Methods for Engineering Sub-Two-Cycle Mode-Locked Lasers

by

Jonathan R. Birge

B.S., Swarthmore College (1996)
M.S., University of Colorado at Boulder (1998)
S.M., Massachusetts Institute of Technology (2008)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2009

Author .....
Department of Electrical Engineering and Computer Science
May 22, 2009 [Corrections May 2, 2011]

Certified by .....
Franz X. Kärtner
Professor
Thesis Supervisor

Accepted by .....
Terry P. Orlando
Chairman, Department Committee on Graduate Theses

Methods for Engineering Sub-Two-Cycle Mode-Locked Lasers

by

Jonathan R. Birge

Submitted to the Department of Electrical Engineering and Computer Science
on May 22, 2009 [Corrections May 2, 2011], in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

We begin by presenting a method to efficiently solve for the steady-state solution of a nonlinear cavity, suitable for simulating a solid-state femtosecond laser. The algorithm directly solves the periodic boundary value problem by using a preconditioned Krylov-Newton shooting solver. The method can be applied to the design and study of mode-locked lasers, as well as the modeling of field enhancement cavities, such as those used in high harmonic generation. In constrast to the standard approach of dynamic simulation, which converges linearly, our algorithm converges quadratically to the stable solution, typically converging two to three orders of magnitude faster than the standard approach.

The second major theme is the control of dispersion in mode-locked lasers. The predominant way to design dispersion compensating optics in the past has been a consideration of the integrated net group delay dispersion (GDD). We propose and implement an alternative spectral quantity based on the energy contained in phase distortions, which we term the Phase Distortion Ratio (PDR). Dispersion compensating mirrors optimized with respect to PDR generally perform significantly better than those where GDD is optimized. We demonstrate this in the design of a dispersion compensating mirror pair capable of compressing single-single pulses.

In the final section, we deal with the unique challenges inherent to measuring sub-two-cycle pulses reliably and accurately. We have recently developed a technique, Two-Dimensional spectral Shearing Interferometry (2DSI), based on spectral shearing, which requires no calibration and does not disperse the pulse being measured. Our method intuitively encodes spectral group delay in a slowly changing fringe in a two-dimensional interferogram. This maximizes use of spectrometer resolution, allowing for complex phase spectra to be measured with high accuracy over extremely large bandwidths, potentially exceeding an octave. We believe that 2DSI is a uniquely cost effective and efficient method for accurately and reliably measuring few- and even single-cycle pulses. While the method is relatively recent, it is well tested and has been successfully demonstrated on several different lasers in two different groups, including one producing 4.9 fs pulses.

Thesis Supervisor: Franz X. Kärtner
Title: Professor

Acknowledgments

First, I'd like to thank my advisor, Prof. Franz Kärtner. His remarkable dedication to his advisees is apparent each day by the constant flow of students through his always open door. He was the ideal advisor, providing freedom and guidance in equal measure. I'd also like to express my appreciation to Profs. Erich Ippen and Leslie Kolodziejski for serving on my committee, being so generous with their scarce time and patient with the last minute scheduling changes. I would also like to thank all the committee members for taking the time to make so many corrections to the draft, and for providing much needed help with the writing.

Prof. Ian Walmsley provided extremely helpful feedback on the SPIDER calibration error work (Chapter 8).

Omid Nohadani started out as a valued collaborator (leading to the work shown in Chapter 7) and has ended up being an even more valued friend.

Christian Jirauchek has always provided great advice on any topic, just one example being his assistance with, and feedback on, the material in Chapter 6.

Working with Prof. Helder Crespo has been a great pleasure, and I owe many of the results in Chapter 9 to his expert laser engineering and design. Likewise, the initial implementation of 2DSI was only made possible by the extremely stable laser built by Richard Ell.

Throughout my stay at MIT I've benefitted greatly from the friendship of Jeff Chen, my fellow member of the class-of-2007-which-somehow-became-2009. Ariel Gordon was also a valued officemate, and one of the many people I've met at MIT who have inspired me to try a little bit harder by their example of scholarship.

Thanks to Andrew Benedick, Jeff Moses and Oli Mueke for many helpful and enjoyable discussions; I count their friendships as one of the best products of my time here.

Thanks to my Little Brother Jaysander, for being a loyal friend and for helping me keep everything in perspective.

And, of course, thanks to my "little" brother David, for helping keep me sane, and always asking about my research no matter how boring it likely was to hear about it.

Thanks to my mother, for teaching me that an accomplishment only means something if done with integrity; and to my father, for showing his young son the adventure of science. The times I spent following him around his lab as a child had more of an effect that he'll even know. In a sense, I'm always working in my father's lab.

Finally, and most importantly, thank you to Michele for being such a wonderful friend and understanding wife.

Curriculum Vitæ

Education

MIT, Ph.D. in Electrical Engineering and Computer Science, June 2009.
MIT, S.M. in Computation for Design and Optimization, February 2008.
University of Colorado at Boulder, M.S. in Electrical Engineering, June 1998.
Swarthmore College, B.S. in Engineering, June 1996.

Patents and Filings

1. J. R. Birge, R. Ell, F. X. Kärtner, “Two-dimensional Spectral Shearing Interferometry for Ultrafast Pulse Characterization,” US patent 7,433,043. Licensed to Menlo Labs, Inc.
2. J. Chen, M. Robinson, G. D. Sharp, J. R. Birge. “Three-panel color management systems and methods,” US patent 7,154,667.
3. J. R. Birge and G. Sharp. “Birefringent Networks,” US patent 7,154,667.
4. J. R. Birge and G. Sharp. “Method for Design of Birefringent Filters,” US invention filing, 2001.
5. J. Chen, M. Robinson, G. Sharp, J. R. Birge. “Compensated Color Management Systems and Methods,” US patents 6,186,309; 6,961,181; 6,961,179.
6. G. Sharp, M. Robinson, J. R. Birge, “Optical system for producing a modulated color image,” US patent 6,704,065.

Refereed Journal Publications

1. J. R. Birge, F. X. Kärtner, “Novel 2DSI method measures sub-two-cycle laser pulses,” Laser Focus World, March, 2009.*
2. Michelle Y. Sander, Jonathan Birge, Andrew Benedick, Helder M. Crespo, and Franz X. Kärtner, "Dynamics of dispersion managed octave-spanning titanium:sapphire lasers," J. Opt. Soc. Am. B 26, 743-749 (2009).
3. J. R. Birge, F. X. Kärtner, “Design and analysis of two-dimensional spectral shearing interferometry,” J. Opt. Soc. Amer. B, submitted, 2009.
4. L.-J. Chen, A. J. Benedick, J. R. Birge, M. Y. Sander, and F. Kärtner, "Octave-spanning, dual-output 2.166 GHz Ti:sapphire laser," Opt. Express 16, 20699-20705 (2008)
5. J. R. Birge, F. X. Kärtner, “Analysis and mitigation of systematic errors in spectral shearing interferometry for few cycle pulse measurement,” J. Opt. Soc. Amer. B, 25, 2008.*
6. H. M. Crespo, J. R. Birge, M. Y. Sander, E. L. Falcão-Filho, A. Benedick, and F. X. Kärtner, "Phase stabilization of sub-two-cycle pulses from prismless octave-spanning Ti:sapphire lasers," J. Opt. Soc. Am. B 25, B147-B154 (2008).
7. Omid Nohadani, Jonathan R. Birge, Franz X. Kärtner, and Dimitris J. Bertsimas, "Robust chirped mirrors," Appl. Opt. 47, 2630-2636 (2008).

8. H. M. Crespo, J. R. Birge, E. L. Falcão-Filho, M. Y. Sander, A. Benedick, and F. X. Kärtner, "Nonintrusive phase stabilization of sub-two-cycle pulses from a prism-less octave-spanning Ti:sapphire laser," Opt. Lett. 33, 833-835 (2008).
9. J. R. Birge, F. X. Kärtner, "Optimization of Multilayer Structures for Utlrafast Optics using Analytic Gradients of Dispersion," Appl. Opt. 46, 14, 2656-2662, 2007.
10. J. R. Birge, R. Ell, F. X. Kärtner, "Two-dimensional spectral shearing interferometry (2DSI) for ultrashort pulse characterization," Optics Letters, 2006.
11. R. Ell, J. R. Birge, M. Araghchini, F. X. Kärtner, "Carrier-envelope phase control by a composite glass plate," Optics Express 14, 5829-36, 2006.
12. R. Ell, G. Angelow, W. Seitz, M. Lederer, H. Huber, D. Kopf, J. R. Birge, F. X. Kaertner, "Quasi-synchronous pumping of mode-locked few-cycle Titanium Sapphire lasers," Optics Letters, 2006.
13. O. D. Mucke, R. Ell, A. Winter, J. Kim, J. R. Birge, L. Matos, F. X. Kartner, "Self-reference 200 MHz octave-spanning Ti:sapphire laser with 45 attosecond carrier-envelope phase jitter," Optics Express, 2005.
14. J. Birge, F. Kaertner, "Efficient Analytic Computation of Dispersion from Layered Media," Applied Optics 45, 1478, 2006.
15. J. Kim, J. R. Birge, V. Sharma, J. G. Fujimoto, F. X. Kaertner, V. Scheuer, and G. Angelow, "Ultrabroadband beam splitter with matched group-delay dispersion," Opt. Lett. 30, 1569, 2005.
16. G. Sharp, M. Robinson, J. Chen, J. Birge, "LCOS Projection Color Management Using Retarder Stack Technology," Displays, 2002.*

Refereed Conference Proceedings

1. J. R. Birge, F. X. Kärtner, "Design of Optimal Dispersive Mirrors for Femtosecond Enhancement Cavities and Compressors by Minimizing Phase Distortion Power," Conf. on Lasers and Electro-Optics, Baltimore, 2009.
2. S. J. Spector, T. M. Lyszczarz, M. W. Geis, D. M. Lennon, J. U. Yoon, M. E. Grein, R. T. Schulein, R. Amataya, J. Birge, J. Chen, H. Byun, F. Gan, C. W. Holzwarth, J. L. Hoyt, E. P. Ippen, F. X. Kärtner, A. Khilo, O. O. Olubuyide, J. S. Orcutt, M. Park, M. Perrott, M. A. Popović, T. Barwicz, M. Dahlem, R. J. Ram, and H. I. Smith, "Integrated Optical Components in Silicon for High Speed Analog-to-Digital Conversion," SPIE Proceedings Photonics West, San Jose, 2008.
3. J. Moses, O. D. Mucke, S. Huang, A. Benedick, E. Falcao-Filho, K. Hong, A. Siddiqui, J. R. Birge, F. Ilday, F. X. Kärtner, "Optimized 2-micron optical parametric chirped pulse amplifier for high harmonic generation," Ultrafast Phenomena, 2008.
4. J. R. Birge, H. Crespo, F. X. Kärtner, "Non-intrusive sub-two-cycle carrier-envelope stabilized pulses using engineered chirped mirrors," Conf. on Lasers and Electro-Optics, San Jose, 2008.
5. J. Moses, O. Mucke, A. Benedick, E. Falcao-Filho, S. Huang, K. Hong, A. Siddiqi, J. R. Birge, F. Ilday, F. X. Kärtner, "2-micron optical parametric chirped pulse ampli-

fier for long-wavelength driven high harmonic generation," Conf. on Lasers and Electro-Optics, San Jose, 2008.

6. A. Khilo, J. R. Birge, F. X. Kärtner, "Adaptive error compensation for photonic analog-to-digital converters," Conf. on Lasers and Electro-Optics, San Jose, 2008.
7. F. X. Kärtner, J. R. Birge, "Two-dimensional spectral shearing interferometry," OSA Annual Conference, Rochester, 2007.*
8. J. R. Birge, F. X. Kärtner, "A preconditioned Newton-Krylov method for computing stationary pulse solutions of mode-locked lasers," Conf. on Lasers and Electro-Optics, Baltimore, 2007.
9. A. Benedick, J. R. Birge, R. Ell, O. D. Mücke, M. Sander, F. X. Kärtner, "Octave Spanning 1GHz Ti:Sapphire Oscillator For HeNe CH4 Based Frequency Combs and Clocks," CLEO Europe, Munich, Germany, June 18-22, 2007.
10. J. R. Birge, " Designing Phase-Sensitive Mirrors by Minimizing Complex Error Energy in the Frequency Domain," in Optical Interference Coatings, OSA Technical Digest (CD) (Optical Society of America, 2007), paper WA7.
11. S. J. Spector, T. M. Lyszczarz, M. W. Geis, D. M. Lennon, J. U. Yoon, M. E. Grein, R. T. Schulein, R. Amataya, J. Birge, J. Chen, H. Byun, F. Gan, C. W. Holzwarth, J. L. Hoyt, E. P. Ippen, F. X. Kärtner, A. Khilo, O. O. Olubuyide, J. S. Orcutt, M. Park, M. Perrott, M. A. Popović, T. Barwicz, M. Dahlem, R. J. Ram, and H. I. Smith, "Integrated Optical Components in Silicon for High Speed Analog-to-Digital Conversion," SPIE Proceedings Photonics West, San Jose, Jan. 21-26, 2007.*
12. F. X. Kartner, A. Benedick, R. Ell, O. Mucke, J. Birge, M. Sander, "Octave-spanning lasers and carrier-envelope phase control," Conference on Advanced Solid-State Photonics (ASSP), Vancouver, CA, Jan28-31, 2007.
13. F. X. Kaertner, A. Benedick, J. Birge, and M. Sander, " Carrier-Envelope Phase Controlled Ultrashort Light Pulses for Nonlinear Optics," in Nonlinear Optics: Materials, Fundamentals and Applications, OSA Technical Digest (CD) (Optical Society of America, 2007), paper MB1.
14. J. R. Birge, R. Ell, F. X. Kärtner, "Two-dimensional spectral shearing interferometry (2DSI) for ultrafast laser optimization," Ultrafast Phenomena XV, Springer, 2006.
15. R. Ell, J. R. Birge, F. X. Kärtner, "Carrier-envelope phase control by a composite glass plate," CLEO, 2006.
16. J. R. Birge, R. Ell, F. X. Kärtner, "Two-dimensional spectral shearing interferometry (2DSI) for ultrashort pulse characterization," CLEO, Long Beach, 2006.
17. O. D. Mucke, R. Ell, A. Winter, J. Kim, J. R. Birge, L. Matos, F. X. Kartner, "Self-Referenced 200 MHz Octave-Spanning Ti:Sapphire Laser With 0.10 Radian Carrier-Envelope Phase Error," ESA International Workshop on Optical Clocks.
18. J. Birge, C. Jirauschek, F. Kaertner, "Efficient Analytic Computation of Group Delay Dispersion from Optical Inteference Coatings," Proc. OSA OIC Top. Mtg., Tucson, 2004.
19. G. Sharp, J. Chen, M. Robinson, J. Birge, "Skew Ray Compensated Retarder-Stack Filters for LCOS Projection," SID Symposium, Vol. 33, 2002.

20. G. Sharp. J. Birge, J. Chen and M. Robinson. “High Throughput Color Switch for Sequential Color Projection,” SID Symposium, Vol. 31, 2000.
21. M. Robinson, J. Korah, G. Sharp and J. Birge. “High Contrast Color Splitting Architecture Using Color Polarization Filters,” SID Symposium, Vol. 31, 2000.
22. G. Sharp and J. Birge. “Retarder Stack Technology for Color Manipulation,” SID Symposium, Vol. 30, 1999.
23. E. Cheever, J. Birge, D. Thomson, W. Santamore and D. George. “A Microprocessor-Based Multi-Channel Muscle Stimulator for Skeletal Muscle Cardiac Assist,” Proc. IEEE EMBS 17th Annual Conf. & 21st Can. Med. & Biol. Eng., Montreal, Canada, 1995.

(*Invited)

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

List of Figures

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

1.1 Background

The first demonstration of the laser in 1960 [4] began an era of increasingly precise control over light. In its simplest form, a laser consists of a gain medium within a low loss optical cavity. In theory, the single resonant frequency which experiences the most net gain is amplified in a laser to the point where the gain medium is saturated to match the cavity loss and only that one mode can survive.[1] However, even when multiple frequencies lase in a continuous wave (CW) laser, the modes have randomly changing relative phases and are spaced unevenly in the frequency domain; the only effect is to broaden the spectrum and reduce the coherence length of the laser, typically to something on the order of centimeters. For most purposes, therefore, conventional lasers can be considered monochromatic.

Mode-locked lasers produce light at remarkable physical extremes of both time and intensity, enabling them to probe and manipulate matter in ways never before available [5]. The few-cycle pulses produced by the best commercially available lasers are some of the shortest electromagnetic events ever created—nearly at the very limit of what is physically possible. The final frontier of laser development is the production of single-cycle pulses directly from an oscillator. Efforts are already underway at several laboratories to produce such pulses. One of the conclusions of this thesis is that perhaps the most important technological hurdle needed to overcome in this endeavor is improving the performance of dispersion compensating mirrors.

Optical pulses lasting on the order of femtoseconds allow for the probing of physical phenomena on a commensurate time scale, sufficient to resolve electronic relaxation processes in molecules or bonding dynamics [6], or control molecular quantum states to guide reactions [7]. With high energy pulses below three cycles or so, effects due to the *absolute* optical phase of the pulse start to appear for the first time, and techniques have emerged [8] to stabilize the carrier envelope phase (CEP) of such pulses relative to their envelopes. Another application of CEP stabilized pulses is in the frequency domain,

[1] In practice, various nonidealities such as inhomogeneous broadening and hole-burning conspire against true single-frequency operation, such that most lasers actually operate with a small cluster of frequencies.

where the stable frequency comb can be used as an optical "clock work" to convert optical frequencies to radio frequencies, allowing one to literally count optical cycles using standard RF electronics [8].

A corollary of the short time span within which the energy is localized in a mode-locked laser is that tremendous intensities are created. The peak focused output coming directly from a mode-locked lasers is already on the order of 10^{12} W/cm^2, which is on the scale of the binding energy of outer electrons in molecules and crystals. As such, the fields associated with the laser pulse are high enough to produce significant nonlinear polarization responses in most materials.[2] The peak intensities of commercially available *amplified* femtosecond lasers are over 10^{15} W/cm^2, sufficient to ionize matter. One novel application of such intense pulses involves creating a plasma in the interior of a transparent material, using the resulting localized heat source to write waveguides embedded in the interior [9].

Another very promising application of such precisely-controlled few-cycle light fields is the ionization, field acceleration and recombination of the outer electron of a noble gas atom to create UV and X-ray pulses, a process known as high harmonic generation (HHG) [10]. Soft X-ray pulses as low as 130 attoseconds [11] have been produced with this technique, which has the potential to create the first lab-scale source of coherent X-rays for use in molecular imaging. In this application, the X-ray production occurs over roughly one-cycle of the driving light field. It is therefore not surprising that the ideal source would be a two- or single-cycle pulse that is carrier envelope stabilized.

1.2 Overview of Thesis

The contributions of this thesis are primarily in three related areas integral to the design of sub-two-cycle mode-locked lasers: the design and measurement of optical filters which control the dispersion inside the cavity; the efficient numerical simulation of laser cavity models; and lastly, the reliable measurement of the pulses produced by such lasers.

1.2.1 Newton-Krylov Nonlinear Cavity Solver

The effective numerical solution of the steady-state solution of a nonlinear cavity is essential to the design and study of mode-locked lasers [1], as well as the modeling of field enhancement cavities, such as those used in high harmonic generation. The standard method for tackling such problems is to develop a numerical model of a cavity round trip, and then use the model to explicitly simulate the dynamic operation of the cavity in question until convergence is reached at some precision [12]. While this has the advantage of demonstrating self-starting and solution stability, dynamics simulation is rather poor when viewed as a numerical algorithm. Transients in cavities inherently decay exponentially, meaning that dynamic simulation exhibits linear numerical convergence to

[2] In a classical sense, the least bound electrons in the material are forced to travel far enough in their light induced oscillations to "sense" the structure of the molecule or crystal and deviate from a linear path.

the final solution. For example, a solid-state laser operating in the dispersion-managed soliton regime can take many thousands of round trips to converge.

