ in (2.15) could be added at the end of the fiber model to make sure that $\boldsymbol{\tau}^{(N)}$ is uniformly distributed on the Poincaré sphere. However, this extra rotation is not necessary here, just as in the case in which only the DGD is biased.

In Fig. 3.3, I show the pdf of the length of the frequency derivative of the polarization dispersion vector $\boldsymbol{\tau}_\omega$ of a fiber with 80 birefringent sections and 10 ps of mean DGD $\langle|\tau|\rangle$. I show the length of $\boldsymbol{\tau}_\omega$ normalized with respect to its expected value $\langle|\boldsymbol{\tau}_\omega|\rangle$. In Fig. 3.3, we combined the results of the 10 biasing distributions with 10^4 samples per bias using the balanced heuristic method. The largest relative variation over the domain $[0, 9]$ is 17%. We observed an excellent agreement between the numerically calculated pdf of the length of the frequency derivative of the polarization dispersion vector obtained using importance sampling and the theoretical pdf of the length of the frequency derivative of the polarization dispersion vector in (2.11) [40], [42].

In Fig. 3.4, I show the results of the joint pdf of the DGD $|\boldsymbol{\tau}|$ and the length of the frequency derivative of the polarization dispersion vector $|\boldsymbol{\tau}_\omega|$ that is obtained with multiple importance sampling in which both the DGD and the length of the frequency

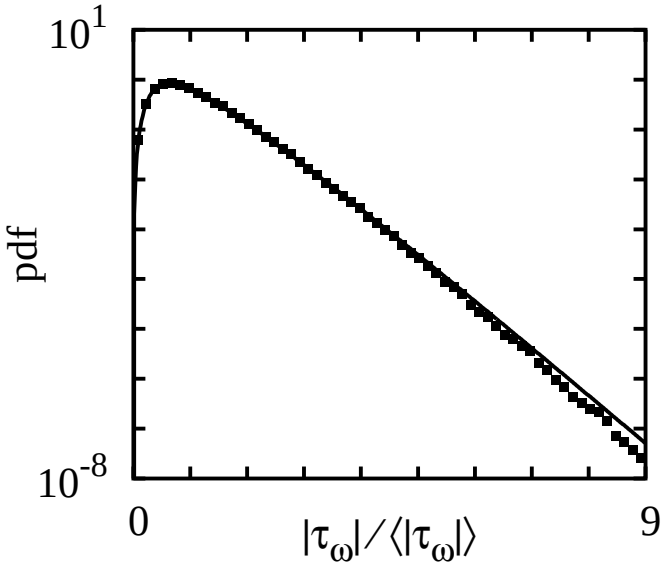

Figure 3.3: The pdf of the normalized $|\boldsymbol{\tau}_\omega|$ plotted on a logarithmic scale with 60 bins. The squares show the results of Monte Carlo simulations with importance sampling in which we biased both the first- and second-order PMD using 10^5 samples. The solid line shows the results of the theoretical distribution of the length of the frequency derivative of the polarization dispersion vector.

derivative of the polarization dispersion vector are biased in comparison with results of standard Monte Carlo simulations with 10^8 samples. We had the same configuration as in Fig. 3.3, except that we used 6×10^4 samples per bias. We observed an excellent agreement between these results. We point out that the relative variation in the joint pdf of the DGD and the length of the frequency derivative of the polarization dispersion vector in the results of standard Monte Carlo simulations depends only on the number of samples used. In addition to the number of samples, the relative variation in the results with importance sampling strongly depends on the set of biases that are combined to produce the numerical joint pdf. As a consequence, the contours of relative variation do not follow the probability contour lines and have a bumpy

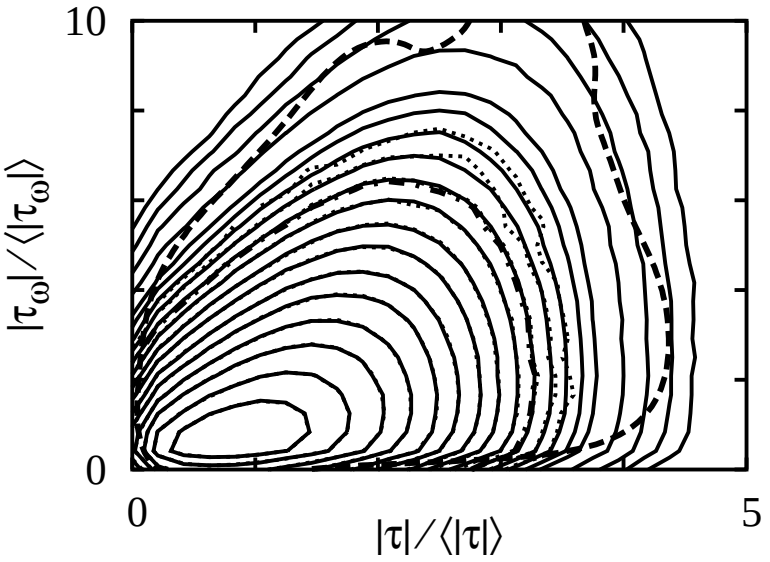

Figure 3.4: The joint pdf of the normalized $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$ with 25×25 bins. The solid lines show the results with importance sampling using 10 biases with 6×10^4 samples in each bias. The dashed line shows the contour level corresponding to 10% relative variation in the results using importance sampling. The dotted lines show the results of standard Monte Carlo simulations using 10^8 samples. The dot-dashed line shows the contour level corresponding to 10% relative variation in the results using standard Monte Carlo simulations with 10^8 samples. The contours of the joint pdf from the bottom to the top of the plot, are at 3×10^{-n} , $n = 1, \cdots, 7$ and 10^{-m}, $m = 1, \cdots, 11$.

structure. I show this behavior in Fig. 3.4 for the 10% relative variation contour. However, we see that the 10% relative variation contour for the biasing simulations lies well outside the 10% relative variation contour for the standard simulations, although the standard simulations have a much larger number of samples. This result supports the conclusion that the biasing procedure is effective.

My colleagues and I observed that one single biasing distribution is insufficient to accurately determine the eye-opening penalty over the entire range of interest in

the $|\boldsymbol{\tau}| - |\boldsymbol{\tau}_\omega|$ plane. It is necessary to combine the statistical result from multiple biasing distributions, since each distribution provides accuracy in different regions of the parameter space. As in the case of importance sampling in which only the DGD is biased, an efficient technique to combine the samples from multiple biasing distributions when using importance sampling in which both $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$ are biased is the balanced heuristic method [55] that was described earlier in Section 3.3.

3.5 Estimators of the mean and of the variance

As I mentioned before, error estimation is important to verify the accuracy of the results estimated with Monte Carlo methods, including importance sampling. When applying importance sampling in which we only bias the DGD, a confidence interval for the estimator of the probability P_I of the indicator function $I(\mathbf{x})$ in (3.3) can be defined from the estimator of the variance of $\hat{P}_I$, which is given by [50]

$$\hat{\sigma}^2_{\hat{P}_I} = \sum_{j=1}^{J} \frac{1}{M_j\,(M_j-1)} \sum_{i=1}^{M_j} \left[I(\mathbf{x}_{j,i}) w_j(\mathbf{x}_{j,i}) L_j(\mathbf{x}_{j,i}) - \hat{P}_{I_j} \right]^2, \tag{3.15}$$

where

$$\hat{P}_{I_j} = \frac{1}{M_j} \sum_{i=1}^{M_j} I(\mathbf{x}_{j,i}) w_j(\mathbf{x}_{j,i}) L_j(\mathbf{x}_{j,i}), \tag{3.16}$$

is the contribution of the samples drawn from j-th biased distribution to the estimator $\hat{P}_I$. The confidence interval of the estimator $\hat{P}_I$ is defined to be the interval $\left(\hat{P}_I - \hat{\sigma}_{\hat{P}_I}, \hat{P}_I + \hat{\sigma}_{\hat{P}_I}\right)$. The relative variation equals $\hat{\sigma}_{\hat{P}_I}/\hat{P}_I$. The key to the derivation of (3.15) is to note that the product between the indicator function $I(\mathbf{x})$ and its corresponding weight $w_j(\mathbf{x})L(\mathbf{x})$ is the random variable in the standard expression of

the variance [58] for the j-th distribution. Then, the variances produced by all the independent biased distributions are added to obtain the expression for the variance of the estimator $\hat{P}_I$ in (3.15).

When applying multiple importance sampling in which both first- and second-order PMD are biased, one can determine the estimator of the average eye-opening penalty $\hat{\mu}$ given a value of the magnitude of first- and second-order PMD, $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$, respectively, and the expected variance $\hat{\sigma}^2_{\hat{\mu}}$ of this estimator in each bin, using the estimators [24]

$$\hat{\mu} \;=\; \frac{\hat{M}}{\hat{C}}, \qquad \hat{\sigma}^2_{\hat{\mu}} \;=\; \frac{\hat{\sigma}^2_{\hat{M}}}{\hat{C}^2} + \frac{\hat{\mu}^2 \hat{\sigma}^2_{\hat{C}}}{\hat{C}^2}, \tag{3.17}$$

where

$$\hat{C} \;=\; \sum_{j=1}^{J} \hat{C}_j, \qquad \hat{M} \;=\; \sum_{j=1}^{J} \hat{M}_j, \tag{3.18}$$

with

$$\hat{C}_j \;=\; \frac{1}{N_j} \sum_{i=1}^{N_j} I\left(\mathbf{x}_{j,i}\right) w_j\left(\mathbf{x}_{j,i}\right) L_j\left(\mathbf{x}_{j,i}\right), \tag{3.19a}$$

$$\hat{M}_j \;=\; \frac{1}{N_j} \sum_{i=1}^{N_j} f\left(\mathbf{x}_{j,i}\right) I\left(\mathbf{x}_{j,i}\right) w_j\left(\mathbf{x}_{j,i}\right) L_j\left(\mathbf{x}_{j,i}\right). \tag{3.19b}$$

Using (3.17), one can generate one-standard-deviation confidence intervals for the average penalty in each bin, which is computed using importance sampling. In (3.17)–(3.19), the parameter N_j is the number of samples drawn from the j-th biased distribution $p^*_j(\mathbf{x})$. The vector $\mathbf{x}_{j,i}$ is the set of biased parameters in the i-th fiber realization of the j-th distribution, where $f\left(\mathbf{x}_{j,i}\right)$ is the associated eye-opening penalty, and J is equal to ten for the results shown in this dissertation, which comprise nine biased and one unbiased simulation. The likelihood ratio of the i-th fiber realization in the

j-th biased distribution is $L_j(\mathbf{x}_{j,i}) = p_j(\mathbf{x}_{j,i})/p_j^*(\mathbf{x}_{j,i})$, where $I(\mathbf{x}_{j,i})$ is the indicator function whose value is one in a bin of interest and zero outside this bin, and $w_j(\mathbf{x}_{j,i})$ are the weights associated with each individual distribution [29]. The distributions $p_j(\mathbf{x})$ and $p_j^*(\mathbf{x})$ are the unbiased and biased pdfs of a random vector $\mathbf{x}$, respectively. Finally,

$$\hat{\sigma}^2_{\hat{C}} = \sum_{j=1}^{J} \frac{1}{N_j(N_j-1)} \sum_{i=1}^{N_j} [I(\mathbf{x}_{j,i})\, w_j(\mathbf{x}_{j,i})\, L_j(\mathbf{x}_{j,i}) - \hat{C}_j]^2, \tag{3.20a}$$

$$\hat{\sigma}^2_{\hat{M}} = \sum_{j=1}^{J} \frac{1}{N_j(N_j-1)} \sum_{i=1}^{N_j} [f(\mathbf{x}_{j,i})\, I(\mathbf{x}_{j,i})\, w_j(\mathbf{x}_{j,i})\, L_j(\mathbf{x}_{j,i}) - \hat{M}_j]^2. \tag{3.20b}$$

It is important to note that the estimators for the mean eye-opening penalty $\hat{\mu}$ and the variance $\hat{\sigma}^2_{\hat{\mu}}$ of the estimator $\hat{\mu}$ given by (3.17) are biased estimators. We obtain the expression for the variance by using the law of propagation of errors [59], where to first-order approximation,

$$\frac{\hat{\sigma}^2_{\hat{\mu}}}{\hat{\mu}^2} = \frac{\hat{\sigma}^2_{\hat{M}}}{\hat{M}^2} + \frac{\hat{\sigma}^2_{\hat{C}}}{\hat{C}^2}. \tag{3.21}$$

The bias in the estimation of $\hat{\mu}$ and $\hat{\sigma}^2_{\hat{\mu}}$ can be reduced by computing $\hat{C}$ with a much larger number of samples than is used to compute $\hat{M}$. This approach is practical because the computational cost of generating fiber realizations to calculate $\hat{C}$ is much smaller than the cost to compute penalties after compensation, which are required to compute $\hat{M}$. For the results that I show in the Section 3.7, we reduced the uncertainty in $\hat{C}$ by computing $\hat{C}$ using 10^7 samples per biased simulation, while $\hat{M}$ was computed using 10^5 samples per biased simulation. Using this approach we note that $\hat{\sigma}^2_{\hat{C}}/\hat{C}^2$ is, in general, two orders of magnitude smaller than $\hat{\sigma}^2_{\hat{M}}/\hat{M}^2$. The maximum value for $\hat{\sigma}^2_{\hat{C}}/\hat{C}^2$ for the results shown in Section 3.7 is 1.21×10^{-4}. The maximum value for

$\hat{\sigma}^2_{\hat{M}}/\hat{M}^2$ is 0.14, but it is much smaller in almost all bins, typically between 0.001 and 0.002.

3.6 Comparison of penalties resulting from first-order and all-order PMD

Using importance sampling in which only the DGD is biased, my colleagues and I compared the eye-opening penalty from a first-order PMD model with the penalty from an all-order PMD model in uncompensated optical fiber transmission systems [26], [27]. Evaluating the performance taking into account only first-order PMD produces a good approximation to the true eye-opening penalty of uncompensated systems when the penalty is low. However, when the penalties are high, this model overestimates the penalty for outage probabilities in the range of interest for systems designers, which is typically approximately 10^{-6} to 10^{-5}. These results demonstrate the importance of accurately modeling PMD by including higher orders.

As I mentioned in Chapter 2, as long as the signal bandwidth is sufficiently narrow, the first-order PMD distortion is the dominant PMD effect. In the first-order approximation, the principal states of polarizations and the DGD do not vary with frequency. The correlation bandwidth [34] $\Delta\nu_{\mathrm{PMD}}$ over which the DGD can be considered constant is inversely proportional to the expected value of the DGD with a constant of proportionality close to 1/2. For pulses with a small bandwidth compared to $\Delta\nu_{\mathrm{PMD}}$, which means that the expected DGD is small compared to the duration of the optical pulse, the first-order model is useful in practice.

Because of the simplicity of the first-order PMD model and its usefulness, it has been frequently used for the performance evaluation of systems [60]–[62]. In this section, I show that performance evaluation taking into account only first-order PMD produces a good approximation of the true eye-opening penalty of uncompensated optical systems when the penalty is low. However, such evaluation overestimates the true eye-opening penalty, *i.e.*, the penalty obtained from a PMD model that takes into account the frequency variation of the DGD and of the principal states of polarizations, at the higher penalties that are associated with outage probabilities of 10^{-6} and 10^{-5}.

In Figs. 3.5–3.7, I show the outage probability of the eye-opening penalty caused by PMD in an uncompensated 10 Gbit/s nonreturn-to-zero (NRZ) system. In Fig. 3.5, we compare the eye-opening penalty from a first-order PMD model with that of an all-order PMD model when the mean DGD $\langle|\boldsymbol{\tau}|\rangle$ is equal to 14 ps, for a typical NRZ system whose pulses have rise and fall times of 30 ps, and whose receiver has an electrical filter with a full width of half maximum (FWHM) bandwidth equal to 8.6 GHz. At this low value of $\langle|\boldsymbol{\tau}|\rangle$ the system penalty is dominated by first-order PMD, so that the eye-opening penalty when one considers all orders of PMD is very close to the eye-opening penalty when only the first order is considered. The PMD-induced penalty is thus highly correlated with the DGD at the central frequency of the channel [3]. In this case, the performance evaluation taking into account only first-order PMD produces a good approximation for the true eye-opening penalty.

In Fig. 3.6, I show the curves of outage probability versus eye-opening penalty for the same system as in Fig. 3.5 with $\langle|\boldsymbol{\tau}|\rangle =$ 20 ps. In Fig. 3.6, we observe that the

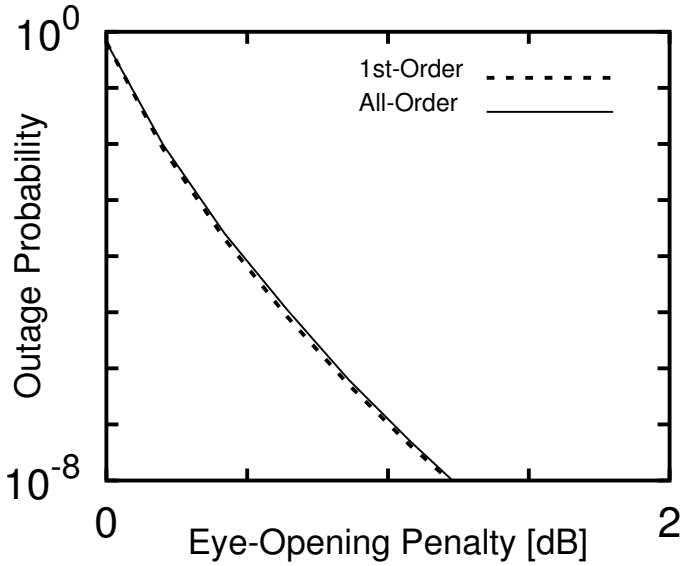

Figure 3.5: Outage probability as a function of the eye-opening penalty margin for a 10 Gbit/s NRZ system with pulse edge rise and fall times of 30 ps and mean DGD, $\langle|\boldsymbol{\tau}|\rangle$, equal to 14 ps. The dashed line shows the outage probability considering only first-order PMD, while the solid line shows the results considering all-order (first- and higher-order) PMD distortion. The outage probability is the probability that the eye-opening penalty exceeds the value displayed on the horizontal axis.

first-order PMD model yields an eye-opening penalty that is slightly greater than that of the all-order PMD model, for outage probability on the order of 10^{-6}. Finally, in Fig. 3.7, I show the eye-opening penalty results when we compare first- and all-order PMD in an NRZ system whose pulses have rise and fall times of 5 ps and whose receiver has an electrical filter with a FWHM bandwidth equal to 10 GHz. As we can see in Fig. 3.7, the difference in eye-opening penalty values between the first- and all-order models is larger than the previous case illustrated in Fig. 3.6 because the pulse format used in this system has a broader bandwidth. In Figs. 3.6 and 3.7, we observe that as higher-order PMD effect become more important, the true

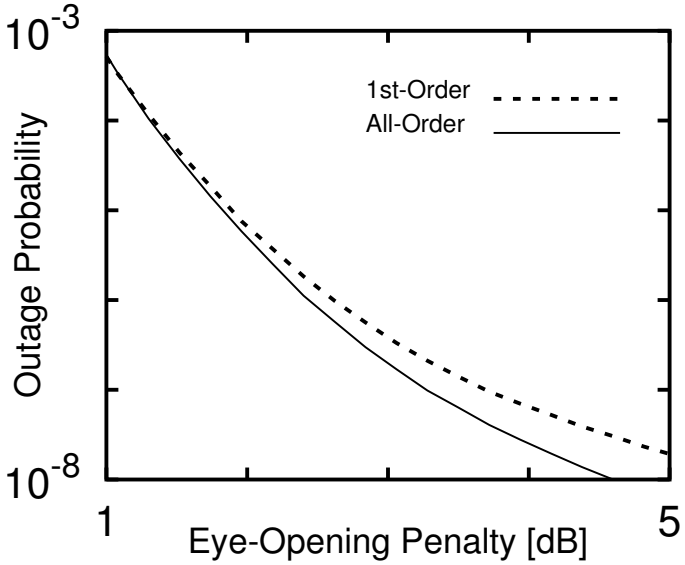

Figure 3.6: Same as Fig. 3.5, except that $\langle|\boldsymbol{\tau}|\rangle = 20$ ps and the pulse edge rise and fall times are 30 ps.

eye-opening penalty tends to be smaller than the penalty produced by the first-order PMD model. In particular, in Fig. 3.7, the first-order PMD model overestimates the true eye-opening penalty that is due to all orders of PMD for the outage probability value of 10^{-6} by approximately 0.8 dB.

