First-principles quantum simulations of nonclassical systems

Simon Kiesewetter

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

Centre for Quantum and Optical Science
Swinburne University of Technology • Melbourne, Australia

July 2017

Abstract

New results in the domain of quantum simulations are presented. After a general overview of state-of-the art techniques for quantum simulations where special emphasis is placed on phase-space methods, these methods are applied to study nonlocality phenomena such as entanglement, the Einstein-Podolsky-Rosen paradox and Bell violations. A proposed experiment in the emerging field of optomechanics, which has already been carried out in the electromechanical domain, is analyzed. This experiment is known to generate entanglement between a massive mechanical oscillator and a radiation field. A full nonlinear analysis based on first-principle methods is carried out. A comparison between different continuous value entanglement parameters is drawn. The experiment is analyzed with regards to the prospect of observing the EPR paradox.

An overview of quantum simulations for continuous variable systems exhibiting the EPR paradox and Bell violation is given. A reflection on Feynman's famous claim on the feasibility of quantum simulations is made. Different simulation methods are analyzed with regards to their accuracy for highly nonlinear systems.

A new type of Bell inequality based on the quaternion and octonion algebra is analyzed for Greenberger-Horne-Zeilinger spin states. The exact detector settings for optimal Bell violations are presented and minimal detector efficiencies are worked out.

An original method for the integration of stochastic differential equations is presented. The method combines a conventional Monte-Carlo approach with deterministic moment hierarchies. This method is capable of producing substantially more accurate results than conventional integration methods by reducing the stochastic error while maintaining linear complexity in the number of samples. Two separate version of this method, for Itō-type stochastic differential equations and for Stratonovich-type stochastic differential equations, are presented.

Acknowledgments

Foremost, I wish to express my sicere gratitute to my principal supervisor, Prof. Peter D. Drummond and my secondary supervisor, Prof. Margaret D. Reid, for countless fruitful discussions which have been of inestimable value during my time as a PhD student. With their immense knowledge, great patience, enthusiam and overall kindness, they have guided me through the PhD process in the best possible way. They originated many great ideas and research proposals. I couldn't have imagined a better team of supervisors. I also wish to thank them for proof-reading my thesis.

Among my fellow physicists at CQOS, I would like to thank, in no particular order, Dr. Bogdan Opanchuk, Dr. Laura Rosales-Zárate, Run Yan Teh, Prof. Bryan Dalton, Alessandro Brolis, Rodney Polkinghorne, Prof. Xia-Ji Liu, Prof. Hui Hu, Umberto Toniolo, King Lun Ng and Dr. Brendan Mulkerin for many intellectually stimulating discussions and collaborations and making the time at CQOS a pleasant and enriching experience. From outside of CQOS, special thanks go to Prof. Qiongyi He from Beijing university and Prof. Khaled Dechoum from Federal Fluminense University in Brazil for the opportunity of scientific collaboration.

I wish to thank my girlfriend Julie Tatiana Cortés Muñoz for her support, especially during my thesis write-up. I also would like to thank both of my parents for always being there for me and supporting me.

Declaration

This is to certify that this thesis

- contains no material which has been accepted for the award to me of any other degree or diploma, except where due reference is made in the text,
- to the best of my knowledge contains no material previously published or written by another person except where due reference is made in the text, and
- where the work is based on joint research or publications, discloses the relative contributions of the respective workers or authors.

Simon Kiesewetter S. Kiesewetter

Date 21 July 2017

Contents

List of Figures . . . xii
List of Tables . . . xiv

1 List of papers as part of the thesis **1**
1.1 Additional papers not part of the thesis . . . 3

2 Introduction **5**
2.1 Literature review on nonclassical systems . . . 5
2.2 Overview of the papers as part of the thesis . . . 7

3 Foundations **11**
3.1 Stochastic methods . . . 11
3.1.1 Fokker-Planck equation and stochastic differential equation . . . 11
3.1.2 Itō and Stratonovich calculus . . . 13
3.1.3 Numerical methods . . . 17
3.2 Phase-space methods . . . 22
3.2.1 Phase-space representation . . . 23
3.2.2 Coherent state . . . 23
3.2.3 Glauber-Sudarshan P-representation . . . 25
3.2.4 s-ordering approach . . . 26
3.2.5 Multi-mode phase-space representations . . . 27
3.2.6 Observables . . . 28
3.2.7 Wigner representation . . . 28
3.2.8 Husimi Q-representation . . . 29
3.2.9 Positive P-representation . . . 29
3.2.10 Operator identities . . . 31
3.2.11 Time-evolution equations . . . 32
3.2.12 Stochastic phase-space methods . . . 34
3.3 Entanglement criteria, EPR steering and Bell inequalities . . . 38
3.3.1 Entanglement criteria . . . 39
3.3.2 EPR-steering criteria . . . 42
3.3.3 Bell inequalities . . . 44
3.3.4 EPR-steering and hierarchy of quantum nonlocality . . . 48
3.4 Cavity Optomechanics . . . 50
3.5 POS . . . 52
3.5.1 Euler-Mayurama version of POS . . . 53
3.5.2 Static ensemble optimization . . . 55

4 Scalable quantum simulation of pulsed entanglement and Einstein-Podolsky-Rosen steering in optomechanics **57**
4.1 Abstract . . . 57
4.2 Introduction . . . 58
4.3 Hamiltonian and operator equations . . . 59
4.3.1 Optomechanical Hamiltonian . . . 60
4.3.2 Master equation . . . 61
4.3.3 Phase-space equations . . . 63

4.4 Adiabatic linearized model . . . 65
4.4.1 Adiabatic Hamiltonian . . . 66
4.4.2 Linearized entanglement generation . . . 67
4.4.3 Linearized readout procedure . . . 68
4.5 Entanglement and EPR-steering measures . . . 70
4.5.1 Entanglement measures . . . 70
4.5.2 EPR-steering criteria . . . 73
4.6 Quantum simulations . . . 75
4.6.1 Electromechanical experiment . . . 75
4.6.2 Optomechanical experiment . . . 79
4.6.3 Numerical methods . . . 81
4.7 Summary . . . 83
4.8 Acknowledgments . . . 83
4.9 Appendix: Effect of losses and inefficiencies . . . 83
4.9.1 Entanglement criteria . . . 84
4.9.2 EPR-steering criteria . . . 85
4.10 Bibliography . . . 86

5 Einstein-Podolsky-Rosen quantum simulations in nonclassical phase-space 91
5.1 Abstract . . . 91
5.2 Introduction . . . 92
5.3 Quantum dynamical simulations . . . 94
5.3.1 Positive P-Representation . . . 94
5.3.2 Dynamical Equations . . . 95
5.4 EPR violations in Nondegenerate down-conversion . . . 96
5.4.1 System Hamiltonian . . . 97
5.4.2 Stochastic Equations . . . 98
5.4.3 Output Quadratures . . . 99
5.5 Recent simulations . . . 101
5.5.1 The Reid-Walls Bell Experiment . . . 101
5.5.2 Bell Inequality Violations . . . 102
5.5.3 Nonlinear EPR-Steering in Optomechanics . . . 104
5.5.4 Nonlinear Corrections . . . 105
5.6 Quantum field simulations . . . 109
5.6.1 Master Equation and Quantum Langevin Form . . . 111
5.6.2 Stochastic Equations in the Positive-P Representation . . . 111
5.7 Summary . . . 114
5.8 Bibliography . . . 115

6 Violations of multisetting quaternion and octonion Bell inequalities 121
6.1 Abstract . . . 121
6.2 Introduction . . . 122
6.3 Shchukin-Vogel multisetting Bell inequalities . . . 124
6.3.1 Quaternion and Octonion Bell inequalities . . . 125
6.4 Quantum predictions for GHZ state . . . 127
6.5 Detection inefficiencies, Noise and heralding . . . 130
6.6 Generalized GHZ states . . . 134
6.7 Discussion and Conclusion . . . 137
6.8 Acknowledgments . . . 137
6.9 Bibliography . . . 138

7 Parallel optimized sampling for stochastic equations 145
7.1 Abstract . . . 145
7.2 Introduction . . . 146
7.3 Stochastic integration error . . . 148
7.3.1 Distributions and observables . . . 148
7.3.2 Fokker-Planck and stochastic equations . . . 150
7.3.3 Sampling and error criteria . . . 150
7.3.4 Optimizing the sample number . . . 152
7.3.5 Effective integration order . . . 153
7.4 Initial conditions for parallel optimized sampling . . . 154
7.4.1 Static POS . . . 155
7.4.2 One-dimensional example of iteration equations . . . 157
7.4.3 Static iteration solution with a matrix pseudo-inverse . . . 157
7.4.4 Improved matrix inversion iterations . . . 158
7.4.5 One-dimensional case: power series method . . . 159
7.5 Numerical examples: Static optimization . . . 159
7.5.1 Static optimization . . . 160
7.5.2 Distribution of the optimized variance reduction . . . 161
7.5.3 Number of iterations and optimization distance . . . 162
7.5.4 Initial optimization cost . . . 164
7.6 Parallel dynamic optimization . . . 165
7.6.1 Euler-Maruyama algorithm . . . 166
7.6.2 One-dimensional case . . . 167
7.6.3 Error propagation . . . 167
7.6.4 Combined optimization . . . 169
7.6.5 Iterative solution to the combined equations . . . 170
7.6.6 Individual optimization . . . 171
7.6.7 Iterative solution to the individual equations . . . 172
7.7 Synthetic one-step benchmarks . . . 173
7.7.1 Combined optimization . . . 174
7.7.2 Number of iterations and convergence distance . . . 175
7.7.3 Individual optimization . . . 176
7.7.4 Number of iterations and convergence distance . . . 178
7.8 Dynamical numerical results . . . 179
7.8.1 Linear drift, additive noise . . . 180
7.8.2 Cumulant and moment variance reduction . . . 181
7.8.3 Nonlinear drift, additive noise . . . 184
7.8.4 Irregular drift, additive noise . . . 186
7.9 Two-dimensional case . . . 187
7.9.1 Error in the equilibrium photon number . . . 189
7.10 Conclusions . . . 190
7.11 Acknowledgements . . . 190
7.12 Bibliography . . . 191

8 Algorithms for integration of stochastic differential equations using parallel optimized sampling in the Stratonovich calculus 193
8.1 Abstract . . . 193
8.2 Introduction . . . 194
8.3 Fokker–Plack equation . . . 196
8.3.1 Ito SDE . . . 196
8.3.2 Stratonovich SDE . . . 198

8.3.3 Observables and moment hierarchies . . . 199
8.3.4 Convergence of moment hierarchies . . . 200
8.4 Stratonovich POS methods . . . 201
8.4.1 Central difference algorithm . . . 201
8.4.2 Initial ensemble optimization . . . 202
8.4.3 Moment equations . . . 204
8.4.4 POS optimization . . . 205
8.4.5 Iterative method for midpoint location . . . 206
8.4.6 Optimization . . . 208
8.4.7 Code examples . . . 209
8.5 Linear Examples . . . 211
8.5.1 Linear oscillator . . . 211
8.5.2 Kubo oscillator . . . 213
8.6 Nonlinear stochastic examples . . . 216
8.6.1 Bistable nonlinear oscillator . . . 217
8.6.2 Cumulant hierarchy method . . . 219
8.6.3 Bistable oscillator with multiplicative noise . . . 222
8.6.4 Laser equation . . . 225
8.7 Summary . . . 226
8.8 Acknowledgments . . . 227
8.9 Appendix: Systems with closed moment hierarchies . . . 228
8.9.1 Linear oscillator . . . 228
8.9.2 Kubo oscillator . . . 228
8.10 Bibliography . . . 230

A POS, ensemble size scaling 233
A.1 Description . . . 233
A.2 Results . . . 236
A.3 Discussion . . . 236

9 Discussion 241
9.1 Concluding remarks . . . 245

B List of Papers 247
B.1 List of papers as part of the thesis including citing articles . . . 247
B.2 List of papers not part of the thesis including citing articles . . . 249

Bibliography 251

List of Figures

4.1 Schematic diagram of the experimental setup . . . 61
4.2 Different entanglement signatures for the optomechanical experiment . . 74
4.3 Entanglement signature for the electromechanical experiment . . . 77
4.4 EPR-steering signature for the electromechanical experiment . . . 78
4.5 Multiplicative entanglement signature for the optomechanical experiment 80
4.6 EPR-steering signatures for the optomechanical experiment . . . 82

5.1 Simulated results for Bell-type inequality violation in a parametric down conversion process . . . 103

5.2 Entanglement signature for a pulsed electromechanical experiment . . . 106
5.3 EPR-steering signature results using different simulation methods I . . . 107
5.4 EPR-steering signature results using different simulation methods II . . . 108

6.1 Threshold detector efficiency for N sites 132
6.2 Effect of measurement noise 135
6.3 Violations for the generalized GHZ state 136
6.4 Threshold detector efficiencies for the generalized GHZ state 136

7.1 Distribution of normalized errors in the static optimization tests for the matrix inversion method 161
7.2 Number of iterations taken by the matrix inversion method, Relative distance D between the optimized vector and the initial randomly sampled vector, Average relative distance depending on the number of trajectories 163
7.3 Dependence of normalized errors on the number of trajectories 163
7.4 Dependence of the cost on the number of trajectories 164
7.5 Distribution of normalized errors for the combined method 174
7.6 number of iterations taken for the combined method, Relative distance between the optimized vector and the initial approximation, average relative distance depending on the number of trajectories in the synthetic optimization tests for the individual expansion method 175
7.7 Average relative distance depending on Δt for the combined method, Average relative distance depending on N_S for the combined method . . 176
7.8 Dependence of errors on the number of trajectories for the combined method, synthetic benchmark 177
7.9 Dependence of average scaled errors on Δt for the combined method, synthetic benchmark 177
7.10 Distribution of normalized errors in the synthetic benchmark for the individual method 178
7.11 Number of iterations taken for the individual method, Relative distance between the optimized vector and the initial approximation, Average relative distance depending on the number of trajectories for the individual method 178
7.12 Dependence of average scaled errors on Δt in the synthetic test for the individual method 179
7.13 Dependence of errors on the number of trajectories in the synthetic test for the expansion method 180
7.14 Linear SDE case, combined method, moment results 182
7.15 Linear SDE case, individual method, moment results 183
7.16 Linear SDE case, individual method, cumulant results 183
7.17 Nonlinear SDE case, combined method, moment results 185
7.18 Nonlinear SDE case, combined method, cumulant results 186
7.19 Nonlinear, non-reguar SDE case, combined method, moment results . . . 187
7.20 Nonlinear, non-reguar SDE case, individual method, cumulant results . . 188
7.21 Laser equation, combined method, as a function of diffusion paramter b . 189

8.1 Linear oscillator results, comparison between conventional and POS-optimized integration and between Ito-type POS and Stratonovich-type POS optimization 212

8.2 Kubo oscillator case 215
8.3 Bistable oscillator case, steady-state accuracies 218
8.4 Bistable oscillator case, accuracies of time-evolution using a non-optimized stochastic method, a POS-optimized method and a truncated cumulant hierarchy method, moment results 220
8.5 Bistable oscillator case, accuracies of time-evolution using a non-optimized stochastic method, a POS-optimized method and a truncated cumulant hierarchy method, cumulant results 221
8.6 Second nonlinear case, steady-state accuracies, moment results 224
8.7 Laser equation case, accuracy as a function of diffusion parameter b . . . 226

A.1 Stochastic and systematic error for nonlinear SDE 237
A.2 Stochastic and systematic error for nonlinear SDE, reduced time-step . . 238

List of Tables

6.1 Evaluation of Pauli products 129
6.2 Optimal angles for 2 and 3 detector settings 130
6.3 Optimal angles for 4 and 5 measurement settings 130
6.4 Optimal angles for 6 and 7 measurement settings 131
6.5 Optimal angles for 8 measurement settings 131

Chapter 1

List of papers as part of the thesis

This thesis was created in accordance with regulations of Swinburne University and with full support of the supervisory panel as a research thesis by publications (written research thesis including associated papers). The publications listed below were written, submitted and accepted by the listed research journals during my candidature period. They consitute chapters 4 – 8 of this document. These have been reformatted into an acceptable format. Each chapter corresponding to a publication has a separate bibliography. The rest of this document has its own bibliography which is seperate from those in chapters 4 – 8 and is given at the end of this thesis.

List of papers:

- S. Kiesewetter, Q. Y. He, P. D. Drummond, and M. D. Reid, "Scalable quantum simulation of pulsed entanglement and Einstein-Podolsky-Rosen steering in optomechanics," Physical Review A **90**, p. 043805; accepted 4 September 2014 (chapter 4)

  For this publication, together with my supervisors I planned the research strategy. I then formulated and solved the phase-space equations, wrote the computer code, and ran the simulations together with comparisons to experiment.

- S. Kiesewetter, K. Dechoum, B. Opanchuk, L. Rosales-Zárate, M. D. Reid, and P. D. Drummond, "Einstein–Podolsky–Rosen quantum simulations in nonclassical phase-space," Journal of the Optical Society of America B **32**, pp. A64-A73; accepted 17 March 2015 (chapter 5)

For this publication, I carried out a literature review and studied the effects of highly nonlinear optomechanical coupling for EPR-steering in the experiment presented in the previous paper. I wrote the code for a positive-P, truncated Wigner as well as a linearized simulation of the system. I ran the simulations and created the respective figures.

- S. Kiesewetter, P. D. Drummond, and M. D. Reid, "Violations of multisetting quaternion and octonion Bell inequalities," Physical Review A **92**, p. 032118; accepted 8 July, 2015 (chapter 6)

  For this publication, I first did a literature review on Bell-type inequalities, particularly CFRD-type and CV-type ones. Together with my supervisors, I then planned the research strategy. I wrote the computer code and carried out a systematic investigation on the optimal detector settings and corresponding detector efficiencies, together with comparisons to previously studied Bell inequalities.

- B. Opanchuk, S. Kiesewetter, P. D. Drummond, "Parallel optimized sampling for stochastic equations," SIAM Journal on Scientific Computing; accepted 5 October, 2016 (chapter 7)

  For this publication, I helped develop the algorithm together with the co-authors. The algorithm is the result of a number of previous attempts at a various reduction method which I tried in coordination with my principal supervisor as well as a gradual refinement process of the ideas presented in this publication through both numerical and theoretical work, which was done in close collaboration between me and the co-authors. During this stage, all authors of this publication wrote and used independent computer codes to verify the correctness of results and be able to try new ideas. The first set of graphs (Figs. 7.1–7.16) was created by Dr. Bogdan Opanchuk and the second set (7.17–7.21) by me. For these, among other things, I wrote GPU-capable implementation of our algorithm to achieve high-precision results.

- S. Kiesewetter, P. D. Drummond, "Algorithms for integration of stochastic differential equations using parallel optimized sampling in the Stratonovich calculus," Computer Physics Communications; accepted 1 November 2016 (chapter 8)

For this publication, I developed the algorithm in close collaboration with my principal supervisor. Through an intensive numerical and theoretial effort, I tried different strategies to combine the ideas of the previous paper with the Stratonovich calculus. After a working algorithm was developed, I then wrote a GPU-capable implementation and ran the simulations for the sample equations presented in the paper.

1.1 Additional papers not part of the thesis

- L. Rosales-Zárate, R. Y. Teh, S. Kiesewetter, A. Brolis, K. Ng, and M. D. Reid, "Decoherence of Einstein–Podolsky–Rosen steering," Journal of the Optical Society of America B **32**, pp. A82-A91; accepted March 2, 2015

Chapter 2

Introduction

The study of nonclassical systems is one of the most active areas of research in physics today, from both an experimental and theoretical point of view. Recent years have seen impressive developments in experimental methods, which have made it possible to realize quantum systems in experiments which could previously only be studied theoretically.

I will begin my thesis with a brief literature review on the early history of quantum physics and more recent experimental developments related to the nonclassical systems which are studied in my papers, before providing an overview of these papers.

2.1 Literature review on nonclassical systems

The first observation and systematic study of a quantum effect is usually considered to be a series of experiments carried out by Balfour Stewart in 1858[1] and independently by Gustav Kirchhoff in 1859[2] on black-body radiation which was first interpreted as a quantum effect by Max Planck in 1900[3]. Another key experiment has been the photoelectric effect, first observed by Heinrich Hertz in 1887[4] and first interpreted as a quantum effect by Albert Einstein in 1905[5], for which he was awarded the Nobel prize in 1921[6].

Although a range of quantum effects can be observed in a room-temperature environment in relatively simple experiments, many other effects in established and emerged areas require a low-temperature environment, a precise setup of microscopic extent, sophisticated trapping techniques, highly sensitive measurement devices and/or control over system parameters and measurements on very short or ultra-short timescales.

Advancements in the field of cryogenic methods play an important part in this ongoing transition from purely theoretical to experimentally realizable systems. One of the most important such cryogenic techniques is the use of laser fields for cooling. These methods are referred to by the umbrella term 'laser cooling'. The first such laser cooling technique, called Doppler cooling, was proposed by Wineland and Dehmelt[7] and Hänsch and Schawlow[8] independently and experimentally realized by Wineland *et al.*[9]. Although different methods of laser cooling have been developed, Doppler cooling remains the most commonly used method[10]. This technique uses laser fields which will produce atomic excitations in trapped sample atoms only if the effective laser frequency is elevated by the Doppler effect when the sample atoms are thermally moving towards the laser source. Further improvements in trapping and cooling techniques have been honored with the Nobel prize in physics in 1997[11], which was awarded to William Phillips, Claude Cohen-Tannoudji and Steven Chu.

An open question since the inception of quantum physics has been its very nature and how it relates to reality. The Copenhagen interpretation[12–15], arguably the most widely accepted interpretation of quantum physics today, was initially received with a lot of skepticism by leading physicists. In 1935, in an influential paper[16], Einstein, Podolsky and Rosen formulated the thought experiment later called "EPR-paradox", conjecturing that quantum mechanics was incomplete. In 1964, Bell[69] gave a counter-argument, demonstrating how EPR's assumption of local realism would lead to results that could be falsified in experiments. For a detailed explanation, refer to sections 3.3.2 and 3.3.3. Several experiments have been carried out to test the argument put forth by Bell, however, up until recently, these always contained at least one loophole and thus could only be considered a partial proof of Bell's argument. It was not until 2015 that an experiment[17] was carried out which is considered truly loophole-free, fully supporting Bell's argument.

An intriguing development in the field of nonclassical systems has been the emergence of the field of optomechanics. In an optomechanical system, the interaction between light and mechanical components due to radiation pressure plays a central role. Among other things, this allows for the study of inherently quantum effects on a mesoscopic physical scale. For a more in-depth description, refer to section 3.4. One of the first groundbreaking works in the field of optomechanics was achieved by Braginsky *et al.*[18,

19]. They successfully demonstrated how in an electromechanical cavity, radiation pressure can be used to achieve damping or anti-damping of the motion of a mechanical oscillator. The first optomechanical experiment involving radiation pressure at optical frequencies was carried out by Dorsel *et al.*[20], who successfully observed optical bistability arising from radiation pressure in an optomechanical cavity. Since then, there have been substantial advances in the fabrication techniques of micromechanical and nanomechanical devices, which have enabled the realization of optomechanical experiments on microscopic scales and led to a dramatic increase in the number of optomechanical experiments and their successes. For example, in 2006, three independent experiments were carried out to demonstrate cavity cooling via radiation pressure[21–23]. In 2011, Chan *et al.*[24] successfully used optomechanical techniques to cool down an optomechanical cavity to its quantum ground-state. For a comprehensive review on cavity optomechanics, see for example [25] and [26].

2.2 Overview of the papers as part of the thesis

These numerous experimental advances, of which I have barely scratched the surface here, lead to a rapid growth of experimentally available data. This sometimes leads to a problem when it comes to interpreting the experimental data: non-classical systems, more often than classical ones, can be highly nonlinear, highly complex or both. In other words, making accurate predictions about experiments involving nonclassical systems can be challenging by itself. This problem is made more difficult by the fact that the quantum Hilbert space dimension typically grows exponentially with the system size, so that traditional expansions in an orthogonal basis can be very large and time-consuming to calculate.

In this situation, one resorts to one of two strategies or a combination of the two:

- Instead of the full problem in its original form, one investigates an abstraction or simpler version of it, based on reasonable and justified assumptions, an approach as old as physics itself.

- One uses numerical methods, typically with the help of computers. If only minimal assumptions are made and numerical methods are exploited intensively instead, such an approach is usually called a simulation by first principles.

While the modeling approach, that is, making approximations where justified, is clearly an invaluable tool, any approximation carries the risk of errors. Furthermore, it is often not clear whether an approximation is justified or not. This is especially the case when the system to be studied is complex in nature, not yet fully or even well understood and predictions with high accuracy are desired. In many cases, a solid estimate regarding the error introduced by the approximation cannot be made.

In my thesis, I analyze several types of nonclassical systems in active research areas and how they can be simulated by first principles. My main focus is on physical phenomena which exhibit nonlocality features, such as entanglement, EPR-steering and Bell violations. Among other things, I analyze how these phenomena can be generated, stored and transferred in the emerging field of optomechanics. Furthermore, I analyze currently established numerical methods and propose an original method called "Parallel optimized sampling" (POS), which allows for the solution of stochastic differential equations with substantially reduced sampling errors.

In the paper "Scalable quantum simulation of pulsed entanglement and Einstein-Podolsky-Rosen steering in optomechanics" (chapter 4), we analyze an experiment first proposed by Hofer *et al.*[27]. The proposed experiment predicts entanglement between the mechanical oscillator of an optomechanical cavity and an impinging laser field. The experiment was carried out successfully in the electrodynamic domain by Palomaki *et al.*[28]. While Hofer *et al.* and Palomaki *et al.* used linearized system dynamics for their predictions, we use exact phase-space methods to simulate the system to confirm both of their findings. We demonstrate the effects that a full nonlinear treatment of the system dynamics leads to. Where Hofer *et al.* and Palomaki *et al.* only used one entanglement criterion, we consider multiple entanglement criteria, demonstrating the superiority of one over the others. Furthermore, we analyze the prospect of observing EPR-steering in the same optomechanical and electromechanical setup. The excellent agreement of first principles simulations with these experiments verifies our methods, which have the advantage that they can be extended into more nonlinear and nonclassical regimes in the future.

In the paper "Einstein-Podolsky-Rosen quantum simulations in nonclassical phase-space" (chapter 5), we give a review of how EPR-steering and Bell inequality violation can be simulated numerically. We first recapitulate the EPR paradox and phase-space

methods and how these can be used in quantum simulations. We summarize a model[29–31] that has been realized experimentally[32] which allows observing EPR steering for continuous variables. We review a proposed experiment[33] in which Bell violations can be observed using parametric down-conversion and how it can be simulated numerically. We further revise the experiment described in the previous paper. We analyze the effect of increasing the nonlinear coupling constant. We demonstrate how, for a sufficiently large value, this will lead to different results when simulated using linearized system dynamics, a truncated-Wigner method and a positive-P method. We further evaluate how quantum field simulations can be used to predict EPR-steering, exemplified with the planar, nondegenerate quantum parametric oscillator.

In the paper "Violations of multisetting quaternion and octonion Bell inequalities" (chapter 6), we study a new class of Bell inequalities, first derived by Cavalcanti *et al.*[34] and subsequently generalized by Shchukin and Vogel[35]. This type of Bell inequality allows for up to 8 measurement settings and arbitrarily many detection sites. We analyze how Bell violation can be observed in an experiment involving GHZ-states and generalized GHZ-states. This gives rise to the question of what the optimal detector settings are, resulting in potentially hundreds of different detector angles which have to be chosen optimally. Using numerical methods, we work out the exact detector settings that result in optimal Bell violations and elucidate the optimal settings for an arbitrary number of sites. We present the minimum detector efficiencies necessary for observing Bell violations and its dependence on the number of detection sites. We further analyze the effect of experimental errors in the detector settings on the prospect of observing Bell violations.

In the paper "Parallel optimized sampling for stochastic equations" (POS) (chapter 7), an original method for the solution of Itō-type stochastic differential equations is introduced that results in substantially reduced sampling errors compared to most traditional methods. We begin by analyzing the error dependence of a traditional, direct approach with the number of time-steps and stochastic samples used before deriving an effective integration order for an optimally chosen number of time-steps and number of samples. After that, our method (POS) is introduced and explained in detail. Two versions of the POS method, which we call "combined optimization" and "individual optimization" are presented. We analyze the error propagation of the POS algorithm and demonstrate how

it can also be used to sample a probability density function with dramatically reduced noises for a set of given observables. The POS method is applied to a number of linear and nonlinear one-dimensional test cases and its ability to achieve substantially reduced sampling errors is demonstrated. We also apply POS to the complex-valued, thereby two-dimensional laser equation, demonstrating that it is capable of treating higher-dimensional stochastic differential equations as well and its usefulness for physically relevant problems.

In the paper "Algorithms for integration of stochastic differential equations using parallel optimized sampling in the Stratonovich calculus" (chapter 8), we introduce a technique based on the approach described in the previous paper, which is capable of integrating Stratonovich-type stochastic differential equations resulting in substantially reduced sampling errors. This approach has the advantage that it has reduced time-step errors and improved stability compared to the Ito-Euler approach described in the previous paper. After a brief review of stochastic differential equations as well as the Itō and Stratonovich calculus, our algorithm is presented and explained in detail. It is applied to a linear problem and the results are compared against those obtained by using the Itō-version of the POS method presented in the previous paper. This POS technique is applied to a number of nonlinear systems, including the two-dimensional laser equation, demonstrating its usefulness. For one of the systems analyzed, we compare the results obtained using the Stratonovich-type POS method against those obtained using a truncated cumulant hierarchy method, clearly demonstrating how the truncated cumulant hierarchy approach can result in large systematic errors. Our algorithm yields results superior to both those obtained by a truncated cumulant hierarchy approach and a traditional, direct approach. Appendix A represents an addendum to this paper. In this appendix, I inspect one of the systems presented in the paper more closely by means of different error measures and a reduced time-step. The results are analyzed and discussed. The material presented in appendix A is original work and unpublished so far.

Chapter 3

Foundations

In this chapter, I will mostly give background information about well-known concepts and methods, except for the last section (section 3.5), where I will explain the basic concepts of an original method developed by us called "Parallel optimized sampling" (POS), which is explained in great detail in chapters 7 and 8. It is important to note that the material presented in section 3.5 is based on original work.

3.1 Stochastic methods

3.1.1 Fokker-Planck equation and stochastic differential equation

A Fokker-Planck equation is a second-order partial differential equation which describes the time-evolution of a probability density function. For the univariate case, it has the form

$$\frac{\partial}{\partial t} p(x,t) = -\frac{\partial}{\partial x}\left[\mu(x,t)\, p(x,t)\right] + \frac{\partial^2}{\partial x^2}\left[\frac{1}{2} D(x,t)\, p(x,t)\right] . \tag{3.1}$$

The term $\mu(x,t)$ is called drift and the term $D(x,t)$ is called diffusion. $D(x,t)$ needs to be nonnegative, hence can be expressed via $D(x,t) \equiv \sigma^2(x,t)$.

It can be shown that the univariate Fokker-Planck equation described by eq. (3.1) is equivalent to the univariate Itō-type stochastic differential equation (SDE)

$$dX(t) = \mu(X(t), t)\, dt + \sigma(X(t), t)\, \mathrm{d}W(t) . \tag{3.2}$$

Here, $X(t)$ is a stochastic, that is, nondeterministic variable subject to a deterministic drift, corresponding to the term $\mu(X(t), t)\, dt$ and a nondeterministic diffusion, corresponding to the term $\sigma(X(t), t)\, \mathrm{d}W(t)$. The term $W(t)$ itself represents a standard Wiener process, also known as standard Brownian motion process. A Wiener process is a nondeterministic, continuous but nondifferentiable Gaussian process satisfying the conditions

$$W(0) = 1 \text{ almost surely} \tag{3.3}$$

$$\mathbb{E}[W(t)] = 0 \tag{3.4}$$

$$\mathrm{Var}(W(t) - W(s)) = t - s \tag{3.5}$$

for all $0 \leq s \leq t$. It has independent increments, in other words, the statistical distribution of $W(t+s) - W(s)$ does not depend on s. It follows that the Wiener process $W(t)$ satisfies

$$W(t)\, W(t') = \delta(t - t')\,. \tag{3.6}$$

A multivariate Fokker-Planck equation is of the form

$$\frac{\partial}{\partial t} p(\boldsymbol{x}, t) = -\sum_{i=1}^{N} \frac{\partial}{\partial x_i} \left[\mu_i(\boldsymbol{x}, t)\, p(\boldsymbol{x}, t)\right] + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\partial^2}{\partial x_i \partial x_j} \left[D_{ij}(\boldsymbol{x}, t)\, p(\boldsymbol{x}, t)\right], \tag{3.7}$$

where $\boldsymbol{x} \equiv (x_1, x_2, ..., x_N)$. The deterministic drift is described by the N-component drift vector $\boldsymbol{\mu}(\boldsymbol{x}, t)$ and the nondeterministic diffusion by the $N \times N$-component diffusion matrix $\boldsymbol{D}(\boldsymbol{x}, t)$. $\boldsymbol{D}(\boldsymbol{x}, t)$ has to be positive semi-definite. It follows that the diffusion matrix can be obtained from an appropriately chosen $N \times M$-component matrix $\boldsymbol{\sigma}$ via

$$D_{ij}(\boldsymbol{x}, t) = \sum_{k=1}^{M} \sigma_{ik}(\boldsymbol{x}, t)\, \sigma_{jk}(\boldsymbol{x}, t)\ . \tag{3.8}$$

Note that the choice for $\boldsymbol{\sigma}$ is not unique.

The corresponding multivariate stochastic differential equation is:

$$\mathrm{d}\boldsymbol{X} = \boldsymbol{\mu}(\boldsymbol{x}, t)\, \mathrm{d}t + \boldsymbol{\sigma}(\boldsymbol{x}, t)\, \mathrm{d}\boldsymbol{W}(t)\ , \tag{3.9}$$

where $\mathrm{d}\boldsymbol{W} \equiv (\mathrm{d}W_1(t), \mathrm{d}W_2(t), ..., \mathrm{d}W_M(t))$ and the M Wiener processes are independent, i.e.

$$\mathrm{d}W_i \mathrm{d}W_j = \delta_{ij} \mathrm{d}t \,. \tag{3.10}$$

3.1.2 Itō and Stratonovich calculus

Due to its nondeterministic nature, a stochastic differential equation is fundamentally different from an ordinary differential equation and its solution cannot be described using ordinary Riemann integration methods. However, it can be described using the concept of a Riemann-Stieltjes integral[36, 37]. This is a generalization of the Riemann integral which shares conceptual similarities to the Riemann integral, particularly the fact that the integration domain is divided up into disjoint intervals and the integral results from taking the limit of infinitely many intervals. An Itō-type stochastic differential equation is defined as a stochastic differential equation whose solution is obtained through the Itō integral. The Itō integral is defined as a Rieman-Stieltjes integral where the diffusion coefficients are evaluated at the beginning of the intervals in the integration domain. To illustrate this, consider the Itō-type univariate martingale process $\mathrm{d}X(t) = \sigma(t)\,\mathrm{d}W(t)$. Its solution is the result of taking the limit

$$X = \lim_{N\to\infty} \sum_{i=1}^{N-1} \sigma(t_i)\left[W_{t+1} - W_t\right] \,. \tag{3.11}$$

Sometimes, the solution to a stochastic differential equation is indicated by an integral, e.g.

$$X(t) = \int \mu(X(t), t)\,\mathrm{d}t + \int \sigma(X(t), t)\,\mathrm{d}W \,. \tag{3.12}$$

In that case, one should keep in mind that the integral is not an ordinary, that is, Riemann-type integral. In the Itō calculus, many rules of ordinary integral calculus do not apply. This can be seen easily by considering the Itō-type stochastic differential equation

$$dX(t) = W(t)\,\mathrm{d}W(t) \,. \tag{3.13}$$

Ordinary integration rules suggest that the solution to the equation above is

$$X(t) = \frac{1}{2}W(t)^2 \ . \tag{3.14}$$

However, this is not the case. By using eq. (3.11), one quickly finds that the true solution is

$$X(t) = \frac{1}{2}W(t)^2 - \frac{1}{2}t \, . \tag{3.15}$$

Closely related to this is Itō's lemma[38]. It describes how an Itō-type stochastic differential equation behaves under variable transformation. For the univariate case with one Wiener process, Itō's lemma states

Theorem 3.1. *Let $X(t)$ be a stochastic variable subject to the Itō-type stochastic differential equation*

$$\mathrm{d}X(t) = \mu(\mathbf{X}, t)\,\mathrm{d}t + \sigma(X, t)\,\mathrm{d}W \, . \tag{3.16}$$

Let $f(X, t)$ be a twice-differentiable function. Then $Y(t) \equiv f(X(t), t)$ is subject to the Itō-type stochastic differential equation

$$\mathrm{d}Y(t) = \left(\frac{\partial f}{\partial t} + \mu(X, t)\frac{\partial f}{\partial X} + \frac{\sigma(X, t)^2}{2}\frac{\partial^2 f}{\partial X^2}\right)\mathrm{d}t + \sigma(X, t)\frac{\partial f}{\partial X}\mathrm{d}W(t) \ . \tag{3.17}$$

For the multivariate case with arbitrarily many Wiener processes, it states[37]

Theorem 3.2. *Let $\mathbf{X}(t) \equiv (X(1), (2), ..., (N))$ be a multivariate stochastic variable subject to the Itō-type stochastic differential equation*

$$\mathrm{d}\mathbf{X}(t) = \boldsymbol{\mu}(\mathbf{X}, t)\,\mathrm{d}t + \boldsymbol{\sigma}(\mathbf{X}, t)\,\mathrm{d}\mathbf{W}(t) \ , \tag{3.18}$$

where $\mu(\mathbf{X}, t) \equiv (\mu_1(\mathbf{X}, t), \mu_2(\mathbf{X}, t), ..., \mu_N(\mathbf{X}, t))$, $\boldsymbol{\sigma}(\mathbf{X}, t)$ is an $N \times M$-component matrix with $\sigma_{ij} = \sigma_{ij}(\mathbf{X}, t)$ and $\mathrm{d}\mathbf{W} \equiv (\mathrm{d}W_1, \mathrm{d}W_2, ..., \mathrm{d}W_M)$, where $W_1, W_2, ..., W_M$ are independent Wiener processes. Let $f(\mathbf{X}, t) : \mathbb{R}^N \to \mathbb{R}$ be a twice-differentiable function.

Then $Y(t) \equiv f(\boldsymbol{X}(t), t)$ *is subject to the Itō-type stochastic differential equation*

$$\mathrm{d}Y(t) = \left(\frac{\partial f}{\partial t} + (\nabla_{\boldsymbol{X}} f)^T \cdot \boldsymbol{\mu}(\boldsymbol{X}, t) + \frac{1}{2}\mathrm{Tr}\left[\boldsymbol{\sigma}(\boldsymbol{X}, t)^T (H_{\boldsymbol{X}} f)\, \boldsymbol{\sigma}(\boldsymbol{X}, t)\right]\right)\mathrm{d}t \tag{3.19}$$

$$+ (\nabla_{\boldsymbol{X}} f)^T \cdot \boldsymbol{\sigma}(\boldsymbol{X}, t) \cdot \mathrm{d}\boldsymbol{W}(t)\ , \tag{3.20}$$

where $(\nabla_{\boldsymbol{X}} f$ *is the gradient of* $f(\boldsymbol{X}, t)$ *w.r.t.* $\boldsymbol{X}$ *and* $H_{\boldsymbol{X}} f$ *is the Hessian matrix of* $f(\boldsymbol{X}, t)$ *w.r.t.* $\boldsymbol{X}$.

Note that in the multivariate stochastic differential equation presented above, the diffusion matrix $\boldsymbol{\sigma}(\boldsymbol{X}, t)$ need not be a diagonal matrix, hence the number of Wiener processes need not be equal to the number of stochastic variables.

While the choice of evaluating the diffusion coefficient at the beginning of the interval in the integration domain is a natural one, it is not imperative. Instead, the diffusion coefficient can be chosen to be evaluated anywhere else in the interval. The Stratonovich integral is defined as the Riemann-Stieltjes integral where the diffusion coefficient is evaluated in the middle of the interval. A Stratonovich-type stochastic differential equation is defined as a stochastic differential equation whose solution is obtained through the Stratonovich-integral. This means that the solution to the Stratonovich-type univariate martingale process $\mathrm{d}X(t) = \sigma(t)\,\mathrm{d}W(t)$ is

$$X = \lim_{N\to\infty} \sum_{i=1}^{N-1} \sigma\left(\frac{t_{i+1} + t_i}{2}\right)\left[W_{t+1} - W_t\right]\ . \tag{3.21}$$

A Stratonovich-type stochastic differential equation and a Stratonovich integral are sometimes indicated through the symbol $\circ$ in front of the Wiener process increment $\mathrm{d}W(t)$ or $\mathrm{d}\boldsymbol{W}(t)$. For example, a univariate Stratonovich-type stochastic differential equation would be indicated via

$$\mathrm{d}X = \mu(X, t)\,\mathrm{d}t + \sigma(X, t) \circ \mathrm{d}W(t) \tag{3.22}$$

and its solution via

$$X(t) = \int \mu(X, t)\,\mathrm{d}t + \int \sigma(X, t) \circ \mathrm{d}W\,. \tag{3.23}$$

While the Stratonovich integral is conceptually more complicated than the Itō integral, it allows for the application of some of the rules of ordinary integration which do not apply to the Itō integral. For example, the Stratonovich-type stochastic differential equation $\mathrm{d}X = W(t) \circ \mathrm{d}W(t)$ has the solution

$$X(t) = \frac{1}{2} W(t)^2 \ , \tag{3.24}$$

in accordance with ordinary integration rules. Also, a stochastic differential equation transforms more "conventionally" in the Stratonovich calculus compared to the Itō calculus. The univariate stochastic differential equation

$$\mathrm{d}X = \mu(X,t)\,\mathrm{d}t + \sigma(X,t) \circ \mathrm{d}W(t) \tag{3.25}$$

results in

$$\mathrm{d}Y(t) = \left(\frac{\partial f}{\partial t} + \mu(X,t) \frac{\partial f}{\partial X} \right) \mathrm{d}t + \sigma(X,t) \frac{\partial f}{\partial X} \circ \mathrm{d}W(t) \ . \tag{3.26}$$

where $Y(t) \equiv f(X,t)$. Note that the somewhat extraneous term $\frac{\sigma(X,t)^2}{2} \frac{\partial^2 f}{\partial X^2} \mathrm{d}t$, which is present in Itō's lemma, is missing in the Stratonovich calculus. A multivariate stochastic differential equation transforms according to the standard chain rule formula in the Stratonovich calculus.

An advantage of the Stratonovich calculus is that numerical integration algorithms typically exhibit a better convergence, that is, smaller time-step errors when compared to algorithms for Itō integrals. This can be understood when considering that the Stratonovich integral is an implicit integral, which requires multi-step algorithms for a solution, which will result in a higher-order dependence of $\mathrm{d}t$. This is true for the most typical algorithms, however algorithms for the solution of Itō-type stochastic differential equations, which have a higher-order dependence of $\mathrm{d}t$ also exist.

Itō-type stochastic differential equations and Stratonovich-type stochastic differential equations can be converted into each other. The univariate Itō-type stochastic differential equation

$$dX = \mu(X,t)\,\mathrm{d}t + \sigma(X,t)\,\mathrm{d}W(t) \tag{3.27}$$

corresponds to the Stratonovich-type stochastic differential equation

$$dX = \left(\mu\left(X,t\right) - \frac{1}{2}\frac{\partial\sigma\left(X,t\right)}{\partial X}\sigma\left(X,t\right) \right) \mathrm{d}t + \sigma\left(X,t\right) \circ \mathrm{d}W\left(t\right) \ . \tag{3.28}$$

The term $\frac{1}{2}\frac{\partial\sigma(X,t)}{X}\sigma\left(X,t\right)\mathrm{d}t$ is sometimes called "Stratonovich-correction".

The multivariate Itō-type stochastic differential equation

$$\mathrm{d}\boldsymbol{X}\left(t\right) = \boldsymbol{\mu}\left(\boldsymbol{X},t\right)\mathrm{d}t + \boldsymbol{\sigma}\left(\boldsymbol{X},t\right)\mathrm{d}\boldsymbol{W}\left(t\right) \tag{3.29}$$

corresponds to the Stratonovich-type stochastic differential equation

$$\mathrm{d}\boldsymbol{X}\left(t\right) = \tilde{\boldsymbol{\mu}}\left(\boldsymbol{X},t\right)\mathrm{d}t + \boldsymbol{\sigma}\left(\boldsymbol{X},t\right) \circ \mathrm{d}\boldsymbol{W}\left(t\right) \ , \tag{3.30}$$

where

$$\tilde{\mu}_i\left(\boldsymbol{X},t\right) \equiv \mu_i\left(\boldsymbol{X},t\right) - \frac{1}{2}\sum_{j=1}^{N}\sum_{k=1}^{N}\frac{\partial\sigma_{ij}\left(\boldsymbol{X},t\right)}{\partial X_k}\sigma_{kj}\left(\boldsymbol{X},t\right) \ . \tag{3.31}$$

3.1.3 Numerical methods

Similar to a conventional partial differential equation, the exact solution to a stochastic differential equation can be found only in exceptional cases using analytic methods. In cases where the exact solution cannot be found using analytic methods, numerical methods can be used to find an approximation to the exact solution.

Typically, this is done through a dual discretization of the time-dependent stochastic variable $X\left(t\right)$ or $\boldsymbol{X}\left(t\right)$, which corresponds to a time-dependent probability density function $P\left(X,t\right)$ or $P\left(\boldsymbol{X},t\right)$:

- The time domain is partitioned into a finite number of intervals whose edges $t_1 < t_2 < \ldots < t_N$ are called time-steps
- The probability density function is approximated using a finite number of elements, called samples.

The samples are chosen for the initial state so that they are distributed according to the probability density function of the initial state. They are then evolved in time so that they continue to approximate the time-evolving probability density function corresponding to

the stochastic differential equation. In most conventional methods, some of which will be described in this chapter, this is done by evolving the samples independently according to the stochastic differential equation. The method of parallel optimized samples (POS), which has been developed by myself and co-authors, takes a different approach and evolves the samples such that their time-evolutions are not independent of each other. This is explained in detail in section 3.5 and in the attached publications in chapters 7 and 8.

Weak and strong convergence

The convergence of a stochastic algorithm can be measured in different ways. Two common criteria are the weak convergence criterion and the strong convergence criterion[37].

Consider a stochastic process $X(t)$. Suppose a given algorithm produces the time-discretized solutions $\tilde{X}(t_i)$ for a given time-step size Δt. In that case, the algorithm is said to be of weak convergence order n if a constant k exists, so that

$$\left|\mathbb{E}\left[X(t_i)\right] - \mathbb{E}\left[\tilde{X}(t_i)\right]\right| \leq k(\Delta t)^n \tag{3.32}$$

and of strong convergence order n if a constant k exists, so that

$$\mathbb{E}\left[\left|X(t_i) - \tilde{X}(t_i)\right|\right] \leq k(\Delta t)^n \ . \tag{3.33}$$

In other words, weak convergence measures the error of the mean while strong convergence measures the mean of the error. In many ways, the strong convergence criterion is the more informative one between the two, particularly when higher-order moments are of interest. Because of that, I will only make reference to the strong convergence of algorithms I present.

Algorithms

A multitude of conventional algorithms, that is, algorithms which involve independent evolution of samples, exist. Depending on whether the stochastic differential equation at hand is Itō-type or Stratonovich-type, the algorithm has to be chosen accordingly. In other words, some algorithms are suitable for Itō-type stochastic differential equations

and some are suitable for Stratonovich-type ones. I will present a selected number of the most common algorithms.

Euler-Mayurama algorithm

The Euler-Mayurama algorithm[37], also called the Euler method, is suitable for the solution of Itō-type stochastic differential equations. It is arguably the simplest example of the algorithms presented in this chapter. Consider the univariate Itō-type stochastic differential equation

$$\mathrm{d}X = \mu\left(X, t\right)\mathrm{d}t + \sigma\left(X, t\right)\mathrm{d}W\left(t\right) \ . \tag{3.34}$$

The stochastic variable X is sampled using N_S elements (samples) according to the probability density function of the initial state. Let $\boldsymbol{X} \equiv \left(X_1, X_2, ..., X_{N_S}\right)$ be a vector composed of the N_S samples. The time-domain is partitioned evenly, resulting in $N+1$ discrete points in time $t_0 < t_1 < ... < t_N$. The separation between the time points $\Delta t \equiv t_{i+1} - t_i$ is called the time step. Using the Euler-Mayurama method, the sample vector at time t_{i+1} follows from the sample vector at time t_i via

$$\boldsymbol{X}\left(t_{i+1}\right) = \boldsymbol{X}\left(t_i\right) + \boldsymbol{\mu}\left(\boldsymbol{X}\left(t_i\right), t_i\right)\Delta t + \boldsymbol{\sigma}\left(\boldsymbol{X}\left(t_i\right), t_i\right)\Delta\boldsymbol{W}, \tag{3.35}$$

where $\Delta\boldsymbol{W} \equiv \left(\Delta W_1, \Delta W_2, ..., \Delta W_N\right)$ is a vector of independent realizations of a normally distributed random variable with mean $\mathbb{E}\left[W_i\right] = 0$ and variance $\mathbb{E}\left[W_i^2\right] = \Delta t$. The vector $\Delta\boldsymbol{W}$ is also called the "noise increment".

As can be seen, in the absence of a stochastic term, the Euler-Mayurama algorithm reduces to a forward Euler scheme for ordinary differential equations.

A generalization to multivariate stochastic differential equation follows naturally by extending the sample vector $\boldsymbol{X}$ to include all stochastic variables.

The Euler-Mayurama algorithm has strong convergence order 0.5.

Heun's method

Heun's method[37] is suitable for Stratonovich-type stochastic differential equations. Given the univariate stochastic differential equation

$$\mathrm{d}X = \mu\left(X, t\right)\mathrm{d}t + \sigma\left(X, t\right) \circ \mathrm{d}W\left(t\right) \ , \tag{3.36}$$

the sample vector $\boldsymbol{X}\left(t_{i+1}\right)$ follows from $\boldsymbol{X}\left(t_i\right)$ via

$$\bar{\boldsymbol{X}}\left(t_{i+1}\right) = \boldsymbol{X}\left(t_i\right) + \boldsymbol{\mu}\left(\boldsymbol{X}\left(t_i\right), t_i\right)\Delta t + \boldsymbol{\sigma}\left(\boldsymbol{X}\left(t_i\right), t_i\right)\Delta \boldsymbol{W} \tag{3.37}$$

$$\begin{aligned}\boldsymbol{X}\left(t_{i+1}\right) = \boldsymbol{X}\left(t_i\right) &+ \frac{1}{2}\left[\boldsymbol{\mu}\left(\boldsymbol{X}\left(t_i\right), t_i\right) + \boldsymbol{\mu}\left(\bar{\boldsymbol{X}}\left(t_{i+1}\right), t_{i+1}\right)\right]\Delta t \\ &+ \frac{1}{2}\left[\boldsymbol{\sigma}\left(\boldsymbol{X}\left(t_i\right), t_i\right) + \boldsymbol{\sigma}\left(\bar{\boldsymbol{X}}\left(t_{i+1}\right), t_{i+1}\right)\right]\Delta \boldsymbol{W},\end{aligned} \tag{3.38}$$

where the variable $\bar{\boldsymbol{X}}\left(t_{i+1}\right)$ is an intermediate step. Note that the noise increments in eqs. (3.37) and (3.38) are identical.

Heun's method has strong convergence order 0.5.

Runge-Kutta method

As is the case for ordinary differential equations, the term "Runge-Kutta method" describes a class of methods, which can be classified systematically, rather than a single method. In this section, I focus on a specific algorithm, which based on the classical 4th-order Runge-Kutta method. The 4th-order Runge-Kutta method is usually regarded an excellent trade-off between computational effort and achievable accuracy.

The stochastic 4th-order Runge-Kutta method[37] is suitable for a Stratonovich-type stochastic differential equation. Given the stochastic differential equation eq. (3.36), the

sample vector $\boldsymbol{X}(t_{i+1})$ follows from $\boldsymbol{X}(t_i)$ via

$$\boldsymbol{K}_1 = \boldsymbol{\mu}(\boldsymbol{X}(t_i), t_i)\,\Delta t + \boldsymbol{\sigma}(\boldsymbol{X}(\boldsymbol{t_i}), t_i)\,\Delta \boldsymbol{W}, \tag{3.39}$$

$$\boldsymbol{K}_2 = \boldsymbol{\mu}\left(\boldsymbol{X}(t_i) + \frac{1}{2}\boldsymbol{K}_1,\ t_i + \frac{\Delta t}{2}\right)\Delta t + \boldsymbol{\sigma}\left(\boldsymbol{X}(t_i) + \frac{1}{2}\boldsymbol{K}_1,\ t_i + \frac{\Delta t}{2}\right)\Delta \boldsymbol{W}, \tag{3.40}$$

$$\boldsymbol{K}_3 = \boldsymbol{\mu}\left(\boldsymbol{X}(t_i) + \frac{1}{2}\boldsymbol{K}_2,\ t_i + \frac{\Delta t}{2}\right)\Delta t + \boldsymbol{\sigma}\left(\boldsymbol{X}(t_i) + \frac{1}{2}\boldsymbol{K}_2,\ t_i + \frac{\Delta t}{2}\right)\Delta \boldsymbol{W}, \tag{3.41}$$

$$\boldsymbol{K}_4 = \boldsymbol{\mu}(\boldsymbol{X}(t_i) + \boldsymbol{K}_3,\ t_{i+1})\,\Delta t + \boldsymbol{\sigma}(\boldsymbol{X}(t_i) + \boldsymbol{K}_3,\ t_{i+1})\,\Delta \boldsymbol{W}, \tag{3.42}$$

$$\boldsymbol{X}(t_{i+1}) = \boldsymbol{X}(t_i) + \frac{1}{6}(\boldsymbol{K}_1 + 2\boldsymbol{K}_2 + 2\boldsymbol{K}_3 + \boldsymbol{K}_4)\ , \tag{3.43}$$

where the variables $\boldsymbol{K}_1, \boldsymbol{K}_2, \boldsymbol{K}_3, \boldsymbol{K}_4$ are intermediate steps and the noise increments $\Delta \boldsymbol{W}$ are identical.

The Runge-Kutta presented here has strong convergence order 0.5.

Central-difference method

Another method which is suitable for Stratonovich-type stochastic differential equations is the central-difference method introduced by Drummond and Mortimer[39] as a weak semi-implicit variant of a family of stochastic algorithms. It uses a number M of intermediate steps, which can be chosen. Using the central-difference method, for the stochastic differential equation eq. (3.36) the sample vector $\boldsymbol{X}(t_{i+1})$ follows from $\boldsymbol{X}(t_i)$ via

$$\bar{\boldsymbol{X}}_0 \equiv \boldsymbol{X}(t_i) \tag{3.44}$$

$$\bar{\boldsymbol{X}}_1 = \boldsymbol{X}(t_i) + \frac{1}{2}\left[\boldsymbol{\mu}\left(\bar{\boldsymbol{X}}_0,\ t_i + \frac{\Delta t}{2}\right)\Delta t + \boldsymbol{\sigma}\left(\bar{\boldsymbol{X}}_0,\ t_i + \frac{\Delta t}{2}\right)\Delta \boldsymbol{W}\right] \tag{3.45}$$

$$\bar{\boldsymbol{X}}_2 = \boldsymbol{X}(t_i) + \frac{1}{2}\left[\boldsymbol{\mu}\left(\bar{\boldsymbol{X}}_1,\ t_i + \frac{\Delta t}{2}\right)\Delta t + \boldsymbol{\sigma}\left(\bar{\boldsymbol{X}}_1,\ t_i + \frac{\Delta t}{2}\right)\Delta \boldsymbol{W}\right] \tag{3.46}$$

$$\ldots$$

$$\bar{\boldsymbol{X}}_M = \boldsymbol{X}(t_i) + \frac{1}{2}\left[\boldsymbol{\mu}\left(\bar{\boldsymbol{X}}_{M-1},\ t_i + \frac{\Delta t}{2}\right)\Delta t + \boldsymbol{\sigma}\left(\bar{\boldsymbol{X}}_{M-1},\ t_i + \frac{\Delta t}{2}\right)\Delta \boldsymbol{W}\right] \tag{3.47}$$

$$\boldsymbol{X}(t_{i+1}) = \boldsymbol{X}(t_i) + \bar{\boldsymbol{X}}_M \tag{3.48}$$

For stochastic differential equations with commutative noise, it has strong convergence order 1. For stochastic differential equations with non-commutative noise, it has strong convergence order of at least 0.5. A stochastic differential equation with diffusion matrix

$\boldsymbol{\sigma}$ is said to have commutative noise iff

$$\sum_{i=1}^{M} \sigma_{i,j_1} \frac{\partial \sigma_{k,j_2}}{\partial x_i} = \sum_{i=1}^{M} \sigma_{i,j_2} \frac{\partial \sigma_{k,j_1}}{\partial x_i} \,. \tag{3.49}$$

Drummond and Mortimer[39] demonstrated that the central-difference method performs remarkably well in the case of the Kubo oscillator.

3.2 Phase-space methods

A complex quantum system refers to a system with a large Hilbert space dimension, typically involving many particles and modes, in contrast to a scenario with exactly one particle or mode.

The objective of the simulation of a complex quantum system is usually to determine the evolution of a quantum state that is subject to a specific set of equations, which describe its dynamics. On the surface, this may seem like a task that is in principle not more difficult than simulating a classical case, which, although might have its own challenges, is generally regarded as feasible and straight-forward in most cases. However, on further inspection, one finds that the situation is quite different for a complex quantum system. In the majority of cases, a straight-forward approach, though in principle not impossible, is unfeasible for even relatively simple cases. The reason for this is the rapid growth of the Hilbert space complexity with the complexity of the setup.

As an example, consider a bosonic case with a maximum occupation number of N_{max} bosons. In order to describe interesting phenomena, it is usually necessary to divide this further into a number of sub-systems, called *modes*. These might, for example, correspond to spatial modes, where the system is considered on a spatial grid, or excitation levels in an atomic or molecular system. In the simplest number representation, for a physical setup that is made up of M different modes, a pure state will be described by a total of $(N_{max} + 1)^M$ complex numbers, assuming the number of particles is not fixed. Because of this exponential dependence, the Hilbert space dimension quickly becomes prohibitively large for a brute-force type computer simulation. To illustrate this, consider a bosonic case with 12 modes and a maximum occupation number of 10. Assuming a size of 8 bytes for a complex number, the memory required to store a single state vector for this hypothetical scenario, despite its seemingly moderate complexity, would be 25 Gigabytes.

For a fermionic case, the Hilbert space complexity scales more slightly more favorably, because in this case, each mode has a maximum occupation number of 1. However, the exponential scaling with the number of modes remains.

In spite of these discouraging numbers, various techniques have been developed for the simulation of quantum systems. Here, I will focus on phase-space methods, which a large part of my research is based on. The most important feature of phase-space techniques is that the associated phase-space dimensionality exhibits a linear scaling with the number of modes, in contrast to the exponential scaling of a brute-force approach. The fact that it is possible to simulate a systems whose complexity grows exponentially with M using a method whose complexity grows linearly with M, might be surprising at first. The difference is that phase-space methods do not keep track of every single quantum state in the system, as in the case for a brute-force approach. Rather, they keep track of the sampled properties which is sufficient for the prediction of expectation values, not unlike a thermodynamic system, where a detailed knowledge of position and momentum of every single particle is not required to make predictions about meaningful observables.

3.2.1 Phase-space representation

Phase-space representations are mappings from a quantum density operator onto a continuous c-number function. Several different types of phase-space representation, distinguished among other things, by an associated operator ordering, exist, as will be explained later in this section. Some of these have associated c-number functions which satisfy the conditions of a probability density function, which is why the term "quasi-probability distribution" is also used in connection with general phase-space representations.

I will start by introducing a particular type of phase-space representation, namely the Glauber-Sudarshan P-representation, before outlining a way to generalize the Glauber-Sudarshan P-representation, which leads to different types of phase-space representations. Before that, I will briefly explain a coherent state.

3.2.2 Coherent state

A coherent state $|\alpha\rangle$, first derived by Schrödinger[40] and later described more systematically by Glauber[41], is defined as an eigenstate of the quantum annihilation operator,

that is, it satisfies the equation

$$\hat{a}\,|\alpha\rangle = \alpha\,|\alpha\rangle\ . \tag{3.50}$$

The eigenvalue α is in general a complex number. It can be represented in the Fock state basis via the infinite sum

$$|\alpha\rangle = e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}}\,|n\rangle\ . \tag{3.51}$$

The set of coherent states is not orthogonal, since its inner product evaluates to

$$\langle\beta|\,\alpha\rangle = e^{-\frac{1}{2}\left(|\beta|^2+|\alpha|^2-2\beta^*\alpha\right)}\ . \tag{3.52}$$

The set of coherent states is, however, an (overcomplete) spanning set, i.e.

$$\frac{1}{\pi}\int \mathrm{d}^2\alpha\ \ |\alpha\rangle\,\langle\alpha| = \mathbb{1}\,, \tag{3.53}$$

in other words, the coherent states constitute a non-orthogonal basis. Application of the creation operator to a coherent state results in

$$\hat{a}^\dagger\,|\alpha\rangle = \left(\frac{\partial}{\partial\alpha} + \frac{\alpha^*}{2}\right)|\alpha\rangle\ . \tag{3.54}$$

The quadrature components $\hat{X} \equiv \frac{\hat{a}+\hat{a}^\dagger}{2}$ and $\hat{P} \equiv \frac{\hat{a}-\hat{a}^\dagger}{2i}$ for a coherent state have quantum uncertainties

$$\Delta X \equiv \sqrt{\langle\alpha|\,\hat{X}^2\,|\alpha\rangle - \langle\alpha|\,\hat{X}\,|\alpha\rangle^2} = \frac{1}{2} \tag{3.55}$$

$$\Delta P \equiv \sqrt{\langle\alpha|\,\hat{P}^2\,|\alpha\rangle - \langle\alpha|\,\hat{P}\,|\alpha\rangle^2} = \frac{1}{2}\,. \tag{3.56}$$

For a coherent state with high occupation number (which corresponds to a large value of $|\alpha|$), relative quadrature uncertainties become small, which is a defining feature of a classical state. A squeezed coherent state exhibits a smaller uncertainty in one quadrature and greater uncertainty in the conjugate quadrature, while the uncertainty product remains unchanged.

3.2.3 Glauber-Sudarshan P-representation

Consider a one-mode bosonic system. The Glauber-Sudarshan P-representation, first introduced by Sudarshan[42] and Glauber[41], (which I will call the P-representation from now on) is motivated by the fact that the coherent states are an (over-)complete set of states, i.e.

$$\hat{\mathbb{1}} = \frac{1}{\pi} \int |\alpha\rangle \langle\alpha| \, \mathrm{d}^2\alpha \tag{3.57}$$

Glauber and Sudarshan proved that the density operator $\hat{\rho}$ of a system can be represented using the coherent state basis via

$$\hat{\rho} = \int \mathrm{d}^2\alpha \, P(\alpha, \alpha^*) \, |\alpha\rangle \langle\alpha| \, . \tag{3.58}$$

$P(\alpha, \alpha^*)$ is called the P-representation or sometimes just the P-function. The notation $P(\alpha, \alpha^*)$ is used to indicate that it need not be an analytic function. The P-representation is real-valued (which follows from the analyticity of $\hat{\rho}$) and satisfies $\int \mathrm{d}^2\alpha \, P(\alpha, \alpha^*) = 1$ (which follows from $\mathrm{Tr}(\hat{\rho}) = 1$). Despite this, $P(\alpha, \alpha^*)$ is not in general a probability density function, because it can be negative.

The P-representation can be singular, which can be easily seen by considering the density operator corresponding to a coherent state $\hat{\rho} = |\alpha_0\rangle \langle\alpha_0|$, in which case $P(\alpha, \alpha^*) = \delta^{(2)}(\alpha - \alpha_0)$. In general, it is even much more singular than that.

The expectation value $\mathrm{Tr}\left(\hat{\rho}\hat{O}\right)$ for an operator $\hat{O}\left(\hat{a}, \hat{a^\dagger}\right)$ can be obtained easily if it is expressed in its normally ordered form

$$\hat{O}_N\left(\hat{a}, \hat{a}^\dagger\right) = \sum_{n,m} c_{n,m} \left(\hat{a}^\dagger\right)^m \hat{a}^n \, , \tag{3.59}$$

in which case it follows that

$$\begin{aligned} \mathrm{Tr}\left(\hat{O}_N\hat{\rho}\right) &= \int \mathrm{d}^2\alpha \, P(\alpha, \alpha^*) \langle\alpha| \hat{O}_N |\alpha\rangle \\ &= \int \mathrm{d}^2\alpha \, P(\alpha, \alpha^*) \, O_N(\alpha, \alpha^*) \, . \end{aligned} \tag{3.60}$$

For a given density operator $\hat{\rho}$, the corresponding P-representation can be found by means of the normally-ordered characteristic function, which is defined as

$$C_1\left(\xi,\xi^*\right) \equiv \mathrm{Tr}\left(\hat{\rho}e^{\xi\hat{a}^\dagger}e^{-\xi^*\hat{a}}\right) . \tag{3.61}$$

The P-representation can then be obtained through the inverse Fourier transform of $C_1\left(\xi\right)$ via

$$P\left(\alpha,\alpha^*\right) = \frac{1}{\pi^2}\int \mathrm{d}^2\xi\, e^{\xi^*\alpha-\xi\alpha^*}C_1\left(\xi,\xi^*\right) . \tag{3.62}$$

3.2.4 s-ordering approach

The s-ordering approach, introduced by Cahill and Glauber[43], associates different types of phase-space representations, including the Glauber-Sudarshan P-representation, the Wigner representation (refer to section 3.2.7) and the Husimi Q-representation (refer to section 3.2.8), which had all been discovered previously, with a specific operator ordering. For a given operator ordering s, the quantum density operator can then be expressed as

$$\hat{\rho} = \int \mathrm{d}^2\alpha W_s\left(\alpha,\alpha^*\right)\hat{\Lambda}_s\left(\alpha,\alpha^*\right) , \tag{3.63}$$

where W_s is the corresponding phase-space representation. $\hat{\Lambda}_s$ is called the associated operator kernel. The associated characteristic functions $C_s\left(\xi,\xi^*\right)$ are

$$C_s\left(\xi,\xi^*\right) = \mathrm{Tr}\left(\hat{\rho}\hat{C}_s\left(\xi,\xi^*\right)\right) , \tag{3.64}$$

where

$$\hat{C}_s\left(\xi,\xi^*\right) = :\exp\left[\xi\hat{a}^\dagger - \xi^*\hat{a}\right]:_s , \tag{3.65}$$

where $::_s$ indicates an s-ordered product. For $s = 1,\ 0,\ -1$, the operator orderings are normal, symmetric and anti-normal ordering, respectively and the associated phase-space representations are the Glauber-Sudarshan P-representation, the Wigner representation and the Husimi Q-representation, respectively. The explicit expressions for $\hat{C}_s\left(\xi,\xi^*\right)$ in

the various orderings are

$$\begin{aligned}
\hat{C}_1(\xi,\xi^*) &= \exp\left[\xi\hat{a}^\dagger\right]\exp\left[-\xi^*\hat{a}\right] \\
\hat{C}_0(\xi,\xi^*) &= \exp\left[\xi\hat{a}^\dagger - \xi^*\hat{a}\right] \\
\hat{C}_{-1}(\xi,\xi^*) &= \exp\left[-\xi^*\hat{a}\right]\exp\left[\xi\hat{a}^\dagger\right]
\end{aligned} \tag{3.66}$$

Using the Campbell-Baker-Hausdorff[44] theorem, the equations above can be expressed as

$$\hat{C}_s(\xi,\xi^*) = e^{s|\xi|^2/2}\hat{C}(\xi,\xi^*)\ , \tag{3.67}$$

where $\hat{C}(\xi,\xi^*) \equiv \hat{C}_0(\xi,\xi^*)$.

The phase-space representations $W_s(\alpha,\alpha^*)$ can be obtained from eq. (3.64) using inverse Fourier transform via

$$P_s(\alpha,\alpha^*) = \frac{1}{\pi^2}\int \mathrm{d}^2\alpha e^{\xi^*\alpha-\xi\alpha^*}C_s(\xi,\xi^*)\ . \tag{3.68}$$

3.2.5 Multi-mode phase-space representations

For a system composed of M different modes, the density operator can be expressed as

$$\hat{\rho} = \hat{\rho}\left(\hat{\boldsymbol{a}},\hat{\boldsymbol{a}}^\dagger\right)\ , \tag{3.69}$$

where $\boldsymbol{a}$ indicates an M-dimensional operator vector $\boldsymbol{a} = (\boldsymbol{a}_1,...,\boldsymbol{a}_M)$. A generalization of the equations described in the previous chapter follows naturally via

$$\begin{aligned}
\hat{\rho} &= \int \mathrm{d}^{2M}\boldsymbol{\alpha}\, P_s(\boldsymbol{\alpha},\boldsymbol{\alpha}^*)\,\hat{\Lambda}_s(\boldsymbol{\alpha},\boldsymbol{\alpha}^*) \\
C_s(\boldsymbol{\xi},\boldsymbol{\xi}^*) &= \mathrm{Tr}\left(\hat{\rho}\hat{C}_s(\boldsymbol{\xi},\boldsymbol{\xi}^*)\right) \\
\hat{C}_s(\boldsymbol{\xi},\boldsymbol{\xi}^*) &= :\exp\left[\boldsymbol{\xi}\cdot\hat{\boldsymbol{a}}^\dagger - \boldsymbol{\xi}^*\cdot\hat{\boldsymbol{a}}\right]:_s \\
P_s(\boldsymbol{\alpha},\boldsymbol{\alpha}^*) &= \frac{1}{\pi^{2M}}\int \mathrm{d}^{2M}\boldsymbol{\xi}\, e^{\boldsymbol{\xi}^*\boldsymbol{\alpha}-\boldsymbol{\xi}\boldsymbol{\alpha}^*}
\end{aligned} \tag{3.70}$$

3.2.6 Observables

For an s-ordered phase-space representation, expectation values $\mathrm{Tr}\left(\hat{\rho}\hat{O}\right)$ can be extracted easily provided the operator $\hat{O}\left(\boldsymbol{a},\boldsymbol{a}^{\dagger}\right)$ is expressed in an s-ordered form $\hat{O}_s\left(\boldsymbol{a},\boldsymbol{a}^{\dagger}\right) = :\hat{O}\left(\boldsymbol{a},\boldsymbol{a}^{\dagger}\right):_s$. In that case, it follows that

$$\mathrm{Tr}\left(\hat{\rho}\hat{O}\right) = \int \mathrm{d}^{2M}\boldsymbol{\alpha}\, W_s\left(\boldsymbol{\alpha},\boldsymbol{\alpha}^*\right) O_s\left(\boldsymbol{\alpha},\boldsymbol{\alpha}^*\right) \tag{3.71}$$

3.2.7 Wigner representation

The Wigner representation, also called Wigner function, first introduced by Wigner[45], is usually indicated by $W\left(\boldsymbol{\alpha},\boldsymbol{\alpha}^*\right)$ and corresponds to a symmetric operator ordering, that is $s = 0$.

Upon further inspection, it can be shown that the Wigner function can be obtained from the corresponding Glauber-Sudarshan P-function by convolution with a Gaussian function via

$$W\left(\alpha,\alpha^*\right) = \frac{2}{\pi}\int \mathrm{d}^2\beta\, P\left(\beta\right) e^{-2|\alpha-\beta|^2} \tag{3.72}$$

for a one-mode system and analogously for an M-mode system. This means that the Wigner function for a one-mode coherent state $\hat{\rho} = |\alpha_0\rangle\langle\alpha_0|$ is

$$W\left(\alpha,\alpha^*\right) = \frac{2}{\pi}e^{-2|\alpha-\alpha_0|^2}\,. \tag{3.73}$$

The fact that the Wigner function is a Gaussian convolution of the P-function means that it will usually be less singular than the P-function for the same quantum system.

As will be explained in more detail later, the evolution equations for a Wigner function for a nonlinear Hamiltonian typically have no second-order derivatives, unless there is dissipation, but instead it has third-order derivatives. As a consequence, these evolution equations do not correspond to a Fokker-Planck equation. Also, the Wigner function is not in general positive and thus does not correspond to a probability density function. However, in many instances, both of these problems can be ignored. In particular, where third-order derivatives are small compared to first- and second-order derivatives, they can be ignored, and the evolution equations can be truncated. This is called the truncated-

Wigner method and turns out to be a useful and reasonably accurate method for many systems.

3.2.8 Husimi Q-representation

The Husimi Q-representation, also called Husimi Q-function or just Q-function, first introduced by Husimi[46], is usually indicated by $Q(\boldsymbol{\alpha}, \boldsymbol{\alpha}^*)$ and corresponds to an anti-normal operator ordering, that is, $s = -1$. It can be shown that the Q-function can be found via

$$Q(\boldsymbol{\alpha}, \boldsymbol{\alpha}^*) = \frac{1}{\pi^M} \langle \boldsymbol{\alpha} | \hat{\rho} | \boldsymbol{\alpha} \rangle \ , \tag{3.74}$$

which means that unlike the types of representation considered so far, it is always a positive semi-definite function and thus corresponds to a true probability density function.

The Q-function results from both the P-function and the Wigner function through a Gaussian convolution via

$$Q(\boldsymbol{\alpha}, \boldsymbol{\alpha}^*) = \frac{1}{\pi^M} \int \mathrm{d}^{2M}\boldsymbol{\beta}\, P(\boldsymbol{\beta}, \boldsymbol{\beta}^*)\, e^{-|\boldsymbol{\alpha}-\boldsymbol{\beta}|^2} \tag{3.75}$$

$$Q(\boldsymbol{\alpha}, \boldsymbol{\alpha}^*) = \frac{1}{\pi^M} \int \mathrm{d}^{2M}\boldsymbol{\beta}\, W(\boldsymbol{\beta}, \boldsymbol{\beta}^*)\, e^{-2|\boldsymbol{\alpha}-\boldsymbol{\beta}|^2} \ . \tag{3.76}$$

Because of this, the Q-function is less singular than both the P-function and the Wigner function.

Even though the Q-function, unlike both the P-function and the Wigner function, corresponds to a true probability density function and is also the most regular, that is, least singular, between the three, it is generally not a useful representation for the purpose of a phase-space simulation, because the evolution equations typically don't exhibit positive definite diffusion terms and thus don't correspond to a Fokker-Planck equation.

3.2.9 Positive P-representation

The Glauber-Sudarshan P-representation corresponds to a continuous-value decomposition of the density operator into the diagonal basis elements of a coherent state $|\boldsymbol{\alpha}\rangle \langle \boldsymbol{\alpha}|$. Drummond and Gardiner[47] first proposed a generalization of this by including the

off-diagonal elements of the form $|\boldsymbol{\alpha}\rangle\langle\boldsymbol{\beta}|$ as well. This leads to a phase-space representation whose dimension is doubled relative to the types of phase-space representation considered so far. In such a generalized positive P-distribution, the density operator becomes

$$\hat{\rho} = \int\int \mathrm{d}^{2M}\boldsymbol{\alpha}\mathrm{d}^{2M}\boldsymbol{\beta}\, P(\boldsymbol{\alpha},\boldsymbol{\beta})\,\hat{\Lambda}_{+}(\boldsymbol{\alpha},\boldsymbol{\beta}) \ , \tag{3.77}$$

where the operator kernel $\hat{\Lambda}_{+}(\boldsymbol{\alpha},\boldsymbol{\beta})$ is

$$\hat{\Lambda}_{+}(\boldsymbol{\alpha},\boldsymbol{\beta}) \equiv \frac{|\boldsymbol{\alpha}\rangle\langle\boldsymbol{\beta}^{*}|}{\langle\boldsymbol{\beta}^{*}|\,\boldsymbol{\alpha}\rangle} \ . \tag{3.78}$$

The denominator $\langle\boldsymbol{\beta}^{*}|\,\boldsymbol{\alpha}\rangle$ ensures normalization of the off-diagonal elements.

For a generalized P-representation as described above, expectation values can be obtained using a normal-ordered operator form $\hat{O}_{N}\left(\hat{a},\hat{a}^{\dagger}\right)$ via

$$\mathrm{Tr}\left(\hat{\rho}\hat{O}\left(\hat{a},\hat{a}^{\dagger}\right)\right) = \int\int \mathrm{d}^{2M}\mathrm{d}^{2M}\, P(\boldsymbol{\alpha},\boldsymbol{\beta})\, O_{N}(\boldsymbol{\alpha},\boldsymbol{\beta}) \ . \tag{3.79}$$

Unlike for the types of phase-space distributions considered so far, a generalized P-distribution is not unique, in other words, generally there exist different functions $P(\boldsymbol{\alpha},\boldsymbol{\beta})$ that correspond to the same density operator $\hat{\rho}$. It can be shown that for a given density operator $\hat{\rho}$, the following function always corresponds to a corresponding generalized P-distribution:

$$P_{+}(\boldsymbol{\alpha},\boldsymbol{\alpha}^{*},\boldsymbol{\beta},\boldsymbol{\beta}^{*}) = \frac{1}{(2\pi)^{2M}} e^{-|\boldsymbol{\alpha}-\boldsymbol{\beta}^{*}|^{2}/4} \left\langle \frac{\boldsymbol{\alpha}+\boldsymbol{\beta}^{*}}{2} \middle| \hat{\rho} \middle| \frac{\boldsymbol{\alpha}+\boldsymbol{\beta}^{*}}{2} \right\rangle \ . \tag{3.80}$$

This choice is clearly a non-negative function. In other words, for a given density operator, it is always possible to find a corresponding non-negative generalized P-function. A non-negative generalized P-function is usually indicated by $P_{+}(\boldsymbol{\alpha},\boldsymbol{\alpha}^{*},\boldsymbol{\beta},\boldsymbol{\beta}^{*})$ and is also called a "positive P-function". Obviously, such a function corresponds to a true probability density function. The positive P-function can generally be chosen in a way that results in a nonsingular function.

An important feature of the positive P-function is that the evolution equations even for a nonlinear Hamiltonian generally result in equations that correspond to a Fokker-Planck equation, which makes the positive P-representation suitable for phase-space simulations even for highly nonlinear systems. The disadvantages of the positive-P function over the

Wigner function include that fact that the dimensionality is the phase-space is twice that of the Wigner function. Also, the positive P-function can sometimes be less localized, that is more spread out in phase-space than the Wigner function. Both of these properties can result in larger sampling errors when compared to the Wigner function. The latter property can result in non-vanishing boundary terms, which may lead to systematic errors. This can be eliminated using more complex gauge-P equations of motion[48].