In contrast, our algorithm converges quadratically to the stable solution, typically requiring the evaluation of less than a hundred round trips to converge to within numerical precision, often two to three orders of magnitude faster than with dynamic simulation [13]. This improvement in convergence rate is achieved by directly solving the periodic boundary value problem for the nonlinear cavity using a Newton-Raphson algorithm. At each Newton step we use a matrix-implicit, preconditioned Krylov subspace method to approximately solve the linearized problem by sending a series of trial pulses through the cavity. The preconditioning is critical to both enabling the quadratic convergence of the overall Newton iterations, as well as the efficiency of solving the linearized system as each step.

While matrix-free Krylov shooting methods have been previously applied to periodically driven RF systems [14, 15], this is the first application of such methods to a *passive* nonlinear system that exhibits periodicity as a result of nonlinear self-organization. Thus, in addition to the domain specific issue of preconditioning, this application is unique in that our period length is unknown at the outset and must be solved for along with the steady-state field.

As a matrix-implicit method, our approach only needs access to a function that computes the action of the nonlinear cavity on an arbitrary input field. As such, the method can be applied to existing problems with little additional effort from the user. In basic terms, our solver operates by sending a series of "trial" perturbations through the cavity, using the observed results to make a series of increasingly accurate estimates of the final solution. A solution can often be obtained to machine precision with less than 100 trial cavity evaluations, whereas the natural laser dynamics might take thousands of cavity round trips to converge to the same level. Our method enables several new opportunities for design and analysis of mode-locked lasers, such as putting the laser model inside an optimization loop or performing Monte Carlo statistical analyses.

1.2.2 Dispersion Compensation

Mode-locked lasers are only made possible by the engineering of anomolous dispersion devices which compensate for the dispersion of the gain material, such that the path length of the cavity as a whole is nearly achromatic. The most precise way to achieve complex negative dispersion curves is through the use of double-chirped dielectric mirror pairs, developed by Kärtner at MIT [16, 17].

Such mirrors are engineered by the use of computationally intensive nonlinear optimization routines, which benefit from the ability to efficiently and accurately compute the theoretical reflectivity and dispersion from a dielectric stack, as well as any associated gradients with respect to layer thickness. In the prior literature, the only method provided for computing dispersion was the use of finite differences, which do not always provide sufficient accuracy to effectly optimize a complex mirror. To fill this void, we have developed numerical methods to efficiently calculate dispersion from a stack using analytic methods and simplifying approximations.

Currently, the predominant approach to designing dispersion compensating optics

considers the group delay dispersion (GDD) as the primary optimization criterion. We propose and implement an alternative spectral quantity based on the energy contained in phase errors, which we term the Phase Distortion Ratio (PDR). This measure can be shown to be more physically relevent than GDD in many applications, and mirrors optimized with respect to PDR generally perform better than those where GDD is optimized (see Chapter 4). In particular, linear applications, most notably pulse compression and enhancement cavities, are provably optimized by the minimization of PDR.

In the context of mode-locked lasers, there has been a great deal of work on the effect of low-order dispersion (e.g. second- and third-order phases). However, little work has been done on the effects of high-order oscillations of phase on mode-locked lasers. Yet, these are the impairments normally dealt with in practice. We investigate these effects through the use of numerical simulation, and find that small phase ripple in the dispersion compensating mirrors are responsible for the spectral structure in current few-cycle lasers. In addition, we use the fast solver discussed above to address the issue of how to best design dispersion compensating filters for mode-locked lasers.

1.2.3 Ultrashort Pulse Metrology

Spectral shearing interferometry has become one of the principal methods used to measure few-cycle pulses. As is the case for any method, however, calibration and stability challenges arise with the standard the standard spectral shearing method (known as SPIDER [18]) as bandwidths approach the single-cycle. We have developed a version of spectral shearing [19] that addresses these issues and ensures that there is no possibility for uncalibrated phase.

Our technique, termed Two-Dimensional spectral Shearing Interferometry (2DSI), requires only the non-critical calibration of the shear frequency and does not perturb the pulse before up-conversion. Rather than encode the spectral group delay in a dense spectral fringe characteristic of other shearing methods, 2DSI encodes phase in a simple two-dimensional fringe by directly scanning the relative phase of the two spectrally sheared components. The lack of a spectral fringe reduces the resolution demands on the spectrometer to that required for proper sampling of the pulse itself, allowing for complex phase spectra to be measured with high accuracy over extremely large bandwidths, potentially exceeding an octave. We believe that 2DSI is a cost effective and efficient method for accurately and reliably measuring few- and even single-cycle pulses. While the method is relatively recent, it is well tested and has been successfully demonstrated on several different lasers, including one producing 4.9 fs pulses with 4.3 fs transform limited bandwidth [20, 21, 22].

Chapter 2

Physics of Ultrafast Optics

In this chapter we briefly cover some of the physics relevent to an understanding of mode-locked lasers. In particular, we discuss nonlinear wave propagation in the context of homogeneous media, deriving the very commonly used general nonlinear Schrödinger equation. We conclude with a very cursory overview of mode-locking, including the master equation model.

2.1 Nonlinear wave propagation

We begin by stating Maxwell's equations in differential form for the case of a nonmagnetic material that is source-free. In Gaussian units[1] the curls of the electric and magnetic fields are related by [23]

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{H}}{\partial t} \tag{2.1a}$$

$$\nabla \times \mathbf{H} = \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t}, \tag{2.1b}$$

where c is the vacuum speed of light and $\mathbf{D}$ is the electric displacement vector. Even though there are no free charges, the motion of paired charges in the material still matters and so we must introduce a separate vector $\mathbf{D}$, which can be viewed as taking into account the possibility of paired charge motion (hence its appearance solely in a time derivative). Optics is fundamentally the study of how charged particles interact at a distance. All of the complexity in optics really occurs in the charge dynamics of the material, and thus in some sense the most important equation is actually that relating $\mathbf{D}$ to $\mathbf{E}$, known as the constitutive relation:

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}(\mathbf{E}), \tag{2.2}$$

[1] Gaussian units are defined such that the electric and magnetic fields are related in terms of the speed of light in vacuum. This eliminates the notion of a vacuum permittivity by making $\mathbf{E}$ and $\mathbf{D}$ equivalent in vacuum. Similarly, vacuum permeability is gone, and $\mathbf{H} = \mathbf{B}$ in a nonmagnetic material.

where **P** is the polarization vector, which represents the material dipole moment at a given point in space. The polarization is driven by the electric field, and this is what allows for nontrivial solutions, especially when the relation between **P** and **E** is nonlinear, as it is in our case.

To derive a wave equation, we take the curl of (2.1a) and substitute in (2.1b) to eliminate the magnetic field, leaving

$$\nabla \times \nabla \times \mathbf{E} + \frac{1}{c^2}\frac{\partial^2 \mathbf{D}}{\partial t^2} = 0. \tag{2.3}$$

We can rewrite the first term of (2.3) by using the vector identity $\nabla \times \nabla \times \mathbf{E} = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E}$ to give us

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} + \frac{1}{c^2}\frac{\partial^2 \mathbf{D}}{\partial t^2} = 0. \tag{2.4}$$

In this thesis, we are concerned with the temporal propagation of a laser beam inside a material. Spatial effects will either be ignored or approximated by other means. We thus assume a transversely polarized infinite planewave propagating in the z direction in a homogeneous medium. This allows us to ignore the divergence of **E** in the first term of (2.4) and collapse the Laplacian to a spatial derivative in z. Furthermore, in the lasers we seek to model, the polarization is invariant so we may simplify (2.4) to the inhomogeneous scalar wave equation,

$$\frac{\partial^2}{\partial z^2}E(z,t) - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}D(z,t) = 0. \tag{2.5}$$

Returning to the question of how to model the material polarizability, we can assume that the vast majority of the polarization is linear with the electric field. Furthermore, given that laser materials are either glasses or centrosymmetric crystals, we can neglect even-order nonlinearities. We thus assume the only appreciable nonlinear polarization is third-order with the electric field. Furthermore, we are only interested in self-modulation, so that we consider only those third-order terms which result in the polarization being driven at the same frequency as the electric field. In the language of perturbation theory, this corresponds to two virtual transitions up in photon energy, one virtual transition down. Given a local monochromatic field written in phasor form, $Ee^{i\omega t} + \text{c.c.}$, this means that we consider the nonlinear contribution to the polarization phasor to be given by

$$P_{\text{NL}} = \chi^{(3)} E^* EE = \chi^{(3)} |E|^2 E. \tag{2.6}$$

In treating the polarizability $\chi^{(3)}$ as a constant scalar (in general, it is a fourth-rank tensor and a function of frequency) we imply an isotropic instantaneous nonlinearity. This is not entirely physical, of course, but it has proven in practice to be a reasonable approximation that captures the salient effects of soliton mode-locking [24]. The nonlinear polarization thus looks like the standard linear polarizability scaled by the amplitude of the field. As such, this effect can be roughly viewed as an "intensity dependent permittivity." We will thus proceed with our derivation of the wave equation by treating D as resulting from a linear filter operating on E plus a perturbative nonlinear term. We

define the transform of a function $G(z,t)$ as

$$\tilde{G}(z,\omega) = \int_{-\infty}^{\infty} dt\, G(z,t)\, e^{-i\omega t}. \tag{2.7}$$

In the frequency domain, the displacement vector D is related to the field E by

$$\tilde{D}(z,\omega) = [\epsilon(\omega) + \Delta\epsilon_{\mathrm{NL}}]\, \tilde{E}(z,\omega), \tag{2.8}$$

where we will treat the nonlinear term as a constant for the time being.

Taking the Fourier transform of (2.5) and substituting in (2.8) gives the harmonic wave equation

$$\frac{\partial^2}{\partial z^2}\tilde{E}(z,\omega) + [\epsilon(\omega) + \Delta\epsilon_{\mathrm{NL}}]\, \frac{\omega^2}{c^2}\tilde{E}(z,\omega) = 0,$$

where we have implicitly allowed the nonlinear term $\Delta\epsilon_{\mathrm{NL}} \propto |E|^2$ to pass through the time derivative despite it being a function of time. This is valid so long as $\partial_t|E| \ll \partial_t E$, not an entirely unreasonable assumption until the envelope starts to approach a single-cycle. The quantity in front of the second term in (2.1) has units of wavenumber, and it is simpler to just write

$$\frac{\partial^2}{\partial z^2}\tilde{E}(z,\omega) + \left[k^2(\omega) + \Delta k^2_{\mathrm{NL}}\right] \tilde{E}(z,\omega) = 0. \tag{2.9}$$

To consider pulse propagation, we assume our electric field can be written as a slowly varying envelope function modulating a monochromatic optical carrier at frequency ω_0,

$$E(z,t) = A(z,t)e^{i(k_0 z - \omega_0 t)} + \text{c.c.},$$

where k_0 is the carrier wave number. The analytic (positive frequency) transform of $E(z,t)$ is related to the transform of $A(z,t)$ by

$$\begin{aligned} \tilde{E}(z,\omega) &= \tilde{A}(z,\omega-\omega_0)e^{ik_0 z} + \tilde{A}^*(z,\omega+\omega_0)e^{-ik_0 z} \\ &\approx \tilde{A}(z,\omega-\omega_0)e^{ik_0 z}, \end{aligned} \tag{2.10}$$

where the approximation is obtained by noting that the envelope cannot, by construction, have significant components at optical frequencies. Finally, we make one more assumption, which is somewhat more subtle. We neglect any backwards traveling waves, and presume that we will follow the forward traveling wave (2.1) in a forward moving frame such that little change occurs in z. Thus, we may neglect all but the lowest order derivatives in z that operate on the envelope.[2] Substituting (2.10) into (2.9) and keeping

[2]This is often referred to as the slowly varying envelope approximation. However, were this really the approximation we were making, we would also have ignored higher-order time derivatives, which have been kept. In fact, it turns out that ignoring higher-order spatial derivatives, as we have done, has the sole effect of limiting us to waves traveling in one direction, and is actually an exact wave equation otherwise.

only the first-order spatial derivative yields

$$\frac{\partial}{\partial z}\tilde{A}(z,\omega) - i\left(\frac{k^2(\omega) + \Delta k_{\mathrm{NL}}^2 - k_o^2}{2k_0}\right)\tilde{A}(z,\omega) = 0. \tag{2.11}$$

If we let $k(\omega_0) = k_0$, then $k(\omega)$ will not deviate much from k_0 over the pulse bandwidth. We can thus write the quantity in brackets in terms of a series expansion of $k(\omega)$,

$$\begin{aligned}\left(\frac{k^2(\omega) + \Delta k_{\mathrm{NL}}^2 - k_o^2}{2k_0}\right) &\approx k(\omega) + \Delta k_{\mathrm{NL}} - k_o \\ &= -ig_0 + \Delta k_{\mathrm{NL}} + \frac{1}{v_g(\omega_0)}(\omega - \omega_0) + \sum_{n=2}^{\infty}\frac{k_n}{n!}(\omega - \omega_0)^n.\end{aligned} \tag{2.12}$$

We have specifically labeled a small signal gain term $g_0 \equiv ik_0$ and a group velocity $v_g(\omega_0) \equiv 1/k_1$. It is now apparent that our approximate treatment of the nonlinearity (specifically, neglecting any time derivatives acting on the nonlinear polarization) restricts the nonlinearity to only affecting the phase velocity. Again, this is only valid to the extent that the pulse envelope varies much more slowly than the opical field. Rewriting (2.11) in terms of (2.12), we have

$$\frac{\partial\tilde{A}}{\partial z} = g_0\tilde{A} + i\Delta k_{\mathrm{NL}}\tilde{A} + \frac{i}{v_g}(\omega - \omega_0)\tilde{A} + i\sum_{n=2}^{\infty}\frac{k_n}{n!}(\omega - \omega_0)^n\tilde{A}. \tag{2.13}$$

Taking the inverse Fourier transform of (2.13) and applying the derivative theorem, the powers of ω become temporal derivatives, and the time domain wave equation is

$$\frac{\partial A}{\partial z} = g_0 A + i\delta|A|^2 A - \frac{i}{v_g}\frac{\partial A}{\partial t} - i\sum_{n=2}^{\infty}\frac{k_n}{n!}\frac{\partial^n A}{\partial t^n}, \tag{2.14}$$

where we've expressed the nonlinear wavenumber perturbation in terms of the intensity, as per (2.6), and an effective parameter δ known as the nonlinear coefficient.

Given that, in practice, we are actually dealing with focused beams and not the ideal plane waves assumed here, δ can be regarded as an empirically determined parameter that includes spatial effects. The nonlinear term acts in quadrature, and simply advances the phase of the field in proportion to the local intensity. The effect due to this term is thus termed self phase modulation (SPM).

Finally, we complete the derivation by rewriting the equation in a frame that is moving with the group velocity, by performing a change of variables into a "local" time $T \equiv t - z/v_g$, and writing the envelope in terms of the local field $u(z,T)$. The result of this transformation is a tilting of the solution in the z-t plane, such that time derivatives are unchanged, but spatial derivatives of A become mixed time and space derivatives in u. Application of the chain rule shows that

$$\frac{\partial A}{\partial z} = \frac{\partial u}{\partial z} - \frac{1}{v_g}\frac{\partial u}{\partial T}.$$

Using a moving frame thus eliminates the group delay delay term, and so the final propagation equation is

$$\boxed{\frac{\partial u}{\partial z} = g_0 u + i\delta|u|^2 u - i\sum_{n=2}^{\infty} \frac{k_n}{n!}\frac{\partial^n u}{\partial T^n}.} \tag{2.15}$$

All series coefficients k_n are assumed to be complex to account for a spectrally arbitrary gain (or loss) mechanism. This equation is known as the generalized nonlinear Schrödinger equation (GNLSE).

2.2 Mode-locking

In a laser consisting of only linear elements, theory states that the longitudinal cavity mode which experiences the strongest gain saturates the amplifying medium [25], leaving it as the only surviving mode.[3] However, adding a strong enough nonlinear filter to the cavity can couple the cavity modes, causing multiple frequencies to not only simultaneously lase, but to "line up" on a regular frequency grid and lock in phase so that they form a uniform comb of frequencies, producing a train of short pulses. This mode of pulsed laser operation is termed mode-locking, and is a prime example of nonlinear self-organization [26]. Given that carrier frequencies of light are on the order of PHz, locking of optical modes allows for the creation of pulses far shorter than those attainable with electronics with negligible relative bandwidth.

The most basic nonlinear mechanism that can be used to promote mode-locking is so-called saturable absorption, whereby a material is placed in the cavity that attenuates light at a rate inverse to the light intensity [27]. In 1966, DeMaria [28] used a saturable absorber dye placed in the cavity of a Nd:glass laser to produce pulses on the order of 10 ps, the first pulses to be produced that were shorter than could be resolved by electronic methods. Thus began the field of ultrafast optics, and a path of rapid progress towards shorter optical pulses.

An important step along the way was the fortuitious discovery of Kerr-lens mode-locking [29]. Self-focusing in the gain medium causes high intensity components to increase their spatial overlap with the pump laser, thus preferentially amplifying pulses over the continuum. This has the net effect of acting as a very fast saturable absorber [30], supporting pulses on the order of tens of femtoseconds even without the help of additional nonlinear broadening.

Researchers later realized that the shortest pulses will be produced by cavities which exhibit a soliton-like shaping mechanism [31], whereby negative dispersion and self-phase modulation will interact to generate extra bandwidth and stabilize the resulting pulse. If the gain filtering, lumped dispersion and nonlinearity in the cavity perturb the pulse little during a round trip, the effects can be considered effectively distributed and the resulting pulse very closely resembles a soliton solution to the GNLSE (2.15). The nonlinear pulse shaping in a soliton laser dominates the saturable absorption shaping,

[3] In practice, this is not strictly true, as spatial hole burning and other nonidealities can cause a narrow band of modes to lase, but this occurs in such a narrow band that most lasers can be well approximated as monochromatic.

leading to pulses that can be much shorter than the SA recovery time, enabling sub-picosecond pulses even from slow saturable absorbers [32]. In soliton lasers, the SA mechanism is needed simply to ensure the stability of the laser against noise, and for self-starting.

The evolution of mode-locking culminated in the 1990s with the development of a variant of the soliton laser, the so-called "dispersion managed soliton" laser [33], explained further in Section 2.2.2. Such lasers are capable of producing pulses below 5 femtoseconds, so short that they contain less than two cycles of the optical field and have bandwidths spanning hundreds of THz [34, 21].

2.2.1 Soliton modelocking master equation

The analytic theory of mode-locking is largely due to the efforts of Hermann Haus, while at Bell Labs and MIT, who formulated a master equation of mode-locking by considering the pulse shaping mechanisms of the laser as a distributed system. To deal with the solitonic regime, the master equation of fast saturable absorber mode-locking was modified by Martinez to include SPM and dispersion [35],

$$\frac{1}{T_R}\frac{\partial}{\partial T}u = \left(g_0 - \ell\right)u + (D_g - iD_2)\frac{\partial^2}{\partial t^2}u + (\gamma + i\delta)|u|^2 u. \tag{2.16}$$

This equation is very similar to the GNLSE (2.15) derived in Section 2.1. We use much the same notation, save that now t is the local time in the moving frame, γ is the SA coefficient, and T is the longer-scale time representing the propagation through the cavity, analogously to z in the GNLSE (2.15). The cavity round trip time is T_R. The cavity is assumed to have only second-order dispersion, given by D_2, and quadratic gain filtering, represented by $D_g = g_0/\Omega_g^2$, where Ω_g is the gain bandwidth. This equation has an analytic steady-state solution [36, 35],

$$u(t) = A_0 \operatorname{sech}\left(\frac{t}{\tau}\right)^{1+i\beta}, \tag{2.17}$$

where β is a chirping parameter and τ the pulse width.