To understand why the first-order PMD model overestimates the penalty for outage probabilities in the range of interest for system designers, which is typically about 10^{-5} to 10^{-6}, we studied how the DGD varies as a function of the angular frequency ω in the neighborhood of a frequency ω_c where the DGD is large. In general, the DGD has a non-zero slope that is due to first- and second-order PMD. However, when the DGD is large, the sign of the slope has no effect on the DGD averaged over the bandwidth of the signal when the pulse's frequency spectrum is symmetric. Consequently, the slope of the DGD will have a small effect on the penalty. Moreover,

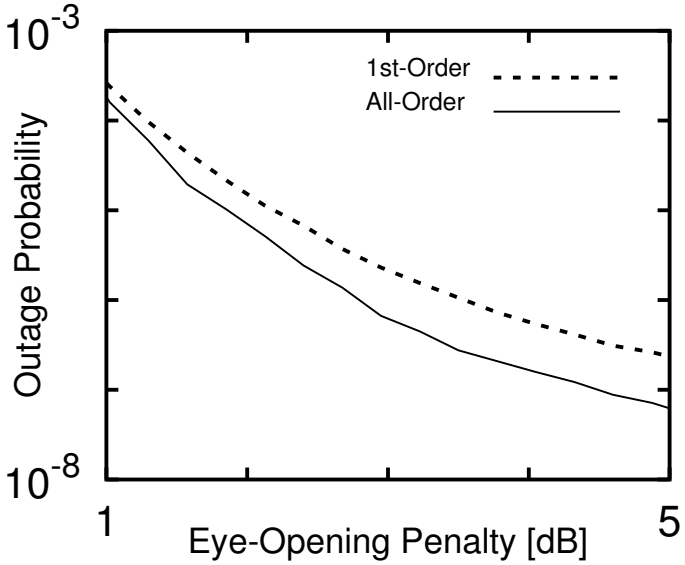

Figure 3.7: Same as Fig. 3.5, except that $\langle|\boldsymbol{\tau}|\rangle = 20$ ps and the pulse edge rise and fall times are 5 ps.

since the effect of the slope causes a small increase in the penalty as often as a small decrease, the expected penalty will depend only weakly on the slope when the DGD is large. In general, the DGD will also have a curvature due to the first three orders of PMD. For large DGD, this curvature will be negative, reducing the DGD when averaged over the bandwidth of the signal, more often than it will be positive. Thus, the expected penalty will be reduced. To quantify this observation, we express the curvature of the DGD $|\boldsymbol{\tau}|_{\omega\omega}$, which depends on the first three orders of PMD, as

$$|\boldsymbol{\tau}|_{\omega\omega} = \frac{|\boldsymbol{\tau}_\omega|^2}{|\boldsymbol{\tau}|} - \frac{(\boldsymbol{\tau}\cdot\boldsymbol{\tau}_\omega)^2}{|\boldsymbol{\tau}|^3} + \frac{\boldsymbol{\tau}\cdot\boldsymbol{\tau}_{\omega\omega}}{|\boldsymbol{\tau}|}. \tag{3.22}$$

Using (3.22), we calculated the conditional expectation of $|\boldsymbol{\tau}|_{\omega\omega}$ at the channel's central frequency ω_c as a function of the DGD, which I show in Fig. 3.8. We used a total of 10^6 Monte Carlo realizations with importance sampling in which only the

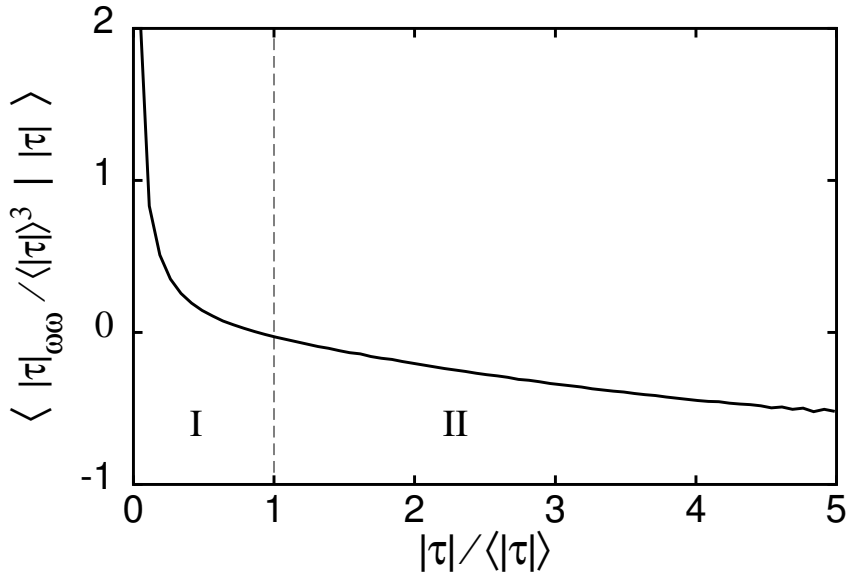

Figure 3.8: Conditional expectation of the second derivative of the DGD, $|\boldsymbol{\tau}|_{\omega\omega}$, given a value of the DGD, $|\boldsymbol{\tau}|$.

DGD is biased [28] and 80 equally spaced bins in the range $0 \leq |\boldsymbol{\tau}| \leq 5$ to generate the curve in Fig. 3.8. We normalized the DGD $|\boldsymbol{\tau}|$ by the mean DGD $\langle|\boldsymbol{\tau}|\rangle$ and $|\boldsymbol{\tau}|_{\omega\omega}$ by $\langle|\boldsymbol{\tau}|\rangle^3$ in order to obtain results that are independent of the mean DGD of the fiber. This conditional expectation $\langle|\boldsymbol{\tau}|_{\omega\omega} \mid |\boldsymbol{\tau}|\rangle$ gives the value of $|\boldsymbol{\tau}|_{\omega\omega}$ when averaged over all the fiber realizations for a given value of $|\boldsymbol{\tau}|$. When the DGD is larger than the mean, corresponding to region II in Fig. 3.8, the value of $\langle|\boldsymbol{\tau}|_{\omega\omega} \mid |\boldsymbol{\tau}|\rangle$ is negative, leading to a reduction of the average DGD when averaged over both the bandwidth of the signal and over fiber realizations, which in turn leads to a reduction of the eye-opening penalty margin at a given outage probability.

This explanation, which only takes into account PMD up to the third order, is merely a heuristic explanation of the results shown in Figs. 3.5–3.7, which include effects at all orders of PMD. When the DGD is large, it is reasonable to suppose

that orders even higher than third might play a significant role. It is possible that the whole notion of a higher-order PMD expansion, which implicitly assumes that the polarization dispersion vector can be written as a convergent Taylor series over the bandwidth of the signal, may break down at the large DGD values that typically produce unacceptable penalties. The resolution of this issue lies beyond the scope of this dissertation. The results do, however, show conclusively that simple theoretical models that take into account only first-order PMD overestimate the penalty when the DGD is large compared to the mean DGD.

3.7 Analysis of the performance of single-section PMD compensators using importance sampling

In this section, I present results in which my colleagues and I used multiple importance sampling to bias the magnitude of both first- and second-order PMD in order to evaluate the performance of single-section PMD compensators in a large region of the first- and second-order PMD plane. I show that importance sampling yields estimates of the average penalty with low variance over the entire region of interest of first- and second-order PMD. I also show that there is little advantage in using a compensator with a variable-DGD element and that the performance of a compensator that minimizes the residual DGD at the central frequency of the channel is considerably worse than a compensator that maximizes the eye opening.

Many performance studies of PMD compensators have focused on the average pulse spreading reduction, and hence the average bit error ratio (BER) of optical systems. However, reducing the average BER may not significantly reduce the outage

probability in the range of 10^{-6}–10^{-5} where real systems must operate [3]. Previous studies that relied on the reduction of the average BER have also not addressed how the penalty is explicitly related to first- and second-order PMD. The use of multiple importance sampling in which both first- and second-order PMD are biased allows one to efficiently study important rare events with large first- and second-order PMD. Therefore, one can accurately calculate outage probabilities for the PMD-induced penalty on the order of 10^{-5} or less in compensated or uncompensated systems. As in Section 3.3, we note that third- and higher-order PMD are also included in the simulations, but we do not specifically bias our simulations toward values of third- and higher-order PMD other than the moderately large values that appear naturally when one biases the first- and second-order PMD values.

In order to compensate for PMD, my colleagues and I considered two types of single-section PMD compensators. The first type consists of a polarization controller (PC) followed by a polarization maintaining fiber that has a fixed DGD element. The second type also uses a PC, but has a variable-DGD element. The parameters of the PC's orientation are the only free parameters that a compensator with a fixed-DGD element possesses, while the value of the DGD is an extra free parameter that the variable-DGD compensator has to control. These compensators have a small number of free parameters to control, as opposed to compensators with multiple sections, which makes them attractive as PMD compensators. The implementation of these compensators was explained in detail in Section 3.2.

Once again, my colleagues and I used the eye opening for performance evaluation of the PMD-compensated systems that we investigated. We computed the joint

pdf of the magnitude of first- and second-order PMD, $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$, using multiple importance sampling applied to first- and second-order PMD [29], and we computed the average value of eye-opening penalty given a value of $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$, where the subscript ω represents the derivative with respect to the angular frequency ω.

We simulated a 10 Gbit/s nonreturn-to-zero system with a mean DGD of 30 ps. The fiber model uses 80 sections of birefringent fiber with the coarse step method, which reproduces first- and higher-order PMD distortions within the probability range of interest. To adequately cover the $|\boldsymbol{\tau}|$–$|\boldsymbol{\tau}_\omega|$ plane, we combined nine biased simulations and one unbiased simulation with 10^5 samples each, using balanced heuristics [55]. For each fiber realization, we computed $|\boldsymbol{\tau}|$, $|\boldsymbol{\tau}_\omega|$, and the eye-opening penalty. Dividing the $|\boldsymbol{\tau}|$–$|\boldsymbol{\tau}_\omega|$ plane into $25 \times 25 = 625$ evenly-spaced bins where we set the maximum values for $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$ as $4\,\langle|\boldsymbol{\tau}|\rangle$ and $6\,\langle|\boldsymbol{\tau}_\omega|\rangle$, respectively, we then determined the estimator of the average eye-opening penalty value and the expected variance of this estimator in each bin, using the estimators described in Section 3.5.

In Fig. 3.9, I show the pdf of the eye-opening penalty for a system with 30 ps mean DGD and a single-section PMD compensator. The compensator consists of a variable DGD element, in which the residual DGD of the system at the central frequency of the channel is canceled after compensation. We computed the pdf using importance sampling in which we only biased the DGD, and we also computed the pdf using importance sampling in which we biased both the first- and second-order PMD. We observed that it is not sufficient to only bias the DGD in order to accurately calculate the compensated penalty and its pdf.

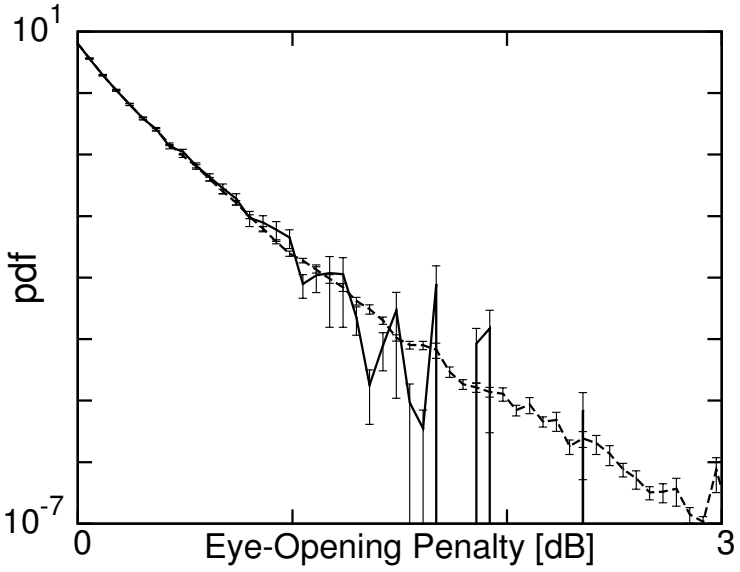

Figure 3.9: Probability density function of the eye-opening penalty for a system with a mean DGD of 30 ps and a single-section compensator. (i) Solid line: Results using importance sampling in which we only biased the DGD. (ii) Dashed line: Results using importance sampling in which we biased both the first- and second-order PMD. The confidence interval is shown with error bars.

In Fig. 3.10, I show contour plots (dotted lines) of the joint pdf of the magnitudes of first- and second-order PMD, $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$ for an uncompensated system, which have been obtained as in [29]. I also show contours of the eye-opening penalty (solid lines) for an uncompensated system and the eye-opening penalty with one-standard-deviation added and subtracted (dashed lines). These three sets of curves are then smoothed using an N-th order Bezier smoothing algorithm [63], where N is the number of points in the contour. The dashed lines represent the one-standard-deviation confidence intervals for the penalty, given by $\hat{\sigma}_{\hat{\mu}}$ in (3.17), and are quite narrow except at the edges of the plot, demonstrating the effectiveness of importance sampling in reducing the variance of the estimator of the penalty in this case. The region of the

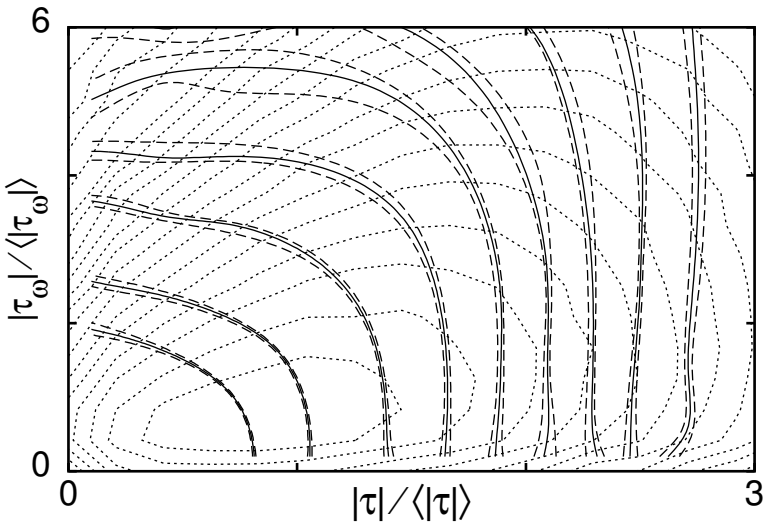

Figure 3.10: Uncompensated system: The dotted lines show contour plots of the joint pdf of the normalized first- $|\boldsymbol{\tau}|$ and second-order PMD $|\boldsymbol{\tau}_\omega|$. The solid and the dashed lines show contour plots of the conditional expectation of the eye-opening penalty and the confidence interval of the contour plots, respectively. The contours of the joint pdf from the bottom to the top of the plot, are at 3×10^{-n} , $n = 1, \cdots, 7$ and 10^{-m}, $m = 1, \cdots, 11$. The curves of the conditional expectation of the eye-opening penalty in dB from the bottom to the top of the plot, are at $0.1, 0.2, 0.4, 0.6, 0.9, 1.2, 1.6, 2.2, 3.2$.

$|\boldsymbol{\tau}|$–$|\boldsymbol{\tau}_\omega|$ plane that is the dominant source of a given penalty is where the corresponding penalty level curve intersects the contour of the joint pdf of $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$ with the highest probability. The contour plots for penalties beyond 1.2 dB are approximately parallel to the second-order PMD axis, indicating the expected result that first-order PMD is the dominant cause of penalty in this uncompensated system.

In Figs. 3.11–3.13, I show the contours of the eye-opening penalties when different PMD compensators are used. The eye-opening penalty contours are plotted as a function of the uncompensated $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$, and I show the same contours of their joint pdf, as in Fig. 3.10. In Fig. 3.11, I show the eye-opening penalty with a

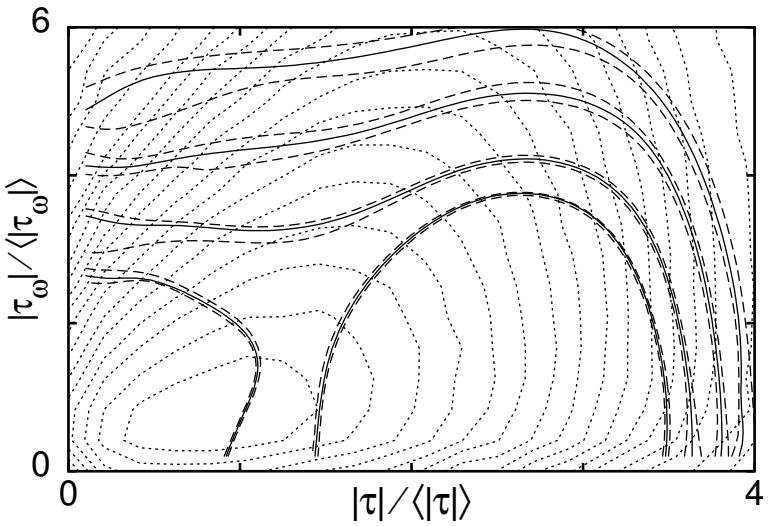

Figure 3.11: Same set of curves of Fig. 3.10 for a compensated system with a fixed-DGD compensator with constant DGD element equal to 2.5 $\langle|\boldsymbol{\tau}|\rangle$. The curves of the conditional expectation of the eye-opening penalty in dB from the bottom to the top of the plot, are at $0.1, 0.2, 0.3, 0.4$.

fixed-DGD compensator with a 75 ps DGD element, in which the polarization transformation produced by the polarization controller has been optimized to maximize the eye opening [64]. In Fig. 3.12, I show the eye-opening penalty with a variable-DGD compensator, in which, once again, the eye opening has been maximized. In Fig. 3.13, I show the eye-opening penalty with a variable-DGD compensator, in which the residual DGD of the system at the central frequency of the channel has been minimized after compensation.