3.2.10 Operator identities

Using the formalism of the s-ordering approach, the density operator can be expressed for any of the above representations, as well as others, which have not been discussed here via

$$\hat{\rho} = \int \mathrm{d}^{2M}\boldsymbol{\alpha}\mathrm{d}^{2M}\boldsymbol{\beta}\, W_s\left(\boldsymbol{\alpha},\boldsymbol{\alpha}^*,\boldsymbol{\beta},\boldsymbol{\beta}^*\right)\hat{\Lambda}_s\left(\boldsymbol{\alpha},\boldsymbol{\beta}\right)\ , \tag{3.81}$$

where s indicates the associated operator ordering. For a singe-dimensional representation, the constraint $\boldsymbol{\beta} \equiv \boldsymbol{\alpha}^*$ is in place, which leads to the collapse of one phase-space integral.

It can be shown that the action of a creation or annihilation operator on the operator kernel $\hat{\Lambda}_s$ corresponds to a linear differential operator involving the corresponding phase-space variable. The exact form of this operator depends on the associated operator ordering, whether the operator is a creation or annihilation operator and whether it is acted from the left or from the right. These operator identities are:

$$\begin{aligned}
\hat{a}_n\hat{\Lambda}_s &= \left(\alpha_n + \frac{s-1}{2}\frac{\partial}{\partial\beta_n}\right)\hat{\Lambda}_s \\
\hat{a}_n^\dagger\hat{\Lambda}_s &= \left(\beta_n + \frac{s+1}{2}\frac{\partial}{\partial\alpha_n}\right)\hat{\Lambda}_s \\
\hat{\Lambda}_s\hat{a}_n &= \left(\alpha_n + \frac{s+1}{2}\frac{\partial}{\partial\beta_n}\right)\hat{\Lambda}_s \\
\hat{\Lambda}_s\hat{a}_n^\dagger &= \left(\beta_n + \frac{s-1}{2}\frac{\partial}{\partial\alpha_n}\right)\hat{\Lambda}_s
\end{aligned} \tag{3.82}$$

3.2.11 Time-evolution equations

For a quantum system undergoing unitary evolution, the quantum dynamics are described by the von Neumann-equation

$$\frac{\partial}{\partial t}\hat{\rho} = -\frac{i}{\hbar}\left[\hat{H},\hat{\rho}\right] . \tag{3.83}$$

To describe losses and gain, eq. (3.83) may include other terms such as the Lindblad superoperator[49]

$$\frac{\partial}{\partial t}\hat{\rho} = -\frac{i}{\hbar}\left[\hat{H},\hat{\rho}\right] + \hat{L}\left[\hat{\rho}\right] . \tag{3.84}$$

Generally speaking, the quantum dynamics are described via

$$\frac{\partial}{\partial t}\hat{\rho} = \hat{\mathcal{L}}\left(\hat{\boldsymbol{a}},\hat{\boldsymbol{a}}^{\dagger}\right)\hat{\rho}\,, \tag{3.85}$$

where $\hat{L}\left(\hat{\boldsymbol{a}},\hat{\boldsymbol{a}}^{\dagger}\right)$ is a multinomial involving the operators $\hat{a}_n, \hat{a}_n^{\dagger}$ acting from both sides.

If $\hat{\rho}$ is expressed through a phase-space representation $W_s\left(\boldsymbol{\alpha},\boldsymbol{\alpha}^*,\boldsymbol{\beta},\boldsymbol{\beta}^*\right)$, eq. (3.84) results in

$$\frac{\partial}{\partial t}\hat{\rho} = \int\int \mathrm{d}^{2M}\boldsymbol{\alpha}\mathrm{d}^{2M}\boldsymbol{\beta}\; W_s\left(\boldsymbol{\alpha},\boldsymbol{\alpha}^*,\boldsymbol{\beta},\boldsymbol{\beta}^*\right)\hat{\mathcal{L}}\left(\hat{\boldsymbol{a}},\hat{\boldsymbol{a}}^{\dagger}\right)\hat{\Lambda}_s\left(\boldsymbol{\alpha},\boldsymbol{\beta}\right) , \tag{3.86}$$

where the superoperator $\hat{\mathcal{L}}$ is now acting on the operator kernel $\hat{\Lambda}_s$. Using the identities eqs. (3.82), this will result in a linear operator L involving the phase-space variables and derivatives of these via

$$\frac{\partial}{\partial t}\hat{\rho} = \int\int \mathrm{d}^{2M}\boldsymbol{\alpha}\mathrm{d}^{2M}\boldsymbol{\beta}\;\left[L\left(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\partial\alpha},\boldsymbol{\partial\beta}\right)W_s\left(\boldsymbol{\alpha},\boldsymbol{\alpha}^*,\boldsymbol{\beta},\boldsymbol{\beta}^*\right)\right]\hat{\Lambda}_s\left(\boldsymbol{\alpha},\boldsymbol{\beta}\right) . \tag{3.87}$$

To illustrate this, consider the following hypothetical von Neumann equation for a one-mode system:

$$\frac{\partial}{\partial t}\hat{\rho} = -\frac{i}{\hbar}\left[\hat{a}^{\dagger},\hat{\rho}\right] . \tag{3.88}$$

In the positive P-representation, this becomes

$$\frac{\partial}{\partial t}\hat{\rho} = -\frac{i}{\hbar}\int\int \mathrm{d}^2\alpha \mathrm{d}^2\beta\, P_+\left(\alpha,\alpha^*,\beta,\beta^*\right)\left[\hat{a}^\dagger,\hat{\Lambda}_+\left(\alpha,\beta\right)\right] \tag{3.89}$$

$$\frac{\partial}{\partial t}\hat{\rho} = -\frac{i}{\hbar}\int\int \mathrm{d}^2\alpha \mathrm{d}^2\beta\, P_+\left(\alpha,\alpha^*,\beta,\beta^*\right)\frac{\partial}{\partial\alpha}\hat{\Lambda}_+\left(\alpha,\beta\right)\ . \tag{3.90}$$

If the phase-space representation satisfies certain conditions, namely if it vanishes faster than any power of α,β at the boundaries of the integration domain, the integral in eq. (3.90) can be carried out via integration by parts and the boundary terms can be ignored. This results in

$$\frac{\partial}{\partial t}\hat{\rho} = -\frac{i}{\hbar}\int\int \mathrm{d}^2\alpha \mathrm{d}^2\beta\, P_+\left(\alpha,\alpha^*,\beta,\beta^*\right)\left(-\frac{\partial}{\partial\alpha}\right)\hat{\Lambda}_+\left(\alpha,\beta\right)\ . \tag{3.91}$$

One can then equate the coefficients of the operator kernel on the left-hand side and right-hand side of eq. (3.91) (after expressing $\hat{\rho}$ on the left-hand side via the phase-space representation) for every α, β to find an evolution equation for $P_+\left(\alpha,\alpha^*,\beta,\beta^*\right)$.

Generalizing the above procedure, one finds the following mappings between an operator acting on $\hat{\rho}$ and a linear differential operator acting on the phase-space representation $W_s\left(\boldsymbol{\alpha},\boldsymbol{\alpha}^*,\boldsymbol{\beta},\boldsymbol{\beta}^*\right)$:

$$\begin{aligned}
\hat{a}_n\hat{\rho} &\to \left(\alpha_n - \frac{s-1}{2}\frac{\partial}{\partial\beta_n}\right)W_s\\
\hat{a}_n^\dagger\hat{\rho} &\to \left(\beta_n - \frac{s+1}{2}\frac{\partial}{\partial\alpha_n}\right)W_s\\
\hat{\rho}\hat{a}_n &\to \left(\alpha_n - \frac{s+1}{2}\frac{\partial}{\partial\beta_n}\right)W_s\\
\hat{\rho}\hat{a}_n^\dagger &\to \left(\beta_n - \frac{s-1}{2}\frac{\partial}{\partial\alpha_n}\right)W_s\,.
\end{aligned} \tag{3.92}$$

This leads to a direct translation of a quantum evolution equation to an evolution equation of the chosen phase-space representation.

The resulting evolution equations for the probability density function are stochastic differential equations. Typically, these will not be easily solvable analytically and need to be treated numerically. Though in principle, it is possible to find the time-evolution numerically by solving for the phase-space representation directly, e.g. by discretizing the phase-space representation, this usually requires immense numerical effort relative

to the achievable accuracy. In most cases, the preferred method is to simulate an set of stochastic differential equations that is equivalent to the phase-space equation of motion instead. This approach, where possible, usually requires drastically lower computational efforts for an achievable accuracy compared to a direct solution of the partial differential equation that is the equation of motion governing the phase-space representation function (W_s) itself.

3.2.12 Stochastic phase-space methods

In order for an equivalent set of stochastic differential equations to exist, the phase-space representation must evolve according to a Fokker-Planck equation, that is, it must not have higher than second-order derivatives. Furthermore, the coefficients of the second-order derivative terms must constitute a positive semi-definite matrix. In general, among the types of phase-space representation discussed here, this is only true for the positive-P representation. This usually precludes the use of the Glauber-Sudarshan P-representation as well as the Husimi Q-representation from stochastic phase-space simulations, unless the specific system to be simulated is a special case resulting in a Fokker-Planck equation for these representations or this problem can be overcome otherwise. At the same time, the typically high degree of singularity and possibility of being negative makes the Glauber-Sudarshan P-representation an even more unlikely candidate for a stochastic phase-space simulation than the Husimi Q-representation. The Wigner representation, on the other hand, often results in an equation of motion that is very similar to a Fokker-Planck equation, except for the fact that it includes derivative terms of third order or higher. Typically, these higher-order derivative terms correspond to strong quantum effects most noticable in the regime of low occupation. This means that for moderate to high occupation numbers, an approximation can be made by ignoring these terms, which results in a Fokker-Planck equation that can then be simulated using the corresponding stochastic differential equations. This method is called "truncated Wigner method" due to the truncation of higher-order derivative terms, while the method of using the positive-P representation, which generally does not require truncation is simply called "positive-P method".

A major advantage of the truncated Wigner method over the positive-P method is its greater simplicity due to the fact that the Wigner representation does not require

phase-space doubling. This means that the truncated Wigner method involves only half as many stochastic equations as the positive-P method. The stochastic equations for the truncated Wigner method themselves closely resemble the Louville equations for the corresponding classical system. Another advantage of the truncated Wigner method is that for a given quantum state, the corresponding Wigner function is unique. On the other hand, a given quantum state often has infinitily many, arbitrarily broad positive-P representations. In many cases, a positive-P simulation leads to a broadening of the distribution in time, usually due to noise originating from nonlinear terms, which leads to an increase of the stochastic error in time.

A major disadvantage of the truncated Wigner method is its approximate nature due to the aforementioned necessity to ignore higher-order derivative terms. Another disadvantage is the fact that the Wigner function has negative values for certain quantum states, in which case the Wigner function does not correspond to a probability density function, which means it cannot be treated in a stochastic simulation. A Wigner function with negative values is typically associated with a highly nonclassical state. The truncated Wigner method is thus unsuitable for simulations involving quantum states corresponding to a Wigner representation with negative values.

The main advantages of the positive-P method are that it corresponds to a probability density function for every quantum state and generally results in a Fokker-Planck equation without the need to truncate terms. As a consequence of this, the positive-P method can be used in many cases where the truncated Wigner method fails, because it involves quantum states with a negative Wigner function or is not accurate enough because of the truncation of higher-order terms. However, the positive-P method has inherent limitations as well. On the one hand, the validity of the positive-P method relies on the assumption that integration by parts performed on eq. (3.90) does not introduce any (non-negligible) boundary terms. Gilchrist *et al.*[50] demonstrated that in general, this is only strictly true in the limit of large damping relative to nonlinear terms present. Nonvanishing boundary terms typically manifest themselves in the form of systematic errors ranging from small or negligible transient contributions to a drastically different evolution, including solutions with incorrect steady-state values. These terms usually originate from certain singularities where trajectories can grow infinitely large due to the phase-space dynamics given by the deterministic part of the equations of motion.

Note that these singularities normally have zero probability to occur yet can still cause nonvanishing boundary terms due to the effect on trajectories in their neighborhoods. Because of these nonvanishing boundary terms, similar to the truncated Wigner method, the positive-P method can strictly speaking not be called an exact method either. However, for virtually all experimentally accessible physical systems, these boundary terms are negligibly small due to large enough damping relative to nonlinear terms. Gilchrist *et al.* analyzed a number of physical systems and found that for every one of these, the system parameters corresponding to nonlinearities would have to be much larger than those of any known physical system for nonlinear boundary terms to have non-negligible effects. This seems to be the case for every feasible experiment today, even though it is impossible to make a definitive statement due to the sheer multitude of conceivable systems.

Another problem of the positive-P method is that for certain systems, the stochastic error can become uncontrollably large. As an example of such a system, consider the anharmonic oscillator whose Hamiltonian is

$$\hat{H} = \frac{1}{2}\hat{a}^{\dagger 2}\hat{a}^2 \,. \tag{3.93}$$

Using appropriately scaled variables, a damped anharmonic oscillator system can be described by the quantum master equation

$$\frac{\partial \hat{\rho}}{\partial \tau} = \frac{\gamma}{2}\left(2\hat{a}\hat{\rho}\hat{a}^\dagger - \hat{a}^\dagger\hat{a}\hat{\rho} - \hat{\rho}\hat{a}^\dagger\hat{a}\right) - \frac{i}{2}\left[\hat{a}^{\dagger 2}\hat{a}^2\right] \,, \tag{3.94}$$

where τ is a rescaled time variable and γ is a dimensionless coefficient measuring the damping rate relative to the nonlinear interaction. The positive-P stochastic phase-space equations corresponding to eq. (3.94) are:

$$d\alpha = \left(-\frac{\gamma}{2}\alpha - i\alpha^2\alpha^+\right)d\tau + \sqrt{-i}\alpha dW_1 \,, \tag{3.95}$$

$$d\alpha^+ = \left(-\frac{\gamma}{2}\alpha^+ + i\alpha^{+2}\alpha\right)d\tau + \sqrt{i}\alpha^+ dW_2 \,, \tag{3.96}$$

where dW_i are independent Wiener processes increments satisfying $\left\langle dW_i dW_j \right\rangle = \delta i, j d\tau$. The occupation number N is given by $N \equiv \alpha\alpha^+$ and satisfies the stochastic differential

equation

$$dN = -\gamma N d\tau + N\left(\sqrt{-i}dW_1 + \sqrt{i}dW_2\right) . \tag{3.97}$$

Taking the expectation value of eq. (3.97) results in

$$d\langle N\rangle = -\gamma\langle N\rangle d\tau . \tag{3.98}$$

The same result can be obtained from eq. (3.94) by means of standard quantum operator techniques and using the fact that

$$\langle N\rangle = \mathrm{Tr}\left[\hat{a}^\dagger\hat{a}\hat{\rho}\right] , \tag{3.99}$$

which means that the positive-P method predicts the correct expectation value for the occupation number. It is worth pointing out that this system is somewhat atypical due to the fact that it is *not* affected by nonvanishing boundary terms, even in the highly nonlinear and low-damping regime. However, on considering the phase-space variable $M \equiv |N|^2$, one finds

$$d\langle M\rangle = 2\left(1-\gamma\right)\langle M\rangle d\tau . \tag{3.100}$$

From eq. (3.100), it follows that if $\gamma < 1$, the standard deviation, in other words the stochastic error ΔN associated with the occupation number will grow exponentially in time. This exponential growth might not be a fundamental limitation, but it poses very serious technical challenges for a stochastic simulation, especially if the simulation time exceeds the e-folding time of eq. (3.100) several times. As a result, there will be a quasi-limit for the simulation time, beyond which the sampling error will render any meaningful prediction intractable, even if massive numerical resources are utilized. As in the case of nonvanishing boundary conditions, the system described here itself exhibits problematic behaviour in the form of uncontrollable stochastic errors only for system parameters, which are well beyond what is experimentally feasible. At the same time, the anharmonic oscillator Hamiltonian is also used in models of interacting Bose-Einstein condensates[51]. Unlike in the case of quantum optics, in the field of Bose-Einstein condensation, for

experimentally feasible system parameters, simulations involving anharmonic oscillator terms can exhibit exponential error growth[52] and become untractable numerically.

In cases where a positive-P simulation is untractable due to nonvanishing boundary terms, uncontrollable sochastic error or a both of these, a relatively new technique called "stochastic gauge method"[48] can be used in an attempt to overcome these problems. This involves the inclusion of an additional field, the so-called gauge field. In addition, two functions of the fields, called drift gauge and diffusion gauge, have to be chosen. This results in a modified and augmented set of stochastic differential equations. The stochastic gauge field then plays the role of a (albeit complex-valued) weight-factor in the retrieval of observables.

It has been demonstrated for a number of example systems that using the stochastic gauge method, the effects of nonvanishing boundary terms can be significantly reduced to a negligible level and that for the anharmonic oscillator system, the simulation time can be substantially increased before the simulation becomes unfeasible due to stochastic errors[53]. This has been demonstrated to be useful in the simulation of a macroscopic Bose-Einstein condensate[54]. However, it seems that even using stochastic gauge methods, it is impossible to increase the simulation time arbitrarily. It seems that the rapid growth of stochastic errors can be mitigated rather than eliminated. A major drawback of stochastic gauge methods is that its success depends on the choice for the drift and diffusion gauge functions. However, there is no straightforward method or prescription how these should be chosen. This means that for every system, a specific set of gauge functions has to be determined based on a careful analysis of the system dynamics as well as intuition and a trial-and-error approach.

3.3 Entanglement criteria, EPR steering and Bell inequalities

In this section, a review of three related, but different concepts, namely entanglement[55], the EPR-paradox[16] and Bell inequalities[69], is given. The EPR paradox and Bell inequalities are physical concepts which elucidate a phenomenon known as quantum nonlocality, that is, the fact that quantum mechanics contradicts a principle called "local realism"[16], which will be described in detail. Nonlocality is one of the most distinctive

and arguably one of the most surprising features of quantum physics which to this day motivates a great amount of theoretical and experimental investigation. Entanglement is the fundamental concept of quantum mechanics that underpins nonlocality. It does not necessarily imply quantum nonlocality, however quantum nonlocality always implies entanglement.

These physical concepts will be explained in detail and various criteria will be presented, which provide a means to quantify the degree to which these phenomena are observed.

3.3.1 Entanglement criteria

Consider a Hilbert space $\mathcal{H}$ that is composed of a number of Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2, ..., \mathcal{H}_N$ via

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes ... \otimes \mathcal{H}_N \,. \tag{3.101}$$

A pure state $|\psi\rangle \in \mathcal{H}$ is said to be separable if it can be expressed via

$$|\psi\rangle = |\psi^1\rangle \otimes |\psi^2\rangle \otimes ... \otimes |\psi^N\rangle \,, \tag{3.102}$$

where $|\psi^1\rangle \in \mathcal{H}_1$; $|\psi^2\rangle \in \mathcal{H}_2$; ...; $|\psi^N\rangle \in \mathcal{H}_N$.

In the following, I will omit the tensor product symbol between states, implying that $|\psi^1\rangle |\psi^2\rangle$ means $|\psi^1\rangle \otimes |\psi^2\rangle$ for $|\psi^1\rangle \in \mathcal{H}_1$, $|\psi^2\rangle \in \mathcal{H}_2$.

The singlet state (called a Bell state)

$$|\psi\rangle = \frac{1}{\sqrt{2}} \left[|\uparrow\rangle_A |\downarrow\rangle_B - |\downarrow\rangle_A |\uparrow\rangle_B \right] \tag{3.103}$$

cannot be factorised into any product state of the form $|\psi\rangle_A |\psi\rangle_B$ and is an example of a non-separable (entangled) state.

A mixed state is a collection of pure states and probabilities $\{p_i, |\psi_i\rangle\}$ where $\sum_i p_i = 1$. It is associated with density operator $\hat{\rho}$ where

$$\hat{\rho} \equiv \sum_i p_i |\psi_i\rangle \langle\psi_i| \,. \tag{3.104}$$

The physical interpretation of a mixed state is that the system is in quantum state $|\psi_i\rangle$ with probability p_i.

A mixed state associated with density operator $\hat{\rho}$ is said to be separable if $\hat{\rho}$ can be expressed through

$$\hat{\rho} = \sum_i p_i \;\; \hat{\rho}_i^1 \otimes \hat{\rho}_i^2 \otimes ... \otimes \hat{\rho}_i^N \,, \tag{3.105}$$

where $\hat{\rho}_i^n$ is the density operator associated with a mixed state whose pure states lie in Hilbert space $\mathcal{H}_n$ and $\sum_i p_i = 1,\ p_i > 0$[56]. A pure or mixed state which is not separable is called an *entangled*[55] state.

A central question related to quantum entanglement is the prospect of measuring entanglement for a given state, that is, determining whether the state is a separable or entangled state or defining a measurement which corresponds to the degree of entanglement of the state. Such a measure is called an entanglement criterion. Entanglement criteria are a subject of ongoing research. The availability and exact form of an entanglement criterion depends on a number of parameters, such as

- Whether the system is composed of Hilbert spaces whose basis sets are finite, countably infinite or uncountably infinite. A Hilbert space with finite or countably infinite basis elements is also called a discrete variable (DV) Hilbert space, while a Hilbert space with uncountably infinite many basis elements is also called a continuous variable (CV) Hilbert space.

- The number of Hilbert spaces (parties) constituting the total Hilbert space,

- Whether the state is a pure state or a mixed state,

- Whether certain assumptions can be made on the states. For example, certain continuous-value entanglement criteria exist which are only valid for Gaussian states.

The simplest case is that of a bipartite pure state with finite-dimensional, thus discrete variable (DV) Hilbert space constituents. In this case, the von Neumann entropy[57] provides a quite natural entanglement criterion. The von Neumann entropy S of a state

with density operator $\hat{\rho}$ is defined as

$$S\left(\hat{\rho}\right) = -\operatorname{Tr}\left[\hat{\rho}\log_2\left(\hat{\rho}\right)\right] . \tag{3.106}$$

In such a case, the entanglement of a state $|\psi\rangle$ can be measured via

$$E\left(|\psi\rangle\right) = S\left(\hat{\rho}^1\right) = S\left(\hat{\rho}^2\right) , \tag{3.107}$$

where $\hat{\rho}^n$ is obtained by tracing out the Hilbert space component of the other constituent, that is

$$\hat{\rho}^1 = \operatorname{Tr}_2\left[|\psi\rangle\langle\psi|\right] , \tag{3.108}$$

$$\hat{\rho}^2 = \operatorname{Tr}_1\left[|\psi\rangle\langle\psi|\right] . \tag{3.109}$$

If $|\psi\rangle$ is separable, then $E\left(|\psi\rangle\right) = 0$. For $E\left(|\psi\rangle\right) > 0$, the state is entangled. If $|\psi\rangle$ is a mixed state, the criterion above does not apply, even for finite-dimensional Hilbert spaces. Measuring entanglement for a mixed state is generally considered a hard problem, even where the Hilbert spaces are relatively simple.

Tan[58] found the following entanglement criterion for a bipartite system using measurements that have a continuous variable (CV) outcome:

$$\Delta^p_{ent} = 2\Delta\left(\hat{X}_1 + \hat{X}_2\right)\Delta\left(\hat{P}_1 - \hat{P}_2\right) , \tag{3.110}$$

where $\hat{X}_i \equiv \frac{1}{2}\left(\hat{a}_i + \hat{a}_i^\dagger\right)$, $\hat{P}_i \equiv \frac{1}{2i}\left(\hat{a}_i - \hat{a}_i^\dagger\right)$ are the quadrature phase amplitudes. In experiments, these are usually measured through homodyne detection[27]. $\left(\Delta\hat{X}\right)^2 \equiv \left\langle\hat{X}^2\right\rangle - \left\langle\hat{X}\right\rangle^2$, sometimes abbreviated $\Delta^2 X$, is the variance of the observable $\hat{X}$ and $\Delta\hat{X} \equiv \sqrt{\left(\Delta\hat{X}\right)^2}$. He demonstrated that $\Delta^p_{ent} < 1$ is a sufficient, but not necessary criterion for a general bipartite two-mode state.

Duan *et al.*[59] found the following entanglement criterion for a bipartite system using measurements that have a continuous variable (CV) outcome:

$$\Delta^a_{ent} = \frac{\left[\Delta\left(\hat{X}_1 - g_x\hat{X}_2\right)\right]^2 + \left[\Delta\left(\hat{P}_1 + g_p\hat{P}_2\right)\right]^2}{\left(1 + g_x g_p\right)/2} , \tag{3.111}$$

where g_x and g_p are arbitrary real constants which can be chosen to minimize Δ^a_{ent}. Usually, the criterion is applied for using $g \equiv g_x = g_p$. They proved that $\Delta^a_{ent} < 1$ is a sufficient, but not necessary criterion for a general bipartite two-mode state, while it is a necessary and sufficient criterion for a bipartite two-mode Gaussian state, for optimally selected quadratures and gains.

Giovannetti *et al.*[60] found a generalization of the criterion formulated by Tan, that is, eq. (3.110), which, when expressed in terms of quadratures $\hat{X}, \hat{P}$, states that

$$\Delta^p_{ent} = \frac{4\Delta\left(\hat{X}_1 - g_x\hat{X}_2\right)\Delta\left(\hat{P}_1 + g_p\hat{P}_2\right)}{\left(1 + g_x g_p\right)} \tag{3.112}$$

with real constants g_x, g_p chosen to minimize Δ^p_{ent}. Note that $\Delta^p_{ent} < 1$ is a sufficient, but not necessary criterion for entanglement. They demonstrated that Δ^p_{ent} is a stronger criterion compared to Δ^a_{ent} in the sense that $\Delta^p_{ent} \leq \Delta^a_{ent}$ for any bipartite CV state. In other words, it is possible that for a given state $|\psi\rangle$, the criterion Δ^a_{ent} does not indicate entanglement, but the criterion Δ^p_{ent} does.

For the entanglement criteria mentioned above, instead of the canonical quadratures $\hat{X}, \hat{P}$, the generalized quadratures $\hat{X}^\theta, \hat{P}^\theta$ can be used, which are defined as

$$\hat{X}^\theta \equiv \frac{1}{2}\left[e^{-i\theta}\hat{a} + e^{i\theta}\hat{a}^\dagger\right] \tag{3.113}$$

$$\hat{P}^\theta \equiv \frac{1}{2}\left[e^{-i(\theta+\pi/2)}\hat{a} + e^{i(\theta+\pi/2)}\hat{a}^\dagger\right], \tag{3.114}$$

in which case $\Delta^a_{ent}, \Delta^p_{ent}$ give the strongest entanglement signatures by considering $\inf\limits_{\theta,\theta'} \Delta^p_{ent}$ or $\inf\limits_{\theta,\theta',g_x,g_p} \Delta^{a/p}_{ent}$, respectively, where θ is the phase associated with state 1 and θ' is the phase associated with state 2. In practice, only the phase difference between the states matters, so the phase can be fixed for one state and only one phase needs to be optimized.

3.3.2 EPR-steering criteria

The Einstein-Podolksy-Rosen (EPR) paradox, first described by Einstein *et al.*[16] is the phenomenon, where quantum entanglement for certain states seemingly leads to an inconsistency between local realism and the completeness of quantum mechanics. In their original work, Einstein *et al.* considered a spatially separated bipartite state with perfectly correlated positions and momenta. For such a state, if the position of particle A were

measured with measurement outcome x, the collapse of the wavefunction would cause the position of particle B to be $x + d$ with 100% certainty, where d is the distance between the two particles. Similarly, if the momentum of particle A were measured to be of value p, the implication would be that particle B has momentum $-p$ with 100% certainty. EPR argued that this Gedankenexperiment itself is a paradox and thus quantum mechanics cannot be considered complete. They based their argument on three assumptions: They assumed quantum mechanics was correct insofar as it would correcly predict the outcome of the Gedankenexperiment described here. They further assumed that any observable, whose outcome could be predicted with 100% certainty, must necessarily correspond to "an element of physical reality", an assumption which is also called "realism". Lastly, they assumed that there could be no "action at a distance", in other words, a measurement on particle A could not have an effect on the quantum state of particle B, since particles A and B are spatially separated. This assumption in combination with the assumption of realism is also called "local realism".

The EPR argument can be summed up as follows: "Assuming local realism, quantum mechanics is incomplete". The argument made by EPR originated many debates about the nature of quantum mechanics and ultimately led to a better understanding of it.

The EPR paradox is based on the fact that a (hypothetical) measurement of an observable of one constituent, such as a particle, of the system allows for the inference of an observable the other constituent. Schrödinger called this effect "steering"[61] - the outcome of a measurement made on one constituent of the system "steers" or "pilots" the quantum state of the other constituent. Unlike entanglement, this steering effect, also called "EPR-steering", is by definition asymmetric, that is, the ability of system A to steer system B need not be equal to the ability of system B to steer system A, even though in EPR's original Gedankenexperiment, this was the case.

EPR-steering originates a new type of quantum nonlocality signature, which is different from entanglement.

A criterion to observe this form of nonlocality for a bipartite continuous variable system, was proposed by Reid[29]. Here, the quadrature phase amplitudes are considered instead of position and momentum. She demonstrates how quadratures of particle A can be inferred from the outcome of a hypothetical quadrature measurement of particle B. Defining an inferred variance in the quadratures through the squared discrepancy

between quadrature and inferred quadrature, i.e.

$$\left(\Delta_{inf}\hat{X}\right)^2 = \left\langle \left[\hat{X}_1 - \hat{X}_{inf}\right]^2 \right\rangle , \tag{3.115}$$

she considers Heisenberg's uncertainty

$$4\Delta_{inf}\hat{X}\Delta_{inf}\hat{P} \geq 1 , \tag{3.116}$$

which becomes

$$4\Delta\left(\hat{X}_A - g\hat{X}_B^\theta\right)\Delta\left(\hat{P}_A + g\hat{P}_B^\theta\right) \geq 1 , \tag{3.117}$$

where g is a real number and $\hat{X}^\theta$, $\hat{P}^\theta$ defined as in the previous section. A violation of eq. (3.117) is indicative of EPR-steering of system A by system B. The strongest EPR-steering criterion can be obtained by considering

$$\inf_{g_x, g_p, \theta} 4\Delta\left(\hat{X}_A - g\hat{X}_B^\theta\right)\Delta\left(\hat{P}_A + g\hat{P}_B^\theta\right) . \tag{3.118}$$

Reid[29] verified that the two-mode squeezed state generates the correlations of the EPR paradox. EPR-steering was first formulated rigorously by Wiseman, Jones and Doherty[62] and Cavalcanti *et al.*[63].

EPR-steering was realised in the optical experiments described in[64]. An experimental investigation of steering for qubits was given by Saunders *et al.*[65]. In general, EPR steering is more difficult to observe than entanglement, particularly where thermal noise is present [66, 67]. A first analysis for EPR steering for an opto-mechanical system using a simple two-mode model was given in by He and Reid[67]. The first demonstration of EPR-steering considered loophole-free by, among other things, realizing space-like separation of all relevant events, was carried out by Wittman *et al.*[68].

The topic of EPR-steering will be further expanded upon in section 3.3.4. To provide an easier flow of reading, the topic of Bell inequalities will be treated before that.

3.3.3 Bell inequalities

The conclusion of many physicists including Einstein from the EPR paradox was that quantum mechanics was incomplete. Their proposition was that it had to be expanded by the addition of local hidden variables, which would have a deterministic effect on measurement outcomes. The inclusion of these local hidden variables, they argued, would make quantum mechanics in agreement with local realism and resolve any seeming paradoxes such as the EPR paradox.

Bell[69] analyzed this proposal and constructed a Gedankenexperiment himself. Instead of the original EPR-paradox, Bell considered a different type of EPR paradox, which was suggested by Bohm and Aharonov[70] after the original paper by EPR. In Bohm's version of the EPR paradox, the same argument is made for a system of two entangled and spatially separated spin states.

Bell considered an experiment where two spatially separated observers (indicated by A and B) measure pairs of particles with perfectly anticorrelated spin, corresponding to the singlet state

$$|\psi\rangle = \frac{1}{\sqrt{2}} \left(|1\rangle_A |-1\rangle_B - |-1\rangle_A |1\rangle_B \right) , \tag{3.119}$$

where $|1\rangle_A$ corresponds to a spin state with z-projection 1 at site A, etc.

He defines the quantity $\mathbb{E}\left[\boldsymbol{a},\boldsymbol{b}\right]$ as the expectation value of $\left(A\left(\boldsymbol{a}\right) \cdot B\left(\boldsymbol{b}\right)\right)$, where $A\left(\boldsymbol{a}\right)$ indicates the outcome of a spin-measurement at site A along the spatial direction indicated by $\boldsymbol{a}$ and $B\left(\boldsymbol{b}\right)$ indicates the outcome of a simultaneous spin-measurement at site B along the spatial direction indicated by $\boldsymbol{b}$.

He further assumes the validity of a local hidden variable theory. He argues that if such a theory is indeed correct, $\mathbb{E}\left[\boldsymbol{a},\boldsymbol{b}\right]$ will result via

$$\mathbb{E}\left[\boldsymbol{a},\boldsymbol{b}\right] = \int A\left(\boldsymbol{a},\lambda\right) B\left(\boldsymbol{b},\lambda\right) p\left(\lambda\right) \mathrm{d}\lambda \, , \tag{3.120}$$

where λ is a local hidden variable. He does not make any assumptions on the nature of λ itself, except for the fact that there exist a probability density $p\left(\lambda\right)$. Using standard arguments from probability calculus as well as the fact that $A\left(\boldsymbol{a},\lambda\right) = \pm 1$, $B\left(\boldsymbol{b},\lambda\right) = \pm 1$,

he derives the formula

$$|\mathbb{E}\left[\boldsymbol{a},\boldsymbol{b}\right] - \mathbb{E}\left[\boldsymbol{a},\boldsymbol{c}\right]| \leq 1 + \mathbb{E}\left[\boldsymbol{b},\boldsymbol{c}\right] \,, \tag{3.121}$$

where $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ are spatial directions. He then demonstrates that, for certain configurations of $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$, the results predicted by quantum mechanics are in direct contradiction to eq. (3.121) and hence, a local hidden variable theory and, by implication, local realism are incompatible with quantum mechanics, independent of the exact nature of the local hidden variables.

Note that his argument does not falsify a local hidden variable theory, but merely states its incompatibility with quantum mechanics. A disproof of a local hidden variables theory can only be established if it can be demonstrated that its predictions are wrong while, presumably, the predictions of quantum mechanics are correct and such a demonstration can only be achieved experimentally. Bell's original equation, that is eq. (3.121), however, is derived on the assumption of perfectly anticorrelated spin-pairs, which are impossible to achieve experimentally.

Since Bell's original publication, several different inequalities have been derived, which have the same implications of local hidden variable theories and local realism, but are generally easier to test experimentally. These are usually referred to as "Bell-type inequalities". One of the most important ones is the Clauser-Horne-Shimony-Holt (CHSH) inequality[71]. The CHSH inequality is, like the original Bell inequality, derived using only standard arguments of probability calculus as well as the triangle inequality. It states that

$$\mathbb{E}\left[\boldsymbol{a},\boldsymbol{b}\right] + \mathbb{E}\left[\boldsymbol{a},\boldsymbol{b}'\right] + \mathbb{E}\left[\boldsymbol{a}',\boldsymbol{b}\right] - \mathbb{E}\left[\boldsymbol{a}',\boldsymbol{b}'\right] \leq 2 \,, \tag{3.122}$$

where $\boldsymbol{a},\boldsymbol{b},\boldsymbol{a}',\boldsymbol{b}'$ are different detector settings. For the choice

$$\begin{aligned} \boldsymbol{a} &= \hat{\boldsymbol{x}} \\ \boldsymbol{b} &= -\frac{\hat{\boldsymbol{z}} + \hat{\boldsymbol{x}}}{\sqrt{2}} \\ \boldsymbol{a}' &= \hat{\boldsymbol{z}} \\ \boldsymbol{b}' &= \frac{\hat{\boldsymbol{z}} - \hat{\boldsymbol{x}}}{\sqrt{2}} \,, \end{aligned} \tag{3.123}$$

assuming a perfectly anticorrelated state, quantum mechanics predicts the left-hand side of ineq. (3.122) to be of value $2\sqrt{2}$, which is a clear violation of the inequality. Unlike in the case of Bell's original inequality, ineq. (3.122) includes the possibility, that is, it is still valid if some of the measurement outcomes on site A and B are 0, e.g. due to deficiencies in the detectors.

A different class of Bell-type equations was derived by Cavalcanti, Foster, Reid and Drummond (CRFD)[34]. Unlike Bell inequalities, these inequalities can be applied to continuous variable systems with no bound on the eigenvalues. Instead of a two-site setup as in the Bell-type inequalities presented so far, they considered an experimental setup with N sites. They define two different observables at site i as $\hat{X}_i$ and $\hat{Y}_i$. They construct the operator

$$\hat{C}_N \equiv \prod_{k=1}^{N} \left(\hat{X}_k + i\hat{Y}_k\right) , \tag{3.124}$$

where i is the imaginary unit. Formally evaluating the real and imaginary part of eq. (3.124) allows for the definition of $\hat{\tilde{X}}_N, \hat{\tilde{Y}}_N$ via

$$\hat{\tilde{X}}_N + i\hat{\tilde{Y}}_N \equiv \prod_{k=1}^{N} \left(\hat{X}_k + i\hat{Y}_k\right) . \tag{3.125}$$

Further they demonstrate that for a local hidden variable theory, for an operator $\hat{F}$, the inequality

$$\left\langle \hat{F} \right\rangle \leq \left\langle \left|\hat{F}\right| \right\rangle \tag{3.126}$$

applies. Applying this to eq. (3.125) then results in the CFRD-inequality

$$\left\langle \hat{\tilde{X}}_N \right\rangle^2 + \left\langle \hat{\tilde{Y}}_N \right\rangle^2 \leq \left\langle \prod_{k=1}^{N} \left(\hat{X}_k^2 + \hat{Y}_k^2\right) \right\rangle . \tag{3.127}$$

Shchukin and Vogel (CV)[35] generalized this approach by also considering the quaternion and octonion algebra, which allows for a maximum of 8 observables at each site, while CFRD limited their analysis to the complex algebra.

A systematic analysis of the possibility of observing violations of the CFRD and CV inequalities, respectively, in an N-site experiment with Greenberger, Horne and Zeilinger

(GHZ) states[72] as well as generalized GHZ states, can be found in chapter 6. An N-site GHZ state is defined as

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|1\rangle^{\otimes N} + |-1\rangle^{\otimes N}\right) , \tag{3.128}$$

which is a special case of the generalized GHZ state, defined as

$$|\psi\rangle = \cos(\alpha)\,|1\rangle^{\otimes N} + \sin(\alpha)\,|-1\rangle^{\otimes N} . \tag{3.129}$$

3.3.4 EPR-steering and hierarchy of quantum nonlocality

Schrödinger[55, 61], who coined the terms "entanglement" as well as "steering" of a system B a system A, rejected the idea of a local hidden variables (LHV) theory, but also the concept of nonlocality. His suggestion was that quantum mechanics was incorrect in its description of spacially separated entangled states. He proposed that in an entangled system, the parties are in definite, local quantum states, which, similar to a local hidden variables theory (LHV), are generally hidden. This model is called a local hidden state (LHS) theory.

Wiseman *et al.*[62] extended the earlier analysis of Reid[29], to systematically analyze the three different types of quantum nonlocality: entanglement, Bell violations (based on the assumption of a LHV theory) and EPR-steering (based on the assumption of a LHS theory), and described them in a unified mathematical formalism, which will be briefly outlined here.

Consider a bipartite system with sites called "Alice" and "Bob", described by a density operator $\hat{\rho}$. Let A be an observable at Alice's site with eigenvalues $\{a\}$ and B be an observable at Bob's site with eigenvalues $\{b\}$. Let the probability to measure a at Alice's site given the density operator $\hat{\rho}$ be denoted by $P(a|A;\hat{\rho})$ with an analogous definition for a measurement on Bob's site. Let the joing probability to measure a at Alice's site and b at Bob's site be denoted by $P(a,b|A,B;\hat{\rho})$. Assuming a local hidden variables (LHV) theory, the joint probability result as

$$P(a,b|A,B;\hat{\rho}) = \sum_{\xi} p(a|A,\xi)\, p(b|B,\xi)\, p_{\xi} , \tag{3.130}$$

where ξ indicates the LHV and $p\left(a|A,\xi\right)$, $p\left(b|B,\xi\right)$ are some probability density functions for the measurement of a at Alice's site, and b at Bob's site, respectively, involving the LHV ξ. If a set of observables A and B at Alice's and Bob's site, respectively, can be found such that no probability density functions $p\left(a|A,\xi\right)$, $p\left(b|B,\xi\right)$ exist to satisfy eq. (3.130), this would falsify a LHV theory.

A quantum state is separable if its density operator $\hat{\rho}$ can be expressed via $\hat{\rho} = \sum_{\xi} \hat{\rho}_{\xi}^{A} \otimes \hat{\rho}_{\xi}^{B}$, where $\hat{\rho}_{\xi}^{A}$ and $\hat{\rho}_{\xi}^{B}$ are pure states of Alice's and Bob's Hilbert space, respectively. Hence, a state with density matrix is separable if for all observables $A,\ B$, the joint probability to measure a and b is

$$P\left(a,b|A,B;\hat{\rho}\right) = \sum_{\xi} P\left(a|A;\hat{\rho}_{\xi}^{A}\right) P\left(b|B;\hat{\rho}_{\xi}^{B}\right) . \tag{3.131}$$

If two observables $A,\ B$ can be found for which eq. (3.131) does not hold, the state is entangled.

Unlike entanglement and Bell violations, EPR-steering is an inherently asymmetric effect, that is, Alice's ability to steer Bob's state and Bob's ability to steer Alice's state are, in principle, two separate effects. Consider the question of Alice steering Bob's state. Assuming a measurement outcome on Alice's site of a, a local hidden states (LHS) theory assumes that prior to a measurement on Alice's site, Bob's system had already been prepared in a specific state $\hat{\rho}_B$. The specific state that $\hat{\rho}_B$ was prepared in, is determined by a hidden variable ξ according to a probability density function $p\left(a|A,\xi\right)$ which results in

$$\tilde{\hat{\rho}}_{B} = \sum_{\xi} p\left(a|A,\xi\right) \hat{\rho}_{\xi}^{B} p_{\xi},, \tag{3.132}$$

The tilde indicates that $\tilde{\hat{\rho}}_{B}$ in eq. (3.132) is unnormalized. For a rigorous, albeit lengthy derivation, refer to [62]. This results in a joint probability $P\left(a,b|A,B;\hat{\rho}\right)$ assuming an LHS model of

$$P\left(a,b|A,B;\hat{\rho}\right) = \sum_{\xi} p\left(a|A,\xi\right) P\left(b|B;\hat{\rho}_{\xi}\right) p_{\xi} . \tag{3.133}$$

If a set of observables $A,\ B$ can be found for which no probability density functions $p\left(a|A,\xi\right)$ exist to satisfy eq. (3.133), then the LHS theory is falsified and EPR-steering

is demonstrated. Based on the formalism of a LHS theory, Wiseman *et al.*[62] showed that eq. (3.117) is a sufficient and necessary EPR-criterion for two-mode Gaussian CV systems, while Cavalcanti *et al.*[63] demonstrated that it is a sufficient, but not necessary criterion for a general two-mode CV system.

Wiseman *et al.*[62] demonstrated that the criteria given by eq. (3.131), eq. (3.133) and eq. (3.130) form a strict hierarchy, that is, every bipartite state that exhibits Bell violations also exhibits EPR-steering and every state that exhibits EPR-steering also exhibits entanglement. This complements prior findings by Werner[56], who demonstrated that every state that exhibits Bell violations implies entanglement but the reverse is not true. Gisin[73] demonstrated that the situation is very different for pure states as any bipartite entangled pure state exhibits EPR-steering as well as Bell violations.

3.4 Cavity Optomechanics

An optomechanical system is a system in which the interaction between an electromagnetic field at an optical frequency and a mechanical component through radiation pressure plays a central role. Many key phenomena of optomechanics can also be achieved using electromagnetic fields at frequencies lower than optical frequencies, which is often included in the term optomechanics as well.

Cavity optomechanics is the study of optomechanical systems which involve one or more optical cavities, also known as Fabry-Pérot resonators or etalons. An optical cavity can be realized in a multitude of ways. In its simplest form, it is made of two highly reflective parallel plates facing each other. If the plates are separated by a distance L, the cavity exhibits a number of longitudinal eigenfrequencies $\omega = m \cdot \pi \frac{c}{L}$. Typically, only the first eigenfrequency is of interest, which I will call the cavity eigenfrequency ω_c.

Although a wide variety of experiments can be considered, which could involve multiple cavities, multiple beamsplitters, nonlinear elements inside the cavity, etc., I will focus here on the simplest setup in cavity optomechanics, also known as the standard optomechanical model[74]. For a comprehensive review of the subject of optomechanics, refer e.g. to [25] and [26].

The standard optomechanical model involves an optical cavity where one of the plates is fixed while the other plate is attached to a mechanical oscillator with eigenfrequency

ω_m. In the limit of small mirror displacement, this system can be described by the Hamiltonian[75]

$$\hat{H}_{sys}/\hbar = \omega_c \hat{a}^\dagger \hat{a} + \omega_m \hat{b}^\dagger \hat{b} + \chi_0 \hat{a}^\dagger \hat{a} \left(\hat{b} + \hat{b}^\dagger\right) , \tag{3.134}$$

where $\hat{a}, \hat{a}^\dagger$ are the cavity mode operators and $\hat{b}, \hat{b}^\dagger$ the mechanical mode operators, respectively. The third term describes the optomechanical coupling, that is, the interaction between the cavity and the mechanical modes due to radiation pressure, where χ_0 is a coupling constant, which results from the system parameters[75].

Further, an external field, such as a laser, with frequency ω_L impinges on the cavity and generates a field amplitude $E(t)$ inside the cavity. Including this term, the Hamiltonian becomes

$$\left(\hat{H}_{sys} + \hat{H}_{laser}\right)/\hbar = \omega_c \hat{a}^\dagger \hat{a} + \omega_m \hat{b}^\dagger \hat{b} + \chi_0 \hat{a}^\dagger \hat{a} \left(\hat{b} + \hat{b}^\dagger\right) + iE(t) \left(\hat{a}^\dagger e^{-i\omega_L t} - \hat{a} e^{i\omega_L t}\right) . \tag{3.135}$$

On defining Δ via

$$\omega_L = \omega_c - \Delta \tag{3.136}$$

and switching into an interaction picture with $H_I/\hbar = \omega_L \hat{a}^\dagger \hat{a}$, the Hamiltonian becomes

$$\hat{H}/\hbar = \Delta \hat{a}^\dagger \hat{a} + \omega_m \hat{b}^\dagger \hat{b} + \chi_0 \hat{a}^\dagger \hat{a} \left(\hat{b} + \hat{b}^\dagger\right) + iE(t) \left(\hat{a}^\dagger - \hat{a}\right) . \tag{3.137}$$

A full description must also include coupling to the environment. Assuming a cavity decay rate of κ and a mechanical decay rate of γ_m and using cavity input-output theory[76], the system dynamics including environment coupling can be described either by the quantum Langevin equations[77]

$$\dot{\hat{a}} = E(t) - (i\Delta + \kappa)\hat{a} - i\chi_0 \hat{a} \left(\hat{b} + \hat{b}^\dagger\right) + \sqrt{2\kappa}\hat{a}^{in} \tag{3.138}$$

$$\dot{\hat{b}} = -(i\omega_m + \gamma_m)\hat{b} - i\chi_0 \hat{a}^\dagger \hat{a} + \sqrt{2\gamma_m}\hat{b}^{in} \tag{3.139}$$

or through a master equation by including appropriate Lindblad operator terms. There can also be intracavity losses which are not coupled to output modes, requiring additional Lindblad operator terms to describe them.

Making an adiabatic approximation and a subsequent linearization of the dynamics described by eqs. (3.138, 3.139)[78] yields an effective, linearized system Hamiltonian

$$\hat{H}_{eff,lin}/\hbar = \Delta\hat{a}^\dagger\hat{a} + \omega_m\hat{b}^\dagger\hat{b} + \chi_{eff}(t)\left(\hat{a} + \hat{a}^\dagger\right)\left(\hat{b} + \hat{b}^\dagger\right), \tag{3.140}$$

where $\chi_{eff}(t)$ is an effective optomechanical coupling constant, which depends on $E(t)$. This linearized Hamiltonian reveals the effect of the optomechanical coupling in the standard optomechanical model for two special values of Δ:

- For $\Delta = -\omega_m$, the interaction term becomes $\hat{H}_{int} = \chi_{eff}\left(\hat{a}\hat{b} + \hat{a}^\dagger\hat{b}^\dagger\right)$ after ignoring off-resonant terms, which is known as the rotating wave approximation. This Hamiltonian is known as the two-mode squeezing Hamiltonian[76]. The two-mode squeezing Hamiltonian generates entanglement between the two modes (here $\hat{a}$, $\hat{b}$). Referring to the optical Doppler shift, the choice $\Delta = \omega_m$ is also called "blue-detuned", that it, the laser frequency is said to be blue-detuned relative to the cavity eigenfrequency.

- For $\Delta = \omega_m$, the interaction term becomes $\hat{H}_{int} = \chi_{eff}\left(\hat{a}\hat{b}^\dagger + \hat{a}^\dagger\hat{b}\right)$ after making the same rotating-wave approximation. This Hamiltonian is a simple beam-splitter Hamiltonian. Its action can be thought of as a state-transfer or state swapping between the optical and mechanical modes. This choice is also called "red-detuned".

The standard optomechanical model in combination with these laser detuning cases forms the basis of my publications on the subject of optomechanics described in chapter 4.

3.5 POS

The method of parallel optimized samples (POS) has been a central part of my research. My research has led to two publications which can be found in chapters 7 and 8. Even though the method is explained in detail in these publications, I will also explain its

central concepts here, because it is a rather complex method. Instead of a comprehensive description, only the key ideas will be described here.

Its central objective is the numerical solution of a stochastic differential equation. Any numerical solution of a stochastic differential equation will have at least two sources of errors: a time-step error ϵ_T, which decreases when the time-step is reduced and a stochastic sampling error ϵ_S, which decreases when more samples are used. While much research has been done to decrease the time-step error of a stochastic integration, the sampling error has been the subject of considerably less research. The objective of POS is to reduce the sampling error.

We have developed two versions of POS. The first version allows for the solution of Itō-type stochastic differential equations while the second one is suitable for Stratonovich-type differential equations and results in lower time-step errors. I will only describe the Itō-type POS algorithm here. The Stratonovich-type algorithm shares its ideas and concepts.

3.5.1 Euler-Mayurama version of POS

We have developed two conceptually similar versions of the POS adaption of the Euler-Mayurama algorithm. One of these versions is called "combined method" and the other "invidiual method". In the interest of brevity, I will only explain the combined method here. For details regarding the individual method, refer to the publication.

Combined method

Consider the univariate Itō-type stochastic differential equation

$$\mathrm{d}X = \mu\left(X\right)\mathrm{d}t + \sigma\left(X\right)\mathrm{d}W\left(t\right) \ . \tag{3.141}$$

For POS, a number of observables $o_1\left(X\left(t\right)\right), o_2\left(X\left(t\right)\right), ..., o_M\left(X\left(t\right)\right)$ are chosen, which are said to be *optimized.* These may or may not coincide with, be a subset of, or overlap with the observables which one wishes to obtain from the stochastic integration.

Let $t_0 < t_1 < ... < t_N$ be discrete time-points with a separation of Δt. Considering Itō's lemma (see 3.1), one can conclude that the observable $o_m\left(X\left(t\right)\right)$ will evolve during

one time-step Δt as

$$o_{i+1}^m = o_i^m + \left[\frac{\partial o_i^m}{\partial X} + \frac{1}{2} \frac{\partial^2 o_i^m}{\partial^2 X} o_i^m \right] \Delta t + \frac{\partial o_i^m}{\partial X} \Delta W , \tag{3.142}$$

where $o_i^m \equiv o_m \left(X \left(t_i \right) \right)$.

Now let $\boldsymbol{X} \left(t \right) \equiv \left(X_1 \left(t \right), X_2 \left(t \right), ..., X_{N_S} \left(t \right) \right)$ be a sampling of the probability density function of the stochastic variable X and $\Delta \boldsymbol{W} \equiv \left(\Delta W_1, \Delta W_2, ..., \Delta W_{N_S} \right)$ a vector of N_S independent noise increments. It follows that the $o_i^m \equiv \left\langle o_m \left(X_n \left(t_i \right) \right) \right\rangle_n$ will evolve approximately as

$$o_{i+1}^m = o_i^m + \sum_{n=1}^{N_S} \left[\frac{\partial o_i^m}{\partial X_n} + \frac{1}{2} \frac{\partial^2 o_i^m}{\partial^2 X_n} o_i^m \right] \Delta t + \sum_{n=1}^{N_S} \frac{\partial o_i^m}{\partial X_n} \Delta W_n . \tag{3.143}$$

Because of the independence of the noise increments, the last term will be zero in the limit $N_s \to \infty$.

Let $\mathcal{L}_i^m \equiv \sum_{n=1}^{N_S} \left[\frac{\partial o_i^m}{\partial X_n} + \frac{1}{2} \frac{\partial^2 o_i^m}{\partial^2 X_n} o_i^m \right] \Delta t$. If it can be ensured that o_i^m evolves as

$$o_{i+1}^m = o_i^m + \mathcal{L}_i^m , \tag{3.144}$$

then the variance in the optimized observables $o \left(X \left(t \right) \right)$ will be reduced throughout the stochastic integration.

Let $\boldsymbol{J}^i$ be the Jacobian matrix where $J_{mn}^i \equiv \frac{\partial o_i^m}{\partial X_n}$. Let $\boldsymbol{X}_{i+1}$ be the solution via the conventional Euler-Mayurama algorithm

$$\boldsymbol{X}_{i+1} = \boldsymbol{X}_i + \boldsymbol{\mu} \left(\boldsymbol{X}_i \right) \Delta t + \boldsymbol{\sigma} \left(\boldsymbol{X}_t \right) , \tag{3.145}$$

where $\boldsymbol{X}_i \equiv \boldsymbol{X} \left(t_i \right)$. In general, $\boldsymbol{X}_{i+1}$ will not satisfy eq. (3.142). Let $\Delta \boldsymbol{X}_{i+1}$ be a correction term to $\Delta \boldsymbol{X}_{i+1}$. Note that

$$\boldsymbol{o} \left(\boldsymbol{X}_{i+1} + \Delta \boldsymbol{X}_{i+1} \right) \approx \boldsymbol{o} \left(\boldsymbol{X}_{i+1} \right) + \boldsymbol{J}^{i+1} \Delta \boldsymbol{X}_{i+1} . \tag{3.146}$$

Note that eq. (3.142) now becomes

$$\boldsymbol{J}^{i+1} \Delta \boldsymbol{X}_{i+1} = \boldsymbol{o}_i + \boldsymbol{\mathcal{L}}_i - \boldsymbol{o}_{i+1} , \tag{3.147}$$

where $\boldsymbol{J}$ is typically nondiagonal. Given that in most cases $N_S > M$, eq. (3.147) is underdetermined. Eq. (3.147) can be solved by means of the Moore-Penrose pseudoinverse

$$\Delta \boldsymbol{X}_{i+1} = \left(\boldsymbol{J}^{i+1}\right)^{+} \left(\boldsymbol{o}_i + \boldsymbol{\mathcal{L}}_i - \boldsymbol{o}_{i+1}\right) . \tag{3.148}$$

The solution of an underdetermined system through the Moore-Penrose pseudoinverse results in the minimal least-squares solution. This ensures minimal change in the sample vector $\boldsymbol{X}$, thus minimizing the possibility of the introduction of a bias. Under certain conditions (refer to the publications for more details), the Moore-Penrose pseudoinverse of a matrix $\boldsymbol{A}$ can be obtained via

$$\boldsymbol{A}^{+} = \boldsymbol{A}^{\dagger} \left(\boldsymbol{A}\boldsymbol{A}^{\dagger}\right)^{-1} . \tag{3.149}$$

By obtaining the Moore-Penrose pseudoinverse in this way for $\boldsymbol{J}^{i+1}$ via

$$\boldsymbol{J}^{i+1} = \left(\boldsymbol{J}^{i+1}\right)^{\dagger} \boldsymbol{u}_{i+1}^{\dagger} \tag{3.150}$$

$$\boldsymbol{u}_i = \boldsymbol{J}^{i} \left(\boldsymbol{J}^{i}\right)^{\dagger} , \tag{3.151}$$

we ensure that the POS algorithm has only linear complexity in N_S.

Because eq. (3.146) is approximate, the sample vector $\boldsymbol{X}_{i+1} + \Delta \boldsymbol{X}_{i+1}$ will satisfy eq. (3.142) only approximately also. Eq. (3.142) can be made to be satisfied to increasing accuracy by applying eq. (3.148) repeatedly in an iterative way. Let $\boldsymbol{X}^{(0)} \equiv \boldsymbol{X}^{i+1}$ (note the distinction between superscript in brackets and without brackets). Then an iterative algorithm which ensures increasingly accurate satisfaction of eq. (3.142) is given via

$$\boldsymbol{X}^{(n+1)} = \boldsymbol{X}^{(n)} + \left(\boldsymbol{J}^{(n)}\right)^{\dagger} \left(\boldsymbol{u}^{(n)}\right)^{-1} \left(\boldsymbol{o}_i + \boldsymbol{\mathcal{L}}_i - \boldsymbol{o}_{(n)}\right) , \tag{3.152}$$

where $\boldsymbol{J}^{(n)} \equiv \boldsymbol{J}\left(\boldsymbol{X}^{(n)}\right)$, $\boldsymbol{o}_{(n)} \equiv \boldsymbol{o}\left(\boldsymbol{X}^{(n)}\right)$ and $\boldsymbol{u}^{(n)} \equiv \boldsymbol{J}^{(n)} \left(\boldsymbol{J}^{(n)}\right)^{\dagger}$. Once a defined stopping criterion is met, the iteration finishes, resulting in the final sample vector for the time-step t_{i+1}.

The generalization for a multivariate stochastic differential equation follows naturally by extending the sample vector $\boldsymbol{X}$ to include the samples of all stochastic variables.

3.5.2 Static ensemble optimization

The ideas from the previous section can be used for the sampling of a probability density function. Consider a sample vector $\boldsymbol{X}$ which was obtained through stochastic sampling of a probability density function. Let $\boldsymbol{o}(X) \equiv (o_1(X), o_2(X), ..., o_M(X))$ be a number of observables for which the exact values $\boldsymbol{m} = (m_1, m_2, ..., m_M)$ for the probability density function which is being sampled, are known. For example, if the probability density function is a normal distribution, $\boldsymbol{m}$ could indicate a number of moments, as the moments of a normal distribution are well-known. Due to the stochastic sampling with a finite number of elements, the $\boldsymbol{o}(\boldsymbol{X})$ will equal $\boldsymbol{m}$ only approximately. Let $\boldsymbol{J}$ be the Jacobian matrix of $\boldsymbol{o}$ w.r.t. $\boldsymbol{X}$ as defined previously. Note that for a small change $\Delta\boldsymbol{X}$ on $\boldsymbol{X}$ leads to

$$\boldsymbol{o}(\boldsymbol{X} + \Delta\boldsymbol{X}) \approx \boldsymbol{o}(\boldsymbol{X}) + \boldsymbol{J}\Delta\boldsymbol{X}. \tag{3.153}$$

By demanding $\boldsymbol{o}(\boldsymbol{X} + \Delta\boldsymbol{X}) = \boldsymbol{m}$, one obtains

$$\boldsymbol{J}\Delta\boldsymbol{X} = \boldsymbol{m} - \boldsymbol{o}(\boldsymbol{X}), \tag{3.154}$$

which can be solved through the Moore-Penrose pseudoinverse, ensuring a minimal least-squares solution for $\Delta\boldsymbol{X}$. Analogously to the previous section, an iterative algorithm can be constructed. Let $\boldsymbol{X}^{(0)} \equiv \boldsymbol{X}$. The iteration follows via

$$\boldsymbol{X}^{(n+1)} = \boldsymbol{X}^{(n)} + \left(\boldsymbol{J}^{(n)}\right)^{+} \left[\boldsymbol{m} - \boldsymbol{o}_{(n)}\left(\boldsymbol{X}^{(n)}\right)\right], \tag{3.155}$$

where $\boldsymbol{J}^{(n)} \equiv \boldsymbol{J}\left(\boldsymbol{X}^{(n)}\right)$ and $\boldsymbol{o}_{(n)} \equiv \boldsymbol{o}\left(\boldsymbol{X}^{(n)}\right)$. This way, $\boldsymbol{X}^{(n)}$ will satisfy $\boldsymbol{o}\left(\boldsymbol{X}^{(\boldsymbol{n})}\right) = \boldsymbol{m}$ with increasing accuracy.

Chapter 4

Scalable quantum simulation of pulsed entanglement and Einstein-Podolsky-Rosen steering in optomechanics

Authors: S. Kiesewetter[1], Q. Y. He[1], P. D. Drummond[1], and M. D. Reid[1]

Affiliations: [1]Centre for Quantum and Optical Science, Swinburne University of Technology, Melbourne, Australia

Published in: Physical Review A **90**, p. 043805; accepted 4 September 2014

4.1 Abstract

We demonstrate a complete, probabilistic quantum dynamical simulation of the standard nonlinear Hamiltonian of optomechanics, including decoherence at finite temperatures. Robust entanglement of a photonic pulse with a mechanical oscillator is predicted. Our exact quantum simulations use the positive-P technique, are scalable to large Hilbert spaces, and give excellent agreement with recent experiments. We predict the required conditions for the next stage in this research. Strong quantum steering between the photonic and mechanical systems is possible, depending on thermal occupation. This is more viable in optical rather than electromechanical experiments.

4.2 Introduction

Optomechanical oscillators provide fundamental tests of mesoscopic quantum mechanics, as well as having potential technological applications to ultra-sensitive measurement. Impressive success in cooling optomechanical systems near their ground state has been reported [1–4], with demonstrations of a number of quantum mechanical effects for mesoscopic systems [5–8], including the observation of entanglement of a microwave field with a mechanical oscillator, in an elegant electromechanical experiment [9]. This was predicted theoretically [10–12] and involves pulsed inputs [12, 13].

The next outstanding goal is to observe the *nonlocal quantum correlations* known as EPR-steering, as predicted by Einstein, Podolsky and Rosen in their famous EPR paradox [14–16], for mesoscopic massive objects. Such a realization would be a precursor to experiments that directly probe the macroscopic reality of an object [17, 18]. It is of fundamental interest not only whether an optical field can entangle with a massive oscillator but also whether the two systems can show these strange directional "spooky action-at-a-distance" [19] effects that Schrödinger called "steering" [20–22].

Here we carry out a scalable, probabilistic quantum simulation of the standard nonlinear optomechanical model, in a regime allowing entanglement and EPR steering of the mechanical oscillator. This fully elucidates the quantum mechanical interplay between entanglement generation and thermal decoherence. We study the dynamical generation of correlations between the oscillator and an output pulse for realistic parameters. In our simulations we utilize the exact positive-P ($+P$) phase-space method [23], which has a positive probability distribution for all quantum states. Our work passes the barrier set by Olsen *et al.* for their pendular cavity quantum simulation [24]. This used the +P method but with a different noise model, and with high temperatures that do not allow entanglement. Other earlier optomechanical calculations make assumptions ranging from linearization [10] to adiabatic approximations [12], or both.

A limitation in current optomechanics experiments is that long interaction times lead to increased decoherence, owing to a coupling to the environment at relatively high temperature. It was recently proposed that this could be overcome by creating and verifying entanglement with pulses of light on fast time scales [12, 13]. This was an adiabatic, linearized study [25]. The authors showed that entanglement is feasible, provided $Qf \gg k_m T_{\text{bath}}/h$, where T_{bath} is the temperature of the environment, f is the

frequency of mechanical oscillator, and Q is the cavity quality. By comparison, we can treat arbitrary pulse shapes, the method is exact rather than linearized, and we use optimized entanglement and EPR measures. We also compare our results with the truncated Wigner (tW) method [26], valid for large photon-numbers [27]. Both methods agree with the recent entanglement observations at JILA [9].

The $+P$ technique, used in other parametric simulations [28, 29], has no approximations apart from those of the standard model [30]. It gives accurate results for realistic experimental parameters. Compared to direct diagonalization [31] or quantum trajectory approaches [32], the method readily scales to large Hilbert spaces [33] and is especially useful for low-order correlations [34]. Neither approximations [35] nor new hardware [36] are required. Such methods have potential for treating a new class of multimode optomechanical devices [37]. This allows a quantitative understanding of the validity of previous methods and predictions of new effects.

Our results validate some earlier predictions for the simplified model in Refs. [12, 38]. They are useful for addressing new EPR-steering experiments in optomechanics. While entanglement of pulse and oscillator is robust, without requiring a low temperature reservoir, thermal noise can prevent steering of the mechanical system when it is thermally excited. This is a fundamentally asymmetric manifestation of nonlocality [38, 39], beautifully illustrated by the oscillator-pulse system.

The paper is organized as follows. Section 4.3 gives our model and simulation equations. Section 4.4 presents approximate analytic results, and Sec. 4.5 summarizes the entanglement and EPR-steering criteria. Section 4.6 presents our exact quantum simulation results, with Sec. 4.7 giving the conclusions.

4.3 Hamiltonian and operator equations

We consider a single-mode optomechanical Fabry-Perot cavity with coherent pumping and damping. A diagram is shown in Fig. 4.1. We treat the regime of low temperatures and low mechanical damping, where quantum effects are observable. To generate entanglement and EPR steering, a blue-detuned pulse is used to entangle the mechanical oscillator with the output field. To verify entanglement, a second red-detuned pulse is

coupled to the cavity-oscillator system to give a readout of the mechanical oscillator position.

4.3.1 Optomechanical Hamiltonian

The standard, single-mode optomechanical model is used [30, 40]. The Hamiltonian includes the energy of the mechanical oscillator mode at ω_m, an input at ω_l, and the optical mode energy at angular frequency $\omega_o = \omega_l + \delta$. We transform to an interaction picture in which the time evolution of the cavity operators at the laser frequency is removed. This subtracts the term $\omega_l \hat{a}^\dagger \hat{a}$, together with corresponding time dependences, leaving

$$\begin{aligned} \hat{H}/\hbar = \delta \hat{a}^\dagger \hat{a} + \omega_m \hat{b}^\dagger \hat{b} + \chi \hat{a}^\dagger \hat{a}(\hat{b} + \hat{b}^\dagger) \\ + iE(t)(\hat{a}^\dagger - \hat{a}) + \hat{H}_r. \end{aligned} \tag{4.1}$$

The first terms give the energy of the optical cavity field and the mechanical oscillator, while the third term is the optomechanical interaction, where χ is the coupling due to radiation pressure. The fourth term is the coupling to the coherent input $E(t)$, and $\hat{H}_r$ describes the coupling to dissipative reservoirs.

We define $k = [b, r] \equiv [\text{blue}, \text{red}]$ for the entangling and readout pulses, respectively. Following the strategy employed in Ref. [9], we first model the input of a blue-detuned laser pulse of duration τ. Thus, the cavity is initially resonant with the lower (Stokes) sideband of the input so that $\omega_o = \omega_l + \delta_b$, where $\delta_b = -\omega_m$. This enhances nonlinear generation of entanglement between a reflected output mode $\hat{A}_b^{\text{out}}$ and the fundamental mode of the mechanical oscillator $\hat{b}$. We have simulated the correlations induced between the entangling pulse and the mechanical oscillator during this process.

After a short delay time τ_{del}, a red-detuned pulse of duration τ with the opposite detuning $\delta_r = \omega_m$ is input, transferring the mechanical oscillator state to the corresponding output field $\hat{A}_r^{\text{out}}$ [12] as described below. The laser driving strength for the kth input pulse is $E(t) = E_0 \varepsilon(t) = \sqrt{2\gamma_o N_k}\varepsilon(t)$. Here N_k is the pulse photon number, and each envelope function $\varepsilon(t)$ is normalized so that $\int_{\tau_k^0}^{\tau_k^1} dt |\varepsilon(t)|^2 = 1$, where τ_k^0 and τ_k^1 are the start and end times of the kth pulse.

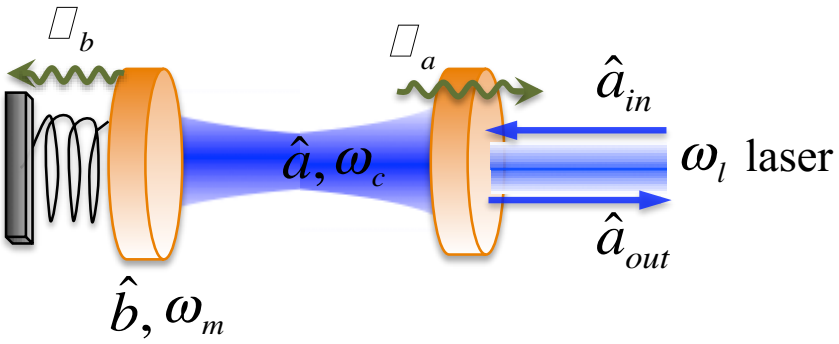

Figure 4.1: Two light pulses enter the cavity and interact with the mirror via radiation pressure. A first blue-detuned pulse entangles, while a second red-detuned pulse gives a readout.

4.3.2 Master equation

This is a driven open system, hence the density matrix must be calculated as the solution of a master equation. The master equation for the reduced density operator is valid [41] for regimes of Markovian behavior and weak damping, so that for $\gamma_j \ll \omega_j$:

$$\begin{aligned} \frac{d\hat{\rho}}{dt} = -\frac{i}{\hbar}[H,\hat{\rho}] + \sum_j \gamma_j \bar{n}_j (2\hat{a}_j^\dagger \hat{\rho} \hat{a}_j - \hat{a}_j \hat{a}_j^\dagger \hat{\rho} - \hat{\rho} \hat{a}_j \hat{a}_j^\dagger) + \\ + \sum_j \gamma_j (\bar{n}_j + 1)(2\hat{a}_j \hat{\rho} \hat{a}_j^\dagger - \hat{a}_j^\dagger \hat{a}_j \hat{\rho} - \hat{\rho} \hat{a}_j^\dagger \hat{a}_j)\,. \end{aligned} \tag{4.2}$$

Here γ_o and γ_m are the optical decay rate and mechanical dissipation rate, while $\bar{n}_j$ are the reservoir thermal occupations. Also, $\hat{\boldsymbol{a}} = (\hat{a}_1, \hat{a}_2) = (\hat{a}, \hat{b})$, with $j = 1,2 \sim o,m$ indexing the optical and mechanical modes respectively.

There are often several distinct types of dissipation in experiments. Here we introduce two types of cavity loss, $\gamma_o = \gamma_{\text{ext}} + \gamma_{\text{int}}$, where γ_{ext} represents the output coupling through the mirrors to external propagating modes, while γ_{int} represents all other types of cavity losses. This distinction is necessary since only the part of the cavity loss that is externally coupled can result in operationally measurable entanglement. We note that Olsen *et al.* used a different, high-temperature master equation in their instructive simulations [24], which are for the strongly dissipative regime of low frequencies and high temperatures typical of gravity-wave detectors.

The master equation used here is equivalent to the following quantum Langevin equation:

$$\begin{aligned}
\dot{\hat{a}} &= E(t) - (i\delta_k + \gamma_o)\,\hat{a} - i\chi\hat{a}(\hat{b} + \hat{b}^\dagger) + \sqrt{2\gamma_o}\hat{a}^{\text{in}}\,, \\
\dot{\hat{b}} &= -\,(i\omega_m + \gamma_m)\,\hat{b} - i\chi\hat{a}^\dagger\hat{a} + \sqrt{2\gamma_m}\hat{b}^{\text{in}}\,, \\
\hat{a}^{\text{out}} &= \sqrt{2\gamma_{\text{ext}}}\hat{a} - \hat{a}^{\text{in}}_{\text{ext}}\,,
\end{aligned} \tag{4.3}$$

where $\hat{a}^{\text{in}}$ and $\hat{b}^{\text{in}}$ are the input quantum noise amplitudes from the optical and mechanical reservoirs. Here $\hat{a}^{\text{out}}$ describes the cavity output field mode, using standard input-output theory [42], and $\langle \hat{a}^{\text{out}\dagger}(t)\hat{a}^{\text{out}}(t)\rangle = \bar{\Phi}^{\text{out}}$ is the mean output flux in photons per second. The input-output relations involve only *external* damping and noise, which leads to an output coupling efficiency, $\eta_o = \gamma_{\text{ext}}/\gamma_o$. There are two different quantum Langevin terms, since while $\hat{a}^{\text{in}}$ is coupled to the internal dynamics, there is also an orthogonal noise term $\hat{a}^{\text{in}}_\perp$ that only couples to the output, as in a beam-splitter:

$$\begin{aligned}
\hat{a}^{\text{in}} &= \sqrt{1-\eta_o}\hat{a}^{\text{in}}_{\text{int}} + \sqrt{\eta_o}\hat{a}^{\text{in}}_{\text{ext}}\,, \\
\hat{a}^{\text{in}}_\perp &= \sqrt{\eta_o}\hat{a}^{\text{in}}_{\text{int}} - \sqrt{1-\eta_o}\hat{a}^{\text{in}}_{\text{ext}}\,.
\end{aligned} \tag{4.4}$$

The input reservoirs are assumed thermal and have correlations given approximately in the Markovian or wide-band noise limit by ideal noise-sources with:

$$\begin{aligned}
\langle \hat{a}_i^{in\dagger}(t)\hat{a}_j^{\text{in}}(t')\rangle &= \bar{n}_{i,\text{th}}\delta_{ij}\delta\,(t-t')\,, \\
\langle \hat{a}_i^{\text{in}}(t)\hat{a}_j^{in\dagger}(t')\rangle &= \left[1 + \bar{n}_{i,\text{th}}\right]\delta_{ij}\delta\,(t-t')\,.
\end{aligned} \tag{4.5}$$

To study output field correlations, we use a temporal mode decomposition to generate single mode output operators with bosonic commutation relations. These are

$$\hat{A}^{\text{in}}_k = \int_0^{\tau_m} dt u^{\text{in}}_k(t)\hat{a}^{\text{in}}(t)\,, \tag{4.6}$$

with similar output mode definitions having in → out. Here τ_m is the maximum interaction time, and the temporal input-output modes $u_k(t)$ have the normalization

$$\int_0^{\tau_m} dt u^*_k(t)u_j(t) = \delta_{kj}\,. \tag{4.7}$$

These operators commute with each other, apart from the bosonic commutation relations within the same mode. They also commute with system operators at *later* times, such as the mechanical oscillator position $\hat{b}$ after an entanglement experiment [42].

However, both the quantum Langevin equations and the master equation are intractable without approximations, due to the large size of the Hilbert space combined with the nonlinear terms in the operator equations.

4.3.3 Phase-space equations

We now introduce an exact quantum simulation method that allows us to calculate dynamical behavior without approximations. This will be used to predict entanglement and EPR steering. As described elsewhere [23], we use standard operator identities in both normally ordered and symmetrically ordered phase-space distributions to transcribe the master equation (4.2) into $c-$number probabilistic phase space equations. These have features similar to phonon-photon coupling equations found in optical fibre simulations [27], as well as in earlier parametric oscillator simulations [28, 29].

These methods allow dynamical quantum simulations of any optomechanical experiment of this type and can readily be generalized to many modes. As an example, in the following sections we analyze pulsed entanglement experiments where the entanglement between the mechanical oscillator and the first (blue-detuned) pulse is generated by the nonlinear interactions of the first pulse. The mechanical oscillator position is read out by the second pulse. This entanglement is then verified by measurements made on the first and second pulses.

Positive-P representation

First we use the positive-P representation [23], which gives a positive phase-space representation for *any* quantum state. This method uses a dimension-doubling approach in order to obtain a probabilistic distribution with a positive-definite diffusion. The transformation requires an assumption of vanishing boundary terms for the phase-space distribution [43]. Such terms are exponentially small for realistic open system parameter values, which we have verified from the low variances of our numerical results.

There are six independent complex phase-space variables $(\alpha, \alpha^+, \beta, \beta^+, \alpha^{\text{out}}, \alpha^{\text{out}+})$ for the cavity mode, the oscillator mode, and the output mode. These represent the

operators $\left(\hat{a}, \hat{a}^\dagger, \hat{b}, \hat{b}^\dagger, \hat{a}^{\text{out}}, \hat{a}^{\text{out}\dagger}\right)$ respectively via a normally ordered correspondence, with $\langle \alpha^+ \alpha \rangle_S = \langle \hat{a}^\dagger \hat{a} \rangle$, $\langle \beta^+ \beta \rangle_S = \langle \hat{b}^\dagger \hat{b} \rangle$, where $\langle . \rangle_S$ is a stochastic and $\langle . \rangle$ a quantum average. After deriving the Fokker-Planck equation [44], one obtains an equivalent set of complex Itô stochastic equations:

$$\begin{aligned}
d\alpha &= \{E(t) - [i\delta_k + i\chi(\beta + \beta^+) + \gamma_o]\,\alpha\}\,dt + dW_1\,, \\
d\beta &= [-(i\omega_m + \gamma_m)\,\beta - i\chi\alpha\alpha^+]\,dt + dW_2\,, \\
d\alpha^+ &= \{E^*(t) + [i\delta_k + i\chi(\beta + \beta^+) - \gamma_o]\,\alpha^+\}\,dt + dW_1^+\,, \\
d\beta^+ &= [(i\omega_m - \gamma_m)\,\beta^+ + i\chi\alpha\alpha^+]\,dt + dW_2^+\,, \\
d\alpha^{\text{out}} &= \sqrt{2\gamma_{\text{ext}}}d\alpha - d\alpha_{\text{ext}}^{\text{in}}\,, \\
d\alpha^{\text{out}+} &= \sqrt{2\gamma_{\text{ext}}}d\alpha^+ - d\alpha_{\text{ext}}^{+in}\,.
\end{aligned} \tag{4.8}$$

The Gaussian noises dW_i are due to both internal nonlinearities and thermal noise inputs, so that $dW_i = dW_i^\chi + dW_i^{\text{th}}$, where

$$\begin{aligned}
dW_1^{\text{th}} &= \sqrt{2\gamma_{\text{int}}}\mathrm{d}\alpha_{\text{int}}^{\text{in}} + \sqrt{2\gamma_{\text{ext}}}\mathrm{d}\alpha_{\text{ext}}^{\text{in}}\,, \\
dW_2^{\text{th}} &= \sqrt{2\gamma_m}\mathrm{d}\beta^{\text{in}}\,.
\end{aligned} \tag{4.9}$$

Here we need to treat the two types of cavity noise separately. In all cases the non-vanishing stochastic thermal correlations are

$$\langle d\alpha_k^{\text{in}} d\alpha_l^{in+} \rangle_S \;=\; \bar{n}_{k,\text{th}}\delta_{kl}dt\,, \tag{4.10}$$

where $k, l = 0, 1, 2 \sim \text{int}, \text{ext}, m$, and $\bar{n}_{k,\text{th}}$ are the mean heat bath occupations. Here we use the notation $d\alpha_m^{\text{in}} \equiv d\beta^{\text{in}}$. In addition, the positive-P method includes quantum noise due to interactions, with correlations

$$\begin{aligned}
\langle dW_i^\chi dW_j^\chi \rangle_S &= -i\delta_{i,3-j}\chi\alpha dt\,, \\
\langle dW_i^{\chi+} dW_j^{\chi+} \rangle_S &= i\delta_{i,3-j}\chi\alpha^+ dt\,.
\end{aligned} \tag{4.11}$$

Input-output mode c-number representations are defined in a similar way to the corresponding operator quantities, as

$$A_k^{\text{in}} = \int_0^{\tau_m} dt u_k^{\text{in}}(t)\alpha^{\text{in}}(t)\,. \tag{4.12}$$

The equations have analytic solutions via stochastic diagram methods [45] and also allow exact probabilistic quantum simulations via an ensemble of trajectories, which can be integrated numerically [46].