Insertion of the ansatz (2.17) into the master equation (2.16) reveals several interesting relationships between the pulse and the cavity parameters. Of chief relevence to this thesis are the pulse width and the chirp. The latter is given by

$$\beta = \frac{3 - 3D_n\delta_n \pm \sqrt{9 - 2D_n\delta_n + 8\delta_n^2 + D_n^2(8 + 9\delta_n^2)}}{2(D_n + \delta_n)}, \tag{2.18}$$

with the normalized quantities $D_n = D_2/D_g$ and $\delta_n = \delta/\gamma$. The pulse width is a function of the chirp (and therefore the nonlinearities) and the dispersions,

$$\tau = \frac{\tau_0}{2}\left(2 - \beta^2 - 3\beta D_n\right), \tag{2.19}$$

with τ_0 the width of the solely saturable absorber mode-locked laser, which is a function

of the dispersion, saturable absorption, and the pulse energy: $\tau_0 = 4D_2/(\gamma W)$. Plugging (2.18) into (2.19) yields a rather dense expression. However, to gain some insight, we can take a series expansion of (2.19) to second order in D_n and δ_n about zero, applicable for small dispersions and nonlinearities,

$$\tau = \tau_0 \left(1 + \frac{15}{27} D_n \delta_n - \frac{6}{27}\delta_n^2 + \frac{21}{27} D_n^2 + \frac{1}{27} D_n^2 \delta_n^2\right) + O[D_n^3 + \delta_n^3]. \tag{2.20}$$

From this we learn that the pulse width decreases with increasing SPM, so long as the dispersion is not allowed to grow too large. Solving for the extremum of (2.20) shows that the shortest pulses will be found at slightly negative dispersions,

$$D_n^* = -\frac{15\delta_n}{2\left(21 + \delta_n^2\right)}. \tag{2.21}$$

From the above, we see that when the SPM is set to zero (SA only) the shortest pulses are found with zero dispersion. For finite SPM, which is always positive, slightly negative dispersion yields the shortest pulses.

2.2.2 Dispersion-managed soliton mode-locking

A dispersion-managed soliton (DMS) laser consists of three primary sections: two negative (anomalous) dispersion regions surrounding a central amplifying region with positive dispersion and self-phase modulation [33], Fig. 2-1. Briefly, the effect of the alternating dispersion is to create a negatively dispersed field as the pulse enters the gain material. When the negatively chirped pulse experiences the SPM of the gain medium, it causes the spectral bandwidth to compress during its travel through the first half of the crystal. This has the effect of squeezing the spectrum to fit the gain bandwidth, thus allowing the laser to support a greater bandwidth than its gain bandwidth would otherwise allow. The discovery of this mechanism was the key breakthrough that allowed femtosecond lasers to operate well below ten femtoseconds.

The dispersion of the whole cavity is balanced such that the dispersion is exactly reversed in sign by the gain material. In a modern laser, negative dispersion is typically achieved by dispersion compensating mirrors which are engineered to take longer to reflect “red” light than “blue” light, discussed further in later chapters. The second half of the crystal expands the spectrum and reverses the sign of the chirp. This phenomenon is illustrated in Fig. 2-2.

In addition, the cavity also must contain a fast saturable absorber of some sort, which supports the initiation of mode-locking and stabilizes the pulses. The shortest pulses are created by lasers which utilize a saturable absorption mechanism which acts instantaneously, with no appreciable recovery time. In practice, fast saturable absorption is achieved through an effect known as Kerr lensing, whereby the spatial self focusing of the beam causes temporal regions with greater intensities to focus themselves into the highly pumped region of the gain material, thereby experiencing greater gain.

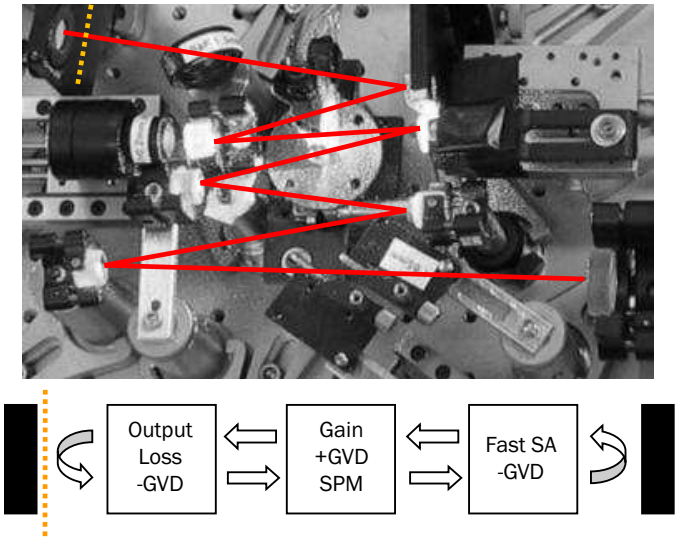

Figure 2-1: An example of a Kerr lens mode-locked laser (top) and a schematic of the 1D model for the cavity used as our test problem. (GVD: group velocity dispersion; SA: saturable absorber; SPM: self-phase modulation.)

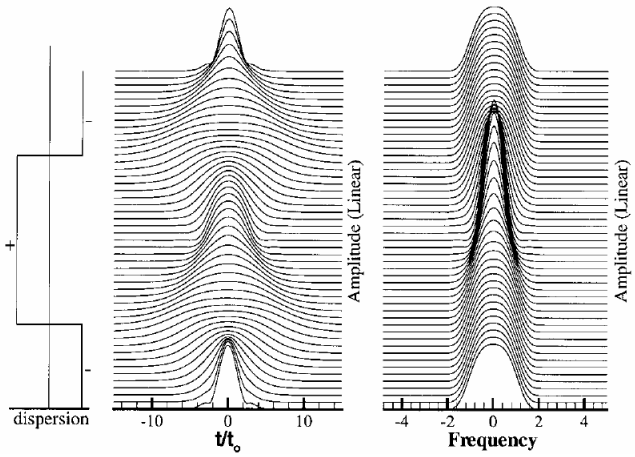

Figure 2-2: Illustration of the pulseshaping mechanism of a dispersion managed soliton laser over one round-trip. (Taken from [1].)

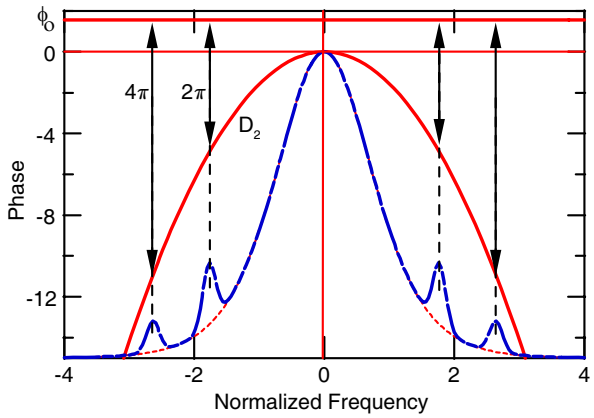

Figure 2-3: Schematic of the origin of Kelly sidebands [2]. Light coupled from the soliton into the continuum by a periodic perturbation is resonant with the soliton when the soliton phase shift and the linear cavity phase differ by a multiple of 2π. (Taken from [3].)

2.3 Results from soliton perturbation theory

Soliton perturbation theory, largely due to the efforts of Haus, was primarily developed to study the effects of noise in solitons [37]. Such issues are beyond the scope of this thesis. However, one of the applications of soliton perturbation theory is a fundamental understanding of the origin of resonant spectral sidebands, which sometimes appear in periodically perturbed soliton systems. A DMS laser, having discrete positive and negative dispersion elements, is susceptible to such resonant purturbative effects. In fact, as discussed later in Chapter 5, such effects explain several unique properties of DMS lasers operating in the positive dispersion regime. Below, we briefly summarize a few relevent results of soliton perturbation theory.

The central idea of soliton perturbation theory is that any change to either the soliton or the system manifests as either perturbations to the soliton degrees of freedom (phase, timing, frequency and amplitude) or as energy transferred to the continuum. A dispersion managed soliton cavity is, of course, a lumped approximation to the NLSE and thus constitutes a series of periodic perturbations to the steady-state "soliton". The amount of energy transferred to the continuum is small at each round trip. However, if the soliton phase differs from the round trip cavity phase by a multiple of 2π, then power can be resonantly pulled from the soliton, leading to a significant buildup of power in a narrow spectral range. This phenomenon is referred to as Kelly sidebands, after the person who first explained their mechanism [2].

Figure 2-3 illustrates Kelly sidebands in a soliton system with pure negative second-order dispersion, though it can happen with any dispersion profile that supports a stable pulse, including one with positive dispersion. A similar phenomenon happens for dispersion managed solitons in the positive dispersion regime, as shown in Chapter 5, where the net phase difference can actually go to zero.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 3

Newton-Krylov Cavity Solver

The standard method of dynamic simulation converges linearly, requiring thousands of propagations through the cavity (demonstrated in Section 3.3). In fact, some lasers require 50,000 or more round trips to converge to significant precision. It is generally the rule that mode-locked lasers perform optimally in regions of marginal stability (i.e. eigenvalues close to zero), further limiting the efficacy of dynamic simulation as a solution method as transients decay slowly.

By treating the solution of the stationary cavity mode as a nonlinear problem to be solved using Newton's method, we hope to converge directly to the solution quadratically.[1] We have no way to efficiently compute the Jacobian of the cavity, however, and so we use a matrix-free shooting method [14], tuned to the specifics of our problem. In many cases, Newton-based shooting methods do not require preconditioning [15], though we find that in our case, the shooting problem is poorly conditioned and does not converge efficiently (or at all) without preconditioning. We derive a diagonal preconditioner that improves the problem conditioning by orders of magnitude. Numerical experiments show that our method converges an order of magnitude faster than dynamic simulation for the kinds of highly nonlinear dispersion-managed soliton cavities employed in few-cycle lasers, and up to four orders of magnitude faster for lasers with less round trip perturbation.

3.1 Split-step method

The GNLSE can be numerically integrated efficiently by a pseudospectral method known as the split-step method [12]. This is the method we use to propagate the pulse through bulk media. In addition, the mathematics behind the split-step method offer some insight into the preconditioning scheme we use in the Krylov solver. It is thus useful to begin with a brief discussion of the split-step method before introducing the cavity solver.

[1] In the context of this chapter, we define computational complexity in terms of round-trip evaluations (i.e. calls to the cavity simulation function).

3.1.1 Derivation

To begin with, we express (2.15) in terms of two operators,

$$\frac{\partial}{\partial z}u(z,T) = \left(\hat{D}+\hat{N}\right)u(z,T), \tag{3.1}$$

where $\hat{D}$ represents all the linear terms (i.e. the gain/loss and dispersion) and $\hat{N}$ is the SPM operator,

$$\hat{D} = g_0 - i\sum_{n=2}^{\infty}\frac{k_n}{n!}\frac{\partial^n}{\partial T^n}, \tag{3.2}$$

$$\hat{N} = i\delta|u|^2. \tag{3.3}$$

The key to the split-step method is the recognition that the dispersion operator is a diagonal matrix in the Fourier domain, and the SPM operator is diagonal (and purely imaginary) in the time domain. Thus, each can be propagated *exactly* in their respective domains with a single exponential. We can efficiently compute an approximate propagation over a distance h by first handling the dispersion alone, and then transforming to the time domain and having the nonlinearity act,

$$u(z+h,T) \approx \exp(h\hat{N})\exp(h\hat{D})u(z,T). \tag{3.4}$$

This single step requires only two Fast Fourier Transform (FFT) operations, as the exponential of a diagonal matrix is a a vector operation. As will be proven in the following section, the accuracy of the iteration can be improved by symmetrizing it to yield the following iteration

$$u(z+h,T) \approx \exp\left(\frac{h}{2}\hat{D}\right)\exp(h\hat{N})\exp\left(\frac{h}{2}\hat{D}\right)u(z,T). \tag{3.5}$$

Other than requiring a single extra dispersion propagation at the end of a computation, this requires no more computation than (3.4), and is thus the scheme used in practice (see Appendix B.2.1).

3.1.2 Convergence

To ascertain the convergence of the symmetric iteration in (3.4), we first consider the exact solution to (3.1) in terms of the operator exponential,

$$u(z+h,T) = \exp\left[h(\hat{D}+\hat{N})\right]u(z,T), \tag{3.6}$$

where $\hat{N}$ is assumed to be invariant in z. The fact that $\hat{N}$ is nonlinear and does not commute with $\hat{D}$ implies that it will, in fact, vary with space. However, for small nonlinearities this will be a negligible effect compared to the error caused by the fact that the split-step method treats everything as commuting operators, and it is this error we are

analyzing.[2]

We can rewrite the product of the exponentials in the split-step iteration (3.5) by breaking the center time-domain step in to two equal pieces and applying the Baker-Hausdorff lemma [38] to each pair to yield

$$\exp\left(\frac{h}{2}\hat{D}\right)\exp(h\hat{N})\exp\left(\frac{h}{2}\hat{D}\right) = \exp\left(\frac{h}{2}\hat{D}+\frac{h}{2}\hat{N}+\frac{1}{2}\frac{h^2}{4}[\hat{D},\hat{N}]+O[h^3]\right) \times \exp\left(\frac{h}{2}\hat{N}+\frac{h}{2}\hat{D}+\frac{1}{2}\frac{h^2}{4}[\hat{N},\hat{D}]+O[h^3]\right). \quad (3.7)$$

The above shows that the error in each half step is $O[h^2]$ due to the fact that the two operators do not commute. Applying the lemma a second time to the right hand side of the above allows the commutators to cancel, since $[a,b]+[b,a]=0$, leaving us with only third-order terms[3] in excess of the exact solution,

$$\exp\left(\frac{h}{2}\hat{D}\right)\exp(h\hat{N})\exp\left(\frac{h}{2}\hat{D}\right) = \exp\left(h\hat{D}+h\hat{N}+O[h^3]\right) \quad (3.8)$$

$$= \exp\left(h\hat{D}+h\hat{N}\right)+O[h^3]. \quad (3.9)$$

Comparing the above with (3.6), we can see that the symmetric split-step method converges as $O[h^3]$.

Another advantage of the split-step approach is that it allows us to handle each effect in the most natural basis. As such, dispersion and gain are not, in actuality, expressed in terms of series coefficients as in (2.15), but are instead simply represented in spectral form as the complex elements of the diagonal matrix representation of $\hat{D}$.

3.2 Laser cavity numerical model

In this thesis, we'll consider a dispersion-managed soliton laser as a test case, considering only the time domain and neglecting any spatial effects (such as self-focusing or diffraction). Thus, we consider only a complex analytic pulse envelope $u(z,T)$, as discussed in Section 2.1. The cavity model we will use is that shown in Figure 2-1.

3.2.1 Dispersion

In our model, the net cavity dispersion is expressed in terms of second- and third-order series coefficients in ω, written as D_2 and D_3, respectively. Where there are no nonlinear elements, dispersive elements (e.g. air, mirrors) may be treated trivially in lumped form.

[2] When actually performing the split-step method, of course, we allow $\hat{N}$ to vary at each step.

[3] The third order terms not explicitly shown in (3.7) involve commutators of commutators, and thus the switching of the order of operators which allowed the second-order terms to cancel does not cause the third-order terms to cancel.

In the Fourier domain this operation is written as

$$U'(\omega) = \exp\left[-i\left(D_2\omega^2 + D_3\omega^3\right)\right] U(\omega), \tag{3.10}$$

where U is the field before and U' after the dispersive element. It would be trivial to handle arbitrary spectral dispersion profiles in the simulation, but for the sake of simplicity in this test model, we limit ourselves to two series terms.

3.2.2 Gain Material

We model the gain material via the GNLSE (2.15), handling propagation using the aforementioned split step. For the levels of nonlinearity typically encountered in a mode-locked laser, a discretization of 30 steps is generally sufficient. The gain is presumed to saturate as a function of the total intracavity power P, given by

$$P = \frac{1}{T_R}\int dT|u|^2, \tag{3.11}$$

with T_R the round trip time. The effect of the gain over a spatial step Δz is handled in the Fourier domain by

$$U(z+\Delta z,\omega) = \exp\left(\frac{g_0}{1+P/P_{\text{sat}}}\frac{\Delta\omega}{\omega+\Delta\omega}\frac{\Delta z}{\ell}\right)U(z,\omega), \tag{3.12}$$

with g_0 the small signal gain, and $\Delta\omega$ the gain bandwidth, P_{sat} the saturation power.

Given the focusing that is occurring in the gain medium, the nonlinear parameter γ is technically a function of z. However, to simplify things we specify the nonlinearity in terms of an empirically determined net nonlinear phase per unit of intensity per pass, known as the SPM coefficient δ. Nonetheless, we still consider the SPM in distributed form, acting throughout the gain crystal. The effect of self phase modulation on the envelope is then given by

$$u(z+\Delta z,T) = \exp\left[i\delta|u(z,T)|^2\frac{\Delta z}{\ell}\right]u(z,T). \tag{3.13}$$

The split-step method is thus implemented by alternating between (3.12) and (3.13).

3.2.3 Fast Saturable Absorber

The full spatio-temporal Kerr lensing mechanism is a prohibitively complex process to model (especially given the standard simulation methods which this thesis aims to improve upon) and thus saturable absorption is not modeled physically, but phenomenologically, using a simple lumped model given by

$$u'(T) = \exp\left(\frac{-q}{1+|u(t)|^2/I_{\text{sat}}}\right)u(T), \tag{3.14}$$

Table 3.1: Parameters for model Titanium:sapphire laser

Parameter	Value
T_R	$1/100\,\text{MHz}^{-1}$
$D^{(2)}_{\text{net}}$	$0.25\,\text{fs}^2$
g_0	0.1
$D^{(2)}_{\text{gain}}$	$232\,\text{fs}^2$
$D^{(3)}_{\text{gain}}$	$160\,\text{fs}^3$
$\Delta\omega$	60 THz
P_{sat}	1 W
δ	$20\,\text{rad/GW}^{-1}$
q	-0.02
I_{sat}	0.3 MW
l_{OC}	0.02

where q is the unsaturated absorption and I_{sat} is the saturation intensity. Both parameters are determined empirically by comparing model results with actual lasers.

3.3 Titanium:sapphire test model

To test our model and solver, we implement a simulation of a typical Titanium:sapphire mode-locked laser. The model parameters we use as a baseline are presented in Table 3.1. For sake of convenience, dispersion is specified in terms of gain material dispersion and the net dispersion of an entire round-trip. We discretize the field as a 256 element vector representing a window of 200 femtoseconds, evaluated at a z slice right before the output coupler (to simulate the pulse as exiting the laser cavity).

In Fig. 3-1, we show the evolution of the cavity over 2000 round-trips, starting from random noise (representing spontaneous emission). The resulting pulse is slightly asymmetric and has a duration (as measured by its full width half maximum) of 8.56 fs. The evolution of the residual, as defined by the "energy" of the difference between the input and output of the cavity relative to the energy of the pulse, is shown in Fig. 3-2. After a period of oscillation as transients die down, the simulation converges linearly overall toward the steady state solution, as expected.