The first observation that my colleagues and I made, comparing Figs. 3.11 and 3.12, is that for penalties above 0.2 dB, the performance of the fixed-DGD compensator is comparable to the variable-DGD compensator as long as $|\boldsymbol{\tau}| \,/\, \langle|\boldsymbol{\tau}|\rangle \leq 3.0$. The domain $|\boldsymbol{\tau}| \,/\, \langle|\boldsymbol{\tau}|\rangle > 3.0$ corresponds to an outage probability in the uncompensated

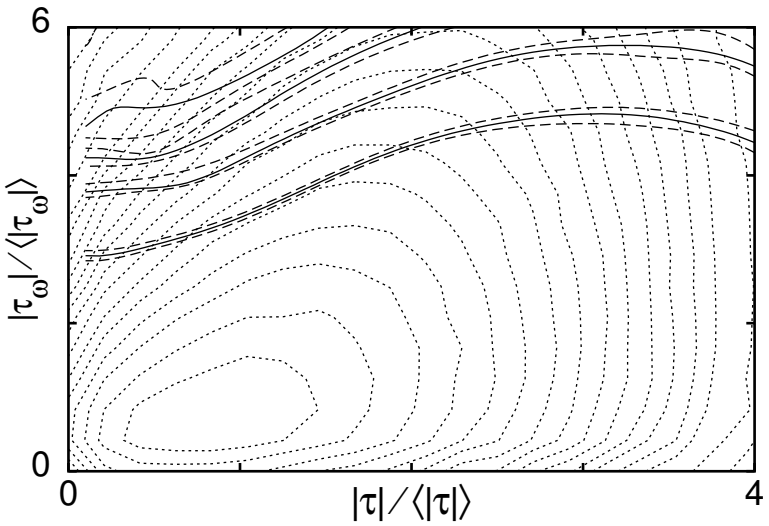

Figure 3.12: Same set of curves of Fig. 3.10 for a compensated system with a variable-DGD compensator with eye opening maximization. The curves of the conditional expectation of the eye-opening penalty in dB from the bottom to the top of the plot, are at $0.1, 0.2, 0.3, 0.4$.

system of less than 10^{-7}, which is usually negligible. As expected, we also observed that the penalty with the variable-DGD compensator is dominated by higher-order PMD. We inferred this result by noting that the contour lines of the penalty are nearly parallel to the $|\boldsymbol{\tau}| \,/\, \langle|\boldsymbol{\tau}|\rangle$–axis, indicating that the penalty is nearly independent of $|\boldsymbol{\tau}|$. Perhaps a bit more surprisingly, we observed the same result with the fixed-DGD compensator as long as the penalty is above 0.2 dB and $|\boldsymbol{\tau}| \,/\, \langle|\boldsymbol{\tau}|\rangle \leq 3.0$. Comparing Figs. 3.12 and 3.13, we observed that a variable-DGD compensator that minimizes the residual DGD performs significantly worse than a compensator that maximizes the eye opening. This result indicates once again the importance of higher-order PMD in the compensator performance.

Finally, in Fig. 3.14, I plot the outage probability ($\hat{P}_o$) as a function of the eye-

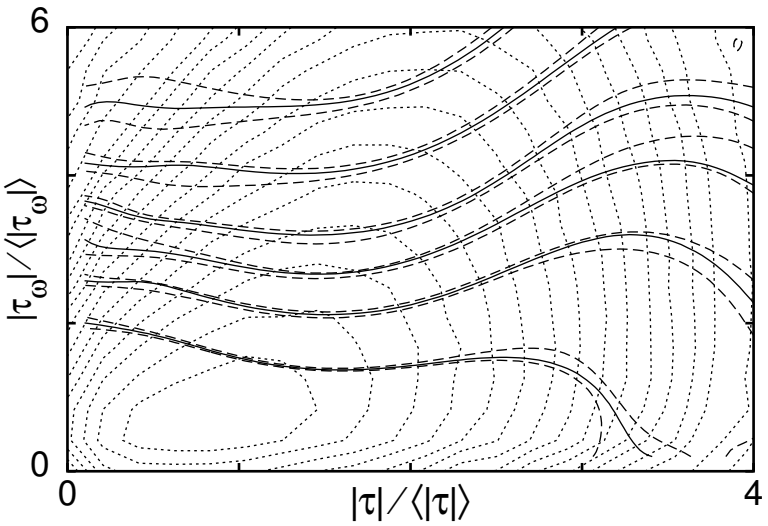

Figure 3.13: Same set of curves of Fig. 3.10 for a compensated system with a variable-DGD compensator with minimized DGD after compensation at the central frequency of the channel. The solid lines show the contours of the conditional expectation of the eye-opening penalty in dB from the bottom to the top of the plot, are at $0.1, 0.2, 0.3, 0.4, 0.6, 0.9$.

opening penalty for the compensators that we studied. The maximum relative error ($\hat{\sigma}_{\hat{P}_o}/\hat{P}_o$) for the curves shown in this plot equals 0.14. This plot confirms the results that we inferred from Figs. 3.11–3.13. The performance of the fixed- and variable-DGD compensators is comparable. The performance of a variable-DGD compensator that minimizes the residual DGD is significantly worse than the performance of a variable-DGD compensator that maximizes the eye opening. This result, which demonstrates the importance of higher-order PMD in determining the penalty, is consistent with [65], where it is shown that a feedback signal provided by a frequency-selective polarimeter is better correlated to the PMD-induced penalty when extracting more values of the polarization dispersion vector over the spectrum of the signal.

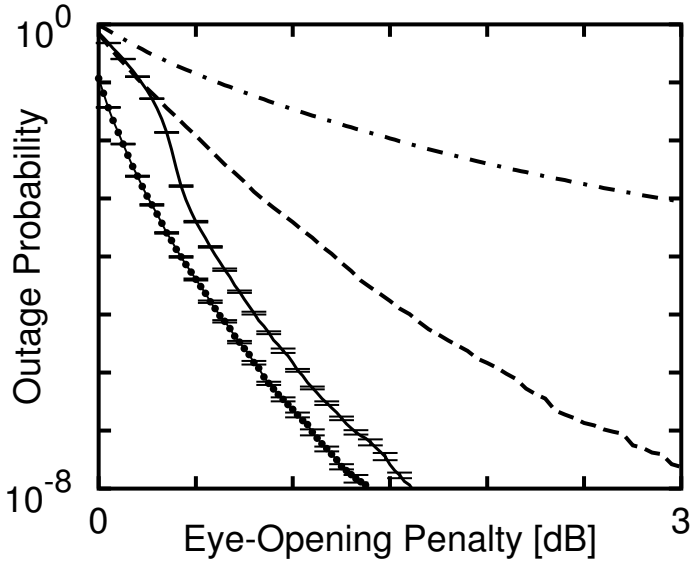

Figure 3.14: Outage probability as a function of the eye-opening penalty margin. The outage probability ($\hat{P}_o$) is the probability that the penalty exceeds the value displayed on the horizontal axis. (i) Dashed-dotted line: Uncompensated case; (ii) Dashed line: Variable-DGD compensator with the compensated DGD minimized at the central frequency of the channel; (iii) Solid line: Fixed-DGD compensator with DGD element equal to 2.5 $\langle|\boldsymbol{\tau}|\rangle$ and maximized eye opening; (iv) Solid-dotted line: Variable-DGD compensator with maximized eye opening. The error bars show the confidence interval for the curves that have at least one bin whose relative error ($\hat{\sigma}_{\hat{P}_o}/\hat{P}_o$) exceeds 10%. For those curves, we show the error bars for one out of three consecutive bins.

The outage probability curve referring to a variable-DGD compensator does not equal 1 at 0 dB because higher-order PMD in the transmission line will chirp the pulses. In most cases, the DGD of the line is small and the variable-DGD element of the compensator can then compress the pulses, which produces an eye-opening at the sampling time that is larger than in the back-to-back case.

Chapter 4

Multicanonical Monte Carlo method for PMD-induced penalty

In this chapter, I briefly review the multicanonical Monte Carlo (MMC) method proposed by Berg and Neuhaus [18], and I describe how my colleagues and I implemented MMC to compute the probability density function (pdf) of the differential group delay (DGD) for PMD emulators. Then, I present results showing the correlation among the histogram bins of the pdf of the DGD that is generated using the MMC method. Finally, I present results with the application of MMC to compute the PMD-induced penalty in uncompensated and single-section compensated system. In particular, I use contours plots to show the regions in the $|\boldsymbol{\tau}|$–$|\boldsymbol{\tau}_\omega|$ plane that are the dominant source of penalties in uncompensated and single-section PMD compensated systems.

4.1 The Multicanonical Monte Carlo method

In statistical physics applications, a conventional canonical simulation calculates expectation values at a fixed temperature T and can, by re-weighting techniques, only be extrapolated to a vicinity of this temperature [66]. In contrast, a single multicanonical

simulation allows one to obtain expectation values over a range of temperatures, which would require many canonical simulations. Hence, the name multicanonical [18], [66]. The multicanonical Monte Carlo method is an iterative method, which in each iteration produces a biased random walk that automatically searches the state space for the important rare events. Within each iteration, the Metropolis algorithm [8] is used to select samples for the random walk based on an estimated pdf of the quantity of interest or control parameter, which is updated from iteration to iteration. Each new sample in the random walk is obtained after a small random perturbation is applied to the previous sample. In each MMC iteration, a histogram of the control parameter is calculated that records how many samples are in each bin. In each iteration, one generates a pre-determined number of samples that can vary from iteration to iteration. Typically, each iteration has several thousand samples. Once the pre-determined number of samples in any iteration has been generated, the histogram of the control parameter is used to update the estimate of the probability of all the bins as in [18], which will be used to bias the following iteration. After some number of iterations, typically 15–50, the number of samples in each bin of the histogram of the control quantity becomes approximately constant over the range of interest, indicating that the estimated pdf of the control quantity is converging to the true pdf.

4.2 MMC implementation to PMD emulators

In the computation of the pdf of the DGD, the state space of the system is determined by the random mode coupling between the birefringent sections in an optical fiber

with PMD, and the control parameter E is the DGD, as in [30]. When applying MMC, the goal is to obtain an approximately equal number of samples in each bin of the histogram of the control quantity. In this dissertation, we compute probabilities by dividing the range of DGD values into discrete bins and constructing a histogram of the values generated by the different random configurations of the fiber sections.

As explained in detail in Chapter 2, the calculations in this dissertation are based on coarse-step PMD emulators consisting of birefringent fiber sections separated by polarization scramblers [37]. We model the fiber using emulators with $N_s = 15$ and $N_s = 80$ birefringent sections. Prior to each section, we use a polarization scrambler to uniformly scatter the polarization dispersion vector on the Poincaré sphere. When polarization scramblers are present, the evolution of the polarization dispersion vector is equivalent to a three-dimensional random walk, and an exact solution [46] is available for the pdf of the DGD that can be compared with the simulations. In unbiased Monte Carlo simulations, the unit matrix $\mathsf{R} = \mathsf{R}_x(\phi)\mathsf{R}_y(\gamma)\mathsf{R}_x(\psi)$ rotates the polarization dispersion vector before each section, such that the rotation angles around the x-axis in the i-th section, ϕ_i and ψ_i, have their pdfs uniformly distributed between $-\pi$ and π, while the cosine of the rotation angle γ_i around y-axis has its pdf uniformly distributed between -1 and 1. Within each MMC iteration, we use the Metropolis algorithm to make a transition from a state k to a state l by making random perturbations $\Delta\phi_i$, $\Delta\gamma_i$, and $\Delta\psi_i$ of the angles ϕ_i, γ_i, and ψ_i in each section, where $\Delta\phi_i$, $\Delta\gamma_i$, and $\Delta\psi_i$ are uniformly distributed in the range $[-\varepsilon\pi, \varepsilon\pi]$. To keep the average acceptance ratio close to 0.5 [10], we choose the coefficient of perturbation $\varepsilon = 0.09$. This perturbation is small, since it does not exceed 10% of

the range of the angles. To obtain the correct statistics in γ_i, since in the coarse step method the cosine of γ_i is uniformly distributed, we accept the perturbation $\Delta\gamma_i$ with probability equal to $\min[1, F(\gamma_i + \Delta\gamma_i)/F(\gamma_i)]$, where $F(\gamma) = 0.5\,(1-\cos^2\gamma)^{1/2}$. When the perturbation is not accepted, we set $\Delta\gamma_i = 0$. The random variable with acceptance probability given by $\min[1, F(\gamma_i + \Delta\gamma_i)/F(\gamma_i)]$ can be implemented by obtaining a random number from a pdf uniformly distributed between 0 and 1, and then accepting the perturbation $\Delta\gamma_i$ if the random number obtained is smaller than $F(\gamma_i + \Delta\gamma_i)/F(\gamma_i)$. To introduce a bias towards large values of the control parameter E, each transition from state k to the state l in the iteration $j+1$ is accepted with probability $P_{\text{accept}}(k \to l) = \min[1, P^j(E_k)/P^j(E_l)]$, and rejected otherwise, where $P^j(E)$ is the estimate of the pdf of DGD obtained after the first j iterations. At the end of each iteration we update $P^j(E)$ using the same recursion algorithm as in [18], so that the number of hits in each bin of the control parameter histogram becomes approximately equal as the iteration number increases.

4.2.1 Summary of the MMC algorithm

In the first iteration we use M_1 samples and set the pdf of the DGD $P^1(E)$ of a PMD emulator with N_s sections as uniform, $P^1(E) = 1/N_b$ (N_b = number of bins). Because every step in the Metropolis algorithm will be accepted with this initial distribution, we more effectively exploit the first iteration by choosing the coefficient of perturbation $\varepsilon = 1$. To update the pdf of the DGD at the end of this iteration we use the recursive equation as in (4.1), which is the same equation used in any other iteration. We then carry out an additional $N-1$ iterations with M_l ($1 < l \leq N$)

samples in each iteration. We note that in general the number of samples in each iteration does not have to be the same. I now present a pseudo-code summary of the algorithm:

Loop over iterations $j = 1$ *to* $N-1$*:*

Loop over fiber realizations (samples) $m = 1$ *to* M_l*:*

(1) start random walk on $\boldsymbol{\phi}$*,* $\boldsymbol{\gamma}$*, and* $\boldsymbol{\psi}$ *with small steps* $\boldsymbol{\Delta\phi}$*,* $\boldsymbol{\Delta\gamma}$*, and* $\boldsymbol{\Delta\psi}$

$$\boldsymbol{\Delta\phi} = \{\Delta\phi_1, \cdots, \Delta\phi_{N_s}\}; \quad \boldsymbol{\Delta\gamma} = \{\Delta\gamma_1, \cdots, \Delta\gamma_{N_s}\}; \quad \boldsymbol{\Delta\psi} = \{\Delta\psi_1, \cdots, \Delta\psi_{N_s}\}$$

(2) compute the provisional value of the DGD (E_{prov}*)*

with the angles $\boldsymbol{\phi} + \boldsymbol{\Delta\phi}$*,* $\boldsymbol{\gamma} + \boldsymbol{\Delta\gamma}$ *and* $\boldsymbol{\psi} + \boldsymbol{\Delta\psi}$*.*

(3) accept provisional step with probability equal to $\min\left[1, P^j(E_m)/P^j(E_{\text{prov}})\right]$

if step accepted: $E_{m+1} = E_{\text{prov}}$

$$\boldsymbol{\phi}_{m+1} = \boldsymbol{\phi}_m + \boldsymbol{\Delta\phi}; \qquad \boldsymbol{\gamma}_{m+1} = \boldsymbol{\gamma}_m + \boldsymbol{\Delta\gamma}; \qquad \boldsymbol{\psi}_{m+1} = \boldsymbol{\psi}_m + \boldsymbol{\Delta\psi}$$

if step rejected: $E_{m+1} = E_m$

$$\boldsymbol{\phi}_{m+1} = \boldsymbol{\phi}_m; \qquad \boldsymbol{\gamma}_{m+1} = \boldsymbol{\gamma}_m; \qquad \boldsymbol{\psi}_{m+1} = \boldsymbol{\psi}_m$$

(4) increment the histogram of E with the sample E_{m+1}

End of loop over fiber realizations

update the pdf of the DGD $P^{j+1}(E)$

restart histogram

go to next iteration j

End

To update $P^j(E)$ at the end of each iteration j we use the recursive equation [18],

$$P_{k+1}^{j+1} = P_k^{j+1} \frac{P_{k+1}^j}{P_k^j} \left(\frac{H_{k+1}^j}{H_k^j} \right)^{\hat{g}_k^j}, \tag{4.1}$$

where $\hat{g}_k^j$, the relative statistical significance of the k-th bin in the j-th iteration, is defined as

$$\hat{g}_k^j = \frac{g_k^j}{\sum_{l=1}^{j} g_k^l}, \quad \text{with} \quad g_k^j = \frac{H_{k+1}^j H_k^j}{H_{k+1}^j + H_k^j}. \tag{4.2}$$

If $H_{k+1}^j + H_k^j = 0$ in a given iteration, then the k-th bin has no statistical significance in this iteration. Therefore, we set $g_k^j = 0$ in that iteration. The statistical significance, $0 \leq \hat{g}_k^j \leq 1$, depends on both previous bins and previous iterations, inducing a significant correlation among P_k^j. Finally, the P_k^j are normalized so that $\sum_{k=1}^{N_b} P_k^j = 1$, where N_b is the number of bins. MMC is an extension of the Metropolis algorithm [8], where the acceptance rule accepts all the transitions to states with lower probabilities, but rejects part of the more likely transitions to states with higher probabilities. As the number of iterations increases, the histogram of the number of hits in each bin will asymptotically converge to a uniform distribution ($H_{k+1}^j / H_k^j \to 1$), and the relative statistical significance will asymptotically converge to zero ($\hat{g}_k^j \to 0$). Consequently, P^{j+1} will asymptotically converge to the true probability of the control parameter. Equations (4.1) and (4.2) were derived by Berg and Neuhaus [18] assuming that the probability distribution is exponentially distributed with a slowly varying exponent that is a function of the control quantity (the temperature in their case and DGD in ours). This assumption is valid in a large number of problems in optical fiber communications, including the pdf of the DGD in fibers with an arbitrary number

of sections [30], [67]. The recursions in (4.1) and (4.2) were derived by applying a quasi-linear approximation to the logarithm of the pdf in addition to a method for combining the information in the current histogram with that of previous iterations according to their relative statistical significance [18], [30].