Wigner representation

Another approach that is simpler – but approximate – is the truncated Wigner distribution [26], which is a symmetrically ordered representation. Here there are three independent complex phase-space variables $(\alpha, \beta, \alpha^{\text{out}})$. These represent the operators via asymmetrically ordered correspondence, with $\langle \alpha^* \alpha \rangle_S = \langle \hat{a}^\dagger \hat{a} + 1/2 \rangle$, $\langle \beta^* \beta \rangle_S = \langle \hat{b}^\dagger \hat{b} + 1/2 \rangle$. After truncating derivatives higher than second order in the Fokker-Planck equation (valid at large photon number), we obtain

$$\begin{aligned} d\alpha &= \{E(t) - [i\delta_k + i\chi(\beta + \beta^*) + \gamma_o]\,\alpha\}\,dt + dW_1^{\text{th}}\,, \\ d\beta &= \left[-(i\omega_m + \gamma_m)\,\beta - i\chi|\alpha|^2\right] dt + dW_2^{\text{th}}\,, \\ d\alpha^{\text{out}} &= \sqrt{2\gamma_{\text{ext}}}\alpha dt - d\alpha_{\text{ext}}^{\text{in}}\,. \end{aligned} \tag{4.13}$$

Here the thermal Gaussian noise correlations are as given in Eq (4.9), except that they now correspond to symmetric ordering of the reservoir operators, with correlations

$$\langle d\alpha_i^{\text{in}} d\alpha_j^{\text{in}*} \rangle_S \quad = \quad (\bar{n}_{i,\text{th}} + 1/2)\delta_{ij} dt\,. \tag{4.14}$$

Input-output mode c-number representations are defined as in Eq (4.12). The stochastic correlations for these quantities are equivalent to symmetrically ordered operator products. These equations also imply that $\langle |\alpha|^2 \rangle_S = \langle \hat{n} + 1/2 \rangle = 1/2$ when there is no driving or coupling. Although approximate, this method is simple. It provides a rigorous justification for the methods used to analyze recent electromechanics experiments [9], but has fewer restrictions than this adiabatic approach.

4.4 Adiabatic linearized model

In the special case of linearized, adiabatic pulses, Hofer *et al.* [12] have shown how to describe optomechanical entanglement using an approximate Hamiltonian. In this subsection we review their approach. It is useful to start with an adiabatic Hamiltonian

[10] that illustrates the physics. This approximation is not used in the exact simulations described later, but it demonstrates the source of the entanglement and read-out couplings.

4.4.1 Adiabatic Hamiltonian

In the adiabatic limit of large detuning, the mean intracavity field is dominated by external driving and damping, so that

$$\left\langle \dot{\hat{a}} \right\rangle \approx E(t) - (i\delta_k + \gamma_o) \langle \hat{a} \rangle \,, \tag{4.15}$$

which has the steady-state solution

$$\langle \hat{a} \rangle = \bar{\alpha}_k \approx \frac{E}{i\delta_k + \gamma_o} \,. \tag{4.16}$$

We now examine the small fluctuations around the steady state, by introducing $\delta\hat{a} = \hat{a} - \langle \hat{a} \rangle$. The derivation of this transformation is well-known [10, 12]. Expanding the Hamiltonian to first order in the interactions of these small fluctuations, and defining the adiabatic coupling $g_{k,a} = i\chi\bar{\alpha}_k$, gives the adiabatic interaction Hamiltonian at large detunings:

$$\hat{H}_a/\hbar \;\approx\; i\left(g^*_{k,a}\delta\hat{a} - g_{k,a}\delta\hat{a}^\dagger\right)(\hat{b} + \hat{b}^\dagger) \,. \tag{4.17}$$

Next, it is convenient to use operators in a frame rotating with δ_k, defined as $\hat{a}^r \equiv \delta\hat{a}e^{i\delta_k t}$, $\hat{b}^r = \hat{b}e^{i\omega_m t}$, and employ the rotating-wave approximation to focus on quasiresonant terms. There are two limits of interest here:

(1) Blue-detuned

Firstly, consider the blue-detuned case of $\delta_b = -\omega_m$, with $|\delta_b| \gg \gamma_o$. The resonant interaction terms are given by the squeezing and entanglement Hamiltonian:

$$\hat{H}_a/\hbar \;=\; i\left(g^*_{k,a}\hat{a}^r\hat{b}^r - g_{k,a}\hat{a}^{r\dagger}\hat{b}^{r\dagger}\right) \,. \tag{4.18}$$

(2) Red-detuned

In the red-detuned case of $\delta_b = \omega_m$, the resonant interaction terms are given by

the beam-splitter Hamiltonian:

$$\hat{H}_a/\hbar = i\left(g^*_{k,a}\hat{a}^r\hat{b}^{r\dagger} - g_{k,a}\hat{a}^{r\dagger}\hat{b}^r\right). \tag{4.19}$$

A normalized output mode in the rotating frame is obtained using the standard cavity input-output relations [42] given in Eq (4.3), together with a mode function designed to match the gain characteristics of the cavity [12]. To achieve this, define input/output mode functions as:

$$u^{\text{in}}_k(t) = \sqrt{\frac{1}{\mathcal{N}_k(\tau)}}e^{i\delta_k t \pm R^{\text{in}}_k(t)}, \tag{4.20}$$

where $R^{\text{in}}_k(t) = \int_0^t G^{\text{in}}_k(t')dt'$ and the normalization is $\mathcal{N}^{\text{in}}_k(\tau) = \int_{\tau^0_k}^{\tau+\tau^0_k} e^{2R^{\text{in}}_k(t)}dt$. There are similar expressions with in $\rightarrow$ out, making four modes in all. The optomechanical gain is $G_k(t) = \pm\left|g_{k,a}\right|^2/\gamma_o = \pm\chi^2 E(t)^2/\gamma_o(\delta_k^2+\gamma_o^2)$. For output modes there is a positive sign for the first (blue-detuned) pulse and negative for the second (red-detuned) pulse. These signs are reversed for the input modes. Since we only are interested in fluctuations, we consistently drop the mean value part of the input-output mode operators.

There is a characteristic phase of $e^{i\phi_b} = -g_{b,a}/\left|g_{b,a}\right|$ for entanglement and $e^{i\phi_r} = g_{r,a}/\left|g_{r,a}\right|$ for readout, where from Eq (4.16), we see that $\phi_k = \tan^{-1}(\gamma_0/\delta_k)$ in both cases. This means that the phases for entanglement and readout have equal magnitudes and opposite signs.

4.4.2 Linearized entanglement generation

In the ideal linearized model, including an adiabatic approximation, the Heisenberg equations resulting from considering only the squeezing or entanglement Hamiltonian, Eq. (4.18), are

$$\begin{aligned}\dot{\hat{a}}^r &= -\gamma_o\hat{a}^r - g_{b,a}\hat{b}^{r\dagger} + \sqrt{2\gamma_o}\hat{a}^{r,\text{in}},\\ \dot{\hat{b}}^r &= -g_{b,a}\hat{a}^{r\dagger}.\end{aligned} \tag{4.21}$$

This has an adiabatic solution, provided $\gamma_o\tau \gg 1$:

$$\begin{aligned}\hat{a}^r &= -\frac{g_{b,a}}{\gamma_o}\hat{b}^{r\dagger} + \sqrt{\frac{2}{\gamma_o}}\hat{a}^{r,\mathrm{in}}\,,\\ \hat{b}^r &= e^{Gt}\hat{b}^r(0) - g_{b,a}\sqrt{\frac{2}{\gamma_o}}e^{Gt}\int_0^\tau ds\, e^{-Gs}\hat{a}^{r,\mathrm{in}}(s)\,,\end{aligned} \tag{4.22}$$

where we have defined $G = G_b^{\mathrm{out}} = \left|g_{b,a}\right|^2/\gamma_o$ as the blue-detuned entanglement gain.

In the rotating frame, with rectangular-shaped pulses and temporal light modes defined in Eq. (4.20), we obtain

$$\begin{aligned}\hat{A}^{\mathrm{in}} &= \sqrt{\frac{2G}{1-e^{-2G\tau}}}\int_0^\tau dt\, e^{-Gt}\hat{a}^{r,in}(t)\,,\\ \hat{A}_\perp^{\mathrm{in}} &= \sqrt{\frac{2G}{e^{2G\tau}-1}}\int_0^\tau dt\, e^{Gt}\hat{a}_\perp^{r,in}(t)\,,\\ \hat{A}^{\mathrm{out}} &= \sqrt{\frac{2G}{e^{2G\tau}-1}}\int_0^\tau dt\, e^{Gt}\hat{a}^{r,out}(t)\,.\end{aligned} \tag{4.23}$$

This leads to the input-output relations

$$\begin{aligned}\hat{A}^{\mathrm{out}} &= \sqrt{\eta_0}\left[e^{G\tau}\hat{A}^{\mathrm{in}} + e^{i\phi_b}\sqrt{e^{2G\tau}-1}\hat{B}^{in\dagger}\right] + \sqrt{1-\eta_0}\hat{A}_\perp^{\mathrm{in}}\,,\\ \hat{B}^{\mathrm{out}} &= e^{G\tau}\hat{B}^{\mathrm{in}} + e^{i\phi_b}\sqrt{e^{2G\tau}-1}\hat{A}^{\mathrm{in}\dagger}\,,\end{aligned} \tag{4.24}$$

where $\hat{B}^{\mathrm{in}} \equiv \hat{b}^r(0)$ and $\hat{B}^{\mathrm{out}} \equiv \hat{b}^r(\tau)$. If $\eta_0 = 1$, then in the limit of $R = G\tau \gg 1$, we get the relations $\hat{A}^{\mathrm{out}} = e^R(\hat{A}^{\mathrm{in}} + e^{i\phi_b}\hat{B}^{in\dagger})$, $\hat{B}^{\mathrm{out}} = e^R(\hat{B}^{\mathrm{in}} + e^{i\phi_b}\hat{A}^{in\dagger})$.

4.4.3 Linearized readout procedure

Analogously to the calculation above, but using the beam-splitter interaction Hamiltonian Eq. (4.19), one obtains the Heisenberg equations for the red-detuned case:

$$\begin{aligned}\dot{\hat{a}}^r &= -\gamma_o\hat{a}^r - g_{r,a}\hat{b}^{r\dagger} + \sqrt{2\gamma_o}\hat{a}^{r,in}\,,\\ \dot{\hat{b}}^r &= g_{r,a}^*\hat{b}^r\,.\end{aligned} \tag{4.25}$$

Introducing the readout gain, $G' = G_r^{\mathrm{out}} = \left|g_{r,a}\right|^2/\gamma_o$, gives the adiabatic solution

$$\begin{aligned}\hat{a}^r &= -\frac{g_{r,a}}{\gamma_o}\hat{b}^r + \sqrt{\frac{2}{\gamma_o}}\hat{a}^{r,in}\,,\\ \hat{b}^r &= e^{-G't}\hat{b}^r(0) - g^*_{r,a}\sqrt{\frac{2}{\gamma_o}}e^{-G't}\int_0^t \mathrm{d}s\, e^{G's}\hat{a}^{r,in}(s)\,.\end{aligned} \tag{4.26}$$

The temporal light-modes of Eq. (4.20) now become

$$\begin{aligned}\hat{A}^{\mathrm{in}}_r &= \sqrt{\frac{2G'}{e^{2G\tau}-1}}\int_{\tau^r_0}^{\tau+\tau^r_0} dt\, e^{G't}\hat{a}^{r,in}\,,\\ \hat{A}^{\mathrm{in}}_\perp &= \sqrt{\frac{2G'}{1-e^{-2G'\tau}}}\int_{\tau^r_0}^{\tau+\tau^r_0} dt\, e^{-G't}\hat{a}^{r,in}_\perp(t)\,,\\ \hat{A}^{\mathrm{out}}_r &= \sqrt{\frac{2G'}{1-e^{-2G\tau}}}\int_{\tau^r_0}^{\tau+\tau^r_0} dt\, e^{-G't}\hat{a}^{r,out}\,.\end{aligned} \tag{4.27}$$

This results in the input-output relation:

$$\begin{aligned}\hat{A}^{\mathrm{out}}_r &= \sqrt{\eta_o}\left[-e^{i\phi_r}\sqrt{\eta_o\,(1-e^{-2G'\tau})}\hat{B}^{\mathrm{in}}_r + e^{-G'\tau}\hat{A}^{\mathrm{in}}_r\right]\\ &\quad+\sqrt{1-\eta_o}\hat{A}^{\mathrm{in}}_{\perp r}\,,\end{aligned} \tag{4.28}$$

where $\hat{B}^{\mathrm{in}}_r \equiv \hat{B}^{\mathrm{out}}$ gives the initial operator for the mechanical oscillator, which is now entangled by the previous interaction, and $\hat{A}^{\mathrm{out}}_r$ and $\hat{A}^{\mathrm{in}}_r$ give the output and input red-detuned pulses. The coupling is such that in the large gain limit $R' \equiv G'\tau \gg 1$, with $\eta_0 = 1$, the red-detuned output pulse $\hat{A}^{\mathrm{out}}_r$ gives a readout of the initial amplitude $\hat{B}^{\mathrm{in}}_r$ of the mechanical oscillator after entanglement.

This is an idealized picture which helps to explain intuitively why these experiments can be regarded as quantum measurements of the oscillator position. In the quantum simulations carried out here, we include a full multi-mode analysis of all output fields, including losses and without adiabatic or single-mode approximations.

4.5 Entanglement and EPR-steering measures

To signify entanglement and EPR steering, it is convenient to consider quadrature measurements. The entangling and readout pulse output field quadratures are given by

$$\hat{X}_k^\theta = \frac{1}{2}\left[e^{-i\theta}\hat{A}_k(\tau) + e^{i\theta}\hat{A}_k^\dagger(\tau)\right] \tag{4.29}$$

respectively. We denote $\hat{P}_k^\theta = \hat{X}_k^{\theta+\pi/2}$, with the angle dropped when $\theta = 0$. We see from the previous section that that for large gain $\hat{A}_r{}^{\text{out}} \to -e^{i\phi_r}\hat{B}_r^{\text{in}}$. For the full entanglement protocol, consisting of a blue-detuned entanglement pulse followed by a red-detuned readout pulse, in the limit $R \gg 1$ as well as $R' \gg 1$, this yields $\hat{X}_b + \hat{X}_r^{\phi_b+\phi_r} \to 0$ and $\hat{P}_b - \hat{P}_r^{\phi_b+\phi_r} \to 0$, where $\hat{X}_b$ ($\hat{P}_b$) and $\hat{X}_r$ ($\hat{P}_r$) mean the quadratures corresponding to the output field of the first (blue-detuned) and the second (red-detuned) pulse, respectively. Hence, the final quadrature of the mechanical oscillator

$$\hat{X}_m^\varphi(\tau) = \frac{1}{2}\left[e^{-i\varphi}\hat{b}^r(\tau) + e^{i\varphi}\hat{b}^{r\dagger}(\tau)\right] \tag{4.30}$$

is transferred to the red-detuned output pulse quadrature, $\hat{X}_r^\theta$, with a sign change.

This leads to measurable entanglement and EPR correlations between the red- and blue-detuned output field quadratures. In the numerical calculations, we optimize the phase choice θ for best entanglement, as is also done experimentally. We calculate both the correlations of the mechanical oscillator with the entangling blue-detuned pulse, and also the correlations between the two outputs, as measured in current experiments.

4.5.1 Entanglement measures

We next consider how to signify entanglement between the reflected entangling field $\hat{X}_b^\theta$ and either the mechanical quadrature $\hat{X}_m^\theta$ or the readout pulse $\hat{X}_r^\theta$. For clarity, we generally write $\hat{X}_m^\theta$ in the following criteria, although we also simulate results with a readout pulse $\hat{X}_r^\theta$, as measured in current experiments. There are numerous possible entanglement measures.

Product criteria

The most robust measures are the *product* signatures first introduced for EPR-steering [15] and applied to entanglement by Tan [47]. The most general inequalities of this type include a measurement gain g. Entanglement is verified if [38, 48]

$$\Delta^{p}_{\text{ent}} = \frac{4\Delta(\hat{X}_m - g\hat{X}^{\theta}_b)\Delta(\hat{P}_m + g\hat{P}^{\theta}_b)}{(1+g^2)} < 1, \tag{4.31}$$

where we introduce the notation $(\Delta\hat{x})^2 \equiv \langle\hat{x}^2\rangle - \langle\hat{x}\rangle^2$, $\langle\hat{x},\hat{y}\rangle = \langle\hat{x}\hat{y}\rangle - \langle\hat{x}\rangle\langle\hat{y}\rangle$, and take $\varphi = 0$. Here g represents real gains used in postprocessing the data. The gain g and phase θ is optimized numerically in some of our simulations, but not all, to show the effect of the optimization.

We assume for simplicity the same gain value for both variances. For example, if we only optimize the gain, then from minimizing Δ^{p}_{ent}, one obtains

$$g = \frac{1}{2a}\left[-b + \sqrt{b^2 - 4ac}\right], \tag{4.32}$$

where $c = \langle\hat{X}_m, \hat{X}_b\rangle = -a$ and $b = (\Delta\hat{X}_b)^2 - (\Delta\hat{X}_m)^2$.

For dealing with the red-detuned readout pulse, we replace $\hat{X}_m$ by $-\hat{X}_r$, and $\hat{P}_m$ by $-\hat{P}_r$, as motivated in the previous section, giving the criterion

$$\Delta^{p}_{\text{ent}} = \frac{\Delta(\hat{X}_r + g\hat{X}_b)\Delta(\hat{P}_r - g\hat{P}_b)}{(1+g^2)/4} < 1. \tag{4.33}$$

Additive criteria

A weaker *additive* criterion can be obtained from this, which is

$$\Delta^{a}_{\text{ent}} = \frac{\Delta^2(\hat{X}_m - g\hat{X}_b) + \Delta^2(\hat{P}_m + g\hat{P}_b)}{(1+g^2)/2} \quad < \quad 1\,, \tag{4.34}$$

where g can again be chosen optimally. The choice $g = 1$ is not always optimal but leads to one of the most widely used entanglement criteria [25]:

$$\Delta^{a}_{\text{ent}} = [\Delta(\hat{X}_m - \hat{X}_b)]^2 + [\Delta(\hat{P}_m + \hat{P}_b)]^2 < 1\,. \tag{4.35}$$

The difference is that $\Delta^p_{\text{ent}} < 1$ is a more robust entanglement signature. This detects entanglement that the weaker criterion of $\Delta^a_{\text{ent}} < 1$ is unable to detect. As described above, for dealing with the red-detuned readout pulse, we replace $\hat{X}_m$ by $-\hat{X}_r$, and $\hat{P}_m$ by $-\hat{P}_r$. We use the more robust product signatures in our numerical results, except for comparison to electromechanical experiments where the additive criterion with $g = 1$ was used to process the data.

Using the input-output relations of the linearized adiabatic model, one obtains after some algebra a theoretical adiabatic prediction for Eq. (4.35) of

$$\Delta^a_{\text{ent,ad}} = \left(\bar{n}_{b,0} + 1\right) \cdot \left(e^R - \sqrt{e^{2R} - 1}\right)^2 . \tag{4.36}$$

However, this simplified picture neglects inefficiencies, as well as the thermal noise added to the mechanical oscillator as it interacts with its thermal reservoir.

Comparisons of entanglement criteria

Fig. 4.2 shows a comparison of the product and additive entanglement criteria for the conditions of an optomechanical experiment described in Section 4.6.2, with a bath temperature chosen as $T = 500$mK. One can see that the gain-optimized product criterion is more robust than the additive criterion where $g = 1$ has been fixed; that is, it is more sensitive to the presence of entanglement. This can be seen in both the adiabatic predictions and the full simulation results.

This graph also shows how the accuracy of the adiabatic model changes with pump strength. While there is good agreement between the simulation results for Δ^a_{ent} and Δ^p_{ent} and the adiabatic predictions for small values of R; that is, for a small pump strength, the simulation results diverge from their theoretical predictions for higher pump strengths. The entanglement signatures begin to deteriorate from a value of $R \approx 1$ for Δ^a_{ent} and $R \approx 0.8$ for Δ^p_{ent}.

Surprisingly, the phase- and gain-optimized product criterion is in very good agreement with the adiabatic prediction for Δ^p_{ent}, even though the adiabatic prediction assumes that $\theta = \varphi = 0$, which in fact are not the optimal parameters.

The third subgraph of Fig. 4.2 shows the optimal phase angle that corresponds to the black curve. The phase shift of $\Delta\varphi \approx -0.075$ is as expected from the adiabatic prediction, as it is approximately equal to $-\gamma_o/\omega_m = -0.0703$.

4.5.2 EPR-steering criteria

The more demanding EPR-steering paradox is established through violation of the Reid-EPR inequality [15, 16]. This tests whether an inferred Heisenberg uncertainty relation is violated under the assumptions of local realism (LR). The paradox is *directional*, so that one obtains different criteria depending on the direction of inference. If one infers the oscillator position from the blue readout, then the paradox occurs if

$$\begin{aligned} E_{m|b} &= 4\Delta_{\text{inf}}\hat{X}_m \Delta_{\text{inf}}\hat{P}_m \,, \\ &= 4\Delta(\hat{X}_m - g\hat{X}_b^\theta)\Delta(\hat{P}_m + g\hat{P}_b^\theta) < 1 \,. \end{aligned} \tag{4.37}$$

Here $\Delta_{\text{inf}}\hat{X}_m$, $\Delta_{\text{inf}}\hat{P}_m$ are the inferred uncertainties, and the optimal gain is given as $g = \langle \hat{X}_m, \hat{X}_b \rangle / (\Delta\hat{X}_b)^2 = -\langle \hat{P}_m, \hat{P}_b \rangle / (\Delta\hat{P}_b)^2$, owing to the symmetries of this problem. For a Reid-EPR paradox achieved by condition (4.37), measurements on the pulse system enable "steering" of the mechanical oscillator m [21, 22]. This criterion with optimal g_x, g_p is necessary and sufficient for EPR steering in two-mode Gaussian systems [15, 21].

A thermal barrier exists for this paradox. Figure 4.6 shows that the mechanical oscillator is steerable by the optical pulse system when $r > r_0$, where a minimum strength r_0 of the gain parameter required, for a given $n_{b,0}$. A thermal barrier means that a threshold level of pulse-oscillator interaction is required for a given thermal occupation $n_{b,0}$ of the oscillator.

An EPR paradox can be shown the other way, by the criterion

$$E_{b|m} = 4\Delta(\hat{X}_b - g\hat{X}_m^\phi)\Delta(\hat{P}_b + g\hat{P}_m^\phi) < 1. \tag{4.38}$$

We similarly modify the EPR-steering criteria when observing two correlated optical outputs, replacing $\hat{X}_m$ by $-\hat{X}_r$, and $\hat{P}_m$ by $-\hat{P}_r$.

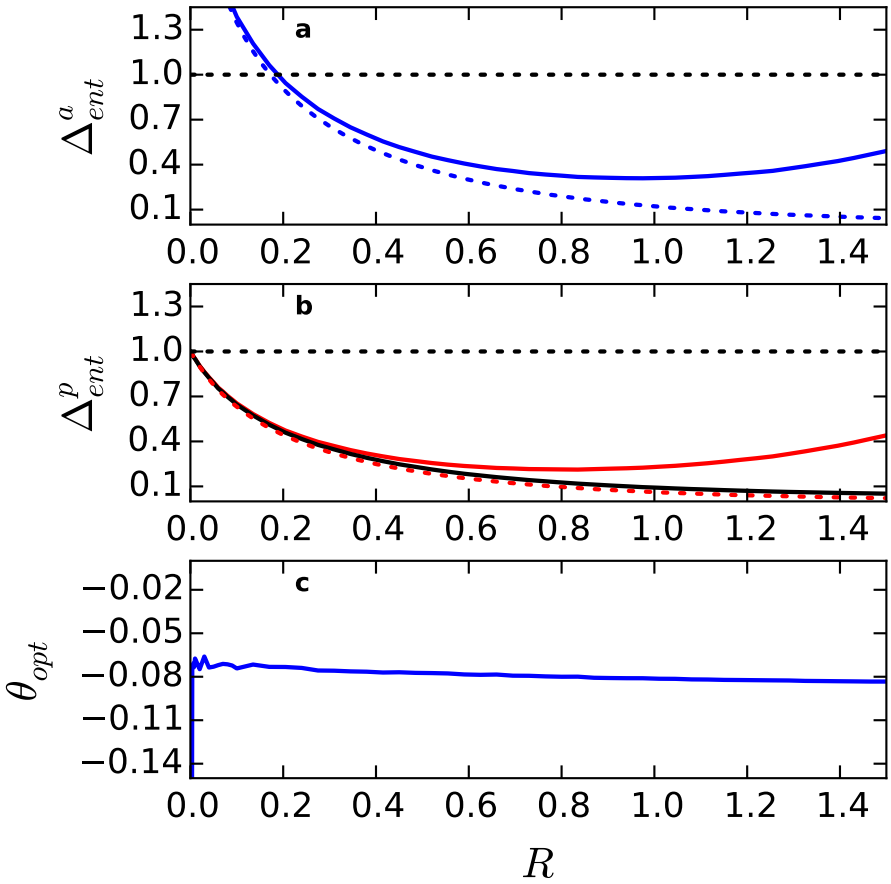

Figure 4.2: Various entanglement signatures are plotted vs R. Plots (a), (b), and (c) show the simulation results and approximate adiabatic predictions for the optomechanical experiment described in Sec. 4.6.2, with an initial and heat bath temperature of $T = 500mK$. Plot (a) shows the additive criterion $\Delta^a_{\text{ent}}$where the gain factor is fixed at $g = 1$, with the solid line representing the simulation results and the dotted line the adiabatic prediction. Plot (b) shows the product criterion Δ^p_{ent} where the gain factor g is chosen optimally to minimize Δ^p_{ent}, but the relative phase is fixed at $\theta = 0$. Here again the solid line represents the simulation results and the dotted line the approximate adiabatic prediction. The black curve shows the simulation result for the product criterion Δ^p_{ent}, where both g and θ are chosen optimally. Plot (c) shows the optimal phase angle θ that corresponds to the black curve.

4.6 Quantum simulations

We assume the input state of both light fields and the mirror to be a thermal state with mean excitation number $\bar{n}_{k,0}$, so that $\Delta\hat{X}_k(0) = \Delta\hat{P}_k(0) = \sqrt{\bar{n}_{k,0}/2 + 1/4}$. Our quantum simulations fully model both the blue- and red-detuned pulses, without linearization or adiabatic assumptions. Plots of the results of our full simulations with the red-detuned output pulse are shown together with comparisons to idealized mechanical oscillator measurements.

4.6.1 Electromechanical experiment

In the pioneering electromechanical experiment carried out recently by Palomaki *et al.* [9], an inductor-capacitor (LC) resonator that interacts with an external microwave field was coupled to a mechanical oscillator. This corresponds to the usual optomechanical experiments, except with an LC resonator replacing the optical cavity and a microwave field replacing to the laser driving field. We give our simulation results for this case first. The good agreement with experimental measurements verifies our quantum simulations. This also indicates some of the limitations of the experiment, due to noise and inefficiency issues.

Electromechanical parameters

We simulated the recent microwave experiment leading to quantum entanglement, using the published parameter values [9] and the reported quantum efficiency of $\eta = 0.2$. The electromechanical (microwave) experimental parameters are $\omega_m/2\pi = 10.34\,\mathrm{MHz}$ for the mechanical oscillator frequency, $\chi_0/2\pi = 200\,\mathrm{Hz}$ for the coupling constant between LC resonator and mechanical oscillator, $\gamma_{\mathrm{int}}/2\pi = 30\,\mathrm{kHz}$ for the internal LC resonator dissipation rate, $\gamma_{\mathrm{ext}}/2\pi = 150\,\mathrm{kHz}$ for the external LC resonator dissipation rate, $\gamma_m/2\pi = 17.5\,\mathrm{Hz}$ for the mechanical oscillator dissipation rate, $\tau_b = \tau_r = 35.5\,\mu\mathrm{s}$ for the pulse duration of both the blue- and red-detuned pulse, and $\tau_{del} = 10\,\mu\mathrm{s}$ for the delay time between the two pulses. The experimental paper quoted intensity decay rates, while we consistently use amplitude decay rates to quantify dissipation.

In the electromechanical experiments, the mechanical oscillator was precooled so that the initial phonon occupation number was $\bar{n}_{m,0} = 0.5 \pm 0.1$. The LC resonator

experienced excess technical noise, giving initial variances of $\bar{n}_{o,0} = 0.12 \pm 0.02$. During the time of the experiment, the mechanical oscillator was in contact with a heat bath at 19 mK, so that $\bar{n}_{m,\text{th}} = 37.8$. There was an initial wait period of 15 μs, during which time we calculate there was a preheating from the heat bath to an occupation number of $\bar{n}_{m,0} = 0.62 \pm 0.1$. An additional electromagnetic noise of $\bar{n}_{o,\text{th}} = 0.12 \pm 0.02$, the same as found initially, was present in both the internal and external input reservoirs of the LC resonator. This additional noise and corresponding relaxation makes for a substantial difference with the optomechanical situation. It is caused by the typically higher technical noises found with microwave sources, as compared with optical laser sources of radiation.

The quantum efficiency of the detector was experimentally measured as $\eta = 0.20 \pm 0.01$. The pump strength of the transfer pulse was fixed at $\Gamma_r/2\pi = 11$ kHz while the pump strength of the entanglement pulse was varied over a range of $\Gamma_b/2\pi = 0 \ldots 18$ kHz. These were defined as intensity gains or losses, so that in our notation $G_k = \pm\Gamma_k/2$.

Entanglement simulation results

The results of the simulations are compared to experiment in Fig 4.3. This used an additive entanglement criterion with unit gain ($g = 1$), and a relative phase adjusted to optimize the correlations.

To simplify comparisons, we use the notation and entanglement signatures used in the experiment. The definition for the pump strength given in the experimental paper was

$$\Gamma_k = \frac{2\chi^2 \bar{n}_{k,ind}}{\gamma_o}, \tag{4.39}$$

where $\bar{n}_{k,ind}$ is the number of coherently induced photons in the cavity under adiabatic conditions, i.e.,

$$\bar{n}_{k,ind} = \left\langle |\alpha|^2 \right\rangle_\infty \approx \frac{E^2}{\delta^2 + \gamma_o^2}. \tag{4.40}$$

Using the relationship $E_k^2 \tau_k = 2\gamma_{\text{ext}} N_k$, which gives the pump amplitude E_k in terms of the total number of external pump photons N_k, we conclude that the experimental parameters Γ_r, Γ_b can be expressed in terms of the earlier definitions as

$$\Gamma_k = \frac{4\chi^2 \eta_o N_k}{(\omega_m^2 + \gamma_o^2)\, \tau_k}, \tag{4.41}$$

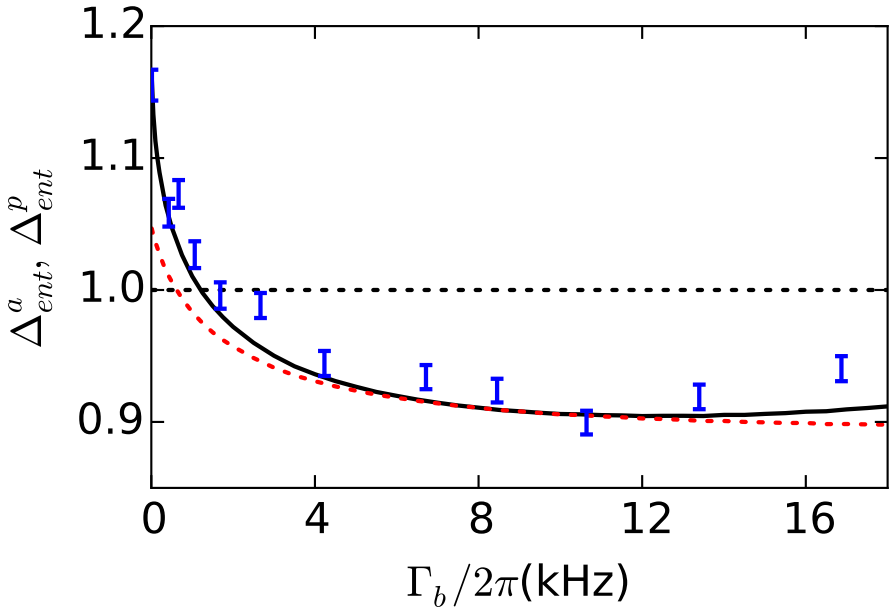

Figure 4.3: Predicted entanglement signatures vs pump-strength Γ_b for the microwave experiment of Ref. [9], using their notation. The black curve gives the additive entanglement signature Δ^a_{ent} at $g = 1$. This shows good agreement with the experimental data, and also excellent agreement with a linearized calculation in Ref. [9], except for small differences at large Γ_b. The circles are experimental values for Δ^a_{ent}. The red dotted line gives the result of the phase- and gain-optimized product signature, Δ^p_{ent}.

where $k = [r, b]$, and $\gamma_o = \gamma_{\text{ext}} + \gamma_{\text{int}}$ denotes the total dissipation rate for the LC resonator. For comparison purposes, our amplitude gain G_k is therefore given by

$$G_k = \pm\Gamma_k/2\,. \tag{4.42}$$

In summary, the figure gives a comparison of the exact simulations vs experimental results for electromechanical pulse entanglement at microwave frequencies. We obtain good agreement between the simulation and the experimental data of the experiment of Palomaki *et al.* [9]. These results are also in agreement with linearized calculations [9], thus verifying that these approximations are valid. However, the exact simulations can also be used in nonlinear regimes where linearization will fail.

EPR-steering violations

It is known that EPR steering is not possible when the efficiency reduces below 0.5 for the steering (pulse) system [49]. As a result, the microwave experimental parameters

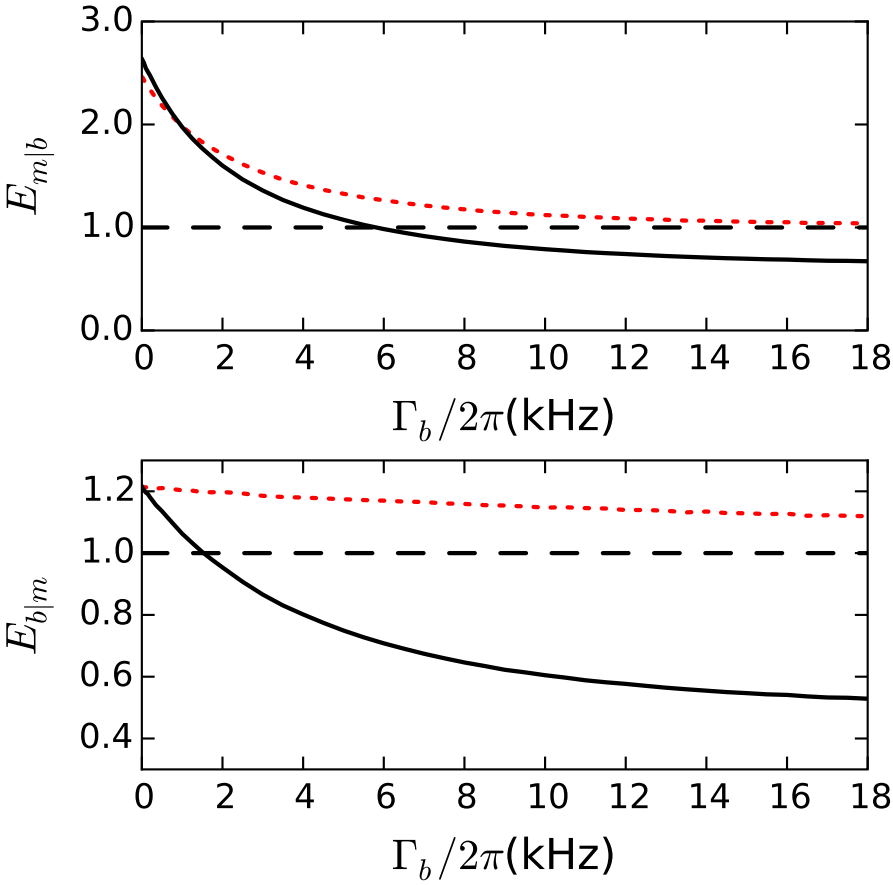

Figure 4.4: (Color online) EPR-steering versus entanglement pump strength Γ_b , for the electromechanical conditions used in Fig. 3, except with $\eta = 0.9$. Dashed lines indicate the result of using optical readouts; solid lines assume direct oscillator measurements. The predictions are for the electromechanical microwave experiment. Although we assume a higher efficiency, the upper curves show how the noisy readout process degrades EPR steering in the output fields. The lower curves assume a direct, low-noise readout of the oscillator position is obtained through some other technique.

are less favorable for EPR steering owing to low quantum efficiencies of $\eta \approx 0.2$. For this reason, we assume a higher efficiency of $\eta = 0.9$ in these EPR calculations. After all, detector efficiency is a moving target, and much higher efficiencies may occur in future.

Even with higher efficiency, however, our simulation reveals no steering for external measurements using the noise parameters of this experiment (inset Figure 4.4), due to the high input noise levels in the measuring pulse inputs. Experiments with lower input noise levels are necessary for EPR-steering observation.

4.6.2 Optomechanical experiment

In the optomechanical simulations, we choose typical parameters that correspond to recent optomechanical experiments. The highest temperatures we simulate are $4K$. This is lower than in some reported data, although this temperature appears to be in the currently accessible range. We calculate the robust product entanglement measure Δ^{p}_{ent}, with optimal gain and phase.

Si optomechanical crystal structures

Parameters are reported in Ref. [2], with $\omega_m/2\pi = 3.7\,\text{GHz}$, $Q_m = \omega_m/\gamma_m = 10^5$, $\gamma_m/2\pi = 37\,\text{KHz}$, $\gamma_o/2\pi = 0.26\,\text{GHz}$, $\chi_0/2\pi = 910\,\text{KHz}$. We choose the photon number for the red transfer pulse as $N_r = 24.6 \times 10^6$, corresponding to a readout gain of $R' = 3$, while for the first pulse a photon number up to $N_b = 8.2 \times 10^6$ is used, corresponding to an entanglement gain $R = 1$. Both pulses have a duration of $\tau = 0.04\,\mu\text{s}$, and the delay time is set to $\tau_{del} = 0.008\,\mu\text{s}$. We assumed $\gamma_{\text{int}} = 0$ – i.e., perfect output coupling – but we include an imperfect detector efficiency of $\eta = 0.9$ for optical detectors. In practice, η should be regarded as including all types of detection losses, including optical coupling losses.

Results for square pulses are presented here, although a variety of pulse shapes ranging from square waves to Gaussians gave strong entanglement and steering. Two different heat bath temperatures were chosen for comparison purposes: either with a cold reservoir at $T_{\text{bath}} = 200\,\text{mK}$, or a "warm" reservoir of $T_{\text{bath}} = 4\,\text{K}$. In both cases, the initial mechanical occupation number was chosen as $\bar{n}_{m,0} = 0.7$, corresponding to an initial precooled oscillator temperature of 200 mK. In the case of these optical simulations, we note that $dW^{\text{in}}_{\text{ext}} = dW^{\text{in}}_{o}$, since we assumed there were no internal loss mechanisms, and the optical thermal occupation is assumed to be negligible.

Entanglement results

The resulting predictions for entanglement in the case of the optomechanical parameters are plotted in Fig. 4.5, where the solid lines indicate results for idealized measurements on the mechanical oscillator, and dotted lines the expected operational measurements using a readout pulse.

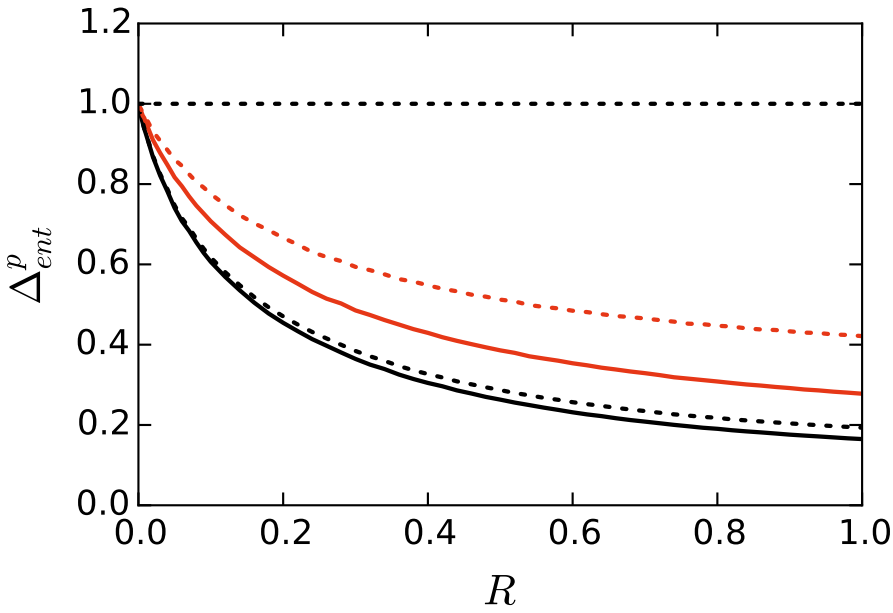

Figure 4.5: The entanglement signature $\Delta^p_{\rm ent}$ is plotted vs R, for an initial mechanical occupation number $\bar{n}_{m,0} = 0.7$ and two values of heat bath occupation $\bar{n}_{m,\rm th} = 0.7$(lower), 22.1(upper) for the oscillator corresponding to temperatures $T = 200\,$mK and $4\,$K. The optical cavity bath is assumed a vacuum, i.e. $\bar{n}_{o,0} = \bar{n}_{o,\rm th} = 0$. The solid lines give the oscillator entanglement; the dotted lines give results of the probe measurements.

Using these entanglement signatures, we have simulated the robust asymmetric EPR entanglement recently predicted in Ref. [38], but without approximations. A graph of the predicted quantum entanglement at low and high temperature is shown in Fig. 4.5 for two experimental scenarios. These calculations simulate the parameter regime of experiments on Si optomechanical crystal structures [2].

Further, we are able to establish the validity of the linearization assumptions, and agreement between the exact positive-P representation and the approximate Wigner method, for both optical and microwave experimental parameter values. Our investigation tells us that nonlinear quantum noise and deviation between exact positive P and approximate methods come into play only for stronger couplings, with $\chi_0 \sim \gamma_o$.

The physical interpretation of the results of Fig. 4.5 is that for any given initial mechanical oscillator occupation number $\bar{n}_{m,0}$, we can always obtain entanglement for $R = \int_0^t G(t')dt' > 0$, provided one uses the asymmetric criteria (4.31) and selects an optimal choice of both phase θ and gain factor g (4.32). This means we can in principle detect entanglement in the presence of thermal mechanical decoherence, without the need to use laser cooling to reduce the value of $\bar{n}_{m,0}$. Figure 4.5 indicates entanglement

at a temperature $T_{\text{bath}} = 4\,\text{K}$ ($\bar{n}_{m,\text{th}} = 22.1$), provided the oscillator is pre-cooled to $T_0 = 200\,\text{mK}$ ($n_{m,0} \sim 0.7$). This is sensitive to the occupation number $\bar{n}_{m,\text{th}}$ of the mechanical heat bath, but is more robust to thermal effects than using the symmetric criterion in Ref. [12].

EPR-steering results

Predictions for optomechanical EPR steering experiments are presented in Fig. 4.6. The efficiency assumed here is $\eta = 0.9$. Generally, these simulations have more favorable conditions for EPR steering, due to the relatively lower thermal occupations for optical inputs as compared to microwave inputs, even assuming the same temperatures. A thermal barrier can still exists for this paradox however.

The results show that the mechanical oscillator is steerable by the optical pulse system when $R > R_0$, where a minimum strength R_0 of the gain parameter required for a given $\bar{n}_{m,0}$. A thermal barrier means that a threshold level of pulse-oscillator interaction is required for a given initial thermal occupation $\bar{n}_{m,0}$ of the mechanical oscillator.

An EPR paradox can also be shown the other way, in which we steer the photons, not the massive oscillator, by the criterion

$$E_{b|m} = 4\Delta(\hat{X}_b - g\hat{X}_m^{\phi})\Delta(\hat{P}_b + g\hat{P}_m^{\phi}) < 1. \tag{4.43}$$

Figure 4.4 shows that this is possible for any value of initial oscillator noise $\bar{n}_{m,0}$, and for any gain. There is no equivalent thermal barrier for the optical pulse "steered" by measurements made on the mechanical system, if the entangling pulse is not thermally excited. We also find that $E_{b|m}$ is less sensitive to mechanical decoherence. This is because we can select optimal gain values g to reduce the effect of the initial thermal noises $\bar{n}_{m,0}$ and the mechanical heat bath $\bar{n}_{m,\text{th}}$. The graphed results also calculate the externally measured criteria, $E_{b|r}$ and $E_{r|b}$, using pulse-probe methods, with reduced correlations.

4.6.3 Numerical methods

A semi-implicit interaction picture stochastic integration method was used [46], with 8×10^5 trajectories for positive-P simulations, up to 2×10^5 for the truncated Wigner simulations, and 10^4 total time steps. The two different phase-space methods gave

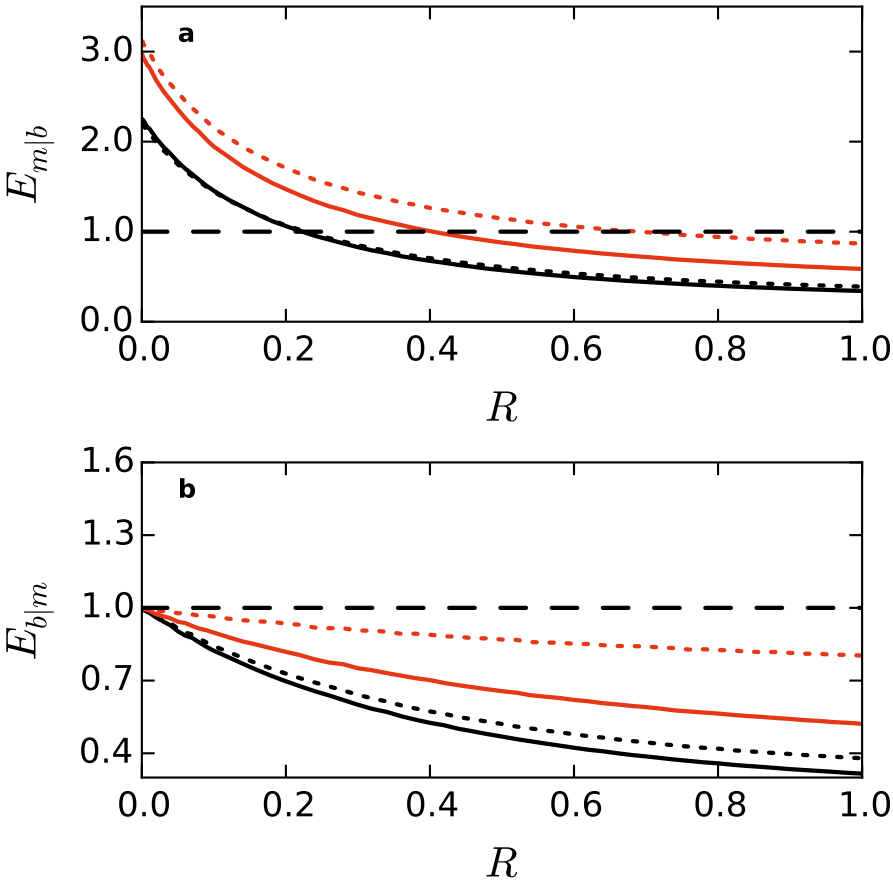

Figure 4.6: (Color online) EPR-steering vs entanglement gain parameter R, for the optomechanical conditions used in Fig. 5. Dashed lines indicate the result of using optical readouts, solid lines assume direct oscillator measurements. The lower black curves are for $T_{\text{bath}} = 200\,\text{mK}$ ($\bar{n}_{m,\text{th}} = 0.7$); the upper red curves are for $T_{\text{bath}} = 4\,\text{K}$ ($\bar{n}_{m,\text{th}} = 22.1$). (a) EPR steering of the mechanical system m is possible if $E_{m|b} < 1$. (b) EPR steering of the optical system is possible if $E_{b|m} < 1$. For both bath temperatures, two-way steering becomes possible once a threshold $R > R_0$ is reached.

identical results for these parameter values, apart from typical sampling errors with relative errors in the mean of up to ± 0.01, hence only one line is plotted for both methods. Truncation errors due to the finite time step were verified to be less than the sampling errors. To minimize software or hardware errors, independent computer codes were written and tested in different languages (C++ and SCILAB) and with different computer hardware, both using double precision floating point. These gave identical results in all cases tested.

4.7 Summary

Optomechanics presents a challenge for exact quantum simulations. It combines a range of occupation numbers and time scales with nonequilibrium and nonlinear open system quantum dynamics. These results demonstrate that the positive-P representation approach can give a useful first principles simulation of the standard model in the quantum regime [30]. For the parameters simulated here, the truncated Wigner approach is also reliable and simple to implement. This method needs to be verified by the more precise positive-P simulations for strong couplings. To the extent we can make comparisons, our results also verify previous analyses using adiabatic techniques.

These simulation techniques can be readily scaled up to study multipartite systems with many oscillator modes [37], strong interactions, and non-adiabatic behavior. Our main result in this work illustrates a fundamental physical principle: *inferred changes* to a massive system as a result of measurement at a distant site will be inhibited by thermal noise. We note that the reverse type of inference is not inhibited, which is typical of the directional property of EPR-steering inference. It suggests that the mechanism for the apparent directional property of quantum measurement could be fundamentally related to noise sources. Direct observation of two-way EPR inference for a massive system is therefore an important next goal in nanomechanics.

4.8 Acknowledgments

We acknowledge support from the Australian Research Council via Discovery and DECRA grants. This work was also supported in part by the National Science Foundation under Grant No. PHYS-1066293 and the hospitality of the Aspen Center for Physics. thank Palomaki *et al.* for details of their experiment.

4.9 Appendix: Effect of losses and inefficiencies

In the detection of either optical or microwave fields, we assume that external detection is carried out with finite efficiency η. Unless stated otherwise, optical detection efficiencies are assumed to be at $\eta = 0.9$, while microwave photons are assumed detected with $\eta = 0.2$, as reported experimentally.

In order to take this into account, we use a simple model, in which detector inefficiency is modeled as a beam splitter that allows only some of the incident photons to reach the detector. For consistency, one must include a vacuum input so that commutation relations are still valid. This gives

$$\hat{a}_k^d = \sqrt{\eta}\hat{a}_k + \sqrt{1-\eta}\hat{a}_k^v ,$$

where $\hat{a}^d$ means the detected amplitude, $\hat{a}^v$ means the vacuum amplitude that is input to the beam-splitter, and $k = [b, r]$ as previously. As the inefficiency terms are often combined with gain terms, because these effects compensate each other, we define $\Gamma \equiv (1-\eta)\left(1+g^2\right)$.

This model yields for the detected quadratures

$$\begin{aligned} \hat{X}_k^d &= \sqrt{\eta}\hat{X}_k + \sqrt{1-\eta}\hat{X}_k^v , \\ \hat{P}_k^d &= \sqrt{\eta}\hat{P}_k + \sqrt{1-\eta}\hat{P}_k^v . \end{aligned}$$

The variance terms corresponding to differences of quadratures in two beams now become

$$\begin{aligned} \sigma^2 &= \left\langle (\hat{X}_r^d - g\hat{X}_b^d)^2 \right\rangle \\ &= \frac{\Gamma}{4} + \eta \left\langle \left(\hat{X}_r - g\hat{X}_b\right)^2 \right\rangle . \end{aligned}$$

4.9.1 Entanglement criteria

For the additive entanglement criterion at unit gain, the inclusion of detector inefficiency in the calculation gives entanglement if $\Delta_{\text{ent}}^a < 1$, where

$$\Delta_{\text{ent}}^a = (1-\eta) + \eta\left\{\Delta^2\left(\hat{X}_r + \hat{X}_b\right) + \Delta^2\left(\hat{P}_r - \hat{P}_b\right)\right\}$$

Including gain, the product criterion indicates entanglement, if $\Delta_{\text{ent}}^p < 1$, where

$$\begin{aligned}\Delta^{p}_{\text{ent}} &= (1-\eta)\left[1+16\eta^2\frac{\Delta^2\left(\hat{X}_r+g\hat{X}_b\right)\Delta^2\left(\hat{P}_r-g\hat{P}_b\right)}{\Gamma^2}\right.\\ &\left.+4\eta\frac{\left[\Delta\left(\hat{X}_r+g\hat{X}_b\right)\right]^2+\left[\Delta\left(\hat{P}_r-g\hat{P}_b\right)\right]^2}{\Gamma}\right]^{\frac{1}{2}}\end{aligned}$$

4.9.2 EPR-steering criteria

Similarly, the *measured* EPR-steering criteria including detection inefficiency become $E^2_{r|b} < 1$ where

$$\begin{aligned}E^2_{r|b} &= \Gamma^2+16\eta^2\Delta^2\left(\hat{X}_r+g\hat{X}_b\right)\Delta^2\left(\hat{P}_r-g\hat{P}_b\right)+\\ &+4\eta\Gamma\left(\Delta^2\left(\hat{X}_r+g\hat{X}_b\right)+\Delta^2\left(\hat{P}_r-g\hat{P}_b\right)\right)\end{aligned}$$

and $r \leftrightarrow b$ for the other steering direction. We note that this is much harder to satisfy when $\eta < 1$ than in the perfect efficiency case. The physical reason is simply that the injection of uncorrelated vacuum noise through the inefficient detection process tends to reduce the strong correlation required to observe an EPR paradox in these experiments.

4.10 Bibliography

[1] A. D. O'Connell, M. Hofheinz, M. Ansmann, R. C. Bialczak, M. Lenander, E. Lucero, M. Neeley, D. Sank, H. Wang, M. Weides, J. Wenner, J. M. Martinis and A. N. Cleland, "Quantum ground state and single-phonon control of a mechanical resonator", Nature **464**, 697–703 (2010) (page 58).

[2] J. Chan, T. P. M. Alegre, A. H. Safavi-Naeini, J. T. Hill, A. Krause, S. Groblacher, M. Aspelmeyer and O. Painter, "Laser cooling of a nanomechanical oscillator into its quantum ground state", Nature **478**, 89–92 (2011) (pages 58, 79, 80).

[3] S. Groblacher, J. B. Hertzberg, M. R. Vanner, G. D. Cole, S. Gigan, K. C. Schwab and M. Aspelmeyer, "Demonstration of an ultracold micro-optomechanical oscillator in a cryogenic cavity", Nat Phys **5**, 485–488 (2009) (page 58).

[4] J. D. Teufel, D. Li, M. S. Allman, K. Cicak, A. J. Sirois, J. D. Whittaker and R. W. Simmonds, "Circuit cavity electromechanics in the strong-coupling regime", Nature **471**, 204–208 (2011) (page 58).

[5] A. H. Safavi-Naeini, J. Chan, J. T. Hill, T. P. M. Alegre, A. Krause and O. Painter, "Observation of quantum motion of a nanomechanical resonator", Phys. Rev. Lett. **108**, 033602 (2012) (page 58).

[6] N. Brahms, T. Botter, S. Schreppler, D. W. C. Brooks and D. M. Stamper-Kurn, "Optical detection of the quantization of collective atomic motion", Phys. Rev. Lett. **108**, 133601 (2012) (page 58).

[7] T. J. Kippenberg and K. J. Vahala, "Cavity optomechanics: Back-action at the mesoscale", Science **321**, 1172–1176 (2008) (page 58).

[8] M. Aspelmeyer, S. Gröblacher, K. Hammerer and N. Kiesel, "Quantum optomechanics – throwing a glance [invited]", J. Opt. Soc. Am. B **27**, A189–A197 (2010) (page 58).

[9] T. A. Palomaki, J. D. Teufel, R. W. Simmonds and K. W. Lehnert, "Entangling mechanical motion with microwave fields", Science **342**, 710–713 (2013) (pages 58–60, 65, 75, 77).

[10] V. Giovannetti, S. Mancini and P. Tombesi, "Radiation pressure induced Einstein–Podolsky–Rosen paradox", EPL (Europhysics Letters) **54**, 559 (2001); S. Mancini, V. Giovannetti, D. Vitali and P. Tombesi, "Entangling macroscopic oscillators exploiting radiation pressure", Phys. Rev. Lett. **88**, 120401 (2002), pages 58, 66.

[11] C. Genes, A. Mari, P. Tombesi and D. Vitali, "Robust entanglement of a micromechanical resonator with output optical fields", Phys. Rev. A **78**, 032316 (2008) (page 58).

[12] S. G. Hofer, W. Wieczorek, M. Aspelmeyer and K. Hammerer, "Quantum entanglement and teleportation in pulsed cavity optomechanics", Phys. Rev. A **84**, 052327 (2011) (pages 58–60, 65–67, 81).

[13] M. R. Vanner, I. Pikovski, G. D. Cole, M. S. Kim, Č. Brukner, K. Hammerer, G. J. Milburn and M. Aspelmeyer, "Pulsed quantum optomechanics", Proceedings of the National Academy of Sciences **108**, 16182–16187 (2011); I. Pikovski, M. R. Vanner, M. Aspelmeyer, M. S. Kim and C. Brukner, "Probing Planck-scale physics with quantum optics", Nat Phys **8**, 393–397 (2012), page 58.

[14] A. Einstein, B. Podolsky and N. Rosen, "Can quantum-mechanical description of physical reality be considered complete?", Phys. Rev. **47**, 777–780 (1935) (page 58).

[15] M. D. Reid, "Demonstration of the Einstein–Podolsky–Rosen paradox using nondegenerate parametric amplification", Phys. Rev. A **40**, 913–923 (1989) (pages 58, 71, 73).

[16] M. D. Reid, P. D. Drummond, W. P. Bowen, E. G. Cavalcanti, P. K. Lam, H. A. Bachor, U. L. Andersen and G. Leuchs, "*Colloquium*: the Einstein–Podolsky–Rosen paradox: From concepts to applications", Rev. Mod. Phys. **81**, 1727–1751 (2009) (pages 58, 73).

[17] E. Schrödinger, "Die gegenwärtige Situation in der Quantenmechanik", Naturwissenschaften **23**, 844 (1935) (page 58).

[18] W. Marshall, C. Simon, R. Penrose and D. Bouwmeester, "Towards quantum superpositions of a mirror", Phys. Rev. Lett. **91**, 130401 (2003) (page 58).

[19] M. Born and H. Born, eds., *The Born–Einstein letters: friendship, politics, and physics in uncertain times: correspondence between Albert Einstein and Max and Hedwig Born from 1916 to 1955 with commentaries by Max Born* (pub-MACMILLAN, 2005) (page 58).

[20] E. Schrödinger, "Discussion of probability relations between separated systems", Proc. Cambridge Philos. Soc. **31**, 555–563 (1935); "Probability relations between separated systems", ibid. **32**, 446–452 (1936), page 58.

[21] H. M. Wiseman, S. J. Jones and A. C. Doherty, "Steering, entanglement, nonlocality, and the Einstein–Podolsky–Rosen paradox", Phys. Rev. Lett. **98**, 140402 (2007); S. J. Jones, H. M. Wiseman and A. C. Doherty, "Entanglement, Einstein–Podolsky–Rosen correlations, Bell nonlocality, and steering", Phys. Rev. A **76**, 052116 (2007), pages 58, 73.

[22] E. G. Cavalcanti, S. J. Jones, H. M. Wiseman and M. D. Reid, "Experimental criteria for steering and the Einstein–Podolsky–Rosen paradox", Phys. Rev. A **80**, 032112 (2009) (pages 58, 73).

[23] P. D. Drummond and C. W. Gardiner, "Generalised P-representations in quantum optics", Journal of Physics A: Mathematical and General **13**, 2353–2368 (1980) (pages 58, 63).

[24] M. K. Olsen, A. B. Melo, K. Dechoum and A. Z. Khoury, "Quantum phase-space analysis of the pendular cavity", Phys. Rev. A **70**, 043815 (2004) (pages 58, 61).

[25] L.-M. Duan, G. Giedke, J. I. Cirac and P. Zoller, "Inseparability criterion for continuous variable systems", Phys. Rev. Lett. **84**, 2722–2725 (2000) (pages 58, 71).

[26] E. Wigner, "On the quantum correction for thermodynamic equilibrium", Phys. Rev. **40**, 749–759 (1932); J. E. Moyal, "Quantum mechanics as a statistical theory", Proc. Cambridge Philos. Soc. **45**, 99–124 (1949); R. Graham and F. Haake, *Quantum statistics in optics and solid-state physics*, 1st ed., Vol. 66 (Springer, New York, 1973); P. Kinsler and P. D. Drummond, "Quantum dynamics of the parametric oscillator", Phys. Rev. A **43**, 6194–6208 (1991), pages 59, 65.

[27] P. D. Drummond and A. D. Hardman, "Simulation of quantum effects in Raman-active waveguides", EPL (Europhysics Letters) **21**, 279–284 (1993) (pages 59, 63).

[28] P. D. Drummond and M. D. Reid, "Correlations in nondegenerate parametric oscillation. II. below threshold results", Phys. Rev. A **41**, 3930–3949 (1990) (pages 59, 63).

[29] S. Chaturvedi, K. Dechoum and P. D. Drummond, "Limits to squeezing in the degenerate optical parametric oscillator", Phys. Rev. A **65**, 033805 (2002); K. Dechoum, P. D. Drummond, S. Chaturvedi and M. D. Reid, "Critical fluctuations and entanglement in the nondegenerate parametric oscillator", ibid. **70**, 053807 (2004), pages 59, 63.

[30] A. F. Pace, M. J. Collett and D. F. Walls, "Quantum limits in interferometric detection of gravitational radiation", Phys. Rev. A **47**, 3173–3189 (1993) (pages 59, 60, 83).

[31] M. Ludwig, B. Kubala and F. Marquardt, "The optomechanical instability in the quantum regime", New Journal of Physics **10**, 095013 (2008); "Enhanced quantum nonlinearities in a two-mode optomechanical system", Phys. Rev. Lett. **109**, 063601 (2012), page 59.

[32] A. Kronwald, M. Ludwig and F. Marquardt, "Full photon statistics of a light beam transmitted through an optomechanical system", Phys. Rev. A **87**, 013847 (2013) (page 59).

[33] P. Deuar and P. D. Drummond, "Correlations in a BEC collision: First-principles quantum dynamics with 150 000 atoms", Phys. Rev. Lett. **98**, 120402 (2007) (page 59).

[34] B. Opanchuk, L. Rosales-Zárate, M. Reid and P. Drummond, "Probabilistic simulation of mesoscopic "Schrödinger cat" states", Physics Letters A **378**, 946–949 (2014) (page 59).

[35] L. Qiu, L. Gan, W. Ding and Z.-Y. Li, "Single-photon generation by pulsed laser in optomechanical system via photon blockade effect", J. Opt. Soc. Am. B **30**, 1683–1687 (2013) (page 59).

[36] S. Lloyd, "Universal quantum simulators", Science **273**, 1073–1078 (1996) (page 59).

[37] J. Chan, A. H. Safavi-Naeini, J. T. Hill, S. Meenehan and O. Painter, "Optimized optomechanical crystal cavity with acoustic radiation shield", Applied Physics Letters **101**, 081115 (2012) (pages 59, 83).

[38] Q. Y. He and M. D. Reid, "Einstein–Podolsky–Rosen paradox and quantum steering in pulsed optomechanics", Phys. Rev. A **88**, 052121 (2013) (pages 59, 71, 80).

[39] V. Handchen, T. Eberle, S. Steinlechner, A. Samblowski, T. Franz, R. F. Werner and R. Schnabel, "Observation of one-way Einstein–Podolsky–Rosen steering", Nat Photon **6**, 596–599 (2012); S. L. W. Midgley, A. J. Ferris and M. K. Olsen, "Asymmetric Gaussian steering: When Alice and Bob disagree", Phys. Rev. A **81**, 022101 (2010); J. Bowles, T. Vértesi, M. T. Quintino and N. Brunner, "One-way Einstein–Podolsky–Rosen steering", arXiv: 1310.0668, page 59.

[40] V. B. Braginskii and A. B. Manukin, "Ponderomotive effects of electromagnetic radiation", Sov. Phys. JETP **25**, 653 (1967); A. Dorsel, J. D. McCullen, P. Meystre, E. Vignes and H. Walther, "Optical bistability and mirror confinement induced by radiation pressure", Phys. Rev. Lett. **51**, 1550–1553 (1983), page 60.

[41] H. J. Carmichael, *Statistical methods in quantum optics 1: Master equations and Fokker–Planck equations* (Springer, Berlin, 2002) (page 61).

[42]B. Yurke, "Squeezed-coherent-state generation via four-wave mixers and detection via homodyne detectors", Phys. Rev. A **32**, 300–310 (1985); C. W. Gardiner and M. J. Collett, "Input and output in damped quantum systems: Quantum stochastic differential equations and the master equation", ibid. **31**, 3761–3774 (1985), pages 62, 63, 67.

[43]A. Gilchrist, C. W. Gardiner and P. D. Drummond, "Positive P representation: Application and validity", Phys. Rev. A **55**, 3014–3032 (1997) (page 63).

[44]P. D. Drummond and M. Hillery, *The quantum theory of nonlinear optics* (Cambridge University Press, Cambridge, 2014) (page 64).

[45]S. Chaturvedi and P. D. Drummond, "Stochastic diagrams for critical point spectra", Eur. Phys. J. B **8**, 251–267 (1999) (page 65).

[46]P. Drummond and I. K. Mortimer, "Computer simulations of multiplicative stochastic differential equations", J. Comp. Phys. **93**, 144–170 (1991) (pages 65, 81).

[47]S. M. Tan, "Confirming entanglement in continuous variable quantum teleportation", Phys. Rev. A **60**, 2752–2758 (1999) (page 71).

[48]V. Giovannetti, S. Mancini, D. Vitali and P. Tombesi, "Characterizing the entanglement of bipartite quantum systems", Phys. Rev. A **67**, 022320 (2003) (page 71).

[49]M. D. Reid, "Monogamy inequalities for the Einstein–Podolsky–Rosen paradox and quantum steering", Phys. Rev. A **88**, 062108 (2013); W. P. Bowen, R. Schnabel, P. K. Lam and T. C. Ralph, "Experimental investigation of criteria for continuous variable entanglement", Phys. Rev. Lett. **90**, 043601 (2003), page 77.

Chapter 5

Einstein-Podolsky-Rosen quantum simulations in nonclassical phase-space

Authors: S. Kiesewetter[1], K. Dechoum[2], B. Opanchuk[1], L. Rosales-Zárate[1], M. D. Reid[1], P. D. Drummond[1]

Affiliations: [1]Centre for Quantum and Optical Science, Swinburne University of Technology, Melbourne, Australia., [2]Instituto de Física da Universidade Federal Fluminense, Boa Viagem, 24210–340, Niterói, Rio de Janeiro, Brazil.

Published in: Journal of the Optical Society of America B **32**, pp. A64-A73; accepted 17 March 2015

5.1 Abstract

We give a brief history of probabilistic quantum simulations of Einstein–Podolsky–Rosen paradoxes. This treats the early origins of the modern proposals using continuous variables, simulation methods using the positive-P representation, and current developments. Recent simulations treated include the behavior of parametric downconversion near the critical point, the simulation of parametric Bell violations, quantum entanglement, and correlations in optomechanics, as well as extensions to quantum field systems with planar interferometers.

5.2 Introduction

The original Einstein–Podolsky–Rosen (EPR) paper [1] used an example of an idealized quantum state to prove the incompatibility of local realism and the completeness of quantum theory. The original authors thought that local realism was the sounder principle. Following the work of Bell [2], most physicists now believe that the contradiction should be resolved in favor of the completeness of quantum mechanics, and against local realism. However, there is a more subtle problem with the EPR argument. Suppose the perfectly correlated idealized states used by EPR as examples never occur in nature. If true, a type of super-selection principle would be at work, acting as a cosmic censor to prevent a paradox from occurring.

Fortunately this is not the case. The Reid inequality [3, 4] signifies the EPR paradox in physically achievable states, thus allowing an experimental investigation. This inequality is the foundation of modern investigations into the EPR paradox. It was later extended to a family of inequalities, generically called the EPR-steering inequalities [5]. Both of these can be violated without requiring perfect correlations. As a consequence, we know that the strange paradoxes of quantum theory have practical applications.

The question still remains of how to predict whether a given physical system will generate such paradoxical quantum states. Most analytical approaches involve linearization, and are therefore approximate. As a result, one cannot easily tell if the analytic calculation is valid. Such problems become more serious when one investigates increasingly nonlinear systems that evidence large EPR-steering violations. The present paper will describe exact phase-space quantum simulation methods that can allow theories to span the linear and nonlinear regimes, thus overcoming this problem.

While traditional phase-space representations are widely known [6–10], these do not generally lead to probabilistic simulations. To simulate quantum dynamics probabilistically, an enlarged *nonclassical* phase-space distribution was developed by Chaturvedi *et al.* [11]. By nonclassical phase space, we mean a phase-space with twice the usual classical dimension. Intuitively, the physical meaning of this approach is that it allows for quantum superpositions of distinct coherent states, in addition to "classical-like" coherent states, thus increasing the phase-space dimension.

The existence and positivity of this distribution, called the positive-P representation, for all bosonic quantum states was proved not long after this [12].An important feature

of these methods is the existence of stochastic equations that describe quantum evolution for a wide range of dynamical problems.

The positive P-representation emerged just before the development of the idea of quantum computing. Feynman wrote a paper on the topic [13], claiming that probabilistic simulations could not be carried out. His argument was that this would be equivalent to a local hidden variable theory and, therefore, could not explain a Bell violation, which can occur in quantum mechanics. This was the motivation for developing quantum computers.

Despite Feynman's objection, quantum models of parametric down conversion were soon developed, and it was shown that these had useful stochastic equations [14]. In a dynamical calculation relevant to experiments, a positive-P stochastic theory was used to calculate EPR violations [15, 16] in parametric downconversion. Subsequent experimental verification of these predictions [17] have shown that EPR paradoxes are experimentally accessible. Such calculations have the merit that they can be readily checked numerically, on solving the corresponding stochastic equations. In this way, issues like nonlinearity can be investigated.

In summary, the situation is not black and white, since Feynman's argument did not take into account positive phase-space distributions. It is now known that Bell inequality violations can be readily simulated probabilistically [18, 19], as phase-space representations are not hidden variable theories.

Early calculations took place in few-mode regimes that are nearly linear [14–16, 20–22]. However, later work extended this to simulate critical fluctuations [23, 24], bistable tunneling [25, 26] and multimode problems [27, 28]. As a consequence, simulation techniques are potentially useful for calculating EPR correlations in nonlinear regimes. This is a promising route toward understanding how one can make the transition from a macroscopic EPR correlation to a full-scale Schrödinger cat.

In the rest of the paper, we review how these quantum simulation methods work, and their use in EPR calculations. We survey the equations that are obtained, and discuss their advantages and disadvantages. Finally, we show how these methods can be employed in quantum simulations of optomechanical entanglement, and the results that are obtained at high nonlinearity, as well as applications to quantum field simulations.

5.3 Quantum dynamical simulations

Because of the large size of Hilbert space in many physical experiments, methods using number states can be forbiddingly large. The widely used operator linearization approximation is not always valid.

Phase-space representations are a useful alternative. Wigner [6] is credited with developing the first such method in quantum mechanics, although his distribution was not a probability. Independently, Moyal [8] used Wigner's distribution to develop a dynamical theory equivalent to quantum mechanics.

The first *probabilistic* phase-space distribution was the Husimi Q-function [7]. The subsequent development of the laser led to a third commonly used method, the normally ordered P-representation [9, 10]. None of these, however, allow probabilistic quantum simulations for general unitary evolution.

5.3.1 Positive P-Representation

Probabilistic quantum simulations require an extension of the phase-space approach to enlarged, nonclassical phase-spaces. This can provide both approximate and numerically exact solutions. In this section, we review some of the main results with this method.

The positive P-representation for a bosonic M-mode system is an expansion of the density matrix of the form

$$\hat{\rho} = \int P(\boldsymbol{\alpha}, \boldsymbol{\alpha}^{+}) \frac{|\boldsymbol{\alpha}\rangle \left\langle (\boldsymbol{\alpha}^{+})^{*} \right|}{\left\langle (\boldsymbol{\alpha}^{+})^{*} \middle| \boldsymbol{\alpha} \right\rangle} d^{2M}\boldsymbol{\alpha} d^{2M}\boldsymbol{\alpha}^{+} \,. \tag{5.1}$$

Here, boldface symbols are M-vectors, so $\boldsymbol{\alpha} = (\alpha_1, \ldots \alpha_M)$, and $\boldsymbol{\alpha}^{+}$ indicates an independent complex vector which is conjugate to $\boldsymbol{\alpha}$ in the mean. We denote this as $\boldsymbol{\alpha}^{+}$, to indicate that it plays the role of a stochastic conjugate. This representation always exists as a positive representation [12], for any quantum density matrix. The states $|\boldsymbol{\alpha}\rangle$ are multimode Glauber coherent states, which are simultaneous eigenstates of the annihilation operator:

$$\hat{a}_n |\boldsymbol{\alpha}\rangle = \alpha_n |\boldsymbol{\alpha}\rangle \,. \tag{5.2}$$

It is worth noting that this method has off-diagonal elements that would be generated if the density matrix included a pure state of form $\left[|\boldsymbol{\alpha}\rangle + \left|(\boldsymbol{\alpha}^{+})^{*}\right\rangle\right]$. Given this inclusion

of off-diagonal elements, one can prove that there is a positive P-distribution [12] for any quantum density matrix, hence the name of the method. As a result of this eigenvalue equation, one can prove that all operator moments are readily obtained through probabilistic sampling:

$$\begin{aligned}\langle \hat{a}_m^\dagger \cdots \hat{a}_n \rangle &= Tr\left[\hat{\rho}\hat{a}_m^\dagger \cdots \hat{a}_n\right] \\ &= \int P(\boldsymbol{\alpha},\boldsymbol{\alpha}^+)[\alpha_m^+ \cdots \alpha_n] d^{2M}\boldsymbol{\alpha} d^{2M}\boldsymbol{\alpha}^+ \,. \end{aligned} \tag{5.3}$$

These moment equations look very similar to hidden variable theory models introduced by Bell [2]. The important difference is that the quantities being averaged are not the actual observables as required by Bell's theorem, even though one can extract any required correlations from these probabilistic averages.

5.3.2 Dynamical Equations

The operator identities given above can be extended to include differential identities. These map any ordering of annihilation and creation operators into a differential equation for P, after integration by parts. Such mappings require that the P-distribution tails are sufficiently bounded [29], and sometimes this necessitates additional regularization [30] to ensure boundedness.

The differential mappings have the form

$$\begin{aligned}\hat{a}_n^\dagger \hat{\rho} &\rightarrow \left[\alpha_n^+ - \frac{\partial}{\partial \alpha_n}\right] P \\ \hat{a}_n \hat{\rho} &\rightarrow \alpha_n P \\ \hat{\rho}\hat{a}_n &\rightarrow \left[\alpha_n - \frac{\partial}{\partial \alpha_n^+}\right] P \\ \hat{\rho}\hat{a}_n^\dagger &\rightarrow \alpha_n^+ P \,. \end{aligned} \tag{5.4}$$

Using these identities, one can transform an arbitrary master equation into a differential equation, according to the equivalence

$$\frac{\partial \hat{\rho}}{\partial t} = \hat{\mathcal{L}}\left[\hat{\rho}\right] \rightarrow \frac{\partial P}{\partial t} = \mathcal{L}\left[P\right] \,, \tag{5.5}$$

where $\mathcal{L}$ is the differential operator obtained from $\hat{\mathcal{L}}$ using the above identities. By Cauchy's theorem, there is a choice of equivalent derivative terms, since the represent-

ation kernel is analytic in $\boldsymbol{\alpha}, \boldsymbol{\alpha}^+$. This allows a choice that ensures a positive definite diffusion, and hence a distribution that remains real and positive. The resulting positive-definite Fokker–Planck equation is mapped into an equivalent stochastic equation according to procedures that are now reasonably standard [31].

In the case of quadratic Hamiltonians, this procedure works equally well for the Wigner representation, provided that the initial distribution is positive. However, stochastic methods for nonlinear Hamiltonians in the Wigner representation require methods involving higher order noise [32] or truncation [33]. This is necessary because the Wigner time evolution equation results in third-order derivatives, with no stochastic equivalent. The truncation or removal of these terms is asymptotically valid as the mode occupation $n \to \infty$. Here we will use the positive P-representation. For many physical problems, either method will give the same result.

The fundamental advantage of probabilistic simulations is that one can treat realistic models without using uncontrolled approximations such as linearization, and these models can be readily scaled to large sizes. Clearly, one still has to deal with sampling error. However, such errors can be calculated, and if not too large, can be reduced simply by taking more samples. In lossless nonlinear dynamics the sampling error typically grows exponentially with time, which creates a maximum time window for practical simulations.