3.4 Cavity Solver Problem Statement

We regard the cavity as a nonlinear operator $\mathbf{g}(\mathbf{u})$ acting on a vector representing the Fourier coefficients of the field in our temporal window. We operate in the basis of the Fourier modes because they are eigenvectors of dispersion and gain, and thus we can trivially invert those operators of the cavity.

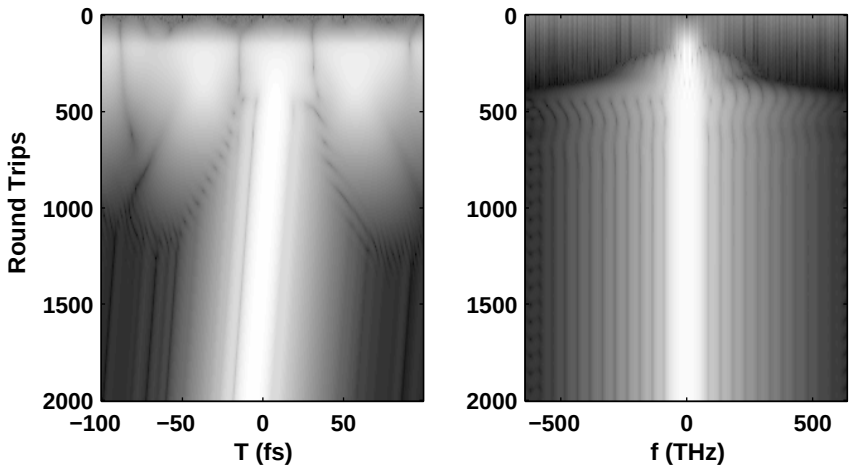

Figure 3-1: Evolution of model laser starting from noise, shown on a log scale. The field intensity as a function of time is on the left, with the power spectral density shown on the right. The pulse does not exactly follow our moving window because of nonlinear effects shifting the spectrum from our assumed center frequency.

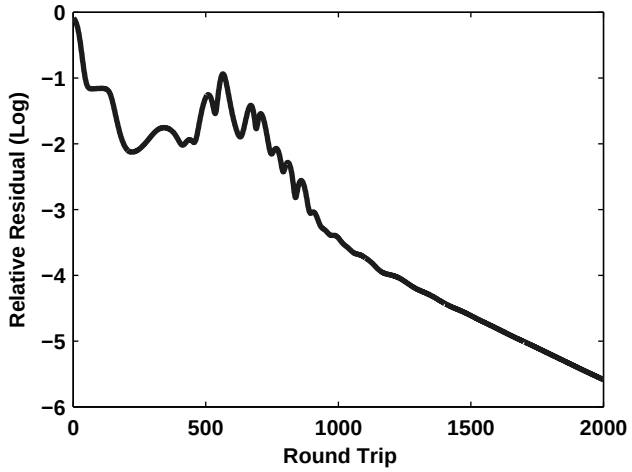

Figure 3-2: Convergence of dynamic cavity evolution. Once the initial transients die, the convergence of the pulse shaping mechanism is linear.

It's important to note that $\mathbf{u}$ need only be sufficient to represent the solution. We can project this vector in and out of a higher dimensional space to perform the actual computation. In fact, given the "breathing" nature of dispersion-managed soliton lasers, for example, it may be necessary to propagate using a much larger temporal window than is necessary to describe the solution.

To solve the problem, we seek an an eigenvector $\mathbf{u}$ to a nonlinear operator representing the change in the pulse over a round trip,

$$\phi(\mathbf{u})\mathbf{g}(\mathbf{u}) \equiv \mathbf{f}(\mathbf{u}) = \mathbf{u}, \tag{3.15}$$

where $\phi(\mathbf{u})$ is a function that takes care of normalizing out a constant and linear phase,

$$\phi_i(\mathbf{u}) = \exp\left[-i(\Delta\phi(\mathbf{u}) - \Delta v_g(\mathbf{u})\omega_i)\right]. \tag{3.16}$$

Due to the presence of nonlinearity, we cannot predict ahead of time what the actual periodicity of the cavity will be, and need to account for a potential perturbation Δv_g to the group delay already assumed in Section 2.1. We also normalize the overall phase of our solution by such that the DC component of the field experiences no phase change. More details can be found in Appendix B.2.2. The core of our method is the repeated solution of Newton-Raphson linearized subproblems given by

$$\left[\mathbf{J_f}(\mathbf{u_k}) - \mathbf{I}\right](\mathbf{u}_{k+1} - \mathbf{u}_k) = -\left[\mathbf{f}(\mathbf{u}_k) + \mathbf{u}_k\right], \tag{3.17}$$

with the cavity Jacobian $\mathbf{J_f}$ defined by

$$(\mathbf{J_f})_{ij} \equiv \frac{\partial f_i(\mathbf{u})}{\partial u_j}. \tag{3.18}$$

With the phase normalization described above, (3.17) becomes singular and so we must use an iterative method to solve it. The bulk of the work in the remainder of this chapter deals with the efficient solution of this problem.

3.5 Jacobian Properties

It turns out that the Jacobian of the full problem $\mathbf{J} = \mathbf{J_f} - \mathbf{I}$ is poorly conditioned and has a full eigenspectrum. Thus, a naive application of Newton's method to (3.15)—using finite differences to compute a full Jacobian—fails due to numerical truncation error in approximating the Jacobian of the cavity. Furthermore, even when a standard application of Newton's method succeeds, the computation of the full Jacobian involves the evaluation of hundreds of round trips, negating much of the computational savings.

In Fig. 3-3 we show the Jacobian of our test problem at a stationary point. The Jacobian has the following salient properties, which will be relevant to a solution of the associated system:

1. Complex. This will affect our choice of solver.

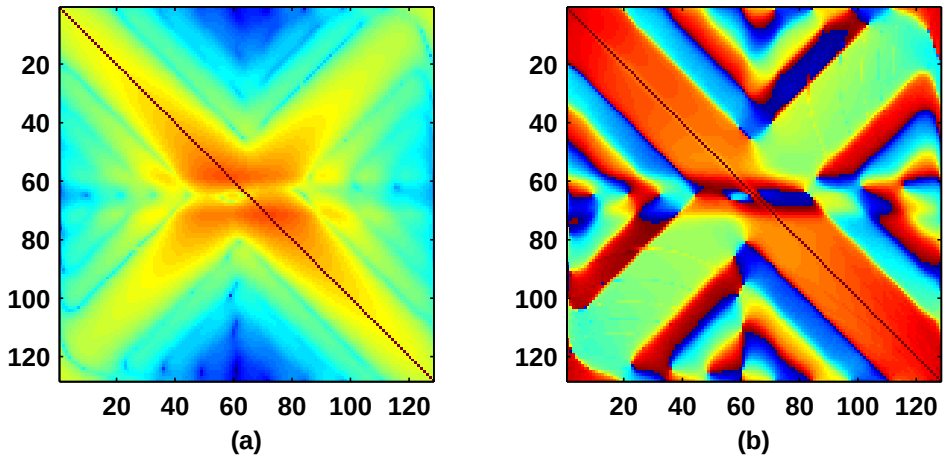

Figure 3-3: Jacobian of model shooting problem (3.29) near stationary point: (a) log magnitude, (b) phase.

2. Nearly diagonal. (Note that the plot in Fig. 3-3 is on a logarithmic scale). This suggests we can effectively precondition the system using a diagonal matrix.
3. Non-Hermitian. We will have to use a generalized solver.

3.6 Diagonal Preconditioner

To derive a diagonal left preconditioner, we begin by approximating the cavity as composed off all Fourier diagonal elements followed by all time diagonal elements, tantamount to approximating the cavity with a single nonsymmetric split-step propagation,

$$\mathbf{f}(\mathbf{u}) \approx \mathbf{N}(\mathbf{u})\mathbf{D}(\mathbf{u})\mathbf{u}. \tag{3.19}$$

As discussed in Chapter 2, everything in the cavity is diagonal in the Fourier domain except for the saturable absorber and SPM, which form the full matrix $\mathbf{N}$. However, we can actually compute the diagonal of $\mathbf{N}$ rather efficiently, and use this to approximate the contribution of SPM and saturable absorption to the diagonal of the Fourier domain Jacobian. The matrix representation of the time diagonal components can be written as

$$\mathbf{N} = \mathbf{F}^{\dagger}\lambda_{\mathbf{N}}\mathbf{F}, \tag{3.20}$$

where $\mathbf{F}$ is the discrete Fourier transform matrix and $\lambda_{\mathbf{N}}$ is the diagonal matrix of the nonlinear operator in the time domain. A diagonal matrix in the time domain is banded in the Fourier domain, and therefore the diagonal of (3.20) is constant. Since the trace

of a matrix is equal to its spectral trace, the diagonal must therefore be

$$N_{ii} = \frac{\mathrm{Tr}\{\mathbf{N}\}}{n} \tag{3.21}$$

$$= \frac{\mathrm{Tr}\{\lambda_{\mathbf{N}}\}}{n} \tag{3.22}$$

$$\equiv \sum_k \exp(n_k)/n \tag{3.23}$$

$$\approx \exp[\sum_k n_k/n], \tag{3.24}$$

where n is the dimension of our system, and n_k are the diagonal elements of the nonlinear operator, expressed in "small signal" form for simplicity and to match how the computations are done in practice. The whole cavity function can thus be approximated by a diagonal operator

$$\mathbf{f}_i(\mathbf{u}) \approx N_{ii} D_{ii} \tag{3.25}$$

$$\approx \exp\left[\sum_k n_k(\mathbf{u}) + d_i(\mathbf{u})\right] u_i, \tag{3.26}$$

with d_i the small signal elements of the Fourier domain operators (i.e. dispersion and gain). If all of the operators are approximately linearly, in the sense that $d'(u) << 1$ and therefore

$$\frac{d}{du} e^{d(u)} u \approx e^{d(u)},$$

then the cavity Jacobian $\mathbf{J_f}$ can be approximated by the diagonal matrix

$$\mathbf{B} = \mathrm{diag}\left\{\exp\left[\sum_k n_k(\mathbf{u}) + \mathbf{d}(\mathbf{u})\right]\right\}. \tag{3.27}$$

In the specific case of our model (Section 3.2) the preconditioner elements will be

$$B_{ii} = \exp\left[\sum_k \frac{q}{1+|\tilde{u}_k|^2/I_{\mathrm{sat}}} + i\delta\Delta z/\ell ||\mathbf{u}||_2^2 \right.$$
$$\left. - i(D_2\omega_i^2 + D_3\omega_i^3) + \frac{g_0}{1+||\mathbf{u}||_2^2/T_R P_{\mathrm{sat}}} \frac{\Delta\omega}{\omega_i + \Delta\omega} \Delta z/\ell \right]. \tag{3.28}$$

Ideally, we'd use a preconditioner than includes more than the diagonal, but this is sufficient for most regimes (with nonlinear phase shifts much less than π) as the Jacobian is itself highly diagonal. With the preconditioner, the linear Newton subproblem (3.17) becomes

$$[\mathbf{B}(\mathbf{u}_k) - \mathbf{I}]^{-1} [\mathbf{J_f}(\mathbf{u}_k) - \mathbf{I}] (\mathbf{u}_{k+1} - \mathbf{u}_k) = -[\mathbf{B}(\mathbf{u}_k) - \mathbf{I}]^{-1} [\mathbf{f}(\mathbf{u}_k) - \mathbf{u}_k]. \tag{3.29}$$

The inversion of the preconditioner $\mathbf{B} - \mathbf{I}$ is trivial for a diagonal $\mathbf{B}$.

Our simple preconditioner (3.27) results in significant reduction in the spectrum of

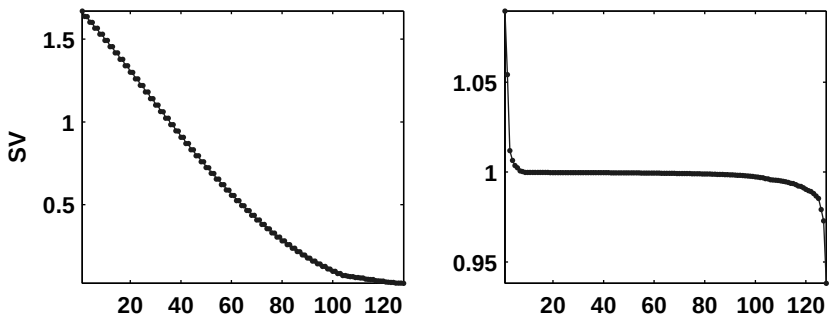

Figure 3-4: SVD spectrum for Newton problem before (left) and after (right) preconditioning. Note the different scales.

the linear subproblem. Fig. 3-4 shows the singular value spectrum for the Newton problem for our soliton laser at its stationary point, both before and after preconditioning. The conditioning improves by roughly a factor of 100 with most of the spectrum collapsing to unity, demonstrating that our simple diagonal preconditioner nearly inverts the system.

3.7 Krylov subspace solver

The fact that we do not have direct access to the analytic Jacobian of our problem—as well as the fact that our problem is amenable to preconditioning—suggests we solve (3.29) using a matrix-implicit Krylov subspace iterative solver. Given a problem $\mathbf{Ax} = \mathbf{b}$, such methods search for a solution within the space spanned by the set of vectors produced by repeated multiplication of the system matrix $\mathbf{A}$ with the residual. Assuming a null initial iterate, this gives

$$\mathscr{K}_m(\mathbf{A}) = \text{span}\left\{\mathbf{b}, \mathbf{Ab}, \mathbf{A}^2\mathbf{b}, \ldots, \mathbf{A}^{m-1}\mathbf{b}\right\}. \tag{3.30}$$

Put another way, the Krylov subspace is the space of all vectors that can be written as $\mathbf{x} = p\{\mathbf{A}\}\mathbf{b}$, where p is a polynomial of degree less than or equal to $m-1$. Essentially, we seek to invert a matrix with a polynomial of the same matrix. The power of the Krylov subspace comes from three facets: (a) it turns out to be a very efficient space in which to search given clustered eigenvalues; (b) such a space arises naturally when iteratively solving a matrix by projection methods; and (c) the solution can be found by computing the operation of a series of vectors through the system, obviating the need to know the full Jacobian.

Among Krylov methods, the most promising for our problem are of the class called *optimal* Krylov solvers, which find solutions $\mathbf{x}$ given by

$$\arg\min_{x \in \mathscr{K}_m} ||\mathbf{b} - \mathbf{Ax}||_2. \tag{3.31}$$

Given the properties of the Jacobian enumerated in Section 3.5, our two main options

are GMRES [39] or GCR [40]. Both methods involve finding orthogonal basis sets over which to solve (3.31). Given the generality of the system, both involve storing up to m vectors. The advantage of GMRES is that it often exhibits superior stability, and requires less space [41]. However, space is not a concern for the size of our problem, and our method will only be worthwhile if we are able to precondition well enough such that only tens of iterations will be needed to solve each linear system. We thus chose GCR for its ability to efficiently compute the residual and updated solution at each iteration. This allows us to minimize the number of round-trip function calls needed.

GCR is a variational method that involves incrementally generating an $\mathbf{A}^\dagger\mathbf{A}$-orthogonal basis using a Gram-Schmidt orthogonalization. The GCR algorithm, slightly modified for complex systems, follows.

Algorithm 1 *Complex Generalized Conjugate Residual (GCR)*

1. *Compute* $\mathbf{r}_0 = \mathbf{b}$. *Set the initial search vector* $\mathbf{p}_0 = \mathbf{r}_0$ *and* $x_0 = 0$.
2. *For* $j = 0, 1, \ldots$ *until* $||r_j||_2 < \epsilon$, *Do*
3. $\alpha_j = \dfrac{\Re(\mathbf{r}_j^\dagger \mathbf{A}\mathbf{p}_j)}{||\mathbf{A}\mathbf{p}_j||_2^2}$
4. $\mathbf{x}_{j+1} = \mathbf{x}_j + \alpha_j \mathbf{p}_j$
5. $\mathbf{r}_{j+1} = \mathbf{r}_j - \alpha_j \mathbf{A}\mathbf{p}_j$
6. *Compute* $\beta_{ij} = \dfrac{\Re[\left(\mathbf{A}\mathbf{r}_{j+1}\right)^\dagger \mathbf{A}\mathbf{p}_i]}{||\mathbf{A}\mathbf{p}_i||_2^2}$ *for* $i = 0, 1, \ldots, j$
7. $\mathbf{p}_{j+1} = \mathbf{r}_{j+1} - \sum_{i=0}^{j} \beta_{ij}\mathbf{p}_i$
8. *End Do*

When implementing GCR for nearly symmetric systems, we can cheat on the orthogonalization somewhat (line 6) by only orthogonalizing the current search direction relative to the last s directions. This method, known as ORTHOMIN(s) [42], turns out to work well for our problem for $s = 15$, and improves both speed and convergence by limiting round-off errors.

When solving for the Newton step using ORTHOMIN(s), we are essentially sending a series of trial perturbations through our cavity, and using the information gleaned from the perturbed output to make an optimal guess as to the best direction in which to move towards a stationary point. Since we only ever need to know the action of the system (3.29) on a single trial vector, we can avoid having to compute the full cavity Jacobian by approximating all the matrix-vector products in the ORTHOMIN(s) program by the forward difference

$$(\mathbf{B}-\mathbf{I})^{-1}(\mathbf{J}_\mathbf{f}-\mathbf{I})\mathbf{p} \approx (\mathbf{B}-\mathbf{I})^{-1}\left(\frac{\mathbf{f}(\mathbf{u}+d\mathbf{p})-\mathbf{f}(\mathbf{u})}{d}-\mathbf{p}\right), \tag{3.32}$$

where **u** is the point around which the finite difference is computed (the current solution guess), **p** is a trial vector, and $d \equiv a/||\mathbf{p}||_2$ with a the difference scale, chosen to balance truncation error with rounding error in the finite difference. The entire algorithm is summarized in a flow chart, Figure 3-5.

3.8 Theoretical Convergence

If we are close enough to a solution, the outer loop should converge quadratically. This is not guaranteed, of course, even if the function is well behaved, as we are using an approximate solver solver. However, the number of outer steps will be small in any event, assuming convergence occurs. The greater concern is how many cavity evaluations will be required to solve each Newton step; that is, how many GCR steps will be required in each inner loop.

The convergence of optimal Krylov methods is a function of the distribution of the eigenvalues. The norm of the residual $\mathbf{r}_n = \mathbf{A}\mathbf{x}_n - \mathbf{b}$ after n iterations is given by [43]

$$||r_n|| \leq \inf_{p \in P_n} ||p_n(\mathbf{A})|| \leq \kappa_2(\mathbf{V}) \inf_{p \in P_n} \max_{\lambda \in \sigma(\mathbf{A})} |p(\lambda)|, \tag{3.33}$$

where P_n is the set of polynomials of degree n with the normalization restriction that $p(0) = 1$, **V** is the eigenvector matrix of **A**, $\sigma(\mathbf{A})$ is the spectrum of **A**, and κ_2 is the condition number. Given that the Krylov subspace for the problem $\mathbf{A}\mathbf{x} = \mathbf{b}$ is the space of all vectors expressible as a polynomial of **A** operating on **b**, it makes sense that the norm of the nth residual is the norm of the best matrix polynomial one can construct out of the Krylov space (the first inequality). The second inequality in (3.33) is not obvious, however. It means that the convergence will be better with lower condition numbers, and with eigenvalues clustered away from zero [41]. In the specific case of **A** having only m distinct eigenvalues, (3.33) states that the algorithm will converge in at most m iterations.