4.3 Correlations

The goal of any scheme for biasing Monte Carlo simulations, including MMC, is to reduce the variance of the quantities of interest. The MMC uses a set of systematic procedures to reduce the variance, which are highly nonlinear as well as iterative and have the effect of inducing a complex web of correlations from sample to sample in each iteration and between iterations. These, in turn, induce bin-to-bin correlations in the histograms of the pdfs. It is easy to see that the use of (4.1) and (4.2) generates correlated estimates for the P_k^j, although this procedure significantly reduces the variance [18]. In this section, I illustrate this correlation by showing results obtained when my colleagues and I applied MMC to compute the pdf of the DGD for a PMD emulator with 80 sections.

We computed the correlation coefficient between bin i and each bin j ($1 \leq j \leq 80$) in the histogram of the normalized DGD by doing a statistical analysis on an ensemble of many independent standard MMC simulations. The normalized DGD, $|\boldsymbol{\tau}| / \langle |\boldsymbol{\tau}| \rangle$, is defined as the DGD divided by its expected value, which is 30 ps in this case. Suppose that on the l-th MMC simulation, we have P_i^l as the probability of the i-th bin and suppose that the average over all L MMC simulations is $\overline{P_i}$. Then, we define

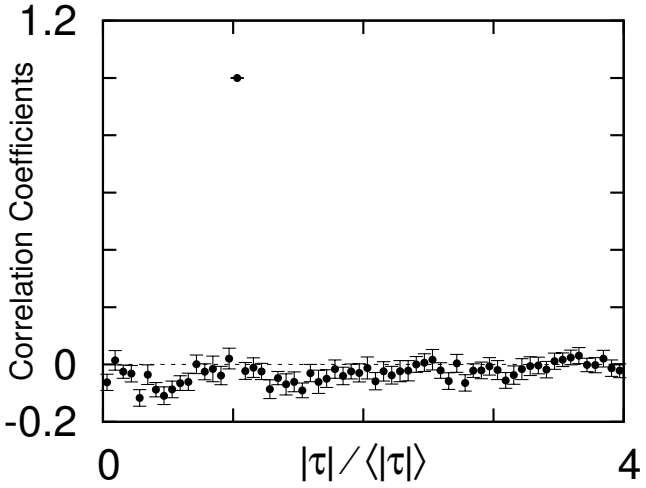

Figure 4.1: Correlation coefficients between bin i and bin j ($1 \leq j \leq 80$) for the 80-section emulator, where the bin i corresponds to $\text{DGD}_i = 30$ ps (1 $\times$ mean DGD). The correlation coefficients are computed using 32 standard MMC simulations. Each standard MMC simulation consists of 30 MMC iterations with $8,000$ samples.

a normalized correlation between bin i and bin j as

$$C(i,j) = \frac{1}{L-1} \sum_{l=1}^{L} \frac{(P_i^l - \overline{P_i})(P_j^l - \overline{P_j})}{\sigma_{P_i} \sigma_{P_j}} \tag{4.3}$$

where σ_{P_i} and σ_{P_j} are the standard deviation of P_i and P_j, respectively. The normalized correlation defined in (4.3) is known as Pearson's correlation coefficient [68].

The values for $C(i,j)$ generated by (4.3) will range from -1 to 1. A value of $+1$ indicates a perfect correlation between the random variables. While a value of -1 indicates a perfect anti-correlation between the random variables. A value of zero indicates no correlation between the random variables.

In Figs. 4.1–4.3, I show the correlation coefficients between bin i and bin j, $1 \leq \ j \ \leq 80$, for the DGD in the bin i, DGD_i, equal to 30 ps, 45 ps, and 75 ps,

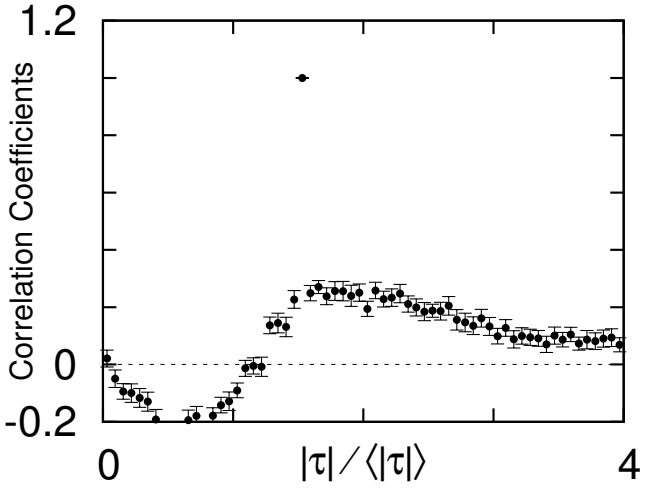

Figure 4.2: Correlation coefficients between bin i and bin j ($1 \leq j \leq 80$) for the 80-section emulator, where the bin i corresponds to $\mathrm{DGD}_i = 45$ ps (1.5 $\times$ mean DGD). The correlation coefficients are computed using 32 standard MMC simulations. Each standard MMC simulation consists of 30 MMC iterations with $8,000$ samples.

respectively. In this case, we used a PMD emulator with 80 sections and the mean DGD is equal to 30 ps. To compute each value of $C(i,j)$ we used $L = 32$ MMC simulations. We computed sample mean $\overline{C(i,j)}$ and standard deviation $\sigma_{C(i,j)}$ using 32 samples of $C(i,j)$. The values of the standard deviation for the results shown in Figs. 4.1–4.3 are in the range from 1.84×10^{-2} to 3.91×10^{-2}. Note that DGD_i equal to 75 ps represents a case in the tail of the pdf of the DGD, where the unbiased Monte Carlo method has very low probability of generating samples, by contrast to a biased Monte Carlo method such as MMC. The results show that the correlations are not significant until we use a large value for DGD_i compared to the mean DGD. However, these values of DGD_i are precisely the values of greatest interest.

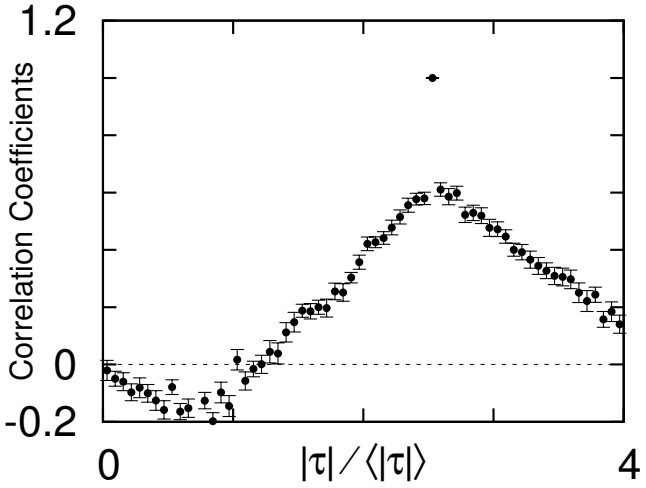

Figure 4.3: Correlation coefficients between bin i and bin j ($1 \leq j \leq 80$) for the 80-section emulator, where the bin i corresponds to $\mathrm{DGD}_i = 75$ ps (2.5 $\times$ mean DGD). The correlation coefficients are computed using 32 standard MMC simulations. Each standard MMC simulation consists of 30 MMC iterations with $8,000$ samples.

4.4 MMC computation of PMD-induced penalty in uncompensated and single-section compensated systems

In this section, we apply the MMC algorithm to compute PMD-induced penalties in a 10 Gbit/s NRZ system using 50 MMC iterations with 2,000 samples each. In Fig. 4.4, I show the outage probability as a function of the eye-opening penalty. The results obtained using the samples in the final iteration of the MMC simulation (dashed and solid lines) are in excellent agreement with the ones obtained using importance sampling (open circles and squares). In this section, I present results in which my colleagues and I used the results computed with importance sampling to validate the

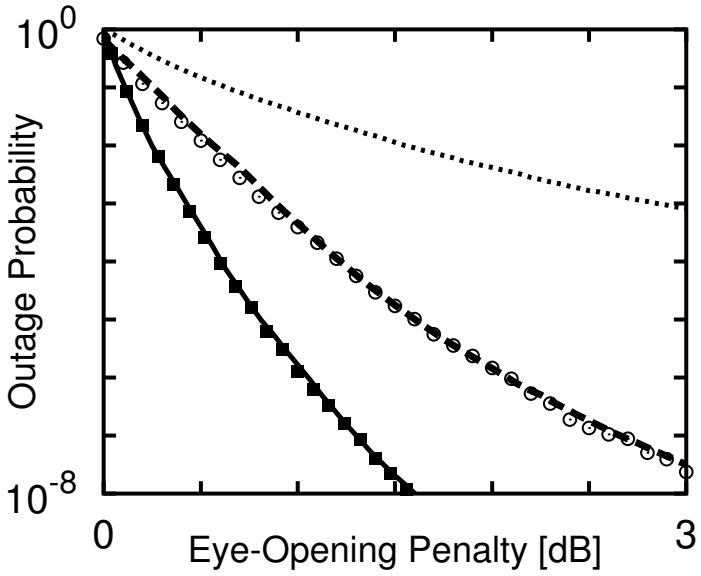

Figure 4.4: Outage probability as a function of the eye-opening penalty. (i) Dotted line: Uncompensated system with a mean DGD of 30 ps. (ii) Dashed line and (iii) Open circles: Results for a variable-DGD compensator, obtained using MMC and IS, respectively, for a system with mean DGD of 30 ps. (iv) Solid line and (v) Squares: Results for an uncompensated system with mean DGD of 15 ps, obtained using MMC and IS, respectively.

results obtained with MMC. The use of importance sampling to compute penalties in PMD single-section compensated systems was already validated with a large number of standard Monte Carlo simulations in Chapter 3 of this dissertation and by Lima *at al.* [3], [50]. Therefore, the results computed with importance sampling can be used to validate the results computed with MMC. Our goal here is to show the applicability of MMC to accurately compute PMD-induced penalties in uncompensated and single-section PMD compensated systems.

In Fig. 4.5, I show contours (dotted lines) of the joint pdf of the magnitude of the uncompensated normalized first- and second-order PMD, $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$, computed

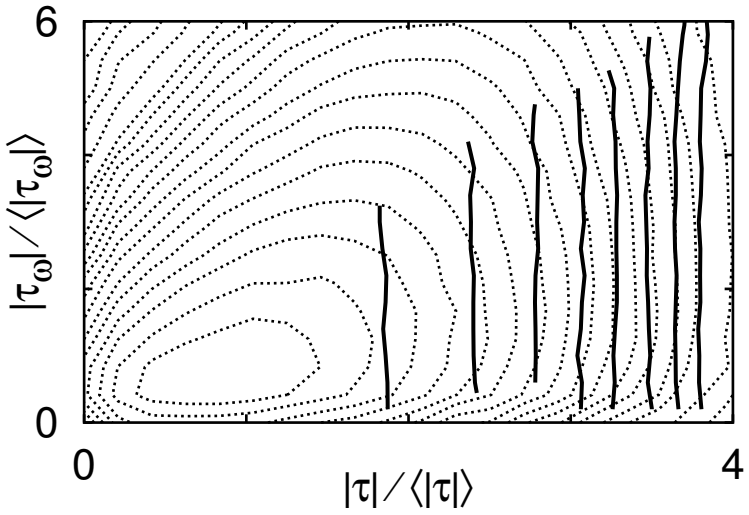

Figure 4.5: Uncompensated system with a mean DGD of 15 ps. The dotted lines show the contour plots of the joint pdf of the normalized $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$, obtained using IS. The solid lines show the average eye-opening penalty given a value of $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$, obtained using MMC. The contours of joint pdf from the bottom to the top of the plot, are at 3×10^{-n} , $n = 1, \cdots, 7$ and 10^{-m}, $m = 1, \cdots, 11$. The penalty contours in dB from the left to the right of the plot, are at 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6.

using importance sampling, as in [23]. I also show contours for the eye-opening penalty (solid lines) of an uncompensated system with a mean DGD, $\langle|\boldsymbol{\tau}|\rangle$, of 15 ps. The penalty contours were produced using the same samples we generated using the MMC method in the computation of the outage probability shown in Fig. 4.4. The fiber realizations obtained using the MMC method are all located in the region of the $|\boldsymbol{\tau}|$–$|\boldsymbol{\tau}_\omega|$ plane that corresponds to the large DGD values that have the highest probability of occurrence, which is the region that is the dominant source of penalties in uncompensated systems. These results are in agreement with what we have inferred from Fig. 3.10 in Section 3.7, where we stated that the contour plots in the region dominating the penalty were approximately parallel to the second-order

PMD axis, indicating that first-order PMD is the dominant cause of penalty in this uncompensated system.

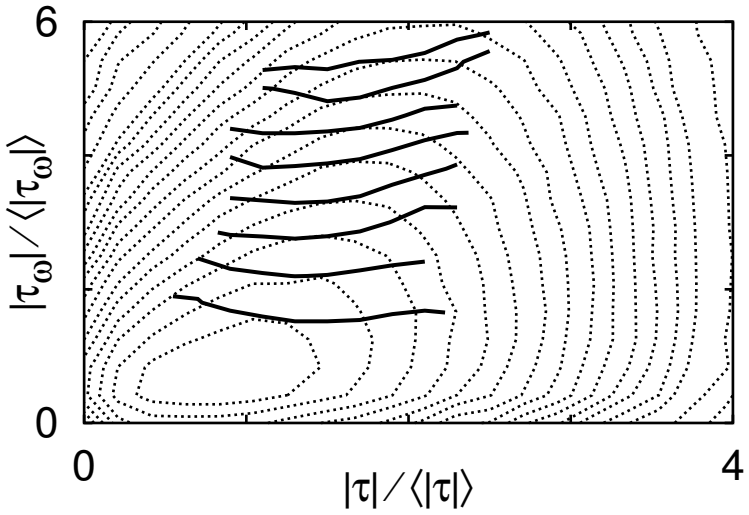

Figure 4.6: Same set of curves of Fig. 4.5 for a compensated system with a variable-DGD compensator. The penalty contours in dB from the bottom to the top of the plot, are at 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6.

In Fig. 4.6, I show contours of the eye-opening penalty (solid lines) of a system with $\langle|\boldsymbol{\tau}|\rangle$=30 ps and a variable-DGD compensator that was programmed to minimize the residual DGD at the central frequency of the channel after compensation. As in Fig. 4.5, the penalty contours were produced using the same samples we generated using the MMC method in the computation of the outage probability shown in Fig. 4.4. In contrast to the results shown in Fig. 4.5, the fiber realizations obtained using the MMC method in Fig. 4.6 are located in the region of the $|\boldsymbol{\tau}|$–$|\boldsymbol{\tau}_\omega|$ plane where $|\boldsymbol{\tau}_\omega|$ is large and the DGD is close to its average, corresponding to the region in the $|\boldsymbol{\tau}|$–$|\boldsymbol{\tau}_\omega|$ plane that is the dominant source of penalties in this compensated system. These results are also in agreement with what we have inferred from Fig. 3.13

in Section 3.7, where we stated that the contour plots in the region dominating the penalty were approximately parallel to the DGD axis, indicating that the penalty is nearly independent of DGD. In Figs. 4.5 and 4.6, the samples obtained using the MMC method are automatically biased towards the specific region of the $|\boldsymbol{\tau}|$–$|\boldsymbol{\tau}_\omega|$ plane that dominates the penalty, *i.e.*, the region where the corresponding penalty level curve intersects the contour of the joint pdf of $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$ with the highest probability.

We did not compute the confidence interval for the results showed in this section. In Chapter 5, I describe a procedure that I developed with the collaboration of my colleagues to efficiently estimate the statistical errors in MMC simulations. Finally, in Chapter 6, I present results obtained with MMC for multi-section PMD compensated systems containing confidence intervals computed with the method described in Chapter 5.

Chapter 5

Estimation of Errors in MMC simulations

In this chapter, I explain why a new error estimation procedure is needed for multicanonical Monte Carlo simulations, and I then present the transition matrix method that I developed in collaboration with my colleagues to efficiently estimate the error in MMC. Finally, I present the validation and application of this method.

5.1 Why a new error estimation procedure ?

Since MMC is a Monte Carlo technique, it is subject to statistical errors, and it is essential to determine their magnitude. In Chapter 3, I showed how to compute errors when using importance sampling (see also [24]). In this chapter, I will show how one can efficiently estimate errors in MMC simulations using a transition matrix method that my colleagues and I developed. In practice, users of Monte Carlo methods often avoid making detailed error estimates. For example, when using an standard, unbiased Monte Carlo simulation to calculate the pdf of a quantity such as the DGD, the number of samples in each bin of the pdf's histogram is independent. Hence, when the histogram is smooth, one can infer that the error is acceptably low. This procedure

is not reliable with MMC simulations because, as I showed in Chapter 4, the MMC algorithm induces a high degree of correlation from bin to bin. While it is always best to estimate error with any Monte Carlo method, it is particularly important in MMC simulations, due to the presence of large sample-to-sample correlations on the tails of the distributions.

As I explained in Chapter 4, the existence of correlations in the samples generated with the MMC method makes calculating the errors in MMC simulations significantly more difficult than in standard Monte Carlo simulations. Also, due to the correlations, one cannot apply to MMC standard error analysis that are traditionally used for simulations with uncorrelated samples. For the same reason, one cannot determine the contribution of the variance from each iteration using standard error propagation methods as in the case with importance sampling simulations. Thus, the MMC variance cannot be estimated by applying a standard error analysis to a single MMC simulation. One can in principle run many independent MMC simulations in order to estimate the error by using the standard sample variance formula [31] on the ensemble of MMC simulations. However, estimating the error of the pdf of the quantity of interest by running many independent MMC simulations is computationally costly and in many cases not feasible. One can overcome this problem with the transition matrix method that I have developed with the collaboration of my colleagues. The transition matrix method is an efficient numerical method to estimate statistical errors in the pdfs computed using MMC. In our method, we use the estimated transition probability matrix to rapidly generate an ensemble of hundreds of pseudo-MMC simulations, which allows one to estimate errors from only one standard MMC simulation. The

transition probability matrix, which is computed from a single, standard MMC simulation, contains all the probabilities that a transition occurs from any bin of the histogram of the quantity of interest to any other bin after a step (or perturbation) in the MMC random walk. The pseudo-MMC simulations are then made using the computed transition matrix instead of running full simulations. Each pseudo-MMC simulation must be made with the same number of samples per iteration and the same number of iterations as in the original standard MMC simulation. Once an ensemble of pseudo-MMC simulations has been calculated, one can use standard procedures to estimate the error. Since the transition matrix that is used in the pseudo-MMC simulations has its own statistical error, it might seem strange at first that it can be used as the basis from which to estimate the error in the MMC simulations. However, bootstrap theory assures us that such is the case [32]. Intuitively, the variation of any statistical quantity among the members of an ensemble of pseudo-MMC simulations is expected to be the same as the variation among members of an ensemble of standard MMC simulations because the simulations are of the same type and the same size.