A related issue is that dynamically generated distributions can have power-law tails. These makes the distributions harder to sample and can in the worst case generate boundary terms in phase-space, causing systematic errors. The effect occurs for strong couplings over long times, and can be tested by examining the distribution tails. One method to solve this problem is to use stochastic gauges [30]. Improved sampling involving parallel optimization is another technique being investigated.

5.4 EPR violations in Nondegenerate down-conversion

An early example of these methods was a realistic model [3, 15, 34] that led to the first practical continuous variable EPR experiment [17] using nondegenerate parametric down-conversion.

5.4.1 System Hamiltonian

Here the system is an idealized interferometer, resonant at three frequencies, $\omega_1, \omega_2, \omega_0$, with corresponding mode operators $\hat{a}_j$. The second harmonic pump mode frequency ω_0 is approximately in two-photon resonance with the other two modes, i.e., $\omega_0 \approx \omega_1 + \omega_2$. The interferometer has an external laser input with amplitude $\mathcal{E}$ at $\omega_0^I = 2\omega_L$, where ω_L is the primary pump frequency, which defines a reference frequency. The two nearly resonant lower frequency modes are called the signal and idler modes in traditional nonlinear optical terminology, although they are interchangeable.

Downconversion of the pump photons occurs because of an $\chi^{(2)}$ nonlinearity, which is usually a parametric nonlinear crystal or waveguide. All modes are damped because of cavity losses, with an amplitude damping of γ_i, and a reservoir operator of $\hat{\Gamma}_j$.

The Hamiltonian that describes this open system [20], including damping, is

$$\hat{H} = \hat{H}_0 + \hat{H}_{\text{int}} + \sum_{j=0,2} \hbar \left(\hat{a}_j \hat{\Gamma}_j^\dagger + \hat{a}_j^\dagger \hat{\Gamma}_j \right) + \hat{H}_R \,. \tag{5.6}$$

Here $\hat{H}_R$ is the free-field reservoir Hamiltonian, and $\hat{H}_0$ defines an interaction picture with

$$\hat{H}_0 = \hbar \sum_{j=0,2} \omega_j^I \hat{a}_j^\dagger \hat{a}_j \,, \tag{5.7}$$

so that operators defined in this picture evolve according to $\hat{a}_j^I = \hat{a}_j \exp\left(-i\omega_j^I t\right)$, where $\hat{a}_j$ are the time-independent Schrödinger picture operators. This method uses reference frequencies ω_i^I determined by the external clock, where $\omega_1^I + \omega_2^I = 2\omega_L$. The relative detuning between the pump laser at $2\omega_L$ and the intracavity pumped mode ω_0 is $\delta_0 = \omega_0 - 2\omega_L$. The downconverted modes have relative detunings $\delta_i = \omega_i - \omega_i^I$. As a result, all explicit time dependence is cancelled in the interaction Hamiltonian.

In the case of resonant down-conversion, with just a second harmonic, this interaction Hamiltonian is then

$$\hat{H}_{\text{int}}/\hbar = i\left[\mathcal{E}\hat{a}_0 + \chi \hat{a}_0 \hat{a}_1^\dagger \hat{a}_2^\dagger - \text{h.c.}\right] + \sum_{j=0,2} \delta_j \hat{a}_j^\dagger \hat{a}_j \,. \tag{5.8}$$

Using standard techniques [35], we obtain the following master equation for the reduced system density operator in the interaction picture:

$$\frac{\partial \hat{\rho}}{\partial t} = \hat{\mathcal{L}}\left[\hat{\rho}\right] = \frac{1}{i\hbar}\left[\hat{H}_{\text{int}}, \hat{\rho}\right] + \sum_j \hat{\mathcal{L}}_j\left[\hat{\rho}\right] , \tag{5.9}$$

where the dissipative part of the Liouville operator is $\hat{\mathcal{L}}_i\left[\hat{\rho}\right] = \gamma_j\left(2\hat{a}_j\hat{\rho}\hat{a}_j^\dagger - \hat{a}_j^\dagger\hat{a}_j\hat{\rho} - \hat{\rho}\hat{a}_j^\dagger\hat{a}_j\right)$, in the limit of a broadband, weakly coupled reservoir, at zero temperature.

The input-output relations that describe the external modes outside the cavity are of the form $\hat{a}_i^{\text{out}} = \sqrt{2\gamma_i^o}\hat{a}_i - \hat{a}_i^{\text{in}}$, with $\hat{a}_i^{\text{in}}$ and $\hat{a}_i^{\text{out}}$ describing input and output fields with input correlations

$$\begin{aligned}\left\langle \hat{a}_i^{\text{in}}(t)\hat{a}_j^{\text{in}\dagger}(t')\right\rangle &= \left(\bar{n}_i^{\text{th}} + 1\right)\delta_{ij}\delta\left(t - t'\right)\\ \left\langle \hat{a}_i^{\text{in}\dagger}(t)\hat{a}_j^{\text{in}}(t')\right\rangle &= \bar{n}_i^{\text{th}}\delta_{ij}\delta\left(t - t'\right) .\end{aligned} \tag{5.10}$$

We will assume in most examples in this paper that the reservoirs are in the vacuum state, $\bar{n}_i^{\text{th}} = 0$. However, non-zero reservoir temperatures can be readily included.

5.4.2 Stochastic Equations

Given vanishing boundary terms, which is typically valid provided that $\chi \ll \gamma_j$, the following stochastic equations are obtained, for any driving field $\mathcal{E}$:

$$\begin{aligned}\frac{d\alpha_0}{dt} &= \mathcal{E} - \tilde{\gamma}_0\alpha_0 - \chi\alpha_1\alpha_2 ,\\ \frac{d\alpha_1}{dt} &= -\tilde{\gamma}_1\alpha_1 + \chi\alpha_2^+\alpha_0 + \sqrt{\chi\alpha_0}\zeta_1(t) ,\\ \frac{d\alpha_2}{dt} &= -\tilde{\gamma}_2\alpha_2 + \chi\alpha_1^+\alpha_0 + \sqrt{\chi\alpha_0}\zeta_2(t) ,\end{aligned} \tag{5.11}$$

where we define $\tilde{\gamma}_j = \gamma_j + i\delta_j$ to obtain a complex damping, and assume the coupling χ is real.

There are also three stochastic conjugate equations, obtained on conjugating any complex coefficients, and replacing $\alpha_j \to \alpha_j^+$, $\zeta_j \to \zeta_j^+$. Here $\zeta_k(t)$, $\zeta_k^+(t)$ represent two independent complex Gaussian stochastic processes. The amplitude of the stochastic fluctuations that act on the signal mode are dependent on the pump field dynamics through the term $\sqrt{\chi\alpha_0}$. The noises have zero average so that $\left\langle \zeta_j(t)\right\rangle = 0$, and their

nonvanishing stochastic correlations are given by

$$\left\langle \zeta_1(t)\zeta_2(t')\right\rangle = \left\langle \zeta_1^+(t)\zeta_2^+(t')\right\rangle \quad = \quad \delta\left(t-t'\right) \ . \tag{5.12}$$

In the classical or mean field approximation, one obtains the well-known classical equations of intracavity parametric oscillation,

$$\begin{aligned} \frac{d\alpha_0}{dt} &= \mathcal{E} - \tilde{\gamma}_0\bar{\alpha}_0 - \chi\bar{\alpha}_1\bar{\alpha}_2 \, , \\ \frac{d\bar{\alpha}_1}{dt} &= -\tilde{\gamma}_1\bar{\alpha}_1 + \chi\bar{\alpha}_1^*\bar{\alpha}_0 \, , \\ \frac{d\bar{\alpha}_2}{dt} &= -\tilde{\gamma}_2\bar{\alpha}_2 + \chi\bar{\alpha}_2^*\bar{\alpha}_0 \, , \end{aligned} \tag{5.13}$$

where $\bar{\alpha}_1 \equiv \langle\hat{a}\rangle = \langle\alpha\rangle$, and these equations assume that mean products of operators or stochastic variables factorize into products of mean values. This analysis predicts there is a critical point $\mathcal{E}_T$ where there is a nonequilibrium phase transition, where $\mathcal{E}_T = \gamma_0\sqrt{\gamma_1\gamma_2}/\chi$ in the resonant case. We define $P = \mathcal{E}/\mathcal{E}_T$ to give a dimensionless measure of the driving field.

5.4.3 Output Quadratures

Calculations can be carried out for the detected output quadratures, as measured in an experiment with twin local oscillators at spatially separated locations. Causal separation is a requirement for the EPR paradox, just as for a Bell violation, although this does not appear to be implemented in experiments as yet. The original stochastic equations, Eq. (5.11), of course include the full quantum correlations. The output fields can be calculated from Fourier transforming the stochastic variables, and we introduce

$$\begin{aligned} \left\langle \alpha_i\left(\omega'\right)\alpha_j\left(\omega\right)\right\rangle &= \frac{c_{ij}\left(\omega\right)}{2\sqrt{\gamma_i\gamma_j}}\delta\left(\omega+\omega'\right) \ , \\ \left\langle \alpha_i^+\left(\omega'\right)\alpha_j\left(\omega\right)\right\rangle &= \frac{s_{ij}\left(\omega\right)}{2\sqrt{\gamma_i\gamma_j}}\delta\left(\omega+\omega'\right) \ , \end{aligned} \tag{5.14}$$

where the the non-zero spectral elements are [16]

$$s_{11}(\omega) = s_{22}(-\omega) = \frac{4\gamma_1\gamma_2\chi^2|\bar{\alpha}_0|^2}{|(i\omega-\tilde{\gamma}_2^*)(i\omega-\tilde{\gamma}_1)-\chi^2|\bar{\alpha}_0|^2|^2},$$

$$\begin{aligned} c_{21}(\omega) &= c_{12}(\omega) \\ &= \frac{2\sqrt{\gamma_1\gamma_2}\chi\bar{\alpha}_0\left(\chi^2|\bar{\alpha}_0|^2+(i\omega-\tilde{\gamma}_2^*)(-i\omega-\tilde{\gamma}_1^*)\right)}{|(i\omega-\tilde{\gamma}_2^*)(i\omega-\tilde{\gamma}_1)-\chi^2|\bar{\alpha}_0|^2|^2}. \end{aligned} \tag{5.15}$$

Here $\bar{\alpha}_0 = \mathcal{E}/\tilde{\gamma}_0$ is the steady state semiclassical result for the pumping mode.

These equations can be treated approximately using linearization, as in the original predictions of EPR correlations. In the symmetric resonant case with $\gamma_1 = \gamma_2 = \gamma$ and zero detunings, we introduce $P_\pm = P \pm 1$, $\nu = \omega/\gamma$ and obtain

$$\begin{aligned} s_{11}(\omega) &= s_{22}(\omega) = P\left[\frac{1}{P_-^2+\nu^2} - \frac{1}{P_+^2+\nu^2}\right], \\ c_{12}(\omega) &= c_{21}(\omega) = P\left[\frac{1}{P_-^2+\nu^2} + \frac{1}{P_+^2+\nu^2}\right]. \end{aligned} \tag{5.16}$$

These quantities correspond to normally ordered correlations of measured output quadrature variables, $\hat{x}_j^\theta(\omega)$, whose real and imaginary parts are experimentally observable, and have a close similarity to Einstein's $\hat{x}$ and $\hat{p}$ variables. One can calculate that a measurement of $\hat{x}_2^\theta$ allows one to infer $\hat{x}_1^{-\theta}$. The optimum inferences variance is

$$\Delta^2(\theta,\omega) = 1 + s_{11}(\omega) + s_{11}(-\omega)\frac{c_{12}(\omega)+c_{21}(\omega)}{1+s_{22}(\omega)+s_{22}(-\omega)}. \tag{5.17}$$

An EPR paradox is found when an inferred Heisenberg uncertainty principle is violated:

$$E_{1|2} = \Delta(0,\omega)\,\Delta\left(\frac{\pi}{2},\omega\right) < 1. \tag{5.18}$$

Here $E_{1|2}$ is the EPR inferred Heisenberg product, when uncertainties at location (1) are inferred from measurements at location (2). All such correlations and fluctuations are greatly enhanced at the critical point where $P \to 1$. This is close to the the point of maximum EPR correlations, and is also the point where linearization breaks down. To treat this regime, a perturbation method involving stochastic diagrams allows higher-order effects to be included [36], which gives the true maximum correlation. Alternatively,

one can simply treat the entire system numerically, to give solutions over the full range of driving fields and detunings [24].

5.5 Recent simulations

Recent work on probabilistic quantum simulations has involved theoretical extensions of the positive-P or related methods to treat new physical systems. This included dynamical and static or equilibrium problems, large-scale simulations with improved computer hardware, and experimental tests. A Gaussian phase-space representation for bosons and fermions was also introduced recently [37, 38]. This gives a unified approach valid for all particle types.

In this section, we review two of these more recent EPR type simulations, including Bell violations, which are even stronger indications of quantum nonlocality, and EPR violations predicted in optomechanics experiments. We also present some new results for highly nonlinear optomechanical systems.

5.5.1 The Reid-Walls Bell Experiment

The Bell theorem shows the existence of even stronger quantum paradoxes than EPR-steering. These rule out all hidden variable theories. Feynman used Bell's theorem to claim in an early paper [13] that probabilistic simulation, while desirable to reduce exponential complexity, cannot be used to simulate quantum mechanics [2]. He argued that a probabilistic simulation would involve a hidden variable theory, which cannot violate a Bell inequality, while quantum mechanics allows this.

However, as explained in the Introduction, phase-space observables like Eq. (5.3) do not have the structure of a hidden variable theory observable, and therefore do not fall into this category. They can indeed violate a Bell inequality probabilistically. Although this was clear from early analyses [39], a computer simulation was only carried out recently.

These Bell violating simulations can be achieved most simply through modeling a Bell experiment using parametric downconversion and following an original proposal of Reid and Walls [40]. Their parametric downconversion idea is the basis of almost all modern Bell violation experiments.

Their original method involved four downconverted modes, with an undepleted pump such that $\langle \hat{a}_0 \rangle = E$, so that

$$\hat{H} = i\hbar\chi E \left(\hat{a}_1^\dagger \hat{a}_2^\dagger + \hat{a}_3^\dagger \hat{b}_4^\dagger - \text{h.c.} \right) . \tag{5.19}$$

The corresponding $+P$ stochastic equations, after applying the identities of Eq. (5.4) are:

$$\begin{aligned} \frac{d\alpha_1}{dt} &= \chi E \alpha_2^+ + \sqrt{\chi E}\zeta_1(t) , \\ \frac{d\alpha_2}{dt} &= \chi E \alpha_1^+ + \sqrt{\chi E}\zeta_2(t) , \\ \frac{d\alpha_3}{dt} &= \chi E \alpha_4^+ + \sqrt{\chi E}\zeta_3(t) , \\ \frac{d\alpha_4}{dt} &= \chi E \alpha_3^+ + \sqrt{\chi E}\zeta_4(t) . \end{aligned} \tag{5.20}$$

In addition, the stochastic correlations are

$$\langle \zeta_1(t)\zeta_2(t') \rangle = \langle \zeta_3(t)\zeta_4(t') \rangle = \delta(t - t') , \tag{5.21}$$

together with the usual Hermitian conjugate stochastic terms.

5.5.2 Bell Inequality Violations

The parametric interaction creates two pairs of entangled outputs. These are then combined to give an output quantum state which includes, *inter alia*, the famous singlet of Bell, Clauser, Aspect, and others. When calculating Bell violating moments, we find that a numerical, probabilistic simulation of these equations gives a clear violation of the Clauser–Horne type Bell inequality, which occurs when $S_{\text{CH}} > 1$. Here S_{CH} is defined as:

$$S_{\text{CH}} = \frac{\tilde{P}(\theta,\phi) - \tilde{P}(\theta,\phi') + \tilde{P}(\theta',\phi) + \tilde{P}(\theta',\phi')}{P^A(\theta') + P^B(\phi)}, \tag{5.22}$$

where $\tilde{P}(\theta,\phi)$ is the joint probability for detecting one up photon at location A with setting θ and one up photon at B with setting ϕ. The other joint probabilities are defined similarly. $P^A(\theta')$ and $P^B(\phi)$ are marginal probabilities for detecting an up photon in location A with setting θ' and an up photon in location B with setting ϕ respectively.

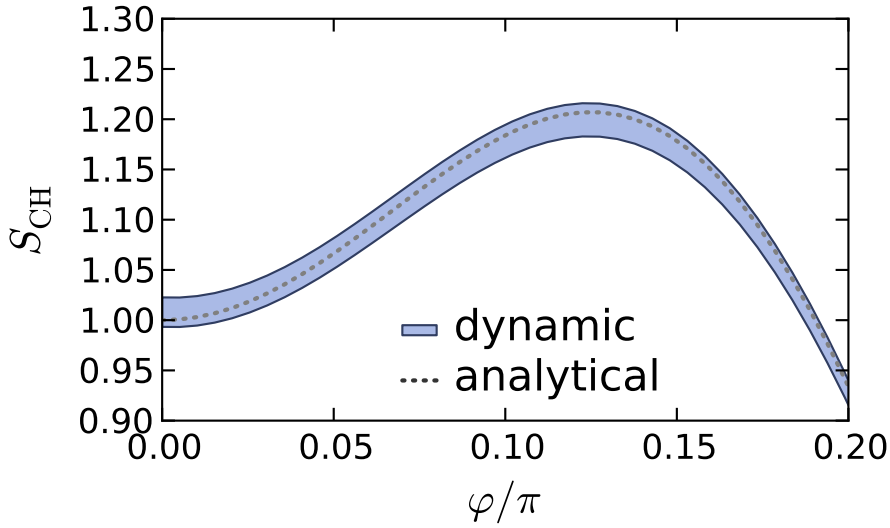

Figure 5.1: (Color online) Dependence of the simulated Clauser–Horne Bell type inequality for the entangled state generated in parametric downconversion process. Plotted is the normalized Bell violation ratio S_{CH}, as a function of the relative polarizer angle φ. The blue filled curve represents the sampling error around the mean, and the dotted line inside this region is the exact value. This result required 2.5×10^5 samples [19].

Choosing a time duration of $\tau = \chi E t = 0.1$, we can obtain predictions for angular correlations with respect to a polarizer angle $\varphi = \phi - \theta = \phi' - \theta' = \theta' - \phi$, and compare the results with an exact analytic solution. This comparison is shown in Fig. 5.1 [19].

Since every result with $S_{\mathrm{CH}} > 1$ violates a Bell inequality, this probabilistic simulation provides a counterexample to Feynman's claim [13]. We note that these simulations give an increasing sampling error as time evolves, because of the increasing variance of the generated distribution. This problem is easily solved if more samples are used. Results in this section typically require 2.5×10^5 samples for useful sampling errors of $\pm 2\%$. Naturally, one can also extend this analysis to treat much more realistic models, and other types of physical system that violate Bell inequalities.

To illustrate quantum simulation of multipartite Bell violations, quantum phase-space simulations of measurements on GHZ states in ion traps were carried out [41]. Although these were not dynamical, the scaling properties showed that there is a classical speedup in this simulation. Measurements can be simulated for multiple noncommuting observables. Thus, a simulation can be exponentially *faster* than an experiment. This speedup is the opposite of what is expected.

5.5.3 Nonlinear EPR-Steering in Optomechanics

At the same time, new experimental systems have emerged, including optomechanical systems. The first optomechanical simulations were of quantum effects in gravity wave detectors [42]. This simulation, however, produced the result that quantum mechanical effects were negligibly small.

The Hamiltonian for optomechanics includes the mechanical oscillator mode energy at frequency ω_m, a pump input at ω_l, and the optical mode frequency $\omega_o = \omega_l + \delta$. As in the previous sections, we transform to an interaction picture in which the time evolution of the cavity operators at the laser frequency is removed. This subtracts the term $\omega_l \hat{a}^\dagger \hat{a}$, together with corresponding time dependences, leaving the standard optomechanical Hamiltonian:

$$\begin{aligned} \hat{H}/\hbar &= \delta \hat{a}^\dagger \hat{a} + \omega_m \hat{b}^\dagger \hat{b} + \chi \hat{a}^\dagger \hat{a}(\hat{b} + \hat{b}^\dagger) \\ &\quad + iE(t)(\hat{a}^\dagger - \hat{a}) + \hat{H}_r \,. \end{aligned} \tag{5.23}$$

The first terms give the energy of the optical cavity field, and the mechanical oscillator, while the third term is the optomechanical interaction, where χ is the coupling resulting from radiation pressure. The fourth term is the coupling to the coherent input $E(t)$, and $\hat{H}_r$ describes the coupling to dissipative reservoirs, which are simple losses to thermal reservoirs at a rate γ_o for the optical, and γ_m for the mechanical mode. The difference between this and the results described above is that the low-frequency mechanical reservoir is at finite temperature.

After deriving the Fokker–Planck equation [43], with the positive-P representation so that $\hat{a} \to \alpha$ and $\hat{b} \to \beta$, one obtains an equivalent set of complex Itō stochastic equations, together with the usual stochastic conjugate equations:

$$\begin{aligned} \frac{d\alpha}{dt} &= E(t) - [i\chi(\beta + \beta^+) + \tilde{\gamma}_o]\,\alpha + \zeta_\alpha(t)\,, \\ \frac{d\beta}{dt} &= -\tilde{\gamma}_m \beta - i\chi\alpha\alpha^+ + \zeta_\beta(t)\,, \end{aligned} \tag{5.24}$$

Here $\tilde{\gamma}_o = \gamma_o + i\delta$, $\tilde{\gamma}_m = \gamma_m + i\omega_m$ are the complex dampings. The Gaussian noises $\zeta_{\alpha,\beta}$ are because of both internal nonlinearities and thermal noise inputs, where

$$\begin{aligned} \zeta_\alpha(t) &= \sqrt{2\gamma_o}\alpha^{\text{in}} + \zeta_\alpha^\chi(t)\,, \\ \zeta_\beta(t) &= \sqrt{2\gamma_m}\beta^{\text{in}} + \zeta_\beta^\chi(t)\,. \end{aligned} \tag{5.25}$$

Here $\zeta_\alpha^\chi(t)$ and $\zeta_B^\chi(t)$ are quantum noises due to interactions with correlations $\langle\zeta_i^\chi(t)\zeta_j^\chi(t')\rangle_S = -i\delta_{i,3-j}\chi\alpha\delta(t-t')$. The nonvanishing stochastic thermal correlations of the reservoir inputs are

$$\langle\alpha_k^{\text{in}}(t)\,\alpha_l^{\text{in}+}(t')\rangle_S = \bar{n}_k^{\text{th}}\delta_{kl}\delta(t-t')\,, \tag{5.26}$$

where $k,l = 1,2 \sim \alpha,\beta$, and $\bar{n}_k^{\text{th}}$ are the mean heat bath occupations numbers. We distinguish between internal and external noise sources in the calculations, by setting $\sqrt{2\gamma_o}\alpha^{\text{in}} = \sqrt{2\gamma_{\text{int}}}\alpha_{\text{int}}^{\text{in}} + \sqrt{2\gamma_{\text{ext}}}\alpha_{ext}^{\text{in}}$. Here $\gamma_o = \gamma_{\text{int}} + \gamma_{\text{ext}}$ is the total amplitude loss rate, composed of internal losses γ_{int} and external losses γ_{ext}. Only the external reservoirs take part in information transfer, since the information lost to internal reservoirs of the cavity is not monitored. Entanglement is caused by the parametric gain because of radiation pressure coupling between optical and mechanical oscillator modes, where $G = \chi^2 E(t)^2/\gamma_o(\delta^2+\gamma_o^2)$ is the gain resulting from parametric coupling.

Since there is a recent experiment on electromechanical entanglement [44], these simulations could be used to make a quantitative comparison between theory and experiment [45]. The results are shown in Fig. 5.2. Note that here we use the experimental pump and probe intensities, together with their notation of $\Gamma_b = 2G$, and a fixed pulse duration of $\tau = 35.5\,\mu$s.

Agreement between theory and experiment is excellent for these entanglement measures. However, due to low efficiency detection of $\sim$ 10%, no EPR paradox could be obtained from the inferred data of this experiment.

5.5.4 Nonlinear Corrections

In optomechanical experiments, expected detection efficiencies are higher that in electromechanical experiments, making these more promising for EPR-steering experiments

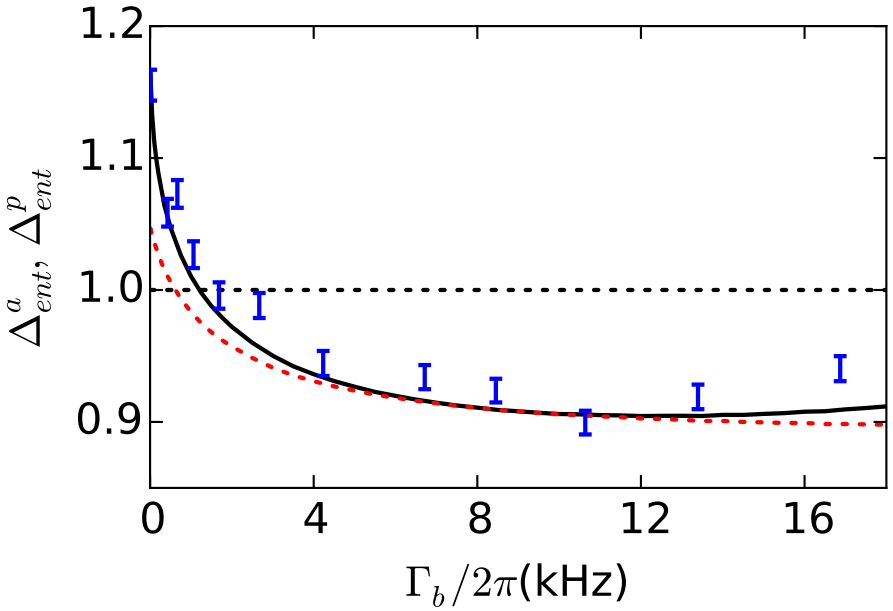

Figure 5.2: Exact positive-P quantum simulations of entanglement in pulsed electromechanics quantum simulations, compared to experimental data in [45]; the black line gives the additive entanglement criterion used in experiments, the red dotted line indicates a more reliable product entanglement criterion method. The maximum sampling error was 7×10^{-4}, which is too small to plot. Blue errors bars correspond to experimental data of Ref. [44].

with massive, macroscopic objects. In this section, we report results that corresponds to an Si optomechanical crystal experiment with parameters close to existing experimental values, except that a very strong coupling regime is utilized to explore the limits of the usual approximations. Here, we have $\omega_m/2\pi = 3.7\,\text{GHz}$, $Q_m = \omega_m/\gamma_m = 10^5$, $\gamma_m/2\pi = 37\,\text{kHz}$, $\gamma_0/2\pi = 0.26\,\text{GHz}$. The mechanical heat bath temperature is $T = 200\,\text{mK}$, corresponding to $\bar{n}^{\text{th}}_m = 0.7$. The pump pulse duration scales linearly with $R = Gt$. Here G is the gain, and in these plots $R = 1$ corresponds to $\tau = 0.04\,\mu\text{s}$, so in plots below, the pulse duration ranges from $\tau = 0 \ldots 0.02\,\mu\text{s}$.

Here the mechanical mode is compared to integrated optical mode directly; that is, no readout pulse is applied. The steering plots were obtained by minimizing both gain and relative phase simultaneously, where a detector efficiency of $\eta = 0.9$ is assumed. In these comparisons, both positive-P (exact) and Wigner (approximate) results are plotted, together with a linearized theory as outlined in the original publications [44, 46]. The results were obtained using 80,000 independent trajectories for the Wigner and linearized simulations, and 7,680,000 trajectories for the exact positive-P simulations.

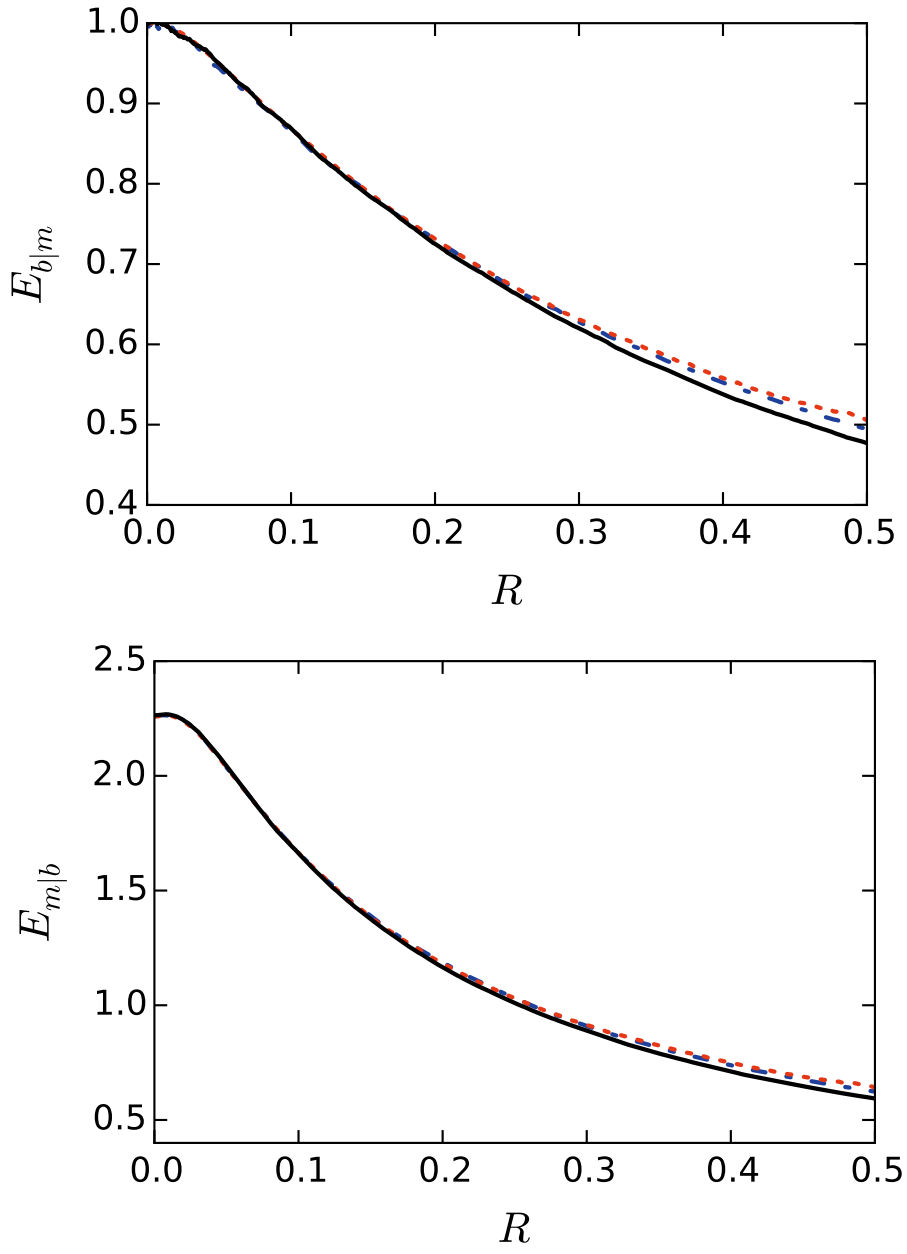

Figure 5.3: Quantum simulations of pulse oscillator entanglement in optomechanical quantum simulations, with $\chi_0/\gamma_0 = 0.35$ [45]. The solid black line indicates linearized simulation results; the dash-dotted blue line indicates Wigner simulation results; and the dotted red line indicates positive-P simulation results.

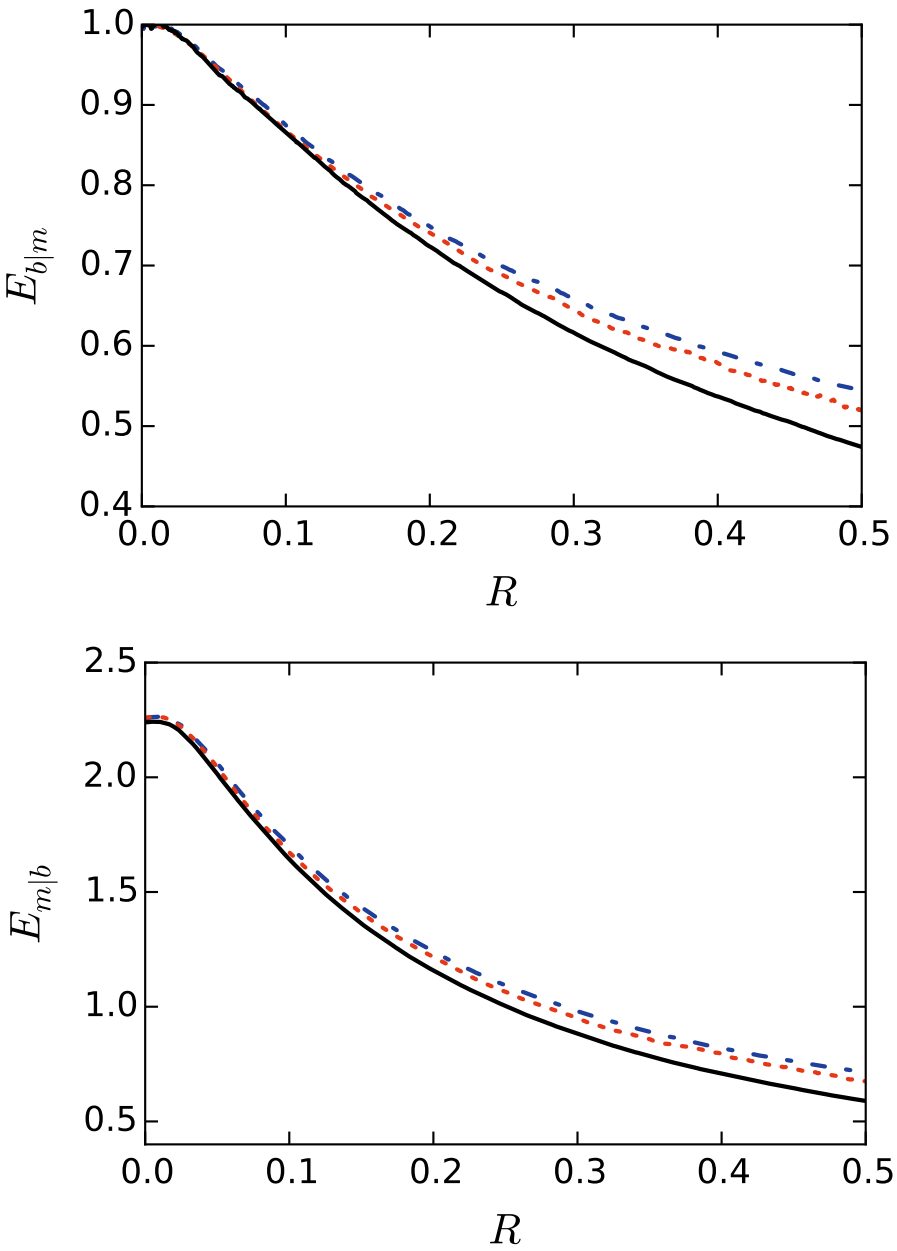

Figure 5.4: Quantum simulations of pulse oscillator entanglement in optomechanical quantum simulations, with $\chi_0/\gamma_0 = 0.7$ [45]. The solid black line indicates linearized simulation results; the dash-dotted blue line indicates Wigner simulation results; and the dotted red line indicates positive-P simulation results. The maximum sampling errors in Figs. (3) and (4) were 9×10^{-3} for the linearized results, 7×10^{-3} for the Wigner results, and 3×10^{-3} for the positive-P results.

The main conclusion we draw from this is that linearization is an excellent approximation in such short pulse experiments with current parameter values, where $\chi_0/\gamma_0 \ll 1$. However, as one increases χ/γ_o closer to unity, as in some recent proposals [47], the discrepancy between exact and linearized approaches becomes increasingly large. Interestingly, the truncated Wigner method remains reliable even at these strong couplings.

5.6 Quantum field simulations

Positive distributions provide an ideal avenue for calculations with large numbers of modes. Since the calculations are sampled over the phase-space, the dimension only grows linearly with the number of modes, rather than exponentially as in orthogonal number state methods. Dynamical simulations of quantum fields are feasible [48], and the initial work in this field used an analytic treatment of stochastic equations to calculate quantum soliton correlations. Subsequently, [49] performed the first experimental test, qualitatively confirming these one-dimensional bosonic quantum soliton squeezing predictions.

Computer simulations using probabilistic methods with stochastic equations were implemented later [50]. Further developments included treatments of Raman-induced decoherence [51], in quantitative agreement with experiments on fiber polarization squeezing.

A system with potential for EPR correlations and which also involves quantum field techniques, is the planar, nondegenerate quantum parametric oscillator. In this case, one simply uses stochastic quantum field simulations to extend the nondegenerate parametric system treated above to the case of a planar, multimode cavity. We calculate the stochastic equations and their predictions here to demonstrate how the method works.

The quantum Hamiltonian in the interaction picture has four main terms that can be summarized by the following expression [52–54]:

$$\hat{H} = \hat{H}_{\text{free}} + \hat{H}_{\text{int}} + \hat{H}_{\text{pump}} + \hat{H}_{\text{res}}\,, \tag{5.27}$$

where there are three fundamental Bose fields, $\hat{A}_i(\boldsymbol{x}, t)$ for $i = 0, 1, 2$. Typically, $\hat{A}_1$, $\hat{A}_2$ represent two orthogonal, polarization modes, while $\hat{A}_0$ is the double frequency pump. The boson fields obey the following equal time commutation relation, where $\boldsymbol{x}$ represents

the location in the plane:

$$\left[\hat{A}_i(\boldsymbol{x},t),\hat{A}_j^\dagger(\boldsymbol{x}',t)\right] = \delta_{ij}\delta\left(\boldsymbol{x}-\boldsymbol{x}'\right) . \tag{5.28}$$

The free evolution Hamiltonian modeling diffraction inside the cavity is

$$\hat{H}_{\text{free}} = \sum_{i=0}^{2} \hbar \int d^2\boldsymbol{x}\, \hat{A}_i^\dagger \left[\omega_i - \frac{v_i^2}{2\omega_i}\nabla^2\right] \hat{A}_i \,. \tag{5.29}$$

As previously, this Hamiltonian describes a planar cavity with intracavity resonant frequencies ω_i. In addition, there are group velocities v_i for the three field envelopes. The two-dimensional Laplacian, $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$. A one-dimensional system can also be treated.

The interaction Hamiltonian inside a crystal with $\chi^{(2)}$ nonlinearity is now given by a spatial integral [55]:

$$\hat{H}_{\text{int}} = i\hbar \int d^2\boldsymbol{x} \left[\chi \hat{A}_0 \hat{A}_1^\dagger \hat{A}_2^\dagger - \chi \hat{A}_0^\dagger \hat{A}_1 \hat{A}_2\right] . \tag{5.30}$$

In terms of dimensions, $\chi \propto \chi^{(2)}/\sqrt{\ell}$, where $\chi^{(2)}$ is the Bloembergen nonlinear polarizability coefficient and ℓ is the intracavity longitudinal mirror spacing.

The Hamiltonian term associated with the input laser pumping in a rotating frame at frequency ω_L is

$$\hat{H}_{\text{pump}} = i\hbar \int d^2\boldsymbol{x} \left[\mathcal{E}^*(\boldsymbol{x}) e^{2i\omega_L t} \hat{A}_0 - \mathcal{E}(\boldsymbol{x}) e^{-2i\omega_L t} \hat{A}_0^\dagger\right] , \tag{5.31}$$

where $\mathcal{E}(\boldsymbol{x})$ is the transverse mode profile of the pump, and it is possible to choose any shape carrying spatial structure for the input pump. The reservoir Hamiltonian is assumed to have the structure

$$\hat{H}_{\text{res}} = \sum_{i=0}^{2} \int d^2\boldsymbol{x} \left[\hat{\Gamma}_j^\dagger \hat{A}_j + \hat{\Gamma}_j \hat{A}_j^\dagger\right] + \hat{H}_{\text{res}}^0 \,. \tag{5.32}$$

Hence, there are local coupling terms to independent external free-field reservoirs for each polarization and each spatial mode. The external fields are described, as usual, by a free Hamiltonian $\hat{H}_{\text{res}}^0$. The non-unitary evolution of the system comes from the coupling between the cavity modes and the output modes, which can be approximated by a

quantum Markovian process. We carry out this calculation in a type of interaction picture, where the interaction picture operators evolve according to a reference Hamiltonian $\hat{H}_0$, such that

$$\hat{H}_0 = \hbar \sum_j \int d^2\boldsymbol{x} \omega_j^I \hat{A}_j^\dagger \hat{A}_j \,, \tag{5.33}$$

where the frequencies ω_j^I are defined in Section 5.4.1.

5.6.1 Master Equation and Quantum Langevin Form

The master equation for the density operator is the same as Eq. (5.9); here the dissipative Liouville super-operator describes an independent reservoir at each spatial location,

$$\hat{\mathcal{L}}_j \left[\hat{\rho}\right] = \gamma_j \int d^2\boldsymbol{x} \left[2\hat{A}_j \hat{\rho} \hat{A}_j^\dagger - \hat{\rho} \hat{A}_j^\dagger \hat{A}_j - \hat{A}_j^\dagger \hat{A}_j \hat{\rho}\right] . \tag{5.34}$$

As previously, we use a rotating frame with the field operators in a frame rotating at frequency ω_i^0.

There are corresponding input-output relations that describe the external modes outside the cavity, of the form $\hat{A}_i^{\text{out}} = \sqrt{2\gamma_i}\hat{A}_i - \hat{A}_i^{\text{in}}$, where $\hat{A}_i^{\text{in}}$ and $\hat{A}_i^{\text{out}}$ describe respectively the input and output fields, respectively, with input correlations that include spatial behavior:

$$\begin{aligned} \left\langle \hat{A}_i^{\text{in}}(\boldsymbol{x},t) \hat{A}_j^{\text{in}\dagger}(\boldsymbol{x}',t') \right\rangle &= \left(\bar{n}_i^{\text{th}} + 1\right) \delta_{ij} \delta\left(\boldsymbol{x} - \boldsymbol{x}'\right) \delta\left(t - t'\right) \\ \left\langle \hat{A}_i^{\text{in}\dagger}(\boldsymbol{x},t) \hat{A}_j^{\text{in}}(\boldsymbol{x}',t') \right\rangle &= \bar{n}_i^{\text{th}} \delta_{ij} \delta\left(\boldsymbol{x} - \boldsymbol{x}'\right) \delta\left(t - t'\right) . \end{aligned} \tag{5.35}$$

In the following calculations, we will assume that the reservoirs are in the vacuum state, $\bar{n}_i^{\text{th}} = 0$. However, non-zero reservoir temperatures can be readily included.

5.6.2 Stochastic Equations in the Positive-P Representation

To treat a master equation of this form, it is convenient to transform it into a Fokker–Planck equation. This can then be mapped into a set of Langevin equations similar to the quantum Heisenberg equations, except for additional stochastic terms.

As explained above, we choose the positive P representation because it generates a genuine (second order) Fokker–Planck equation with positive-definite diffusion, provided the distribution vanishes sufficiently rapidly at the phase-space boundaries. Another

alternative is the Wigner representation, but this leads to a third order differential equation. While this is exact in principle, it cannot be mapped into Langevin equations.

In the positive P-representation, following standard operator identities and phase-space partial integration, stochastic partial differential equations are obtained for normally ordered stochastic representations of quantum fields. Similar equations in the degenerate case, but with no driving field, have been treated in studies of molecular BEC EPR correlations [56–58]. In the present case, by an extension of earlier work in photonics [59] we obtain

$$\begin{aligned}
\frac{\partial A_0}{\partial t} &= -\tilde{\gamma}_0 A_0 + \mathcal{E}(\boldsymbol{x}) - \chi A_1 A_2 + \frac{i v_0^2}{2\omega_0}\nabla^2 A_0 \,, \\
\frac{\partial A_1}{\partial t} &= -\tilde{\gamma}_1 A_1 + \chi A_0 A_2^+ + \frac{i v_1^2}{2\omega_1}\nabla^2 A_1 + \sqrt{\chi A_0}\xi_1 \,, \\
\frac{\partial A_2}{\partial t} &= -\tilde{\gamma}_2 A_2 + \chi A_0 A_1^+ + \frac{i v_2^2}{2\omega_2}\nabla^2 A_2 + \sqrt{\chi A_0}\xi_2 \,,
\end{aligned} \tag{5.36}$$

The stochastic fields ξ_k that describe quantum noise are complex and Gaussian, with $\langle \xi_k(t) \rangle = 0$, and nonvanishing correlations given by

$$\langle \xi_1(\boldsymbol{x}, t)\xi_2(\boldsymbol{x}', t') \rangle = \delta(\boldsymbol{x} - \boldsymbol{x}')\delta(t - t') \,. \tag{5.37}$$

The equations that correspond to the Hermitian conjugate fields are obtained as usual by conjugating the constant terms, and replacing stochastic and noise fields so that $A_i \to A_i^+$ and $\xi_i \to \xi_i^+$, where ξ_i and ξ_j^+ are independent gaussian complex noises, while A_i and A_i^+ are six independent, complex c-number fields. One can treat these equations in several ways, either through direct numerical simulation, linearization, or using higher order stochastic diagrams methods. The results near the critical point display some of the character of certain thermodynamic and magnetic ordering critical points.

There is a subtlety, in that the EPR correlations are a quantum correction to a much larger critical point fluctuation behavior, which destroys these strong quantum correlations as the critical point is approached.

We do not have space in this brief review to give all the detailed results of this analysis, which will appear elsewhere in greater detail. However, the important point of interest here is that, although these equations describe an exponentially complex system, displaying highly nonclassical, EPR-correlated behavior, they can be numerically simulated with

no approximation apart from sampling error, which can be calculated. As an example of the corresponding analytic calculations, Eq. (5.36) can be used to draw parallels with the zero-dimensional system from Section 5.4. We can linearize them by assuming a constant pumping field $\mathcal{E}(\boldsymbol{x}) \equiv \mathcal{E}$, and substituting the steady state below-threshold solution $A_0 = \mathcal{E}/\tilde{\gamma}_0$. After a Fourier transform in space $\alpha_i(\boldsymbol{k}, t) = (2\pi)^{-1/2} \int A_i(\mathrm{x}, t) e^{i\mathbf{k}\cdot\mathbf{x}} d\mathrm{x}$ we then have

$$\begin{aligned} \frac{\partial \alpha_1}{\partial t} &= -\left(\tilde{\gamma}_1 + \frac{i v_1^2 |\mathbf{k}|^2}{2\omega_1}\right)\alpha_1 + \frac{\chi\mathcal{E}}{\tilde{\gamma}_0}\alpha_2^+ + \sqrt{\frac{\chi\mathcal{E}}{\tilde{\gamma}_0}}\zeta_1 , \\ \frac{\partial \alpha_2}{\partial t} &= -\left(\tilde{\gamma}_2 + \frac{i v_2^2 |\mathbf{k}|^2}{2\omega_2}\right)\alpha_2 + \frac{\chi\mathcal{E}}{\tilde{\gamma}_0}\alpha_1^+ + \sqrt{\frac{\chi\mathcal{E}}{\tilde{\gamma}_0}}\zeta_2 , \end{aligned} \tag{5.38}$$

where the noise is correlated as

$$\langle \zeta_1(\boldsymbol{k}, t)\zeta_2(\boldsymbol{k}', t')\rangle = \delta(\boldsymbol{k} + \boldsymbol{k}')\delta(t - t') . \tag{5.39}$$

These equations are very similar to the zero-dimensional equations (5.11), with additional $\mathbf{k}$-dependent detuning. The output fields are now correlated as

$$\begin{aligned} \left\langle \alpha_i(\mathbf{k}', \omega')\, \alpha_j(\mathbf{k}, \omega)\right\rangle &= \frac{c_{ij}(\mathbf{k}, \omega)}{2\sqrt{\gamma_i\gamma_j}}\delta(\omega + \omega')\,\delta(\boldsymbol{k} + \boldsymbol{k}') , \\ \left\langle \alpha_i^+(\mathbf{k}', \omega')\, \alpha_j(\mathbf{k}, \omega)\right\rangle &= \frac{s_{ij}(\mathbf{k}, \omega)}{2\sqrt{\gamma_i\gamma_j}}\delta(\omega + \omega')\,\delta(\boldsymbol{k} + \boldsymbol{k}') , \end{aligned} \tag{5.40}$$

with the spectral elements given by Eq. (5.15). In dimensionless variables P and ν (the same as the ones used in (5.16)), assuming $\delta_1 = \delta_2$, $\gamma_1 = \gamma_2$, $\omega_1 = \omega_2$ and $v_1 = v_2$, and introducing $\beta(\mathbf{k}) = \left(\delta_1 + v_1^2|\mathbf{k}|^2/(2\omega_1)\right)/\gamma$,

$$\begin{aligned} s_{11}(\omega) &= s_{22}(\omega) = \frac{4P^2}{\left(1 + \beta^2 - \nu^2 - P^2\right)^2 + 4\nu^2} , \\ c_{12}(\omega) = c_{21}(\omega) &= \frac{2P\left(P^2 + \nu^2 + 1 - \beta^2 - 2i\beta\right)}{\left(1 + \beta^2 - \nu^2 - P^2\right)^2 + 4\nu^2} . \end{aligned} \tag{5.41}$$

This result is similar to the standard EPR result given in Eq (5.16), except that the presence of detunings from the transverse momentum both reduces the extent of the correlations and rotates the quadrature phase.

5.7 Summary

In summary, we have reviewed the application of the positive P-representation to a variety of physical systems in quantum optics and optomechanics. These all display the highly quantum correlated behavior that characterizes the EPR-steering paradox. These methods are neither restricted to linear systems nor to systems with only a few spatial modes.

We note that Feynman's claim that probabilistic quantum simulations would not be possible without "hocus-pocus" appear to be resolved somewhat in favor of the "hocus-pocus". Not only EPR correlations, but even more fundamental violations of Bell inequalities are readily simulated with these methods.

Australian Research Council Discovery Grant.

5.8 Bibliography

[1] A. Einstein, B. Podolsky and N. Rosen, "Can quantum-mechanical description of physical reality be considered complete?", Phys. Rev. **47**, 777–780 (1935) (page 92).

[2] J. S. Bell, "On the Einstein–Podolsky–Rosen paradox", Physics **1**, 195–200 (1964) (pages 92, 95, 101).

[3] M. D. Reid, "Demonstration of the Einstein–Podolsky–Rosen paradox using nondegenerate parametric amplification", Phys. Rev. A **40**, 913–923 (1989) (pages 92, 96).

[4] M. D. Reid, P. D. Drummond, W. P. Bowen, E. G. Cavalcanti, P. K. Lam, H. A. Bachor, U. L. Andersen and G. Leuchs, "*Colloquium*: the Einstein–Podolsky–Rosen paradox: from concepts to applications", Rev. Mod. Phys. **81**, 1727–1751 (2009) (page 92).

[5] E. G. Cavalcanti, S. J. Jones, H. M. Wiseman and M. D. Reid, "Experimental criteria for steering and the Einstein–Podolsky–Rosen paradox", Phys. Rev. A **80**, 032112 (2009) (page 92).

[6] E. Wigner, "On the quantum correction for thermodynamic equilibrium", Phys. Rev. **40**, 749–759 (1932) (pages 92, 94).

[7] K. Husimi, "Some formal properties of the density matrix", Proc. Phys. Math. Soc. Jpn. **22**, 264–314 (1940) (pages 92, 94).

[8] J. E. Moyal, "Quantum mechanics as a statistical theory", Math. Proc. Cambridge **45**, 99 (1947) (pages 92, 94).

[9] R. Glauber, "Coherent and incoherent states of the radiation field", Phys. Rev. **131**, 2766–2788 (1963) (pages 92, 94).

[10] E. Sudarshan, "Equivalence of semiclassical and quantum mechanical descriptions of statistical light beams", Phys. Rev. Lett. **10**, 277–279 (1963) (pages 92, 94).

[11] S. Chaturvedi, P. D. Drummond and D. F. Walls, "Two photon absorption with coherent and partially coherent driving fields", Journal of Physics A: Mathematical and General **10**, L187–L192 (1977) (page 92).

[12] P. D. Drummond and C. W. Gardiner, "Generalised P-representations in quantum optics", Journal of Physics A: Mathematical and General **13**, 2353–2368 (1980) (pages 92, 94, 95).

[13] R. P. Feynman, "Simulating physics with computers", Int. J. Theor. Phys. **21**, 467–488 (1982) (pages 93, 101, 103).

[14]P. D. Drummond, K. J. McNeil and D. Walls, "Non-equilibrium transitions in sub/second harmonic generation", Optica Acta: International Journal of Optics **28**, 211–225 (1981) (page 93).

[15]M. D. Reid and P. D. Drummond, "Quantum correlations of phase in nondegenerate parametric oscillation", Phys. Rev. Lett. **60**, 2731–2733 (1988) (pages 93, 96).

[16]P. D. Drummond and M. D. Reid, "Correlations in nondegenerate parametric oscillation. II. Below threshold results", Phys. Rev. A **41**, 3930–3949 (1990) (pages 93, 99).

[17]Z. Y. Ou, S. F. Pereira, H. J. Kimble and K. C. Peng, "Realization of the Einstein–Podolsky–Rosen paradox for continuous variables", Phys. Rev. Lett. **68**, 3663–3666 (1992) (pages 93, 96).

[18]P. D. Drummond, B. Opanchuk, L. Rosales-Zárate and M. D. Reid, "Simulating Bell violations without quantum computers", Physica Scripta **2014**, 014009 (2014) (page 93).

[19]L. Rosales-Zárate, B. Opanchuk, P. D. Drummond and M. D. Reid, "Probabilistic quantum phase-space simulation of Bell violations and their dynamical evolution", Phys. Rev. A **90**, 022109 (2014) (pages 93, 103).

[20]K. J. McNeil and C. W. Gardiner, "Quantum statistics of parametric oscillation", Phys. Rev. A **28**, 1560–1566 (1983) (pages 93, 97).

[21]P. D. Drummond and S. J. Carter, "Quantum-field theory of squeezing in solitons", J. Opt. Soc. Am. B **4**, 1565–1573 (1987) (page 93).

[22]P. Kinsler and P. D. Drummond, "Quantum dynamics of the parametric oscillator", Phys. Rev. A **43**, 6194–6208 (1991) (page 93).

[23]S. Chaturvedi, K. Dechoum and P. D. Drummond, "Limits to squeezing in the degenerate optical parametric oscillator", Phys. Rev. A **65**, 033805 (2002) (page 93).

[24]K. Dechoum, P. D. Drummond, S. Chaturvedi and M. D. Reid, "Critical fluctuations and entanglement in the nondegenerate parametric oscillator", Phys. Rev. A **70**, 053807 (2004) (pages 93, 101).

[25]M. D. Reid and B. Yurke, "Effect of bistability and superpositions on quantum statistics in degenerate parametric oscillation", Phys. Rev. A **46**, 4131–4137 (1992) (page 93).

[26]M. D. Reid and L. Krippner, "Macroscopic quantum superposition states in nondegenerate parametric oscillation", Phys. Rev. A **47**, 552–555 (1993) (page 93).

[27]P. Deuar and P. D. Drummond, "Correlations in a BEC collision: First-principles quantum dynamics with 150 000 atoms", Phys. Rev. Lett. **98**, 120402 (2007) (page 93).

[28]C. M. Savage and K. V. Kheruntsyan, "Spatial pair correlations of atoms in molecular dissociation", Phys. Rev. Lett. **99**, 220404 (2007) (page 93).

[29]A. Gilchrist, C. W. Gardiner and P. D. Drummond, "Positive P representation: Application and validity", Phys. Rev. A **55**, 3014–3032 (1997) (page 95).

[30]P. P. Deuar and P. D. Drummond, "Gauge P representations for quantum-dynamical problems: Removal of boundary terms", Phys. Rev. A **66**, 033812 (2002) (pages 95, 96).

[31]C. W. Gardiner, *Stochastic methods*, 1st (Springer, 1985) (page 96).

[32]L. I. Plimak, M. K. Olsen, M. Fleischhauer and M. J. Collett, "Beyond the Fokker–Planck equation: Stochastic simulation of complete Wigner representation for the optical parametric oscillator", Europhys. Lett. **56**, 372–378 (2001) (page 96).

[33]R. Graham, "Statistical theory of instabilities in stationary nonequilibrium systems with applications to lasers and nonlinear optics", in *Springer tracts in modern physics: Ergebnisse der exakten Naturwissenschaften; volume 66*, edited by G. Höhler (Springer Berlin Heidelberg, Berlin, Heidelberg, 1973), pp. 1–97 (page 96).

[34]M. D. Reid and P. D. Drummond, "Correlations in nondegenerate parametric oscillation: Squeezing in the presence of phase diffusion", Phys. Rev. A **40**, 4493–4506 (1989) (page 96).

[35]C. W. Gardiner and P. Zoller, *Quantum noise*, 2nd (Springer, 2000) (page 98).

[36]S. Chaturvedi and P. D. Drummond, "Stochastic diagrams for critical point spectra", Eur. Phys. J. B **8**, 251–267 (1999) (page 100).

[37]J. F. Corney and P. D. Drummond, "Gaussian quantum operator representation for bosons", Phys. Rev. A **68**, 063822 (2003) (page 101).

[38]J. F. Corney and P. D. Drummond, "Gaussian quantum Monte Carlo methods for fermions and bosons", Phys. Rev. Lett. **93**, 260401 (2004) (page 101).

[39]P. D. Drummond, "Violations of Bell's inequality in cooperative states", Phys. Rev. Lett. **50**, 1407–1410 (1983) (page 101).

[40]M. D. Reid and D. F. Walls, "Violation of Bell's inequalities in quantum optics", Phys. Rev. Lett. **53**, 955–957 (1984) (page 101).

[41] M. D. Reid, B. Opanchuk, L. Rosales-Zárate and P. D. Drummond, "Quantum probabilistic sampling of multipartite 60-qubit Bell-inequality violations", Phys. Rev. A **90**, 012111 (2014) (page 103).

[42] M. K. Olsen, A. B. Melo, K. Dechoum and A. Z. Khoury, "Quantum phase-space analysis of the pendular cavity", Phys. Rev. A **70**, 043815 (2004) (page 104).

[43] P. D. Drummond and M. Hillery, *The quantum theory of nonlinear optics* (Cambridge University Press, Cambridge, 2014) (page 104).

[44] T. A. Palomaki, J. D. Teufel, R. W. Simmonds and K. W. Lehnert, "Entangling mechanical motion with microwave fields", Science **342**, 710–713 (2013) (pages 105, 106).

[45] S. Kiesewetter, Q. Y. He, P. D. Drummond and M. D. Reid, "Scalable quantum simulation of pulsed entanglement and Einstein–Podolsky–Rosen steering in optomechanics", Phys. Rev. A **90**, 043805 (2014) (pages 105–108).

[46] S. G. Hofer, W. Wieczorek, M. Aspelmeyer and K. Hammerer, "Quantum entanglement and teleportation in pulsed cavity optomechanics", Phys. Rev. A **84**, 052327 (2011) (page 106).

[47] M. Ludwig, A. H. Safavi-Naeini, O. Painter and F. Marquardt, "Enhanced quantum nonlinearities in a two-mode optomechanical system", Phys. Rev. Lett. **109**, 063601 (2012) (page 109).

[48] S. J. Carter, P. D. Drummond, M. D. Reid and R. M. Shelby, "Squeezing of quantum solitons", Phys. Rev. Lett. **58**, 1841–1844 (1987) (page 109).

[49] M. Rosenbluh and R. M. Shelby, "Squeezed optical solitons", Phys. Rev. Lett. **66**, 153–156 (1991) (page 109).

[50] P. D. Drummond and A. D. Hardman, "Simulation of quantum effects in Raman-active waveguides", EPL (Europhysics Letters) **21**, 279–284 (1993) (page 109).

[51] J. F. Corney, P. D. Drummond, J. Heersink, V. Josse, G. Leuchs and U. L. Andersen, "Many-body quantum dynamics of polarization squeezing in optical fibers", Phys. Rev. Lett. **97**, 023606 (2006) (page 109).

[52] A. Gatti and L. Lugiato, "Quantum images and critical fluctuations in the optical parametric oscillator below threshold", Phys. Rev. A **52**, 1675–1690 (1995) (page 109).

[53] A. Gatti, H. Wiedemann, L. A. Lugiato, I. Marzoli, G.-L. Oppo and S. M. Barnett, "Langevin treatment of quantum fluctuations and optical patterns in optical parametric oscillators below threshold", Phys. Rev. A **56**, 877–897 (1997) (page 109).

[54] A. Gatti, L. A. Lugiato, G.-L. Oppo, R. Martin, P. Di Trapani and A. Berzanskis, "From quantum to classical images", Opt. Express **1**, 21–30 (1997) (page 109).

[55] M. Hillery and L. D. Mlodinow, "Quantization of electrodynamics in nonlinear dielectric media", Phys. Rev. A **30**, 1860–1865 (1984) (page 110).

[56] K. V. Kheruntsyan and P. D. Drummond, "Quantum correlated twin atomic beams via photodissociation of a molecular Bose–Einstein condensate", Phys. Rev. A **66**, 031602 (2002) (page 112).

[57] K. V. Kheruntsyan, M. K. Olsen and P. D. Drummond, "Einstein–Podolsky–Rosen correlations via dissociation of a molecular Bose–Einstein condensate", Phys. Rev. Lett. **95**, 150405 (2005) (page 112).

[58] C. M. Savage, P. E. Schwenn and K. V. Kheruntsyan, "First-principles quantum simulations of dissociation of molecular condensates: Atom correlations in momentum space", Phys. Rev. A **74**, 033620 (2006) (page 112).

[59] P. D. Drummond and K. Dechoum, "Universality of quantum critical dynamics in a planar optical parametric oscillator", Phys. Rev. Lett. **95**, 083601 (2005) (page 112).

Chapter 6

Violations of multisetting quaternion and octonion Bell inequalities

Authors: S. Kiesewetter[1], P. D. Drummond[1], and M. D. Reid[1]

Affiliations: [1]Centre for Quantum-Atom Optics, Swinburne University of Technology, Melbourne, Australia.

Published in: Physical Review A **92**, p. 032118; accepted 8 July, 2015

6.1 Abstract

We examine the N-partite quaternion and octonion dichotomic Bell inequalities derived by Vogel and Shchukin utilizing the square identities of Euler and Degen, which apply to experiments with M settings at each spatially separated site. We reveal these inequalities to be violated by the Greenberger-Horne-Zeilinger state, for $N \geq 3$ and $2 \leq M \leq 8$. Violations are robust with respect to loss, the threshold detection efficiency being $\eta > 2^{\frac{2}{N}-1}$ for all M, implying violations for efficiencies as low as $\eta \sim 50\%$ at each site, as $N \to \infty$.

6.2 Introduction

Bell derived his famous inequalities as a way to distinguish the predictions of quantum mechanics from those of local hidden variable (LHV) theories [1, 2]. Bell's original work examined two spatially separated spin-$1/2$ (qubit) systems. Experiments have supported quantum mechanics, which predicts a violation of the Bell inequality, thus falsifying all LHV theories [3–5]. Work on Bell nonlocality for multipartite systems (sites) began with the analyses of Svetlichny [6] and Greenberger, Horne, and Zeilinger (GHZ) [7]. Mermin showed that the degree of violation increased exponentially with the number of sites, N [8]. This result translated to a surprising robustness to loss and noise for larger N [9]. Experiments have realized N-partite GHZ states for $N \sim 6 - 8$ [10, 11] and investigated their nonlocal properties for $N \sim 3$ [12].

Studies of Bell nonlocality to date have been mainly restricted to experiments with two measurement settings at each site [2, 13, 14]. Relatively little is known about Bell nonlocality for experiments with *multiple measurement settings.* Yet, multisetting experiments, first studied by Gisin and Collins [15, 16], could have an important application to quantum information tasks. For example, quantum key distribution protocols depend on implementation of random independent switching between measurement settings at the relevant sites. Random switching between more measurement settings could increase security in some cases, by making it more difficult for Eve to gain correct knowledge of measurement choices. Also important is that for some applications such as device-independent cryptography, one may require rigorous violation of the relevant Bell inequality [17]. This is also essential for the rigorous falsification of all LHV theories [18, 19]. Work by Brunner *et al.* [20] and Pal *et al.* [21] reveal multisetting Bell inequalities to be promising for this purpose.

Recent experiments have violated a Bell inequality without detection of fair sampling assumptions, using two measurement settings and detectors with efficiencies of $\eta \sim 75\%$ [19]. This violation was obtained using a nonmaximally entangled state [22, 23]. Examination of LHV models gives a lower bound on the efficiencies η required for violations of M-setting Bell inequalities as (at least) $\eta \geq 1/M$ [24–27], suggesting it may be possible to rigorously violate a Bell inequality at low efficiencies by using more than two settings. In fact, three-setting inequalities have been proposed for the two-site case, to allow a loophole-free violation with 43% at one detector (for the nonmaximally

entangled state) [20, 28]. Multisetting inequalities have been used for the loophole-free detection of the nonlocality associated with the Eistein-Podolsky-Rosen (EPR) paradox ("EPR steering" [29]), for efficiencies well below 50% [30, 31] and consistent with the bound $\eta > 1/M$ [26]. For the maximally entangled state, LHV models exist to imply 50% as a feasible efficiency for symmetric Bell experiments [32, 33], but it was initially unclear how one could violate a Bell inequality at this efficiency [20]. Subsequently, Cavalcanti *et al.* presented such an inequality for two settings that allowed violations at 50% for N-partite GHZ states as $N \to \infty$ [26, 34]. In 2012, Pal *et al.* confirmed the advantage of multisettings, by deriving a family of multisetting Bell inequalities for N-partite GHZ states that predict threshold efficiencies as low as 38% for accessible parameters [21].

In this paper, we present a second set of multisetting multipartite Bell inequalities that are predicted to be violated for reasonable efficiencies. Shchukin and Vogel (SV) have derived two multisetting Bell inequalities, whose structure is closely related to the algebra of quaternions and octonions [35]. The two-setting version of these inequalities was originally derived by Cavalcanti-Foster-Reid-Drummond (CFRD) [36] and studied in the qubit form by Salles *et al.* [37] and Cavalcanti *et al.* [34]. For two settings, violations for N-partite qubit GHZ states are possible provided $N > 3$. These violations are different from those predicted by Mermin-Ardehali-Belinskii-Klyshko (MABK) [8, 38, 39], in the sense that they allow Bell nonlocality for reduced detection efficiencies, approaching the limit $\eta \to 50\%$ as $N \to \infty$ [26, 34]. Salles *et al.* reported three-setting violations of generalizations of these inequalities, achieved using the GHZ state [37]. Up to now, however, as far as we know, there has been no explicit demonstration of violations of the quaternion or octonion Shchukin-Vogel Bell inequalities for higher settings.

Here, we show violation of the Shchukin-Vogel quaternion and octonion Bell inequalities using the GHZ state, for all $N \geq 3$ and for experiments involving $M = 2$–8 measurement settings. We show that the violations are predicted for efficiencies $\eta > 2^{\frac{2}{N}-1}$, independent of the number of measurement settings M. Although this does not give the optimistically expected dependence on M [24, 25], the efficiency approaches 50% for large number of sites for all M and is obtained for the maximally entangled GHZ state based on inequalities similar to those of MABK. The case of the generalized GHZ state has been examined for multisetting generalizations of the MABK inequalities, where it was found that more than two settings can provide an advantage in detecting nonloc-

ality [40–42]. We discuss this case for the SV inequalities, as well as the potential for experimental implementation, noting that heralding would be required.

6.3 Shchukin-Vogel multisetting Bell inequalities

We consider N spatially separated systems, and spacelike separated measurements performed on them. Supposing there are two measurement settings at each site labeled by k, so that observables $\hat{A}_k$ and $\hat{B}_k$ are measured by each setting. Then if local hidden variable (LHV) theories are valid, CFRD showed that the following inequality will hold [36]:

$$\left|\langle\prod_{k=1}^{N}(A_k + iB_k)\rangle\right|^2 \quad \leq \quad \langle\prod_{k=1}^{N}\{(A_k)^2 + (B_k)^2\}\rangle, \tag{6.1}$$

where A_k and B_k are the results of the measurements $\hat{A}_k$ and $\hat{B}_k$. The moments that are required to be measured on the left side of the inequality are determined from the expansion involving the complex numbers $f_k = A_k + iB_k$. The moments for an arbitrary number of sites may be written using a recursive relation. Inequalities of this type have been studied further in Refs. [26, 34, 43–46] and adapted to give inequalities for other forms of nonlocality in Refs. [34, 46, 47]. Mermin originally formulated inequalities using the imaginary construction [8], and the left side of this inequality is closely related to that of the MABK inequalities [38, 39]. For two sites, the CFRD inequality is

$$\begin{aligned}\langle A_1A_2 - B_1B_2\rangle^2 + \langle B_1A_2 + A_1B_2\rangle^2 \\ \leq \langle(A_1^2 + B_1^2) \times (A_2^2 + B_2^2)\rangle\,.\end{aligned} \tag{6.2}$$

Schukin and Vogel (SV) developed the theory further to derive a four and eight observable form of these inequalities [35]. We adopt their notation, denoting the outcomes of four observables that can be measured at each site k by A_k, B_k, C_k, and D_k, and the outcomes of eight observables by A_k, B_k, C_k, D_k, E_k, F_k, G_k, and H_k. The quaternion multipartite inequalities are evaluated by forming the quaternion $q_k = A_k + iB_k + jC_k + kD_k$ on the left side and using quaternion multiplication. The octonion multipartite inequalities are similarly developed using the multiplication laws for octonions.

6.3.1 Quaternion and Octonion Bell inequalities

For two sites, the quaternionic Bell inequality is [35]

$$\begin{aligned}
&\langle A_1A_2 - B_1B_2 - C_1C_2 - D_1D_2\rangle^2 \\
&+\langle B_1A_2 + A_1B_2 - D_1C_2 + C_1D_2\rangle^2 \\
&+\langle C_1A_2 - B_1D_2 + A_1C_2 + D_1B_2\rangle^2 \\
&+\langle D_1A_2 + A_1D_2 - C_1B_2 + B_1C_2\rangle^2 \\
&\leq \langle (A_1^2 + B_1^2 + C_1^2 + D_1^2) \\
&\times (A_2^2 + B_2^2 + C_2^2 + D_2^2)\rangle\,.
\end{aligned} \tag{6.3}$$

For the two-side case, we are considering the Bell inequality

$$\begin{aligned}
\left|\left\langle\left(\hat{A}_1 + i\hat{B}_1 + j\hat{C}_1 + k\hat{D}_1\right)\left(\hat{A}_2 + i\hat{B}_2 + j\hat{C}_2 + k\hat{D}_2\right)\right\rangle\right|^2 &\leq \\
\left\langle\left(\hat{A}_1^2 + \hat{B}_1^2 + \hat{C}_1^2 + \hat{D}_1^2\right)\left(\hat{A}_2^2 + \hat{B}_2^2 + \hat{C}_2^2 + \hat{D}_2^2\right)\right\rangle\,,
\end{aligned} \tag{6.4}$$

where here the well-known quaternion algebra $i^2 = j^2 = k^2 = ijk = -1, ij = -ji = k, jk = -kj = i$ applies. This inequality holds based on the fact that for any observable $\hat{A}$, $\left\langle\hat{A}\right\rangle^2 \leq \left\langle\hat{A}^2\right\rangle$. Recursive application of (6.4) allows for arbitrarily many (N) sites. The product $\left(\hat{A}_1 + i\hat{B}_1 + j\hat{C}_1 + k\hat{D}_1\right)\left(\hat{A}_2 + i\hat{B}_2 + j\hat{C}_2 + k\hat{D}_2\right)$ can be expanded into

$$\begin{aligned}
\left(\hat{A}_1 + i\hat{B}_1 + j\hat{C}_1 + k\hat{D}_1\right)\left(\hat{A}_2 + i\hat{B}_2 + j\hat{C}_2 + k\hat{D}_2\right) &= \\
\hat{I}_r + i\cdot\hat{I}_i + j\cdot\hat{I}_j + k\cdot\hat{I}_k
\end{aligned} \tag{6.5}$$

where

$$\begin{aligned}
\hat{I}_r &= \hat{A}_1\hat{A}_2 - \hat{B}_1\hat{B}_2 - \hat{C}_1\hat{C}_2 - \hat{D}_1\hat{D}_2\,, \\
\hat{I}_i &= \hat{A}_1\hat{B}_2 + \hat{B}_1\hat{A}_2 + \hat{C}_1\hat{D}_2 - \hat{D}_1\hat{C}_2\,, \\
\hat{I}_j &= \hat{A}_1\hat{C}_2 - \hat{B}_1\hat{D}_2 + \hat{C}_1\hat{A}_2 + \hat{D}_1\hat{B}_2\,, \\
\hat{I}_k &= \hat{A}_1\hat{D}_2 + \hat{B}_1\hat{C}_2 - \hat{C}_1\hat{B}_2 + \hat{D}_1\hat{A}_2\,.
\end{aligned} \tag{6.6}$$

The terms I_r, I_i, I_j, I_k satisfy the Euler square identity

$$\hat{I}_r^2 + \hat{I}_i^2 + \hat{I}_j^2 + \hat{I}_k^2 = (\hat{A}_1^2 + \hat{B}_1^2 + \hat{C}_1^2 + \hat{D}_1^2)(\hat{A}_2^2 + \hat{B}_2^2 + \hat{C}_2^2 + \hat{D}_2^2)\,, \tag{6.7}$$

which is fundamental for inequality (6.4) to hold [35].

The three-site quaternion Bell inequality is derived similarly, and its expanded form is written explicitly as [35]

$$\begin{aligned} &(I_r)^2 + (I_i)^2 + (I_j)^2 + (I_k)^2 \\ \leq \langle &(A_1^2 + B_1^2 + C_1^2 + D_1^2) \times (A_2^2 + B_2^2 + C_2^2 + D_2^2) \\ &\times (A_3^2 + B_3^2 + C_3^2 + D_3^2)\,, \rangle \end{aligned} \tag{6.8}$$

where

$$\begin{aligned} I_r \;=\; &\langle (A_1A_2 - B_1B_2 - C_1C_2 - D_1D_2)A_3 \\ &-(B_1A_2 + A_1B_2 - D_1C_2 + C_1D_2)B_3 \\ &-(C_1A_2 - B_1D_2 + A_1C_2 + D_1B_2)C_3 \\ &-(D_1A_2 + A_1D_2 - C_1B_2 + B_1C_2)D_3\,, \rangle \end{aligned}$$

$$\begin{aligned} I_i \;=\; &\langle (A_1A_2 - B_1B_2 - C_1C_2 - D_1D_2)B_3 \\ &+(B_1A_2 + A_1B_2 - D_1C_2 + C_1D_2)A_3 \\ &+(C_1A_2 - B_1D_2 + A_1C_2 + D_1B_2)D_3 \\ &-(D_1A_2 + A_1D_2 - C_1B_2 + B_1C_2)C_3\,, \rangle \end{aligned}$$

$$\begin{aligned} I_j \;=\; &\langle (A_1A_2 - B_1B_2 - C_1C_2 - D_1D_2)C_3 \\ &-(B_1A_2 + A_1B_2 - D_1C_2 + C_1D_2)D_3 \\ &+(C_1A_2 - B_1D_2 + A_1C_2 + D_1B_2)A_3 \\ &+(D_1A_2 + A_1D_2 - C_1B_2 + B_1C_2)B_3\,, \rangle \end{aligned}$$

$$
\begin{aligned}
I_k &= \langle (A_1A_2 - B_1B_2 - C_1C_2 - D_1D_2)D_3 \\
&\quad + (B_1A_2 + A_1B_2 - D_1C_2 + C_1D_2)C_3 \\
&\quad - (C_1A_2 - B_1D_2 + A_1C_2 + D_1B_2)B_3 \\
&\quad + (D_1A_2 + A_1D_2 - C_1B_2 + B_1C_2)A_3 \,.\rangle
\end{aligned} \tag{6.9}
$$

The N-site inequality is readily evaluated in a recursive form that enables computation.

We define the quantity $B_{N,M}$ as the ratio of the left-hand side (*lhs*) to the right-hand side (rhs) of the Bell inequality [Eq. (6.1), Eq. (6.3), or Eq. (6.8), etc.] that has M settings and N sites, so that

$$B_{N,M} > 1 \tag{6.10}$$

indicates a violation of the Bell inequality. The expressions for the octonion Bell inequalities are derived similarly, using the eight square identity as explained by Shchukin and Vogel [35]. A setup with three, five, six and seven settings can be achieved by simply defining the remaining settings as identical to zero. For example, a setup of three settings can be achieved by setting all $D_i = 0$.

6.4 Quantum predictions for GHZ state

We suppose the multipartite system is N spin-1/2 subsystems, prepared in a GHZ state. The GHZ state

$$|\Psi\rangle = \frac{1}{\sqrt{2}}\{|\uparrow\rangle^{\otimes N} \pm |\downarrow\rangle^{\otimes N}\} \tag{6.11}$$

is a maximally entangled state for N spin-1/2 systems. The Pauli spin observables applying to a measurement on the system at the kth site are denoted $\hat{\sigma_x}^k$, $\hat{\sigma_y}^k$, $\hat{\sigma_z}^k$. Here, $|\uparrow\rangle_k$ and $|\downarrow\rangle_k$ denote the eigenstates of $\hat{\sigma}_z^k$. We suppose the possible measurements are the set of all possible spin observables. The GHZ superposition state has been generally defined here with either sign in the superposition of probability amplitudes. However, as explained later, the final results for the Bell violations do not depend on the sign.

The right side of the inequalities (6.1) and (6.4) simplify, for the Pauli spin experiment where the results of the measurements are always ± 1. For the case of two settings (observables), the right side of the inequality reduces to 2^N. For four settings, the right side reduces to 4^N, and for eight settings, it reduces to 8^N.