Given how close our model system is to symmetric, the distribution of eigenvalues can be approximated by the singular values. As we've already shown in Figure 3-4, the preconditioning greatly improves both the conditioning and spectral distribution of the system. Based on the fact that roughly 90% of the singular values are clustered very close to one, we'd expect that the subproblems should converge roughly an order of magnitude faster than would be required for a full matrix. For our model problem, (3.33) implies that each inner loop should only require on the order of ten round trip evaluations. Assuming the Newton steps converge in less than ten iterations, we should thus expect to solve the cavity using on the order of 100 round trip evaluations.

Whether or not this is much of an improvement over naive simulation depends, of course, on the convergence of the natural system. As a rule of thumb, the more dispersion in the system, the slower it naturally converges. On the other hand, the more nonlinear the cavity, the faster it will converge. The utility of this algorithm them, hinges on the granularity of the cavity relative to the pulse shaping mechanisms. On one extreme, weakly nonlinear solid-state lasers can take 50,000 round trips to converge, whereas our algorithm will only require around 30 round trips. At the other end of the spectrum, fiber

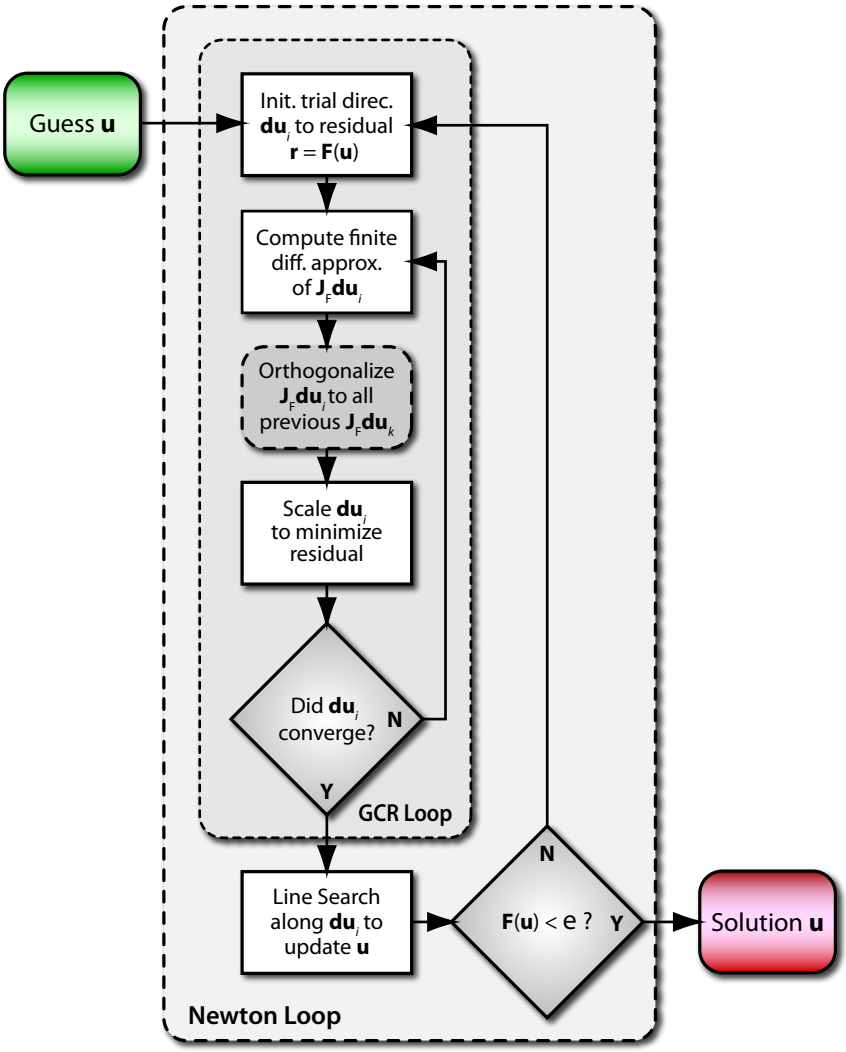

Figure 3-5: Flow chart for entire Newton-Krylov solver. Dashed borders indicate iterative loops. In the figure, ϵ is the final termination tolerance, and F is the function which computes one cavity round trip. (Figure courtesy of Robert R. Birge.)

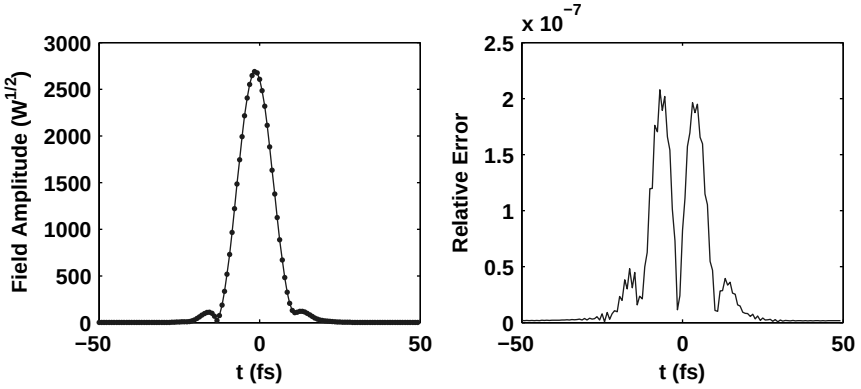

Figure 3-6: Left: comparison of the amplitude of the final solutions obtained by our method (dots) and standard dynamical evolution (solid line) for our test mode-locked laser model. Right: absolute value of difference between the two solutions.

lasers are capable of significant pulse shaping over one pass, and they tend to converge within hundreds of round trips, offering little opportunity for improvement given the expected slow convergence of our inner loop problems in the presence of high nonlinearity.

3.9 Accuracy

Given that the convergence test of the algorithm is the same convergence test for the standard method (i.e. that the cavity reproduces the solution to within the some level of precision) and given that our method uses the same cavity model, the accuracy of our method is not really an issue and is limited by the cavity model. However, it is technically possible that our method could converge to a quasi-stable cavity mode that would not be found by a dynamic simulation. In Fig. 3-6 we compare the final solution found by our method with that found by dynamic simulation. As expected, they are roughly similar to within the convergence criteria of 10^{-8}.

3.10 Empirical convergence

As with all Newton solvers, the convergence of the outer loop is dependent on the starting point. As such, our solver is best used to refine an initial rough guess to high precision. However, the guess can be off by a significant amount (as illustrated in Fig. 3-8) and in some cases can converge starting from noise. Nonetheless, the better the guess the more reliable the convergence. However, this is not inconsistent with the two main intended uses of this algorithm: (a) to compute the results of many different cavity parameters, where the output of one simulation may be used as a seed to the next; and (b) in an optimization loop, where the parameters evaluated will be highly correlated with those from preceding computations. If a solution starting from zero is desired, the best approach is to run the dynamic simulation for a hundred round trips or so, and then let

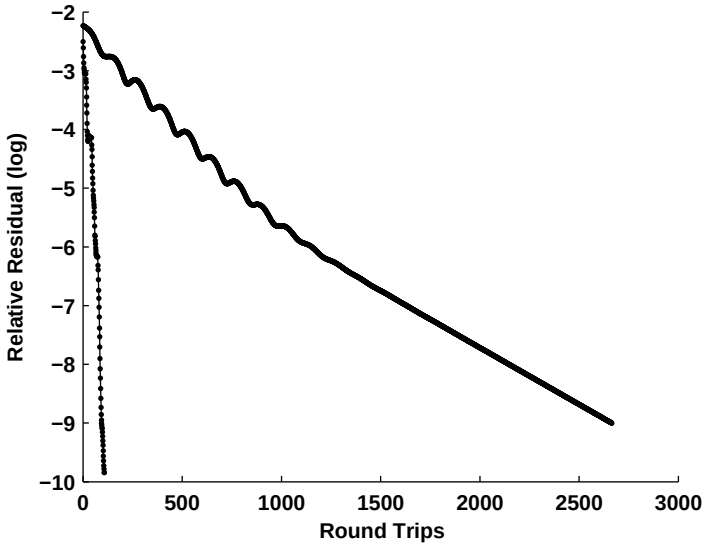

Figure 3-7: Comparison of convergence between our method (dotted line) and standard dynamical evolution (solid) for our test mode-locked laser model. All cavity evaluations are included, including those used to solve the linear Newton subproblems.

the algorithm take over once the pulse has evolved sufficiently.

This was the approach we took to test the algorithm's convergence relative to dynamic simulation. In Fig. 3-7, we compare the convergence of our algorithm to that obtained with standard simulation. Both the Newton-Krylov solver and the cavity dynamic simulation were seeded with a very rough starting pulse obtained from iterating the cavity 200 times on a starting impulse. Continuing with the dynamic simulation required over 2500 round trips to converge to within 10^{-9}. Our method was more than 34 times faster, and was able to converge to 10^{-10} while requiring only 76 cavity round-trip evaluations. The quadratic convergence of the outer Newton process is apparent in Fig. 3-7.

To provide a visualization of the kind of paths taken by the Newton-Krylov solver, we plot the exact sequence of trial pulses sent through the cavity solver in terms of the log of their residuals, Fig. 3-9. Note that initially the rough pulse shape is found and further refinement simply involves small perturbations and scaling. It thus makes sense that searching along paths in the direction of the residual would work well. The observed fact that the solution tends to overshoot during early steps suggests that convergence could be improved markedly were care taken in determining the Newton step length.

3.11 Future Work

The method appears to work quite well up to roughly π radians of nonlinear phase per round trip. This is good enough to deal with most few-cycle lasers, but the algorithm becomes less effective for lasers with extreme amounts of nonlinearity per round-trip, converging at roughly the same rate as dynamic simulation when the cavity nonlinear

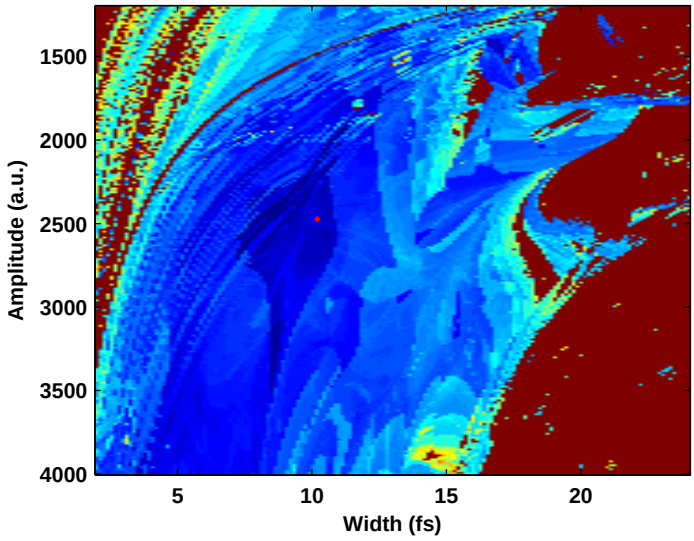

Figure 3-8: Convergence map of Newton-Krylov solver for Gaussian starting iterates of various widths and amplitudes for a ten femtosecond laser model (chosen for its quick convergence). The darker colors represent fewer steps, with the outer red region starting points that did not ever converge. The vertical axis is the amplitude and the horizontal the width of the starting Gaussian. The actual solution to the model is best approximated by the dot at roughly $(2500, 10)$. This map suggests that the method will converge for a wide range of starting guesses, but that starting with energetic short pulses is a good strategy in the absence of any information about the true solution.

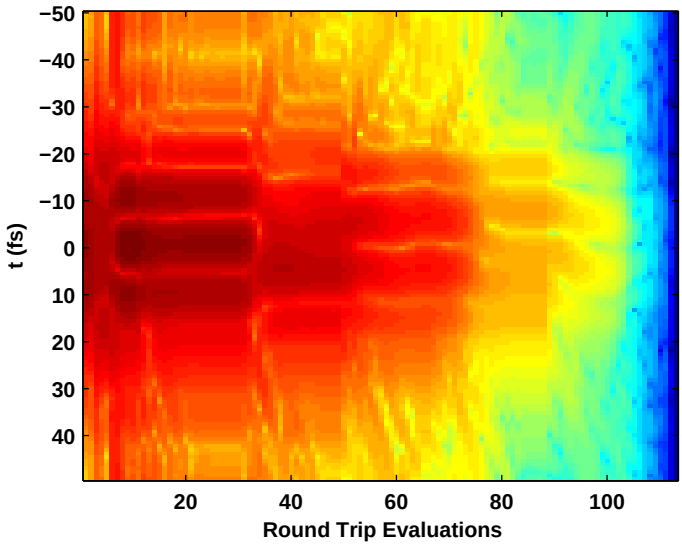

Figure 3-9: The sequence of residuals produced by the direct solver as a function of the number of round trip evaluations, log scaled.

phase approaches one cycle. There are several avenues of improvement that could be pursued to enhance the applicability and convergence of the algorithm, however. In addition, there is potential for this algorithm to be applied to more general cavity models incorporating spatial effects.

Better preconditioning for high SPM. The reason for the failure at high levels of nonlinearity is most likely the fact that we are using a diagonal preconditioner in the Fourier domain, which ignores all but a constant factor contribution from the nonlinearity. Given that the computational cost of the algorithm is almost completely dominated by the costly propagation through the gain medium, especially given the small vectors needed to represent our solutions, it might be worthwhile to do a more involved preconditioning. We could easily compute several stripes of the banded matrix representing the lumped nonlinearity, and its inversion would likely be negligible compared to the cavity round trip evaluations.

Reduced basis sets. There is no reason why we must compute a round trip in the same basis that we represent our search. However, we have not yet taken advantage of the potentially significant performance improvements via this route. In general, the representation that is optimal for accurately computing a round trip will be of significantly higher dimension than that needed to sufficiently represent the solution. Not only do we waste effort operating on wider temporal windows than we need (which end up filled with zeros) but we also hurt our convergence with the resulting poorly conditioned system. This could provide significant improvement in convergence, and is the most promising next direction to take.

Alternative linear solvers. So far we've only tried GCR and ORTHOMIN. Given that we experienced improvement by moving to ORTHOMIN it's likely we are, in fact, running into numerical stability problems. Despite our hope otherwise given the small number of iterations, the numerical stability issues suggest we may have chosen the wrong solver. It may be productive to investigate other Krylov solvers, such as ORTHODIR and GMRES.

Better line search implementation. It's possible that our region of convergence could be greatly improved by improving the Newton steps. The result shown in Fig. 3-9 and discussed in Section 3.10 suggest that a line search could significantly improve the rate of convergence.

Phase normalization handling. Our solution to the problem of having no *a priori* knowledge of the cavity phase or group velocity (see the discussion in Section 3.4) was the addition of an ad hoc normalization function (3.15). Undoubtedly, this affects the convergence of our Krylov method to some extent as it significantly breaks the symmetry of the system. There may be better ways to handle this novel facet of our problem, such as solving for the ideal normalizing phases in a separate step, and then taken them as fixed during the solution of the Newton subproblems.

Parallelization. The approach we've taken here has a potentially significant unrealized benefit over dynamic simulation: the ability to be parallelized. Dynamic simulation is inherently a serial process, whereas the solution to our linear subproblems can take advantage of parallel linear numerical techniques. At the extreme where we have at least as many processors as dimensions in our field, we can solve each linear problem in the time it takes to compute two round trips. This would allow another order of magnitude

advantage in computational time over standard simulation.

Extension to full spatio-temporal model. An unfortunate truth is that while we can phenomenologically model Kerr lens mode-locked lasers sufficiently well to capture their salient features, we cannot simulate them well enough to use simulation as a primary design tool. Designers of these lasers thus currently have to include significant margins in the laser as built, and getting the laser to mode lock is largely a matter of trial-and-error, adjusting cavity parameters such as dispersion and gain. A significant aspect of this is our lack of quantitative understanding of Kerr lensing, driven by our lack of ability to effectively model it in a cavity.

The method presented in this thesis could be applied in a straightforward way to the simulation of a spatio-temporal cavity model. Preconditioning would work similarly, with the Jacobian for the linear cavity elements being diagonal and computable analytically if we operate in a basis composed of temporal and spatial frequency modes. Were an extension of this method to work with similar effectiveness on a spatial model, it might allow us to refine a quantitative model of Kerr lens mode locking by matching the output of the simulation with spatial measurements from actual lasers. Such an "optimization" would require a fast cavity solver.

The ability to fully model a nonlinear cavity would provide a significant benefit to the field of ultrafast optics, as it would allow for the precise engineering of mode-locked lasers without need for the trial-and-error tweaking involved today in the development of a laser. An even more auspicious goal would be the development of a model sufficiently accurate to act as a testbed for new laser development and research. Laser development is currently at the point where models do not adequately predict the operation of the shortest pulsed lasers. Providing a computational method capable of accurately modeling the full physics could revolutionize the development of mode-locked lasers, which currently require the construction of prototypes costing hundreds of thousands of dollars to build.

Chapter 4

Phase Distortion in Ultrafast Optics

Historically, the performance of phase sensitive optical devices has been measured in terms of group delay dispersion (GDD), and designs optimized by minimizing the integrated deviation of GDD from the ideal desired response. We begin this chapter by arguing that GDD is not always a meaningful criteria to consider for ultrahigh bandwidth devices. We then derive an alternative spectral quantity with which to optimize a phase-sensitive filter, which we refer to as the phase distortion ratio (PDR). This approach expresses both reflectivity and phase errors in units of power, putting them on an equal basis. This provides a rational way to include knowledge of the expected power spectral density of the source, an important factor to consider that has heretofore been largely neglected in dispersive filter optimization. We show that this approach yields greatly improved results in applications such as pulse compression, cavity enhancement, and mode-locked lasers.

4.1 Introduction

Optical thin-film mirrors are critical components of many modern ultrafast optical systems and communications devices. In some situations a specific group delay dispersion (GDD) profile is desired, such as in dispersion compensating mirrors [16, 44], and in others the dispersion is made as small as possible to minimize its effect on short pulses. In either event, phase sensitive filters have historically been designed by numerically minimizing a weighted integral of reflectance and GDD error.

For problems where a given dispersion curve can be matched very closely, it generally does not matter what one chooses for an optimization criterion (i.e. merit function) so long as the merit function has a local minimum at the ideal response. Certainly, integrated GDD satisfies this condition. However, for difficult filter engineering problems which push the limits of what is possible in terms of bandwidth and/or precision, performance tradeoffs are necessary, and the only solutions available are those which roughly approximate the desired outcome. In such cases, the optimization process must balance among a large number of factors (i.e. the filter performance at a series of frequencies) to arrive at the best compromise available. The choice of merit function plays a critical role in determining the performance of the final design. If not, the optimization may

converge to non-optimal points. Unfortunately, it is generally not computationally feasible to use our ultimate goals as merit functions. For example, in the case of dispersion compensating mirrors for a mode-locked laser, we may wish to find mirrors which minimize the output pulse width of the laser. The stable pulse of a laser cannot be computed quickly enough to be used in an optimization loop, so we must use some other criteria as a proxy. The main task involved in optimization is the selection of computationally efficient models for our problem that is as close as possible to a linear function of our underlying objectives.

4.2 Group delay dispersion background

The concept of GDD arose in the context of the slowly varying envelope approximation, where the relative bandwidth is, by definition, assumed to be small enough such that dispersive broadening is well described by the local second-order phase. In these cases, minimizing the GDD at a single point is a perfectly valid way to minimize dispersive broadening. As bandwidths increased, such that a series expansion of phase at a single point was no longer valid, the concept of GDD was extended by considering GDD as a function of wavelength. The standard way to design wideband phase compensating filters has historically been to minimize the mean squared deviation of GDD from the desired profile [45].

It is not always clear, however, how to ascribe physical meaning to the frequency domain integral of squared GDD over a wide bandwidth where a series approximation is not valid. Optimizing integrated GDD error would certainly minimize the mean distortion of a set of independent sources each of narrow bandwidth (e.g. WDM channels). However, in the case where a wide bandwidth of frequencies will be coherently interfered (as with a laser pulse), RMS GDD error may not be an ideal measure of performance.