To illustrate the transition matrix method, we calculated the pdf of DGD due to PMD and the associated confidence interval for two types of PMD emulators [44]. We validated our method by comparison to the results obtained by using a large ensemble of standard MMC simulations. We tested our method by applying it to PMD emulators because it was the first random phenomenon in optical fiber communication to which MMC was applied [30] and has become essential for testing biasing Monte Carlo methods. Moreover, it is computationally feasible to validate the proposed

method with a large ensemble of standard MMC simulations. That is not the case for most other problems, *e.g.*, the error rate due to optical noise [69] and the residual penalty in PMD-compensated systems [19].

5.2 New error estimation procedure

In this dissertation, I introduce an efficient numerical procedure that we refer to as the transition matrix method, to compute statistical errors in MMC simulations that properly accounts for the contributions of all MMC iterations. The transition matrix method is a bootstrap resampling method [32], [70] that uses a computed estimate of the probability of a transition from bin i to bin j of the histogram of the DGD. In a bootstrap method, one estimates a complex statistical quantity by extracting samples from an unknown distribution and computing the statistical quantity. In the case of computing the pdf of the DGD in PMD emulators, the complex statistical quantity is the probability of each bin in the histogram of the DGD, the pseudo-samples are the DGD values obtained in the pseudo-MMC simulations, and the unknown distribution is the true transition matrix. One then repeatedly and independently draws an ensemble of pseudo-samples with replacement from each original sample and computes the statistical quantity of interest using the same procedure by which the statistical quantity was first estimated. We describe the Bootstrap method with more detail in Section 5.2.2. One can then estimate the variance of the quantity of interest from these pseudo-samples using standard techniques. The bootstrap method is useful when it is computationally far more rapid to resample the original

set of samples than to generate new samples, allowing for an efficient estimate of the variance.

5.2.1 The transition matrix method

In this section, I explain the transition matrix method in the context of computing errors in the pdf of the DGD for PMD emulators. The transition matrix method has two parts. In the first part, one obtains an estimate of the pdf of the DGD and an estimate of the one-step transition probability matrix $\mathbf{\Pi}$. To do so, one runs a standard MMC simulation, as described in Section 4.2.1. At the same time, one computes an estimate of the transition probability $\pi_{i,j}$, which is the probability that a sample in the bin i will move to the bin j after a single step in the MMC algorithm. *I stress that a transition attempt must be recorded whether or not it is accepted by the Metropolis algorithm after the fiber undergoes a random perturbation.* The transition matrix is a matrix that contains the probability that a transition will take place from one bin to any other bin when applying a random perturbation. It is independent of the procedure for rejecting or accepting samples, which is how the biasing is implemented in the MMC method. An estimate of the transition matrix that is statistically as accurate as the estimate of the pdf using MMC can be obtained by considering all the transitions that were attempted in the MMC ensemble. One uses this information to build a $N_b \times N_b$ one-step transition probability matrix, where N_b is the number of bins in the histogram of the pdf. The transition matrix $\mathbf{\Pi}$ consists of elements $\pi_{i,j}$, where the sum of the row elements of $\mathbf{\Pi}$ equals 1. The elements $\pi_{i,j}$

are computed as

$$\pi_{i,j} = \frac{\sum_{m=1}^{M_t-1} I_i(E_m) I_j(E_{m+1})}{\sum_{m=1}^{M_t-1} I_i(E_m)}, \quad \text{if} \quad \sum_{m=1}^{M_t-1} I_i(E_m) \neq 0, \tag{5.1}$$

and $\pi_{i,j} = 0$, otherwise. In (5.1), M_t is the total number of samples in the MMC simulation and E_m is the m-th DGD sample. The indicator function $I_i(E)$ is chosen to compute the probability of having a DGD sample inside the bin i of the histogram. Thus, $I_i(E)$ is defined as 1 inside the DGD range of the bin i, otherwise $I_i(E)$ is defined as 0.

In the second part of the procedure, one carries out a series of MMC simulations, that my colleagues and I refer to as pseudo-MMC simulations. In each step, if one starts for example in bin i of the histogram, one picks a new provisional bin j using a procedure to sample from the pdf π_i, where $\pi_i(j) = \pi_{i,j}$. This procedure is explained in detail below. One then accepts or rejects this provisional transition using the same criteria as in full, standard MMC simulations, and the number of samples in the bins of histogram is updated accordingly. Thus, one is using the transition matrix $\mathbf{\Pi}$ to emulate the random changes in the DGD that result from the perturbations $\Delta\phi_i$, $\Delta\gamma_i$, and $\Delta\psi_i$ that were used in the original standard MMC simulation. In all other respects, each pseudo-MMC simulation is like the standard MMC simulation. In particular, the metric for accepting or rejecting a step, the number of samples per iteration, and the number of iterations must be kept the same. It is possible to carry out hundreds of these pseudo-MMC simulations in a fraction of the computer time that it takes to carry out a single standard MMC simulation. This procedure

requires us to hold the entire transition matrix in memory, which could in principle be memory-intensive, although this issue did not arise in any of the problems that we considered. This procedure will be useful when evaluating a transition using the transition matrix requires far less computational time than calculating a transition using the underlying physics. That will typically be the case and was certainly the case for the problems that we considered.

An estimate of the pdf of the DGD is obtained in the final iteration of each pseudo-MMC simulation. Since the estimates of the probability in a given bin in the different pseudo-MMC simulations are independent, one may apply the standard formula for computation of the variance $\sigma^2_{p^\star_i}$ of the i-th bin

$$\sigma^2_{p^\star_i} = \frac{1}{(B-1)} \sum_{b=1}^{B} \left(p^\star_{i,b} - \overline{p^\star_i} \right)^2, \quad \text{with} \quad \overline{p^\star_i} = \frac{1}{B} \sum_{b=1}^{B} p^\star_{i,b}, \tag{5.2}$$

where $p^\star_{i,b}$ is the probability of the i-th bin in the histogram of the DGD obtained in the b-th pseudo-MMC simulation and B is the total number of pseudo-MMC simulations. Thus, $\sigma_{p^\star_i}$ is an estimate of the error in the i-th bin in the histogram of the DGD obtained in a single MMC simulation.

I now illustrate the details of how we choose the provisional transition from bin i to bin j with the following pseudo-code:

bin DGD of current sample = i

use random number to generate x from a uniform pdf between 0 *and* 1*:* $x \leftarrow U[0,1]$

for j=1 to N_b

if $(x < \pi^{\mathrm{cdf}}_{i,j})$

new bin = j

break

end if

end for

current bin = *new bin*

where $\pi_{i,j}^{\text{cdf}} = \sum_{m=1}^{j} \pi_{i,m}$ is the cumulative transition probability. This procedure is used to sample from the pdf π_i, where $\pi_i(j) = \pi_{i,j}$, and with $\pi_{i,j}$ defined as the probability that a sample in the bin i will move to the bin j.

5.2.2 Bootstrap method

Efron's bootstrap [32] is a well-known general purpose technique for obtaining statistical estimates without making *a priori* assumptions about the distribution of the data. Suppose one draws a random vector $\mathbf{x} = (x_1, x_2, ..., x_n)$ with n samples from an unknown probability distribution F and one wishes to estimate the error in a parameter of interest $\hat{\theta} = s(\mathbf{x})$. Since there is only one sample of $\hat{\theta}$, one cannot use the sample standard deviation formula to compute the error. However, one can use the random vector $\mathbf{x}$ to determine an empirical distribution $\hat{F}$ from F (unknown distribution). Then, one can generate bootstrap samples from $\hat{F}$, $\mathbf{x}^\star = (x_1^\star, x_2^\star, ..., x_n^\star)$, to obtain $\hat{\theta}^\star = s(\mathbf{x}^\star)$ by drawing n samples with replacement from $\mathbf{x}$. The quantity $s(\mathbf{x}^\star)$ is the result of applying the same function $s(.)$ to $\mathbf{x}^\star$ as was applied to $\mathbf{x}$. For example, if $s(\mathbf{x})$ is the median of $\mathbf{x}$, then $s(\mathbf{x}^\star)$ is the median of the bootstrap resampled data set. The star notation indicates that $\mathbf{x}^\star$ is not the actual data set $\mathbf{x}$, but

rather a resampled version of $\mathbf{x}$ obtained from the estimated distribution $\hat{F}$. Note that one can generate as many bootstrap samples $\mathbf{x}^\star$ as one needs, and then generate independent bootstrap sample estimates of $\hat{\theta}$, $\hat{\theta}_1^\star = s(\mathbf{x}_1^\star)$, ... , $\hat{\theta}_B^\star = s(\mathbf{x}_B^\star)$, where B is the total number of bootstrap samples. Then, one can estimate the error in $\hat{\theta}$ using the standard deviation formula on the bootstrap samples $\hat{\theta}^\star$.

The transition matrix method that I am describing in this dissertation is related to the bootstrap resampling method as follows:

1) $\hat{F}$ is an estimate of the transition matrix obtained from a single standard MMC simulation;

2) $\mathbf{x}_1^\star$, ..., $\mathbf{x}_B^\star$, are the collection of samples that is obtained from the ensemble of pseudo-MMC simulations. We note that $\mathbf{x}_b^\star$ should be computed using the exact same number of iterations and the exact same number of samples per iteration as in the original standard MMC simulation;

3) Each $\hat{\theta}_b^\star$, where $b = 1, 2, ..., B$, is a value for the probability $p_k^\star$ of the k-th bin of the histogram of the DGD obtained from each of the pseudo-MMC simulations;

4) Given that one has B independent $p_k^\star$, one can obtain an error estimate for each bin in the estimated pdf of the DGD using the traditional sample standard deviation formula [31], [32]

$$\sigma_{\hat{\theta}^\star} = \left[\frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\theta}_b^\star - \overline{\hat{\theta}^\star} \right)^2 \right]^{1/2}, \tag{5.3}$$

where,

$$\overline{\hat{\theta}^\star} = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}_b^\star. \tag{5.4}$$

5.2.3 Assessing the error in the MMC error estimation

The estimate of the MMC variance also has an error, which depends on the number of samples in a single standard MMC simulation and on the number of pseudo-MMC simulations (bootstrap samples) [71]. In the work that I present in this dissertation, the error due to the bootstrap resampling is minimized by using $1,000$ bootstrap pseudo-MMC simulations. Therefore, the residual error is due to the finite number of samples used to estimate both the pdf of the DGD and the transition matrix in the single standard MMC simulation, *i.e.*, in the first part of the transition matrix method. Thus, there is a variability in the estimate of the MMC variance due to the variability of the transition matrix $\hat{\mathbf{\Pi}}$ as an estimate of the true transition matrix $\mathbf{\Pi}$.

To estimate the error in the estimate of the MMC variance, we apply a procedure known in the literature as *bootstrapping the bootstrap* or *iterated bootstrap* [72]. The procedure in based on the principle that if the bootstrap can estimate errors in one statistical parameter using $\hat{\mathbf{\Pi}}$, one can also use bootstrap to check the uncertainty in the error estimate using bootstrap resampled transition matrices $\hat{\mathbf{\Pi}}^{\star}$. The procedure consists of:

1) Running one standard MMC simulation;

2) Generating $N_B = 100$ pseudo-MMC simulations and computing transition matrices for each of the pseudo-MMC simulation. Therefore, we obtain N_B transition matrices that we call pseudo-transition matrices $\hat{\mathbf{\Pi}}^{\star}_B$;

3) For each pseudo-transition matrix $\hat{\mathbf{\Pi}}^{\star}_B$ we calculate $N_B = 100$ pseudo-MMC simulations (N_B values for the probability of any given bin of the estimated pdf of the DGD,

$p^{\star\star}$). The double star notation indicates quantities computed with bootstrap resampling from a pseudo-transition matrix. We then estimate the error for the probability of any given bin in the estimated pdf of the DGD, $\sigma_{p^{\star\star}}$, for each pseudo-transition matrix;

4) Since we have $N_B = 100$ pseudo-transition matrices, we repeat step 3 N_B times and obtain N_B values for $\sigma_{p^{\star\star}}$. Then, we compute the double bootstrap confidence interval $\Delta p^{\star\star}$ of the relative variation of the error of p (statistical error in p, where p is the probability of any given bin in the estimated pdf of the DGD computed using a single standard MMC simulation):

$$\Delta p^{\star\star} = \left[\frac{\overline{\sigma_{p^{\star\star}}} - \sigma_{\left(\sigma_{p^{\star\star}}\right)}}{p},\ \frac{\overline{\sigma_{p^{\star\star}}} + \sigma_{\left(\sigma_{p^{\star\star}}\right)}}{p} \right], \tag{5.5}$$

where,

$$\sigma_{\left(\sigma_{p^{\star\star}}\right)} = \left[\frac{1}{N_B - 1} \sum_{n=1}^{N_B} \left(\sigma_{p^{\star\star}}^{(n)} - \overline{\sigma_{p^{\star\star}}} \right)^2 \right]^{1/2}, \tag{5.6}$$

and

$$\overline{\sigma_{p^{\star\star}}} = \frac{1}{N_B} \sum_{n=1}^{N_B} \sigma_{p^{\star\star}}^{(n)}. \tag{5.7}$$

In (5.6) and (5.7), $\sigma_{p^{\star\star}}^{(n)}$ is the standard deviation of $p^{\star\star}$ computed using the n-th pseudo-transition matrix.

In Fig. 5.1, I show the relative variation of $p^{\star\star}$ and its confidence interval $\Delta p^{\star\star}$ for a PMD emulator with 15 sections. My colleagues and I used 14 MMC iterations with $4,000$ samples each (total of $56,000$ samples). The confidence interval of the relative variation is defined in (5.5). We used a total of 80 evenly-spaced bins where

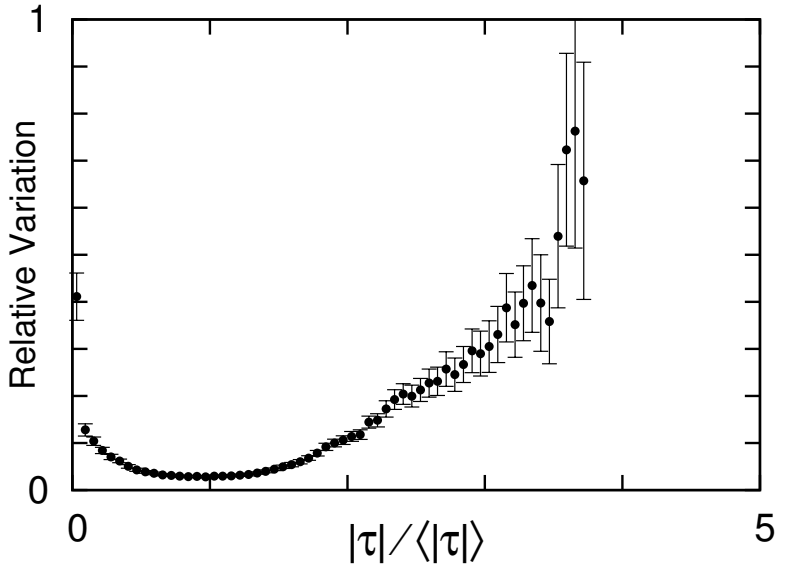

Figure 5.1: Relative variation ($\hat{\sigma}_{\hat{P}_{\mathrm{DGD}}}/\hat{P}_{\mathrm{DGD}}$) of the pdf of the normalized DGD, $|\boldsymbol{\tau}|/\langle|\boldsymbol{\tau}|\rangle$, for the 15-section PMD emulator using 14 MMC iterations with $4,000$ samples. The confidence interval is given by (5.5) when we compute an ensemble of standard deviations using bootstrap resampling for each of the 100 pseudo-transition matrices.

we set the maximum value for the normalized DGD as five times the mean DGD. We also use the same number for bins for all the figures shown in this chapter. As expected, we observed that the error in the estimate of the MMC variance is large when the MMC variance is also large. The confidence interval $\Delta p^{\star\star}$ is between $(2.73\times10^{-2}, 3.19\times10^{-2})$ and $(3.61\times10^{-1}, 4.62\times10^{-1})$ for $|\boldsymbol{\tau}|/\langle|\boldsymbol{\tau}|\rangle < 2$. It increases to $(2.68\times10^{-1}, 4.48\times10^{-1})$ when $|\boldsymbol{\tau}|/\langle|\boldsymbol{\tau}|\rangle = 3$ and to $(4.05\times10^{-1}, 9.09\times10^{-1})$ at the largest value of $|\boldsymbol{\tau}|/\langle|\boldsymbol{\tau}|\rangle$. We concluded that the estimate of the relative variation of the probability of a bin is a good estimate of its own accuracy. This result is similar to what is observed with the standard analysis of standard Monte Carlo simulations [31]. Intuitively, one expects the relative error and the error in the

estimated error to be closely related because both are drawn from the same sample space.

In Fig. 5.1, we also observe that the relative variation increases with the DGD for values larger than the mean DGD, especially in the tail of the pdf. This phenomenon occurs because the regions in the configuration space that contribute to the tail of the pdf of the DGD are only explored by the MMC algorithm after several iterations. As the number of iterations increases, the MMC algorithm allows the exploration of less probable regions of the configuration space. Because less probable regions are explored in the last iterations, there will be a significantly smaller number of hits in the regions that contribute to the tail of the pdf of the DGD. As a consequence, the relative variation will increase as the DGD increases.

5.3 Application and validation

We estimated the pdf of the normalized DGD ($\hat{P}_{\mathrm{DGD}}$) and its associated confidence interval $\Delta\hat{P}_{\mathrm{DGD}}$ for PMD emulators comprised of 15 and 80 birefringent fiber sections with polarization scramblers at the beginning of each section. The normalized DGD, $|\boldsymbol{\tau}| / \langle |\boldsymbol{\tau}| \rangle$, is defined as the DGD divided by its expected value, which is equal 30 ps. We used 14 MMC iterations with $4,000$ samples each to compute the pdf of the normalized DGD when we used a 15-section emulator and 30 MMC iterations with $8,000$ samples each when we used an 80-section PMD emulator.