To evaluate the left side of the inequalities, we need to calculate the moments of the GHZ state. For example, for $N = 2$, we substitute in (6.4) the operators $\hat{A}_k = \hat{n}_{1k} \cdot \hat{\sigma}^k$, $\hat{B}_k = \hat{n}_{2k} \cdot \hat{\sigma}^k$, $\hat{C}_k = \hat{n}_{3k} \cdot \hat{\sigma}^k$, and $\hat{D}_k = \hat{n}_{4k} \cdot \hat{\sigma}^k$, where $\hat{\sigma}^k$ is understood to act only on particle k, and $\hat{n}_{ij}$ are arbitrary unit-length vectors. One can rewrite $\hat{n}_\alpha \cdot \hat{\sigma}$ as $\sin(\theta_\alpha)\cos(\varphi_\alpha)\,\hat{\sigma}_x + \sin(\theta_\alpha)\sin(\varphi_\alpha)\,\hat{\sigma}_y + \cos(\theta_\alpha)\,\hat{\sigma}_z$. Considering a typical term $\langle\Psi^\pm|\,\hat{\sigma}_i^1\hat{\sigma}_j^2\,|\Psi^\pm\rangle$ where $i,j = \{x,y,z\}$ and $|\Psi^\pm\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow\rangle_1|\uparrow\rangle_2 \pm |\downarrow\rangle_1|\downarrow\rangle_2\right)$, as they appear in the left side of (6.4), one finds that the only nonzero terms are $\langle\hat{\sigma}_x^1\hat{\sigma}_x^2\rangle = \pm 1$, $\left\langle\hat{\sigma}_y^1\hat{\sigma}_y^2\right\rangle = \mp 1$, and $\langle\hat{\sigma}_z^1\hat{\sigma}_z^2\rangle = 1$. Hence, a term $\left\langle\hat{X}_{i1}\hat{X}_{j2}\right\rangle$ will reduce to

$$\begin{aligned}\left\langle\hat{X}_{i1}\hat{X}_{j2}\right\rangle &= \sin(\theta_{i1})\sin(\theta_{j2}) \\ &\quad\times\left(\pm\cos(\varphi_{i1})\cos(\varphi_{j2}) \mp \sin(\varphi_{i1})\sin(\varphi_{j2})\right) \\ &\quad+\cos(\theta_{i1})\cos(\theta_{j2}) \qquad (6.12)\end{aligned}$$

where $\hat{X}_1 = \hat{A}$, $\hat{X}_2 = \hat{B}$, $\hat{X}_3 = \hat{C}$, and $\hat{X}_4 = \hat{D}$. Similarly, for evaluating the predictions at larger N, only some correlations will be nonzero. The nonzero terms are as given in Table 6.1. Note that because of the alternating nature of the terms in Table 6.1, for a system with all polar angles fixed at $\theta = \frac{\pi}{4}$, the relative sign in the GHZ state produces an overall sign that affects all terms in Table I equally. This means that for such a setup, the left side of the SV inequality will be independent of the sign of the GHZ state due to the squaring of its terms.

When $N = 2$, the GHZ state reduces to a Bell state. In this case, no violation of the two-site inequality is possible, consistent with results in Ref. [37]. As an example, we could take $A = \hat{\sigma}_x$ and $B = \hat{\sigma}_y$, $C_1 = \frac{1}{\sqrt{2}}\{-\hat{\sigma}_x + \hat{\sigma}_y\}$, $D_1 = \frac{1}{\sqrt{2}}\{-\hat{\sigma}_x - \hat{\sigma}_y\}$, and $C_2 = \frac{1}{\sqrt{2}}\{\hat{\sigma}_x + \hat{\sigma}_y\}$, $D_2 = \frac{1}{\sqrt{2}}\{\hat{\sigma}_x - \hat{\sigma}_y\}$. Then we find $C_1C_2 - D_1D_2 = A_1A_2 - B_1B_2 = \hat{\sigma}_x^1\hat{\sigma}_x^2 - \hat{\sigma}_y^1\hat{\sigma}_y^2$, which gives the maximum of 4, and $-D_1C_2 + C_1D_2 = B_1A_2 + A_1B_2 = \hat{\sigma}_x^1\hat{\sigma}_y^2 + \hat{\sigma}_y^1\hat{\sigma}_x^2$ (which will give 0), and $C_1A_2 - B_1D_2 + A_1C_2 + D_1B_2 = 0$ and $D_1A_2 + A_1D_2 - C_1B_2 + B_1C_2$, so the left and right sides of the inequality are equal. That there is no violation is confirmed using the full search over all possible spin measurement choices.

Violations of the SV Bell inequality become possible for $N \geq 3$. We can verify algebraically that the choice of measurements of $C_1 = C_2 = \frac{1}{2}\left\{\hat{\sigma}_x + \sqrt{3}\hat{\sigma}_y\right\}$, $D_1 = -D_2 = \frac{1}{2}\left\{\sqrt{3}\hat{\sigma}_x - \hat{\sigma}_y\right\}$, $C_3 = \frac{1}{2}\left\{-\hat{\sigma}_x + \sqrt{3}\hat{\sigma}_y\right\}$, and $D_3 = \frac{1}{2}\left\{\sqrt{3}\hat{\sigma}_x + \hat{\sigma}_y\right\}$, will give

N	
3	$\langle\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\rangle = \pm 1$ $\langle\hat{\sigma}_x\hat{\sigma}_y\hat{\sigma}_y\rangle = \mp 1$
4	$\langle\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\rangle = \pm 1$ $\langle\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_y\hat{\sigma}_y\rangle = \mp 1$ $\langle\hat{\sigma}_y\hat{\sigma}_y\hat{\sigma}_y\hat{\sigma}_y\rangle = \pm 1$ $\langle\hat{\sigma}_z\hat{\sigma}_z\hat{\sigma}_z\hat{\sigma}_z\rangle = 1$
5	$\langle\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\rangle = \pm 1$ $\langle\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_y\hat{\sigma}_y\rangle = \mp 1$ $\langle\hat{\sigma}_x\hat{\sigma}_y\hat{\sigma}_y\hat{\sigma}_y\hat{\sigma}_y\rangle = \pm 1$
6	$\langle\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\rangle = \pm 1$ $\langle\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_y\hat{\sigma}_y\rangle = \mp 1$ $\langle\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_y\hat{\sigma}_y\hat{\sigma}_y\hat{\sigma}_y\rangle = \pm 1$ $\langle\hat{\sigma}_y\hat{\sigma}_y\hat{\sigma}_y\hat{\sigma}_y\hat{\sigma}_y\hat{\sigma}_y\rangle = \mp 1$ $\langle\hat{\sigma}_z\hat{\sigma}_z\hat{\sigma}_z\hat{\sigma}_z\hat{\sigma}_z\hat{\sigma}_z\rangle = 1$
7	$\langle\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\rangle = \pm 1$ $\langle\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_y\hat{\sigma}_y\rangle = \mp 1$ $\langle\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_y\hat{\sigma}_y\hat{\sigma}_y\hat{\sigma}_y\rangle = \pm 1$ $\langle\hat{\sigma}_x\hat{\sigma}_y\hat{\sigma}_y\hat{\sigma}_y\hat{\sigma}_y\hat{\sigma}_y\hat{\sigma}_y\rangle = \mp 1$

Table 6.1: Evaluation of Pauli products, showing the nonzero moments predicted by the GHZ state for various N. Here, every combination of $\hat{\sigma}_{xyz}$ acting on different particles must be considered as nonzero. For example, the term $\langle\hat{\sigma}_x\hat{\sigma}_y\hat{\sigma}_y\rangle$ means that $\langle\hat{\sigma}_x^1\hat{\sigma}_y^2\hat{\sigma}_y^3\rangle$, $\langle\hat{\sigma}_y^1\hat{\sigma}_x^2\hat{\sigma}_y^3\rangle$, and $\langle\hat{\sigma}_y^1\hat{\sigma}_y^2\hat{\sigma}_x^3\rangle$ must be taken into consideration.

a -4 for both of the first two terms in the I_r . However, the last two terms will be identically zero for this choice of measurement, so $I_r = -8$. We see $I_i = 0$, $I_j = -4$ and $I_k = -4\sqrt{3}$. The result is that the left side of the inequality is 2×8^2 and the right side is 4^3, which gives a significant violation. The numerical search over all possible angles confirmed the result as that giving the maximum violation.

For N sites and M settings, the right side of the SV Bell inequality becomes rhs $= M^N$. Our numerical analysis has shown to great accuracy for $N = 2, 3, \ldots, 7$ sites and $M = 2, 3, \ldots, 8$ settings that the angle choices stated in Tables 6.2–6.5 yield a maximal value for the left-hand side of the SV Bell inequalities, which is given by lhs $= M^N \times 2^{N-2}$. The rhs of the inequality is rhs $= M^N$. Not taking detector inefficiencies into account, we get a Bell violation value of $\frac{\text{lhs}}{\text{rhs}} = 2^{N-2}$, as plotted in Fig. 6.1, regardless of the number of settings used. We note that while the SV Bell inequalities are derived based on four and eight observables, using quaternion and octonion algebra, respectively, we can by setting one or more of the observables to zero in each of them arrive at Bell inequalities

N	2 settings ϕ			3 settings ϕ			
3		A	B		A	B	C
	1	0	$\frac{\pi}{2}$	1	0	$\frac{2\pi}{3}$	$\frac{\pi}{3}$
	2	0	$\frac{\pi}{2}$	2	0	$\frac{\pi}{3}$	$\frac{2\pi}{3}$
	3	0	$\frac{\pi}{2}$	3	0	$\frac{2\pi}{3}$	$\frac{\pi}{3}$
4		A	B		A	B	C
	1	0	$\frac{\pi}{2}$	1	0	$\frac{2\pi}{3}$	$\frac{\pi}{3}$
	2	0	$\frac{\pi}{2}$	2	0	$\frac{\pi}{3}$	$\frac{2\pi}{3}$
	3	0	$\frac{\pi}{2}$	3	0	$\frac{2\pi}{3}$	$\frac{\pi}{3}$
	4	0	$\frac{\pi}{2}$	4	0	$\frac{\pi}{3}$	$\frac{2\pi}{3}$

Table 6.2: Optimal measurement settings ϕ_{kn}, $k = A, B, C, D$, at each site $n = 1, \ldots, N$. The settings are given in spherical coordinates, with all $\theta \equiv 0$ at all sites, so that, e.g., $\hat{A}_1 = (\cos\phi_{A1}, \sin\phi_{A1}, 0)^T \times \hat{\sigma}^1 = \cos\phi_{A1}\hat{\sigma}^1_x + \sin\phi_{A1}\hat{\sigma}^1_y$. A higher number of sites N can be treated by simply continuing the alternating pattern. The settings for higher M are given in Tables 6.3–6.5.

N	4 settings ϕ					5 settings ϕ					
3		A	B	C	D		A	B	C	D	E
	1	0	$\frac{\pi}{6}$	$-\frac{\pi}{3}$	$\frac{\pi}{2}$	1	0	$\frac{\pi}{6}$	$-\frac{\pi}{3}$	$\frac{4\pi}{3}$	$\frac{2\pi}{3}$
	2	0	$\frac{5\pi}{6}$	$\frac{4\pi}{3}$	$\frac{\pi}{2}$	2	0	$\frac{5\pi}{6}$	$\frac{4\pi}{3}$	$-\frac{\pi}{3}$	$\frac{\pi}{3}$
	3	0	$\frac{\pi}{6}$	$-\frac{\pi}{3}$	$\frac{\pi}{2}$	3	0	$\frac{\pi}{6}$	$-\frac{\pi}{3}$	$\frac{4\pi}{3}$	$\frac{2\pi}{3}$
4		A	B	C	D		A	B	C	D	E
	1	0	$\frac{\pi}{6}$	$-\frac{\pi}{3}$	$\frac{\pi}{2}$	1	0	$\frac{\pi}{6}$	$-\frac{\pi}{3}$	$\frac{4\pi}{3}$	$\frac{2\pi}{3}$
	2	0	$\frac{5\pi}{6}$	$\frac{4\pi}{3}$	$\frac{\pi}{2}$	2	0	$\frac{5\pi}{6}$	$\frac{4\pi}{3}$	$-\frac{\pi}{3}$	$\frac{\pi}{3}$
	3	0	$\frac{\pi}{6}$	$-\frac{\pi}{3}$	$\frac{\pi}{2}$	3	0	$\frac{\pi}{6}$	$-\frac{\pi}{3}$	$\frac{4\pi}{3}$	$\frac{2\pi}{3}$
	4	0	$\frac{5\pi}{6}$	$\frac{4\pi}{3}$	$\frac{\pi}{2}$	4	0	$\frac{5\pi}{6}$	$\frac{4\pi}{3}$	$-\frac{\pi}{3}$	$\frac{\pi}{3}$

Table 6.3: Optimal angles for four and five measurement settings, as described for Table 6.2.

for fewer settings. In this way, the quaternion Bell inequality is also a Bell inequality for $M = 2, 3, 4$ settings, reducing in the case of $M = 2$ to the CFRD Bell inequality for spins [34, 37]. Similarly, the octonion SV Bell inequality gives rise to the Bell inequalities with $M = 5, 6, 7$, and 8 settings.

6.5 Detection inefficiencies, Noise and heralding

We now examine the important question of the effect of detection inefficiencies on the Bell violations. The following analysis will model the loss that takes place due to detector inefficiencies and propagation through the medium. We extend the analysis given in Ref.

N	6 settings ϕ							7 settings ϕ							
3		A	B	C	D	E	F		A	B	C	D	E	F	G
	1	0	$\frac{\pi}{8}$	$\frac{3\pi}{8}$	$\frac{5\pi}{8}$	$\frac{7\pi}{8}$	$\frac{\pi}{2}$	1	0	$\frac{\pi}{8}$	$\frac{3\pi}{8}$	$\frac{5\pi}{8}$	$\frac{7\pi}{8}$	$-\frac{\pi}{3}$	$\frac{4\pi}{3}$
	2	0	$\frac{7\pi}{8}$	$\frac{5\pi}{3}$	$\frac{3\pi}{8}$	$\frac{\pi}{8}$	$\frac{\pi}{2}$	2	0	$\frac{7\pi}{8}$	$\frac{5\pi}{8}$	$\frac{3\pi}{8}$	$\frac{\pi}{8}$	$\frac{4\pi}{3}$	$-\frac{\pi}{3}$
	3	0	$\frac{\pi}{8}$	$\frac{3\pi}{8}$	$\frac{5\pi}{8}$	$\frac{7\pi}{8}$	$\frac{\pi}{2}$	3	0	$\frac{\pi}{8}$	$\frac{3\pi}{8}$	$\frac{5\pi}{8}$	$\frac{7\pi}{8}$	$-\frac{\pi}{3}$	$\frac{4\pi}{3}$
4		A	B	C	D	E	F		A	B	C	D	E	F	G
	1	0	$\frac{\pi}{8}$	$\frac{3\pi}{8}$	$\frac{5\pi}{8}$	$\frac{7\pi}{8}$	$\frac{\pi}{2}$	1	0	$\frac{\pi}{8}$	$\frac{3\pi}{8}$	$\frac{5\pi}{8}$	$\frac{7\pi}{8}$	$-\frac{\pi}{3}$	$\frac{4\pi}{3}$
	2	0	$\frac{7\pi}{8}$	$\frac{5\pi}{3}$	$\frac{3\pi}{8}$	$\frac{\pi}{8}$	$\frac{\pi}{2}$	2	0	$\frac{7\pi}{8}$	$\frac{5\pi}{8}$	$\frac{3\pi}{8}$	$\frac{\pi}{8}$	$\frac{4\pi}{3}$	$-\frac{\pi}{3}$
	3	0	$\frac{\pi}{8}$	$\frac{3\pi}{8}$	$\frac{5\pi}{8}$	$\frac{7\pi}{8}$	$\frac{\pi}{2}$	3	0	$\frac{\pi}{8}$	$\frac{3\pi}{8}$	$\frac{5\pi}{8}$	$\frac{7\pi}{8}$	$-\frac{\pi}{3}$	$\frac{4\pi}{3}$
	4	0	$\frac{7\pi}{8}$	$\frac{5\pi}{3}$	$\frac{3\pi}{8}$	$\frac{\pi}{8}$	$\frac{\pi}{2}$	4	0	$\frac{7\pi}{8}$	$\frac{5\pi}{8}$	$\frac{3\pi}{8}$	$\frac{\pi}{8}$	$\frac{4\pi}{3}$	$-\frac{\pi}{3}$

Table 6.4: Optimal angles for six and seven measurement settings, as described for Table 6.2.

N	8 settings ϕ								
3		A	B	C	D	E	F	G	H
	1	0	$\frac{\pi}{8}$	$\frac{3\pi}{8}$	$\frac{5\pi}{8}$	$\frac{7\pi}{8}$	$-\frac{\pi}{8}$	$-\frac{5\pi}{8}$	$\frac{3\pi}{2}$
	2	0	$\frac{7\pi}{8}$	$\frac{5\pi}{8}$	$\frac{3\pi}{8}$	$\frac{\pi}{8}$	$\frac{9\pi}{8}$	$\frac{13\pi}{8}$	$-\frac{\pi}{2}$
	3	0	$\frac{\pi}{8}$	$\frac{3\pi}{8}$	$\frac{5\pi}{8}$	$\frac{7\pi}{8}$	$-\frac{\pi}{8}$	$-\frac{5\pi}{8}$	$\frac{3\pi}{2}$
4		A	B	C	D	E	F	G	H
	1	0	$\frac{\pi}{8}$	$\frac{3\pi}{8}$	$\frac{5\pi}{8}$	$\frac{7\pi}{8}$	$-\frac{\pi}{8}$	$-\frac{5\pi}{8}$	$\frac{3\pi}{2}$
	2	0	$\frac{7\pi}{8}$	$\frac{5\pi}{8}$	$\frac{3\pi}{8}$	$\frac{\pi}{8}$	$\frac{9\pi}{8}$	$\frac{13\pi}{8}$	$-\frac{\pi}{2}$
	3	0	$\frac{\pi}{8}$	$\frac{3\pi}{8}$	$\frac{5\pi}{8}$	$\frac{7\pi}{8}$	$-\frac{\pi}{8}$	$-\frac{5\pi}{8}$	$\frac{3\pi}{2}$
	4	0	$\frac{7\pi}{8}$	$\frac{5\pi}{8}$	$\frac{3\pi}{8}$	$\frac{\pi}{8}$	$\frac{9\pi}{8}$	$\frac{13\pi}{8}$	$-\frac{\pi}{2}$

Table 6.5: Optimal angles for eight measurement settings, as described for Table 6.2.

[34]. We represent the experimental observables as the Schwinger spin observables,

$$\begin{aligned}
\hat{\sigma}_z &= \hat{a}_+^\dagger \hat{a}_+ - \hat{a}_-^\dagger \hat{a}_- \\
\hat{\sigma}_x &= \hat{a}_+^\dagger \hat{a}_- + \hat{a}_+ \hat{a}_-^\dagger \\
\hat{\sigma}_y &= (\hat{a}_+^\dagger \hat{a}_- - \hat{a}_+ \hat{a}_-^\dagger)/i \\
\hat{\sigma}^2 &= \hat{n}(\hat{n}+2) \\
\hat{n} &= \hat{a}_+^\dagger \hat{a}_+ + \hat{a}_-^\dagger \hat{a}_-
\end{aligned} \tag{6.13}$$

where $\hat{a}_\pm$ is the destruction boson operator for two orthogonal modes $\pm$ at the site k. Here, $\hat{\sigma}^2 = (\hat{\sigma}_x)^2 + (\hat{\sigma}_y)^2 + (\hat{\sigma}_z)^2$, and $\hat{n}$ is the total number operator for each site. The superscripts k denote which site is being referred to. Usually, the modes $\pm$ correspond to orthogonal field polarization and $\hat{a}_\pm^{k\dagger}\hat{a}_\pm^k$ is the number operator for a photonic mode $\pm$ at site k. At each site, the spin measurement $\hat{\sigma}_\phi^k$ is performed by a

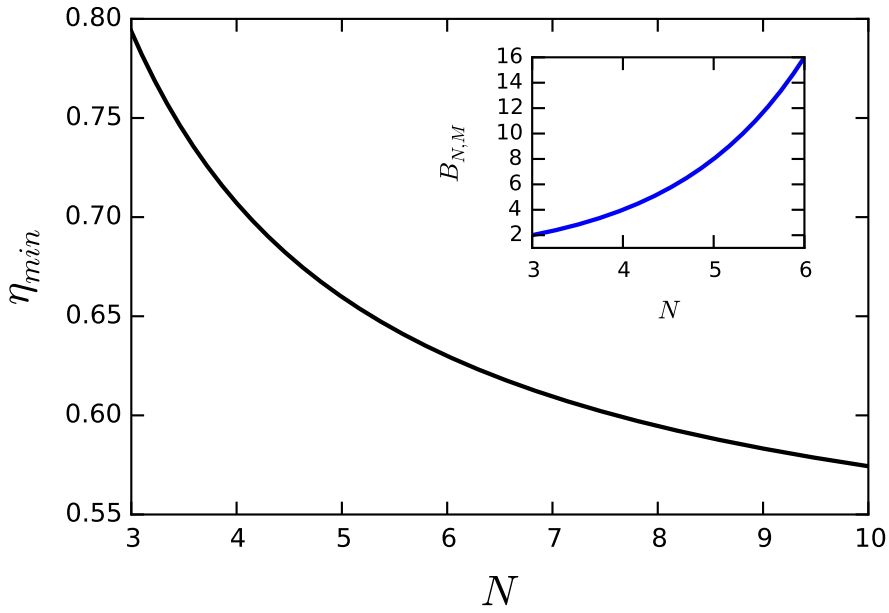

Figure 6.1: Threshold detector efficiency for N sites: The minimal detector efficiency $\eta_{\min}$ needed to observe Bell violations for a given number of sites N. Inset: The Bell value $B_{N,M}$ vs N. $B_{N,M} > 1$ indicates violation of the Bell inequality. The results are independent of the number of measurement settings M.

polarization measurement chosen at a suitable angle (and possibly with phase shifts) [48]. For example, the modes at the detectors after the polarizer measurement for $\hat{\sigma}_\phi$ can be represented by the transformed modes $\hat{c}_+$, $\hat{c}_-$which are linear combinations of $\hat{a}_\pm$ and for which $\hat{c}_+^\dagger \hat{c}_+ - \hat{c}_-^\dagger \hat{c}_- = \hat{\sigma}_\phi$. The photonic GHZ state is

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|1\rangle^{\otimes N} + e^{i\phi}|0\rangle^{\otimes N}) \,, \tag{6.14}$$

where $|n\rangle^{\otimes N} \equiv \bigotimes_{k=1}^{N} |n\rangle_k$ and $|n\rangle_k$ are the eigenstates of $\hat{n}^k$. We see that components of the photonic GHZ state are mapped into the ± 1 eigenstates of $\hat{\sigma}_z$: $|\downarrow\rangle_k \to |0\rangle_{+k}|1\rangle_{-k}$ and $|\uparrow\rangle_k \to |1\rangle_{+k}|0\rangle_{-k}$.

With nonideal detection efficiencies, not all photons that impinge on the detector will be recorded. For each emission event, there are three outcomes at each site. We follow the traditional example [2] and denote the outcomes of the observables A, B, etc. as $+1$, -1, or 0 if the photon is detected "up," "down," or "not at all," respectively. There are two detectors at each site, placed to determine the number of photons (0 or 1) corresponding to modes $\hat{c}_+^{(k)}$ and $\hat{c}_-^{(k)}$, respectively (see Refs. [2, 49] for details). We model the inefficiency (loss) of each detector using a simple beam-splitter model, where we evaluate the moments

of detected fields $\hat{a}_{det}$ given by

$$\hat{a}_{det} = \sqrt{\eta}\hat{a} + \sqrt{1-\eta}\hat{a}_{\text{vac}} \tag{6.15}$$

Here, η is the probability that an incoming photon is detected. The $\hat{a}$ is the boson operator for the incoming field mode, and $\hat{a}_{\text{vac}}$ is the boson operator for a vacuum reservoir mode that couples to the incoming field and into which quanta are lost.

Equation (6.15) enables calculation of the predicted effect of detection inefficiencies on the moments needed for the Bell inequalities. It is important to understand how the experiment would be carried out. In a LHV theory, the quantities $A, B, \ldots$ that appear in the inequality are numbers that now have the possible values $\pm 1, 0$. The derivation of the Bell SV inequalities does not restrict the values that the variables can attain, and hence the same inequalities will still hold rigorously in the lossy scenario, to allow a loophole-free test of LHV theories. In an ideal situation, one *heralds* the emission event and detects the value of the observables $A, B, \ldots$ at each of the sites, some being $+1$ or -1, and some being 0. One evaluates from the data the moments that are defined on the left and right sides of the inequality. If these moments violate the inequalities, then we have confirmed the failure of all LHV theories to describe the experiment.

What do we expect to obtain for a given detection inefficiency η? On the lhs of the inequalities, we measure moments like those in the Clauser-Horne-Shimony-Holt (CHSH) Bell inequalities [2, 13], such as $\langle A_1 A_2 \ldots A_N \rangle$. Consider first two sites. We denote the joint probability for the outcome $+1$ and -1 at sites 1 and 2, respectively, by $P(+,-)$, etc. Thus,

$$\langle A_1 A_2 \rangle = P(+,+) + P(-,-) - P(+,-) - P(-,+) \tag{6.16}$$

The probabilities where the outcome is the null event 0 do not directly contribute because their value is $A = 0$, but the null events contribute to the total number of counts and hence to the normalization of the probabilities. The joint probabilities $P(+,+)$, etc. involve a single photon detection at each site, and hence will scale as η^2. More generally, with N sites, there will be N photon detections and we will obtain a scaling of η^N. Hence we obtain, for the left side of the Bell inequalities, lhs $= \eta^{2N} M^N \times 2^{N-2}$. On the other

hand, the moments on the rhs involve the expectation values of the A^2 values. The A^2 value is always 1 if the photon is detected either "up" or "down," and is always 0 if the photon is not detected. Denoting the probabilities for the outcomes $+1$, -1, and 0 at site k by $P_k(+)$, $P_k(-)$, and $P_k(0)$, respectively, we note that

$$\langle A_1^2 A_2^2 \rangle = P(+,+) + P(+,-) + P(-,+) + P(-,-) \,. \tag{6.17}$$

The moment scales with efficiency as η^2. Extending to the product of results for $\langle A_1^2 A_2^2 \ldots \rangle$ based on measurement at the N sites, we see the prediction is rhs $= \eta^N M^N$. This means that an experiment involving N sites requires a minimum detector efficiency of $\eta_{\min} = 2^{\frac{2}{N}-1}$ in order to detect Bell violation, regardless of the number of settings M used (Fig. 6.1).

We note that because of the square on the lhs of the SV Bell inequalities, the violation cannot straightforwardly be obtained without determining the probability of the zero detection event at all sites. Thus, in the experiment described, there is a need for heralding the emission events in order to obtain a violation of the inequalities [50]. This provides a substantial limitation. Alternative experimental arrangements could be explored in the future. For example, it is possible that an analysis of the type presented by Garg and Mermin for the CHSH inequalities may be useful in providing an alternative experiment, without the need for heralding [51].

As a further calculation of the sensitivity of the Bell violations to possible noise sources, we show in Fig. 6.2 the effect of Gaussian random noise on the measurement angle. Regardless of the number of settings, the violations are robust to this noise source.

6.6 Generalized GHZ states

Now we turn to the generalized GHZ state

$$|\Psi\rangle = \cos\alpha|\uparrow\rangle^{\otimes N} + \sin\alpha|\downarrow\rangle^{\otimes N} \,. \tag{6.18}$$

It has been shown for the bipartite ($N = 2$) case that the nonmaximally entangled state (which has $\alpha < \pi/4$) allows violations of Bell inequalities for *lower* efficiency thresholds

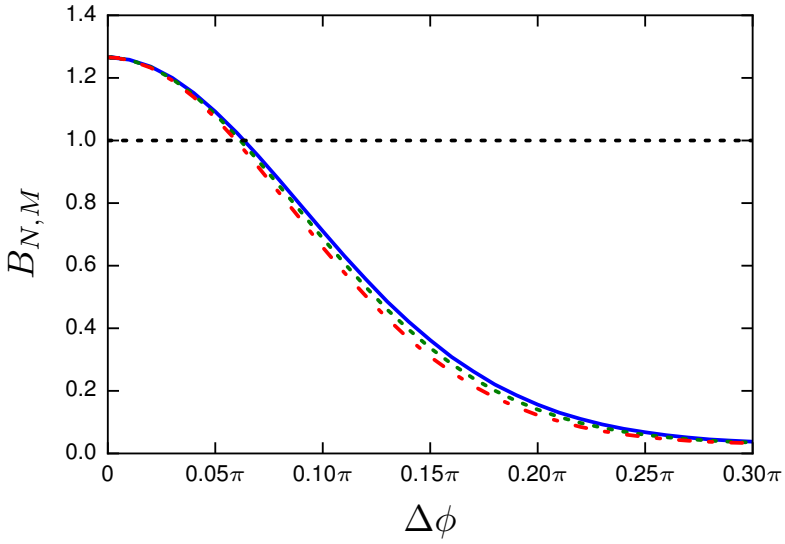

Figure 6.2: Effect of measurement noise: The Bell value $B_{N,M}$ achievable for $N = 4$ sites and (from top to lower line) $M = 2, 3, 8$ measurement settings when the ensemble averages are evaluated with a random Gaussian fluctuation of standard deviation $\Delta\phi$ in the measurement angles. Here, detector efficiency is $\eta = 0.75$. Violation of the Bell inequality is achieved when $B_{N,M} > 1$. The mean angle setting is optimized according to the values given in the Tables 6.2 and 6.5, respectively.

than the maximally entangled state ($\alpha = \pi/4$) [23]. These violations are achieved with asymmetric Clauser-Horne-type inequalities [19, 20, 22, 28]. Also interesting is that the MABK Bell inequalities do not allow any violation for the generalized GHZ state $|\Phi\rangle = \cos\alpha|000\rangle + \sin\alpha|111\rangle$ for some values of α, namely $\sin 2\alpha \leq 1/\sqrt{2^{N-1}}$ and N odd [40]. However, a violation becomes possible on using extensions of the MABK inequalities for four measurement settings [41, 42].

We are thus motivated to ask the following question: For what range of α does the two-setting CFRD-type Bell inequality allow a test of nonlocality, and can this range be extended (as in the MABK case) by considering multiple settings? Our answer to this question is shown in Figs. 6.3 and 6.4.

For detection efficiencies at $\eta \sim 0.65$, one would require GHZ states with $N \geq 6$ to obtain a violation of the Bell inequalities. This would not seem unrealistic, given the experiments of Refs. [10] which report creation of photonic GHZ states for $N = 6 - -8$. A challenging feature of the experimental implementation would also be to correctly measure the probability of the null (zero-detection) event.

We find that violations are possible over a significant range of α, though the range is a more restricted range than that obtainable for the MABK inequalities. This is not

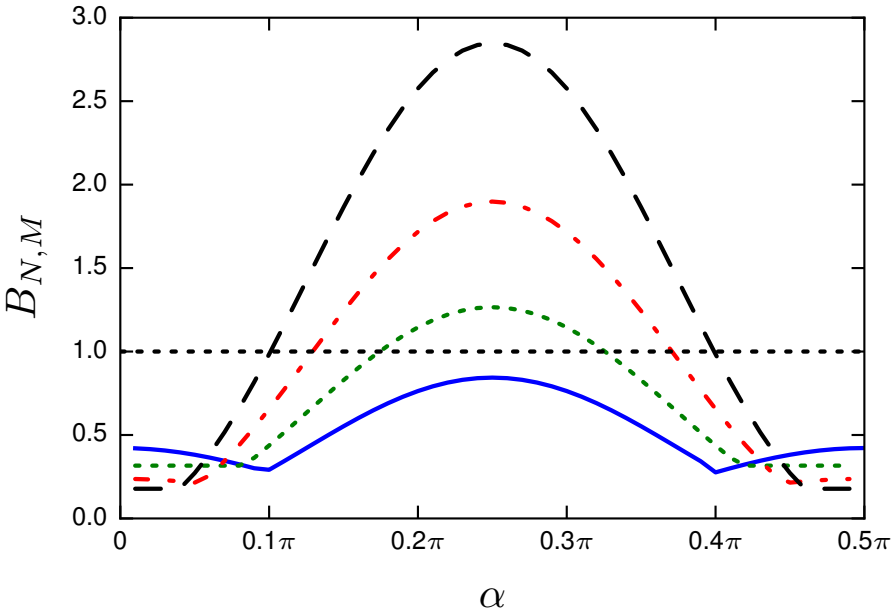

Figure 6.3: Violations for the generalized GHZ state: Here $B_{N,M} > 1$ indicates violation of the Bell inequality with (from lower to top line) $N = 3, 4, 5, 6$ sites and $M = 2$ measurement settings for a detector efficiency of $\eta = 0.75\%$. The results for settings $M = 2 - -8$ are indistinguishable.

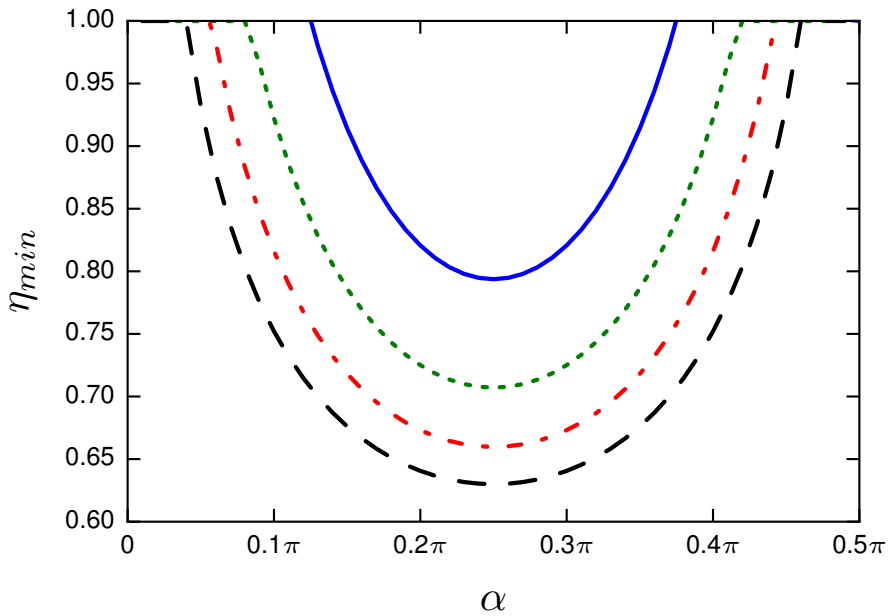

Figure 6.4: Threshold detector efficiencies for the generalized GHZ state: Here, $B_{N,M} > 1$ indicates violation of the Bell inequality with (from top to lower line) $N = 3, 4, 5, 6$ sites and $M = 2$ measurement settings. The results for settings $M = 2 - -8$ are indistinguishable.

unexpected, given that the inequalities here have a smaller violation ratio B and do not violate LHV theories at all for $N = 2$. As with the MABK case, the range of α giving a violation expands with N, but in this case the results are essentially unchanged with M. Our numerical results suggest that the respective ranges of α coincide up to a shift of N by one; that is, our numerical results are consistent with the relation $\sin 2\alpha \leq 1/\sqrt{2^{N-2}}$ for the SV inequalities.

6.7 Discussion and Conclusion

We have presented multisetting Bell inequalities that allow violation for an N-partite GHZ state when the efficiencies at each site approach $\eta \to 50\%$. This supplements the earlier work by Pal *et al.*, which reported multisetting Bell violations for reasonable efficiencies (as low as 38% for eight sites and 11 settings) [21]. We show that for the inequalities considered in this paper, an efficiency limit of arbitrarily close to 50% is attainable independently of the number of measurement settings and for the maximally entangled N-partite GHZ state that has a symmetric weighting of its composite spin states, provided one uses a sufficiently large number of sites N. The inequalities used here are different from those of Pal *et al.*, in that the "no detection" event is assigned the outcome of zero that is distinct from the Pauli spin values.

The violations are obtained using the spin version of the Shchukin-Vogel inequalities. These, however, require three or more sites ($N \geq 3$) in order to allow a test of Bell nonlocality even at ideal efficiencies. By comparison, our studies for this case suggest that violations are not possible at such low-efficiency thresholds for multipartite W states. We also note that the natural multipartite extensions of Gisin's original multisetting inequalities [15] would not lead to lower threshold efficiencies.

6.8 Acknowledgments

We wish to acknowledge funding support from the Australian Research Council through the Discovery Grants program.

6.9 Bibliography

[1]J. S. Bell, "On the Einstein–Podolsky–Rosen paradox", Physics **1**, 195–200 (1964) (page 122).

[2]J. F. Clauser and A. Shimony, "Bell's theorem. Experimental tests and implications", Reports on Progress in Physics **41**, 1881 (1978) (pages 122, 132, 133).

[3]S. J. Freedman and J. F. Clauser, "Experimental test of local hidden-variable theories", Phys. Rev. Lett. **28**, 938–941 (1972) (page 122).

[4]A. Aspect, J. Dalibard and G. Roger, "Experimental test of Bell's inequalities using time-varying analyzers", Phys. Rev. Lett. **49**, 1804–1807 (1982) (page 122).

[5]G. Weihs, T. Jennewein, C. Simon, H. Weinfurter and A. Zeilinger, "Violation of Bell's inequality under strict Einstein locality conditions", Phys. Rev. Lett. **81**, 5039–5043 (1998) (page 122).

[6]G. Svetlichny, "Distinguishing three-body from two-body nonseparability by a Bell-type inequality", Phys. Rev. D **35**, 3066–3069 (1987); D. Collins, N. Gisin, S. Popescu, D. Roberts and V. Scarani, "Bell-type inequalities to detect true n-body nonseparability", Phys. Rev. Lett. **88**, 170405 (2002), page 122.

[7]D. M. Greenberger, M. Horne and A. Zeilinger, *Bell's theorem, quantum theory and conceptions of the universe*, edited by M. Kafatos (Kluwer Dordrecht, The Netherlands, 1989) (page 122).

[8]N. D. Mermin, "Extreme quantum entanglement in a superposition of macroscopically distinct states", Phys. Rev. Lett. **65**, 1838–1840 (1990) (pages 122–124).

[9]S. L. Braunstein and A. Mann, "Noise in Mermin's *n*-particle Bell inequality", Phys. Rev. A **47**, R2427–R2430 (1993); G. Brassard, A. Broadbent and A. Tapp, "Recasting Mermin's multi-player game into the framework of pseudo-telepathy", Quantum Info. Comput. **5**, 538–550 (2005), page 122.

[10]C.-Y. Lu, X.-Q. Zhou, O. Guhne, W.-B. Gao, J. Zhang, Z.-S. Yuan, A. Goebel, T. Yang and J.-W. Pan, "Experimental entanglement of six photons in graph states", Nat Phys **3**, 91–95 (2007); Y.-F. Huang, B.-H. Liu, L. Peng, Y.-H. Li, L. Li, C.-F. Li and G.-C. Guo, "Experimental generation of an eight-photon Greenberger–Horne–Zeilinger state", Nature Communications **2**, Article, 546 (2011); X.-C. Yao, T.-X. Wang, P. Xu, H. Lu, G.-S. Pan, X.-H. Bao, C.-Z. Peng, C.-Y. Lu, Y.-A. Chen and J.-W. Pan, "Observation of

eight-photon entanglement", Nat Photon **6**, 225–228 (2012); H. Lu, L.-K. Chen, C. Liu, P. Xu, X.-C. Yao, L. Li, N.-L. Liu, B. Zhao, Y.-A. Chen and J.-W. Pan, "Experimental realization of a concatenated Greenberger–Horne–Zeilinger state for macroscopic quantum superpositions", ibid. **8**, Letter, 364–368 (2014), pages 122, 135.

[11]J.-W. Pan, D. Bouwmeester, M. Daniell, H. Weinfurter and A. Zeilinger, "Experimental test of quantum nonlocality in three-photon Greenberger–Horne–Zeilinger entanglement", Nature **403**, 515–519 (2000) (page 122).

[12]J. Lavoie, R. Kaltenbaek and K. J. Resch, "Experimental violation of Svetlichny's inequality", New Journal of Physics **11**, 073051 (2009); H.-X. Lu, J.-Q. Zhao, X.-Q. Wang and L.-Z. Cao, "Experimental demonstration of tripartite entanglement versus tripartite nonlocality in three-qubit Greenberger–Horne–Zeilinger–class states", Phys. Rev. A **84**, 012111 (2011); B. P. Lanyon, M. Zwerger, P. Jurcevic, C. Hempel, W. Dür, H. J. Briegel, R. Blatt and C. F. Roos, "Experimental violation of multipartite Bell inequalities with trapped ions", Phys. Rev. Lett. **112**, 100403 (2014), page 122.

[13]J. F. Clauser, M. A. Horne, A. Shimony and R. A. Holt, "Proposed experiment to test local hidden-variable theories", Phys. Rev. Lett. **23**, 880–884 (1969) (pages 122, 133).

[14]N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani and S. Wehner, "Bell nonlocality", Rev. Mod. Phys. **86**, 419–478 (2014) (page 122).

[15]N. Gisin, "Bell inequality for arbitrary many settings of the analyzers", Physics Letters A **260**, 1–3 (1999) (pages 122, 137).

[16]D. Collins and N. Gisin, "A relevant two qubit Bell inequality inequivalent to the CHSH inequality", Journal of Physics A: Mathematical and General **37**, 1775 (2004) (page 122).

[17]A. K. Ekert, "Quantum cryptography based on Bell's theorem", Phys. Rev. Lett. **67**, 661–663 (1991); V. Scarani and N. Gisin, "Quantum communication between N partners and Bell's inequalities", ibid. **87**, 117901 (2001); J. Barrett, L. Hardy and A. Kent, "No signaling and quantum key distribution", ibid. **95**, 010503 (2005); A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio and V. Scarani, "Device-independent security of quantum cryptography against collective attacks", ibid. **98**, 230501 (2007), page 122.

[18]P. M. Pearle, "Hidden-variable example based upon data rejection", Phys. Rev. D **2**, 1418–1425 (1970) (page 122).

[19]M. Giustina, A. Mech, S. Ramelow, B. Wittmann, J. Kofler, J. Beyer, A. Lita, B. Calkins, T. Gerrits, S. W. Nam, R. Ursin and A. Zeilinger, "Bell violation using entangled photons without the fair-sampling assumption", Nature **497**, Letter, 227–230 (2013); B. G. Christensen, K. T. McCusker, J. B. Altepeter, B. Calkins, T. Gerrits, A. E. Lita, A. Miller, L. K. Shalm, Y. Zhang, S. W. Nam, N. Brunner, C. C. W. Lim, N. Gisin and P. G. Kwiat, "Detection-loophole-free test of quantum nonlocality, and applications", Phys. Rev. Lett. **111**, 130406 (2013), pages 122, 135.

[20]N. Brunner, N. Gisin, V. Scarani and C. Simon, "Detection loophole in asymmetric Bell experiments", Phys. Rev. Lett. **98**, 220403 (2007) (pages 122, 123, 135).

[21]K. F. Pál, T. Vértesi and N. Brunner, "Closing the detection loophole in multipartite Bell tests using Greenberger–Horne–Zeilinger states", Phys. Rev. A **86**, 062111 (2012) (pages 122, 123, 137).

[22]J. F. Clauser and M. A. Horne, "Experimental consequences of objective local theories", Phys. Rev. D **10**, 526–535 (1974) (pages 122, 135).

[23]P. H. Eberhard, "Background level and counter efficiencies required for a loophole-free Einstein–Podolsky–Rosen experiment", Phys. Rev. A **47**, R747–R750 (1993) (pages 122, 135).

[24]S. Massar and S. Pironio, "Violation of local realism versus detection efficiency", Phys. Rev. A **68**, 062109 (2003) (pages 122, 123).

[25]T. Vértesi, S. Pironio and N. Brunner, "Closing the detection loophole in Bell experiments using qudits", Phys. Rev. Lett. **104**, 060401 (2010) (pages 122, 123).

[26]M. D. Reid, "Two-setting multisite Bell inequalities for loophole-free tests with up to 50% loss", Phys. Rev. A **87**, 062108 (2013) (pages 122–124).

[27]M. D. Reid, "Monogamy inequalities for the Einstein–Podolsky–Rosen paradox and quantum steering", Phys. Rev. A **88**, 062108 (2013) (page 122).

[28]A. Cabello and J.-Å. Larsson, "Minimum detection efficiency for a loophole-free atom-photon Bell experiment", Phys. Rev. Lett. **98**, 220402 (2007) (pages 122, 135).

[29]H. M. Wiseman, S. J. Jones and A. C. Doherty, "Steering, entanglement, nonlocality, and the Einstein–Podolsky–Rosen paradox", Phys. Rev. Lett. **98**, 140402 (2007); S. J. Jones, H. M. Wiseman and A. C. Doherty, "Entanglement, Einstein–Podolsky–Rosen correlations, Bell nonlocality, and steering", Phys. Rev. A **76**, 052116 (2007); E. G. Cavalcanti, S. J. Jones, H. M. Wiseman and M. D. Reid, "Experimental criteria for steering and the Einstein–Podolsky–Rosen paradox", ibid. **80**, 032112 (2009), page 123.

[30]A. J. Bennet, D. A. Evans, D. J. Saunders, C. Branciard, E. G. Cavalcanti, H. M. Wiseman and G. J. Pryde, "Arbitrarily loss-tolerant Einstein–Podolsky–Rosen steering allowing a demonstration over 1 km of optical fiber with no detection loophole", Phys. Rev. X **2**, 031003 (2012) (page 123).

[31]D. H. Smith, G. Gillett, M. P. de Almeida, C. Branciard, A. Fedrizzi, T. J. Weinhold, A. Lita, B. Calkins, T. Gerrits, H. M. Wiseman, S. W. Nam and A. G. White, "Conclusive quantum steering with superconducting transition-edge sensors", Nature Communications **3**, Article, 625 (2012); B. Wittmann, S. Ramelow, F. Steinlechner, N. K. Langford, N. Brunner, H. M. Wiseman, R. Ursin and A. Zeilinger, "Loophole-free Einstein–Podolsky–Rosen experiment via quantum steering", New Journal of Physics **14**, 053030 (2012), page 123.

[32]N. Gisin and B. Gisin, "A local hidden variable model of quantum correlation exploiting the detection loophole", Physics Letters A **260**, 323–327 (1999) (page 123).

[33]J.-Å. Larsson and J. Semitecolos, "Strict detector-efficiency bounds for $\boldsymbol{n}$-site Clauser-Horne inequalities", Phys. Rev. A **63**, 022117 (2001); A. Cabello, D. Rodríguez and I. Villanueva, "Necessary and sufficient detection efficiency for the Mermin inequalities", Phys. Rev. Lett. **101**, 120402 (2008), page 123.

[34]E. G. Cavalcanti, Q. Y. He, M. D. Reid and H. M. Wiseman, "Unified criteria for multipartite quantum nonlocality", Phys. Rev. A **84**, 032115 (2011) (pages 123, 124, 130, 131).

[35]E. Shchukin and W. Vogel, "Quaternions, octonions, and Bell-type inequalities", Phys. Rev. A **78**, 032104 (2008) (pages 123–127).

[36]E. G. Cavalcanti, C. J. Foster, M. D. Reid and P. D. Drummond, "Bell inequalities for continuous-variable correlations", Phys. Rev. Lett. **99**, 210405 (2007) (pages 123, 124).

[37] A. Salles, D. Cavalcanti, A. Acín, D. Pérez-García and M. M. Wolf, "Bell inequalities from multilinear contractions", Quantum Info. Comput. **10**, 703–719 (2010) (pages 123, 128, 130).

[38] M. Ardehali, "Bell inequalities with a magnitude of violation that grows exponentially with the number of particles", Phys. Rev. A **46**, 5375–5378 (1992) (pages 123, 124).

[39] A. V. Belinskiĭ and D. N. Klyshko, "Interference of light and Bell's theorem", Physics-Uspekhi **36**, 653 (1993); N. Gisin and H. Bechmann-Pasquinucci, "Bell inequality, Bell states and maximally entangled states for n qubits", Physics Letters A **246**, 1–6 (1998), pages 123, 124.

[40] M. Żukowski, Č. Brukner, W. Laskowski and M. Wieśniak, "Do all pure entangled states violate Bell's inequalities for correlation functions?", Phys. Rev. Lett. **88**, 210402 (2002) (pages 124, 135).

[41] X.-H. Wu and H.-S. Zong, "Violation of local realism by a system with n spin-$\frac{1}{2}$ particles", Phys. Rev. A **68**, 032102 (2003) (pages 124, 135).

[42] W. Laskowski, T. Paterek, M. Żukowski and Č. Brukner, "Tight multipartite Bell's inequalities involving many measurement settings", Phys. Rev. Lett. **93**, 200401 (2004) (pages 124, 135).

[43] Q. Y. He, E. G. Cavalcanti, M. D. Reid and P. D. Drummond, "Bell inequalities for continuous-variable measurements", Phys. Rev. A **81**, 062106 (2010) (page 124).

[44] Q. Y. He, E. G. Cavalcanti, M. D. Reid and P. D. Drummond, "Testing for multipartite quantum nonlocality using functional Bell inequalities", Phys. Rev. Lett. **103**, 180402 (2009) (page 124).

[45] K.-P. Marzlin and T. A. Osborn, "Continuous-variable Bell inequalities in phase space", arXiv: 1202.2534 (2012) (page 124).

[46] Q. Y. He, P. D. Drummond and M. D. Reid, "Entanglement, EPR steering, and Bell-nonlocality criteria for multipartite higher-spin systems", Phys. Rev. A **83**, 032120 (2011) (page 124).

[47] Q. Y. He, P. D. Drummond, M. K. Olsen and M. D. Reid, "Einstein–Podolsky–Rosen entanglement and steering in two-well Bose–Einstein-condensate ground states", Phys. Rev. A **86**, 023626 (2012); "Einstein-Podolsky-Rosen entanglement strategies in two-well Bose–Einstein condensates", Phys. Rev. Lett. **106**, 120405 (2011), page 124.

[48] M. D. Reid, W. J. Munro and F. De Martini, "Violation of multiparticle Bell inequalities for low- and high-flux parametric amplification using both vacuum and entangled input states", Phys. Rev. A **66**, 033801 (2002) (page 132).

[49] L. Rosales-Zárate, B. Opanchuk, P. D. Drummond and M. D. Reid, "Probabilistic quantum phase-space simulation of Bell violations and their dynamical evolution", Phys. Rev. A **90**, 022109 (2014) (page 132).

[50] P. Walther, M. Aspelmeyer and A. Zeilinger, "Heralded generation of multiphoton entanglement", Phys. Rev. A **75**, 012313 (2007) (page 134).

[51] A. Garg and N. D. Mermin, "Detector inefficiencies in the Einstein–Podolsky–Rosen experiment", Phys. Rev. D **35**, 3831–3835 (1987) (page 134).

Chapter 7

Parallel optimized sampling for stochastic equations

Authors: Bogdan Opanchuk[1], Simon Kiesewetter[1] and Peter D. Drummond[1]

Affiliations: [1]Centre for Quantum-Atom Optics, Swinburne University of Technology, Melbourne, Australia.

To be published in: SIAM Journal on Scientific Computing; accepted 5 October, 2016

7.1 Abstract

Stochastic equations play an important role in computational science, due to their ability to treat a wide variety of complex statistical problems. However, current algorithms are strongly limited by their sampling variance, which scales proportionate to $1/N_S$ for N_S samples. In this paper, we obtain a new class of variance reduction methods for treating stochastic equations, called parallel optimized sampling. The objective of parallel optimized sampling is to reduce the sampling variance in the observables of an ensemble of stochastic trajectories. This is achieved through calculating a finite set of observables — typically statistical moments — in parallel, and minimizing the errors compared to known values. The algorithm is both numerically efficient and unbiased. Importantly, it does not increase the errors in higher order moments, and generally reduces such errors as well. The same procedure is applied both to initial ensembles and to changes in a finite time-step. Results of these methods show that errors in initially optimized moments can be reduced to the machine precision level, typically around 10^{-16} in current hardware.

For nonlinear stochastic equations, sampled moment errors during time-evolution are larger than this, due to error propagation effects. Even so, we provide evidence for error reductions of up to two orders of magnitude in a nonlinear equation example, for low order moments, which is a large practical benefit. The sampling variance typically scales as $1/N_S$, but with the advantage of a very much smaller prefactor than for standard, non-optimized methods.

7.2 Introduction

Stochastic differential equations (SDEs) play a universal and important role in many disciplines requiring quantitative modeling [1–9]. Their practical advantage is that, when used to solve large statistical problems, the ability to randomly sample greatly reduces the complexity of treating the full distribution function. They are employed for treating problems ranging from statistical physics, chemistry and engineering through to economics, biology and financial modeling. As a result of this large field of applications, there is substantial literature on the algorithms used to solve them [10–15], and these are generally different to methods for ordinary differential equations (ODEs).

This utility is not without a price. Such numerical algorithms typically utilize many independent trajectories. Any statistical result will therefore have a sampling variance that scales $1/N_S$ for N_S random samples. This causes to a typical error ϵ_S that scales as $1/\sqrt{N_S}$, giving an error that only decreases slowly with the total computation time. Numerical algorithms generally aim to reduce the truncation error ϵ_T due to the discretization in time, giving an error of order $(\Delta t)^{-p}$ for methods of order p. Yet reducing the truncation error cannot reduce the total error to less than the sampling error. Since the sampling error is often a large part of the total error one has to deal with, reducing the truncation error has a limited effectiveness.

Here we propose a new class of reduced variance SDE algorithms called *parallel optimized sampling* or POS algorithms for short. This approach can provide extremely useful variance reduction compared to independent random sampling methods. The method locally eliminates sampling errors over a finite set of moments, up to numerical precision. Higher order moment errors are reduced also. Global sampling error improvement over a finite time interval is not as large as this, due to error propagation effects, but is still very

substantial. Other variance reduction techniques that can also improve sampling errors include the use of low discrepancy (or quasi-random) sequences [16], and an extrapolated hierarchy of simulations [17], which we do not treat here.

To achieve our results, we introduce a modification of the standard independent stochastic trajectory approach. A set of N_S trajectories is solved *in parallel* using an optimized algorithm. This gives an improvement that depends on the order of the statistical moment being calculated. The present method is a highly practical one, since the run time is still only linear in N_S. This means that the error reductions are of real value. If the run time scaled as N_S^2, and errors were reduced to $1/N_S$, the reductions would not be so useful, since the total computational overhead at fixed error would not change.

There is a close relationship between the present approach and methods of moment hierarchies [5] or cumulant expansions. These existing cumulant based methods have a major drawback. One must close the infinite hierarchy with an arbitrary ansatz at some finite order. If this is incorrect, the method will fail [18], and several examples exist for this problem. In fact, such cumulant hierarchies cannot in general be truncated consistently [19]. By contrast, our method requires no ansatz; it gracefully reduces in a consistent way to the usual, unbiased stochastic method at large sample sizes. It is a unification of moment hierarchy and stochastic methods.

We focus mostly on simple one-dimensional cases here for simplicity, except for one higher-dimensional example to illustrate the approach. Extensions of the POS method to cases of higher-order convergence in time will be treated elsewhere. However, as one example of potential applications, sampling errors can play a major role in quantum simulations using phase-space representations. The difficulty is that distribution tails may only be weakly bounded [20]. In such cases, sampling error can determine whether a quantum simulation is feasible or not. One approach to solving this problem is using weighted trajectory methods [21]. Another method, which was successfully used in treating long-range order in the fermionic Hubble model, is to impose global conservation laws [22]. The present approach has a more general applicability than either of these, but can also be combined with such earlier techniques.

The organization of the paper is as follows. In Section 7.3, we analyze the stochastic errors that can arise in computational methods for stochastic differential equations. In

Section 7.4 sampling errors are reduced in initial observables by introducing a new type of stochastic noise generator. Numerical examples of this are treated in Section 7.5, showing that variance in selected moments can be reduced to machine precision. Next, in Section 7.6, this approach is extended to dynamical optimization of stochastic equation algorithms, with numerical examples given in the remaining sections. In Section 7.7 we start by treating single-step optimization, which ignores error-propagation. In Section 7.8, we demonstrate the global, or multi-step performance of these methods for a number of one-dimensional linear and nonlinear cases, showing that global sampling error variance is substantially reduced compared to traditional independent trajectory approaches, typically by at least an order of magnitude. Finally, Section 7.9 treats a two-dimensional example which has very similar behaviour. Section 7.10 gives our summary and conclusions.

7.3 Stochastic integration error

Stochastic integration involves both truncation errors due to finite step-size, and errors due to finite samples. In this section we analyze the performance of independent SDE methods by considering the overall resources needed to solve the equations for some given error. To explain the background to our approach, we show that increasing the convergence order is a less effective way to reduce errors than one might expect, because of a trade-off between the resources needed to reduce both truncation and sampling errors. As large ensembles are required to reduce sampling errors, we find that higher-order methods for SDEs, while undoubtedly useful, are not as effective as in ODE integration.

In the rest of this paper, we show how to overcome this problem by using parallel algorithms that reduce sampling errors.

7.3.1 Distributions and observables

We start by establishing our notation. The observables of a probability distribution function $P(t, \mathbf{x})$ are the fundamental objects of interest in stochastic calculations. These are probability-weighted integrals of functions $o_m(\mathbf{x})$ over $\mathbf{x}$:

$$\langle o_m \rangle = \int \mathrm{d}\mathbf{x} o_m(\mathbf{x}) P(t, \mathbf{x}) . \tag{7.1}$$

In general, $\mathrm{d}\mathbf{x}$ is a Euclidean measure over a d-dimensional real or complex space of variables $\mathbf{x}$. The numerical examples given here will focus mainly on moments of one-dimensional distributions with $o_m(x) = x^m$, giving a vector whose error should be minimized:

$$\mathbf{o} = \{o_1, \dots, o_M\}. \tag{7.2}$$

With sampling methods, the time-evolution is solved by generating a number N_S of independent sample paths $\mathbf{x}^{(n)}(t)$. This procedure is equivalent to approximating the true distribution by the sampled distribution defined as

$$P_S(t, \mathbf{x}) = \frac{1}{N_S} \sum_{n=1}^{N_S} \delta\left(\mathbf{x} - \mathbf{x}^{(n)}(t)\right). \tag{7.3}$$

We note that the initial value problem consists of defining an initial static distribution $P(\mathbf{x})$ at $t = t_0$, and then sampling it with initial points $\mathbf{x}^{(m)}(t_0)$. For dynamics, the sampled paths are defined at N_T discrete times t_i, with a spacing Δt, such that exact ensemble average values are recovered in the double limit of infinite sample and zero step-size:

$$\langle o_m(t_i)\rangle = \lim_{\Delta t \to 0} \lim_{N_S \to \infty} \frac{1}{N_S} \sum_{n=1}^{N_S} o_m\left(\mathbf{x}^{(n)}(t_i)\right). \tag{7.4}$$

These paths are obtained by numerical algorithms which propagate $\mathbf{x}^{(n)}(t_i)$ to $\mathbf{x}^{(n)}(t_{i+1})$, with a truncation error vanishing as $\Delta t \to 0$. If we call N_S the sample number, the phase-space is made up of a set of distinct samples or trajectories. These form an extended vector of parallel trajectories,

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}^{(1)T} & \dots & \mathbf{x}^{(N_S)T} \end{pmatrix}^T, \tag{7.5}$$

of size $N_S d$. It is convenient to define the sampled observables as a function $\bar{\mathbf{o}}$ of the extended vector of sampled trajectories,

$$\bar{o}_m(\mathbf{X}) = \frac{1}{N_S} \sum_{n=1}^{N_S} o_m\left(\mathbf{x}^{(n)}\right). \tag{7.6}$$

Throughout this paper, we will use the standard notation that all extended vectors or matrices involving a dimension of the order of the sample size N_S are written in upper

case like $\mathbf{X}$, while objects that do not depend on the sample size are written in lower case.

7.3.2 Fokker-Planck and stochastic equations

To explain the problem of interest here, we give a detailed analysis of a stochastic differential equation or SDE. The underlying distribution function $P(t, \mathbf{x})$ satisfies the Fokker-Planck equation [6],

$$\frac{\partial P(t,\mathbf{x})}{\partial t} = \left[-\sum_i \frac{\partial}{\partial x_i} a_i(\mathbf{x}) + \frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j} \mathsf{d}_{ij}(\mathbf{x}) \right] P(t,\mathbf{x}) \ . \tag{7.7}$$

We wish to minimize the errors for solving the corresponding Itō SDE [4, 6], given in a standard form by:

$$\mathrm{d}\mathbf{x} = \mathbf{a}(\mathbf{x})\,\mathrm{d}t + \mathbf{b}(\mathbf{x})\,\mathrm{d}\mathbf{w}, \tag{7.8}$$

where $\mathbf{d} = \mathbf{b}\mathbf{b}^T$, and the Gaussian distributed real noise $\mathrm{d}\mathbf{w}$ has correlations given by:

$$\begin{aligned} \left\langle \mathrm{d}w_i \mathrm{d}w_j \right\rangle_\infty &= \delta_{ij}\mathrm{d}t \\ \left\langle \mathrm{d}w_i \right\rangle_\infty &= 0. \end{aligned} \tag{7.9}$$

Here the $\langle \dots \rangle$ notation means ensemble average, so that

$$\langle \mathbf{x} \rangle_{N_S} \equiv \frac{1}{N_S} \sum_{n=1}^{N_S} x^{(n)}, \tag{7.10}$$

with $\langle \dots \rangle_\infty$ being the infinite ensemble limit.

7.3.3 Sampling and error criteria

To evaluate the computational error, we must define an error criterion relevant to the entire sample ensemble $\mathbf{X}$. In practical terms, this means one must calculate errors in each required observable, rather than just the errors in a single trajectory. By minimizing the total error over a finite number of samples, one should be able to arrive at the most efficient algorithm that utilizes a given computational resource.

To quantify the error we will use a weighted norm $\|P\|_W = \sqrt{(P,P)} \geq 0$. The integration error of a sampled, calculated stochastic distribution P_S relative to the exact

distribution P_E is then defined as

$$\epsilon = \|P_S - P_E\|_W, \tag{7.11}$$

which depends on the choice of weighted norm.

The computed averages, $\langle o_m \rangle_{P_S}$, are obtained through sampling, which means that typically one is interested in calculating specific observables with minimum error. To evaluate the computational accuracy relevant to these observables of interest, we choose a particular set of observable quantities, o_m, for $m = 1, \ldots M$, to define the distribution norm, so that:

$$\|P\|_W^2 \equiv \sum_{m=1}^{M} W_m \left|\langle o_m \rangle_P\right|^2. \tag{7.12}$$

Here W_m is the relative weight assigned to observable m. We assume the error is evaluated at a fixed time t, otherwise one may wish to minimize the errors at each of a set of sample times. In most of this paper, the observables are a finite set of one-dimensional moments, $o_m \equiv x^m$, although other measures are certainly possible. For simplicity, we choose $W_m = 1$ from now on.

With this definition, the *total* error is

$$\epsilon = \sqrt{\sum_{m=1}^{M} \left|\langle o_m \rangle_{P_S} - \langle o_m \rangle_{P_E}\right|^2}, \tag{7.13}$$

The integration error ϵ clearly must be calculated from the entire vector of stochastic trajectories $\mathbf{x}$. The task of defining an optimal stochastic integration method is to obtain $\mathbf{x}(t)$ with a procedure that minimizes the total error ϵ relative to some computational resource. Using the definition that

$$\mu_m = \langle o_m \rangle_{P_E}, \tag{7.14}$$

we see that:

$$\epsilon = \sqrt{\sum_{m=1}^{M} |\bar{o}_m(\mathbf{X}) - \mu_m|^2}. \tag{7.15}$$

Although different to the usual probability norms, we will use the above definition throughout this paper, since it corresponds to the operational requirements of a norm

for a sampled distribution. For a single observable, say x, ϵ^2 is simply the usual sample variance in x — provided there are no other errors present as well. However, there are usually other errors, including errors due to arithmetic roundoff and the finite step-size of the integration algorithm.

7.3.4 Optimizing the sample number

What is the optimal strategy to solve this SDE, given fixed total computational resources and a goal of minimizing the total error? With independent paths, the computational resource utilized is proportional to $N = N_S N_T$, where N_S is the number of samples and N_T is the number of time-steps. One can invest total processing time either in reducing the time-step, or in increasing the number of samples, but reducing the step-size means using less samples, and vice-versa.

To understand these issues quantitatively, we first consider the optimal strategy for error-reduction with traditional methods that involve using independent sample paths. We will show that it is only important to reduce the step-size or discretization error to the point where it is some fraction of the sampling error.

Consider the total error for N_T time-steps with a stochastic integration method of global order p, that utilizes N_S *independent* sample paths. From the central limit theorem, the global sampling error ϵ_S due to the use of a finite sample, is

$$\epsilon_S = \sigma N_S^{-1/2}, \tag{7.16}$$

where σ is the standard deviation of the measured quantity, and the same type of scaling holds for sums over errors in multiple observables.

There are additional errors ϵ_T over a fixed time interval $T = N_T \Delta t$, due to the finite time-step Δt. If the local error of one time-step is $\epsilon_{\Delta t}$, then at best $\epsilon_T \approx N_T \epsilon_{\Delta t}$. The discretization order p is therefore defined so that the global time-step truncation error ϵ_T scales as

$$\epsilon_T = c N_T^{-p} \propto \Delta t^p, \tag{7.17}$$

where c is an algorithm-dependent constant that depends on the set of computed observables.

Since the two error sources are independent, the overall error is given approximately by a sum of two terms:

$$\epsilon = \epsilon_T + \epsilon_S = c{N_T}^{-p} + \sigma N_S^{-1/2}. \tag{7.18}$$

We wish to constrain the total processor time — parallel or otherwise — so that $N_S N_T = N$ is bounded. The total CPU time is then $T_{\text{CPU}} = N\tau$, if one step in time takes a real time duration τ, and other overheads are negligible.

Next, consider how to optimize the tradeoff between the truncation and sampling errors. Letting $\tilde{c} = cN^{-p}$, and keeping N fixed, one obtains:

$$\epsilon = \tilde{c} N_S^p + \sigma N_S^{-1/2} \tag{7.19}$$

This is minimized by choosing a combination of time-step and sample number such that:

$$\begin{aligned} N_S &= N^{\frac{2p}{2p+1}} \left[\frac{\sigma}{2pc}\right]^{\frac{2}{2p+1}} \\ N_T &= N^{\frac{1}{2p+1}} \left[\frac{\sigma}{2pc}\right]^{-\frac{2}{2p+1}} \end{aligned} \tag{7.20}$$

The total error is then:

$$\epsilon = \left[\frac{c\sigma^{2p}}{2pN^p}\right]^{\frac{1}{2p+1}} \left[1 + \frac{1}{2p}\right] \tag{7.21}$$

7.3.5 Effective integration order

From the analysis above, the best that one can do to reduce errors is to obtain a scaling of $\epsilon \propto N^{-p_{\text{eff}}}$, where the effective order p_{eff} is

$$p_{\text{eff}} = \frac{p}{2p+1}. \tag{7.22}$$

This is because a higher order integration technique has no effect on the sampling error, which eventually dominates the error calculation. The best effective order, even with $p = \infty$, is $p_{\text{eff}} = 1/2$. Given that stochastic higher order algorithms are complex and slow, higher p values are not always an advantage.

This optimal effective order requires the use of an optimal ratio of samples to steps, which is:

$$\frac{N_S}{N_T} = N^{\frac{2p-1}{2p+1}} \left[\frac{\sigma}{2pc} \right]^{\frac{4}{2p+1}} . \tag{7.23}$$

This ratio is difficult to calculate a-priori, since these constants are not usually known in advance. One can estimate the sampling error numerically by measuring the sample standard deviation σ, which is known from the statistics of the simulation. The truncation error constant c is also measurable by changing the step-size. Thus, both c and σ can be measured in software. An optimum point is reached when the ratio between discretization and sampling error is given by the extremely simple result that:

$$\frac{\epsilon_T}{\epsilon_S} = \frac{1}{2p} . \tag{7.24}$$

In summary, it is only productive to reduce the step-size to the point where the discretization error of $1/2p$ of the sampling error. After this point is reached, one is better off to use more samples. Another way to describe this is that the least computational cost for error ϵ is

$$N \propto \epsilon^{-1/p_{\text{eff}}} = \epsilon^{-(2p+1)/p}. \tag{7.25}$$

There are multi-scale methods [17] that can achieve a cost of $N \propto \epsilon^{-2} (\log \epsilon)^2$, which is a useful further improvement.

In summary, from the traditional analyses of ODE integration, one might expect that the error would be reduced extremely rapidly with a higher order method. This is not the case with an SDE. Because the independent sampling error varies slowly with resource, this becomes the dominant error with increasing resources.

Our conclusion is that sampling error strongly limits the effectiveness of independent sampling methods for SDEs.

7.4 Initial conditions for parallel optimized sampling

We now describe a family of variance reduction methods using parallel optimized sampling (POS), which locally eliminate sampling errors for a finite set of observables $\mathbf{o}$, thus improving computational efficiency. As described above, for N_S samples, there is a $N_S d$-dimensional extended vector of stochastic variables, $\mathbf{X} = \begin{pmatrix} \mathbf{x}^{(1)T} & \cdots & \mathbf{x}^{(N_S)T} \end{pmatrix}^T$,

which is optimized so that $\mathbf{X} \to \mathbf{X}^{(\mathrm{opt})}$. Both initial samples and the resulting stochastic trajectories are optimized in parallel, thus making use of increasingly parallel hardware capabilities.

Before deriving the POS algorithm, we need to address an issue arising from the known initial condition for the SDE, with $\mathbf{x}$ having a distribution $P(\mathbf{x})$ at $t = t_0$. When we derive the optimization conditions, one of the assumptions will be that the moments of $\mathbf{x}$ are already optimized at the *start* of each time step, that is $\langle o_m(t)\rangle_{N_S} = \mu_m(t)$ for $m = 1 \ldots M$. At every step of the SDE integration, if we obtain the optimized stochastic vector from the previous step, then, since we optimize the differentials of the moments, the resulting stochastic variables will be optimized as well as possible, apart from any time-step errors, which is a separate issue.

But how do we ensure the *initial* set of moments of the distribution $P(\mathbf{x})$ is optimized at the first integration step?

7.4.1 Static POS

We emphasize that our main goal is to minimize the variance for the dynamical stochastic equations, treated next. As a consequence, we do not use methods like Gaussian quadratures for the static variance reduction problem. These are limited in the diversity of initial distributions treated, and they cannot be easily generalized to the dynamic case. Instead we take a set of random samples from $P(\mathbf{x})$, and optimize these initial sampled variables to reduce the variance in moments of interest.

We call this static POS, as it does not involve dynamical time-evolution. In other words, we wish to sample the initial values for the trajectories **so that M chosen observables are equal to their known exact values**, that is,

$$\bar{\mathbf{o}}(\mathbf{X}) = \boldsymbol{\mu} \equiv \langle \mathbf{o} \rangle_\infty, \tag{7.26}$$

where $\boldsymbol{\mu}$ is an M-dimensional vector. That is, we wish to set $\epsilon_P = \epsilon_P(t_0) = 0$, where ϵ_P is the total error (7.15) in the initial distribution, using an observable-based error measure. In the examples given in this paper, we treat all moments up to a given order. More generally, at high dimension d, treating all correlations or moments to a fixed

order is exponentially hard. Therefore, it is necessary to select moments according some importance criterion, the effects of which will be treated in a later publication.

This is a classic example of nonlinear, multidimensional root-finding, with optimization over the dimensionality of the full space of samples. It can be a challenge to solve these equations efficiently, with resources no greater than $\mathcal{O}(N_S)$, and with minimal changes to the original sampled estimate for $\mathbf{X}(t_0)$, which is labelled $\mathbf{X}_{(0)}$. However, in examples treated here, we find that the numerical problem can be very accurately solved, naturally with machine-limited numerical precision. Here we use an iterative, modified Newton-Raphson method for the solution. Other techniques are available as well.

In an iterative Newton-Raphson approach, we use stochastic methods for the first estimate, $\mathbf{X}_{(0)}$. At each iteration, we set $\mathbf{X} \to \mathbf{X} + \delta\mathbf{X}$ to obtain the next estimate $\mathbf{X}_{(i+1)}$, which has a better fit to the required moments, so that $\mathbf{X}_{(i+1)} = \mathbf{X}_{(i)} + \delta\mathbf{X}_{(i+1)}$, where $\|\delta\mathbf{X}_{(i+1)}\| \ll \|_{(i)}\|$, and $\| \ldots \|$ denotes the Euclidean norm.

Expanding to first order in $\delta\mathbf{X}_{(i+1)}$, the conditions we require are:

$$\begin{aligned} \bar{\mathbf{o}}^{(\mathrm{opt})} &= \boldsymbol{\mu} \\ &= \bar{\mathbf{o}}\left(\mathbf{X}_{(i)}\right) + \mathbf{J}\left(\mathbf{X}_{(i)}\right)\delta\mathbf{X}_{(i+1)} + \mathcal{O}\left(\delta\mathbf{X}^2_{(i+1)}\right), \end{aligned} \tag{7.27}$$

where $\mathbf{J}$ is a Jacobian of $\bar{\mathbf{o}}$:

$$\mathrm{J}_{mn}\left(\mathbf{X}\right) = \frac{\partial \bar{o}_m\left(\mathbf{X}\right)}{\partial X_n}. \tag{7.28}$$

In matrix form, the iteration equation is therefore

$$\mathbf{J}_{(i)}\delta\mathbf{X}_{(i+1)} = \boldsymbol{\mu} - \bar{\mathbf{o}}_{(i)}. \tag{7.29}$$

where we use the obvious notation that $\bar{\mathbf{o}}_{(i)} \equiv \bar{\mathbf{o}}\left(\mathbf{X}_{(i)}\right)$ and $\mathbf{J}_{(i)} \equiv \mathbf{J}\left(\mathbf{X}_{(i)}\right)$. We note that out method is scale-invariant in the sense that it will result in the exact same solutions if we rescale any observable and its exact value as $o_i(\mathbf{X}) \to c_i o_i(\mathbf{X})$, where c_i is a constant.

Each step of the iterative solution is underdetermined, since the relevant matrices are not square, and hence not invertible. Such problems are relatively common in optimization, and here we analyze two different techniques to solve them.

7.4.2 One-dimensional example of iteration equations

We will use as an illustrative example the situation when the observables are moments of a one-dimensional stochastic equation. In this example $o_m = x^m$, so the conditions are:

$$\begin{aligned} \left[\mathbf{X}^{(\mathrm{opt})}\right]^m \cdot \mathbf{1} &= N_S \mu_m \\ &= \mathbf{X}_{(i)}^m \cdot \mathbf{1} + m\mathbf{X}_{(i)}^{m-1} \cdot \delta\mathbf{X}_{(i+1)} + \mathcal{O}\left(\delta\mathbf{X}_{(i+1)}^2\right), \end{aligned} \tag{7.30}$$

where the powers of matrices and vectors are understood in the element-wise (Hadamard) sense (e.g., $\mathbf{X}^2 \equiv \mathbf{X} \circ \mathbf{X}$). Hence, in this case the observable vector and matrix of derivatives are:

$$\begin{aligned} \bar{\mathbf{o}}\left(\mathbf{X}\right) &= \frac{1}{N_S} \sum_{n=1}^{N_S} \begin{pmatrix} x^{(n)} \\ \cdots \\ [x^{(n)}]^M \end{pmatrix}, \\ \mathbf{J}\left(\mathbf{X}\right) &= \frac{1}{N_S} \begin{pmatrix} 1 & \cdots & 1 \\ & \cdots & \\ M\left[x^{(1)}\right]^{M-1} & \cdots & M\left[x^{(N_S)}\right]^{M-1} \end{pmatrix}. \end{aligned} \tag{7.31}$$

7.4.3 Static iteration solution with a matrix pseudo-inverse

The simplest method to solve the iteration equation (7.29) is to use a matrix pseudo-inverse. This has another advantage, since the pseudo-inverse has the property that it generates a least-squares solution, with minimal changes, as required for this algorithm.

Thus, we have the iterative procedure

$$\mathbf{X}_{(i+1)} = \mathbf{X}_{(i)} + \mathbf{J}_{(i)}^{+}\left(\boldsymbol{\mu} - \bar{\mathbf{o}}_{(i)}\right). \tag{7.32}$$

where $\mathbf{J}^{+}$ indicates the pseudo-inverse of the $M \times N_S$ Jacobian $\mathbf{J}$.

This pseudo-inverse method is simple, general and effective. It satisfies our requirements given above. However, standard pseudo-inverse software using singular value decomposition (SVD) methods has drawbacks. It is relatively slow, can have stability problems and may not be available on all programming platforms.

7.4.4 Improved matrix inversion iterations

To improve efficiency and overcome problems with the use of general purpose pseudo-inverses, there is a faster alternative method for solving (7.29).

The pseudo-inverse is a limit. If $\mathbf{J}^\dagger$ denotes the conjugate transpose of the matrix $\mathbf{J}$, then:

$$\mathbf{J}^+ \equiv \lim_{\alpha \to 0} \mathbf{J}^\dagger \left(\mathbf{J}\mathbf{J}^\dagger + \alpha \mathbf{I} \right)^{-1} . \tag{7.33}$$

Hence, provided $\mathbf{u} = \mathbf{J}\mathbf{J}^\dagger$ is invertible, one can directly obtain that $\mathbf{J}^+ = \mathbf{J}^\dagger \mathbf{u}^{-1}$. It is important to note here that $\mathbf{u}$ is only an $M \times M$ matrix, regardless of how many samples were used originally.

This gives as many unknown coefficients as there are equations. As a result, we only need solve M linear equations in M unknowns, for which there are efficient techniques, resulting in the iteration procedure

$$\mathbf{X}_{(i+1)} = \mathbf{X}_{(i)} + \mathbf{J}^\dagger_{(i)} \mathbf{u}^{-1}_{(i)} \left(\boldsymbol{\mu} - \bar{\mathbf{o}}_{(i)} \right) . \tag{7.34}$$

Here the matrix $\mathbf{u}$ being inverted is an $M \times M$ positive-semidefinite matrix. We use the lower case for it to indicate that it does not take an extended dimensionality. It is generally invertible and relatively small compared to the sample size, resulting in a highly efficient optimization algorithm. Numerical investigations show that this simple method is convergent for large sample number N_S, which corresponds to cases where the initial sample has moments close to their ideal values.

As the stopping condition in the numerical examples given later, we choose the condition that the changes are small relative to the modulus of the current $\mathbf{X}$ value, i.e. $\|\delta \mathbf{X}_{(i+1)}\| / \|\mathbf{X}_{(i)}\| < \eta = 10^{-8}$, where η^2 is the numerical machine precision, as previously. An iteration limit of $I_{\max} = 50$ was used, although as we see later, the typical iteration number is much lower than this.