Consider, for example, a dispersion compensating mirror used to compress an ultrashort pulse. The dominant nonideality in such mirrors is the creation of satellite pulses caused by an impedance mismatch at the mirror surface. It can be shown [46] that a spurious reflection of magnitude a at a relative delay τ causes a GDD error proportional to a^2 (to lowest order in a). In many applications, however, it is the *energy* contained in any satellite pulses that is of relevance, not their delay. The quadratic scaling of GDD with delay, however, means that integrated GDD error is not a monotonic function of satellite pulse energy, and thus GDD minimization will generally converge to a suboptimal solution. This is illustrated dramatically in Section 4.5.

One avenue around these issues is to optimize the mirror in the time domain [47, 48]. However, thin film filters are most naturally analyzed in the frequency domain (where they are diagonalized) and thus time domain optimization presents its own difficulties. Requiring one FFT per wavelength at each optimization step (to compute the gradient) involves significant computation, and operating in the time domain renders analytic gradients of merit functions infeasible. Thus, a single optimization step will require an effort that scales as $O(n_\lambda \log n_\lambda n^2)$, with wavelength count n_λ and layer count n. This is significantly greater than the $O(n_\lambda n)$ complexity experienced using analytic gradients

in the frequency domain [49], and that is notwithstanding the potentially slower convergence rate from using finite difference gradients.

In the following, we propose an alternative method to optimize a reflective filter in the frequency domain in terms of a simple quantity we refer to as the phase distortion ratio (PDR). This approach expresses both reflectivity and phase errors in units of power, putting them on an equal basis, which facilitates weighting between them and eliminates the ambiguity inherent in adding GDD errors and reflectivity errors (which have incompatible units). Furthermore, this approach provides a rational way to consider the expected power spectral density of the source, an important factor to consider that has heretofore been neglected in dispersive filter optimization. We believe this approach will yield improved results for many applications in communications and ultrafast optics. As a bonus, the method avoids the costly evaluation of GDD and GDD gradients during optimization.

4.3 Phase Distortion Ratio

In lieu of GDD, we propose a simple spectral quantity based on the fractional energy lost to phase distortions. We will show in subsequent sections that this criterion turns out to produce optimal designs in certain linear cases of pulse compressors and enhancement cavities. While results cannot be proved for the case of a general nonlinear cavity, we show through simulation that phase distortions are more predictive of cavity performance than GDD errors.

4.3.1 Derivation

While we assert that integrated GDD error is not a valid basis for optimization, spectral GDD is an ideal way to *specify* a desired dispersion curve. We will thus express the design goal in terms of the ideal GDD curve, written as $\hat{D}_2(\omega)$. The GDD of our device being optimized will be written at $D_2(\omega)$, with the net phase error defined as

$$\Delta\phi = \iint_{\omega_1}^{\omega 2} d\omega^2 \, D_2(\omega) - \hat{D}_2(\omega), \tag{4.1}$$

where ω_1 and ω_2 are the bounds of the optimization. Likewise, the optimal reflectivity will be $\hat{R}(\omega)$ and that of the design denoted as $R(\omega)$.

Without making any claims as to its validity for any purpose, we begin by considering the total power contained in the difference between the ideal mirror response and that of a given design. The phase error in (4.1) will, in general, include constant and linear phase terms which we wish to ignore.[1] These phases will be dealt with in the following by the inclusion of floating terms $\phi_0 + \phi_1\omega$, with ϕ_0 and ϕ_1 chosen to minimize the error power. By Parseval's theorem, then, the total power contained in the difference between

[1] It is possible that optical comb enhancement cavity mirrors might eventually be concerned with an absolute phase term, but for the present discussion we will focus on applications where the only goal is an ideal spectral GDD.

the pulses we want and those actually reflected is

$$\begin{aligned}\text{Error Power} &= \min_{\phi_0,\phi_1} \int d\omega\, P(\omega) \left| e^{im\hat{\phi}(\omega)} - r^m(\omega) e^{im[\phi(\omega)+\phi_0+\phi_1\omega]} \right|^2 \\ &= \int d\omega\, P(\omega) \left[1 + R^m(\omega)\right] + 2 \min_{\phi_0,\phi_1} \int d\omega\, P(\omega) R^m(\omega) \cos[m(\Delta\phi(\omega) + \phi_0 + \phi_1\omega)], \quad (4.2)\end{aligned}$$

where $r(\omega)$ is the absolute value of the field reflection coefficient, $P(\omega)$ is the expected power spectral density of the source, and m is the number of reflections we are considering.

The cosine term is problematic. First, its periodicity would impede convergence in a local gradient optimization. Second, it makes it impossible to solve directly for the optimal values of the floating phase coefficients. Because of the floating terms, however, the argument to the cosine at the end of an optimization will generally be much smaller than π (assuming a reasonably good solution is found). As such, we can very effectively replace the cosine with an expression that is second-order in phase. In addition, for any reasonable solution the $R(\omega)$ factor can be well approximated by unity and thus dropped from the second term,

$$\text{EP} = \int d\omega\, P(\omega) \left[1 - r^m(\omega)\right]^2 + \min_{\phi_0,\phi_1} \int d\omega\, P(\omega) \underbrace{m^2 \left(\Delta\phi(\omega) + \phi_0 + \phi_1\omega\right)^2}_{\text{Phase Distortion Ratio}}. \quad (4.3)$$

The first term above is the power lost to mirror leakage. The second term is the power that is directed into phase distortions. Thus, the under-bracketed term is a spectral quantity that represents the fraction of power lost to phase distortion at each wavelength, which we call the spectral *phase distortion ratio* (PDR). This quantity turns out to be very useful in the design and optimization of phase compensating filters, as we will show in following sections.

To eliminate the variables of the minimization in (4.3), we take the gradient of the second term in (4.3) with respect to the coefficients ϕ_0 and ϕ_1. Setting both gradient elements to zero yields a set of coupled equations for the optimal constant and linear phases,

$$\begin{aligned}\phi_0 + \langle\omega\rangle \phi_1 &= -\left\langle \Delta\phi(\omega) \right\rangle \\ \langle\omega\rangle \phi_0 + \left\langle \omega^2 \right\rangle \phi_1 &= -\left\langle \omega\, \Delta\phi(\omega) \right\rangle,\end{aligned} \quad (4.4)$$

with $\langle\cdot\rangle$ denoting the power weighted mean defined by $\left\langle f \right\rangle = \int d\omega\, P(\omega) f(\omega) / \int d\omega\, P(\omega)$. Solving for the optimal floating phase terms and substituting back into (4.3) gives us a quantity that is entirely a function of mirror parameters, and thus can be used directly in an optimization. The final expression for the PDR is

$$\boxed{\text{PDR}(\omega) \equiv m^2 \left[\Delta\phi(\omega) - \frac{\Delta\omega^2 + \omega_0^2 - \omega\,\omega_0}{\Delta\omega^2} \left\langle \Delta\phi \right\rangle + \frac{\omega - \omega_0}{\Delta\omega^2} \left\langle \omega\, \Delta\phi \right\rangle \right]^2}, \quad (4.5)$$

where $\Delta\omega = \sqrt{\langle\omega^2\rangle - \langle\omega\rangle^2}$ is the source bandwidth and $\omega_0 = \langle\omega\rangle$ is the source center frequency. Note that the coefficients in front of the $\Delta\phi$ averages are all constants that only need to be computed once during an optimization.

4.3.2 Practical issues of optimizing with PDR

In practice, directly computing the phase of the filter can be complicated by phase unwrapping issues, as it is often the case that filters have large group delay offsets.[2] It is best to compute the phase error by a numerical integration of group delay. As numerical quadrature and the gradient are both linear operators, analytic gradients of the above merit functions can be computed in a straightforward manner from gradients of group delay and reflectivity, as in [49]. While the final analytic expression is cumbersome to detail in closed form, it is straightforward algorithmically and does not pose a significant computational burden, as numerical integration is linear in complexity. An example merit function computation is shown in Appendix C.

4.4 Enhancement cavity design

Enhancement cavities for laser combs have received considerable attention recently, both in the context of maximizing nonlinear conversion through field enhancement [50], as well as for repetition rate (comb spacing) enhancement for spectroscopy [51]. In either case, a crucial aspect of such cavities is the mirror design, which must exactly compensate for intracavity dispersion (if any) such that the cavity modes remain equidistant. In enhancement cavities, the only concern is the absolute shifting of individual resonances, and any oscillation of the detuning as a function of frequency is irrelevant. Optimizing for minimum GDD, as is usually done, places a significant unwarranted penalty on designs with quickly varying phase oscillations, again leading to suboptimal designs as the phase response is compromised needlessly.

4.4.1 Cavity transmission is a function of PDR

Assuming the reflectivities of the mirrors maintain a sufficient cavity finesse for field enhancement and/or mode suppression, the dominant mechanism affecting the cavity throughput will be the detuning of cavity resonances due to mirror dispersion nonideality. We begin by considering the transmission of a Fabry-Perot resonator [25],

$$T(\omega) = \frac{[1-R(\omega)]^2}{[1-R(\omega)]^2 + 4R(\omega)\sin^2[\phi(\omega)/2]}, \tag{4.6}$$

where $R(\omega)$ is the aggregate reflectivity of the cavity mirrors, and $\phi(\omega)$ is the round trip phase of the entire cavity, including free space propagation, intracavity materials, and

[2] By phase unwrapping, we refer to the process of removing spurious discontinuities in the reflection phase that occur when the reflection phase ranges over more than 2π.

the mirror under consideration. Phase distortions in the mirrors will affect the enhancement cavity by pulling resonances of the cavity off of comb lines. The spectral transmission of a single comb through a cavity mode can be very closely approximated (assuming reasonable transmission to begin with) by taking a second order series expansion of (4.6) with respect to local mirror phase deviation $\Delta\phi(\omega) = \pi m - \phi(\omega)$, with $m \in Z$ an index representing the laser comb line nearest ω. This gives an expression for the local change in transmission,

$$\delta T\Big|_{\omega} = \frac{R}{[1-R]^2}\Delta\phi^2 + O[\Delta\phi^4]. \tag{4.7}$$

Any global linear phase term in the mirror can be taken up by changing the length of the cavity. A constant phase can be accounted for by changing the f_{CEO} of the laser being locked to the cavity.[3] Therefore, (4.7) implies that the global transmission attenuation of an enhancement cavity mirror system is proportional to the PDR, so the overall transmission of a laser comb nominally repetition-rate-matched to the cavity is

$$T(\omega) = 1 - \frac{R(\omega)}{[1-R(\omega)]^2}\mathrm{PDR}(\omega) + O[\mathrm{PDR}^2(\omega)]. \tag{4.8}$$

This is valid at any frequency, assuming that the phase oscillations are much wider than the mode spacing, and that the cavity resonances shift by much less than their width. While the latter criterion will, by neccesity, hold for any useful cavity mirror design and thus (4.8) is appropriate as an optimization merit function. If the cavity modes differ from the laser comb spacing by an integer multiple, such as with a rate enhancement cavity, (4.8) is to be interpreted as the fractional tranmission of those modes which fit the cavity spacing.

4.4.2 Cavity merit function

While the specific optimization approach will depend on the application of the enhancement cavity in question, the preceding shows that the relevant underlying spectral quantity to consider is the PDR. For example, in the case of a repetition rate enhancement cavity, a suitable approach would be to maximize (4.8) subject to the constraint that the mirror reflectivity was above a certain minimum required for unwanted mode supression.

If the desired goal is to simply maximize the total field intensity inside the cavity, (4.8) implies that the appropriate merit function to minimize is

$$Z = \int d\omega\, P(\omega)\frac{R(\omega)(2 - R + \mathrm{PDR}(\omega)) - 1}{[1-R]^3}. \tag{4.9}$$

In practice, for relatively narrow relative bandwidths, less than one tenth, enhancement cavities can be designed for low dispersion by using mirrors based on Bragg stacks. To progress beyond that, however, enhancement cavities with chirped mirrors will have

[3] Were this not the case, and the carrier envelope phase of the laser fixed to some desired value, it is not difficult to formulate an alternative spectral parameter that would allow for an arbitrary linear phase but would hold a constant phase, modulo 2π. This might be termed a spectral dephasing ratio.

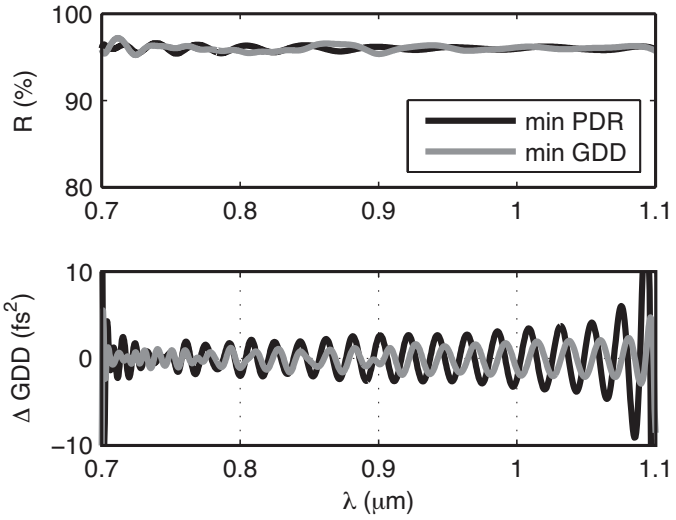

Figure 4-1: Enhancement cavity mirror GDD error comparison.

to be used, with intracavity positive dispersion elements introduced. Given the extreme sensitivity of wide bandwidth cavities to resonance shifts, proper design of the mirrors will be imperative to successful implementation, and the performance of the mirrors will be the limiting factor to bandwidth.

4.4.3 Example optimization

To demonstrate the efficacy of PDR optimization of cavity mirrors, we consider the case of an ultra-broadband rate enhancement cavity with a modest finesse factor of 77, created by a chirped mirror pair. The cavity was assumed to have roughly 1 mm of fused silica providing internal positive dispersion. Such a cavity would be useful for increasing the mode spacing of a femtosecond comb for high precision astrophysical spectroscopy, as in [51].

A double chirped mirror pair was designed using the standard GDD optimization approach. Another mirror pair was then optimized using the criterion in (4.8), starting from the same starting point used in the GDD design. These two designs are compared in Figure 4-1. It is clear that the GDD optimized design (grey) has less GDD ripple than the minimum PDR design (black). In fact, the PDR optimized design is four times worse in terms of RMS GDD.

The theoretical cavity transmission (assuming perfect cavity locking) is shown in Figure 4-2. Despite the significantly lower dispersion ripple of the GDD optimized mirror, it is virtually useless, admitting only a few nanometers of bandwidth into the cavity before the mirrors' phase errors dephase the cavity from the comb. The PDR optimized cavity, however, transmits an average of around 90% over the entire 400 nm bandwidth.

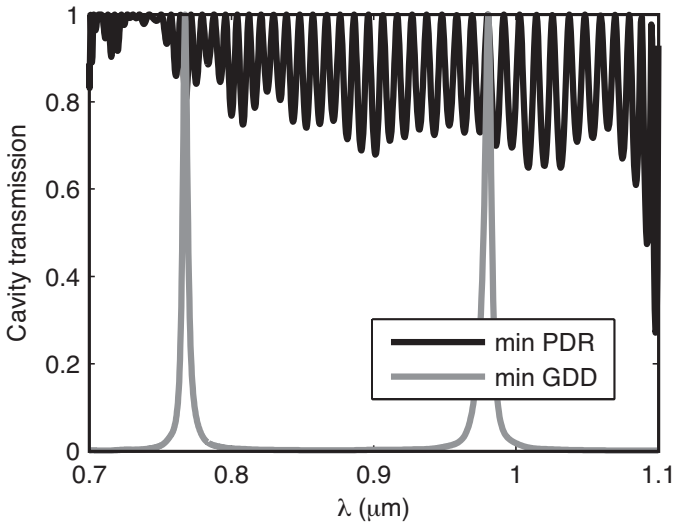

Figure 4-2: Enhancement cavity transmission comparison. This transmission curve represents the fraction of power transmitted for those comb lines which matche the cavity spacing. That is, the full cavity transmission will be attenuated from the shown curve by a factor equal to the ratio of the cavity spacing to the laser mode spacing.

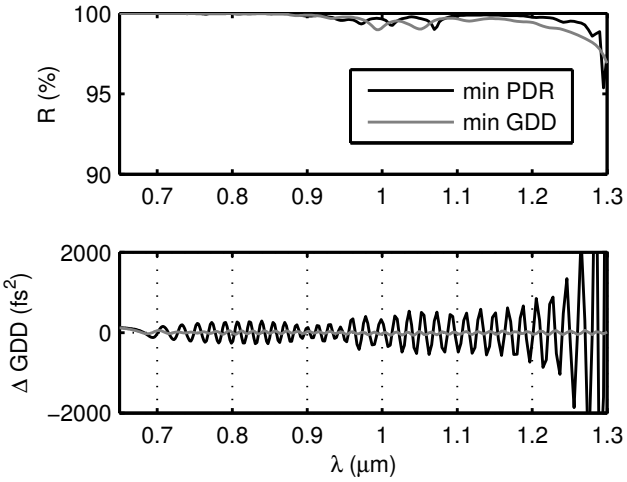

Figure 4-3: Compressor GDD comparison.

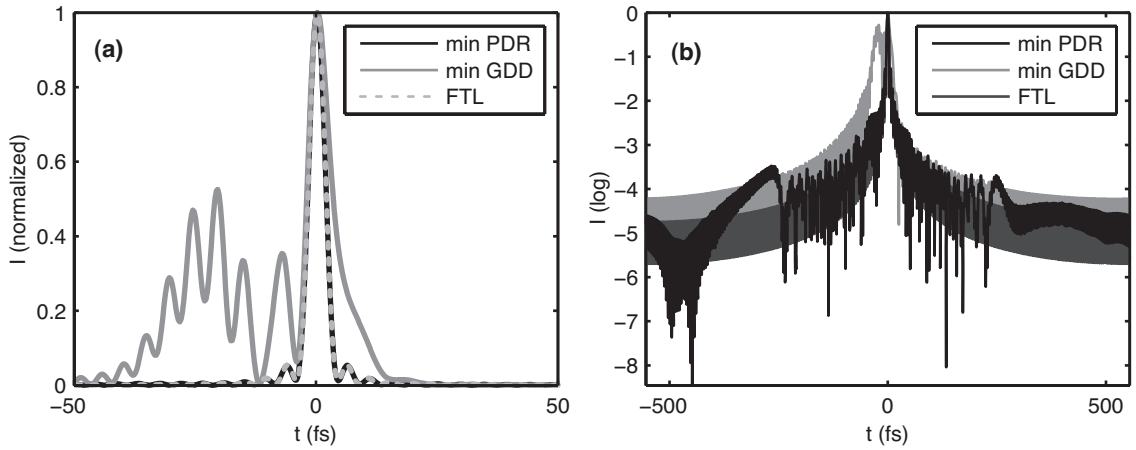

Figure 4-4: Compressor pulse comparison on a linear (a) and log scale (b).

4.5 Compressor Design

As an additional linear test case, we also consider the design of a six mirror dispersion compensation system comprised of two complementary mirror types, intended to recompress a single-cycle NIR pulse after traveling through 2 mm of fused silica. As before, we performed two designs, one which minimizes GDD and one which minimizes PDR. The two designs are compared in Figure 4-3.