We monitored the accuracy of our computation by calculating the relative variation of the pdf of the normalized DGD. The relative variation is defined as the ratio

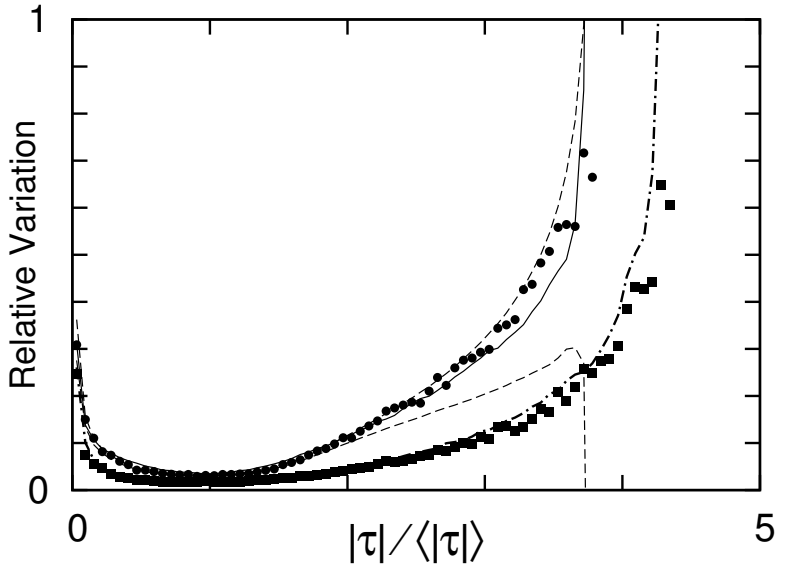

Figure 5.2: Relative variation ($\hat{\sigma}_{\hat{P}_{\rm DGD}}/\hat{P}_{\rm DGD}$) of the pdf of the normalized DGD, $|\boldsymbol{\tau}|/\langle|\boldsymbol{\tau}|\rangle$. (i) Circles: Transition matrix method based on a single standard MMC simulation for the 15-section PMD emulator; (ii) Solid: 10^3 standard MMC simulations for the 15-section emulator; (iii) Dashed: Confidence interval of the relative variation of the error estimated using the transition matrix method for the 15-section PMD emulator; (iv) Squares: Transition matrix method based on a single standard MMC simulation for the 80-section PMD emulator; (v) Dot-dashed: 10^3 standard MMC simulations for the 80-section PMD emulator.

between the standard deviation of the pdf of the normalized DGD and the pdf of the normalized DGD ($\hat{\sigma}_{\hat{P}_{\rm DGD}}/\hat{P}_{\rm DGD}$). In Fig. 5.2, I show the relative variation when we used PMD emulators with 15 and with 80 birefringent sections. The symbols show the relative variation when we applied the procedure that I described in Section 5.2 with $1,000$ pseudo-MMC simulations based on a single standard MMC simulation and the transition matrix method, while the solid and the dot-dashed lines show the relative variation when we used $1,000$ standard MMC simulations. The circles and

the solid line show the results for a 15-section PMD emulator, while the squares and dot-dashed line show the results when we used an 80-section PMD emulator. As expected, the result from an ensemble of pseudo-MMC simulations shows a systematic deviation from the result from an ensemble of standard MMC simulations for both emulators. The systematic deviation changes depending on which standard MMC simulation is used to generate the pseudo ensemble.

In Fig. 5.2, the two dashed lines show the confidence interval of the relative variation with the 15-section PMD emulator computed using the transition matrix method, *i.e.*, the confidence interval for the results that are shown with the circles. The confidence interval $\Delta p^{\star\star}$ is between $(3.04\times10^{-2}, 3.28\times10^{-2})$ and $(2.76\times10^{-1}, 3.62\times10^{-1})$ for $|\boldsymbol{\tau}|/\langle|\boldsymbol{\tau}|\rangle < 2$. It increases to $(2.39\times10^{-1}, 4.31\times10^{-1})$ when $|\boldsymbol{\tau}|/\langle|\boldsymbol{\tau}|\rangle = 3$ and to $(2.69\times10^{-1}, 9.88\times10^{-1})$ at the largest value of $|\boldsymbol{\tau}|/\langle|\boldsymbol{\tau}|\rangle$. While the relative variation that is computed using the transition matrix method from a single MMC simulation will vary from one standard MMC simulation to another, the results obtained from different standard MMC simulations are likely to be inside this confidence interval with a well-defined probability. The confidence interval of the relative variation was obtained using a procedure similar to the one discussed in the Section 5.2.3, except that we computed the relative variation of the probability of a bin using the transition matrix method for every one of the $1,000$ standard MMC simulations. Therefore, we effectively computed the true confidence interval of the error estimated using the transition matrix method. We have verified that the confidence interval calculated using the double bootstrap procedure on a single standard MMC simulation agrees well with the true confidence interval in all the cases that we investigated.

We observed an excellent agreement between the results obtained with the transition matrix method based on a single standard MMC simulation and the results obtained with $1,000$ standard MMC simulations for both 15 and 80 fiber sections when the relative variation ($\hat{\sigma}_{\hat{P}_{\mathrm{DGD}}}/\hat{P}_{\mathrm{DGD}}$) is smaller than 15%. For larger relative variation, the true error is within the confidence interval of the error, which can be estimated using the double bootstrap method described in Section 5.2.3. The curves for the 80-section PMD emulator have a larger DGD range because a fiber with 80 birefringent sections is able to produce larger DGD values than is possible with a fiber with 15 birefringent fiber sections [44].

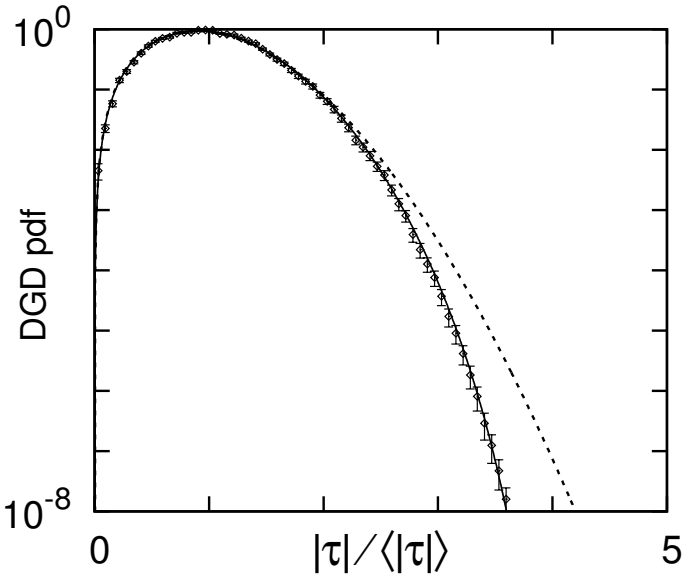

Figure 5.3: The pdf of the normalized DGD, $|\boldsymbol{\tau}|/\langle|\boldsymbol{\tau}|\rangle$, for the 15-section PMD emulator using 14 MMC iterations with $4,000$ samples. (i) Diamonds: DGD pdf with error estimation using the transition matrix method, (ii) Dashed line: Maxwellian pdf, (iii) Solid line: Analytical pdf of the DGD for the 15-section PMD emulator.

In Figs. 5.3 and 5.4, I show with symbols the results for the pdf of the normalized

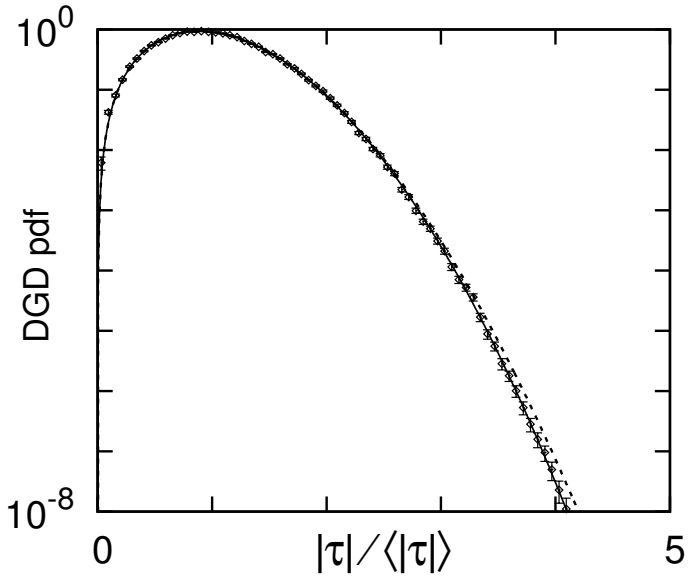

Figure 5.4: The pdf of the normalized DGD, $|\boldsymbol{\tau}|/\langle|\boldsymbol{\tau}|\rangle$, for the 80-section PMD emulator using 30 MMC iterations with $8,000$ samples. (i) Diamonds: DGD pdf with error estimation using the transition matrix method, (ii) Dashed line: Maxwellian pdf, (iii) Solid line: Analytical pdf of the DGD for the 80-section PMD emulator.

DGD and its confidence interval using the numerical procedure that I presented in Section 5.2. The solid line shows the pdf of the normalized DGD obtained analytically using the solution presented in (2.21) (see also [46]) for 15 and 80 concatenated birefringent fiber sections with equal length. For comparison, we also show the Maxwellian pdf for the same mean DGD. In table 5.1, I present selected data points from the curves shown in Fig. 5.3. For both 15- and 80-section emulators, we find that the MMC yields estimates of the pdf of the normalized DGD with a small confidence interval. In Figs. 5.3 and 5.4, we see that the standard deviation ($\hat{\sigma}_{\hat{P}_{\mathrm{DGD}}}$) for the DGD pdf is always small compared to the DGD pdf. The values of the relative variation ($\hat{\sigma}_{\hat{P}_{\mathrm{DGD}}}/\hat{P}_{\mathrm{DGD}}$) ranges from 0.016 to 0.541. We used only $56,000$ MMC

$\lvert\boldsymbol{\tau}\rvert / \langle\lvert\boldsymbol{\tau}\rvert\rangle$	P_{DGD}	$\hat{P}_{\mathrm{DGD}}$	$\hat{\sigma}_{\hat{P}_{\mathrm{DGD}}}$	$\hat{\sigma}_{\hat{P}_{\mathrm{DGD}}} / \hat{P}_{\mathrm{DGD}}$
0.031	3.00×10^{-3}	4.50×10^{-3}	1.35×10^{-3}	0.301
0.344	3.16×10^{-1}	2.84×10^{-1}	1.75×10^{-2}	0.062
0.719	8.56×10^{-1}	8.57×10^{-1}	2.83×10^{-2}	0.033
1.094	8.63×10^{-1}	8.50×10^{-1}	2.76×10^{-2}	0.033
1.469	4.64×10^{-1}	4.66×10^{-1}	2.16×10^{-2}	0.046
1.844	1.43×10^{-1}	1.36×10^{-1}	1.21×10^{-2}	0.089
2.219	2.50×10^{-2}	2.32×10^{-2}	3.37×10^{-3}	0.145
2.594	2.26×10^{-3}	2.15×10^{-3}	4.43×10^{-4}	0.206
2.969	8.70×10^{-5}	7.57×10^{-5}	2.16×10^{-5}	0.286
3.344	8.92×10^{-7}	8.13×10^{-7}	3.49×10^{-7}	0.430
3.594	1.10×10^{-8}	1.59×10^{-8}	8.63×10^{-9}	0.541

Table 5.1: Selected data points from the curves shown in Fig. 5.3. The columns from left to right show: the normalized DGD value, the analytical probability density function, the estimated probability density function, the standard deviation computed using the transition matrix method, and the relative variation.

samples to compute the pdf of the DGD in a 15-section emulator, but we were able nonetheless to accurately estimate probabilities as small as 10^{-8}. Since the relative error in unbiased Monte Carlo simulations is approximately given by $N_I^{-1/2}$, where N_I is the number of hits in a given bin, it would be necessary to use on the order of 10^9 unbiased Monte Carlo samples to obtain a statistical accuracy comparable to the results that I show in the bin with lowest probability in Figs. 5.3 and 5.4.

I conclude this chapter by stressing that the computational time that is required to estimate the errors using the transition matrix method does not scale with the

time needed to carry out a single standard MMC simulation. It takes approximately 17.5 seconds of computation using a Pentium 4.0 computer with 3 GHz of clock speed to estimate the errors in the pdf of the DGD for the 80-section emulator using $1,000$ pseudo-MMC simulations with the transition matrix method, once the transition matrix is available. The computational time that is required to compute the pdf of the DGD using only one standard MMC simulation is 60 seconds. To obtain $1,000$ standard MMC simulations would require about 16.6 hours of CPU time in this case.

I also stress that it is difficult to estimate the statistical errors in MMC simulations because the algorithm is iterative and highly nonlinear. I introduced the transition matrix method that allows us to efficiently estimate the statistical errors from a single standard MMC simulation, and I showed that this method is a variant of the bootstrap procedure. My colleagues and I applied this method to calculate the pdf of the DGD and its expected error for 15-section and 80-section PMD emulators. Finally, we validated this method in both cases by comparing the results to estimates of the error from ensembles of $1,000$ independent standard MMC simulations. The agreement was excellent. In Chapter 6, we apply the transition matrix method to estimate errors in the outage probability of PMD uncompensated and compensated systems. We anticipate that the transition matrix method will allow one to estimate errors with any application of MMC including the computation of the pdf of the received voltage in optical communication systems [69] and the computation of rare events in coded communication systems [73].

Chapter 6

Comparison of two biasing Monte Carlo methods for calculating outage probabilities in systems with multi-section PMD compensators

In this chapter, I present the work in which my colleagues and I evaluate the performance of single-section and three-section polarization mode dispersion compensators using the biasing Monte Carlo methods of importance sampling and MMC, as in [25]. In Chapter 3, I showed that standard importance sampling in which only first-order PMD is biased is insufficient to compute penalties in most compensated systems. In Chapter 4, I used MMC to show the regions in the $|\boldsymbol{\tau}|$–$|\boldsymbol{\tau}_\omega|$ plane that are the dominant source of penalties in uncompensated and single-section PMD compensated systems. Here, I show that both multiple importance sampling in which both first- and second-order PMD are biased and MMC work well with all the compensators that we investigated. I show that multiple importance sampling works well even in a system with a three-section compensator, when both first- and second-order PMD are compensated. The applicability of importance sampling in these systems is consis-

tent with the existence of a large correlation between the first- and second-order PMD of the transmission line and higher orders of PMD after compensation, so that the first two orders, even when compensated, remain highly correlated with the residual penalty. My colleagues and I directly demonstrated the existence of this correlation.

6.1 MMC and importance sampling to compute PMD-induced penalties

As I mentioned earlier, the large PMD penalties of interest to system designers cannot be efficiently computed using standard, unbiased Monte Carlo simulations, since they are very rare. To overcome this hurdle, advanced Monte Carlo methods such as importance sampling [28], [29] and multicanonical Monte Carlo [18] have recently been applied to compute these penalties [19], [24] using a much smaller number of samples.

In optical fiber communication systems without PMD compensators, the penalty is correlated with the differential group delay due to PMD. As a consequence, one may apply importance sampling in which one only biases the DGD [28] to compute PMD-induced penalties. However, biasing the DGD alone is inadequate to compute penalties in compensated systems. On the other hand, the use of multiple importance sampling, in which both first- and second-order PMD are biased [29], allows one to efficiently study important rare events with large first- and second-order PMD. In [24], [50], we used multiple importance sampling in which the first- and second-order PMD are biased to compute the outage probability due to PMD in uncompensated systems and in compensated systems with a single-section compensator.

As discussed in Chapters 3 and 4, the development of importance sampling requires some *a priori* knowledge of how to bias a given parameter in the simulations. In this particular problem, the parameter of interest is the penalty. However, to date there is no importance sampling method that directly biases the penalty. Instead of directly biasing the penalty, one relies on the correlation of the first- and second-order PMD with the penalty, which may not hold in all compensated systems. In contrast to importance sampling, MMC does not require *a priori* knowledge of which rare events contribute significantly to the penalty distribution function in the tails. MMC is an iterative method, which in each iteration produces a biased random walk that automatically searches the state space for the important rare events. This knowledge is accumulated, allowing the penalty distribution function to be obtained further out in the tail from one iteration to the next.

In this chapter, I present results in which my colleagues and I used multiple importance sampling in which both the first- and second-order PMD are biased. We also used MMC. We used both methods to investigate the performance of single-section and three-section PMD compensators. We show that both methods are appropriate to compute outage probabilities with the compensators that we investigated. They yield the same results within the limit of their statistical errors, and importance sampling yields lower errors for comparable run times. It may appear surprising at first that importance sampling works well with a three-section compensator in which both first- and second-order PMD are compensated by the feedback process. We will show however that the residual PMD is highly correlated with the first two compensated orders in a three-section compensator. The applicability of importance sampling in

this case is consistent with the existence of a large correlation between first- and second-order PMD of the transmission line and the higher orders of PMD after compensation. Thus, even when the first two orders are compensated, they remain highly correlated with the residual penalty.

My colleagues and I investigated the effectiveness of higher-order PMD compensators by comparing a single-section, variable-DGD compensator to a three-section compensator. We focused our study on three-section compensators [74] because they are the simplest multi-section compensators that allows one to compensate for the first- and second-order PMD. We found that the use of a three-section compensator does significantly improve the compensation when compared to a single-section PMD compensator. However, the improvement is less than a factor of two (in dB), despite the large increase in complexity; the three-section compensator has 7 feedback parameters [74], while the single-section has 3 feedback parameters. We show that the residual PMD is highly correlated with the first two compensated orders in a three-section compensator. We attribute the diminished returns with increased complexity to this correlation. This correlation also explains the success of importance sampling in which both first- and second-order PMD are biased.

6.2 PMD Compensators

We investigated a single-section PMD compensator [75], which is a variable-DGD compensator that was programmed to eliminate the residual DGD at the central frequency of the channel after compensation, and a three-section PMD compensator proposed in [74], which compensates for first- and second-order PMD. I described

details of the implementation of the single-section compensator that we used in Section 3.2. The three-section compensator consists of two fixed-DGD elements that compensate for the second-order PMD and one variable-DGD element that eliminates the residual DGD at the central frequency of the channel after compensation. The three-section compensator that we used has the first- and second-order PMD as feedback parameters. This compensator can also in principle operate in a feedforward configuration.

6.2.1 Three-section compensator

Second-order PMD has two components: Polarization chromatic dispersion (PCD) and the principal states of polarization rotation rate (PSPRR) [74]. Let $\boldsymbol{\tau}_1$ be the polarization dispersion vector of the transmission line, and let $\boldsymbol{\tau}_2$ and $\boldsymbol{\tau}_3$ be the polarization dispersion vectors of the two fixed-DGD elements of the three-section compensator. Using the concatenation rule [49], the first- and second-order PMD vector of these three concatenated fibers are given by

$$\boldsymbol{\tau}_{\text{tot}} = \mathsf{R}_3\,\mathsf{R}_2\,\boldsymbol{\tau}_1 + \mathsf{R}_3\boldsymbol{\tau}_2 + \boldsymbol{\tau}_3, \tag{6.1}$$

$$\boldsymbol{\tau}_{\text{tot},w} = (\boldsymbol{\tau}_3 + \mathsf{R}_3\boldsymbol{\tau}_2) \times \mathsf{R}_3\,\mathsf{R}_2\,\tau_1\hat{\mathbf{q}}_1 + \boldsymbol{\tau}_3 \times \mathsf{R}_3\boldsymbol{\tau}_2 + \mathsf{R}_3\,\mathsf{R}_2\,\tau_{1w}\hat{\mathbf{q}}_1 + \mathsf{R}_3\,\mathsf{R}_2\,\tau_1\hat{\mathbf{q}}_{1w}, \tag{6.2}$$

where R_2 and R_3 are the rotation matrices of the polarization controllers before the first and the second fixed-DGD elements of the compensator, respectively. In (6.2), $\tau_{1w}\hat{\mathbf{q}}_1$ and $\tau_1\hat{\mathbf{q}}_{1w}$ are the transmission line PCD and the PSPRR components, respectively, where we express the polarization dispersion vector of the transmission fiber

as $\boldsymbol{\tau}_1 = \tau_1 \hat{\mathbf{q}}_1$. Here, the variable τ is the DGD and $\hat{\mathbf{q}} = \boldsymbol{\tau}/|\boldsymbol{\tau}|$ is the Stokes vector of one of the two orthogonal principal states of polarization.