We note that this simplified method is only useful when $\mathbf{u}$ is invertible, otherwise the full SVD techniques or other more robust algorithms should be used. This is not a major limitation: as long as the first derivatives of our observables are linearly independent, $\mathbf{u}$ will be invertible. As often the case with such numerical methods, performance can be further improved through linear scaling or translation operations.

However, sometimes the Newton-Raphson method doesn't converge, not because $\mathbf{u}$ lacks invertibility (which is rare), but because the initial guess is too far away from the solution. This can be detected by checking if the norm of $\boldsymbol{\Delta}$ increases. In such cases, one can back-track and try again with a different initial estimate. In the case of time-evolution dynamics, treated next, a smaller time-step can also be used to ensure the initial guess is close to hte solution.

7.4.5 One-dimensional case: power series method

As an equivalent way to understand our approach in the one-dimensional case, we can expand $\delta\mathbf{X}$, which is the iterative **change** in $\mathbf{X}$, using the derivative matrix and a vector of changes in observables, $\delta\mathbf{o}$. With observables as moments, this is a series in $\mathbf{X}^{\dagger}$ up to order $M-1$, with M coefficients written as a vector $\boldsymbol{\delta}$. The expansion is given by:

$$\delta\mathbf{X} = \mathbf{J}^{\dagger}\delta\bar{\mathbf{o}} = \frac{1}{N_S}\sum_{m=1}^{M} m\left(\mathbf{X}^{\dagger}\right)^{m-1}\delta\bar{o}_m. \tag{7.35}$$

Defining a square, positive definite matrix:

$$\mathbf{u}_{(i)} = \mathbf{J}_{(i)}\mathbf{J}^{\dagger}_{(i)}, \tag{7.36}$$

the basic iteration equation (7.29) becomes:

$$\mathbf{u}_{(i)}\delta\bar{\mathbf{o}}_{(i+1)} = \boldsymbol{\mu} - \bar{\mathbf{o}}_{(i)}. \tag{7.37}$$

Since the solution for $\boldsymbol{\delta}_{(i+1)}$ requires the inversion of $\mathbf{u}_{(i)}$, this *leads to the same final algorithm* that is defined above, in Eq (7.34).

7.5 Numerical examples: Static optimization

In this section, we perform numerical optimizations of initial conditions, following the methods of the previous section. This is important because the initial sampling error can dominate the sampling error found throughout the calculation, if it is not optimized. For definiteness, in most of the numerical examples of POS algorithms analyzed here, the

m-th observable is a finite moment of a one-dimensional real distribution. Hence,

$$\bar{o}_m \equiv \langle o_m \rangle_{N_S} = \frac{1}{N_S} \sum_{n=1}^{N_S} \left[x^{(n)} \right]^m = \frac{\mathbf{X}^m \cdot \mathbf{1}}{N_S}, \tag{7.38}$$

where $\mathbf{1}$ is an N_S-dimensional unit column vector. We optimize moments, for $m = 1, \dots M$. As well as calculating the resulting error in these optimized moments, we also obtain the errors in other moments. An important issue is that the non-optimized moments should not have increased errors. We will also consider the ensuing results for two less regular observables: $o_{\exp}(x) = \exp(x)$ (a combination of an infinite number of moments) and $o_{||}(x) = |x|$ (an irregular function).

7.5.1 Static optimization

In this section we will test the performance of the iterative algorithm from Section 7.4 applied to the most common initial condition — a normal distribution. However, we emphasize that this choice is purely for convenience, and any initial distribution could be used. In this section we will use $\mu = \mu_1 = 0$ for all the tests and set the variance $\sigma^2 = \mu_2 = 1$. This helps to avoid precision loss when calculating central moments, which is important, since the method is extremely efficient and optimizes the moments almost to the limit of numerical precision, as seen below. If maximum precision is required for cases when $\mu \neq 0$, one can subtract the mean of the initial distribution so that $\mu = 0$ in the new variables.

With $\mu = 0$, the moments of the test distribution are given by

$$\mu_m = \begin{cases} 0, & m \text{ is odd}, \\ \sigma^m \, (m-1)!!, & m \text{ is even}. \end{cases} \tag{7.39}$$

The number of moments being optimized here is $M = 6$, but this is not essential. For the additional observables the expected values are the expectation of the log-normal distribution

$$\mu_{\exp} = \exp\left(\frac{\sigma^2}{2}\right), \tag{7.40}$$

and the expectation of the folded normal distribution

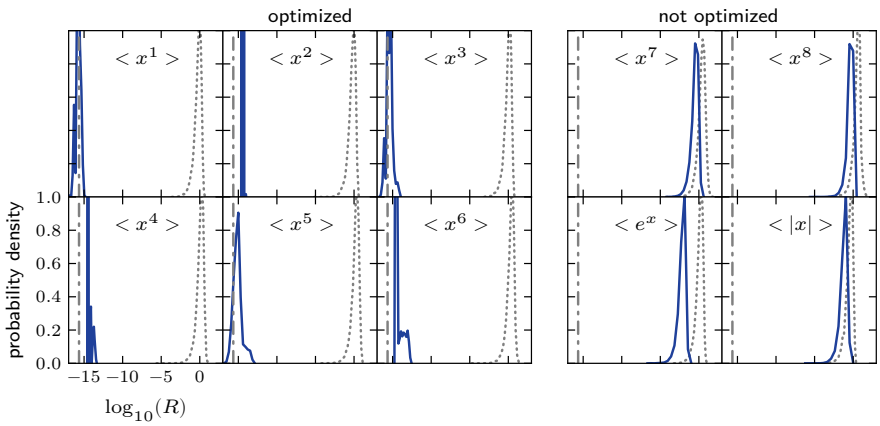

Figure 7.1: Distribution of normalized errors in the static optimization tests for the matrix inversion method with $\sigma = 1$, $N_S = 1000$ (solid blue lines). The grey dotted curves show the normalized errors of a randomly sampled vector. The dash-dotted line denotes the numerical precision limit ($\epsilon_{\text{num}} \approx 2.22 \times 10^{-16}$ for double precision floating point numbers). Zero error results are removed from binning for ease of plotting.

$$\mu_{\|} = \sigma\sqrt{\frac{2}{\pi}}. \tag{7.41}$$

We test the behavior of the algorithms in 10000 independent optimization attempts, each with $N_S = 1000$ parallel samples, in order to collect statistics. In each attempt, a N_S-dimensional vector $\mathbf{X}$ is generated using random sampling and the matrix inversion iterative method from Section 7.4.4 is applied to it. We then plot the distribution of measures of interest, which indicate the performance of the method. This demonstrates how well the algorithm performs when tested over a range of different initial random samples, each drawn from the same underlying normal distribution.

7.5.2 Distribution of the optimized variance reduction

There are several quantities we are interested in. We first investigate the distribution of the optimized variance. This gives an idea of not just how much the variance is reduced, but also the probability of a given amount of variance reduction.

In more detail, we plot the numerically obtained distribution of final central moments of the vector compared to the desired values:

$$\tilde{R}_m(\mathbf{X}) = \left| \frac{1}{N_S} o(\mathbf{X}) \cdot \mathbf{1} - \mu_m \right|, \tag{7.42}$$

where we will compare this quantity for initial randomly sampled vectors and optimized vectors. It is convenient to scale this to the expectation of the error in the sampled moment (for random samples) for the normal distribution:

$$R_m = \frac{\tilde{R}_m}{\sigma^m \sqrt{m!/N_S}}, \tag{7.43}$$

so the average of R_m for the initial randomly sampled vectors will be close to 1 for all m. Fig. 7.1 shows this quantity both for the moments we are optimizing ($m = 1 \dots M$) and for two higher moments that are not optimized. The errors for the additional observables are normalized simply as

$$R_{\text{exp},\|} = \frac{\tilde{R}_{\text{exp},\|}}{\sqrt{N_S}}. \tag{7.44}$$

The optimization procedure was able to reduce errors almost to the limit of numerical precision of $R_m \sim 10^{-15}$ in many cases. The moments that were not directly optimized show some improvement as well, but typically closer to a single order of magnitude. Since we use logarithmic graphs, and R_m can sometimes be zero, we excluded these zero results from the binning.

7.5.3 Number of iterations and optimization distance

It is also useful to know the number of iterations performed in each of the algorithms. This gives an idea of the numerical cost, which we investigate in greater detail below. If too many iterations are needed, the time taken will increase to the point that one may as well use more traditional methods.

Fig. 7.2(a) demonstrates that only a small number of iterations — typically four — are necessary for the method to converge, given a reasonably large ensemble. The third quantity of interest is the relative distance of the optimized vector from the initial randomly sampled one:

$$D = \frac{\|\mathbf{X}^{(\text{opt})} - \mathbf{X}\|}{\|\mathbf{X}\|}. \tag{7.45}$$

We would like this to be much less than 1, which means that the optimized vector is still approximately "random". Fig. 7.2(b) shows that it is indeed the case, and the average deviation is $D \approx 5 \times 10^{-3}$ for $\sigma = 1$. Fig. 7.2(c) shows that the results have the desirable

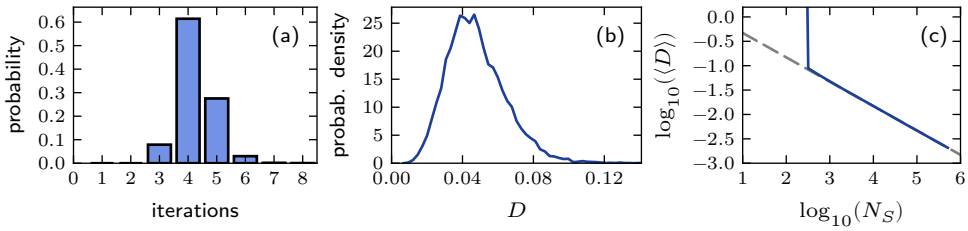

Figure 7.2: (a) Number of iterations taken by the matrix inversion method, $\sigma = 1$, $N_S = 1000$. (b) Relative distance D between the optimized vector and the initial randomly sampled vector in the static optimization tests for the matrix inversion method, $\sigma = 1$, $N_S = 1000$. (c) Average relative distance $\langle D \rangle$ depending on the number of trajectories N_S in the static optimization tests for the matrix inversion method, $\sigma = 1$. The grey dashed line denotes the slope of a $1/\sqrt{N_S}$ dependence.

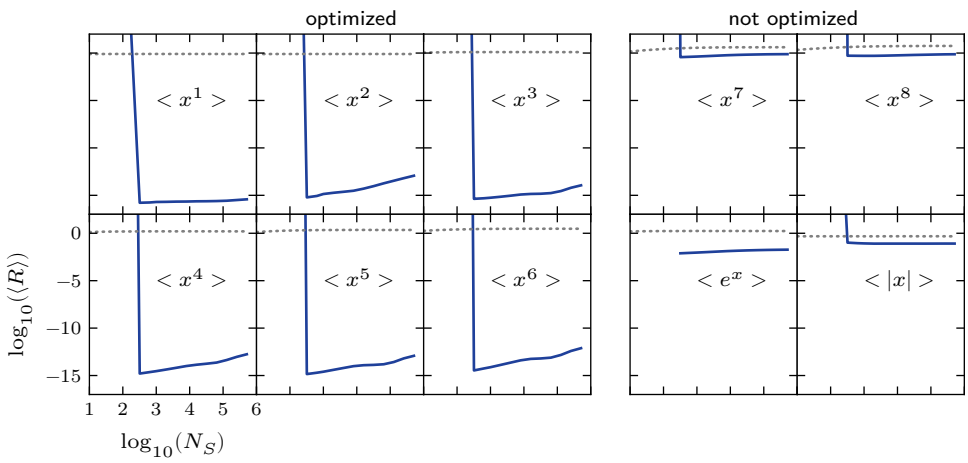

Figure 7.3: Dependence of normalized errors on the number of trajectories N_S in the static optimization tests for the expansion method, $\sigma = 1$. The dotted curve shows the dependence for randomly sampled vectors.

property that the relative distance between the optimized and non-optimized trajectories scales with $1/\sqrt{N_S}$. This is expected from the fact that the initial sampled moment relative errors scale as $1/\sqrt{N_S}$, hence one expects the minimum required change to be of this order.

However, if there are too few trajectories, the simple matrix inversion iteration method diverges, producing unacceptably large changes D in the distribution. For these small N_S values, the iteration limit of $I_{\max} = 50$ was reached, indicating a lack of convergence.

To investigate this more carefully, the dependence of the final error R on the number of trajectories N_S, is plotted in Fig. 7.3. The results also display a sharp cut-off in the number of trajectories below which the iterative approach does not converge. More robust minimization algorithms could presumably solve this, but the issue was not investigated here for space reasons.

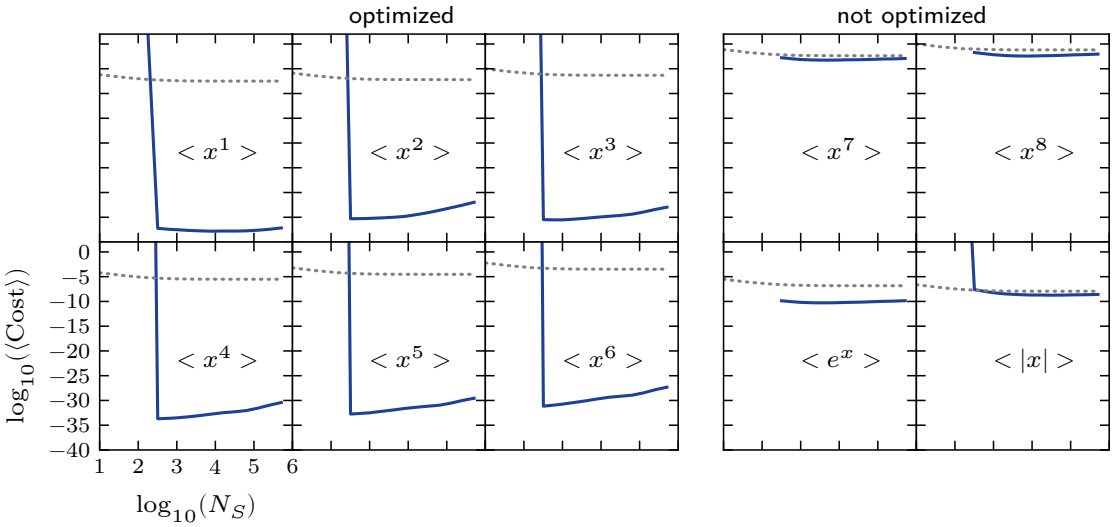

Figure 7.4: Dependence of the cost on the number of trajectories N_S in the static optimization tests for the expansion method, $\sigma = 1$. The dotted curve shows the dependence for randomly sampled vectors.

7.5.4 Initial optimization cost

For practical applications, the time taken for the optimization is also important. More accurately, the quantity of interest is the efficiency [23], which in our case can be expressed as $\mathrm{Eff} = 1/\left(\mathrm{T}_{\mathrm{CPU}}\tilde{\mathrm{R}}^2\right)$, where T_{CPU} is the time required to generate the vector $\mathbf{X}^{(\mathrm{opt})}$.

Since in our case in a significant fraction of optimization attempts $\tilde{R}$ for optimized moments is exactly zero, we plot the cost

$$\mathrm{Cost} = \frac{1}{\mathrm{Eff}} \equiv T_{\mathrm{CPU}}\tilde{R}^2. \tag{7.46}$$

averaged over attempts. Fig. 7.4 shows this quantity for all investigated observables. For the optimized moments ($m = 1 \ldots 6$) the cost of POS is significantly lower than the cost of the purely random sampling (starting from a certain critical number of trajectories). The situation is different for non-optimized observables, where POS still demonstrates a decreased cost, but the difference is not as overwhelming.

In conclusion, the expansion iterative method performs well for a wide range of parameters, while being fast and easy to implement. Convergence is extremely rapid when the number of stochastic trajectories is increased above a critical value. However, we note that many other methods of nonlinear root-finding exist, and we do not exclude the possibility of better techniques being available.

We use this approach in the SDE examples given below, to pre-optimize the starting conditions for the SDE integration tests.

7.6 Parallel dynamic optimization

The general approach here is to assume that samples with an optimized set of observables are known at a given time t. Given this, a noise ensemble is chosen to give the lowest possible errors in the observables at a later time $t + \Delta t$. Overall convergence is then achieved by taking a large N_S limit. The goal of the algorithm is that the observable moments are obtained with minimal error. For efficiency, the resulting calculation complexity should be linear in N_S.

In summary, the POS algorithm must therefore meet three essential requirements:

1. Equations for a finite set of M observables are optimized both initially, and for local changes in time to a given order in Δt.

2. Non-optimized observables still have their correct values to the same order in Δt in the limit of $N_S \to \infty$.

3. The computational cost of the algorithm should be no worse than $\mathcal{O}(N_S)$.

If the observables chosen are moments, then, since higher order moments depend on lower order ones, these will experience a degree of optimization as well, even though not directly optimized. POS methods can also be combined with higher-order time-step methods, but here we focus on sampling error reduction for ease of explanation. Other types of observable also experience a degree of variance reduction as well. Typically this is not as great as experienced by the optimized moments, but this depends on the observable and the stochastic equation.

There is an analogy between this approach and the method of moment hierarchies in statistical physics [5]. The difference is that rather than an arbitrary truncation of the moment hierarchy, the higher order moments are estimated in an unbiased way via sampling.

We now describe how to extend this initial optimization algorithm to treat dynamical optimization of moments during stochastic time evolution of $\mathbf{x}$. This involves stochastic noise terms and deterministic drift terms.

7.6.1 Euler-Maruyama algorithm

For simplicity, we only treat the optimization of the Euler-Maruyama integration scheme for an SDE in Itō form [13], which has truncation order $p = 1$ for convergence of moments. This is a discrete expression of the standard Ito SDE, so that for a finite step in time Δt,

$$\Delta \mathbf{x} = \mathbf{a}\Delta t + \mathbf{b}\Delta \mathbf{w}. \tag{7.47}$$

Next, we introduce $\mathbf{A} = \begin{pmatrix} \mathbf{a}^T\left(\mathbf{x}^{(1)}\right) & \cdots & \mathbf{a}^T\left(\mathbf{x}^{(N_S)}\right) \end{pmatrix}^T$ as an extended vector of drift coefficients, $\mathbf{B} = \mathrm{diag}\begin{pmatrix} \mathbf{b}\left(\mathbf{x}^{(1)}\right) & \cdots & \mathbf{b}\left(\mathbf{x}^{(N_S)}\right) \end{pmatrix}$ (a block-diagonal matrix with $\mathbf{b}\left(\mathbf{x}^{(i)}\right)$ elements) as an extended matrix of noise terms, and $\Delta\mathbf{W} = \begin{pmatrix} \left(\Delta\mathbf{w}^{(1)}\right)^T & \cdots & \left(\Delta\mathbf{w}^{(N_S)}\right)^T \end{pmatrix}^T$ as an extended vector of noises. The parallel sample vector $\mathbf{X}$ will satisfy the following vector SDE, now of dimension $N_S d$:

$$\Delta\mathbf{X} = \mathbf{A}\Delta t + \mathbf{B}\Delta\mathbf{W}. \tag{7.48}$$

One requires that the sampled observables are given as closely as possible by the known infinite ensemble results, that is,

$$\bar{\mathbf{o}}\left(\mathbf{X} + \Delta\mathbf{X}\right) - \bar{\mathbf{o}}\left(\mathbf{X}\right) = \left\langle \Delta\left(\mathbf{o}(\mathbf{x})\right)\right\rangle_\infty + \mathcal{O}(\Delta t^2), \tag{7.49}$$

where $\left\langle \Delta\left(\mathbf{o}(\mathbf{x})\right)\right\rangle_\infty$ is the ideal observable change for an infinite ensemble to order Δt. This is given either by an application of a single-step approximation to the Fokker-Planck equation (7.7) followed by partial integration, or equivalently by an application of Itō's formula [6] to the stochastic equation (7.8):

$$\left\langle \Delta o_m\left(\mathbf{x}\right)\right\rangle_\infty = \left\langle \frac{\partial o_m}{\partial x_i} a_i + \frac{1}{2}\frac{\partial^2 o_m}{\partial x_i \partial x_j}\mathrm{d}_{ij}\right\rangle_\infty \Delta t + \mathcal{O}\left(\Delta t^2\right) \tag{7.50}$$

In order to be able to calculate these moments in practice we must make another approximation, since we only know the required averages for a finite ensemble of trajectories. From the central limit theorem, if the individual variances are finite then:

$$\left\langle \frac{\partial o_m}{\partial x_i} a_i + \frac{1}{2}\frac{\partial^2 o_m}{\partial x_i \partial x_j}\mathrm{d}_{ij}\right\rangle_\infty = \left\langle \frac{\partial o_m}{\partial x_i} a_i + \frac{1}{2}\frac{\partial^2 o_m}{\partial x_i \partial x_j}\mathrm{d}_{ij}\right\rangle_{N_S} + \mathcal{O}\left(1/\sqrt{N_S}\right). \tag{7.51}$$

In some cases the equality is exact; see the discussion of the error introduced by this approximation in Section 7.6.3.

7.6.2 One-dimensional case

To illustrate how the method works for a single variable, we note that in the one-dimensional moment-based case, this result reduces to:

$$\langle \Delta (x^m) \rangle_\infty \approx m \left\langle x^{m-1} a + \frac{m-1}{2} x^{m-2} \mathrm{b}^2 \right\rangle_{N_S} \Delta t. \tag{7.52}$$

These methods can also be extended to higher orders in the step-size, but here we treat the equations up to the first order in Δt, and truncate higher orders for simplicity. Higher order algorithms will be treated elsewhere.

At this stage, we point out the existence of more than one possible strategy for satisfying the moment equations. Since each has its own distinct advantages and disadvantages, two particular approaches will be treated here. They are described in the following two subsections.

Both of the resulting POS methods have the following useful features:

- Uniqueness — the matrix iteration equations have unique results.
- Parameter independence — only the observables that are optimized are specified, not an arbitrary parameter.
- Linear complexity — the time taken scales as N_S, since the inner products only require N_S operations.

7.6.3 Error propagation

With any dynamical POS algorithm, the goal of the optimization is to remove any difference between the sampled observables and the infinite trajectory observables. This can be achieved exactly for static optimization, but there are error propagation effects to be considered in the dynamical case. Before treating the step-wise optimization method, we briefly consider the potential effect of error propagation.

To understand this, we note that error propagation effects depend on both the choice of observables and the structure of the equations themselves. Ideally, the optimized set of

observables form a closed set of equations, but this is rarely the case in practical nonlinear calculations.

We illustrate the effects of error propagation by considering the one-dimensional moment-based equations. Initially, the error in the p-th moment is:

$$\epsilon_p = \left|\langle x^p + \Delta x^p_{(0)}\rangle_{N_S} - \langle x^p + \Delta x^p\rangle_\infty\right| \neq 0, \tag{7.53}$$

which varies as $1/\sqrt{N_S}$ for random noises. The aim of our optimization is to set the difference to zero, apart from time-step errors. Can this be achieved exactly, assuming that the stochastic variables themselves are initially optimized at the start of the simulation, that is, $\langle x^p\rangle_{N_S} = \langle x^p\rangle_\infty$ for $p = 1, \dots M$?

To answer this, note that the drift and the diffusion terms can be expanded in the powers of x as

$$a(x,t) = \sum_{j=0}^{\infty} f_j(t) x^j, \quad b^2(x,t) = \sum_{j=0}^{\infty} g_j(t) x^j. \tag{7.54}$$

After substituting the expansions into the condition expressions, and expanding to order (Δt),

$$\begin{aligned} \epsilon_p = \quad & p^2 \Bigg| \left\langle x^{p-1}\Delta w + \frac{p-1}{2} x^{p-2}\left(\Delta w^2 - \mathsf{b}^2 \Delta t\right)\right\rangle_{N_S} \\ & + \sum_{j=M-p+2}^{\infty} f_j(t)\left(\langle x^{p+j-1}\rangle_{N_S} - \langle x^{p+j-1}\rangle_\infty\right)\Delta t \\ & + \frac{p-1}{2}\sum_{j=M-p+3}^{\infty} g_j(t)\left(\langle x^{p+j-2}\rangle_{N_S} - \langle x^{p+j-2}\rangle_\infty\right)\Delta t \Bigg|^2 . \end{aligned} \tag{7.55}$$

The last two terms depend on unoptimized and thus unknown differences $\langle x^j\rangle_{N_S} - \langle x^j\rangle_\infty$ for $j > M$. In deriving the POS algorithm, we assume that all the moments of the sampled distribution are correct at the start of the step in time, which allows us to neglect these higher-order moment differences.

One possibility is that the coefficients f and g for the corresponding indices are in fact zero. This is equivalent to having a at most linear in x, and b^2 at most quadratic in x. If this condition is satisfied, then higher order moments remain uncoupled to lower order

moments, and no error propagation occurs apart from the usual error propagation in an ODE method.

If this condition is not satisfied, neglecting these terms will result in unoptimized high order moments "leaking in" to low order moments over time, with the effect more noticeable for orders closer to M. We ignore this effect in deriving the algorithm, which means that we can only *locally* remove sampling errors exactly, as we show later. As a result, higher-order moment errors gradually degrade the optimized moments. This results in globally increased errors for nonlinear SDE solutions, which is demonstrated in the numerical examples described in the next section. The consequence is that it is no longer possible to reach machine precision. However, substantial variance reduction is still possible.

7.6.4 Combined optimization

In combined optimization, all time orders in the sampled moments are combined together, to give a moment estimate in an analogous form to that treated previously for the initial conditions. The advantage of this approach is its ease of implementation and simplicity.

However, unlike the case of the initial distribution, we typically do not know the infinite ensemble average moments exactly, that is, to all orders in Δt. Hence, in this approach the generated probability distribution differs from the stochastic distribution by terms of order Δt^2, even in the infinite ensemble limit.

Hence, we use the parallel SDE to provide the *initial* estimate for the stochastic trajectory, prior to error optimization, of:

$$\mathbf{X}_{(0)}\left(t+\Delta t\right)=\mathbf{X}\left(t\right)+\mathbf{A}\left(t\right)\Delta t+\mathbf{B}\left(t\right)\Delta\mathbf{W}\,. \tag{7.56}$$

Obtaining the improved trajectory estimates is then essentially identical to the static requirement given already. We can write the moment requirement in the form (7.26):

$$\bar{\mathbf{o}}\left(\mathbf{X}\left(t+\Delta t\right)\right)=\mathbf{c}, \tag{7.57}$$

where $\mathbf{c}$ is a vector of ideal observables, and $\mathbf{X}\left(t+\Delta t\right)=\mathbf{X}_{(0)}\left(t+\Delta t\right)+\delta\mathbf{X}$. Since we do not know these "ideal" averages over infinite number of trajectories, we approximate $\mathbf{c}$ by assuming we know all the moments exactly at the start of the time interval based

on knowing $\mathbf{X}(t)$. Changes in moments are estimated up to terms of order Δt using the approximations (7.50) and (7.51) so that, in general:

$$c_m = \left\langle o_m + \left[\frac{\partial o_m}{\partial x_i} a_i + \frac{1}{2} \frac{\partial^2 o_m}{\partial x_i \partial x_j} \mathsf{d}_{ij} \right] \Delta t \right\rangle_{N_S} . \tag{7.58}$$

In the one-dimensional example case, from Eq. (7.52),

$$c_m = \left\langle x^m + m \left[x^{m-1} a + \frac{m-1}{2} x^{m-2} \mathsf{b}^2 \right] \Delta t \right\rangle_{N_S} . \tag{7.59}$$

At the i-th step in time, we assume that $\mathbf{X}_{(i+1)}(t+\Delta t) = \mathbf{X}_{(i)}(t+\Delta t) + \delta \mathbf{X}_{(i+1)}$, so that

$$\bar{\mathbf{o}}\left(\mathbf{X}_{(i+1)}(t+\Delta t)\right) = \bar{\mathbf{o}}\left(\mathbf{X}_{(i)}(t+\Delta t)\right) + \mathbf{J}_{(i)} \delta \mathbf{X}_{(i+1)} + O(\delta \mathbf{X}^2_{(i+1)}). \tag{7.60}$$

7.6.5 Iterative solution to the combined equations

The iterative equations to be solved are:

$$\mathbf{J}_{(i)} \delta \mathbf{X}_{(i+1)} = \mathbf{c} - \bar{\mathbf{o}}_{(i)}. \tag{7.61}$$

Here we define the matrices $\mathbf{J}_{(i)}$ as in Eq. (7.31). The moment equation is then solved iteratively using the techniques outlined above, that is, either by pseudo-inverse iterations, or more efficiently by the matrix inversion method. With pseudo-inverses, the iteration equations are then simply:

$$\mathbf{X}_{(i+1)} = \mathbf{X}_{(i)} + \mathbf{J}^{+}_{(i)} \left(\mathbf{c} - \bar{\mathbf{o}}_{(i)}\right). \tag{7.62}$$

For cases where $\mathbf{u} = \mathbf{J}\mathbf{J}^{\dagger}$ is invertible, we can define, just as in the static case:

$$\mathbf{X}_{(i+1)} = \mathbf{X}_{(i)} + \mathbf{J}^{\dagger}_{(i)} \mathbf{u}^{-1}_{(i)} \left(\mathbf{c} - \bar{\mathbf{o}}_{(i)}\right). \tag{7.63}$$

Numerical results using this method are given in the next Section.

7.6.6 Individual optimization

In the individual optimization approach, both the finite ensemble moments and the infinite ensemble moments are calculated to the same order in Δt, which allows us to optimize each order in time individually. This strategy permits a clear separation of ensemble and time-step errors. For this purpose, it is convenient to define an effective noise term, $\Delta\mathbf{V} = \mathsf{B}\Delta\mathbf{W}$. This is optimized to give the final change in $\mathbf{X}$.

From the moment equations of Eq. (7.49) and (7.50), one has the required observable changes to order Δt:

$$\sum_n \frac{\partial \bar{o}_m}{\partial X_n}\Delta V_n(\mathbf{X}) + \frac{1}{2}\sum_{n,p}\frac{\partial^2 \bar{o}_m}{\partial X_n \partial X_p}\left[\Delta V_n(\mathbf{X})\Delta V_p(\mathbf{X}) - \Delta t \mathsf{D}_{np}(\mathbf{X})\right] = e_m = \mathcal{O}\left(\Delta t^2\right) \tag{7.64}$$

Here, since we wish to set every moment error to zero over a finite set of moments, it is convenient to define an error vector, $\mathbf{e} = (e_1, \dots e_M)^T$. For example, in the one-dimensional case, from Eq. (7.49) the extended diffusion matrix is simply diagonal, and one has that:

$$\frac{m}{N_S}\sum_n \left[X_n^{m-1}\Delta V_n + \frac{m-1}{2}X_n^{m-2}\left(\Delta V_n^2 - \mathsf{D}_{nn}\Delta t\right)\right]_{N_S} = e_m = \mathcal{O}\left(\Delta t^2\right), \tag{7.65}$$

The error terms can be viewed as two distinct "mean" and "variance" conditions analogous to those in Eq. (7.9), and are equal to zero in the limit of an infinite ensemble. The error requirement is satisfied provided two individual conditions are met, representing terms of order $\sqrt{\Delta t}$ and Δt respectively:

$$\begin{aligned} \mathbf{e}^{(1)} &= \mathsf{J}\Delta\mathbf{V}^{(\mathrm{opt})} = \mathbf{0} \\ \mathbf{e}^{(2)} &= \frac{1}{2}\boldsymbol{\mathcal{H}} : \left[\Delta\mathbf{V}^{(\mathrm{opt})}\left(\Delta\mathbf{V}^{(\mathrm{opt})}\right)^T - \mathsf{D}\Delta t\right] = \mathbf{0}, \end{aligned} \tag{7.66}$$

where ":" stands for double contraction, so $\left(\boldsymbol{\mathcal{H}} : \mathbf{V}\mathbf{V}^T\right)_i \equiv \sum_{jk}\mathcal{H}_{ijk}V_jV_k$. Here, we define J as previously, and $\boldsymbol{\mathcal{H}}$ as a Hessian tensor:

$$\mathcal{H}_{mnp}(\mathbf{X}) = \frac{\partial^2 \bar{o}_m(\mathbf{X})}{\partial X_n \partial X_p}. \tag{7.67}$$

In the one-dimensional moment-based example, the results are simpler. Due to the absence of cross-derivative terms, the second derivative tensor becomes a matrix,

$$\mathsf{H}_{mn}(\mathbf{X}) = \frac{\partial^2 \bar{o}_m(\mathbf{X})}{\partial X_n^2} = m\frac{m-1}{N_S}X_n^{m-2}. \tag{7.68}$$

The variance condition then becomes:

$$\mathbf{e}^{(2)} \equiv \frac{1}{2}\mathsf{H}\left(\Delta\mathbf{V}^{(\mathrm{opt})} \circ \Delta\mathbf{V}^{(\mathrm{opt})} - \mathbf{D}\Delta t\right) = \mathbf{0}, \tag{7.69}$$

where $D_n \equiv \mathsf{D}_{nn}$.

7.6.7 Iterative solution to the individual equations

These are nonlinear equations, and to solve them we implement an iterative approach. For a large ensemble, the optimizing equations are nearly satisfied to zero-th order, up to errors of order $1/\sqrt{N_S}$. Hence, it is efficient to iteratively solve the equations by linearization, giving a Newton-Raphson approach similar to the combined method given above.

Defining the difference $\delta\mathbf{X}_{(i+1)} \equiv \Delta\mathbf{V}_{(i+1)} - \Delta\mathbf{V}_{(i)}$, the separate conditions are linearized by assuming that $\|\delta\mathbf{X}_{(i+1)}\| \ll \|\Delta\mathbf{V}\|$, to give

$$\begin{aligned} \mathsf{J}\left(\Delta\mathbf{V}_{(i)} + \delta\mathbf{X}_{(i+1)}\right) &= \mathbf{0}, \\ \frac{1}{2}\mathcal{H} : \left(\Delta\mathbf{V}_{(i)}\Delta\mathbf{V}_{(i)}^T + \delta\mathbf{X}_{(i+1)}\Delta\mathbf{V}_{(i)}^T + \Delta\mathbf{V}_{(i)}\delta\mathbf{X}_{(i+1)}^T - \mathsf{D}\Delta t\right) &= \mathbf{0}. \end{aligned} \tag{7.70}$$

Noting that $\mathcal{H}$ is symmetric with respect to its last two indices ($\mathcal{H}_{mnp} = \mathcal{H}_{mpn}$), we can rewrite the terms with $\delta\mathbf{X}_{(i+1)}$ in the second equation as

$$\frac{1}{2}\mathcal{H} : \left(\delta\mathbf{X}_{(i+1)}\Delta\mathbf{V}_{(i)}^T + \Delta\mathbf{V}_{(i)}\delta\mathbf{X}_{(i+1)}^T\right) \equiv \left(\mathcal{H}\Delta\mathbf{V}_{(i)}\right)\delta\mathbf{X}_{(i+1)}, \tag{7.71}$$

which allows us to concatenate the two systems into a single matrix equation

$$\tilde{\mathsf{J}}_{(i)}\delta\mathbf{X}_{(i+1)} = \tilde{\mathbf{R}}\left(\Delta\mathbf{V}_{(i)}\right). \tag{7.72}$$

Here the doubly extended matrix $\tilde{\mathbf{J}}$ is:

$$\tilde{\mathbf{J}} = \begin{pmatrix} \mathbf{J} \\ \mathcal{H}\Delta\mathbf{V} \end{pmatrix}. \tag{7.73}$$

and the remainder vector $\tilde{\mathbf{R}}\left(\Delta\mathbf{V}\right)$ is defined as:

$$\tilde{\mathbf{R}}\left(\Delta\mathbf{V}\right) = \begin{pmatrix} -\mathbf{J}\Delta\mathbf{V} \\ \frac{1}{2}\mathcal{H} : \left(\mathbf{D}\Delta t - \Delta\mathbf{V}_{(i)}\Delta\mathbf{V}_{(i)}^{T}\right) \end{pmatrix}. \tag{7.74}$$

Starting with $i = 0$, we now take $\Delta\mathbf{V}_{(i)}$ and solve these equations iteratively to obtain $\Delta\mathbf{V}_{(i+1)} = \delta\mathbf{X}_{(i+1)} + \Delta\mathbf{V}_{(i)}$. To obtain a linear solution for each iterative step, we can use a pseudo-inverse as before, so that:

$$\Delta\mathbf{V}_{(i+1)} = \Delta\mathbf{V}_{(i)} + \tilde{\mathbf{J}}_{(i)}^{+}\tilde{\mathbf{R}}\left(\Delta\mathbf{V}_{(i)}\right). \tag{7.75}$$

As previously, this reduces to ordinary matrix inversion in most cases of interest since, we can define a positive semi-definite matrix

$$\tilde{\mathbf{u}}\left(\Delta\mathbf{v}\right) \quad = \quad \tilde{\mathbf{J}}\left(\Delta\mathbf{V}\right)\tilde{\mathbf{J}}^{\dagger}\left(\Delta\mathbf{V}\right). \tag{7.76}$$

Provided this is invertible, this gives an iterative procedure for the individual POS algorithm, which is:

$$\Delta\mathbf{V}_{(i+1)} = \Delta\mathbf{V}_{(i)} + \tilde{\mathbf{J}}_{(i)}^{\dagger}\tilde{\mathbf{u}}_{(i)}^{-1}\mathbf{R}_{(i)}. \tag{7.77}$$

This method can be implemented to scale as $\mathcal{O}\left(MN_S\right)$ by pre-calculating repeating matrix elements.

7.7 Synthetic one-step benchmarks

Before we apply our POS method to solving an SDE with error-propagation over many successive minimization steps, it is informative to see how effective it is at finding a solution for the corresponding single step equation.

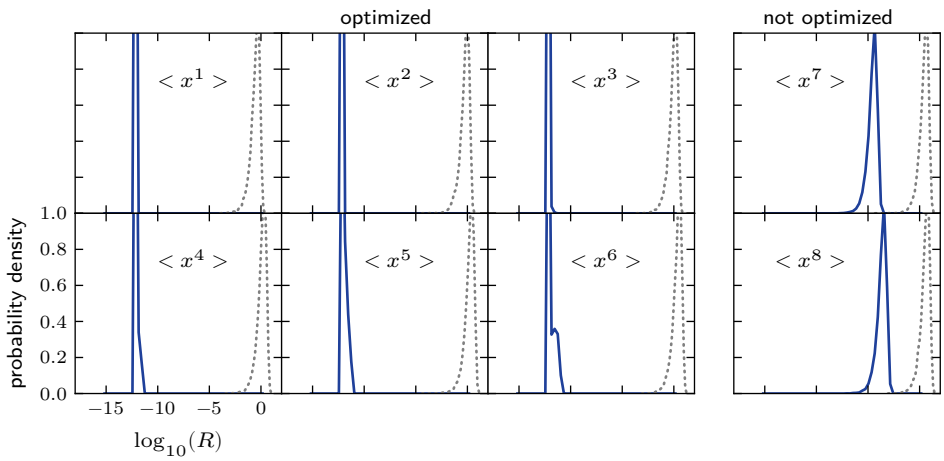

Figure 7.5: Distribution of normalized errors in the synthetic benchmark for the combined method with $\Delta t = 10^{-4}$, $N_S = 1000$. Blue solid lines show the distribution of R_m for the optimized vector, grey dotted lines show the distribution of R_m for the initial approximation $\mathbf{X}_{(0)}(t + \Delta t)$.

In order to evaluate the statistical performance of these two methods at single-step error optimization, we implement the combined and individual optimization methods with 10000 separate initial ensembles, each of 1000 parallel samples of $\mathbf{X}$ and initial stochastic noises $\mathbf{W}$. In all the tests, the initial elements of $\mathbf{X}$ are normally distributed with the mean 1 and standard deviation 0.1, $a_i = 0.5$ for all i (constant drift), $\mathrm{b}_i = 0.5$ for all i (additive noise).

Results for more complicated synthetic tests show similar behavior, so we do not include them here.

7.7.1 Combined optimization

In this section we are interested in how well we can solve the target equation (7.57). The error measure in this case is essentially the same as in Section 7.5.1:

$$\tilde{R}_m(\mathbf{X}) = \left| \frac{1}{N_S} \mathbf{X}^m \cdot \mathbf{1} - c_m \right|, \tag{7.78}$$

where the target moments c_m are given by Eq. (7.59).

Similarly to the previous section, we will scale the errors as

$$R_m = \frac{\tilde{R}_m}{\sqrt{\Delta t / N_S}} \tag{7.79}$$

in order to bring the errors for the unoptimized, randomly sampled vectors close to 1.

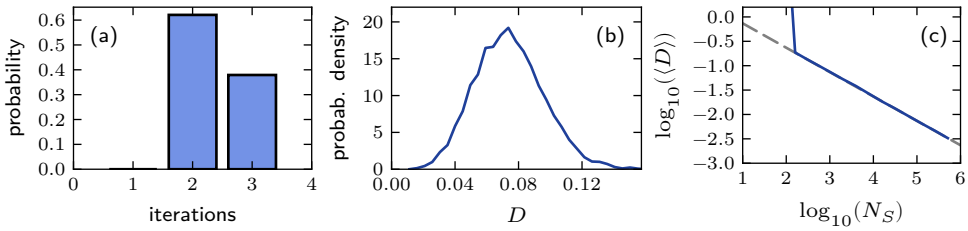

Figure 7.6: (a) number of iterations taken in the synthetic test for the combined method, $\Delta t = 10^{-4}$, $N_S = 1000$, (b) Relative distance D between the optimized vector $\mathbf{X}^{(\mathrm{opt})}(t + \Delta t)$ and the initial approximation $\mathbf{X}_{(0)}(t + \Delta t)$, and (c) average relative distance $\langle D \rangle$ depending on the number of trajectories N_S in the synthetic optimization tests for the individual expansion method, $\Delta t = 10^{-4}$. The grey dashed line denotes the slope of a $1/\sqrt{N_S}$ dependence.

The results in Fig. 7.5 show that a significant improvement is achieved for all 6 moments being optimized, with normalized sampling errors reduced to 10^{-12}, or around twelve orders of magnitude in these examples. Even the non-optimized moments of orders $m = 7, 8$ have errors reduced to 10^{-5}, or around five orders.

7.7.2 Number of iterations and convergence distance

The method requires only a few iterations to converge, as Fig. 7.6(a) demonstrates.

Next, we calculate the distance of the variance-reduced solution from the original estimate. It is important that this quantity is as small as possible, consistent with the targeted variance reduction, to eliminate systematic errors in higher-order, non-minimized moments. We compute:

$$D = \frac{\|\Delta\mathbf{X}^{(\mathrm{opt})} - \Delta\mathbf{X}_{(0)}\|}{\|\Delta\mathbf{X}_{(0)}\|} \equiv \frac{\|\mathbf{X}^{(\mathrm{opt})}(t + \Delta t) - \mathbf{X}_{(0)}(t + \Delta t)\|}{\|\mathbf{X}_{(0)}(t + \Delta t) - \mathbf{X}(t)\|} \tag{7.80}$$

to estimate how far the solution is from the initial approximation. Fig. 7.6(b) shows that the deviation of the optimized noise from the original is 0.07 on average, which is of order $1/\sqrt{N_s}$ as expected.

The difference in the orders of Δt in the left and the right parts of the target equation (7.57) leads to the distance D being bounded by a finite value, depending on either Δt or $1/N_S$. This is unlike the behavior demonstrated by the static optimization in the previous section. Our tests show that for very small Δt, the asymptotic dependence of $\langle D \rangle$ on N_S is $\langle D \rangle \propto 1/\sqrt{N_S}$, as shown in Fig. 7.7(a). Similarly, for very large N_S the dependence is $\langle D \rangle \propto \Delta t^{3/2}$, as seen in Fig. 7.7(b).

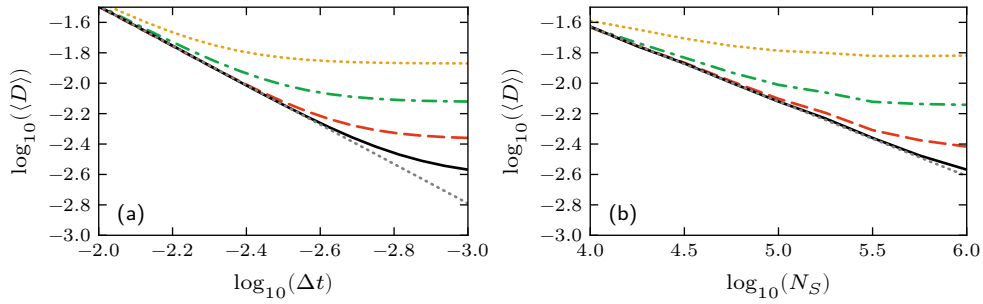

Figure 7.7: (a) Average relative distance $\langle D \rangle$ depending on Δt in the synthetic optimization tests for the combined method, with $N_S = 10^{4.5}$ (dotted yellow line), $N_S = 10^5$ (dash-dotted green line), $N_S = 10^{5.5}$ (dashed red line) and $N_S = 10^6$ (solid black line). The grey dashed line denotes the least squares fit for the linear part of the $N_S = 10^6$ line and corresponds to the dependence $\langle D \rangle \propto \Delta t^{1.29}$. (b) Average relative distance $\langle D \rangle$ depending on N_S in the synthetic optimization tests for the combined method, with $\Delta t = 10^{-2.25}$ (dotted yellow line), $\Delta t = 10^{-2.5}$ (dash-dotted green line), $\Delta t = 10^{-2.75}$ (dashed red line) and $\Delta t = 10^{-3}$ (solid black line). The grey dashed line denotes the least squares fit for the linear part of the $\Delta t = 10^{-3}$ line and corresponds to the dependence $\langle D \rangle \propto N_S^{-0.49}$.

This asymptotic error is essentially due to the fact that the moment equations are themselves an expansion in Δt, leading to a residual truncation error.

The tests of the dependence on the noise scale (Δt) show that the method is stable for a wide range of values, but breaks down at larger values of Δt as seen in Fig. 7.8. The dependence on the number of trajectories in Fig. 7.9 displays a similar cutoff as seen in the tests of the static POS, in the results of the previous subsection.

7.7.3 Individual optimization

In this subsection we will use the individual method from Section 7.6.6, for a similar type of synthetic benchmark. Since the target equations we try to satisfy are different from the previous section, the definition of the error changes to simply

$$\tilde{R}_m(\mathbf{X}) = |e_m^{(1)}| + |e_m^{(2)}|, \tag{7.81}$$

where the vectors $\mathbf{e}^{(1)}$ and $\mathbf{e}^{(2)}$ are defined by Eq. (7.66).

Similarly to the approach of the previous subsection, we normalize the error as

$$R_m = \frac{\tilde{R}_m}{\sqrt{\Delta t / N_S}}. \tag{7.82}$$

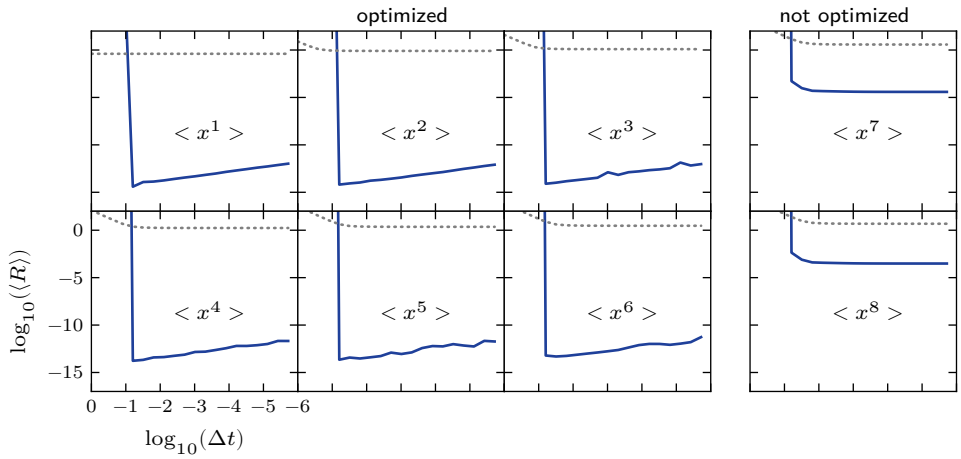

Figure 7.8: Dependence of errors on the number of trajectories N_S in the synthetic test for the combined method, $\Delta t = 10^{-4}$. Blue solid lines show the average of R_m for the optimized vector, grey dotted lines show the average of R_m for the initial approximation $\mathbf{X}_{(0)}\ (t + \Delta t)$.

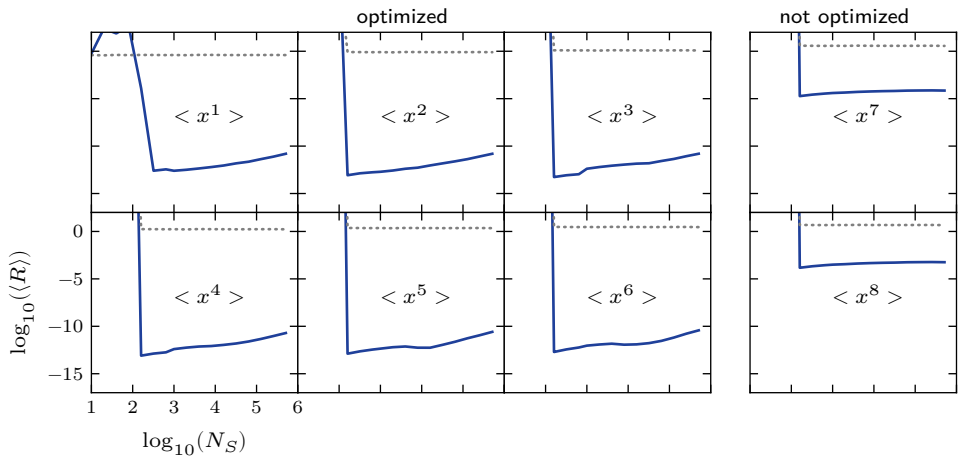

Figure 7.9: Dependence of average scaled errors on Δt in the synthetic test for the combined method, $N_S = 1000$. Blue solid lines show the average of R_m for the optimized vector, grey dotted lines show the average of R_m for the initial approximation $\mathbf{X}_{(0)}\ (t + \Delta t)$.

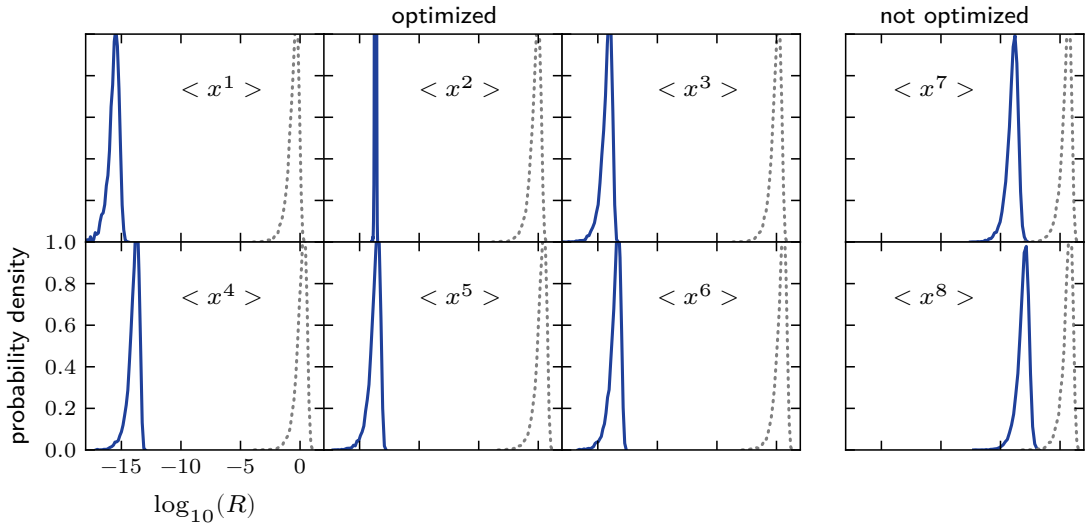

Figure 7.10: Distribution of normalized errors in the synthetic benchmark for the individual method with $\Delta t = 0.1$, $N_S = 1000$. Blue solid lines show the distribution of R_m for the optimized vector, grey dotted lines show the distribution of R_m for the initial effective noise term $\Delta\mathbf{V}$.

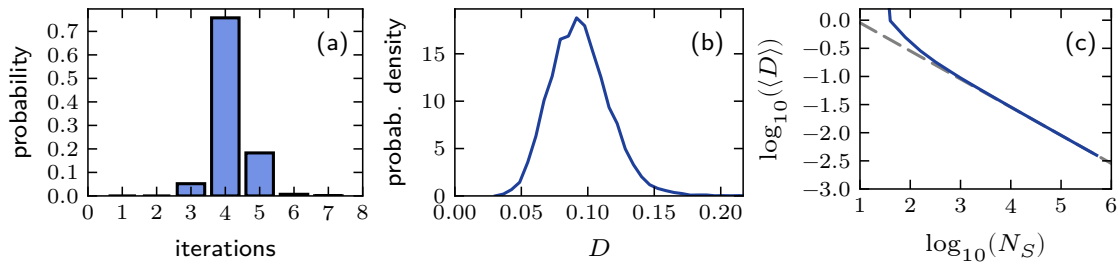

Figure 7.11: (a) The number of iterations taken in the synthetic test for the individual method, $\Delta t = 0.1$, $N_S = 1000$. (b) Relative distance D between the optimized vector $\Delta\mathbf{V}^{(\mathrm{opt})}$ and the initial approximation $\Delta\mathbf{V}$. (c) Average relative distance $\langle D \rangle$ depending on the number of trajectories N_S in the synthetic optimization tests for the individual method, $\Delta t = 0.1$. The grey dashed line denotes the slope of a $1/\sqrt{N_S}$ dependence.

The results in Fig. 7.10(a) show that a significant improvement is achieved for the first 6 moments being optimized, with results only limited by the numerical precision. To check that numerical precision was the limiting factor, tests were run with the same method in quadruple precision. This demonstrated that the improvement was indeed limited only by the numerical precision, with errors now as low as 10^{-30}. The stopping condition must be changed to $\eta = 10^{-15}$ to reach this greater numerical precision limit.

7.7.4 Number of iterations and convergence distance

We also calculate the distance

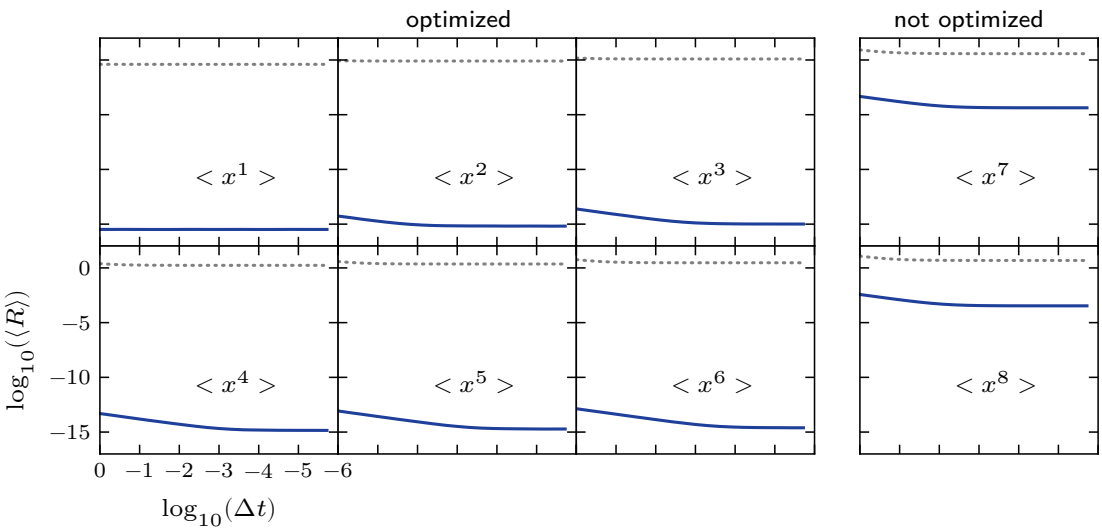

Figure 7.12: Dependence of average scaled errors on Δt in the synthetic test for the individual method, $N_S = 1000$. Blue solid lines show the average of R_m for the optimized vector, grey dotted lines show the average of R_m for the initial approximation $\Delta \mathbf{V}_{(0)}$.

$$D \equiv \frac{\|\Delta \mathbf{V}^{(\mathrm{opt})} - \Delta \mathbf{V}\|}{\|\Delta \mathbf{V}\|} \tag{7.83}$$

and the number of iterations in our tests. The method also requires only a few iterations to converge, as Fig. 7.11(a) demonstrates. Fig. 7.11(b) shows that the deviation of the optimized noise from the original one is small. Unlike the combined method, there is no mismatch between the orders of Δt in Eq. (7.66), and the distance therefore converges as $\langle D \rangle \propto 1/\sqrt{N_S}$, as shown in Fig. 7.11(c).

The tests of the dependence on the noise scale (Δt) show that this method is exceptionally stable over a wide range of step-size values, as seen in Fig. 7.12. The dependence on the number of trajectories in Fig. 7.13 displays a similar cutoff to those seen in the tests of the static POS.

7.8 Dynamical numerical results

Next, putting all the preliminary results together, tests were performed of complete stochastic differential equation solutions. Although all results demonstrate strong variance reduction, the amount and nature of variance reduction depends on the precise equation, as explained below.

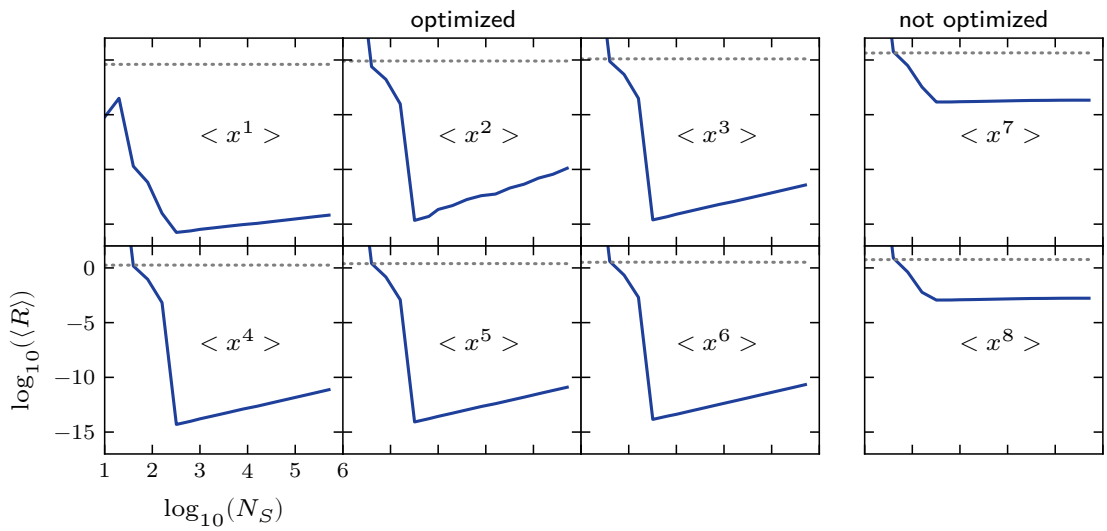

Figure 7.13: Dependence of errors on the number of trajectories N_S in the synthetic test for the expansion method, $\Delta t = 0.1$. Blue solid lines show the average of R_m for the optimized vector, grey dotted lines show the average of R_m for the initial approximation $\Delta \mathbf{V}_{(0)}$.

7.8.1 Linear drift, additive noise

It is instructive to first consider an "ideal" linear SDE for the application of POS, with the target observables being the moments of x as previously. An ideal linear case is one with a linear a and b. This is preferred as a first test because the errors from non-optimized moments do not get mixed with the optimized ones, as explained in Section 7.6.3. In these benchmarks, both the combined and individual methods were used, giving very similar performance overall.

As an example, we used a one-dimensional Ornstein-Uhlenbeck process

$$\mathrm{d}x = (f - gx)\,\mathrm{d}t + b\mathrm{d}w, \tag{7.84}$$

where $\langle \mathrm{d}w \rangle = 0$ and $\langle \mathrm{d}w^2 \rangle = dt$, and f, g and b are constants. It has well-known exact solutions for every moment. Namely, for Gaussian starting conditions the mean is given by:

$$\langle x \rangle = \frac{f}{g}\left(1 - e^{-gt}\right) + e^{-gt}\langle x \rangle_{t=0}, \tag{7.85}$$

and for the variance,

$$\langle (x - \langle x \rangle)^2 \rangle = e^{-2gt} \langle (x - \langle x \rangle)^2 \rangle_{t=0} + \frac{b^2}{2g} \left(1 - e^{-2gt}\right), \tag{7.86}$$

Higher order central moments are then calculated according to

$$\langle (x - \langle x \rangle)^m \rangle = 0, \quad m = 3, 5, \ldots \tag{7.87}$$

and:

$$\langle (x - \langle x \rangle)^m \rangle = m!! \langle (x - \langle x \rangle)^2 \rangle^{m/2}, \quad m = 4, 6, \ldots \tag{7.88}$$

Consequently, the raw moments are calculated from the central ones using the recursive formula

$$\langle x^m \rangle = \langle (x - \langle x \rangle)^m \rangle - \sum_{p=0}^{m-1} \begin{pmatrix} m \\ p \end{pmatrix} \langle x^p \rangle \left(-\langle x \rangle\right)^{m-p}. \tag{7.89}$$

We also use cumulants as benchmarks, since these are a commonly used alternative statistical measure to the ordinary moment. The cumulants are calculated from the moments recursively, as

$$\kappa_m = \langle x^m \rangle - \sum_{p=1}^{m-1} \begin{pmatrix} m-1 \\ p-1 \end{pmatrix} \kappa_p \langle x^{m-p} \rangle. \tag{7.90}$$

In our tests we took $f = 1$, $g = 0.2$, $b = 0.5$. At initial times, $x(0)$ is normally distributed with mean 0.5 and standard deviation 0.1. We pre-optimized the starting distributions using the static method from Section 7.4.4 in order to satisfy the requirements of POS methods.

7.8.2 Cumulant and moment variance reduction

For our tests, we compare the deviations — due to sampling errors — from the exact solution in the first 8 moments and first 8 cumulants of x at $t = 1$ for the POS integrator. We use both combined and individual methods, with a reference SDE solver having the same number of samples, for comparison. The latter uses the same integration method (explicit Euler) and the same number of trajectories as POS. We set up POS to optimize

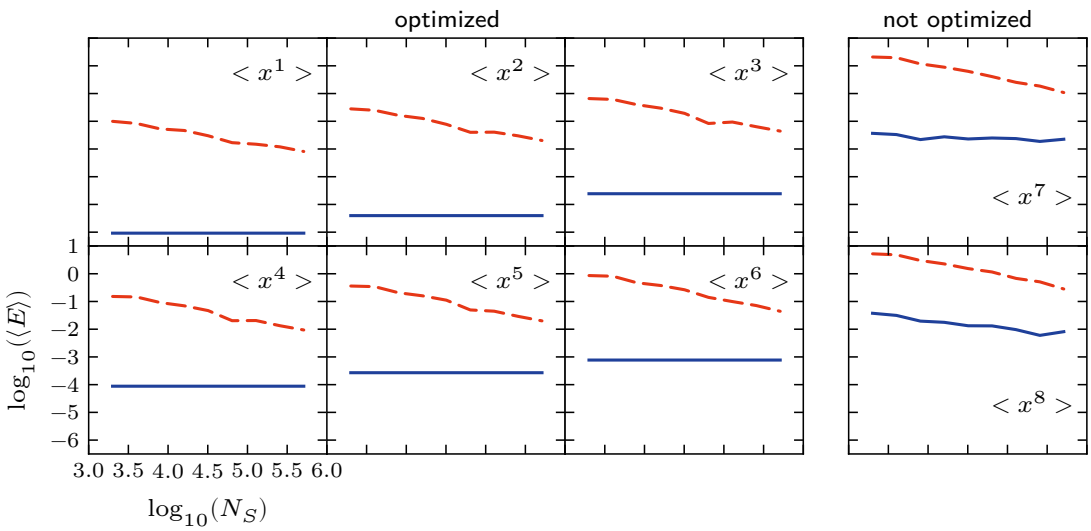

Figure 7.14: Linear SDE case with $M = 6$ optimized moments, showing the difference of the first 8 moments from the exact solution at $t = 1$ for the POS integrator using the combined method (solid blue lines) and a reference explicit Euler integrator (dashed red lines), with identical numbers of trajectories and samples, averaged over 8 tests. Both integrators use $N_T = 8 \times 10^4$ time steps.

the first 6 moments, so that we could look at the behavior of both optimized and non-optimized moments.

In Figs. 7.14 and 7.15 we plotted the error in the first moments

$$E\left[\langle x^m \rangle\right] = \left|\langle x^m \rangle - \langle x^m \rangle^{(\text{exact})}\right| \tag{7.91}$$

for the reference and the two POS integrators, averaged over 8 independent tests. In case of the POS methods the error is dominated by the time step error ϵ_T, while the error for the reference integrator decreases proportional to $1/\sqrt{N_S}$. One can see that the POS error is significantly decreased even for the moments that were not directly optimized. The error improvement is largest for the individual method, but is still several orders of magnitude even for the simpler combined method. For the remaining graphs we will mostly focus on the individual method which give better results, to minimize the number of graphs, as both have similar general behavior.

The general behavior in the linear case can be understood better if we plot the errors in the cumulants $E\left[\kappa_p\right]$ instead, as in Fig. 7.16. It shows that the error in the first 6 cumulants is greatly reduced, while the non-optimized cumulants stay roughly the same. In Fig. 7.15, 7-th and 8-th moments, while not being optimized directly, still depend on

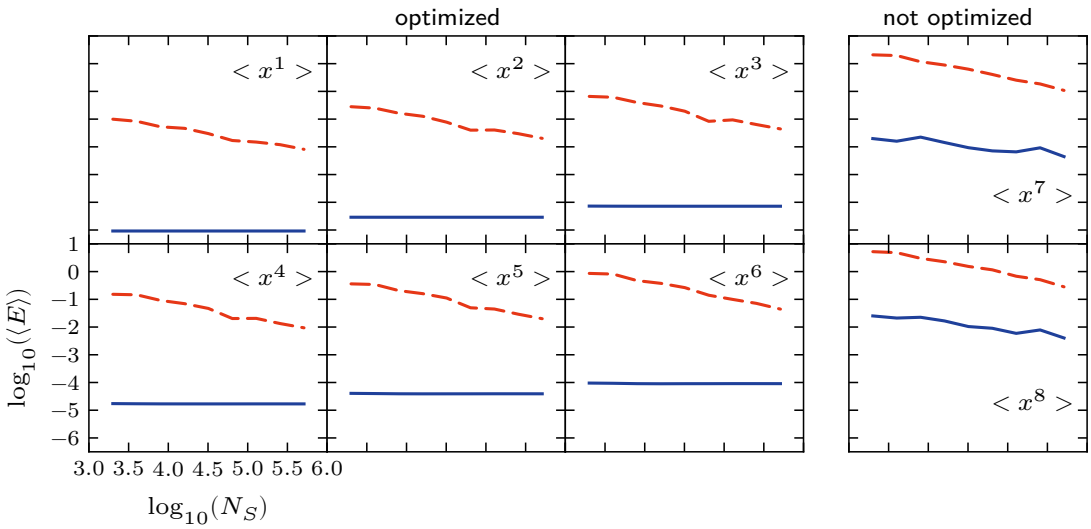

Figure 7.15: Linear SDE case with $M = 6$ optimized moments, showing the difference of the first 8 moments from the exact solution at $t = 1$ for the POS integrator using the individual method (solid blue lines) and a reference explicit Euler integrator (dashed red lines), with identical numbers of trajectories and samples, averaged over 8 tests. Both integrators use $N_T = 8 \times 10^4$ time steps.

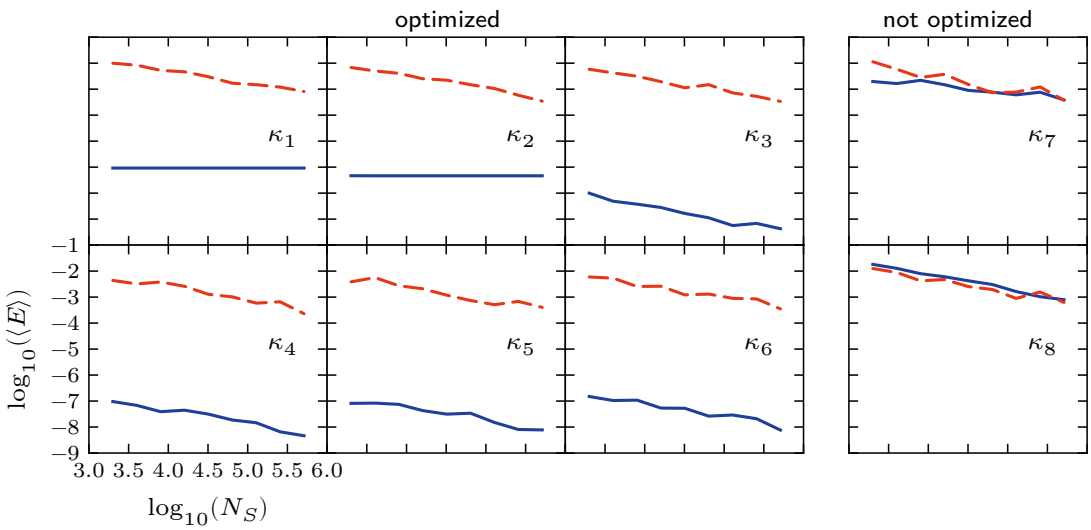

Figure 7.16: Linear SDE case with $M = 6$ optimized moments, showing the difference of the first 8 cumulants from the exact solution at $t = 1$ for the POS integrator using the individual method (solid blue lines) and a reference explicit Euler integrator (dashed red lines), with identical numbers of trajectories and samples, averaged over 8 tests. Both integrators use $N_T = 8 \times 10^4$ time steps.

the lower cumulants, and thus benefit from their reduced error. This behavior is caused by the fact that in a linear SDE, cumulants of different orders are not coupled with each other; however moments *are* coupled, leading to the observed error performance.

7.8.3 Nonlinear drift, additive noise

As an example of a real-world, nonlinear SDE we can apply the POS method to, we take an equation with a nonlinear drift term and additive noise:

$$\mathrm{d}x = \left(x - x^3\right)\mathrm{d}t + \mathrm{d}w, \tag{7.92}$$

where $\langle \mathrm{d}w \rangle = 0$ and $\langle \mathrm{d}w^2 \rangle = 1$. This is similar to a known case where moment hierarchy methods fail to give correct results [18].

For large times t, Eq. (7.92) converges to a steady-state solution. For any given observable $\langle f(x) \rangle$, we can find its steady-state value using the formula

$$\langle f(x) \rangle_{ss} = \frac{\int_{-\infty}^{\infty} f(x)\, e^{x^2 - \frac{1}{2}x^4} \mathrm{d}x}{\int_{-\infty}^{\infty} e^{x^2 - \frac{1}{2}x^4} \mathrm{d}x}. \tag{7.93}$$

We initialize our ensemble with normally-distributed numbers with mean 0 and standard deviation $\frac{1}{\sqrt{2}}$ and integrate Eq. (7.92) using a combined POS-method as well as an unoptimized Ito-Euler method until $t = 25$ and then compare with the steady-state values. The POS-optimized integration includes an initial static optimization. Integrating until $t = 25$ ensures that Eq. (7.92) predicts observables sufficiently close to the steady-state values. We assume that after some integration time, the time evolution of the observables will be dominated by a transient which is exponentially converging to the steady-state solution. A least-squares fit of the time-evolution of the 8'th moment is in excellent agreement with the time-evolution resulting from our simulations. From the fitting parameters, we can read off a decay of more that 50 mean lifetimes at $t = 25$. We conclude that the deviation from the steady-state value due to finite integration time is negligible compared to sampling errors. As in the previous section, we plot the first 8 moments and first 8 cumulants of x at $t = 25$, but we optimize only the first 6 moments when the POS is used. In order to show the behaviour of non-polynomial observables, we also plot the exponential and absolute value of x.

In Fig. 7.17 we are plotting differences from the accurate solution for the first 8 moments:

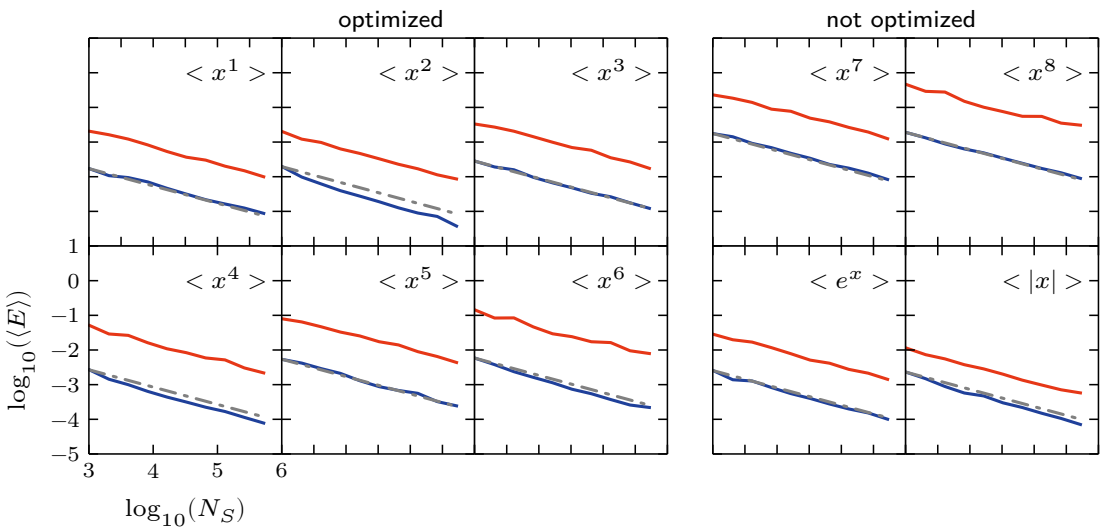

Figure 7.17: Nonlinear SDE case with $M = 6$ optimized moments, showing the difference of the first 8 moments, as well as the exponential and absolute value function from the steady-state solution at $t = 25$ for the POS integrator using the combined method (solid blue lines). We also include the errors in an explicit Euler integrator (solid red lines), with identical numbers of trajectories and samples, for comparison. Both integrators use $N_T = 12.5 \times 10^3$ time steps. The dash-dotted grey line shows the slope for $\propto N_S^{-1/2}$.

$$E\left[\langle x^m \rangle\right] = \left|\langle x^m \rangle - \langle x^m \rangle_{ss}\right| \tag{7.94}$$

as well as for the exponential and absolute value function defined analogously to Eq. (7.94) averaged over 120 independent test runs for the POS integration and 600 runs for the reference integration.

The POS integrator consistently outperforms the reference integrator, with errors reduced by a factor of $10 - 100$**.**

A sampling error reduction of this order of magnitude requires $10^2 - 10^4$ times more samples using standard integrators.

In Fig. 7.18 we show results for the cumulants in the nonlinear case, for completeness. Unlike the linear case, where the improvement was confined to only the explicitly optimized cumulants, we see that in the nonlinear case even the non-optimized cumulants are optimized as well. This is because the nonlinear equations couple all cumulants with each other. The results show that POS results in dependency slightly better than $N_S^{-1/2}$ for the even cumulants. This is an interesting and noteworthy results though we don't have an explanation for it yet.

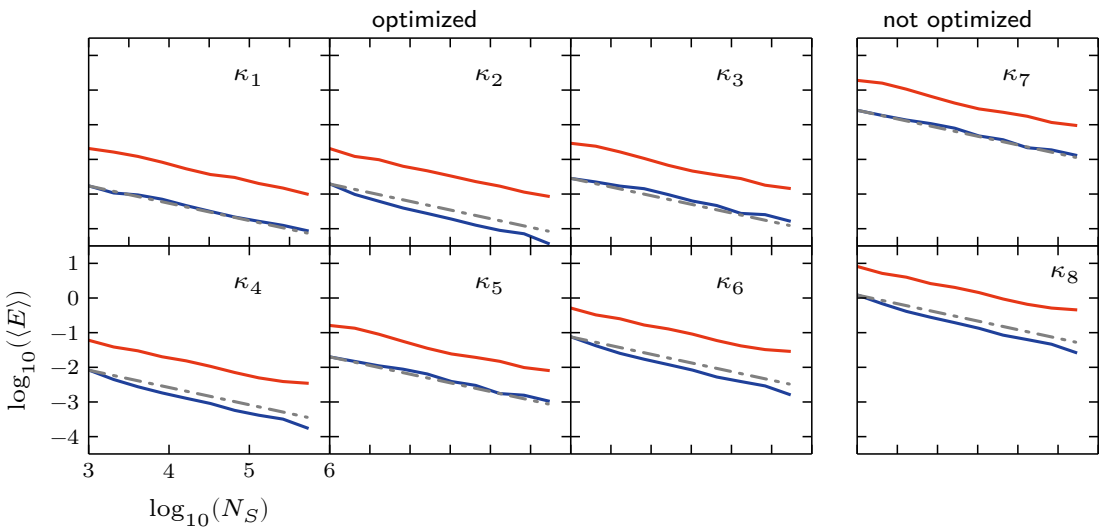

Figure 7.18: Nonlinear SDE case, showing the difference of the first 8 cumulants from the steady-state values at $t = 25$ for the POS integrator using the combined method (solid blue lines) and a reference explicit Euler integrator (dashed red lines). Both integrators use $N_T = 12.5 \times 10^3$ time steps. The dash-dotted grey lines shows the slope for $\propto N_S^{-1/2}$.

In a separate analysis, we have integrated Eq. (7.92) until $t = 4$ with a time-step of $\Delta t = 4 \times 10^{-4}$ using both the combined and the individual POS-method as well as an explicit Ito-Euler reference method. We have compared the values of the first 8 consecutive moments and cumulants with that of a regular SDE integrator with $N_T = 8 \times 10^4$ time steps and $N_S = 10^9$ trajectories, using a central difference integration method [24]. The large number of trajectories for the regular SDE integrator ensured that we obtained a sufficiently accurate estimate for the true value of the moments and cumulants at $t = 4$. The comparison showed that individual POS method performed very similarly to the combined POS method, resulting in a similarly good accuracy improvement. In the interest of space, we are limiting our results to the combined POS method from now on.

7.8.4 Irregular drift, additive noise

We now investigate the properties of an SDE whose functional behaviour is not a regular polynomial. We chose the SDE

$$\mathrm{d}x = x\,(1 - |x|)\,\mathrm{d}t + \mathrm{d}w, \tag{7.95}$$

where $\langle \mathrm{d}w \rangle = 0$ and $\langle \mathrm{d}w^2 \rangle = 1$.