The PDR optimization produced a GDD curve with 18 times more RMS GDD error than the GDD design. However, when looking at the effect on a pulse after compression, the minimum energy design performs significantly better, theoretically yielding output pulses that are indistinguishable from fourier transform limited pulses. Figure 4-4(a) shows the time domain error signal for the two filters relative to a theoretically perfect mirror.

The GDD design has virtually no fast phase ripple, but the phase itself has longer scale variations which destroy the pulse. Insight into why the GDD optimization failed can be seen from a logarithmic plot of the pulse, Figure 4-4(b), that the PDR design produces small phase distortions far from the main pulse, around 250 fs away. While the energy contained in these satellite pulses is negligible, the quadratic GDD scaling with delay discussed earlier prevents the GDD optimization from considering this a valid solution.

Finally, we note that the minimum PDR design required no human intervention to pick weighting values, necessitating less user intervention than the GDD design, which required iterative adjustment of the weighting between GDD and reflectivity to achieve sufficient reflectivity performance.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 5

Dispersion Compensation of Mode-Locked Lasers

In this chapter, we investigate the effects of intracavity phase distortions on mode-locked laser performance. Unfortunately, analytic results from the master equation do not shed a great deal of light on this problem, as phase distortions constitute a very high order effect that cannot be effectively handled by a perturbative approach. We are thus forced to turn to numerical experiments. Fortunately, even one-dimensional laser cavity models are capable of capturing the essential dynamics of dispersion-managed soliton lasers [52].

We find that even small phase mirror phase oscillations of less than 0.1% of a wave can profoundly affect the overall spectral shape of the laser. In addition, the standard dispersion managed cavity arrangement of having small net positive dispersion is particularly vulnerable to phase oscillations. This suggests significant improvement in laser performance is still potentially possible with advances in mirror technology.

Finally, we conclude this chapter by addressing the question of how to best optimize cavity mirrors, using a statistical approach—enabled by our Krylov solver—to evaluate the relative performance of PDR versus GDD optimization. The results suggest that neither method is ideal, but that PDR is significantly more predictive of cavity performance.

5.1 Effects of intracavity phase ripple

Phase ripple in intracavity dispersion compensating mirrors manifests in several ways. The most simple is as an effective cavity loss. With sufficient SPM, satellite pulses are not resonant with the main pulse and do not build up appreciably. If the satellite pulses do not mix with the main pulse, the net effect of phase distortion is simply to pull a certain amount of power out of the main pulse each round trip. This is the case for relatively narrow band pulses which do not experience significant dispersion during a round trip. In such situations, the integrated PDR is clearly a very relevent optimization criterion.

On the other hand, for dispersion managed systems with large swings in pulse chirp, it is possible for the pulses to be broadened sufficiently during transit through the gain medium for the main pulses and satellite pulses to overlap with each other and exchange

energy via third-order nonlinearities. In this case, which clearly applies for sub-two-cycle lasers, the effects of phase ripple can be quite complicated, explaining why neither GDD nor PDR is found to be perfectly correlated with the pulse spectrum in Section 5.3. While increasing the ripple magnitude past a certain point will always destroy mode-locking, simulations show that increasing phase ripple will sometimes actually *increase* the bandwidth, confirming that the only way to truly optimize dispersion compensating mirrors is to consider a full cavity simulation. As we show in the following, phase distortions are responsible for the characteristic M-shaped spectral structure of few-cycle lasers.

5.1.1 SPM and phase distortion

In the absense of significant SPM, simulations show that the presense of phase ripples of even 1/1000 of a wave can result in spectral amplitude ripples of nearly full modulation. That this should be the case is not obvious, because the reflectivity spectra of the mirrors are flat. Were only linear effects taking place, the satellite pulses would not manifest as amplitude oscillations. Absent SPM, the phase velocity of the satellite pulses is matched to the main pulse, and thus power can be resonantly fed into the satellite pulses, allowing them to grow to appreciable intensity. Furthermore, these satellite pulses experience the same pulse shaping dynamics as the main pulse, a process that compresses them and changes their phase relationship to the main pulse. In cavities with little or no SPM, satellite pulses can be as short as the main pulse and nearly as intense. The interference of the two results in large spectral amplitude oscillations.

To overcome these effects, SPM must be used. The introduction of SPM quells the build up of phase distortions by establishing an intensity dependent round trip (soliton) phase. Satellite pulses, having significantly less intensity, will not be phase matched to the main pulse and will thus not grow appreciably. In the case of negative dispersion cavities, even small amounts of SPM can nearly completely eliminate the effect of phase distortions, rendering phase distortions nothing more than another source of loss. Cavity mirrors for such lasers should therefore be designed by minimizing the power weighted sum of PDR and transmission.

5.1.2 Resonant phase distortions in positive dispersion cavities

A soliton-continuum net phase plot (such as that shown in Figure 2-3 in the context of soliton perturbation theory of Section 2.3) turns out to be very useful in understanding why positive dispersion regime lasers are more susceptible to phase distortions than those operating in the negative dispersion regime.

Figure 5-1 shows a simulated spectrum from a 5 fs dispersion managed soliton laser operating in the net positive dispersion regime, with parameters similar to the laser of [53]. This is qualitatively representative of state-of-the-art few-cycle lasers, using simulated DCMs with phase ripple commensurate to those currently available (0.5% of a wave peak-to-peak). Above the spectrum, we show a plot of the net phase difference between the "soliton" and the continuum (i.e. the linear phase of the cavity) over a single round trip. Unlike the case for negative dispersion, a positive dispersion laser (assuming

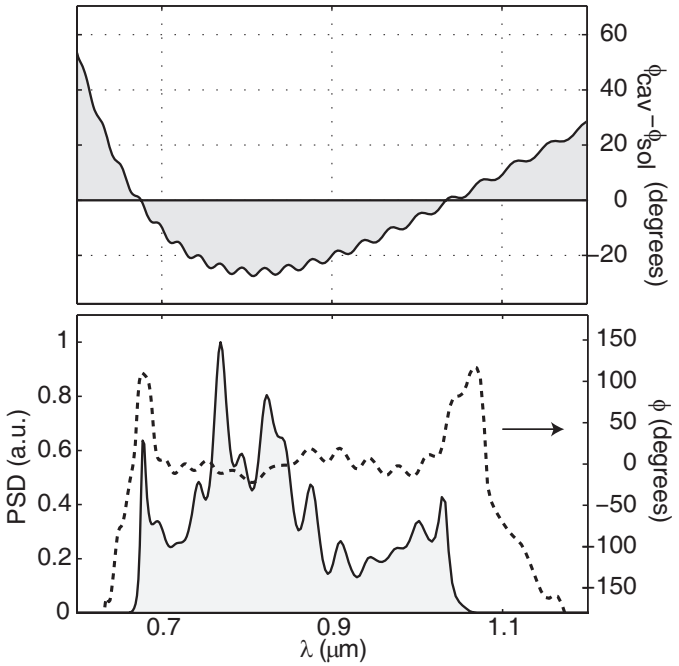

Figure 5-1: Simulated spectrum of few-cycle laser operating in the high-SPM, positive dispersion regime typical of solid state oscillators. The resonant interaction of the phase distortions with the main pulse result in the "M"-shaped power spectrum characteristic of few-cycle oscillators.

sufficient dispersion) tends to operate with the soliton always close to resonance with the continuum. Thus, even a small amount of ripple corresponds to a significant change in the net soliton-cavity phase, bringing the continuum significantly closer to resonance with the solition. Contrast this with the negative dispersion regime, illustrated in Figure 5-2, where the phase margin is well away from zero, despite the cavity having the same nonlinear coefficient. Negative dispersion cavities are much less susceptible to phase ripple as a result.

One possible interpretation of the amplitude ripples is interference from satellite pulses at a temporal distance corresponding to the inverse ripple period. As discussed earlier, this is certainly the case with saturable absorber mode-locked lasers with low or zero SPM. However, in the positive dispersion and high-SPM regime, well-formed satellite pulses cannot grow, as the nonlinearity renders them poorly phase matched to the main pulse. Inspection of both simulation and experimental measurement of actual lasers confirm that there are indeed no appreciable satellite pulses, only a complicated pedestal structure (see, for example, Figure 9-8 in Chapter 9). The mechanism behind the spectral peaks in a positive dispersion DMS laser is evidently fundamentally different from that of a saturable absorber-only laser, the most prominent evidence of this being the fact that the ripples increase towards the wings in the positive dispersion regime but decrease in the SA-only or negative dispersion regime.

Our proposed explanation for these observations is that when there is significant positive dispersion, the main soliton can be nearly phase-matched to narrowband quasi-

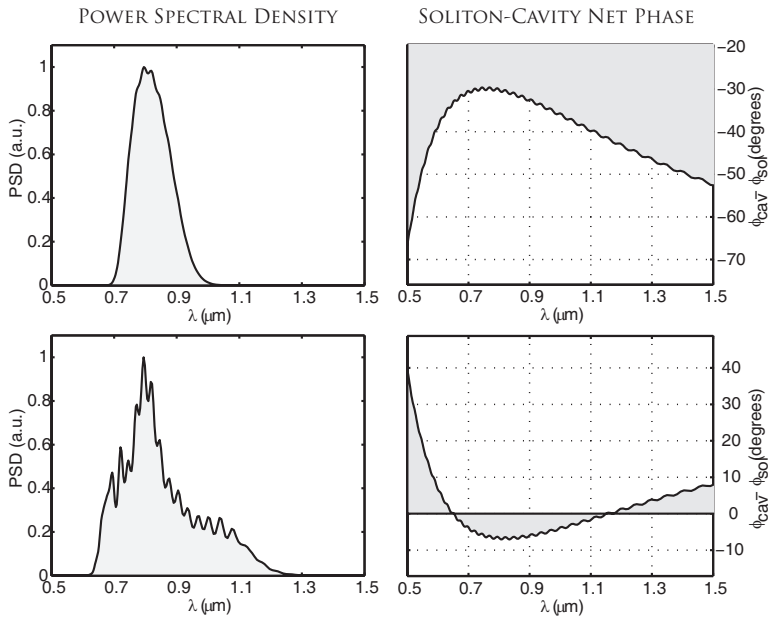

Figure 5-2: Comparison between two simulated laser cavities with mirror phase distortions. The two cavities differ only in thesign of net dispersion. The negative dispersion cavity operates with a significant margin between the soliton and the cavity, greatly attenuating the effects of the mirror's phase distortions.

solitons in the wings, which are supported by local negative dispersion caused by the phase oscillations. These quasi-solitons have their own (much smaller) soliton phase which is why they are more coupled at wavelengths where the main soliton approaches resonance with the continuum, but only while the margin is negative (as it would be for a normal soliton).

The ripples in a positive dispersion laser are thus somewhat analogous to Kelly sidebands [2]. However, they are fundamentally distinct in several ways. For one, they are more broadband than Kelly sidebands and correspond to coupling to a narrowband pulse, rather than a sharp continuum state. (To wit, in Figure 5-1, sharp Kelly sideband-like features can be seen at the actual zero-crossings, in addition to the wider peaks found at the local phase maxima.) Second, phase oscillation-induced peaks manifest only in positive dispersion lasers where the phase margin between soliton and continuum is negative. As also seen in Figure 5-1, once the margin becomes positive, the spectral oscillations go away. This further supports the notion that it is not a simple continuum resonance.

5.1.3 Origin of M-shaped spectra in few-cycle lasers

Few-cycle dispersion managed soliton lasers have a characteristic M-shaped spectrum, as illustrated in Figures 5-1 and 9-8. At first glace, it might seem that this is caused solely by the kind of dynamics seen in All Normal Dispersion (ANDi) lasers. Such lasers have been studied by Bélanger and Kalashnikov [54, 55], and realized experimentally by

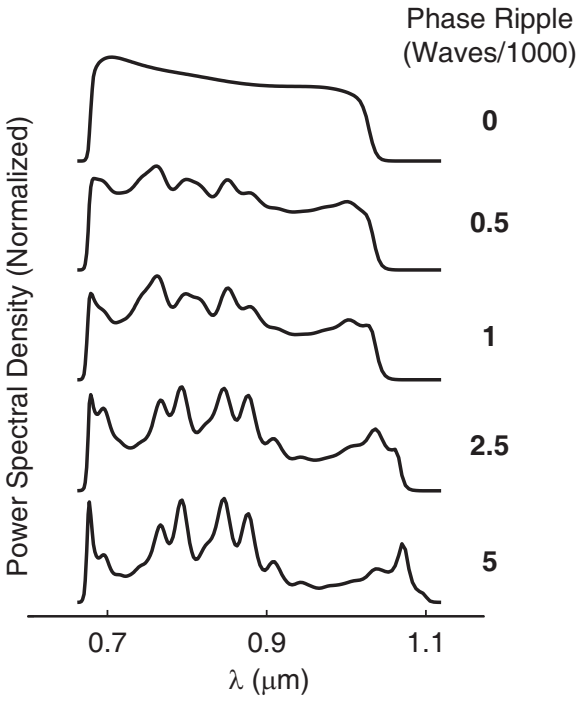

Figure 5-3: Set of simulated spectra for a DM soliton laser with varying amounts of intracavity phase distortion. The top plot is for perfect mirrors, the last corresponds to phase ripples typically produced by a well-matched DCM pair. The M-shaped spectrum is due to increasingly resonant coupling between the main pulse and narrow-band peaks supported by negative local dispersion.

Chong [56].

However, the spectral peaks predicted by the theory of Bélanger occur in distributed Ginzburg-Landau models without periodic perturbations or phase ripples. According to Kalashnikov, the M-shaped spectra in such stretched pulse lasers occurs due to fourth-order dispersion. This suggests another mechanism is at work in the case of few-cycle solid-state lasers.

The resonant soliton-soliton coupling effects proposed in Section 5.1.2 may provide an explanation for the characteristic 'M'-shaped spectra. Towards the edges of the spectrum, the positive dispersion causes the cavity phase oscillation peaks to be increasingly resonant with the soliton. This causes power to preferentially build in the wings, at the expense of the region just outside the gain bandwidth. The result is a spectrum that is high in the middle (due to gain) and high at the edges (due to resonant coupling).

To verify this theory, we performed a series of simulations for a range of phase ripple magnitudes. The results are shown in Figure 5-3, and lend support to the notion that few-cycle lasers owe their spectral structure to the mechanism just described. For zero phase distortion, the spectrum is perfectly smooth, with no wing enhancement, and power decreasing away from the gain center. It is only as phase ripple is introduced that power is pulled out of the center and into the wings, culminating in spikes for the phase oscillations closest to the soliton-soliton resonance. Note that the dispersion is the same

in each case; only the magnitude of the fast phase oscillations were changed.

5.2 Toward single-cycle pulses from an oscillator

The presence of significant ripple in dispersion compensating mirrors forces cavities to be operated at relatively high levels of SPM and positive dispersion. The large nonlinearity is needed to quell satellite pulses. It is well known that large nonlinearities lead to pulse instability [57, 3], however. This means that larger amounts of dispersion are required than would be ideal from a bandwidth standpoint.

Current sub-two-cycle lasers are especially subject to phase distortions; present thin film mirror manufacturing technology is not capable of producing octave spanning mirrors without significant phase oscillations (i.e. below 0.1% of a wave). Few-cycle lasers thus must be operated in a high-SPM/high-dispersion regime. Unfortunately, the large nonlinearity results in significant spatio-temporal coupling, which degrades output beam quality. The high nonlinearity also manifests as an M-shaped spectral phase (see Figure 5-1), hindering compression of the pulses to their fourier limit.

Since mirror phase distortions are evidently limiting laser performance, it is motivating to consider what would be possible with improved mirrors. Figure 5.2 shows a set of three cavity simulations for mirrors, simulated to have roughly one fifth the ripple of current designs.

In the lower panel of Figure 5.2, we show what is possible if the mirror dispersion is tuned to increase the resonant wing enhancement, pushing the zero-crossing further out. The resulting bandwidth is sufficient for a single-cycle pulse, and yet this is achieved with a smaller nonlinearity than that of the laser simulated in Figure 5-1. With less phase distortion, the cavity can be operated with less dispersion, and thus less nonlinearity is needed to achieve a given bandwidth. If such a laser could be realized, it would have higher beam quality and a smoother spectral phase, simplifying compression.

It should be stressed that the results here are from theoretical simulations, and therefore should be taken with some degree of skepticism, especially with regard to the possiblity for single-cycle pulses. In particular, shock terms were not taken into account in the simulation, and SPM was assumed to be instantaneous. Nonetheless, given the excellent agreement between similar simulations and experimentally measured two-cycle lasers [52], it is reasonable to conclude that significant advances in laser performance are to be had with further improvements in mirror manufacturing and design. Improvements in the latter can come partly from the insights made here, as well as work underway to improve the numerical optimization process. However, improvements must also occur in coating manufacturing tolerances if dispersion compensating mirrors—and thus modelocked lasers—are to reach their full potential. While much effort has been focused lately on external cavity compression, we should not forget that there is still much room left to improve oscillator technology. The potential for pulses approaching a single-cycle directly from an oscillator warrants further effort to improve dispersion compensating mirror technology.

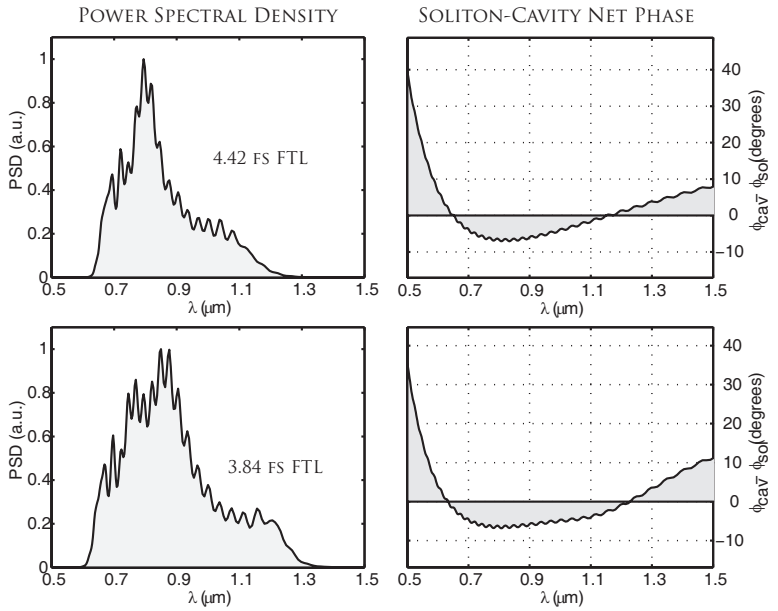

Figure 5-4: (top) A simulated laser with low phase distortion mirrors (0.1% waves of phase ripple) allowing for a relatively small positive net cavity dispersion (15 microns-worth of BaF). (bottom) The same laser with DCMs slightly modified to push out the soliton-continuum phase crossing, leading to a single-cycle pulse directly from the oscillator.

5.3 PDR versus GDD optimization

In the two examples of PDR optimization from the previous chapter, the optical systems were entirely linear. It was thus possible to use a merit function that directly represented the performance of the system. The situation is not so simple in the case of a nonlinear system, such as a mode-locked laser cavity, where it is impossible to derive an analytic function of the mirror layers that is monotonic to the pulse width or the bandwidth.

Nevertheless, there is reason to suspect that a minimal PDR design will at least be better than a GDD-based design. In the absense of other considerations, minimizing the energy taken out of the pulse through perturbations should be prioritized over minimizing the delay of satellite pulses. Of course, energy taken out can be fed back into the pulse through the interplay of dispersion and SPM, as discussed in Section 5.1 and thus we do not expect to find a simple relation between laser performance and PDR.

In the absence of analytic solutions, we must find some way to estimate the relative efficacies of one or more candidate functions to be used as a proxy for laser performance. In other words, while minimizing neither GDD or PDR will directly maximize our laser bandwidth, can we at least figure out which is superior in most cases? The general question may be cast as follows.