The three-section PMD compensator has two operating points [74]. For the first operating point, the term $\boldsymbol{\tau}_3 \times \mathsf{R}_3 \boldsymbol{\tau}_2$ in (6.2) is used to cancel the PSPRR component $\mathsf{R}_3\,\mathsf{R}_2\,\tau_1 \hat{\mathbf{q}}_{1w}$, provided that we choose R_3 and R_2 so that $\mathsf{R}_3^{\dagger}\boldsymbol{\tau}_3 \times \boldsymbol{\tau}_2$ and $\mathsf{R}_2 \tau_1 \hat{\mathbf{q}}_{1w}$ are antiparallel, where $\mathsf{R}_3^{\dagger}$ is the Hermitian conjugate of R_3. Note that with this configuration one cannot compensate for PCD.

For the second operating point, $\boldsymbol{\tau}_3 \times \mathsf{R}_3 \boldsymbol{\tau}_2$ in (6.2) is used to compensate for PCD by choosing $\mathsf{R}_3^{\dagger}\boldsymbol{\tau}_3 \times \boldsymbol{\tau}_2$ and $\mathsf{R}_2 \tau_{1w} \hat{\mathbf{q}}_1$ to be antiparallel. Moreover, we can add an extra rotation to R_2 so that $\left[\left(\mathsf{R}_3^{\dagger}\boldsymbol{\tau}_3 + \boldsymbol{\tau}_2\right) \times \mathsf{R}_2\,\tau_1 \hat{\mathbf{q}}_1\right]$ and $\mathsf{R}_2\,\tau_1 \hat{\mathbf{q}}_{1w}$ are also antiparallel. In this way, the compensator can also reduce the PSPRR term. In our simulations, we computed the reduction of the PCD and PSPRR components for the two operating points and we selected the one that presented the largest reduction of the second-order PMD. Finally, the third, variable-DGD, section of the compensator cancels the residual DGD $\boldsymbol{\tau}_{\mathrm{tot}}$ after the first two sections.

6.3 Simulation results and discussions

We evaluated the performance of a single-section and a three-section PMD compensator in a 10 Gbit/s nonreturn-to-zero system with a mean DGD of 30 ps. We used perfectly rectangular pulses filtered by a Gaussian shape filter that produces a rise time of 30 ps. We simulated a string with 8 bits generated using a pseudorandom binary sequence pattern. We modeled the fiber using the coarse step method with

80 birefringent fiber sections, which reproduces first- and higher-order PMD distortions within the probability range of interest [50]. The results of our simulations can also be applied to 40 Gbit/s systems by scaling down all time quantities by a factor of four. As in Chapters 3 and 4, we used the eye opening for performance evaluation. The three-section compensator has two fixed-DGD elements of 45 ps and one variable-DGD element. The results that I present in this section were obtained using 30 MMC iterations with $8,000$ samples each and using importance sampling with a total of 2.4×10^5 samples. We estimated the errors in MMC using the transition matrix method that I described in Chapter 5, while we estimated the errors in importance sampling as described in Chapter 3 and in [24].

In Fig. 6.1, I show the outage probability for a 1-dB penalty as function of the DGD element (τ_c) for a system with the three-section compensator that we used. The mean DGD of the system before compensation is 30 ps. We observed that there is an optimum value for τ_c that minimizes the outage probability, which is close to 45 ps. We set the values for the two fixed-DGD elements of the three-section PMD compensator that we used to this optimum value. The reason why the outage probability rises when τ_c becomes larger than this optimum is because large values of τ_c add unacceptable penalties to fiber realizations with relatively small second-order PMD values that could be adequately compensated at lower values of τ_c. We also observed that there is a relatively small dependence of the outage probability on τ_c. That is because the third, variable-DGD section of the compensator cancels the residual DGD after the first two sections, which significantly mitigates the penalty regardless of the value of τ_c.

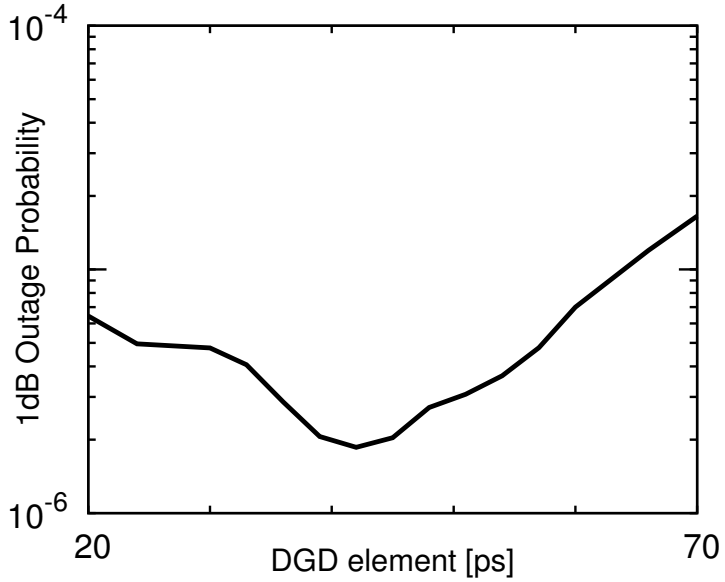

Figure 6.1: Outage probability for a 1-dB penalty as function of the DGD element (τ_c) of the three-section compensator for a system with mean DGD of 30 ps.

In Fig. 6.2, I plot the outage probability ($\hat{P}_{\rm op}$) as a function of the eye-opening penalty for the compensators that we studied. The histogram of the penalty was divided into 34 evenly spaced bins in the range -0.1 and 2 dB, even though I show results from 0 to 1.5 dB of penalty. The maximum relative error ($\hat{\sigma}_{\hat{P}_{\rm op}}/\hat{P}_{\rm op}$) for the curves computed with MMC shown in this plot equals 0.13. The relative error for the curves computed with importance sampling is smaller than with MMC, and is not shown in the plot. The maximum relative error for the curves computed with importance sampling equals 0.1. The results obtained using MMC (solid lines) are in agreement with the ones obtained using importance sampling (symbols). The agreement between the MMC and importance sampling results was expected for the case that we used a single-section compensator, since this type of compensator can

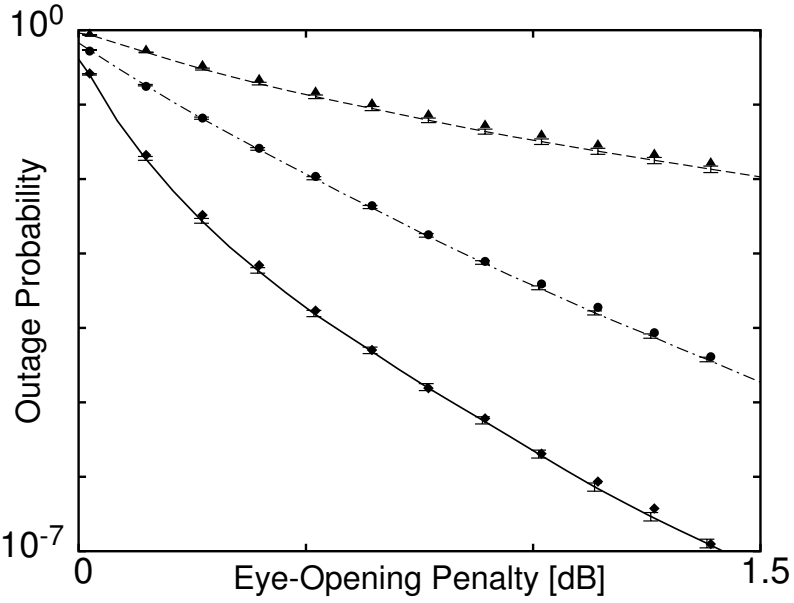

Figure 6.2: Outage probability as a function of the eye-opening penalty for a system with mean DGD of 30 ps. (i) Dashed line (MMC) and triangles (IS): Uncompensated system. (ii) Dot-dashed line (MMC) and circles (IS): System with a single-section compensator. (iii) Solid line (MMC) and diamonds (IS): System with a three-section compensator. The error bars show the confidence interval for the MMC results.

only compensate for first-order PMD [19], so that the dominant source of penalty after compensation is the second-order PMD of the transmission line. Hence, it is expected that MMC and importance sampling give similar results. I showed similar results in Chapter 4. My colleagues and I also observed good agreement between the MMC and importance sampling results for the three-section compensator. This level of agreement indicates that three-section compensators that compensate for the first two orders of the Taylor expansion of the transmission line PMD produce residual third and higher orders of PMD that are significantly correlated with the first- and second-order PMD of the transmission line. That is why the use of importance

sampling to bias first- and second-order PMD is sufficient to accurately compute the outage probability in systems where the first two orders of PMD of the transmission line are compensated.

Significantly, we observed that the performance improvement with the addition of two sections, from the single-section compensator to the three-section compensator, is not as large as the improvement in the performance when one section is added, from the uncompensated to the single-section compensator. The diminishing returns that we observed for increased compensator complexity is consistent with the existence of correlations between the residual higher orders of PMD after compensation and the first two orders of PMD of the transmission line that are compensated by the three-section compensator.

Figures 6.3–6.5 quantify the correlation between the lower and higher orders of PMD. In Fig. 6.3, I show the conditional expectation of the magnitude of second-order of PMD both before and after the three-section compensator given a value of the DGD of the transmission line. In these figures, the DGD $|\boldsymbol{\tau}|$ is normalized by the mean DGD $\langle|\boldsymbol{\tau}|\rangle$ and $|\boldsymbol{\tau}_\omega|$ is normalized by $\langle|\boldsymbol{\tau}_\omega|\rangle$ to obtain results that are independent of the mean DGD and of the mean of the magnitude of second-order PMD. We observed a large correlation between $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$ before compensation, while after compensation $|\boldsymbol{\tau}_\omega|$ is significantly reduced and is less correlated with the DGD, demonstrating the effectiveness of the three-section compensator in compensating for second-order PMD. In Figs. 6.4 and 6.5, I show the conditional expectation of the magnitude of the third-order PMD and of the fourth-order PMD, respectively, before and after the three-section compensator, given a value of the DGD of the transmission

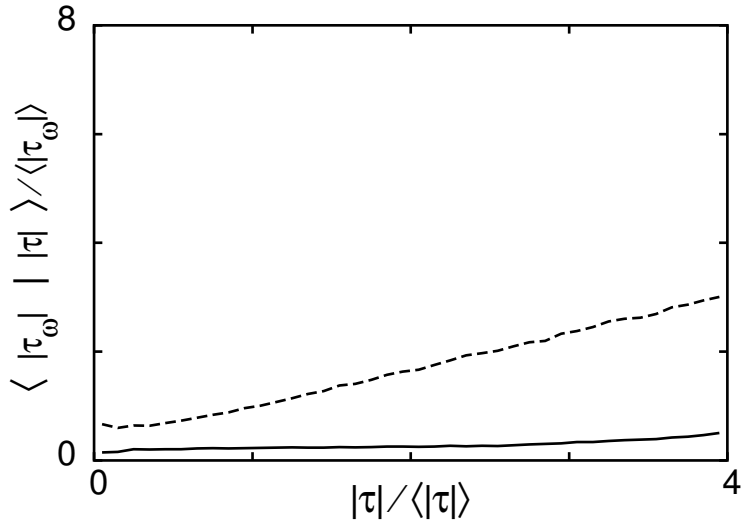

Figure 6.3: Conditional expectation of the magnitude of the normalized second-order PMD, $|\boldsymbol{\tau}_\omega|$, given a value of the DGD of the transmission line, $|\boldsymbol{\tau}|$. Conditional expectation before (dashed) and after (solid) the three-section compensator.

line. In both cases, we observed a high correlation of the third- and the fourth-order PMD with the DGD before and after compensation. In addition, we observed a significant increase of these higher-order PMD components after compensation, which leads to a residual penalty after compensation that is correlated to the original first- and second-order PMD.

In Fig. 6.6, I show contour plots of the conditional expectation of the penalty with respect to the first- and second-order PMD for a system with a three-section PMD compensator [74]. These results show that the residual penalty after compensation is significantly correlated with the first- and second-order PMD. The correlation between the higher orders of PMD with the DGD that I show in Figs. 6.3–6.5 can be estimated from the concatenation rule [49], which explicitly indicates a dependence

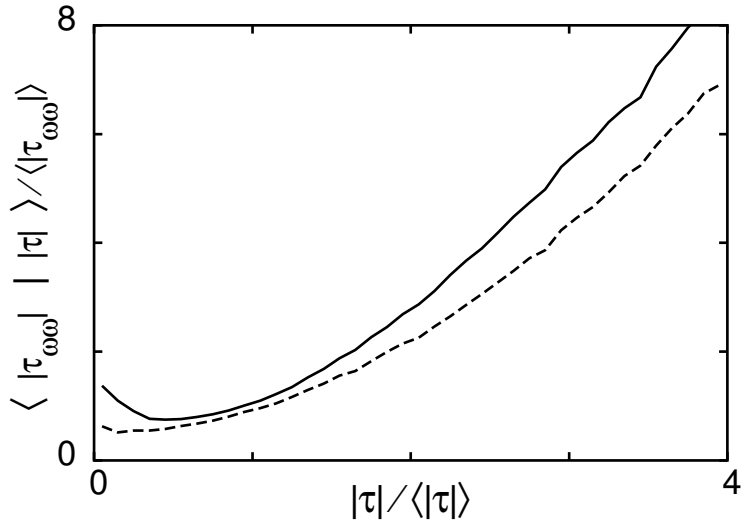

Figure 6.4: Conditional expectation of the magnitude of the normalized third-order PMD, $|\boldsymbol{\tau}_{\omega\omega}|$, given a value of the DGD of the transmission line, $|\boldsymbol{\tau}|$. Conditional expectation before (dashed) and after (solid) the three-section compensator.

of the higher-order PMD components on the lower order components. The increase in these higher-order components after compensation is also due to our choice of the operating point of this compensator, which is set to compensate only for first- and second-order PMD, regardless of the higher-order PMD components. It is possible that this three-section PMD compensator would perform better if all 7 parameters of the compensator are adjusted to achieve the global penalty minimum. However, finding this global optimum is unpractical due to the large number of local optima in such a multidimensional optimization space, as my colleagues and I found in our investigation of single-section PMD compensators [50]. On the other hand, the compensation of first- and second-order PMD using the three-section compensator that we studied here, which was proposed by Zheng, *et al.* [74], can be implemented in practice.

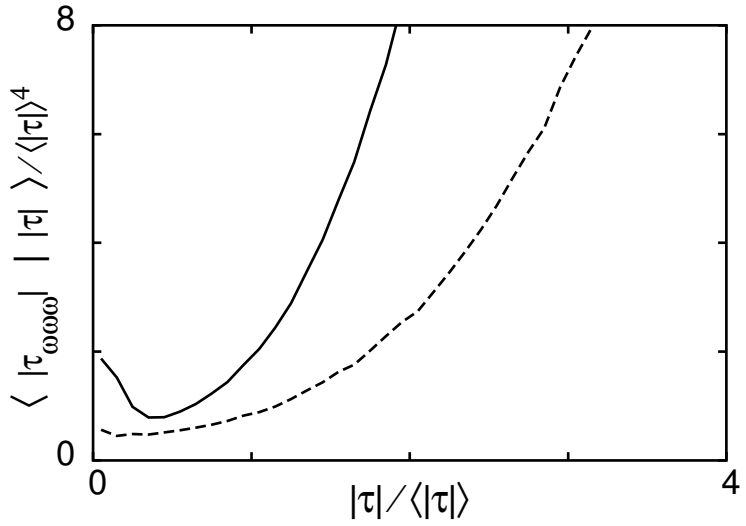

Figure 6.5: Conditional expectation of the magnitude of the normalized fourth-order PMD, $|\boldsymbol{\tau}_{\omega\omega\omega}|$, given a value of the DGD of the transmission line, $|\boldsymbol{\tau}|$. Conditional expectation before (dashed) and after (solid) the three-section compensator.

In this chapter, I showed that both multiple importance sampling and MMC can be used with all the compensators that we investigated to reduce the computation time for the outage probability due to PMD in optical fiber communication systems. Importance sampling in which both the first- and second-order PMD are biased can be used to efficiently compute the outage probability even with a three-section PMD compensator in which both first- and second-order PMD are compensated, which is consistent with the existence of a large correlation between first- and second-order PMD of the transmission line and higher orders of PMD after compensation. We directly verified the existence of these correlations. In contrast to what I presented in Chapter 4, where importance sampling was used to validate the results with MMC, in this chapter, I used MMC to validate the results obtained with importance sampling. We used MMC to validate the results obtained with importance sampling

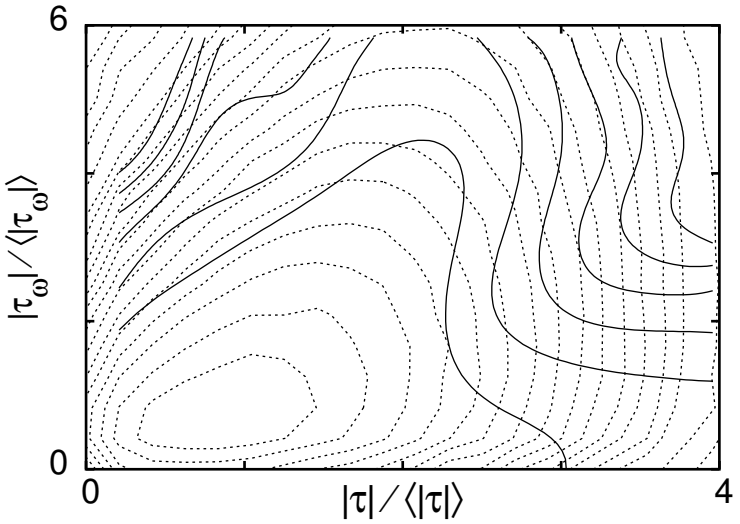

Figure 6.6: Three-section compensated system. The dotted lines are contour plots of the joint pdf of the normalized $|\boldsymbol{\tau}|$ and $|\boldsymbol{\tau}_\omega|$ from the bottom to the top of the plot, are at 3×10^{-n}, with $n = 1, \cdots, 7$ and 10^{-m}, with $m = 1, \cdots, 11$. The solid lines are contour plots of the conditional expectation of the eye-opening penalty in dB from the bottom to the top of the plot, are at 0.1, 0.2, 0.3, 0.4, 0.5, 0.6.

because MMC can be used to compute penalties induced by all orders of PMD and not just penalties correlated to first- and second-order PMD as is the case with the importance sampling method. I showed that MMC yields the same results as importance sampling, within the statistical errors of both methods. Finally, I showed that the three-section compensator offers less than twice the advantage (in dB) of single-section compensators. We attribute the diminishing returns with increased complexity to the existence of correlations between the first two orders of PMD prior to compensation and higher orders of PMD after compensation.