We can obtain the steady-state solutions for a given observable $\langle f(x) \rangle$ via

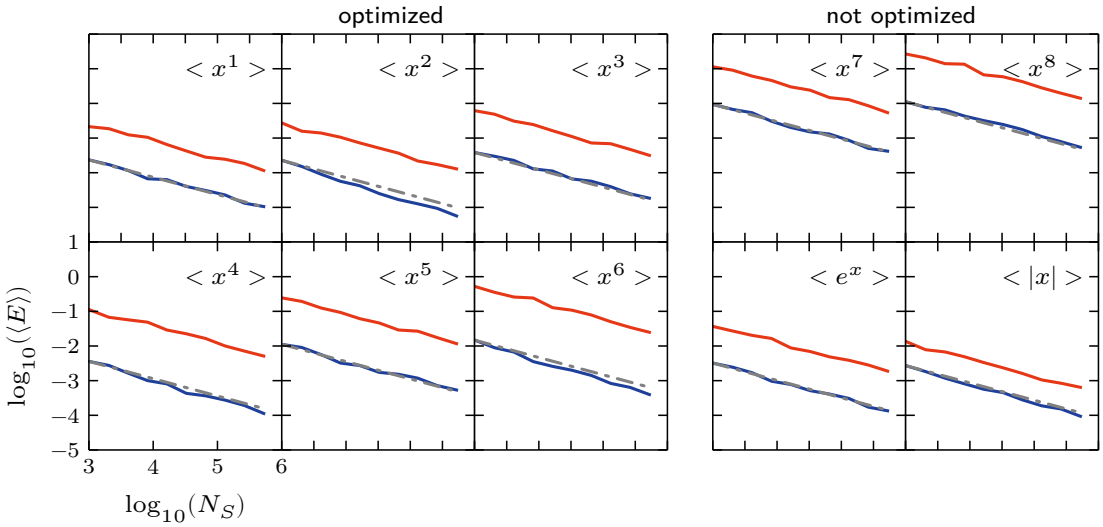

Figure 7.19: Nonlinear SDE case, showing the difference of the first 8 moments, the exponential and absolute value function from the steady-state solution at $t = 25$ for the POS integrator using the combined method (solid blue lines) and a reference explicit Euler integrator (solid red lines). Both integrators use $N_T = 12.5 \times 10^3$ time steps. The dash-dotted grey line shows the slope for $\propto N_S^{-1/2}$.

$$\langle f(x) \rangle_{ss} = \frac{\int_{-\infty}^{\infty} f(x)\, e^{x^2 - \frac{2}{3}|x|x^2} \mathrm{d}x}{\int_{-\infty}^{\infty} e^{x^2 - \frac{2}{3}|x|x^2} \mathrm{d}x}. \tag{7.96}$$

We run a simulation identical to the one in Section 7.8.3, where we replace the SDE by Eq. (7.95). An analysis analogous to that in Section 7.8.3 reveals an equally good convergence to the steady-state values at $t = 25$. The results given in Fig. 7.19 show that the overall performance of POS for the non-regular SDE is very similar to that for the nonlinear SDE studied in Section 7.8.3. Despite the fact that Eq. (7.95) is non-regular, we observe virtually the same rate of improvement through POS as we did for the regular SDE.

7.9 Two-dimensional case

Finally, to demonstrate that POS can be applied to SDE's with dimension higher than 1, we briefly consider the so-called laser equation. This is a simplified model for the single-mode quantum statistics in a laser system. Although it can be written as a single complex equation, it is an example of an SDE with dimension higher than one in real

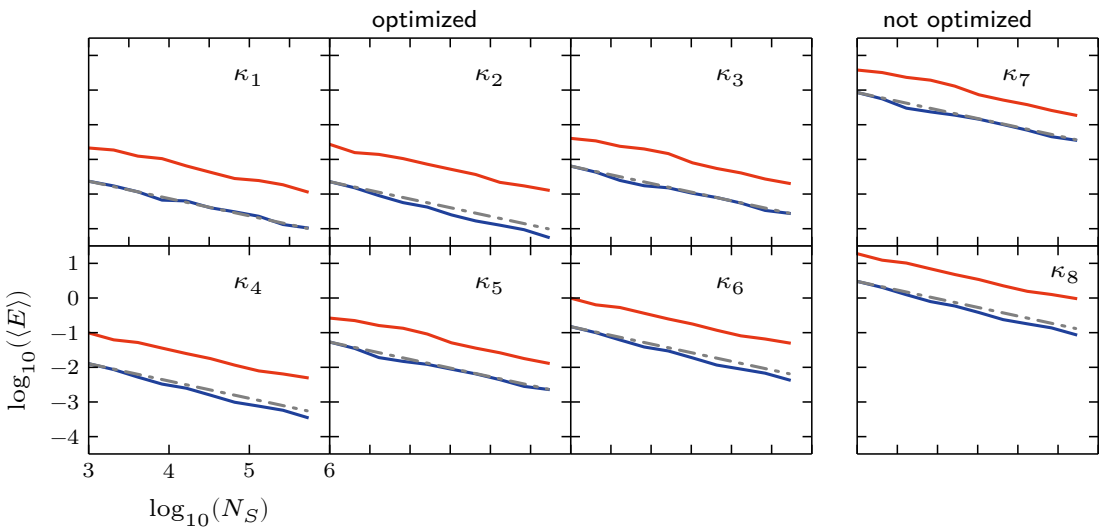

Figure 7.20: Nonlinear SDE case, showing the difference of the first 8 cumulants from the accurate solution at $t = 25$ for the POS integrator using the individual method (solid blue lines) and a reference explicit Euler integrator (solid red lines). Both integrators use $N_T = 12.5 \times 10^3$ time steps. The dash-dotted grey line shows the slope for $\propto N_S^{-1/2}$.

variables. We consider

$$da = \left(1 - |a|^2\right) a \, dt + b \, dW_c, \tag{7.97}$$

where both a and dW_c are complex-valued. The noise obeys $\langle dW_c dW_c^* \rangle = 2dt$. Here a is proportional to the mode amplitude, and b is a real-valued parameter that depends on the steady-state photon number of the laser mode [25]. By treating the real and imaginary parts of a as separate variables, Eq. (7.97) constitutes a two-dimensional real-valued SDE.

The scaled photon number n is given by $n = |a|^2$. A simple calculation reveals the steady-state value for n to be

$$n_{ss} = 1 + \sqrt{\frac{2}{\pi}} \frac{b}{\exp\left[1/\left(2b^2\right)\right]\left[1 + \operatorname{erf}\left(1/\left(\sqrt{2}b\right)\right)\right]} \,. \tag{7.98}$$

Using $N_S = 131072$ trajectories for the real and imaginary part $\Re\{a\}$ and $\Im\{a\}$, resulting in $N_S = 262144$ trajectories in total, we initialize the ensemble with normally-distributed numbers with mean 0 and standard deviation $\frac{1}{\sqrt{2}}$ for $\Re\{a\}$ and $\Im\{a\}$ and carry out a POS-optimized integration using the combined method as well as an explicit Ito-Euler reference integration with an integration time of $t = 25$. An analysis similar to that in Sections 7.8.3 and 7.8.4 reveals an equally good convergence to the steady-state at $t = 25$. We then compare the value of $n = |a|^2$ at $t = 25$ with the steady-state value.

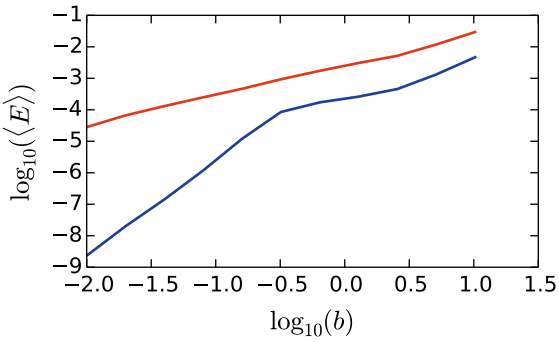

Figure 7.21: Accuracy of the scaled photon number n_s for the laser equation using a reference explicit Euler integrator (red line) and a POS-optimized integration (blue line) as a function of the parameter b. Both integrators use $N_T = 12.5 \times 10^3$ time steps.

7.9.1 Error in the equilibrium photon number

We are optimizing (constraining) every moment and cross-moment for the real and imaginary part of a up to order 4. In other words, we are optimizing the first 4 consecutive moments of $\Re\{a\}$, of $\Im\{a\}$ and cross-moments of the form $\Re\{a\}^n \Im\{a\}^m$, with a combined power $n + m$ not greater than 4. This results in a total of 18 constrained moments. The POS optimization includes the optimization of the initial trajectories.

We vary the parameter b over a range from $1 \cdot 10^{-2}$ to 10.24 in 11 steps. As in Sections 7.8.3 and 7.8.4 we average over 120 independent test runs for the POS integration and 600 runs for the reference integration.

As a typical physical example of a scientifically important quantity in a stochastic calculation, Fig. 7.21 shows the error in calculating n_s, the equilibrium photon number in the laser cavity. Here we use the combined Ito POS method (blue line), and a non-optimized stochastic integration (red line) for the laser equation for different values of b.

This example shows that POS can be successfully used to improve the accuracy of complex-valued SDEs and, more generally speaking, real-valued multivariate SDEs by several orders of magnitude.

7.10 Conclusions

In summary, we have proposed and implemented a novel variance reduction technique for solving stochastic differential equations, using parallel optimized sampling. The essential feature is that it unifies moment hierarchy and independent stochastic methods.

A finite moment hierarchy condition is imposed as a nonlinear constraint on the random noises generated at each step in the integration. This gives a dramatic reduction in sampling error for all moments calculated. In the case of linear equations, we find that the low order moments have sampling errors reduced to machine accuracy. While higher order moments also have their errors reduced, the higher order cumulants are not affected.

For nonlinear equations, the error reduction is not as large, but it occurs over all moments and cumulants studied, even including the non-optimized cumulants of higher order than the optimization limit. We emphasize that the proposed algorithms have a general applicability to all types of stochastic equations, and can be extended in principle to higher order methods as well as the simple Euler integration treated here.

7.11 Acknowledgements

We wish to acknowledge financial support from the Australian Research Council through the Discovery Grant program, as well as useful discussions with Rodney Polkinghorne, Piotr Deuar and Brian Dalton.

7.12 Bibliography

[1]P. Langevin, "Sur la théorie du mouvement brownien", C. R. Acad. Sci. (Paris) **146**, 530–533 (1908) (page 146).

[2]D. S. Lemons, "Paul Langevin's 1908 paper "on the theory of Brownian motion" ["sur la théorie du mouvement brownien," C. R. Acad. Sci. (Paris) 146, 530–533 (1908)]", Am. J. Phys. **65**, 1079 (1997) (page 146).

[3]G. Uhlenbeck and L. Ornstein, "On the theory of the Brownian motion", Phys. Rev. **36**, 823–841 (1930) (page 146).

[4]L. Arnold, *Stochastic differential equations: theory and applications*, reprint (Folens Publishers, 1992), p. 228 (pages 146, 150).

[5]H. Risken, *The Fokker–Planck equation*, 2nd (Springer-Verlag, Berlin, 1996) (pages 146, 147, 165).

[6]C. W. Gardiner, *Handbook of stochastic methods*, 2nd (Springer-Verlag, Berlin, 1985), p. 442 (pages 146, 150, 166).

[7]I. Karatzas and S. Shreve, *Brownian motion and stochastic calculus*, 2nd (Springer, 1991), p. 470 (page 146).

[8]N. G. Van Kampen, *Stochastic processes in physics and chemistry*, 3rd (North Holland, 2007), p. 464 (page 146).

[9]P. Glasserman, *Monte Carlo methods in financial engineering*, reprint (Springer, 2010), p. 596 (page 146).

[10]G. N. Mil'shtein, "Approximate integration of stochastic differential equations", Theor. Probab. Appl. **19**, 557–562 (1975) (page 146).

[11]D. L. Ermak and H. Buckholz, "Numerical integration of the Langevin equation: Monte Carlo simulation", J. Comput. Phys. **35**, 169–182 (1980) (page 146).

[12]D. Talay and L. Tubaro, "Expansion of the global error for numerical schemes solving stochastic differential equations", Stoch. Anal. Appl. **8**, 483–509 (2007) (page 146).

[13]P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations* (Springer-Verlag, Berlin, 1992) (pages 146, 166).

[14]K. Burrage and T. Tian, "A note on the stability properties of the euler methods for solving stochastic differential equations", New Zeal. J. Math. **29**, 115–127 (2000) (page 146).

[15]D. J. Higham, "An algorithmic introduction to numerical simulation of stochastic differential equations", SIAM Rev. **43**, 525–546 (2001) (page 146).

[16]W. J. Morokoff and R. E. Caflisch, "Quasi-Monte Carlo integration", J. Comput. Phys. **122**, 218–230 (1995) (page 147).

[17]M. B. Giles, "Multilevel Monte Carlo path simulation", Oper. Res. **56**, 607–617 (2008) (pages 147, 154).

[18]M. Morillo, J. Gómez-Ordóñez and J. M. Casado, "Checking the validity of truncating the cumulant hierarchy description of a small system", in *Localized excitations in nonlinear complex systems*, Vol. 7, edited by R. Carretero-González, J. Cuevas-Maraver, D. Frantzeskakis, N. Karachalios, P. Kevrekidis and F. Palmero-Acebedo, Nonlinear Systems and Complexity (Springer International Publishing, 2014), pp. 377–387 (pages 147, 184).

[19]P. Hänggi and P. Talkner, "A remark on truncation schemes of cumulant hierarchies", J. Stat. Phys. **22**, 65–67 (1980) (page 147).

[20]A. Gilchrist, C. W. Gardiner and P. D. Drummond, "Positive P representation: Application and validity", Phys. Rev. A **55**, 3014–3032 (1997) (page 147).

[21]P. P. Deuar and P. D. Drummond, "Gauge P representations for quantum-dynamical problems: Removal of boundary terms", Phys. Rev. A **66**, 033812 (2002) (page 147).

[22]T. Aimi and M. Imada, "Gaussian-basis Monte Carlo method for numerical study on ground states of itinerant and strongly correlated electron systems", J. Phys. Soc. Jpn. **76**, 084709 (2007) (page 147).

[23]P. L'Ecuyer, "Efficiency improvement and variance reduction", in Proceedings of the 26th conference on winter simulation (1994), pp. 122–132 (page 164).

[24]M. J. Werner and P. D. Drummond, "Robust algorithms for solving stochastic partial differential equations", J. Comput. Phys. **132**, 312–326 (1997) (page 186).

[25]P. D. Drummond and M. Hillery, *The quantum theory of nonlinear optics* (Cambridge University Press, 2014), p. 380 (page 188).

Chapter 8

Algorithms for integration of stochastic differential equations using parallel optimized sampling in the Stratonovich calculus

Authors: Simon Kiesewetter[1], and Peter D. Drummond[1]

Affiliations: [1]Centre for Quantum-Atom Optics, Swinburne University of Technology, Melbourne, Australia.

Published in: Computer Physics Communications; accepted 1 November 2016

8.1 Abstract

A variance reduction method for stochastic integration of Fokker–Planck equations is derived. This unifies the cumulant hierarchy and stochastic equation approaches to obtaining moments, giving a performance superior to either. We show that the brute force method of reducing sampling error by just using more trajectories in a sampled stochastic equation is not the best approach. The alternative of using a hierarchy of moment equations is also not optimal, as it may converge to erroneous answers. Instead, through Bayesian conditioning of the stochastic noise on the requirement that moment equations are satisfied, we obtain improved results with reduced sampling errors for a given number of stochastic trajectories. The method used here converges faster in

time-step than Ito–Euler algorithms. This parallel optimized sampling (POS) algorithm is illustrated by several examples, including a bistable nonlinear oscillator case where moment hierarchies fail to converge.

8.2 Introduction

Fokker–Plank equations or FPEs [1, 2], combining first order drift and second order diffusion terms, are among the most universally encountered equations in physics. First obtained in work of Fokker, Planck and Langevin [3] on particle motion in a random environment, their use has now extended to many areas of physics and other sciences [4–6]. They are equivalent to stochastic differential equations or SDEs [7–9]. Functional Fokker–Plack equations are equivalent to the increasingly useful stochastic partial differential equations [10].

In this work, we derive and analyze a novel, hybrid approach to solving these equations. Already, there are a variety of methods [5]. The most common, due to its ease of use, is the direct solution of the equivalent stochastic equation through numerical simulation. Yet stochastic methods are greatly restricted by sampling error. As a result, the random error is no better than $1/\sqrt{N}$ for N trajectories. While there are methods to reduce the error due to time discretization [11–15], variance reduction to reduce sampling errors remains problematic [16, 17].

Another, apparently more efficient method, is to use a hierarchy of moment equations [18–20]. This approach requires truncation, as the hierarchy is generically infinite. The truncation approach has a severe drawback, however: truncated hierarchies of moment equations are fundamentally inconsistent, and can have multiple solutions [21]. As a result, moment equation hierarchies often may converge to an incorrect answer [22]. This is a very severe limitation. A truncated hierarchy could appear to converge, yet the answer obtained may be wrong.

We introduce an an alternative approach for solving Fokker–Plack equations called parallel optimized sampling (POS). This unifies the two earlier approaches, giving the efficiency of moment equations while retaining the correctness of stochastic methods. In the present paper, we show how to obtain a POS algorithm with moment equations from

the Stratonovich form of a stochastic equation [8]. This is a common form of equation for a physical stochastic process.

Stratonovich equations have the merit that they are well-suited to numerical integration. They are the broad-band limit of a colored noise - i.e., finite band-width - stochastic process. They follow the rules of standard calculus, and algorithms similar to the usual ordinary differential equation methods can be used. In principle, one can transform these to the Ito form, which is preferred by mathematicians, and then use methods designed for Ito equations. This approach is treated here[23]. However, there can be numerical advantages to using algorithms that are close to ordinary methods of calculus.

In order to demonstrate how these Stratonovich-POS methods work, we analyze the application of this method to some well-known problems: the linear oscillator, the Kubo oscillator, stochastic motion in a bistable well, and a complex-valued laser equation. In the bistable case, the truncated moment hierarchy method gives incorrect moments even when taken to very high order. This illustrates the advantage of the hybrid POS approach. The complex-valued example demonstrates that these methods are applicable in more than one dimension.

We show that POS methods, unifying moment and stochastic methods, not only converge to the correct result, but give orders of magnitude greater accuracy than stochastic methods with the same trajectory number. Rather than trying to obtain precision with more independent trajectories, it is greatly advantageous to use the POS approach with fewer trajectories.

Sampled Monte Carlo methods [24] are the mathematical foundation of many other methods in statistical physics and quantum many-body theory. Hierarchies of moments are also the basis of 'dynamical mean-field' [25], 'cluster expansion' [26] or 'multi-configurational [27] types of approach. Accordingly, it is important to understand such methods. These approaches appear to have the same fundamental limitation, i.e. that direct simulations have large sampling errors, while nonlinear hierarchies of equations may have multiple, incorrect solutions. Thus, while we focus on stochastic equations here, it is not impossible that POS techniques could be useful for treating other statistical problems.

8.3 Fokker–Plack equation

The fundamental problem addressed here is that we wish to find a numerical algorithm to efficiently simulate the observables of a distribution function $P(t, \mathbf{x})$ over a $D-$ $-$dimensional real space, that satisfies the Fokker–Plack equation (FPE) [8],

$$\frac{\partial P(t, \mathbf{x})}{\partial t} = \left[-\frac{\partial}{\partial x_i} a_i(\mathbf{x}) + \frac{1}{2} \frac{\partial^2}{\partial x_i \partial x_j} \mathsf{d}_{ij}(\mathbf{x}) \right] P(t, \mathbf{x}) \ . \tag{8.1}$$

Here we use the Einstein summation convention to sum over doubled indices. In the case of complex variables, the equations can be conveniently re-expressed either by using $2D$ real variables, or equivalently on defining $z_i = x_i$, $z_{i+D} = x_i^*$, , $a_{i+D} = a_i^*$, and extending the summation to $2D$ complex variables. In either case $\mathsf{d}_{ij}(\mathbf{x})$ then becomes a $2D \times 2D$ matrix.

The observables that one wishes to calculate are mean values of functions $\mathbf{o}(\boldsymbol{x}) = (o_1(\boldsymbol{x}), \dots o_M(\boldsymbol{x}))$ at some time t. In this paper, we use the notation $\langle o_m(t) \rangle_\infty$ to indicate mean values calculated from the full probability distribution using a volume measure $d\mathbf{x}$, according to:

$$\langle o_m(t) \rangle_\infty = \int o_m(\mathbf{x}) P(t, \mathbf{x}) \, d\mathbf{x}. \tag{8.2}$$

The corresponding stochastic differential equation (SDE) describes the evolution $\mathbf{x}^{(n)}(t)$ of a number of sampled trajectories of the initial probability, where $n = 1, \dots, N$. In the calculations of mean values, one must therefore make use of a finite ensemble average of N sample trajectories, such that

$$\langle o_m(t) \rangle_N = \frac{1}{N} \sum_n o_m\left(\mathbf{x}^{(n)}(t)\right). \tag{8.3}$$

Any SDE can be written following standard methods using Stratonovich or Ito stochastic calculus [7]. Here we focus on methods that apply parallel optimized sampling (POS) methods to the Stratonovich calculus.

8.3.1 Ito SDE

First, we briefly review the Ito approach to stochastic equations, as this the simplest form. It is also useful in obtaining moment equations. The Ito SDE [7, 8] corresponding to Eq.

(8.1) is given in either real or complex form by:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{a}(\mathbf{x}(t)) + \mathbf{b}(\mathbf{x}(t))\,\boldsymbol{\xi}(t) \tag{8.4}$$

where $\mathbf{x}$ and $\boldsymbol{a}$ are column vectors of real or complex variables, and the diffusion matrix must satisfy $\mathbf{d} = \mathbf{b}\mathbf{b}^T$, where $\mathbf{b}$ is an $M \times M'$ matrix. The Gaussian distributed real noise $\boldsymbol{\xi}$ is a column vector of dimension M', with correlations in the limit of an infinite ensemble given by:

$$\begin{aligned} \left\langle \xi_i(t)\,\xi_j(t') \right\rangle_\infty &= \delta_{ij}\delta(t-t') \\ \left\langle \xi_i(t) \right\rangle_\infty &= 0. \end{aligned} \tag{8.5}$$

In the simplest, Ito–Euler algorithm, one samples the trajectories as a finite set, which is essentially an extended vector:

$$\mathbf{X} = (\mathbf{x}^{(1)}, \ldots \mathbf{x}^{(N)})^T. \tag{8.6}$$

Next, for a finite step in time Δt, the Ito SDE has noise terms evaluated at the starting point, so that for finite differences:

$$\begin{aligned} \mathbf{X}(t+\Delta t) &= \mathbf{X}(t) + \Delta_I \mathbf{X} \\ &= \mathbf{X}(t) + \mathbf{A}(\mathbf{X}(t))\,\Delta t + \mathbf{B}(\mathbf{X}(t))\,\mathbf{W}, \end{aligned} \tag{8.7}$$

where the extended equation coefficients are defined as:

$$\begin{aligned} \mathbf{A}(\mathbf{X}) &= (\mathbf{a}\left(\mathbf{x}^{(1)}\right), \ldots \mathbf{a}\left(\mathbf{x}^{(N)}\right))^T \\ \mathbf{B}(\mathbf{X}) &= diag\left(\mathbf{b}\left(\mathbf{x}^{(1)}\right), \ldots \mathbf{b}\left(\mathbf{x}^{(N)}\right)\right) \\ \mathbf{W} &= (\mathbf{w}^{(1)}, \ldots \mathbf{w}^{(N)})^T, \end{aligned} \tag{8.8}$$

given that $\mathbf{w}^{(n)}$ is a finite time-integral over the noise inputs in the nth sample trajectory:

$$\mathbf{w}^{(n)} = \int_t^{t+\Delta t} \boldsymbol{\xi}^{(n)}(t')\,dt'. \tag{8.9}$$

This is known as a *sampled forward difference* equation, and corresponds to the simplest possible algorithm for solving an SDE. For variable changes, one must use Ito's rules [8], which differ from those of standard calculus.

In the complex variable case, one can write a single complex equation of the form:

$$\frac{d\mathbf{z}(t)}{dt} = \mathbf{a}(\mathbf{z}(t)) + \mathbf{b}(\mathbf{z}(t))\,\boldsymbol{\xi}(t)\,, \tag{8.10}$$

where $\mathbf{z} = \boldsymbol{x} + i\boldsymbol{y}$, and the corresponding real, symmetric $2M \times 2M$ diffusion matrix is:

$$\boldsymbol{d} = \begin{bmatrix} \mathbf{b}^x \\ \mathbf{b}^y \end{bmatrix} \left[\mathbf{b}^{xT}, \mathbf{b}^{yT}\right]. \tag{8.11}$$

8.3.2 Stratonovich SDE

The Stratonovich SDE is an alternative form of stochastic equation which directly corresponds to the infinite bandwidth limit of a physical, noise driven process. It has a conventional behavior under changes of variable. This is written in a similar way, except that when $\mathbf{b}(\mathbf{x})$ is not a constant function, the drift term is modified so that $\mathbf{a} \to \tilde{\mathbf{a}}$, where:

$$\frac{d\mathbf{x}(t)}{dt} = \tilde{\mathbf{a}}(\mathbf{x}(t)) + \mathbf{b}(\mathbf{x}(t))\,\boldsymbol{\xi}(t)\ . \tag{8.12}$$

The Stratonovich SDE is mathematically defined as the limiting behavior of a stochastic process whose derivatives are evaluated at the midpoint $\bar{\mathbf{x}}$, where $\bar{\mathbf{x}} = \mathbf{x} + \Delta_S\mathbf{x}/2$. This subtle difference is important because the noise term is non-differentiable. If the Ito and Stratonovich SDE correspond to the same Fokker–Plack equation, there is a well-known relationship between the two formalisms, which is that

$$\tilde{a}_i \;=\; a_i - \frac{1}{2}\sum_{j,m}\frac{\partial b_{ij}}{\partial x_m} b_{mj}\,. \tag{8.13}$$

All moments can be calculated in either formalism. However, the original Fokker–Plack equation, which contains Ito-type terms directly, is usually more useful for calculating moments analytically.

On the other hand, the Stratonovich SDE is often more useful as a starting point for numerical integration of the stochastic equations, as it can be used in ordinary differential equation methods like midpoint algorithms [12]. In this approach, which generally gives

lower time-discretization and sampling errors than the Ito–Euler method, one uses a finite step defined implicitly as:

$$\Delta_S \mathbf{X} = \tilde{\mathbf{A}}\left(\bar{\mathbf{X}}(t)\right)\Delta t + \mathbf{B}\left(\bar{\mathbf{X}}(t)\right)\mathbf{W}, \tag{8.14}$$

For multiplicative noise terms, numerical experiments show that this midpoint Stratonovich method typically also gives a lower sampling error for the calculated moments [12] than a simple Ito–Euler approach. Many other methods can also be used [11, 13, 15], however, they all have sampling errors.

8.3.3 Observables and moment hierarchies

An alternative approach to solving problems of stochastic evolution is to use a hierarchy of coupled cumulant or moment equations. The general evolution equations for any set of observables $\mathbf{o} = (o_1, \dots, o_M)$ can be written directly from Ito's rule for variable changes, using the fact that noise and variable averages factorize at the same time.

The result for the mean values of a vector $\mathbf{o}$ of stochastic observables in an infinite ensemble is an equation that generally couples moments of different orders to each other:

$$\frac{d\langle \mathbf{o}\rangle_\infty}{dt} = \left\langle a_i(\mathbf{x})\frac{\partial \mathbf{o}}{\partial x_i} + \frac{1}{2}\mathsf{d}_{ij}(\mathbf{x})\frac{\partial^2 \mathbf{o}}{\partial x_i \partial x_j}\right\rangle_\infty . \tag{8.15}$$

For later use in understanding parallel optimization stochastic methods (POS), we can write this equation in a compact form as:

$$\frac{d\langle \mathbf{o}\rangle_\infty}{dt} = \langle \boldsymbol{\mathcal{L}}(\mathbf{x})\rangle_\infty . \tag{8.16}$$

As an example, if the observables are moments x^m of a one-variable equation, then:

$$\langle \mathcal{L}_m(\mathbf{x})\rangle_\infty = \left\langle a(x)\, m x^{m-1} + \frac{m(m-1)}{2} d(x) x^{m-2}\right\rangle_\infty . \tag{8.17}$$

It is important to notice that if either $a(x)$ or $\sqrt{d(x)}$ has nonlinear terms, then the ensemble averages on the right hand side involve higher order moments than m. This

leads to a hierarchy of moment equations which is not closed, which we discuss next in more detail.

To establish our notation for the case of a finite ensemble, we define sampled means as in Eq (8.3), so that:

$$\bar{o}_m\left(\mathbf{X}\left(t\right)\right) = \left\langle o_m(t)\right\rangle_N \equiv \frac{1}{N}\sum_n o_m(\mathbf{x}^{(n)}\left(t\right)) \,. \tag{8.18}$$

In a stochastic integration, the infinite ensemble results are approximated by the sampled observables, with a sampling error of order $1/\sqrt{N}$. We can therefore write the sampled moment differential equations as:

$$\frac{d\bar{\mathbf{o}}\left(\mathbf{X}\right)}{dt} = \bar{\boldsymbol{\mathcal{L}}}\left(\mathbf{X}\right) + O\left(1/\sqrt{N}\right) \,, \tag{8.19}$$

where:

$$\bar{\boldsymbol{\mathcal{L}}}\left(\mathbf{X}\right) = \left\langle a_i\left(\mathbf{x}\right)\frac{\partial \mathbf{o}\left(\mathbf{x}\right)}{\partial x_i} + \frac{1}{2}\mathsf{d}_{ij}\left(\mathbf{x}\right)\frac{\partial^2 \mathbf{o}\left(\mathbf{x}\right)}{\partial x_i \partial x_j}\right\rangle_N \,. \tag{8.20}$$

8.3.4 Convergence of moment hierarchies

Truncated moment hierarchies are sometimes used as a method for solving a Fokker–Plack equation. As explained above, the moment equations are generally not a closed set, except in the case of linear equations. Hence one must have a closure relation when dealing with a computable, finite set of moments. Closure relations are commonly used in many fields employing these methods, ranging from chemistry, physics and engineering to biology, neurology and economics. The most common closure relations assume that either cumulants or moments simply vanish above a certain order. Note that in order for moment hierarchies to be useful, including more moments before such truncation ought to reduce the error below any predetermined amount.

Despite this issue, the moment hierarchy approach is often thought of as a relatively well-controlled approximation. However, this is not always the case. Careful analysis shows that this approach can introduce uncontrolled errors. This problem is caused by the fact that the truncation of a moment hierarchy - usually by setting cumulants above a given order to zero - is generally inconsistent with having a positive probability, except for the special case of a Gaussian distribution [21]. In addition, the steady-state solution

to a truncated moment hierarchy equations can be multivalued. Since there may be infinitely many solutions, a numerical solution can easily converge to the wrong answer.

This problem has been treated in detail in the case of nonlinear, bistable Fokker–Plack equations. In this simple case, it is known that apparently converging, yet incorrect results for moments are still obtained with moment equations taken even up to 60-th order [18, 22]. Similar problems occur in multi-mode generalizations of these equations [28]. We shall use a nonlinear bistable equation as a test case for the algorithm derived here, and show that POS methods not only give correct answers, but have lower sampling errors than the usual stochastic methods for this problem.

8.4 Stratonovich POS methods

Parallel optimization stochastic methods (POS) for an SDE unify the moment and stochastic methods. In this variance reduction method, parallel sets of random trajectories are generated conditionally, depending on the requirement that they satisfy moment equations up to a specified order. The closure of the moment hierarchy is due to the fact that the trajectories provide sampled estimates of moments to all orders.

This greatly reduces the sampling error, which is typically the dominant error term that limits a stochastic integration. The algorithm also removes the moment closure problem that limits moment hierarchy methods. The basic idea is explained elsewhere for Euler type integration of Ito SDEs[23].

Here we analyze results for an implicit central-difference method. This approach is both more robust and has much lower lower step-size errors than Ito–Euler integration, making it more suitable for practical physics applications.

8.4.1 Central difference algorithm

Our main goal is to analyze the optimization of the central difference integration scheme [12] for an SDE in Stratonovich form [7, 8], which for a finite step in time Δt, is written

$$\mathrm{x}(t+\Delta t) = \mathrm{x}^{(0)} + \tilde{\boldsymbol{a}}\left(\bar{\mathrm{x}}\right)\Delta t + \mathbf{b}\left(\bar{\mathrm{x}}\right)\mathrm{w}. \tag{8.21}$$

Here $\mathrm{x}^{(0)} = \mathrm{x}(t)$, $\tilde{\boldsymbol{a}}$ indicates that the functional form of the drift includes a Stratonovich correction, while $\bar{\mathrm{x}}$ is the midpoint, $\bar{\mathrm{x}} = \left(\mathrm{x}(t+\Delta t) + \mathrm{x}(t)\right)/2$, and w is a

vector of Gaussian noises defined so that $\langle w_i w_j \rangle = \delta_{ij}\Delta t$. Without POS optimization, this can simply be treated iteratively, which provides a very robust algorithm for integrating an SDE. It has lower discretization errors than the Euler method. When iterated to convergence - typically requiring $3-4$ iterations - this method can give results with an error as small as $\mathcal{O}(\Delta t^3)$, although this depends on the precise form of the noise term. Generally, it also has lower sampling errors for high order moments than Euler type methods.

We note that many other stochastic methods exist as well as this one [13], including higher order Runge–Kutta methods, and these may give further improvements in terms of truncation or step-size error. However, since our focus is on the question of sampling error - which is generally the largest error one must deal with - the examples used here rely on this simple yet robust technique.

In parallel optimized stochastic (POS) methods, the generated noise fields are conditioned on moment equations of a finite ensemble of parallel trajectories. To analyze this, we define a finite ensemble of trajectories $\mathbf{X}$, as previously, with corresponding drift coefficients and noises, which are all extended column vectors.

If the starting point is $\mathbf{X}^{(0)}$, the resulting sample vector $\mathbf{X}$ after a step Δt will satisfy the following finite difference SDE, now of dimension MN:

$$\mathbf{X} = \mathbf{X}^{(0)} + \tilde{\mathbf{A}}\left(\bar{\mathbf{X}}\right)\Delta t + \mathbf{B}\left(\bar{\mathbf{X}}\right)\mathbf{W}, \tag{8.22}$$

where:

$$\begin{aligned} \bar{\mathbf{X}} &= \frac{1}{2}\left[\mathbf{X} + \mathbf{X}^{(0)}\right] \\ \tilde{\mathbf{A}} &= \left(\tilde{\boldsymbol{a}}\left(\mathbf{x}^{(1)}\right), \quad \cdots \quad \tilde{\boldsymbol{a}}\left(\mathbf{x}^{(N)}\right)\right). \end{aligned} \tag{8.23}$$

This method, by itself, leads to the usual sampling error issues, and next we turn to methods for optimizing moment sampling errors.

8.4.2 Initial ensemble optimization

All stochastic methods require an initial finite sample or ensemble of trajectories. Typically, the trajectory vector is initialized according to some probability distribution, corresponding to the initial condition of the FPE. The finite sampling naturally leads to

sampling errors in the observables $\bar{\mathbf{o}}\left(\mathbf{X}_{in}\right)$, even before the SDE has been integrated in time.

Since we wish to minimize the sampling error at all times, it is essential to start with an initial ensemble of trajectories $\mathbf{X}_{in}$ whose sampling error is reduced. Usually, if the initial probability distribution is known exactly, one can also find the exact expectation values for the observables $\boldsymbol{\mu} \equiv \langle \bar{\mathbf{o}}\left(\mathbf{X}_{in}\right)\rangle$.

We can now use a POS-like algorithm to find $\mathbf{X}_{in}^{opt}$ such that

$$\bar{\mathbf{o}}\left(\mathbf{X}_{in}^{opt}\right) \quad = \quad \boldsymbol{\mu}\,. \tag{8.24}$$

We define $\mathbf{X}_{in}^{(0)}$ as the initial sampling and $\mathbf{X}_{in}^{(i+1)} = \mathbf{X}_{in}^{(i)} + \Delta\mathbf{X}_{in}^{(i)}$. From a Taylor expansion, clearly:

$$\bar{\mathbf{o}}\left(\mathbf{X}_{in}^{(i+1)}\right) \approx \bar{\mathbf{o}}\left(\mathbf{X}_{in}^{(i)}\right) + \mathbf{J}_{in}\left(\mathbf{X}_{in}^{(i)}\right)\Delta\mathbf{X}_{in}^{(i)}\,, \tag{8.25}$$

where

$$J_{ij}^{in}\left(\mathbf{X}\right) = \frac{\partial\bar{o}_i}{\partial X_j} = \bar{o}_{i,j}\left(\mathbf{X}\right)\,. \tag{8.26}$$

Note that we have called the Jacobian $\mathbf{J}_{in}$ here to distinguish it from later Jacobians $\mathbf{J}$ in section 8.4.6.

To solve the resulting equations we note that J_{ij} is a non-square matrix, and can be inverted using the Moore-Penrose pseudo-inverse $\mathbf{J}^{+}$, to give the solution with minimal least squares change.

Eq. 8.25 then leads to the Newton–Raphson iteration prescription

$$\mathbf{X}_{in}^{(i+1)} \quad = \quad \mathbf{X}_{in}^{(i)} + \mathbf{J}_{in}^{+}\left(\mathbf{X}_{in}^{(i)}\right)\left[\boldsymbol{\mu} - \bar{\mathbf{o}}\left(\mathbf{X}_{in}^{(i)}\right)\right]\,. \tag{8.27}$$

Next, we recall that the pseudo-inverse is defined as:

$$\mathbf{J}^{+} = \lim_{\epsilon\to 0}\mathbf{J}^{\dagger}\left[\mathbf{J}\mathbf{J}^{\dagger} + \epsilon\right]^{-1}\,. \tag{8.28}$$

For cases where $\mathbf{u}_{in} = \mathbf{J}_{in}\mathbf{J}_{in}^{\dagger}$ is invertible, we can therefore obtain that:

$$\mathbf{J}_{in}^{+} = \mathbf{J}_{in}^{\dagger}\mathbf{u}_{in}^{-1}, \tag{8.29}$$

which allows us to solve Eq. 8.27 via

$$\mathbf{X}_{in}^{(i+1)} \quad = \quad \mathbf{X}_{in}^{(i)} + \mathbf{J}_{in}^{\dagger}\mathbf{u}_{in}^{-1}\left(\mathbf{X}_{in}^{(i)}\right)\left[\boldsymbol{\mu} - \bar{\mathbf{o}}\left(\mathbf{X}_{in}^{(i)}\right)\right] .$$

This way, the algorithmic complexity is only linear in the ensemble size N, which is of great benefit given that in stochastic simulations, one typically deals with a large number of samples. Note that $\mathbf{u}$ is invertible provided the first derivatives of our observables are linearly independent.

More generally, the solution can be reliably obtained using the Paige-Saunders LSQR algorithm [29].

8.4.3 Moment equations

For POS methods, we assume that the specified moments have their correct values initially, through an initial optimization of the sampled moments as described above. Next, we impose an *additional* requirement on the new trajectory value $\mathbf{X}$ after each step. We require that the sampled observables evolve according to the specified moment equations, that is,

$$\bar{\mathbf{o}}\left(\mathbf{X}\right) - \bar{\mathbf{o}}\left(\mathbf{X}^{(0)}\right) \quad \approx \quad \langle \mathbf{o}(\mathbf{x}, t + \Delta t) - \mathbf{o}(\mathbf{x}, t)\rangle \ . \tag{8.30}$$

This is essentially a constraint on $\mathbf{X} = \Delta\mathbf{X} + \mathbf{X}^{(0)}$. In principle, we could impose this constraint on moments to all orders, but in practice, one can only constrain a finite subset of the moments. Above this order, the moment hierarchy needs a closure relation as usual. Rather than applying an arbitrary truncation, the moments are now estimated by an ensemble average, which exists and is consistent to all orders. In other words, the higher moments needed to close the moment equations are approximated by ordinary sampling of the parallel set of trajectories which are propagated stochastically.

Since we wish to use an accurate approximation to the moment equations, we solve these moment equations using a second-order Runge–Kutta algorithm. This is also known as a collocation or trapezoid method. To describe the approach, we require a stochastic estimate of the trajectory change $\Delta\mathbf{X}$. Once this is obtained, we take:

$$\begin{aligned} \bar{\mathbf{o}}\left(\mathbf{X}\right) - \bar{\mathbf{o}}\left(\mathbf{X}^{(0)}\right) &= \frac{\Delta t}{2}\left[\bar{\boldsymbol{\mathcal{L}}}\left(\mathbf{X}^{(0)}\right) + \bar{\boldsymbol{\mathcal{L}}}\left(\mathbf{X}\right)\right] \\ &= \Delta\bar{\mathbf{o}}\left(\mathbf{X}\right) . \end{aligned} \tag{8.31}$$

We note that this equation requires solving for a new trajectory value $\mathbf{X}$ that must now satisfy both the stochastic equation 8.22 and the moment equation 8.31. However, the stochastic equation has as many equations as unknowns. Accordingly, the noise term must be chosen conditionally on the moment equations, in order for a solution to simultaneously exist for both.

8.4.4 POS optimization

We can now write the Stratonovich equations in the form of a conditional stochastic equation. This is defined so that additional noises $\boldsymbol{\Delta}$ are chosen conditionally, to satisfy the moment equations:

$$\mathbf{X} = \mathbf{X}^{(0)} + \tilde{\mathbf{A}}\left(\bar{\mathbf{X}}\right)\Delta t + \mathbf{B}\left(\bar{\mathbf{X}}\right)\mathbf{W} + \boldsymbol{\Delta}\big|_{\Delta\bar{\mathbf{o}}} . \tag{8.32}$$

This can be carried out most simply as a direct method in which the new noises $\boldsymbol{\Delta}$ are only constrained by moment equations. Alternatively, one can use a Bayesian algorithm, in which $\boldsymbol{\Delta} = \mathbf{B}\left(\bar{\mathbf{X}}\right)\Delta\mathbf{W}$ - hence $\mathbf{B}\left(\bar{\mathbf{X}}\right)\mathbf{W} + \boldsymbol{\Delta} = \mathbf{B}\left(\bar{\mathbf{X}}\right)\tilde{\mathbf{W}}$, and so $\tilde{\mathbf{W}}$ is simply a modification of the original noise term $\mathbf{W}$. In this case, the new noises are just a different choice of the standard noises. The Bayesian approach can be explained using the fact that stochastic equations are effectively a form of sampled Monte-Carlo integration of a path-integral solution to the Fokker–Plack equation, which is now being conditioned on the requirement that it satisfies the moment equations.

In this approach, one method of optimization is to define the probability of observing a noise $\tilde{\mathbf{W}}$, given a required moment $\Delta\bar{\mathbf{o}}$, by the Bayesian formula:

$$P\left(\Delta\tilde{\mathbf{W}}\,|\Delta\bar{\mathbf{o}}\right) = \frac{P\left(\Delta\bar{\mathbf{o}}\,|\tilde{\mathbf{W}}\right)P\left(\tilde{\mathbf{W}}\right)}{P\left(\Delta\bar{\mathbf{o}}\right)}. \tag{8.33}$$

Here $P\left(\Delta\bar{\mathbf{o}}\,|\tilde{\mathbf{W}}\right)$ is the probability of finding the required moment change given a specified noise, while $P\left(\tilde{\mathbf{W}}\right)$ is the a-priori noise probability, and $P\left(\Delta\bar{\mathbf{o}}\right)$ is the moment probability.

The Bayesian approach gives a comb-like, highly peaked distribution, in which the Gaussian background probability is modified by localized peaks where moment equations are satisfied. The localized peaks become more dense as sample numbers increase. Therefore, an algorithm which is asymptotically correct at large sample number is to first sample the underlying Gaussian $P\left(\tilde{\mathbf{W}}\right)$, then correct this with a method which finds the nearest noise that satisfies the moment equation.

From the computational perspective, the direct method is simpler to implement, while the Bayesian method is useful in enforcing any natural conservation laws implicit in the noise choice, but it is generally harder to guarantee convergence. For this reason, we give a derivation here for the direct case, noting that the Bayesian approach has a similar behavior except for additional noise constraints. We give examples both for constant $\mathsf{B}\left(\bar{\mathbf{X}}\right)$ (additive noise) and for $\mathsf{B}\left(\bar{\mathbf{X}}\right)$ not constant (multiplicative noise), demonstrating that results converge to the exact solutions in both cases.

8.4.5 Iterative method for midpoint location

For a POS algorithm, the moment hierarchy solutions must match the stochastic moments to the given order. Dynamically, the two equations to be solved for each step in time are:

$$\begin{aligned} \mathbf{X} - \mathbf{X}^{(0)}(t) - \tilde{\mathbf{A}}\left(\bar{\mathbf{X}}\right)\Delta t - \mathsf{B}\left(\bar{\mathbf{X}}\right)\mathbf{W} - \boldsymbol{\Delta} &= 0 \\ \bar{\mathbf{o}}\left(\mathbf{X}\right) - \bar{\mathbf{o}}\left(\mathbf{X}^{(0)}\right) - \Delta\bar{\mathbf{o}}\left(\mathbf{X}\right) &= 0, \end{aligned} \tag{8.34}$$

where the drift and diffusion are evaluated at the midpoint, $\bar{\mathbf{X}} = \left(\mathbf{X} + \mathbf{X}^{(0)}\right)/2$. Here $\Delta\bar{\mathbf{o}}$ is the moment hierarchy calculated *change*, which depends on the final value, and $\bar{\mathbf{o}}\left(\mathbf{X}\right)$ is the calculated moment vector as a function of the new value $\mathbf{X} = \mathbf{X}(\boldsymbol{\Delta})$.

We rewrite this for convenience as:

$$\begin{aligned} \mathbf{X} - \mathbf{F}_1(\mathbf{X},\boldsymbol{\Delta}) &= 0 \\ \mathbf{c}\,(\mathbf{X}) &= 0\,. \end{aligned} \tag{8.35}$$

In general, we solve the moment equations by varying the noise with a method that gives us a *minimal least-squares* change in the trajectories. The POS condition is a nonlinear vector equation. Both the optimized change $\boldsymbol{\Delta}$ and the new trajectory location $\mathbf{X} = \mathbf{X}(\mathbf{t} + \mathbf{t})$ appear. Since the trajectory is an implicit function of the noise when combined with the implicit midpoint condition, this has the form:

$$\begin{pmatrix} \mathbf{X}(\boldsymbol{\Delta}) - \mathbf{F}_1(\mathbf{X}(\boldsymbol{\Delta}),\boldsymbol{\Delta})) \\ \mathbf{c}\,(\mathbf{X}(\boldsymbol{\Delta})) \end{pmatrix} = 0. \tag{8.36}$$

A number of strategies are possible. An important issue is that the dimension of the moment equations is typically smaller than that of the trajectory equations, and the system of equations is underdetermined. This means that it is numerically most efficient to use simple function iteration to solve the trajectory equations, combined with a more sophisticated root-finding method for the moment equations.

We therefore use an iterative method for $\mathbf{X}^{(n)},\boldsymbol{\Delta}^{(n)}$, although other methods may be employed as well, as long as they converge to the nearest possible set of trajectories that satisfy the POS equations. First we define $\mathbf{X}^{(0)} = \mathbf{X}(t)$ as the starting point, and $\boldsymbol{\Delta}^{(0)} = 0$ as the initial correction term. The update for the trajectory is given using a functional iteration method, by:

$$\mathbf{X}^{(n)} = \mathbf{F}_1(\mathbf{X}^{(n-1)},\boldsymbol{\Delta}^{(n)})\,. \tag{8.37}$$

This locates the next approximate trajectory value, $\mathbf{X}^{(n)} = \mathbf{X}\left(\boldsymbol{\Delta}^{(n)}\right)$ and hence allows an estimate of the midpoint position.

8.4.6 Optimization

Next, the optimized noise term is calculated, using an iterative Newton–Raphson method. To obtain this, we Taylor expand the moment equations around the optimum point where $\mathbf{c}(\mathbf{X}) = 0$, noting that $\mathbf{X}$ depends on the noise:

$$\mathbf{c}(\mathbf{X}(\boldsymbol{\Delta})) = \mathbf{J}\left[\boldsymbol{\Delta} - \boldsymbol{\Delta}^{opt}\right] + O\left[\left|\boldsymbol{\Delta} - \boldsymbol{\Delta}^{opt}\right|^2\right], \tag{8.38}$$

where we define the Jacobian using the chain rule, as:

$$J_{ij} = \sum_k \frac{\partial c_i}{\partial X_k}\frac{\partial X_k}{\partial \Delta_j}. \tag{8.39}$$

The moments and the Jacobian depend on the optimized noise $\boldsymbol{\Delta}$ via the functional dependence of $\mathbf{X}$ on $\boldsymbol{\Delta}$, and can be written in a compact form as functions of the noise:

$$\mathbf{c}(\boldsymbol{\Delta}) = \bar{\mathbf{o}}(\mathbf{X}(\boldsymbol{\Delta})) - \bar{\mathbf{o}}\left(\mathbf{X}^{(0)}\right) - \Delta\bar{\mathbf{o}}(\mathbf{X}(\boldsymbol{\Delta})). \tag{8.40}$$

Hence, using the chain rule:

$$J_{ij}(\boldsymbol{\Delta}) = \sum_k \left[\bar{o}_{i,k}(\boldsymbol{\Delta}) - \Delta\bar{o}_{i,k}\right]\frac{\partial X_k}{\partial \Delta_j}, \tag{8.41}$$

where the derivatives of X_k with respect to the noise are obtained from the first line of Eq (8.34):

$$\frac{\partial X_k}{\partial \Delta_j} = \frac{1}{2}\sum_l \left[\tilde{A}_{k,l}\Delta t + B_{km,l}W_m\right]\frac{\partial X_l}{\partial \Delta_j} + \delta_{kj}. \tag{8.42}$$

For simplicity, we drop small terms of order Δt and W, although these can be retained to give more rapid convergence. This gives the result that:

$$J_{ij} \approx \bar{o}_{i,j} + O\left(\sqrt{\Delta t}\right) \tag{8.43}$$

where $\bar{o}_{i,j}$ is a matrix of the moment derivatives, which is a function of the optimized noise $\boldsymbol{\Delta}$.

The Newton–Raphson iteration equations are then of the form:

$$\mathbf{c}\left(\boldsymbol{\Delta}^{(n)}\right) = \mathbf{J}\left(\boldsymbol{\Delta}^{(n)}\right)\left[\boldsymbol{\Delta}^{(n)} - \boldsymbol{\Delta}^{(n+1)}\right] \tag{8.44}$$

hence:

$$\boldsymbol{\Delta}^{(n+1)} = \boldsymbol{\Delta}^{(n)} - \mathbf{J}^{+}(\mathbf{X}^{(n)})\,\mathbf{c}(\mathbf{X}^{(n)})\,. \tag{8.45}$$

where $\mathbf{J}^{+}$ denotes the pseudo-inverse as mentioned in chapter 8.4.2. As described previously, Eq. 8.45 can be calculated via

$$\boldsymbol{\Delta}^{(n+1)} = \boldsymbol{\Delta}^{(n)} - \mathbf{J}^{\dagger}\mathbf{u}^{-1}(\mathbf{X}^{(n)})\,\mathbf{c}(\mathbf{X}^{(n)}) \tag{8.46}$$

where $\mathbf{u} = \mathbf{J}\mathbf{J}^{\dagger}$ provided the first derivatives of the observables are linearly independent.

Rather than carrying out each iteration separately, we find it is efficient to use a combined functional and Newton–Raphson iterative solution of:

$$\begin{aligned} \mathbf{X}^{(n)} &= \mathbf{F}_1(\mathbf{X}^{(n-1)}, \boldsymbol{\Delta}^{(n)}) \\ \boldsymbol{\Delta}^{(n+1)} &= \boldsymbol{\Delta}^{(n)} - \mathbf{J}^{+}(\mathbf{X}^{(n)})\,\mathbf{c}(\mathbf{X}^{(n)})\,. \end{aligned} \tag{8.47}$$

8.4.7 Code examples

In order to better explain the Stratonovich POS algorithm, we present a code listing written in pseudocode.

The algorithm presented will advance the N_S-dimensional trajectory vector $\mathbf{x}$ subject to the stochastic differential equation in Stratonovich form $\dot{\mathbf{x}}(t) = \mathbf{a}(\mathbf{x}(t)) + \mathbf{b}(\mathbf{x}(t))\ (t)$ by one time-step Δt while constraining the vector of observables $\mathbf{o}(\mathbf{x}(t))$. The moment propagator $\bar{\mathcal{L}}(\mathbf{x})$, as outlined in section 8.3.3, describes the infinitesimal time-evolution of $\mathbf{o}(\mathbf{x}(t))$ and can be found using Ito's lemma. The matrix $\mathbf{J}(\mathbf{x})$ represents the Jacobian matrix of the observable vector as a function of the trajectories $J_{ij} = \partial o_i / \partial x_j$. The vector $\boldsymbol{W}$ is a vector consisting of normally-distributed random numbers with mean 0 and variance Δt, while the vector $\boldsymbol{\Delta}$ represents the correction term.

The Stratonovich POS algorithm works in the following way:

```
1  function StratPOS_step(x, W)
2  {
3     Δ ← 0
4     o_0 ← o(x)
5     L̄_0 ← L̄(x)
6     x_i ← x + a(x)Δt + b(x)W
7     while (stopping criterion not met)
8     {
9        x̄_i ← (x + x_i)/2
10       x_i ← x + a(x̄_i)Δt + b(x̄_i)W + Δ
11       L̄_i ← L̄(x_i)
12       õ_i ← o_0 + (L̄_0 + L̄_i)Δt/2
13       Δo_i ← õ_i − o_i
14       J_i ← J(x_i)
15       Δx_i ← J_i^† ((J_i J_i^†)^{-1} Δo_i)
16       x_i ← x_i + Δx_i
17       Δ ← Δ + Δx_i
18    }
19    x ← x_i
20 }
```

During each iteration in the while-loop, a new midpoint $\bar{\mathbf{x}}_i$ is calculated in line 10, resulting in a new approximation for the final time-evolved trajectory vector $\mathbf{x}_i$ in line 11. This trajectory vector is then subject to a new moment equation which stipulates $\Delta\mathbf{o}_i$ (calculated in line 13) to be zero. This moment equation results is a least-squares correction $\Delta\mathbf{x}_i$ to the trajectory vector $\mathbf{x}_i$. Note that the matrix inversion in line 15 is usually not a numerically expensive operation, because $\mathbf{J}\mathbf{J}^\dagger$ is a square matrix where the number of rows and column is equal to the number of elements in $\mathbf{o}$, rather than the total number of samples or trajectories - which is typically much larger still. Furthermore, the matrix $\mathbf{J}\mathbf{J}^\dagger$ is invertible if $\mathbf{J}$ has full rank, in other word if the derivatives of all observables are linearly independent. The vector $\boldsymbol{\Delta}$ is the accumulated change in the trajectory vector for every iteration.

A good stopping criterion has proven to be a lower limit for the mean square of the correction in $\mathbf{x}_i$, which can be obtained from $\Delta\mathbf{x}_i$, but other choices are possible as well, such as a fixed number of iterations.

Note that because the algorithm is partially based on a Newton–Raphson iteration, if the initial guess for $\mathbf{x}_i$ lies outside of the basin of attraction, the algorithm will not converge. This is always indicated by a rising mean square in $\Delta\mathbf{x}_i$. In such a case, one can reduce the time-step successively until the algorithm converges. It is important to use the same noise-vector and maintain consistency with the initial random number generating seed for reduced time-steps in order to prevent the introduction of a bias by simply avoiding certain noise-vectors.

8.5 Linear Examples

In this section, we give linear examples that illustrate the POS technique. We start with examples of linear additive and multiplicative stochastic equations. In these cases moment evolution equations can be exactly obtained. In addition, moment hierarchies are closed, and in principle could be used directly. However, these are useful cases for testing performance and illustrating the POS method.

8.5.1 Linear oscillator

The linear oscillator stochastic equation has additive noise, and in its simplest form it has $\tilde{a} = a = -x$, $b = 1$. Hence, the Ito and Stratonovich equations are identical:

$$dx = -x \cdot dt + dw\,. \tag{8.48}$$

For this system, the moment hierarchy is closed, so it is straightforward to find the analytic solution for the moments. An outline of this can be found in the Appendix.

Fig. 8.1 shows the time evolution of the first moment of a linear oscillator with initial mean $x_0 = 0.5$ and standard deviation $\sigma_0 = 0$, integrated using different methods with $N_S = 200$ trajectories and a step-size of $\Delta t = 2 \cdot 10^{-2}$. For Fig. 8.1(a), we have integrated the stochastic equations directly, using a conventional Runge–Kutta Monte Carlo method, more precisely a two-step Heun's method.

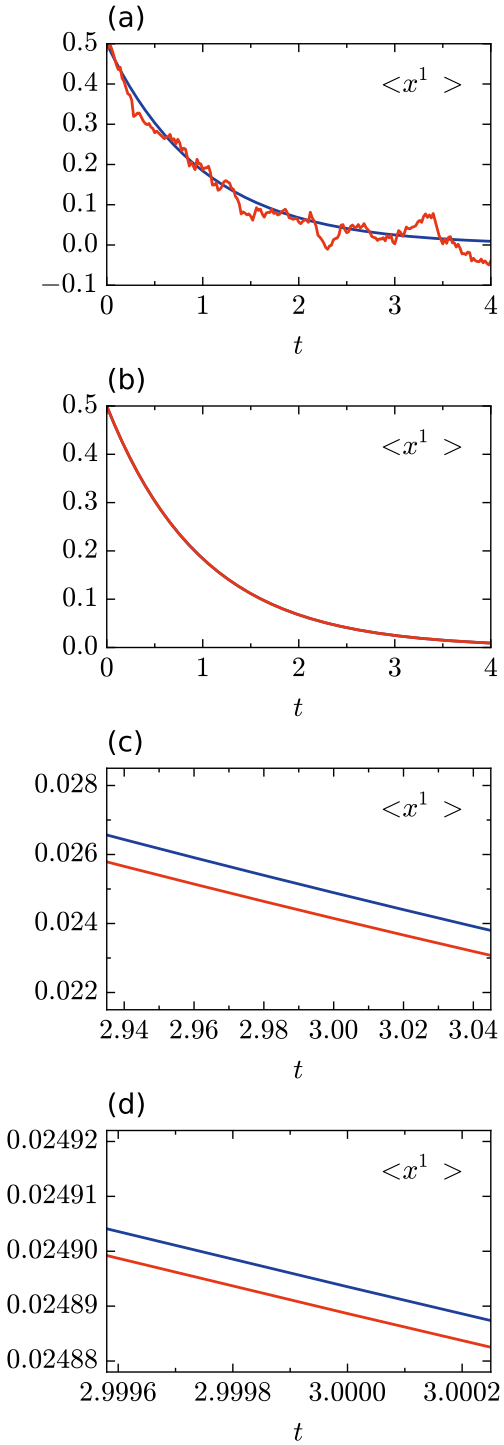

Figure 8.1: (a) Simulation results for the linear oscillator obtained using a non-optimized Runge–Kutta method for $N = 200$ trajectories (red line) and analytical value (blue line). (b) Simulation results for the same system, but using an Ito version of POS, for the same number of trajectories. (c) Magnification of the results in (b). (d) Simulation results for the same system, but using the Stratonovich version of POS described here, for the same number of trajectories, magnified. Note the greatly increased accuracy of (d) compared to (c).

For Fig. 8.1(b), we have used an Ito version of POS with 5 POS iterations per time-step and an initial ensemble optimization. The Ito version of POS is similar to the method described here, but it is based on an Ito–Euler integration scheme instead of a Stratonovich one. A full description will be published elsewhere. Fig. 8.1(c) shows a magnification of Fig. 8.1(b) to highlight the accuracy.

Lastly, in Fig. 8.1(d), we have used the Stratonovich version of POS with 5 POS iterations per time-step an initial ensemble optimization. As in Fig. 8.1(c), we have magnified the resulting graph to highlight the greater accuracy achievable with this algorithm, due to its higher order convergence in time. In these results, the POS methods eliminate sampling errors completely, leaving only residual errors due to the finite step-size in time.

8.5.2 Kubo oscillator

The Kubo oscillator is a complex, multiplicative noise equation with a stochastic frequency. It is a Stratonovich equation with $\tilde{a} = 0$, $b = iz$:

The Stratonovich equation is:

$$dz = izdw\,. \tag{8.49}$$

The Stratonovich correction is:

$$\delta_S a_i \equiv -\frac{1}{2}\frac{\partial b}{\partial z} b = \frac{1}{2} z\,. \tag{8.50}$$

Hence the Ito equation is:

$$dz = -\frac{1}{2} zdt + izdw\,. \tag{8.51}$$

In real variables form, one has:

$$\begin{aligned} b^x &= -y \\ b^y &= x\,, \end{aligned} \tag{8.52}$$

hence the real diffusion matrix is:

$$\boldsymbol{d} = \begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix}. \tag{8.53}$$

As for the linear oscillator, the Kubo oscillator has a closed moment hierarchy. An outline of its solution can be found in the appendix.

Kubo stochastic integration with POS

The simplest way to treat this as a stochastic equation is to transform the Stratonovich equation to a minimal dimensional subspace, namely the phase variable:

$$\begin{aligned} \dot{\theta} &= \zeta(t) \\ z &= z_o e^{i\theta}. \end{aligned} \tag{8.54}$$

With this approach, the amplitude moments become invariant to all orders even after sampling, so we only have to constrain the analytic moments. Note, if

$$o_m = z^m \tag{8.55}$$

then:

$$J_{mk} = \frac{\partial \bar{o}_m}{\partial \theta^{(k)}} = \frac{im}{N}\left[z^{(k)}\right]^m . \tag{8.56}$$

To make this more definite, suppose we only wish to optimize real moments like $\Re(z^m)$. Then:

$$\begin{aligned} J_{mk} &= \frac{\partial \bar{o}_m}{\partial \theta^{(k)}} \\ &= \frac{\partial}{\partial \theta^{(k)}} \frac{m}{2N}\left\{ i\left[z^{(k)}\right]^m - i\left[z^{(k)*}\right]^m \right\} \\ &= \frac{m}{N}\Re\left[iz^{(k)}\right]^m . \end{aligned} \tag{8.57}$$

POS simulation results

Fig. 8.2 shows the accuracy that was achieved integrating the Kubo oscillator using a conventional Monte-Carlo algorithm (red line) and using the Stratonovich POS method (blue line) as a function of the number of trajectories used. Using the transformation $z_k = e^{i\theta_k}$, we simulate the Kubo oscillator by integration the real-valued phase angles θ_k. This way, we reduce the complex-valued equation of motion for the trajectories into a

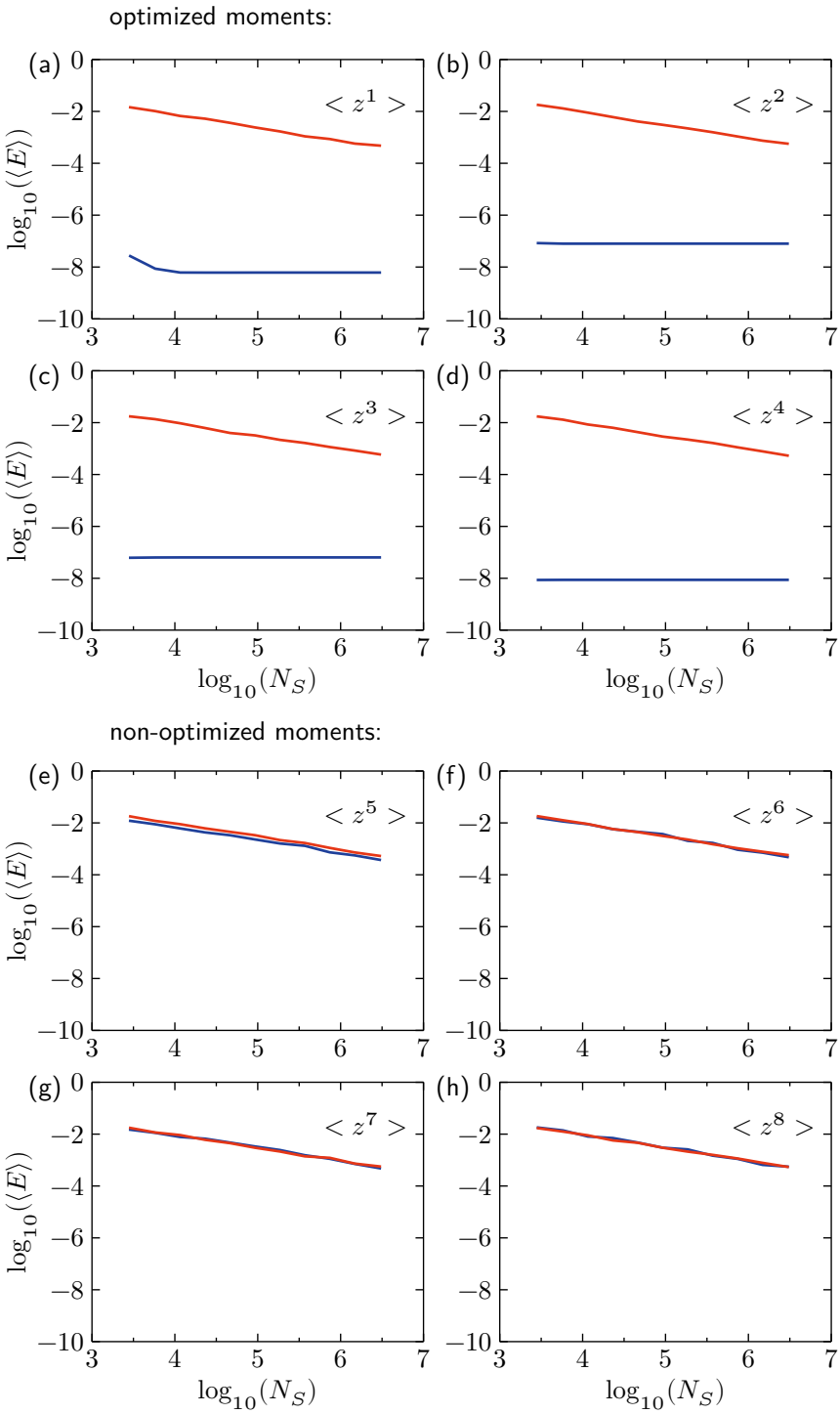

Figure 8.2: Accuracies in the moments for a non-optimized midpoint method Stratonovich simulation (red line) and the Stratonovich midpoint POS simulation (blue line) for the Kubo oscillator as a function of the number of trajectories in a double-logarithmic plot. The POS simulation was carried out optimizing the real parts and the imaginary parts of the first 4 moments, resulting in 8 optimized quantities in total and included the optimization of the initial trajectories.

real-valued equation of motion for the phase angles, thus reducing the dimensionality of the system.

We initialize the phase angles with mean $\theta_0 = 0.1$ and standard deviation $\sigma_{\theta,0} = 0.25$. The step-size is $\Delta t = 2.5 \cdot 10^{-3}$ and the simulation time is $t_{max} = 1$.

The non-optimized integration was carried out using a semi-implicit midpoint method with 3 iteration steps per time-step [12], while the POS-optimized integration was carried out using the Stratonovich POS algorithm with 5 POS iterations per time-step and an initial ensemble optimization. We optimize the real parts as well as the imaginary parts of the first 4 consecutive moments of z, resulting in a total of 8 optimized (constrained) observables.

The quantity $\langle E \rangle$ plotted in Fig. 8.2 is the absolute deviation of the numerically obtained m-th moment in z from its theoretical expectation value at $t = t_{max}$. Because the moments are complex-valued, we use the absolute value of the deviation. In order to reduce statistical fluctuations, several values for E were obtained and averaged to obtain $\langle E \rangle$. In short, the accuracy for the m-th moment, $\langle E^{(m)} \rangle$, is obtained via $\langle E^{(m)} \rangle \equiv \left\langle \sqrt{\left\langle E_r^{(m)} \right\rangle^2 + \left\langle E_i^{(m)} \right\rangle^2} \right\rangle$, where $\langle E_r^{(m)} \rangle \equiv \left\langle \left| \langle \Re\left(z^m\right) \rangle - \langle \Re\left(z^m\right) \rangle^{(exact)} \right| \right\rangle$ and $\left\langle E_i^{(m)} \right\rangle \equiv \left\langle \left| \langle \Im\left(z^m\right) \rangle - \langle \Im\left(z^m\right) \rangle^{(exact)} \right| \right\rangle$.

The POS-optimization results in an impressive increase of accuracy in the optimized moments of up to more than 6 orders of magnitudes for the same number of trajectories. The accuracy achieved is largely independent of the number of trajectories used.

It is not surprising that, since there is no coupling of different orders of moments, only the optimized moments have reduced sampling errors. The non-optimized moments see relatively little change in their sampling errors.

8.6 Nonlinear stochastic examples

Next, we treat cases involving nonlinear stochastic equations. In these cases the moment hierarchies are not closed, and therefore there is the possibility that sampling error can 'leak' into the optimized moments via the coupling of moment equations to each other. It is important to understand the consequence of this. As we see, sampling errors are still reduced, and in fact even the non-optimized moments are improved.

8.6.1 Bistable nonlinear oscillator

We consider a nonlinear oscillator with additive noise. This widely studied case is either bistable or monostable, depending on the nonlinear term. The form treated here is bistable: it has $\tilde{a} = a = x - x^3$, $b = 1$. Unlike the previous two examples, a finite set of moment equations is not closed, so this example tests a situation where the moment hierarchy closure is obtained from coupling to the high-order stochastic moments obtained from sampling, but which are not themselves optimized through POS techniques.

Stochastic equations

As the noise is additive, the Ito and Stratonovich equations are therefore identical:

$$dx = \left(x - x^3\right)\, dt + dw\,. \tag{8.58}$$

The moment propagator is:

$$\mathcal{L}_m\,(x) \quad = \quad m(x^m - x^{m+2}) + \frac{1}{2}m(m-1)x^{m-2}\,. \tag{8.59}$$

In Fig. 8.3, we show the accuracy achieved for the nonlinear oscillator using a conventional Stratonovich method (red line) and the Stratonovich POS method (blue line) as a function of the number of trajectories used. The simulations were carried out using a step-size of $\Delta t = 2 \cdot 10^{-3}$, a maximum time of $t_{max} = 10$, an initial mean of $x_0 = 0$ und an initial standard deviation of $\sigma_0 = 0.9$. As in section 8.5.2, the non-optimized integration was carried out using a 3-step semi-implicit midpoint method. The POS-optimized integration was carried out using the Stratonovich-type POS method described here with 5 iterations per time-step. Integrating up to $t_{max} = 10$ ensures that the resulting moments are sufficiently close to the steady-state value of the nonlinear oscillator. The steady-state value can be obtained analytically via

$$\langle x^m \rangle^{(ss)} = \frac{\int_{-\infty}^{\infty} x^m e^{(x^2 - x^4/2)} dx}{\int_{-\infty}^{\infty} e^{(x^2 - x^4/2)} dx}\,. \tag{8.60}$$

Using the steady-state value as a sufficiently good approximation of the exact value, we then define the accuracy $\langle E \rangle$ analogously to section 8.5.2, the only difference being

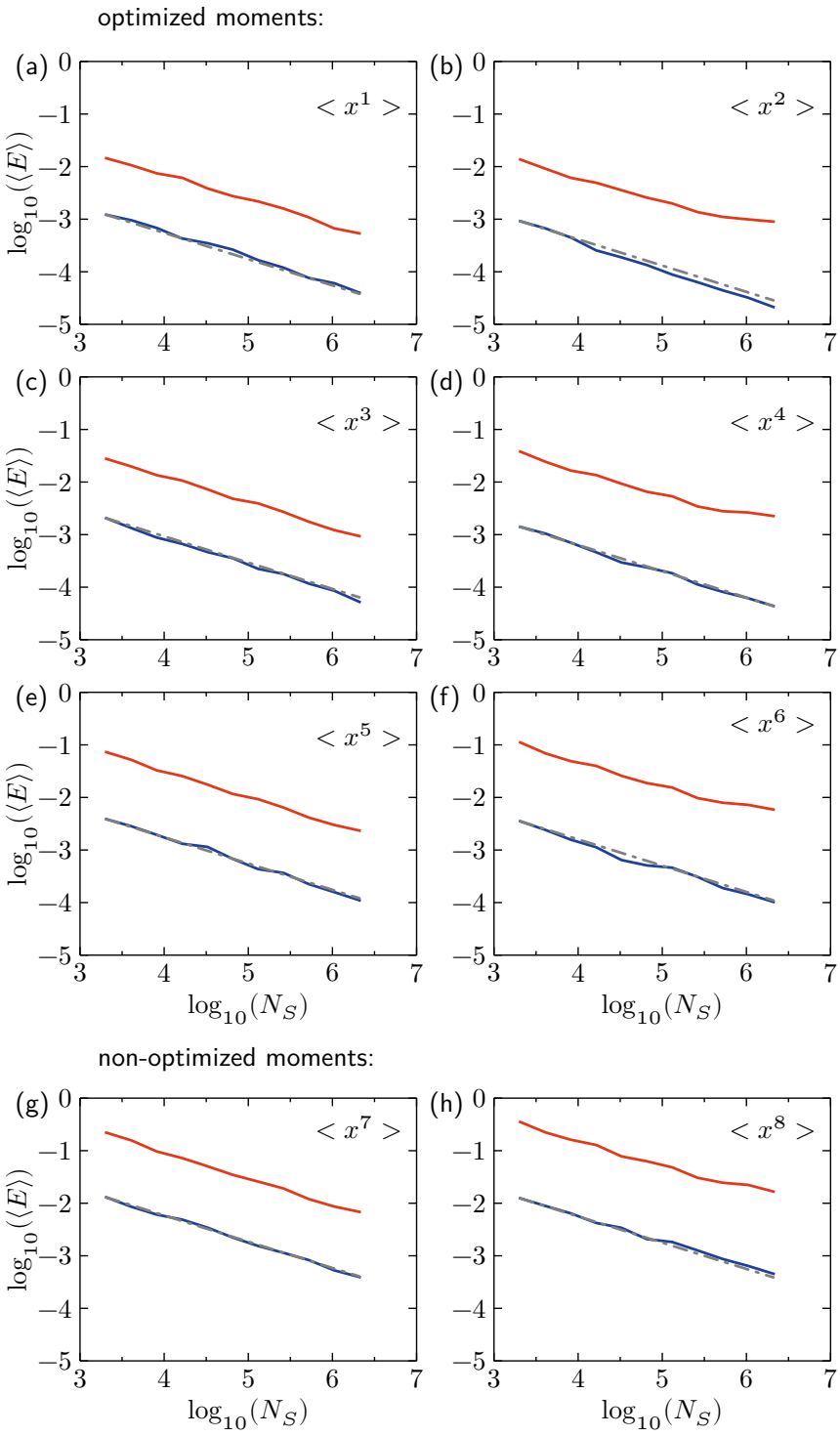

Figure 8.3: Accuracies in the moments for a non-optimized midpoint method Stratonovich simulation (red line) and a Stratonovich midpoint POS simulation (blue line) for the nonlinear oscillator as a function of the number of trajectories in a double-logarithmic plot. The simulation uses a step-size of $\Delta t = 2 \cdot 10^{-3}$, initial conditions $x_0 = 0$, $\sigma_0 = 0.9$ and a simulation time of $t_{max} = 10$. The POS simulation was carried out optimizing the first 6 moments using 5 iterations per time-step as well as optimizing the initial ensemble.

that the moment are purely real for this problem. In short, the accuracy for the m-th moment $\langle E^{(m)} \rangle$ is defined as $\langle E^{(m)} \rangle = \left\langle \left| \langle x^m \rangle - \langle x^m \rangle^{(exact)} \right| \right\rangle$.

The grey dash-dotted lines in Fig. 8.3 is superimposed and represents a $N_s^{-\frac{1}{2}}$ slope. Note that using the POS method, the error can be reduced by significantly more than one order of magnitude for the same number of trajectories. The accuracy of POS in this case exhibits a clear $N_s^{-\frac{1}{2}}$ scaling, indicating that although the sampling error is greatly reduced, it is limited by the residual sampling error of the non-optimized moments.

8.6.2 Cumulant hierarchy method

For this bistable equation, what is the accuracy achieved with the truncated cumulant hierarchy method?