Chapter 7

Conclusions

Polarization mode dispersion is a major source of impairments in optical fiber communication systems. Since PMD is a random process, Monte Carlo simulations are often used to compute PMD-related probability distribution functions. However, the large PMD penalties of interest to system designers cannot be efficiently computed using standard Monte Carlo simulations, since they are very rare. Advanced Monte Carlo methods such as importance sampling and the multicanonical Monte Carlo are statistical methods that make the computation of rare penalties feasible.

In this dissertation, I described the contributions that I made with the collaboration of my colleagues to the field of optical fiber communications. I showed how to apply advanced Monte Carlo techniques of importance sampling and MMC to accurately and efficiently estimate penalties induced by PMD in uncompensated and compensated systems. These techniques make more efficient use of Monte Carlo simulations to compute the probability of rare events that lead to penalties of interest to system designers. Using these two advanced Monte Carlo methods, my colleagues and I evaluated the performance of PMD compensators and compared the efficiency of im-

portance sampling and MMC to compute penalties in different types of compensated systems. We also applied importance sampling to compare the penalty resulting from first-order and all-order PMD models, demonstrating the importance of accurately modeling PMD including higher orders. We showed that evaluating the performance taking into account only first-order PMD produces a good approximation to the true eye-opening penalty of uncompensated systems when the penalty is low. However, when the penalties are high, this model overestimates the penalty for outage probabilities in the range of interest for systems designers, which is typically in the range from 10^{-6} to 10^{-5}.

We also concluded that importance sampling in which only the DGD is biased is sufficient to accurately calculate the uncompensated penalties and their pdfs, but it is not sufficient to accurately calculate the compensated penalties and their pdfs. To study compensated systems where second-order PMD also plays a role it is necessary to use an importance sampling method capable of biasing both the DGD and the length of the frequency derivative of the polarization dispersion vector. The use of importance sampling in which both the DGD and magnitude of the second-order PMD are biased allowed us to conclude that the performance of single-section fixed- and variable-DGD compensators is comparable. In addition, the performance of a variable-DGD compensator that minimizes the residual DGD at the central frequency of the channel is significantly worse than the performance of a fixed- or a variable-DGD compensator that maximizes the eye opening, which indicates the importance of higher-order PMD in determining the penalty since a feedback signal is better correlated to the PMD-induced penalty when extracting more values of the polarization

dispersion vector over the spectrum of the channel.

We compared the two advanced Monte Carlo methods of importance sampling and MMC for calculating outage probabilities in systems with single-section and multi-section PMD compensators. We showed that both methods can be used to speed up the computation of outage probability due to PMD in optical fiber communication systems with all the compensators that we investigated and that multiple importance sampling can be used to efficiently compute the outage probability even with a three-section PMD compensator in which both first- and second-order PMD are compensated, which demonstrates the existence of a large correlation between first- and second-order PMD of the transmission line and higher orders of PMD after compensation. We also showed that the three-section compensator that we studied offers less than twice the advantage (in dB) of single-section compensators. We attributed the diminishing returns with increased complexity to the existence of correlations between the first two orders of PMD prior to compensation and higher orders of PMD after compensation.

Error estimates are essential to verify the accuracy of results obtained with any Monte Carlo method. In this dissertation, I showed how to estimate the statistical errors when using importance sampling and multicanonical Monte Carlo methods. My colleagues and I used standard error estimation procedure and first-order error propagation method to estimate the error in importance sampling simulations. The MMC method, on the other hand, is iterative and highly nonlinear, which induces a web of correlations in the samples. Therefore, one cannot apply to MMC standard error analysis that are traditionally used for simulations with uncorrelated samples.

We developed an efficient numerical method to estimate statistical errors when using MMC, which we refer to as the transition matrix method. The transition matrix method allowed us to efficiently estimate the statistical errors from a single standard MMC simulation. We applied this method to estimate errors in the pdf of the DGD and in the outage probability of uncompensated and compensated systems with PMD.

We anticipate that the transition matrix method will allow one to estimate errors with any application of MMC including the computation of the pdf of the received voltage in optical communication systems [69] and the computation of rare events in coded communication systems [73].

Bibliography

[1] V. J. Mazurczyk and J. L. Zyskind, "Polarization dependent gain in erbium doped-fiber amplifiers," *IEEE Photon. Technol. Lett.*, vol. 6, no. 5, pp. 616–618, 1994.

[2] E. Lichtman, "Limitations imposed by polarization-dependent gain and loss on all-optical ultralong communication systems," *J. Lightwave Technol.*, vol. 13, no. 5, pp. 906–913, 1995.

[3] I. T. Lima Jr., G. Biondini, B. S. Marks, W. L. Kath, and C. R. Menyuk, "Analysis of PMD compensators with fixed DGD using importance sampling," *IEEE Photon. Technol. Lett.*, vol. 14, no. 5, pp. 627–629, 2002.

[4] R. Srinivasan, *Importance Sampling: Applications in Communications and Detection*, New York: Springer, 2002.

[5] L. Rayleigh, "On James Bernoulli's theorem in probabilities," *Philosophical Magazine*, vol. 47, no. 1, pp. 246–251, 1899.

[6] N. Metropolis and S. Ulam, "The Monte Carlo method," *Journal of American Statistical Association*, vol. 44, no. 247, pp. 335–341, 1949.

[7] J. H. Hammersley and D. C. Handscomb, *Monte Carlo Methods*, New York: Wiley, 1964.

[8] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, no. 6, pp. 1087–1092, 1953.

[9] J. R. Howell, "Application of Monte Carlo to heat transfer problems," *Advances in Heat Transfer*, vol. 5, no. 1, pp. 1–54, 1968.

[10] M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods*, New York, NY: John Wiley and Sons, 1986.

[11] M. N. Varma and A. Chatterjee, *Computational Approaches in Molecular Radiation Biology: Monte Carlo Methods*, New York: Plenum Press, 1994.

[12] N. G. Cooper, *From Cardinals to Chaos*, Cambridge: Cambridge University Press, 1989.

[13] C. Z. Mooney, *Monte Carlo Simulation*, California: Sage Publications, 1997.

[14] I. M. Sobol, *The Monte Carlo Method*, Chicago: The University of Chicago Press, 1974.

[15] G. Marsaglia, *Computer Science and Statistics*, Amsterdam: L. Billard, 1985.

[16] H. Sunnerud, C. Xie, M. Karlsson, R. Samuelssonm, and P. A. Andrekson, "A comparison between different PMD compensation techniques," *J. Lightwave Technol.*, vol. 20, no. 3, pp. 368–378, 2002.

[17] R. Y. Rubinstein, *Simulation and the Monte Carlo Method*, New York: John Wiley & Sons, 1981.

[18] B. A. Berg and T. Neuhaus, "The multicanonical ensemble: a new approach to simulate first-order phase transitions," *Phys. Rev. Lett.*, vol. 68, no. 1, pp. 9–12, 1992.

[19] A. O. Lima, I. T. Lima Jr., J. Zweck, and C. R. Menyuk, "Efficient computation of PMD-induced penalties using multicanonical Monte Carlo simulations," in *Proceedings ECOC 2003*, We364, 2003, pp. 538–539.

[20] A. O. Lima, I. T. Lima Jr., T. Adalı, and C. R. Menyuk, "A novel polarization diversity receiver for PMD mitigation," *IEEE Photon. Technol. Lett.*, vol. 14, no. 4, pp. 465–467, 2002.

[21] A. O. Lima, T. Adalı, I. T. Lima Jr., and C. R. Menyuk, "Polarization diversity and equalization for PMD mitigation in optical communication systems," in *Proceedings ICASSP 2002*, III, 2002, pp. 2721–2724.

[22] A. O. Lima, T. Adalı, I. T. Lima Jr., and C. R. Menyuk, "Polarization diversity receiver for PMD mitigation," in *Proceedings OFC 2002*, WI7, 2002, pp. 241–242.

[23] A. O. Lima, I. T. Lima Jr., B. S. Marks, C. R. Menyuk, and W. L. Kath, "Performance analysis of single-section PMD compensators using multiple importance sampling," in *Proceedings OFC 2003*, ThA3, 2003, pp. 419–421.

[24] A. O. Lima, I. T. Lima Jr., C. R. Menyuk, G. Biondini, B. S. Marks, and W. L. Kath, "Statistical analysis of the performance of PMD compensators

using multiple importance sampling," *IEEE Photon. Technol. Lett.*, vol. 15, no. 12, pp. 1716–1718, 2003.

[25] A. O. Lima, I. T. Lima Jr., C. R. Menyuk, and J. Zweck, "Performance evaluation of single-section and three-section PMD compensators using extended Monte Carlo methods," in *Proceedings OFC 2005*, OME27, 2005.

[26] A. O. Lima, I. T. Lima Jr., T. Adalı, and C. R. Menyuk, "Comparison of power penalties due to first- and all-order PMD distortions," in *Proceedings ECOC 2002*, 7.1.2, 2002, pp. 1–2.

[27] A. O. Lima, I. T. Lima Jr., C. R. Menyuk, and T. Adalı, "Comparison of penalties resulting from first-order and all-order polarization mode dispersion distortions in optical fiber transmission systems," *Optics Lett.*, vol. 28, no. 5, pp. 310–312, 2003.

[28] G. Biondini, W. L. Kath, and C. R. Menyuk, "Importance sampling for polarization-mode dispersion," *IEEE Photon. Technol. Lett.*, vol. 14, no. 3, pp. 310–312, 2002.

[29] S. L. Fogal, G. Biondini, and W. L. Kath, "Multiple importance sampling for first- and second-order polarization-mode dispersion," *IEEE Photon. Technol. Lett.*, vol. 14, no. 9, pp. 1273–1275, 2002.

[30] D. Yevick, "The accuracy of multicanonical system models," *IEEE Photon. Technol. Lett.*, vol. 15, no. 2, pp. 224–226, 2003.

[31] R. Walpole and R. Myers, *Probability and Statistics for Engineers and Scientists*, New York: Macmillian, 1993.

[32] B. Efron, "Bootstrap methods: another look at the Jackknife," *The Annals of Statistics*, vol. 7, no. 1, pp. 1–26, 1979.

[33] I. P. Kaminow, "Polarization in optical fibers," *IEEE J. Quantum Electron.*, vol. 17, no. 1, pp. 15–22, 1981.

[34] C. D. Poole and J. Nagel, *Optical Fiber Telecommunications*, vol. III-A, San Diego, U.S.A: Academic, 1997.

[35] C. D. Poole, J. H. Winters, and J. A. Nagel, "Dynamical equation for polarization dispersion," *Optics Lett.*, vol. 16, no. 6, pp. 372–374, 1991.

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

[37] D. Marcuse, C. R. Menyuk, and P. K. A. Wai, "Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence," *J. Lightwave Technol.*, vol. 15, no. 9, pp. 1735–1746, 1997.

[38] G. P. Agrawal, *Fiber Optic Communication Systems*, New York: Wiley-Interscience, 2nd edition, 1997.

[39] M. H. Eiselt, "PMD compensation: A system perspective," in *Proceedings OFC 2004*, ThF4, 2004, pp. 28–30.

[40] I. P. Kaminow and T. Li, *Optical Fiber Telecommunications*, vol. IV-B, San Diego, CA: Academic, 2002.

[41] G. J. Foschini, L. E. Nelson, R. M. Jopson, and H. Kogelnik, "Probability densities of second-order polarization mode dispersion including polarization dependent chromatic fiber dispersion," *IEEE Photon. Technol. Lett.*, vol. 12, no. 3, pp. 293–295, 2000.

[42] P. Ciprut, N. Gisin, R. Passy, J. P. Von der Weid, F. Prieto, and C. W. Zimmer, "Second-order polarization mode dispersion: Impact on analog and digital transmissions," *J. Lightwave Technol.*, vol. 16, no. 5, pp. 757–771, 1998.

[43] H. Goldstein, *Classical Mechanics*, Reading, MA: Addison-Wesley, 1980.

[44] I. T. Lima Jr., R. Khosravani, P. Ebrahimi, E. Ibragimov, A. E. Willner, and C. R. Menyuk, "Comparison of polarization mode dispersion emulators," *J. Lightwave Technol.*, vol. 19, no. 12, pp. 1872–1881, 2001.

[45] G. Poole and D. L. Favin, "Polarization dispersion measurements based on transmission spectra through a polarizer," *J. Lightwave Technol.*, vol. 12, no. 6, pp. 917–929, 1994.

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

[47] T. Takahashi, T. Imai, and M. Aiki, "Automatic compensation technique for timewise fluctuation polarization mode dispersion in in-line amplifier systems," *Electron. Lett.*, vol. 30, no. 4, pp. 348–349, 1994.

[48] F. Heismann, "Automatic compensation of first-order polarization mode dispersion in a 10 Gbit/s transmission system," in *Proceedings ECOC 1998*, 7.1.2, 1998, pp. 329–330.

[49] J. P. Gordon and H. Kogelnik, "PMD fundamentals: Polarization mode dispersion in optical fibers," *Proc. Nat. Acad. Sci.*, vol. 97, no. 9, pp. 4541–4550, 2000.

[50] I. T. Lima Jr., A. O. Lima, G. Biondini, C. R. Menyuk, and W. L. Kath, "A comparative study of single-section polarization-mode dispersion compensators," *J. Lightwave Technol.*, vol. 22, no. 4, pp. 1023–1032, 2004.

[51] H. Bülow, "System outage probability due to first- and second-order PMD," *IEEE Photon. Technol. Lett.*, vol. 10, no. 5, pp. 696–698, 1998.

[52] P. R. Trischitta and E. L. Varma, *Jitter in Digital Transmission Systems*, Boston: Artech House, 1989.

[53] E. Polak, *Computational Methods in Optimization*, New York: Academic Press, 1971.

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

[55] E. Veach, *Robust Monte Carlo methods for light transport simulation*, Ph.D. Dissertation, Stanford University, Palo Alto, CA, December 1997.

[56] I. T. Lima Jr., A. O. Lima, J. Zweck, and C. R. Menyuk, "Efficient computation of outage probabilities due to polarization effects in a WDM system using a reduced Stokes model and importance sampling," *IEEE Photon. Technol. Lett.*, vol. 15, no. 1, pp. 45–47, 2003.

[57] G. Biondini, W. L. Kath, and C. R. Menyuk, "Non-Maxwellian DGD distributions of PMD emulators," in *Proceedings OFC 2001*, ThA5, 2001, pp. 1–3.

[58] R. E. Walpole and R. H. Myers, *Optical Fiber Communication Systems*, New York, NY: Macmillan Publishing Co., 1993.

[59] J. Mandel, *The Statistical Analysis of Experimental Data*, New York: Dover, 1964.

[60] C. D. Poole, R. W. Tkash, A. R. Chraplyvy, and D. A. Fishman, "Fading in lightwave systems due to polarization-mode dispersion," *IEEE Photon. Technol. Lett.*, vol. 3, no. 1, pp. 68–70, 1991.

[61] H. Bülow, "Operation of digital transmission system with minimal degradation due to polarization-mode dispersion," *Electron. Lett.*, vol. 31, no. 3, pp. 214–215, 1995.

[62] A. Djupsjöbacka, "Calculation of signal outage due to polarization mode dispersion," *IEEE Photon. Technol. Lett.*, vol. 13, no. 7, pp. 660–662, 2001.

[63] B. Falcidieno, I. Herman, and C. Pienovi, *Computer Graphics and Mathematics*, New York: Springer-Verlag, 1992.

[64] F. Buchali, S. Lanne, J.-P. Thiery, W. Baumert, and H. Bülow, "Fast eye monitor for 10 Gbit/s and its application for optical PMD compensation," in *Proceedings OFC 2001*, TuP5, 2001, pp. 1–3.

[65] I. Roudas, G. Piech, Y. Zhu, and D. Q. Chowdhury, "Coherent heterodyne frequency-selective polarimeter for error signal generation in higher-order PMD compensators," in *Proceedings OFC 2002*, WQ2, 2002, pp. 299–301.

[66] B. A. Berg, "Multicanonical simulations step by step," *Comp. Phys. Commum.*, vol. 153, no. 3, pp. 397–407, 2003.

[67] D. Yevick, "Multicanonical communication system modeling—application to PMD statistics," *IEEE Photon. Technol. Lett.*, vol. 14, no. 11, pp. 1512–1514, 2002.

[68] S. Kachigan, *Multivariate Statistical Analysis: A Conceptual Introduction*, New York: Radius Press, 1991.

[69] R. Holzlöhner and C. R. Menyuk, "The use of multicanonical Monte Carlo simulations to obtain accurate bit error rates in optical communications systems," *Opt. Lett.*, vol. 28, no. 20, pp. 1894–1897, 2003.

[70] K. Singh, "On the asymptotic accuracy of Efron's Bootstrap," *The Annals of Statistics*, vol. 9, no. 6, pp. 1187–1195, 1981.

[71] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*, New York, NY: Chapman and Hall, 1993.

[72] J. Booth and P. Hall, "Monte Carlo approximation and the Iterated Bootstrap," *Biometrika*, vol. 81, no. 2, pp. 331–340, 1994.

[73] R. Holzlöhner, A. Mahadevan, C. R. Menyuk, J. M. Morris, and J. Zweck, "Evaluation of the very low BER of FEC codes using dual adaptive importance sampling," *Comm. Lett.*, vol. 9, no. 2, pp. 163–165, 2005.

[74] Y. Zheng, B. Yang, and X. Zhang, "Three-stage polarization mode dispersion compensator capable of compensating second-order polarization model dispersion," *IEEE Photon. Technol. Lett.*, vol. 14, no. 10, pp. 1412–1414, 2002.

[75] R. Noé, D. Sandel, M. Yoshida-Dierolf, S. Hinz, V. Mirvoda, A. Schöpflin, C. Glingener, E. Gottwald, C. Scheerer, G. Fisher, T. Weyrauch, and W. Haase, "Polarization mode dispersion compensation at 10, 20, and 40 Gb/s with various optical equalizers," *J. Lightwave Technol.*, vol. 17, no. 9, pp. 1602–1616, 1999.