Truncating the cumulant hierarchy of the nonlinear oscillator for orders higher than 6 results in the following equations of motion:

$$
\begin{aligned}
\dot{k}_1 &= -k_3 - 3k_1k_2 - k_1^3 + k_1 \\
\dot{k}_2 &= -2k_4 - 6k_1k_3 - 6k_2^2 + \left(2 - 6k_1^2\right) k_2 + 1 \\
\dot{k}_3 &= -3k_5 - 9k_1k_4 + \left(-27k_2 - 9k_1^2 + 3\right) k_3 - 18k_1k_2^2 \\
\dot{k}_4 &= -4k_6 - 12k_1k_5 + \left(-48k_2 - 12k_1^2 + 4\right) k_4 - 36k_3^2 \\
&\quad -72k_1k_2k_3 - 24k_2^3 \\
\dot{k}_5 &= -15k_1k_6 - 75k_2k_5 - 15k_1^2k_5 + 5k_5 - 150k_3k_4 \\
&\quad -120k_1k_2k_4 - 90k_1k_3^2 - 180k_2^2k_3 \\
\dot{k}_6 &= -108k_2k_6 - 18k_1^2k_6 + 6k_6 - 270k_3k_5 - 180k_1k_2k_5 \\
&\quad -180k_4^2 - 360k_1k_3k_4 - 360k_2^2k_4 - 540k_2k_3^2 \,.
\end{aligned}
\tag{8.61}
$$

Figs. 4 and 5 show the accuracy of the cumulant hierarchy method (blue line), together with that of a non-optimized stochastic integration (green line) and the Stratonovich POS method (red line), for the nonlinear oscillator as a function of time. Fig. 8.4 shows the accuracy in the moments while Fig. 8.5 that of the corresponding cumulants. The system is initialized with a mean of $x_0 = 0.25$ and standard deviation $\sigma_0 = 0.5$. The cumulant

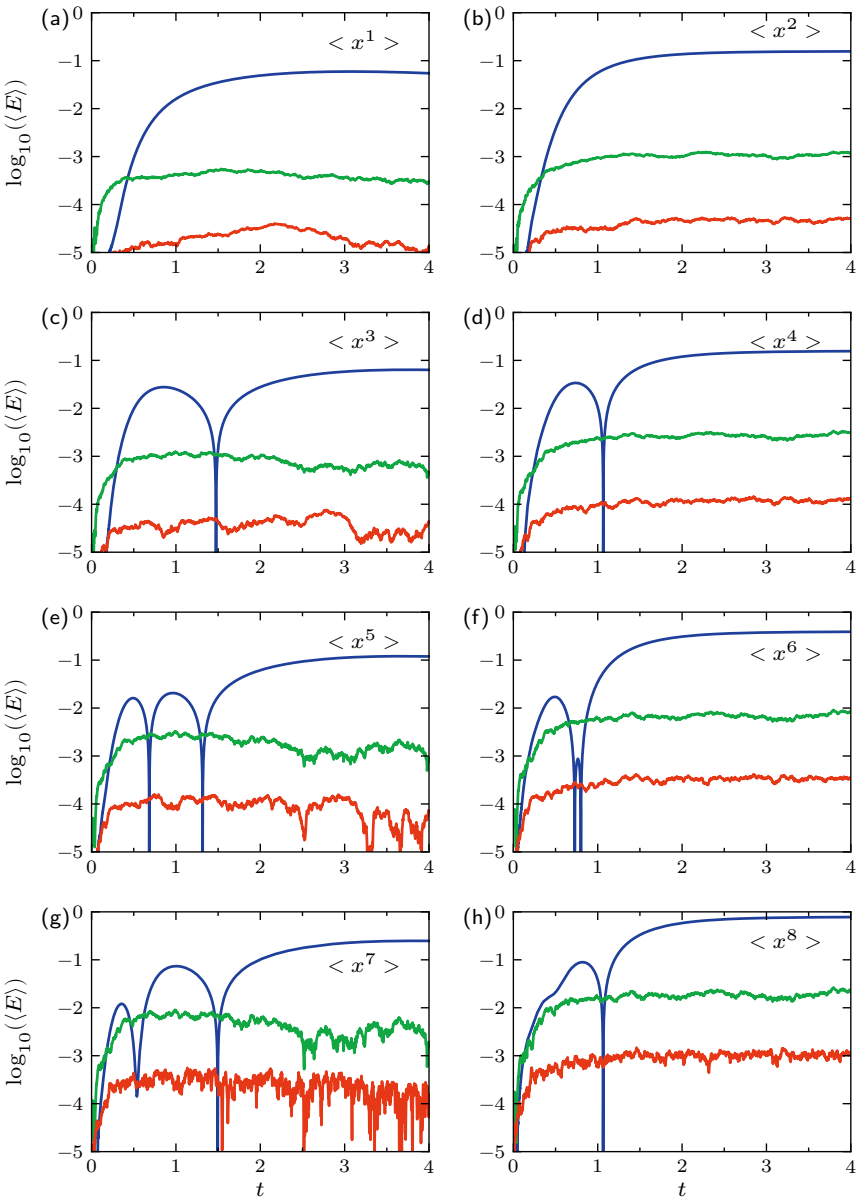

Figure 8.4: Accuracy of the moments for the nonlinear oscillator using a truncated cumulant hierarchy method to order 6 (blue line), a non-optimized stochastic integration (green line) and a POS-optimized stochastic integration, optimizing 6 moments (red line) as a function of time. Low error excursions of the moment hierarchy are due to crossings of exact and computed values.

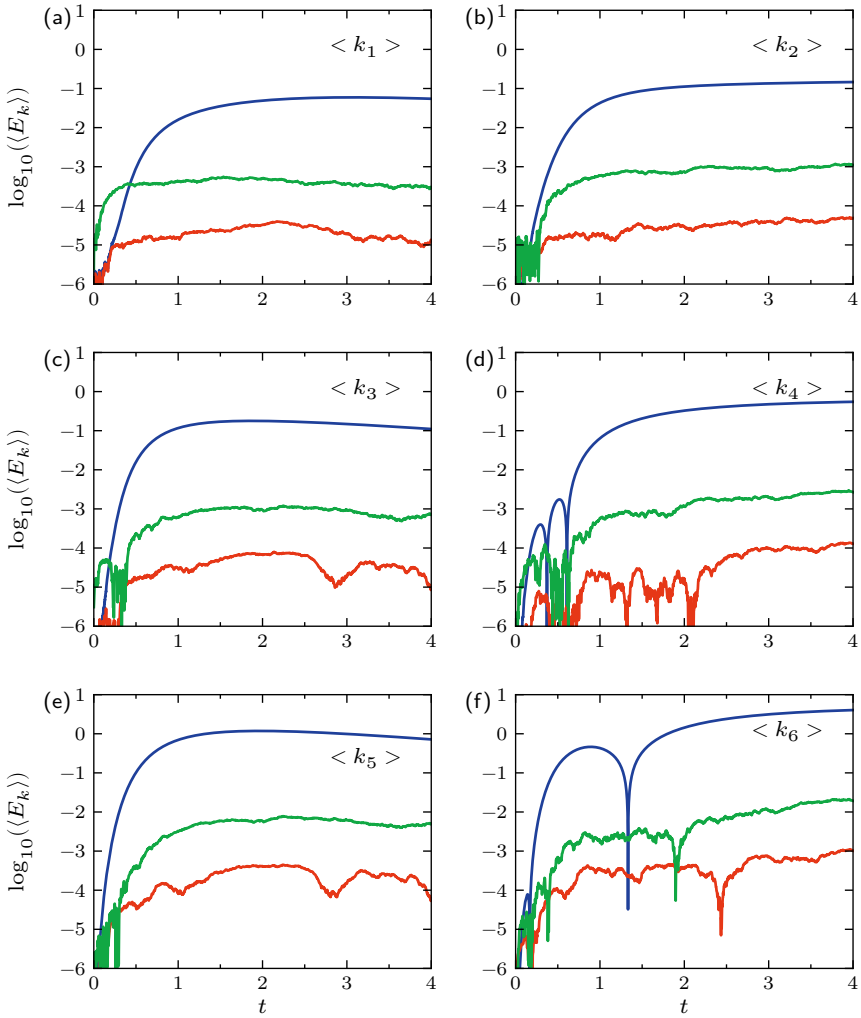

Figure 8.5: Accuracy of the cumulants for the nonlinear oscillator using a truncated cumulant hierarchy method to order 6 (blue line), a non-optimized stochastic integration (green line) and a POS-optimized stochastic integration, optimizing 6 moments (red line) as a function of time.

hierarchy was integrated using a conventional 4th order Runge–Kutta algorithm with a time-step of $\Delta t = 10^{-5}$. The stochastic integrations were carried out using $N = 524288$ trajectories and a time-step of $\Delta t = 2.5 \cdot 10^{-3}$. For the non-optimized stochastic integration, we used a 3-step semi-implicit midpoint method. For the POS-optimized integration, we used the Stratonovich-type POS method described here optimizing (constraining) the first 6 consecutive moments.

In order to have an approximation of the exact solution of the non-linear oscillator, we carried out another stochastic integration using a 3-step midpoint method, but this time with a time-step of $\Delta t = 10^{-4}$ and $N = 2.5 \cdot 10^{10}$ trajectories. Due to the low step-size and extremely high number of trajectories, we can use the result of this simulation as a sufficiently good approximation of the exact (theoretical) value, for which we don't have an analytical expression. As previously, we define the accuracy of the simulation results as the mean absolute deviation from the exact value.

We see that the conventional stochastic integration method results in several orders of magnitude better accuracy when compared to the cumulant hierarchy method. The reason for this is that the cumulant hierarchy method simply doesn't converge to the correct answer. Even when taken to sixtieth order - much higher than the present results [22] - the steady-state value for the second-order moment is known to remain at variance with the correct value. In the results calculated here, we show how this discrepancy evolves in time, and also that it is steadily worse for higher order cumulants. The errors exceed 100% for the sixth-order cumulant, when this is the largest term included in the truncation scheme.

Using the POS-optimized method not only is seen to converge to the correct results for all moments calculated, but this method also increases the accuracy significantly more than stochastic methods with the same number of trajectories.

8.6.3 Bistable oscillator with multiplicative noise

Next, we consider the case of multiplicative noise, where the diffusion is a function of position. We use the same drift term as in section 8.6.1, but choose a diffusion term of $b(x) = \sqrt{1 + x^2}$. The Ito equation of motion is

$$dx_{(I)} = \left(x - x^3\right) dt + \sqrt{1 + x^2}\, dw \tag{8.62}$$

and the Stratonovich equation of motion is

$$dx_{(S)} = \left(\frac{x}{2} - x^3\right) dt + \sqrt{1 + x^2}\, dw\,. \tag{8.63}$$

The moment propagator is:

$$\mathcal{L}_m(x) = m(x^m - x^{m+2}) + \frac{1}{2}m(m-1)\left(x^m + x^{m-2}\right)\,. \tag{8.64}$$

In Fig. 8.6, we show the accuracy achieved for the nonlinear oscillator with multiplicative noise using a conventional Stratonovich method (red line) and the Stratonovich POS method (blue line) as a function of the number of trajectories used. The simulations were carried out using a step-size of $\Delta t = 2 \cdot 10^{-3}$, a maximum time of $t_{max} = 10$, an initial mean of $x_0 = 0$ und an initial standard deviation of $\sigma_0 = 0.9$. As in section 8.5.2, the non-optimized integration was carried out using a 3-step semi-implicit midpoint method. The POS-optimized integration was carried out using the Stratonovich-type POS method described here with 5 iterations per time-step. Integrating up to $t_{max} = 10$ ensures that the resulting moments are sufficiently close to the steady-state value of the nonlinear oscillator. The steady-state value can be obtained analytically via

$$\langle x^m \rangle^{(ss)} = \frac{\int_{-\infty}^{\infty} x^m \left(1 + x^2\right) e^{-x^2} dx}{\int_{-\infty}^{\infty} \left(1 + x^2\right) e^{-x^2} dx} = \begin{cases} 0 & \text{for } m \text{ odd} \\ \frac{(m+3)(m-1)!!}{3 \cdot 2^{m/2}} & \text{for } m \text{ even}\,. \end{cases} \tag{8.65}$$

The accuracies $\langle E \rangle$ are defined analogously to section 8.5.2, that is $\left\langle E^{(m)} \right\rangle = \left\langle \left| \langle x^m \rangle - \langle x^m \rangle^{(exact)} \right| \right\rangle$. The grey dash-dotted lines indicate a $N_s^{-\frac{1}{2}}$ slope.

Using the POS mothod results in more accurate results for all of the moments and all ensemble size studied, with the exception of the second moment for the smallest ensemble size of $N_S = 2048$ samples. Here the accuracy achieved using the POS method is roughly equal to that using a conventional Stratonovich integration method. Unlike

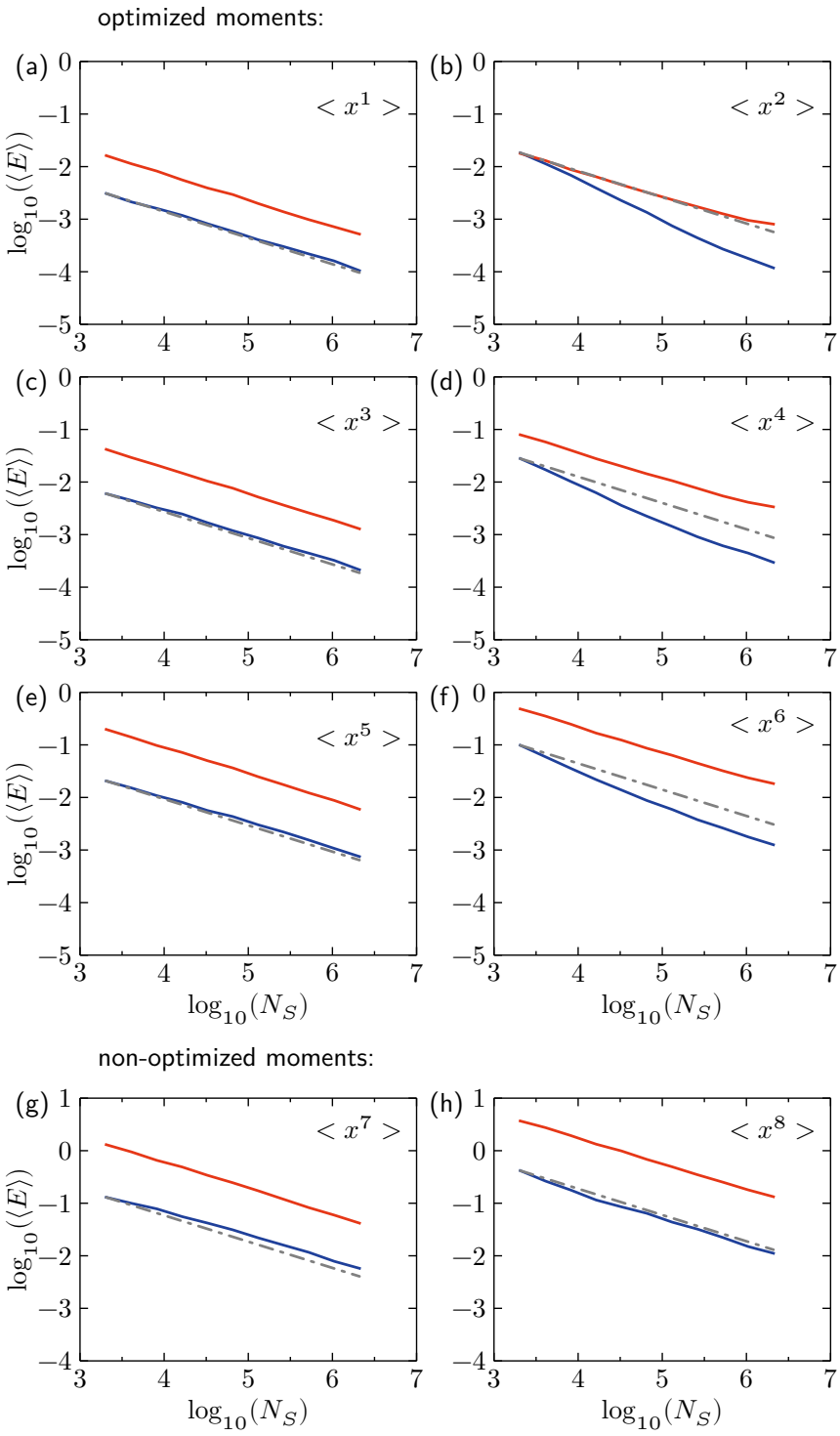

Figure 8.6: Accuracies in the moments for a non-optimized midpoint method Stratonovich simulation (red line) and a Stratonovich midpoint POS simulation (blue line) for the nonlinear oscillator with multiplicative noise as a function of the number of trajectories in a double-logarithmic plot. The simulation uses a step-size of $\Delta t = 2 \cdot 10^{-3}$, initial conditions $x_0 = 0$, $\sigma_0 = 0.9$ and a simulation time of $t_{max} = 10$. The POS simulation was carried out optimizing the first 6 moments using 5 iterations per time-step as well as optimizing the initial ensemble.

the results for the bistable oscillator with additive noise presented in section 8.6.1, some moments, most notably the even moments up to 6'th order, exhibit an ensemble-size scaling that differs from the familiar $N_S^{-\frac{1}{2}}$ one. While the accuracies in the 2nd, 4th and 6th moments scale significantly more strongly at a rate of about $N_S^{-0.72}$, the accuracy in the 7th moment scales slightly less strongly at about $N_S^{-0.42}$. The exact origin of this different scaling behaviour remains somewhat unclear. The higher order moments are not optimized, however, and therefore are clearly not expected to behave as well as the optimized, low order moments.

8.6.4 Laser equation

A simplified model for the single-mode quantum statistics in a laser system can be found to be:

$$da = \left(1 - |a|^2\right) a\, dt + b\, dW_c\,, \tag{8.66}$$

where both a and dW_c are complex-valued. The noise obeys $\langle dW_c dW_c^* \rangle = 2dt$. Here a is proportional to the mode amplitude, and b is a real-valued parameter that depends on the steady-state photon number of the laser mode [30].

The scaled photon number n_s is given by $n = |a|^2$. A simple calculation reveals the steady-state value for n to be

$$n_{ss} = 1 + \sqrt{\frac{2}{\pi}} \frac{b}{\exp\left[1/\left(2b^2\right)\right]\left[1 + \operatorname{erf}\left(1/\left(\sqrt{2}b\right)\right)\right]}\,. \tag{8.67}$$

Fig. 8.7 shows the error in calculating n_s with the Stratonovich POS-method described here (blue line) and a non-optimized stochastic integration (red line) for the laser equation for different values of b.

The simulations were carried out using $N = 131072$ trajectories and a time-step of $\Delta t = 2 \cdot 10^{-3}$ with an integration time of $t_{max} = 25$. The long integration time ensures that the system is sufficiently close to the steady state.

As with the systems mentioned previously, we are using a 3-step midpoint method to carry out the non-optimized simulation. Because the Stratonovich POS method outlined

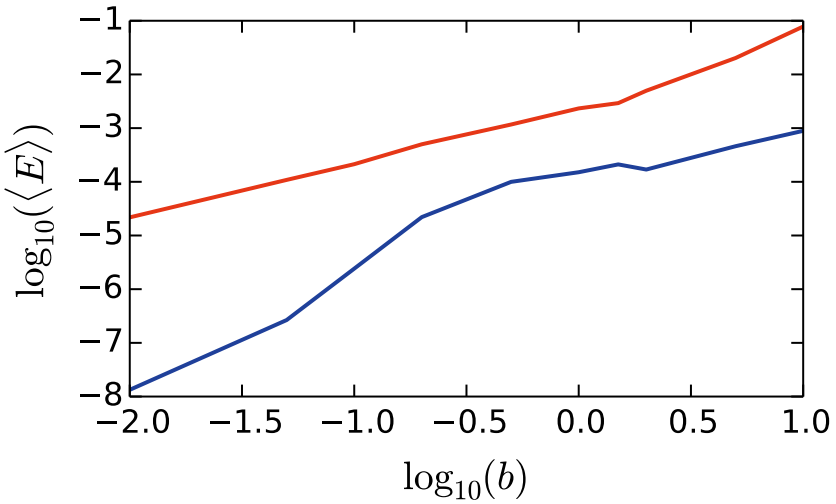

Figure 8.7: Accuracy of the scaled photon number n_s for the laser equation using a non-optimized stochastic integration (red line) and a POS-optimized integration (blue line) as a function of the parameter b.

in this article is defined for real-valued stochastic differential equations, we treat the complex-valued laser equation as a 2-variate real-valued SDE for the real and imaginary part of a. We are optimizing (constraining) every moment and cross-moment for the real and imaginary part of a up to order 4. In other words, we are optimizing the first 4 consecutive moments of $\Re\{a\}$, of $\Im\{a\}$ and cross-moments of the form $\Re\{a\}^n \Im\{a\}^m$, with a combined power $n + m$ not greater than 4. This results in a total of 18 constrained moments. The POS optimization includes the optimization of the initial trajectories.

As before, we define the accuracy as the average deviation of n at $t = 25$ from its steady-state value n_{ss}, which we assume is a sufficiently good approximation to the correct value.

This example shows that POS can be successfully used to improve the accuracy of complex-valued SDEs and, more generally speaking, real-valued multivariate SDEs by several orders of magnitude.

8.7 Summary

There are many known algorithms to reduce discretization error in time for stochastic equations. However, there are very few useful techniques that reduce sampling error. As a result, this type of error is a central issue that limits the applicability of these methods

to many important problems in physics. Since these sampling methods are used as well in many other disciplines, this is an important issue.

No matter how high the convergence order is, it is always the case that reducing the step-size below a certain threshold is less useful than using more trajectories, owing to the dominance of the sampling error in the results. The alternative of using truncated hierarchies of moment equations is even more problematic, owing to systematic errors. This cannot easily be checked by simply using higher orders of moments, as the series may not converge to the correct answer.

We have shown that it is advantageous to unify these two approaches through a type of Bayesian conditioning of samples of the stochastic trajectories on moment equations. The use of a set of random trajectories automatically guarantees the existence of statistical moments to all orders, thus removing the need to truncate either cumulants or moments at any order. It is important that in the nonlinear example treated here, moments of higher order than the optimized moments have reduced errors as well as the optimized moments.

All examples treated here converge rapidly to the correct answer for all orders checked. However, we cannot exclude other possibilities and the precise domain of equations for which these methods apply merits further investigation. In particular, we note that systematic errors may occur at small ensemble sizes. Our results so far indicate that any such errors become rapidly negligible when a large enough ensemble is used for POS error minimization. In summary, we have shown that even with nonlinear stochastic equations, the use of parallel optimized sampling, combined with the relatively more precise nature of moment equations, allows one to greatly improve overall sampling errors in stochastic trajectories, for both the optimized and the non-optimized moments.

8.8 Acknowledgments

S. K. and P.D.D. wish to acknowledge useful discussions with B. Opanchuk and B. Dalton, as well as research funding from the Australian Research Council Discovery program and Swinburne University of Technology.

8.9 Appendix: Systems with closed moment hierarchies

8.9.1 Linear oscillator

The moment propagator is:

$$\mathcal{L}_m(x) = -mx^m + \frac{1}{2}m(m-1)x^{m-2}\,. \tag{8.68}$$

Here the moment coupling is closed, and a finite set of moments can be computed in a sequence from $1, \ldots, N$. For example, the mean value is

$$\langle x(t)\rangle = \langle x(0)\rangle \exp(-t)\ . \tag{8.69}$$

Using Ito's rules for higher moments, a stochastic moment equation is

$$\frac{do_m}{dt} = -mo_m + \frac{1}{2}m(m-1)o_{m-2}\,. \tag{8.70}$$

Hence, for example,

$$\frac{do_2}{dt} = -2(o_2 - 1/2)\,. \tag{8.71}$$

This has the exact solution that:

$$o_2 = [o_2(0) - 1/2]\,e^{-2t} + 1/2\,. \tag{8.72}$$

8.9.2 Kubo oscillator

Since noise and variable are uncorrelated in Ito equations, the first moment equation is:

$$\langle \dot{z}(t)\rangle = -\frac{1}{2}\langle z(t)\rangle\ , \tag{8.73}$$

which has the solution that:

$$\langle z(t)\rangle = \langle z(0)\rangle \exp(-t/2)\ . \tag{8.74}$$

More generally, the moment propagator is:

$$\begin{aligned}\mathcal{L}_m(z) &= a(z)\, m z^{m-1} + \frac{m(m-1)}{2}\mathsf{d}(z)\, z^{m-2} \\ &= -\frac{1}{2}m z^m - \frac{1}{2}m(m-1)z^m \\ &= -\frac{1}{2}m^2 z^m\,. \end{aligned} \tag{8.75}$$

That is, we have that each analytic moment decays independently according to:

$$\frac{d}{dt}\langle z^m\rangle = -\frac{1}{2}m^2\langle z^m\rangle\,. \tag{8.76}$$

Hence, since the Gaussian increments are all independent of the initial condition:

$$\langle z^m(t)\rangle = \langle z^m(0)\rangle \exp\left[-m^2 t/2\right]\,. \tag{8.77}$$

Note however that these are not all the independent moments. The most general moment is written:

$$\langle z^m(t) z^{*n}(t)\rangle = \langle z^m(0)\rangle \exp\left[-(m-n)^2 t/2\right]\,. \tag{8.78}$$

8.10 Bibliography

[1] A. D. Fokker, "Die mittlere Energie rotierender elektrischer Dipole im Strahlungsfeld.", Annalen der Physik **43**, 810–820 (1914) (page 194).

[2] M. Planck, Sitzber Preuss. Akad. Wiss. Physik. Math. **K1**, 324 (1917) (page 194).

[3] P. Langevin, "Sur la théorie du mouvement brownien", C. R. Acad. Sci. (Paris) **146**, 530–533 (1908) (page 194).

[4] I. Karatzas and S. Shreve, *Brownian motion and stochastic calculus*, 2nd (Springer, 1991), p. 470 (page 194).

[5] H. Risken, *The Fokker–Planck equation*, 2nd (Springer-Verlag, Berlin, 1996) (page 194).

[6] P. Glasserman, *Monte Carlo methods in financial engineering*, reprint (Springer, 2010), p. 596 (page 194).

[7] L. Arnold, *Stochastic differential equations: Theory and applications*, reprint (Folens Publishers, 1992), p. 228 (pages 194, 196, 201).

[8] C. W. Gardiner, *Handbook of stochastic methods*, 2nd (Springer-Verlag, Berlin, 1985), p. 442 (pages 194–196, 198, 201).

[9] N. G. Van Kampen, *Stochastic processes in physics and chemistry*, 3rd (North Holland, 2007), p. 464 (page 194).

[10] M. J. Werner and P. D. Drummond, "Robust algorithms for solving stochastic partial differential equations", J. Comput. Phys. **132**, 312–326 (1997) (page 194).

[11] G. N. Mil'shtein, "Approximate integration of stochastic differential equations", Theor. Probab. Appl. **19**, 557–562 (1975) (pages 194, 199).

[12] P. D. Drummond and I. K. Mortimer, "Computer simulations of multiplicative stochastic differential equations", J. Comput. Phys. **93**, 144–170 (1991) (pages 194, 198, 199, 201, 216).

[13] P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations* (Springer-Verlag, Berlin, 1992) (pages 194, 199, 202).

[14] D. J. Higham, "An algorithmic introduction to numerical simulation of stochastic differential equations", SIAM Rev. **43**, 525–546 (2001) (page 194).

[15]D. Talay and L. Tubaro, "Expansion of the global error for numerical schemes solving stochastic differential equations", Stoch. Anal. Appl. **8**, 483–509 (2007) (pages 194, 199).

[16]M. B. Giles, "Multilevel Monte Carlo path simulation", Oper. Res. **56**, 607–617 (2008) (page 194).

[17]A. Abdulle and A. Blumenthal, "Stabilized multilevel Monte Carlo method for stiff stochastic differential equations", Journal of Computational Physics **251**, 445–460 (2013) (page 194).

[18]R. M. Wilcox and R. Bellman, "Truncation and preservation of moment properties for Fokker–Planck moment equations", Journal of Mathematical Analysis and Applications **32**, 532–542 (1970) (pages 194, 201).

[19]K. Sture, J. Nordholm and R. Zwanzig, "Strategies for fluctuation renormalization in nonlinear transport theory", Journal of Statistical Physics **11**, 143–158 (1974) (page 194).

[20]N. G. Van Kampen, "A cumulant expansion for stochastic linear differential equations. I", Physica **74**, 215–238 (1974) (page 194).

[21]P. Hänggi and P. Talkner, "A remark on truncation schemes of cumulant hierarchies", J. Stat. Phys. **22**, 65–67 (1980) (pages 194, 200).

[22]R. V. Bobryk, "Closure schemes in stochastic nonlinear dynamics: a validation case study", Physical Review E **83** (2011) (pages 194, 201, 222).

[23]B. Opanchuk, S. Kiesewetter and P. D. Drummond, "Parallel optimized sampling for stochastic equations", arXiv: 1502.07186 (2015) (pages 195, 201).

[24]D. M. Ceperley, "Path integrals in the theory of condensed helium", Reviews of Modern Physics **67**, 279–355 (1995) (page 195).

[25]A. Georges, G. Kotliar, W. Krauth and M. J. Rozenberg, "Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions", Reviews of Modern Physics **68**, 13–125 (1996) (page 195).

[26]J. E. Mayer and E. Montroll, "Molecular distribution", The Journal of Chemical Physics **9**, 2–16 (1941) (page 195).

[27]O. E. Alon, A. I. Streltsov and L. S. Cederbaum, "Multiconfigurational time-dependent Hartree method for bosons: Many-body dynamics of bosonic systems", Phys. Rev. A **77**, 033613 (2008) (page 195).

[28] M. Morillo, J. Gómez-Ordóñez and J. M. Casado, "Checking the validity of truncating the cumulant hierarchy description of a small system", in *Localized excitations in nonlinear complex systems* (Springer, 2014), pp. 377–387 (page 201).

[29] C. C. Paige and M. A. Saunders, "Lsqr: An algorithm for sparse linear equations and sparse least squares", ACM Trans. Math. Soft. **8**, 43–71 (1982) (page 204).

[30] P. D. Drummond and M. Hillery, *The quantum theory of nonlinear optics* (Cambridge University Press, 2014) (page 225).

Appendix A

POS, ensemble size scaling

In this appendix, I will present new results supplementing and expanding those presented in chapter 8.

A.1 Description

In chapter 7, an original method, called "Parallel optimized sampling" was presented. In its simplest form, it can be used to impose exact values for a selected set of observables, such as moments, on an ensemble comprised of samples, provided that the initial values are already close to the exact values to be imposed. It is an iterative procedure that minimizes the euclidean norm of the ensemble vector increments between iteration steps. These ideas were then used to construct an algorithm for the integration of stochastic differential equations. It is a combination of a standard Euler-Mayurama algorithm and the method of imposing specific values on a set of observables referred to previously. In chapter 8, a similar method is presented, which is based on a Stratonovich-type integration algorithm and thus suitable for Stratonovich-type stochastic differential equations. Its concrete implementation is arguably significantly more complex than its Euler-Mayurama counterpart, yet the basic idea is still based on the same idea.

Both methods, that is, the Ito-type and Stratonovoch-type POS methods have been applied to a linear stochstic differential equation with the result that the stochastic error is almost completely eliminated. However, this is not representative of the overall performance of POS. Instead, it is much more insightful to measure the performance of POS for nonlinear stochastic differential equations. Both methods have been tested for

a bistable nonlinear oscillator with additive noise, given by the stochastic differential equation

$$dx = \left(x - x^3\right) dt + dw\,. \tag{A.1}$$

In both cases, eq. (A.1) has been integrated for an integration time long enough to assume that a steady-state has been reached. An error measure $\langle E \rangle$ is then defined as the mean absolute deviation of an observable resulting from a specific trajectory from its theoretically predicted steady state value. In other words, for an observable $f(x)$, the error measure $\langle E \rangle$ is defined as

$$\langle E \rangle \equiv \left\langle \left| f(x_i) - f_{\text{theor}} \right| \right\rangle_i\,, \tag{A.2}$$

where $\langle \rangle_i$ indicates an average over the ensemble vector and f_{theor} the theoretically predicted steady-state value of the observable $f(x)$. For both methods, the first 8 consecutive moments have been analyzed and in both cases, the result was that using the POS method, the error measure $\langle E \rangle$ is reduced by between 1 and 1.5 order of magnitude relative to a conventional, that is, non-POS optimized, integration method using the same system parameters. Furthermore, both the POS method and a conventional integration method scale exhitit a $N_s^{-1/2}$-scaling, where N_S is the number of stochastic samples used in the simulation. The results can be found in figs. 7.17 and 8.3.

The Stratonovich-type algorithm introduced in capter 8 was also tested against a bistable nonlinear oscillator with multiplicative noise, given by the Ito-type stochastic differential equation

$$dx_{(I)} = \left(x - x^3\right) dt + \sqrt{1 + x^2} dw\,, \tag{A.3}$$

which is equivalent to the Stratonovich-type stochastic differential equation

$$dx_{(S)} = \left(\frac{x}{2} - x^3\right) dt + \sqrt{1 + x^2} dw\,. \tag{A.4}$$

The results for this simulation can be found in fig. 8.6. As in the case with additive noise, the results obtained using the POS-method still exhibit a lower error measure $\langle E \rangle$ than those obtained using a reference integration. However, compared to eq. (A.1),

the improvement of a POS-integration over a reference integration is notably smaller. More importantly, some observables, specifically the 2nd, 4th and 6th momentm exhibit a scaling notably different from the familiar $N_s^{-1/2}$ one. A scaling different from $N_s^{-1/2}$ is indicative of errors which are non-stochastic in nature. At the same time, the moments of order other than 2, 4 and 6 exhibit a relatively clear error scaling corresponding to stochastic errors.

In order to investigate the exact nature of the errors resulting from eq. (A.3), I have defined two new error signatures E_{stoch} and E_{system}, which are meant to correspond to the stochastic and systematic error of the simulation, respectively. For these new error measurement, an essential aspect is that the simulation, involving N_S stochastic samples x_i, is carried out multiple (M) times. The systematic error is then defined as

$$E_{\text{system}} \equiv \left| \langle \langle f(x_i) \rangle - f_{\text{theor}} \rangle_j \right| , \tag{A.5}$$

where the outer mean $\langle \rangle_j$ indicates an averaging over the M independent runs of the simulation. Eq. (A.5) is now an measurement of the total deviation not on the basis of individual trajectories, but instead of an entire ensemble.

Furthermore, I define a stochastic error as

$$E_{\text{stoch}} = \sqrt{\frac{2}{\pi}} \sigma \left(\langle f(x_i) \rangle \right)_j , \tag{A.6}$$

where $\sigma(\,)_j$ means the standard deviation of the values numerated by the index j. Assuming a set of values $y_i \equiv f(x_i)$ which are normally distributed according to

$$P(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} , \tag{A.7}$$

the stochastic error would then result via

$$E_{\text{stoch}} = \int_{-\infty}^{\infty} |(y-\mu) \cdot P(y)| \, dy = \sigma \sqrt{\frac{2}{\pi}} , \tag{A.8}$$

which is where the factor $\sqrt{\frac{2}{\pi}}$ in eq. (A.6) comes from.

A.2 Results

Fig. A.1 shows the errors as defined by eqs. (A.5) and (A.6) resulting from a simulation of eq. (A.4) using the Stratonovich midpoint POS-method introduced in chapter 8 as well as for a conventional 3-step semi-implicit midpoint method[79] as a function of the number of stochastic samples per ensemble in a double-logarithmic plot. For both methods, the ensemble was initialized using a normal distribution with mean $\mu_0 = 0$ and standard deviation $\sigma_0 = 0.9$. The total simulation time was $T_{textmax} = 10$ with a time-step of $\Delta t = 2 \cdot 10^{-3}$. For the POS simualation, the first 6 consecutive moments were optimized using a fixed number of 5 iterations. The red lines show the conventional integration results, while the results of the POS-integration are given by the blue lines. Solid lines correspond to the stochastic error E_{stoch} as defined in eq. (A.6), while dashed lines correspond to the systematic error E_{system} as defined in eq. (A.5). The gray dash-dotted lines indicate a $N_S^{-1/2}$-scaling. For non-POS stochastic errors (solid red lines), no scaling indicators are shown, because these lines already exhibit a near perfect $N_S{-}1/2$-scaling. The results were obtained by running approx. $M = 680$ independent POS-optimized and reference integrations.

Fig. A.2 shows results obtained in the same way as those in fig. A.1, but with a reduced time-step of $\Delta t = 5 \cdot 10^{-4}$. For fig. A.2, approx. $M = 7800$ non-POS integrations and $M = 4400$ POS-integrations were used.

A.3 Discussion

In the author's opinion, the results in figs. A.1 and A.2 contain both expected and unexpected, conclusive and somewhat puzzling information. The results of the reference integration, indicated by red lines, are not very surprising and quite easy to understand: The stochastic error follows an almost perfect $N_S^{-1/2}$-trend as it should (a trendline has not been included in the graph here because it would not be distinguishable from the error line) while the systematic error, especially in the case of the even moments, is roughly constant. Furthermore, reducing the time-step results in an unchanged stochastic error, but reduces the systematic error, which remains mostly constant. This leads to the conclusion that for the reference integration, the systematic error is mostly the result of the discretization in time. In other words, it seems that in the case of the reference integration, the error

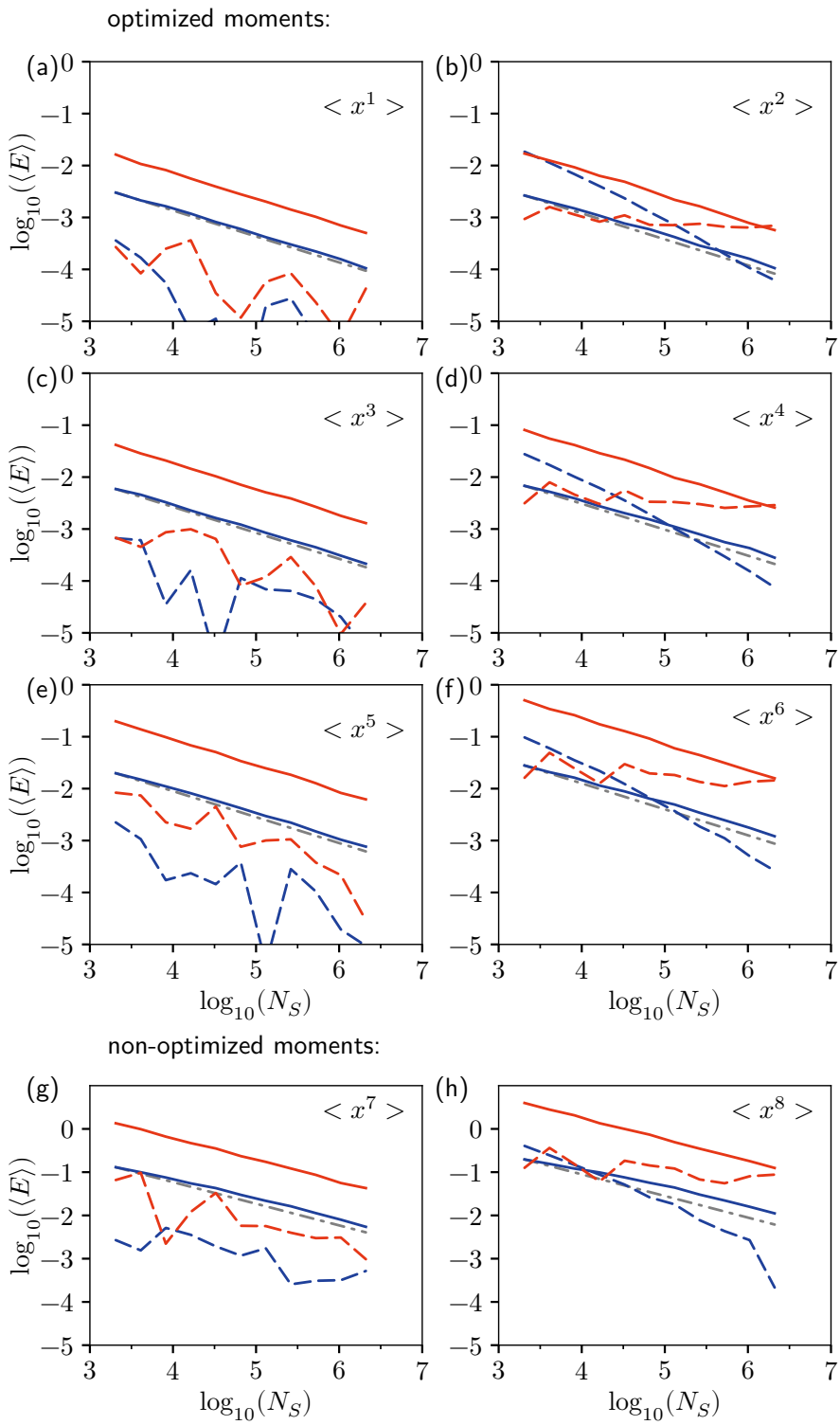

FIGURE A.1: Stochastic error (solid lines) and systematic error (dashed lines) for eq. (A.4) using a POS-optimized integration method (blue lines) and a conventional reference integration method (red lines) for the first 8 consecutive moments as a function of the ensemble size N_S. The simulation was carried out for an integration time $T_{\max} = 10$ and a time-step of $\Delta t = 2 \cdot 10^{-3}$.

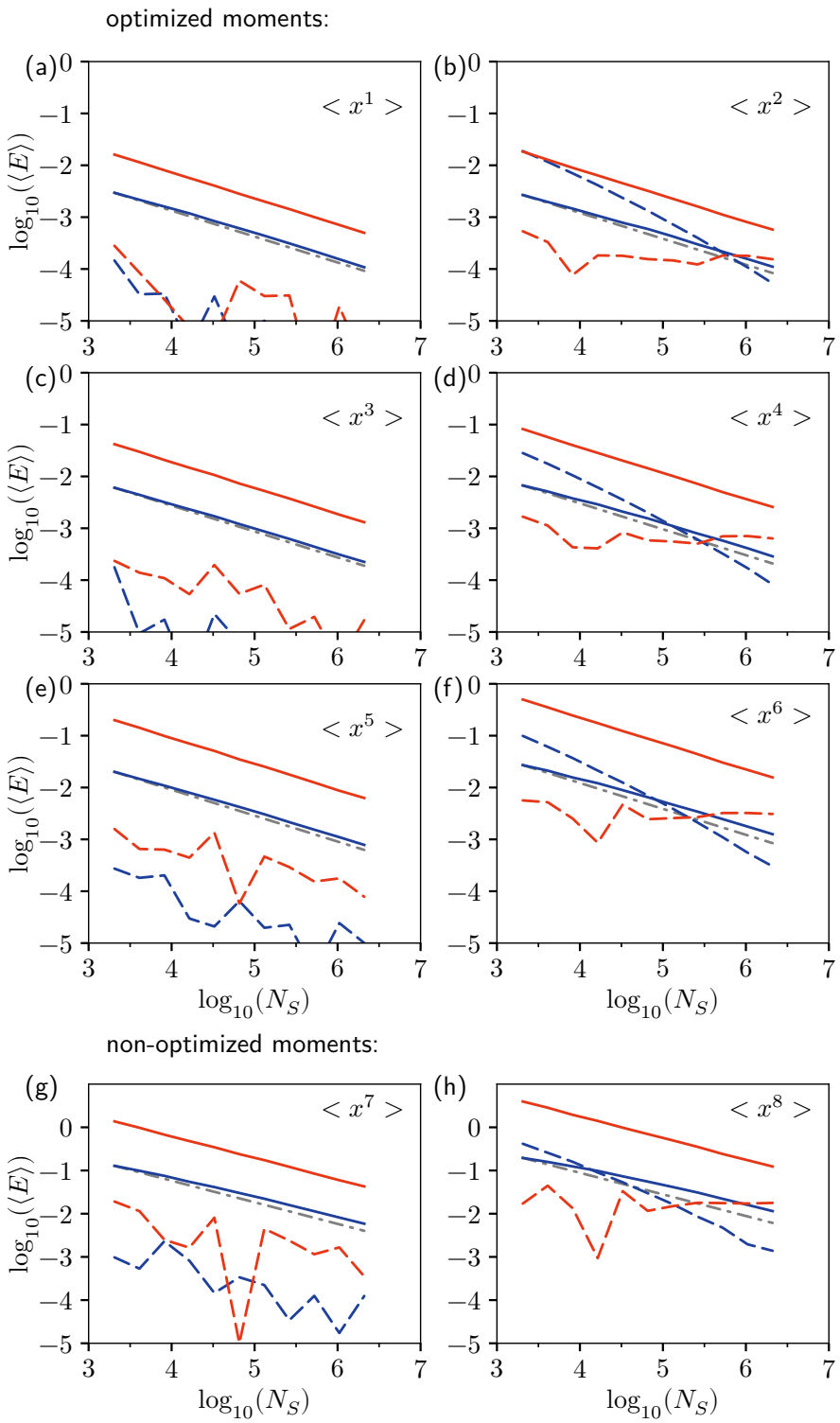

FIGURE A.2: Stochastic error (solid lines) and systematic error (dashed lines) for eq. (A.4) using a POS-optimized integration method (blue lines) and a conventional reference integration method (red lines) for the first 8 consecutive moments as a function of the ensemble size N_S. The simulation was carried out for an integration time $T_{\max} = 10$ and a time-step of $\Delta t = 5 \cdot 10^{-4}$.

measures introduced in eqs. (A.5, A.6) allow for the separation between stochastic and time-step error.

In the case of the POS-integration, the situation is less straightforward. The POS integration method results in a consistently reduced stochstic error compared to the stochastic error resulting from a reference integration. It is noteworthy that the stochastic error resulting from the POS-integration deviates from a $N_S^{-1/2}$ and instead exhibits a scaling with a slightly lower negative power than that. Probably the most surprising results are the systematic errors, particularly those of the even moments, resulting from the POS-integration. For one, they are not constant as in the case of the reference integration, but rather exhibit a strong N_S-scaling with a negative power notably higher than $1/2$. Furthermore, for all of the even moments presented in figs. A.1 and A.2, the systematic error resulting form the POS-integration is greater than both the stochastic error resulting from the POS-integration and the systematic error resulting from the reference integration for a low number of samples. At the same time, it is lower than both of these errors for a high number of samples. Equally surprising, while reducing the time-step leads to substantially reduced systematic errors in the case of the reference integration, for the POS-integration, the systematic error for even moments remains virtually unchanged. For the odd moments, the systematic error seems to be reduced for both the POS-integration and the reference integration.

Clear, the scaling behaviour of the stochastic and systematic errors resulting from the POS-integration is quite different from the conventional one. It seems that in addition to reducing the stochastic error, the POS algorithm also has a profound effect on the systematic error. Remarkably, using the POS algorithm, the systematic error can be reduced to level which is substantially below that of the time-step error of a conventional integration method with the same system parameters if a large enough sample size is used. At the same time, a sample size which is too small will lead to an even larger systematic error than that of a reference integration.

Chapter 9

Discussion

In the paper "Scalable quantum simulation of pulsed entanglement and Einstein-Podolsky-Rosen steering in optomechanics," a full nonlinear analysis of the experiment proposed by Hofer *et al.*[27] was carried out, where previously, the analysis had mostly been limited to linearized system dynamics. We demonstrated that a full nonlinear simulation is possible with conventional and available methods. For the experiments we studied, it was shown that overall, the results obtained through a full nonlinear analysis agree with those obtained through linearized system dynamics. However, in principle, one can never be sure beforehand that this is the case. Also, it was demonstrated how a full nonlinear simulation reveals important aspects of the experiment, such as phase shifts, which remain hidden where linearized system dynamics are used. For these reasons, we conclude that in general, a full treatment of the nonlinear system dynamics from first principles should be preferred over approximate methods such as linearized system dynamics, when the aim is to make precise experimental predictions as this approach has the most predictive power with very few or no disadvantages. This approach is expected to become more important with the development of optomechanical systems having greater nonlinearity, and with coupled systems of nonlinear optomechanics.

Furthermore, 3 different entanglement criteria were compared in this experiment. Our results clearly demonstrate the already known fact, that the multiplicative entanglement criterion is more robust and stronger than the additive entanglement criterion. In spite of this, the additive entanglement criterion seems to be the more commonly used choice in the field of quantum optics. We conclude that the product criterion, which was derived earlier, should be used instead.

Also, the prospect of observing EPR-steering in this experimental setup was analyzed. We considered an experiment in the electromechanical domain previously carried out by Palomaki *et al.*[28] as well as an experiment that has not been carried out, based on system parameters corresponding to an experiment carried out by Chan *et al.*[24]. It was demonstrated that, while both experiments will exhibit an excellent degree of entanglement, EPR-steering can not be observed in the experiment carried out by Palomaki *et al.*, however, it could clearly be observed in a hypothetical experimental setup based on the parameters of Chan *et al.*[24]. This confirms the well-known fact that EPR-steering is an effect which is much more sensitive to thermal noise than entanglement[66, 80]. It also merits future efforts on realizing an experiment of the type proposed by Hofer *et al.* in the optomechanical domain as this has the potential to demonstrate EPR-steering between two mesoscopic objects.

In the paper "Einstein-Podolsky-Rosen quantum simulations in nonclassical phase-space," a review of numerical simulations of EPR-steering was given. We presented the background of the material in a concise way and illustrated through a number of examples how EPR-steering and Bell violations can be simulated using exact and first-principles phase-space methods. This gives new insight into Feynman's claim[81] that probabilistic quantum simulations are not possible without "hocus-pocus", because, as he argued, a quantum simulation without "hocus-pocus" would be akin to a local hidden variable theory and therefore would not be equivalent to quantum mechanics, in particular could not explain Bell violation. As was highlighted to in our paper, it is perfectly feasible to carry out a quantum simulation using phase-space methods which are capable of correctly predicting Bell violations without the addition of any "hocus-pocus". This is because theories like the positive-P method are not local hidden variables (LHV) theories, but rather use a non-classical phase-space. We also demonstrated that for a sufficiently large nonlinear coupling, linearized system dynamics and approximate methods such as the truncated-Wigner method will yield different results from the exact positive-P method, giving further merit to an exact, fully nonlinear treatment using the positive-P method in general. While current optomechanical experiments typically involve coupling constants which are small enough to justify a linearized treatment, it is conceivable that future experiments will reach the domain where a full nonlinearized treatment using exact methods such as the positive-P method will be necessary to achieve accurate predictions.

In the paper "Violations of multisetting quaternion and octonion Bell inequalities," a new type of Bell inequality[34, 35] was analyzed that intriguingly involves the abstract mathematical concepts of the quaternion and octonion algebra, allowing for up to 8 different measurement settings and arbitrarily many detector sites. The inequality was analyzed in a setup involving GHZ-states[72] and generalized GHZ-states. We worked out the exact detector settings, which potentially involves hundreds of phase angles, that would result in an optimal detection of Bell violation. We analyzed the minimal required detector efficiencies in order to detect Bell violations. Our results indicate that the number of measurement settings has no influence on the minimal detector efficiencies. Although this does not confirm our optimistic expectation of a reduced minimum detector efficiency through a higher number of measurement settings, our results further indicate that the minimal detector efficiency η has a dependence of $\eta = 2^{\frac{2}{N}-1}$, where N is the number of measurement sites. This results in a limit of $\eta \to \frac{1}{2}$ for $N \to \infty$. We believe that these results, while not necessarily giving rise to superior experiments, are nevertheless an important contribution to the field of Bell violations. We have conclusively and comprehensively analyzed the SV-type Bell inequalities for GHZ-states, which had not been done previously.

In the paper "Parallel optimized sampling for stochastic equations," our original method for the integration of Itō-type stochastic differential equations was presented. Many techniques which have been developed with the aim of improving state-of-the-art techniques for the integration of stochastic differential equations focus exclusively on time-step errors. By deriving an effective integration order that takes the sampling error into consideration, we demonstrated that the usefulness of such an approach is limited because the sampling error tends to become the dominating source of error fairly quickly. Because of that, in developing our algorithm, we focused explicitly on the sampling error. The aim of our method is the integration of Itō-type stochastic differential equations with a reduced sampling error. We have outlined the basic idea of our approach and demonstrated via a number of sample equations that our method is capable of achieving substantially reduced sampling errors when compared to a traditional Euler-Mayurama algorithm, while at the same time maintaining a linear complexity in the number of samples used. The sample equations contain linear and nonlinear equations as well as univariate and multivariate ones. In addition to that, a variation of our method can also

be used to for the task of sampling probability density functions to reduce sampling errors based on known moments or other observables. It is important to point out that the number of methods available which have the same objective, that is, variance reduction, is rather limited. Many variance reduction techniques are only useful for very specific types of problems. In contrast, our approach is a general purpose one and can be applied to a wide range of equations, including those encountered in the field of nonclassical systems, where phase-space methods are applied, because these require the integration of a multivariate stochastic differential equation.

In the paper "Algorithms for integration of stochastic differential equations using parallel optimized sampling in the Stratonovich calculus," the work of the previous paper was expanded to construct an algorithm capable of solving Stratonovich-type stochastic differential equations. Similar to the previous paper, we outlined the basic idea, described the algorithm in detail and applied our technique to a number of sample equations. It was demonstrated that this method is capable of not only substantially reducing the sampling error, but also results in a smaller time-step error when compared to the Itō-type algorithm presented in the previous paper. This is a direct consequence of the fact that this approach is based on a Stratonovich-type algorithm. The fact that the ideas of our previous papers can be used to construct a substantially more complex Stratonovich-type algorithm exhibiting the same variance-reduction potential highlights the flexibility of our approach. It is conceivable that the ideas included in our method will be applied to different problems in the theory of stochastic processes at some point.

We also compared the accuracies achieved for a particular sample equation when treated via our approach, a traditional Stratonovich-type algorithm and a truncated cumulant hierarchy approach. In a truncated cumulant hierarchy approach, the deterministic equations of motion of the cumulants are obtained from the stochastic differential equation. Generally, this results in a hierarchy of infinitely many equations which is non-closed, that is, the equations of motion of lower-order cumulants will depend on the values of higher-order cumulants. The hierarchy is then truncated so that all cumulants higher than a given order are set to zero. The differential equations can then conveniently be solved using integration techniques for deterministic differential equations. It was demonstrated that the cumulant hierarchy method results in large systematic errors for

the sample equation considered, while the traditional Stratonovich-type method results in substantially lower errors and our algorithm results in an even greater error reduction.

These results are consistent with the work of Bobryk[82] has analyzed the truncated cumulant hierarchy method among other methods for similar equations. He demonstrated that even for large truncation orders, the cumulant hierarchy method still results in rather large systematic errors, suggesting that the cumulant hierarchy method might not converge for these equations. In many fields of physics, such as statistical physics and quantum many-body theory, several methods exist which are conceptually similar to a truncated cumulant hierarchy method, such as dynamical mean-field[83], cluster expansion[84] or multi-configurational[85] methods. Bobryk's findings suggest that the validity of these methods should be questioned. We believe that our approach or its ideas could constitute a viable alternative to these.

9.1 Concluding remarks

Quantum simulations are an invaluable tool in the study of nonclassical systems. While a direct simulation of the entire Hilbert space of a physical system is possible in principle, it clearly becomes unfeasible rather quickly in practice. On the other hand, phase-space methods such as the truncated Wigner method and the positive-P method provide a framework in which a quantum simulation is both possible and feasible. While Feynman[81] believed that a quantum simulation would either be prohibitively resource-intensive or unable to predict all phenomena of quantum mechanics, this is clearly not the case, among other things due to the existence of phase-space methods. It seems clear that quantum simulations will only become more important in the future. Ongoing advances in computer technology will make it easier to carry out numerical simulations. Improvements in experimental methods and detector technology will result in an increased demand for predictions at high precision. This will certainly be true for systems which cannot be treated with available models or other approximations for the required accuracies.

The emerging field of optomechanics bears great potential for the future, among other things, because it allows for the observing of quantum effects on a mesoscopic scale. Its potential applications include the detection of gravitational waves, realizations of a

Schrödinger cat and a systematic study of decoherence. The study of nonlocality appears to be a particularly intriguing application of quantum optics because of its potential to generate nonlocality between mesoscopic objects. For example, an experimental demonstration of entanglement, EPR-steering or even Bell violation between two massive mechanical oscillators would be another milestone in the field of nonlocality and quantum mechanics in general.

We have successfully demonstrated via a number of toy problems that our original method called "parallel optimized sampling" (POS) can lead to substantially more accurate results in a Monte-Carlo integration of a stochastic differential equation for an equal number of stochastic samples. In chapter A, using two distinct error measure, named "stochastic error" and "systematic error", we dicovered that using the POS method can lead to both greater and substantially smaller systematic errors compared to a conventional integration method, depending on the number of samples used. While this on the one hand reveals that POS not only has the potential to reduce stochastic errors, but systematic errors as well, it also shows that under certain conditions, it can perform worse than conventional methods. Though we have tested POS on a number of linear, nonlinear, univariate and multivariate, Itō-type and Stratonovich-type equations, further tests are certainly desirable. Also, a more rigorous mathematical analysis of the method should be carried out. Particularly, it would be useful to be able to predict or estimate the performance, that is, achievable accuracy, for a POS-simulation given a specific stochastic differential equation.

We believe that POS has many potential applications in fields where stochastic differential equations play a role, including quantum simulations, and could have a great impact in the field of stochastic processes itself. Due to time constraints, we have been unable to apply POS to a quantum simulation such as those we have carried out in the field of quantum optics, even though we believe this would be feasible. It is worth pointing out that POS is a new method and very much in its infancy.

Appendix B

List of Papers

B.1 List of papers as part of the thesis including citing articles

- S. Kiesewetter, Q. Y. He, P. D. Drummond, and M. D. Reid, "Scalable quantum simulation of pulsed entanglement and Einstein-Podolsky-Rosen steering in optomechanics," Physical Review A **90**, p. 043805; accepted 4 September 2014

  Citing articles:

  - Q. Y. He, Q. H. Gong, M. D. Reid, "Classifying directional Gaussian entanglement, Einstein-Podolsky-Rosen steering, and discord," Physical Review Letters **114**, p. 060402
  - R. Y. Teh and M. D. Reid, "Criteria for genuine N-partite continuous-variable entanglement and Einstein-Podolsky-Rosen steering," Physical Review A **90**, p. 062337
  - K. Sun, X.-J. Ye, J.-S. Xu, X.-Y. Xu, J.-S. Tang, Y.-C. Wu, J.-L. Chen, C.-F. Li, and G.-C. Guo, "Experimental Quantification of Asymmetric Einstein-Podolsky-Rosen Steering," Physical Review Letters **116**, p. 160404
  - L. Rosales-Zárate, R. Y. Teh, S. Kiesewetter, A. Brolis, K. Ng, and M. D. Reid, "Decoherence of Einstein–Podolsky–Rosen steering," Journal of the Optical Society of America B **32**, pp. A82-A91
  - O. Suchoi, K. Shlomi, L. Ella, and E. Buks, "Time-resolved phase-space tomography of an optomechanical cavity," Physical Review A **91**, p. 043829

- H. Tan, X. Zhang, and G. Li, "Steady-state one-way Einstein-Podolsky-Rosen steering in optomechanical interfaces," Physical Review A **91**, p. 032121
- P. D. Drummond and S. Chaturvedi, "Quantum simulations in phase-space: from quantum optics to ultra-cold physics," Physica Scripta **91**, Number 7
- A. A. Rakhubovsky and R. Filip, "Robust entanglement with a thermal mechanical oscillator," Physical Review A **91**, p. 062317
- Q. Lin and B. He, "Optomechanical entanglement under pulse drive," Optics Express **23**, pp. 24497-24507
- V. Kupčík and R. Filip, "Continuous-variable entanglement mediated by a thermal oscillator," Physical Review A **92**, p. 022346
- R. Y. Teh, L. Rosales-Zárate, B. Opanchuk, M. D. Reid, "Signifying the nonlocality of NOON states using Einstein-Podolsky-Rosen steering inequalities," Physical Review A **94**, p. 042119
- S. Kiesewetter, K. Dechoum, B. Opanchuk, L. Rosales-Zárate, M. D. Reid, and P. D. Drummond, "Einstein–Podolsky–Rosen quantum simulations in nonclassical phase-space," Journal of the Optical Society of America B **32**, pp. A64-A73
- M. D. Reid, "Interpreting the macroscopic pointer by analyzing the Einstein-Podolsky-Rosen steering of an entangled macroscopic superposition state," arXiv:1604.00623
- Q. Zheng, Y. Yao, and Y. Li, "Optimal quantum parameter estimation in a pulsed quantum optomechanical system," Physical Review A **93**, p. 013848

- S. Kiesewetter, K. Dechoum, B. Opanchuk, L. Rosales-Zárate, M. D. Reid, and P. D. Drummond, "Einstein–Podolsky–Rosen quantum simulations in nonclassical phase-space," Journal of the Optical Society of America B **32**, pp. A64-A73; accepted 17 March 2015
- S. Kiesewetter, P. D. Drummond, and M. D. Reid, "Violations of multisetting quaternion and octonion Bell inequalities," Physical Review A **92**, p. 032118; accepted 8 July, 2015

- B. Opanchuk, S. Kiesewetter, P. D. Drummond, "Parallel optimized sampling for stochastic equations," SIAM Journal on Scientific Computing; accepted 5 October, 2016

  Citing articles:

  - S. Kiesewetter, P. D. Drummond, "Algorithms for integration of stochastic differential equations using parallel optimized sampling in the Stratonovich calculus," Computer Physics Communications

- S. Kiesewetter, P. D. Drummond, "Algorithms for integration of stochastic differential equations using parallel optimized sampling in the Stratonovich calculus," Computer Physics Communications; accepted 1 November 2016

B.2 List of papers not part of the thesis including citing articles

- L. Rosales-Zárate, R. Y. Teh, S. Kiesewetter, A. Brolis, K. Ng, and M. D. Reid, "Decoherence of Einstein–Podolsky–Rosen steering," Journal of the Optical Society of America B **32**, pp. A82-A91; accepted March 2, 2015

  Citing articles:

  - K. Sun, X.-J. Ye, J.-S. Xu, X.-Y. Xu, J.-S. Tang, Y.-C. Wu, J.-L. Chen, C.-F. Li, and G.-C. Guo, "Experimental Quantification of Asymmetric Einstein-Podolsky-Rosen Steering," Physical Review Letters **116**, p. 160404
  - R. Y. Teh, L. Rosales-Zárate, B. Opanchuk, M. D. Reid, "Signifying the nonlocality of NOON states using Einstein-Podolsky-Rosen steering inequalities," Physical Review A **94**, p. 042119
  - M. D. Reid, "Interpreting the macroscopic pointer by analyzing the Einstein-Podolsky-Rosen steering of an entangled macroscopic superposition state," arXiv:1604.00623
  - B. Dalton, J. Goold, B. Garraway, M. Reid, "Quantum Entanglement for Systems of Identical Bosons. I General Theory," arXiv:1506.06906

Bibliography

[1]B. E. Stewart, "I.—an account of some experiments on radiant heat, involving an extension of Prevost's theory of exchanges", Transactions of the Royal Society of Edinburgh **22**, 1–20 (1861) (page 5).

[2]G. Kirchhoff, "Ueber die Fraunhoferschen Linien", Monatsberichte der Königliche Preussische Akademie der Wissenschaften zu Berlin, 662–665 (1860) (page 5).

[3]M. Planck, "Ueber das Gesetz der Energieverteilung im Normalspectrum", Annalen der Physik **309**, 553–563 (1901) (page 5).

[4]H. Hertz, "Ueber einen Einfluss des ultravioletten Lichtes auf die electrische Entladung", Annalen der Physik **267**, 983–1000 (1887) (page 5).

[5]A. Einstein, "Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt", Annalen der Physik **322**, 132–148 (1905) (page 5).

[6]N. Foundation, *The Nobel prize in physics*, 1921 (page 5).

[7]D. J. Wineland and H. Dehmelt, "Proposed $10^{14}\Delta v < v$ laser fluorescence spectroscopy on TI^+ mono-ion oscillator III", Bulletin of the American Physical Society **20**, 637 (1975) (page 6).

[8]T. W. Hänsch and A. L. Schawlow, "Cooling of gases by laser radiation", Optics Communications **13**, 68 (1975) (page 6).

[9]D. J. Wineland, R. E. Drullinger and F. L. Walls, "Radiation-Pressure Cooling of Bound Resonant Absorbers", Physical Review Letters **40**, 1639–1642 (1978) (page 6).

[10]P. Törmä and K. Sengstock, *Cold atoms vol. 3: Quantum gas experiments* (Imperial College Press, 2015), p. 19 (page 6).

[11]N. Foundation, *The Nobel prize in physics*, 1997 (page 6).

[12]W. Heisenberg, "Ueber den anschaulichen Inhalt der quantentheoretischen Kinematik and Mechanik", Zeitschrift für Physik **43**, 172–198 (page 6).

[13]W. Heisenberg, *The physical principles of the quantum theory* (Chicago, Ill.,The University of Chicago Press, 1930) (page 6).

[14]N. Bohr, "The quantum postulate and the recent development of atomic theory", Nature **121**, 580–590 (page 6).

[15]P. A. M. Dirac, *Principles of quantum mechanics* (Oxford university press, 1958) (page 6).

[16]A. Einstein, B. Podolsky and N. Rosen, "Can quantum-mechanical description of physical reality be considered complete?", Phys. Rev. **47**, 777–780 (1935) (pages 6, 38, 42).

[17]B. Hensen, H. Bernien, A. E. Dreau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellan, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau and R. Hanson, "Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres", Nature **526**, Letter, 682–686 (2015); M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-Å. Larsson, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, J. Beyer, T. Gerrits, A. E. Lita, L. K. Shalm, S. W. Nam, T. Scheidl, R. Ursin, B. Wittmann and A. Zeilinger, "Significant-loophole-free test of Bell's theorem with entangled photons", Phys. Rev. Lett. **115**, 250401 (2015); L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, K. J. Coakley, S. D. Dyer, C. Hodge, A. E. Lita, V. B. Verma, C. Lambrocco, E. Tortorici, A. L. Migdall, Y. Zhang, D. R. Kumor, W. H. Farr, F. Marsili, M. D. Shaw, J. A. Stern, C. Abellán, W. Amaya, V. Pruneri, T. Jennewein, M. W. Mitchell, P. G. Kwiat, J. C. Bienfang, R. P. Mirin, E. Knill and S. W. Nam, "Strong loophole-free test of local realism", ibid. **115**, 250402 (2015), page 6.

[18]V. B. Braginsky and A. B. Manukin, "Ponderomotive effects of electromagnetic radiation", Sov. Phys. JETP **25**, 653 (page 6).

[19]V. B. Braginsky, A. B. Manukin and M. Y. Tikhonov, "Investigation of dissipative ponderomotive effects of electromagnetic radiation", Sov. Phys. JETP **31**, 829 (page 6).

[20]A. Dorsel, J. D. McCullen, P. Meystre, E. Vignes and H. Walther, "Optical bistability and mirror confinement induced by radiation pressure", Phys. Rev. Lett. **51**, 1550–1553 (1983) (page 7).

[21] O. Arcizet, P.-F. Cohadon, T. Briant, M. Pinard and A. Heidmann, "Radiation-pressure cooling and optomechanical instability of a micromirror", Nature **444**, 71–74 (2006) (page 7).

[22] S. Gigan, H. R. Bohm, M. Paternostro, F. Blaser, G. Langer, J. B. Hertzberg, K. C. Schwab, D. Bauerle, M. Aspelmeyer and A. Zeilinger, "Self-cooling of a micromirror by radiation pressure", Nature **444**, 67–70 (2006) (page 7).

[23] A. Schliesser, P. Del'Haye, N. Nooshi, K. J. Vahala and T. J. Kippenberg, "Radiation pressure cooling of a micromechanical oscillator using dynamical backaction", Phys. Rev. Lett. **97**, 243905 (2006) (page 7).

[24] J. Chan, T. P. M. Alegre, A. H. Safavi-Naeini, J. T. Hill, A. Krause, S. Groblacher, M. Aspelmeyer and O. Painter, "Laser cooling of a nanomechanical oscillator into its quantum ground state", Nature **478**, 89–92 (2011) (pages 7, 242).

[25] M. Aspelmeyer, T. J. Kippenberg and F. Marquardt, "Cavity optomechanics", Rev. Mod. Phys. **86**, 1391–1452 (2014) (pages 7, 50).

[26] P. Meystre, "A short walk through quantum optomechanics", Annalen der Physik **525**, 215–233 (2013) (pages 7, 50).

[27] S. G. Hofer, W. Wieczorek, M. Aspelmeyer and K. Hammerer, "Quantum entanglement and teleportation in pulsed cavity optomechanics", Phys. Rev. A **84**, 052327 (2011) (pages 8, 41, 241).

[28] T. A. Palomaki, J. D. Teufel, R. W. Simmonds and K. W. Lehnert, "Entangling mechanical motion with microwave fields", Science **342**, 710–713 (2013) (pages 8, 242).

[29] M. D. Reid, "Demonstration of the Einstein–Podolsky–Rosen paradox using nondegenerate parametric amplification", Phys. Rev. A **40**, 913–923 (1989) (pages 9, 43, 44, 48).

[30] M. D. Reid and P. D. Drummond, "Quantum correlations of phase in nondegenerate parametric oscillation", Phys. Rev. Lett. **60**, 2731–2733 (1988) (page 9).

[31] M. D. Reid and P. D. Drummond, "Correlations in nondegenerate parametric oscillation: Squeezing in the presence of phase diffusion", Phys. Rev. A **40**, 4493–4506 (1989) (page 9).

[32] Z. Y. Ou, S. F. Pereira, H. J. Kimble and K. C. Peng, "Realization of the Einstein–Podolsky–Rosen paradox for continuous variables", Phys. Rev. Lett. **68**, 3663–3666 (1992) (page 9).

[33]M. D. Reid and D. F. Walls, "Violation of Bell's inequalities in quantum optics", Phys. Rev. Lett. **53**, 955–957 (1984) (page 9).

[34]E. G. Cavalcanti, C. J. Foster, M. D. Reid and P. D. Drummond, "Bell inequalities for continuous-variable correlations", Phys. Rev. Lett. **99**, 210405 (2007) (pages 9, 46, 243).

[35]E. Shchukin and W. Vogel, "Quaternions, octonions, and Bell-type inequalities", Phys. Rev. A **78**, 032104 (2008) (pages 9, 47, 243).

[36]T. J. Stieltjes, "Recherches sur les fractions continues", Ann. d. fac. d. sciences Toulouse **8**, J1–J122 (1894) (page 13).

[37]P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations* (Springer-Verlag, Berlin, 1992) (pages 13, 14, 18–20).

[38]K. Itō, "Stochastic integral", Proc. Imperial Acad. Tokyo **20**, 519–524 (1944) (page 14).

[39]P. Drummond and I. K. Mortimer, "Computer simulations of multiplicative stochastic differential equations", J. Comp. Phys. **93**, 144–170 (1991) (pages 21, 22).

[40]E. Schrödinger, "Der stetige Übergang von der Mikro- zur Makromechanik", **14**, 664–666 (1926) (page 23).

[41]R. Glauber, "Coherent and incoherent states of the radiation field", Phys. Rev. **131**, 2766–2788 (1963) (pages 23, 25).

[42]E. Sudarshan, "Equivalence of semiclassical and quantum mechanical descriptions of statistical light beams", Phys. Rev. Lett. **10**, 277–279 (1963) (page 25).

[43]K. Cahill and R. Glauber, "Ordered expansions in boson amplitude operators", Phys. Rev. **177**, 1857–1881 (1969) (page 26).

[44]J. E. Campbell, Proc. Lond. Math. Soc. **28**, 381–390 (1897); ibid. **29**, 14–32 (1897); H. Baker, Proc. Lond. Math. Soc. (1) **34**, 347–360 (1902); ibid. **35**, 333–374 (1903); Proc. Lond. Math. Soc. (Ser 2) **3**, 24–47 (1905); F. Hausdorff, Ber. Verh. Saechs. Akad. Wiss. Leipzig **58**, 19–48 (1906), page 27.

[45]E. P. Wigner, "On the quantum correction for thermodynamic equilibrium", Phys. Rev. **40**, 749–759 (1932) (page 28).

[46]K. Husimi, "Some formal properties of the density matrix", Proc. Phys. Math. Soc. Jpn. **22**, 264–314 (1940) (page 29).

[47]P. D. Drummond and C. W. Gardiner, "Generalised P-representations in quantum optics", J. Phys. A: Math. Gen. **13**, 2353–2368 (1980) (page 29).

[48]P. P. Deuar and P. D. Drummond, "Gauge P representations for quantum-dynamical problems: Removal of boundary terms", Phys. Rev. A **66**, 033812 (2002) (pages 31, 38).

[49]G. Lindblad, "On the generators of quantum dynamical semigroups", Comm. Math. Phys. **48**, 119–130 (1976) (page 32).

[50]A. Gilchrist, C. W. Gardiner and P. D. Drummond, "Positive P representation: Application and validity", Phys. Rev. A **55**, 3014–3032 (1997) (page 35).

[51]A. J. Leggett, "Bose-einstein condensation in the alkali gases: some fundamental concepts", Rev. Mod. Phys. **73**, 307–356 (2001) (page 37).

[52]M. Steel, M. K. Olsen, L. I. Plimak, P. D. Drummond, S. Tan, M. J. Collett, D. F. Walls and R. Graham, "Dynamical quantum noise in trapped Bose–Einstein condensates", Phys. Rev. A **58**, 4824–4835 (1998) (page 38).

[53]P. D. Drummond and P. Deuar, "Quantum dynamics with stochastic gauge simulations", Journal of Optics B: Quantum and Semiclassical Optics **5**, S281 (2003) (page 38).

[54]P. P. Deuar and P. D. Drummond, "Correlations in a BEC collision: First-principles quantum dynamics with 150 000 atoms", Phys. Rev. Lett. **98**, 120402 (2007) (page 38).

[55]E. Schrödinger, "Die gegenwärtige Situation in der Quantenmechanik", Naturwissenschaften **23**, 807–812 (1935) (pages 38, 40, 48).

[56]R. F. Werner, "Quantum states with Einstein–Podolsky–Rosen correlations admitting a hidden-variable model", Phys. Rev. A **40**, 4277–4281 (1989); A. Peres, "Separability criterion for density matrices", Phys. Rev. Lett. **77**, 1413–1415 (1996), pages 40, 50.

[57]J. Von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Springer, 1932); *Mathematical foundations of quantum mechanics* (Princeton University Press, 1955); U. Fano, "Description of states in quantum mechanics by density matrix and operator techniques", Rev. Mod. Phys. **29**, 74–93 (1957), page 40.

[58]S. M. Tan, "Confirming entanglement in continuous variable quantum teleportation", Phys. Rev. A **60**, 2752–2758 (1999) (page 41).

[59]L.-M. Duan, G. Giedke, J. I. Cirac and P. Zoller, "Inseparability criterion for continuous variable systems", Phys. Rev. Lett. **84**, 2722–2725 (2000) (page 41).

[60]V. Giovannetti, S. Mancini, D. Vitali and P. Tombesi, "Characterizing the entanglement of bipartite quantum systems", Phys. Rev. A **67**, 022320 (2003) (page 42).

[61]E. Schrödinger, "Discussion of probability relations between separated systems", Proc. Cambridge Philos. Soc. **31**, 555–563 (1935) (pages 43, 48).

[62]H. M. Wiseman, S. J. Jones and A. C. Doherty, "Steering, entanglement, nonlocality, and the Einstein–Podolsky–Rosen paradox", Phys. Rev. Lett. **98**, 140402 (2007); S. J. Jones, H. M. Wiseman and A. C. Doherty, "Entanglement, Einstein–Podolsky–Rosen correlations, Bell nonlocality, and steering", Phys. Rev. A **76**, 052116 (2007), pages 44, 48–50.

[63]E. G. Cavalcanti, S. J. Jones, H. M. Wiseman and M. D. Reid, "Experimental criteria for steering and the Einstein–Podolsky–Rosen paradox", Phys. Rev. A **80**, 032112 (2009) (pages 44, 49).

[64]M. D. Reid, P. D. Drummond, W. P. Bowen, E. G. Cavalcanti, P. K. Lam, H. A. Bachor, U. L. Andersen and G. Leuchs, "*Colloquium* : the Einstein–Podolsky–Rosen paradox: From concepts to applications", Rev. Mod. Phys. **81**, 1727–1751 (2009) (page 44).

[65]D. J. Saunders, S. J. Jones, H. M. Wiseman and G. J. Pryde, "Experimental EPR-steering using Bell-local states", Nat Phys **6**, 845–849 (2010) (page 44).

[66]L. Rosales-Zárate, R. Y. Teh, S. Kiesewetter, A. Brolis, K. Ng and M. D. Reid, "Decoherence of Einstein–Podolsky–Rosen steering", J. Opt. Soc. Am. B **32**, A82–A91 (2015) (pages 44, 242).

[67]Q. Y. He and M. D. Reid, "Einstein–Podolsky–Rosen paradox and quantum steering in pulsed optomechanics", Phys. Rev. A **88**, 052121 (2013) (page 44).

[68]B. Wittmann, S. Ramelow, F. Steinlechner, N. K. Langford, N. Brunner, H. M. Wiseman, R. Ursin and A. Zeilinger, "Loophole-free Einstein–Podolsky–Rosen experiment via quantum steering", New Journal of Physics **14**, 053030 (2012) (page 44).

[69]J. S. Bell, "On the Einstein–Podolsky–Rosen paradox", Physics **1**, 195–200 (1964); N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani and S. Wehner, "Bell nonlocality", Rev. Mod. Phys. **86**, 419–478 (2014), pages 6, 38, 45.

[70]D. Bohm and Y. Aharonov, "Discussion of experimental proof for the paradox of Einstein, Rosen, and Podolsky", Phys. Rev. **108**, 1070–1076 (1957) (page 45).

[71]J. F. Clauser, M. Horne, A. Shimony and R. Holt, "Proposed experiment to test local hidden-variable theories", Phys. Rev. Lett. **23**, 880–884 (1969) (page 46).

[72]D. M. Greenberger, M. Horne and A. Zeilinger, *Bell's theorem, quantum theory and conceptions of the universe*, edited by M. Kafatos (Springer, 1989), p. 348 (pages 47, 243).

[73]N. Gisin, "Bell's inequality holds for all non-product states", Physics Letters A **154**, 201–202 (1991) (page 50).

[74]M. Aspelmeyer, T. J. Kippenberg and F. Marquardt, *Cavity optomechanics - nano- and micromechanical resonators interacting with light* (Springer Berlin Heidelberg, 2014) (page 50).

[75]A. F. Pace, M. J. Collett and D. F. Walls, "Quantum limits in interferometric detection of gravitational radiation", Phys. Rev. A **47**, 3173–3189 (1993) (pages 50, 51).

[76]P. D. Drummond and M. Hillery, *The quantum theory of nonlinear optics* (Cambridge University Press, Cambridge, 2014) (pages 51, 52).

[77]C. W. Gardiner and P. Zoller, *Quantum noise*, 2nd (Springer, 2000) (page 51).

[78]S. Mancini and P. Tombesi, "Quantum noise reduction by radiation pressure", Phys. Rev. A **49**, 4055–4065 (1994) (page 51).

[79]P. D. Drummond and I. K. Mortimer, "Computer simulations of multiplicative stochastic differential equations", J. Comput. Phys. **93**, 144–170 (1991) (page 236).

[80]R. J. Lewis-Swan and K. V. Kheruntsyan, "Sensitivity to thermal noise of atomic Einstein–Podolsky–Rosen entanglement", Phys. Rev. A **87**, 063635 (2013) (page 242).

[81]R. P. Feynman, "Simulating physics with computers", Int. J. Theor. Phys. **21**, 467–488 (1982) (pages 242, 245).

[82]R. V. Bobryk, "Closure schemes in stochastic nonlinear dynamics: a validation case study", Physical Review E **83** (2011) (page 245).

[83]A. Georges, G. Kotliar, W. Krauth and M. J. Rozenberg, "Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions", Reviews of Modern Physics **68**, 13–125 (1996) (page 245).

[84]J. E. Mayer and E. Montroll, "Molecular distribution", The Journal of Chemical Physics **9**, 2–16 (1941) (page 245).

[85]O. E. Alon, A. I. Streltsov and L. S. Cederbaum, "Multiconfigurational time-dependent Hartree method for bosons: Many-body dynamics of bosonic systems", Phys. Rev. A **77**, 033613 (2008) (page 245).