Our mirror is parametrized by a state vector $\mathbf{x}$ of layer thicknesses. Let $g(\mathbf{x})$ be a function representing some simulated performance figure of a laser composed of such mirrors. Let $f(\mathbf{x})$ be a potential mirror merit function we wish to consider as a stand-in for g, such as integrated PDR, for instance. How may we decide the extent to which f is

a valid approximation to g in the context of an optimization?

We take a statistical approach that involves computing $f(\mathbf{x})$ and $g(\mathbf{x})$ for a random set of $\mathbf{x}$ in the region of a feasible solution. If we consider each value of g and f in the monte carlo sample to be values taken from random variables G and F, respectively, we may cast the problem as one of random variable estimation. Ideally, f will be a close as possible to a linear function of g so that they share the same extremal points. We thus wish for F to be a good linear estimator of G. From elementary statistical estimation theory [58] we know that the least squares linear estimator of G will be given by

$$\mathbf{E}[G] + \rho(G,F)\frac{\sigma_G}{\sigma_F}\big(F - \mathbf{E}[F]\big), \tag{5.1}$$

where ρ is the correlation coefficient

$$\rho(G,F) = \frac{\mathrm{cov}(G,F)}{\sigma_G \sigma_F}. \tag{5.2}$$

It is convenient to normalize and shift our samples so that their means are zero and their variances are unity. Let $\bar{G} = (G - \mathbf{E}[G])/\sigma_G$, and so on. Inspection of (5.1) shows that the best estimator of $\bar{G}$ is simply $\bar{F}$ multiplied by the correlation coefficient. From estimation theory we then have the root mean squared error

$$\text{RMS Estimation Error} = \sqrt{1 - \rho^2\left(\bar{G},\bar{F}\right)}. \tag{5.3}$$

Assuming our Monte Carlo sampled points $\mathbf{x}$ are representative of mirrors actually seen during optimization, the absolute value of the correlation coefficient will be a good predictor of how well $f(\mathbf{x})$ will approximate our desired goal. This method should be generally applicable to any situation where we wish to optimize a physical observable in a system that is not practical to directly compute.

If we take a large but representative set of samples of our proposed merit function f and our actual goal g, we would expect f to be a good minimizer of g if F is statistically well-correlated to G, and this is exactly what the above is saying. Furthermore, this approach provides for an intuitive method of computation. If we plot all the normalized values of F and G and perform a linear regression, (5.1) implies that the correlation coefficient will be approximated by the slope of the fit line (if sufficient points are used). An ideal merit function will fall tightly on a line, and a completely useless merit function will result in a random scatter of points. Inspection of such a plot may provide information not only regarding the correlation, but how it fails and perhaps how to correct it.

As a test case, we take a sub-10 femtosecond dispersion managed soliton cavity simulated using an actual DCM design. We take as our target variable G the RMS change in the pulse spectrum given random perturbation to the mirror layers. For each perturbation, we compute the resulting integrated PDR and GDD of the mirror, considering these samples from random variables F_1 and F_2. Computing 5000 such points[1] using

[1] Computing the 5000 cavity perturbations using the Newton-Krylov solver took roughly 10 minutes on an eight core computer. Doing so with a standard dynamic solver would have taken about two hours.

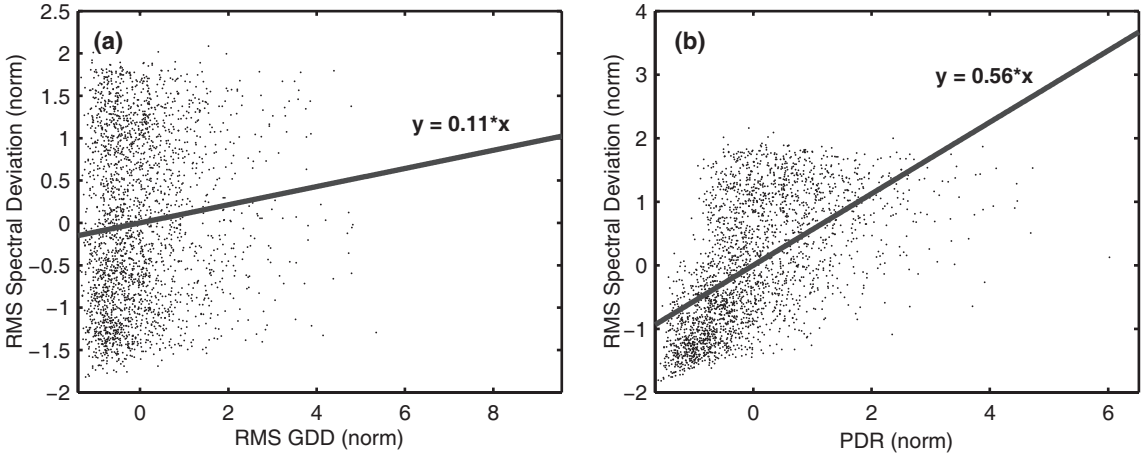

Figure 5-5: Monte Carlo results showing the correlation between cavity bandwidth and PDR and GDD, for a sub-10 fs laser.

the Newton-Krylov solver and normalizing the output as explained earlier, we generate the plots shown in Figures 5.3(a) and (b). Obviously, neither PDR nor GDD is a perfect predictor of perturbation to the steady state pulse. However, it is visually clear that PDR is far more predictive; the GDD error is essentially uncorrelated with the change in spectrum. A linear fit to each is shown, and the correlation coefficient between PDR and the RMS spectral deviation is 0.56, five times the correlation coefficient of GDD.

We conclude that using PDR to optimize cavity mirrors is preferable to using GDD. Having said that, neither is particularly well-correlated with laser performance, nor would we expect them to be given the complex nonlinear interactions possible between satellite pulse and main pulse. Developing the ability to directly optimize the cavity performance by putting a model of the laser somehow "in the loop" is the only way to truly maximize mirror performance. Given the tremendous gains possible from improving mirror performance (as demonstrated in Section 5.2) this is a goal well worth pursuing.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 6

Thin Film Phase Response Optimization

In Chapter 4 we found that group delay dispersion is usually not the best choice as an optimization criterion. Nonetheless, because of difficulties dealing with phase directly (since phase must always be computed via an inverse tangent, leading to branch cut discontinuities) is is best to compute the phase by integrating the group delay after taking out any delay offset. In this chapter, we discuss the specifics of efficiently computing the group delay of a dielectric stack, including analytic gradients of the same. Analytic gradients are essential to the rapid convergence of optimization routines, as finite differences suffer from either round-off or truncation error.

6.1 Introduction

Dispersion compensating dielectric mirrors [59, 16] have played a critical role in the development of mode-locked lasers, with state-of-the-art mirror pairs allowing for femtosecond group delay control over nearly an octave of bandwidth.[17] Such precise control of optical phase has enabled pulses containing only a few optical cycles directly from an oscillator [34, 60]. Furthermore, the compression or manipulation of pulses outside of the laser cavity requires the design of mirrors with prescribed group delay dispersion over extremely wide bandwidths [61]. Thus, the synthesis of multilayer filters with prescribed phase properties has received increasing interest in the past decade or so [62, 63].

When numerically optimizing a thin film structure, the majority of the computational effort is dedicated to repeatedly computing the gradient of the merit function, and perhaps also the merit function alone (such as during line searches with numerical derivatives). Should the merit function include the spectral dispersion, one must be able to compute the gradient of phase derivatives. While analytic methods have been published for computing gradients of simple reflectivity [64], no work has been shown on analytically computing gradients of dispersion. To our knowledge, this is the first published algorithm for computing analytic gradients of dispersion, approximate or otherwise.

We recently demonstrated an inductive method[65, 66] for computing analytic derivatives of multilayer phase to any order. Here, we extend this method to computing the

full analytic gradient of such phase derivatives. This was the first published algorithm to compute the exact analytic gradient of dispersion. The method is $O[nm^2]$ in terms of matrix multiplications, where n is the number of layers and m is the dispersion order. Furthermore, we show an approximation that allows for the accurate computation of dispersion gradients in only $O[nm]$, with significant improvement in practice even for $m = 1$ (group delay).

Computing analytic gradients is important for optimizing multilayer coatings for two reasons. First, the use of analytic derivatives avoids the issues of numerical stability associated with finite differences, improving accuracy and convergence.[67] Second, and perhaps most importantly, computing gradients using finite differences results in a gradient algorithm that scales as $O[n^2]$ in the number of layers, making it rather inefficient for complicated mirror systems.

While the general scheme shown in this paper can be applied to the computation of any order of dispersion, the implementation complexity increases for higher orders. Fortunately, as we be shown in later chapters, group delay is generally all that is needed.

6.2 Analytic Computation of Stack Phase Derivatives

Here, we review the analytic computation of phase derivatives for a dielectric stack. We leave out most details here, and only cover computation of first-order phase derivatives (group delay). However, the method can be directly extended to any order of dispersion. Further details and a discussion of the validity of the constant coupling approximation can be found in Ref. [66].

6.2.1 General Case

In this paper we will follow the convention established in the Ref. [66] and consider a dielectric stack whose total transfer matrix is written as

$$\mathbf{T}(k) = \begin{pmatrix} T_{11}(k) & T_{12}(k) \\ T_{12}^*(k) & T_{11}^*(k) \end{pmatrix}. \tag{6.1}$$

In our notation, $\mathbf{T}_\ell$ refers to the transfer matrix of the ℓth layer, which is defined to include only the interface reflection between it and the previous medium and propagation through the layer. We'll write $\mathbf{T}_{(\ell_2,\ell_1)} \equiv \mathbf{T}_{\ell_2} \cdots \mathbf{T}_{\ell_1+2}\mathbf{T}_{\ell_1+1}$ to refer to the matrix that goes from the end of layer ℓ_1 to the end of layer ℓ_2. The substrate can be handled as a final layer with a thickness of zero.

For convenience, we depart from computing phase derivatives in terms of frequency, as was done in Ref. [66], and use vacuum wavenumber, $k = \omega/c$, instead. This simply avoids having c appear in intermediate formulas, which will help when considering gradients. This also more closely matches the way computation is done in practice, making it easier to compare the results of this paper with the code provided.

In our notation, the transfer matrix operates on a vector whose components are the forward and reverse propagating wave amplitudes, respectively.[68] For reasons that will become clear later, we will write the matrix for the ℓth layer as the product of a full matrix

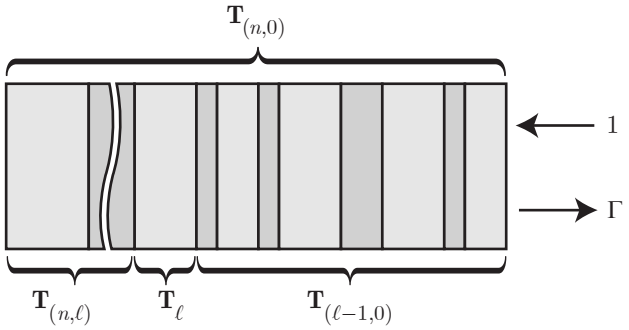

Figure 6-1: Diagram showing transfer matrix notation.

$\mathbf{D}_\ell$, which handles the transfer across the interface, followed by a diagonal matrix $\mathbf{P}_\ell$ that propagates through the layer:

$$\mathbf{T}_\ell \equiv \mathbf{P}_\ell \mathbf{D}_\ell \tag{6.2}$$

$$= \begin{pmatrix} e^{-i\tilde{n}_\ell(k)d_\ell k} & 0 \\ 0 & e^{i\tilde{n}_\ell(k)d_\ell k} \end{pmatrix} \times \frac{1}{2}\begin{pmatrix} 1+p_\ell(k) & 1-p_\ell(k) \\ 1-p_\ell(k) & 1+p_\ell(k) \end{pmatrix}, \tag{6.3}$$

where $\tilde{n}_\ell(k) \equiv n_\ell(k)\cos\theta_\ell$ is the effective index (which takes into account the propagation angle θ_ℓ of the wave) and $p_\ell(k)$ is the ratio

$$p_\ell(k) \equiv \begin{cases} \frac{\tilde{n}_{\ell-1}(k)}{\tilde{n}_\ell(k)} & \text{TE polarization,} \\ \frac{\tilde{n}_{\ell-1}(k)n_\ell^2(k)}{\tilde{n}_\ell(k)n_{\ell-1}^2(k)} & \text{TM polarization.} \end{cases} \tag{6.4}$$

The complex transmission and reflection coefficients are given from the elements of the transfer matrix (6.2) by

$$\Gamma(k) = -\frac{T_{12}^*(k)}{T_{11}^*(k)}, \tag{6.5}$$

$$T(k) = T_{11}(k) - \frac{|T_{12}(k)|^2}{T_{11}^*(k)}, \tag{6.6}$$

respectively.

To determine the mth frequency derivative of phase (either in reflection or transmission) one must know the zeroth through mth derivatives of the elements of the transfer matrix. As discussed in Ref. [66], this can be done in $O[nm^2]$ operations by inductively computing the matrices $\mathbf{T}_{(\ell,0)}$ for ℓ from one to n. In the case of group delay, for example, this means repeatedly computing matrices of the form

$$\mathbf{T}_{(\ell,0)} = \mathbf{T}_\ell \mathbf{T}_{(\ell-1,0)}, \tag{6.7}$$

$$\frac{\partial \mathbf{T}_{(\ell,0)}}{\partial k} = \frac{\partial \mathbf{T}_\ell}{\partial k}\mathbf{T}_{(\ell-1,0)} + \mathbf{T}_\ell \frac{\partial \mathbf{T}_{(\ell-1,0)}}{\partial k}. \tag{6.8}$$

6.2.2 Constant Coupling Approximation

Assuming $p'_\ell(k) \to 0$ yields a significant decrease in complexity for computing dispersion, from $O[m^2]$ to $O[m]$, as experimentally verified in Table 1 of Ref. [66]. This implies that we neglect the derivatives of the $\mathbf{D}_\ell(k)$ matrices that couple between forward and backward waves, hence the name. The reason why this is more efficient can be seen by considering the derivative of (6.2) under the approximation,

$$\mathbf{T}'_\ell(k) \approx \mathbf{P}'_\ell \mathbf{D}_\ell \tag{6.9}$$

$$= \begin{pmatrix} -id_\ell\left[\tilde{n}_\ell(k) + k\tilde{n}'_\ell(k)\right] & 0 \\ 0 & id_\ell\left[\tilde{n}_\ell(k) + k\tilde{n}'_\ell(k)\right] \end{pmatrix} \mathbf{P}'_\ell \mathbf{D}_\ell \tag{6.10}$$

$$\equiv -id_\ell\left[\tilde{n}_\ell(k) + k\tilde{n}'_\ell(k)\right]\boldsymbol{\sigma}_3 \mathbf{T}_\ell(k). \tag{6.11}$$

For convenience, we have used the Pauli matrix

$$\boldsymbol{\sigma}_3 \equiv \begin{pmatrix} +1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{6.12}$$

though there is obviously no implied connection between the present application and spinors except where the commutation relations may prove useful (e.g two such derivative approximations in succession cancel to a scalar). The reason for the simple form is that $\mathbf{P}_\ell$ is diagonal, and so taking the derivative of it is equivalent to left multiplying it with another diagonal matrix. Since the symmetry of transfer matrices is such that we only need to keep track of one row, left multiplication by a diagonal matrix is computationally equivalent to a single scalar multiplication (though it still does not commute, of course, so it cannot be lumped with other scalars). The net result is that the first derivative matrix can be computed using only one matrix multiplication instead of two.

6.2.3 Accuracy of constant coupling approximation

Empirically, works well for computing at least second order dispersion for any mirror. Moreover, there is theoretical reason to believe that this should be the case, especially for the chirped structures used in dispersion compensating mirrors. Using results from the theory of double chirped mirrors developed by Kärtner and Matuschek[16, 17], we show that the constant coupling assumption is physically justified. This is supported by results from a computation on an actual chirped mirror design.

It is well known that a dielectric stack can be modeled using coupled-mode theory[69]. In this context, the local coupling coefficient $\kappa(\omega)$ is approximately proportional to the Fresnel reflection between layers:

$$\kappa(\omega) \approx -2r(\omega) = -2\frac{1-p(\omega)}{1+p(\omega)}, \tag{6.13}$$

where we have dropped the ℓ subscript for this section. Thus, setting $p'(\omega) = 0$ is tantamount to assuming that $\kappa'(\omega) = 0$. Matuschek showed that a chirped mirror can be transformed into a weakly-inhomogeneous coupled-mode problem which can then be

solved using the WKB approximation. The group delay of a chirped mirror is then found to be (see equation (52) in [69])

$$\tau_g(\omega) = -2\int_0^{m_t(\omega)} \frac{\partial}{\partial\omega}\sqrt{\delta^2(m)-\kappa^2(m)}\,dm, \tag{6.14}$$

where m is a normalized spatial variable which parameterizes location within the mirror, $m_t(\omega)$ is the location of the classical turning point, and $\delta(m)$ is the coupled-mode detuning coefficient. Matuschek argues that in the classically accessible region of the mirror, the detuning coefficient dominates the coupling coefficient and so, to zeroth order, we have

$$\begin{aligned}\tau_g^{(0)}(\omega) &= 2\int_0^{m_t(\omega)} \frac{\partial}{\partial\omega}|\delta(m)|\,dm,\\ &= 2\pi\int_0^{m_t(\omega)} \frac{1}{k_B}\,dm.\end{aligned} \tag{6.15}$$

Finally, if the Bragg wavelength k_B is taken to first order as

$$k_B(m) = k_0 + k_1 m, \tag{6.16}$$

then the group delay can be found analytically to be approximately

$$\tau_g^{(0)}(\omega) = \frac{2\pi}{ck_1}\left[\ln\left(1-\frac{\kappa(\omega)}{\pi}\right)+\ln\left(\frac{ck_0}{\omega}\right)\right]. \tag{6.17}$$

The contribution of $\kappa(\omega)$ to the group delay will be minor for $r \ll 1$. Furthermore, $\kappa'(\omega)$ only appears in GDD or higher. Thus, we would expect that neglecting $p'(\omega)$ would be a good approximation for GD, with fair but decreasing accuracy for higher orders of dispersion. This is borne out in the examples shown in [66]. Nonetheless, neglecting $\kappa'(\omega)$ is still only approximately correct for GD due to approximations made in going from (6.14) to (6.15).

This can all be understood intuitively by recognizing that the effective round trip optical path length is dominated by the material indices and the location of the classical turning point, the latter of which is not strongly affected by the Fresnel reflection coefficient, $r(\omega)$. This is the case even when the material dispersion itself is large such that local index derivatives must be considered. Thus, despite the fact that $p'(\omega)$ and $n'(\omega)$ contribute to GD at the same order, and both appear to the same order in ω (since $c/d_\ell \sim \omega$), it is only necessary to consider $n'(\omega)$ when computing $\phi'(\omega)$.

6.3 Gradients of $\mathbf{T}(k)$

At the core of computing the gradient of group delay is the problem of computing the gradients of $\mathbf{T}_{(n,0)}(k)$ and $\mathbf{T}'_{(n,0)}(k)$ with respect to the n layer thicknesses, denoted as d_ℓ.

To begin with, we factor the total transfer matrix to isolate the ℓth layer:

$$\mathbf{T} = \mathbf{T}_{(n,\ell)}\,\mathbf{T}_\ell\,\mathbf{T}_{(\ell-1,0)}. \tag{6.18}$$

The ℓth gradient element is then simply

$$\frac{\partial \mathbf{T}}{\partial d_\ell} = \mathbf{T}_{(n,\ell)}\,\frac{\partial \mathbf{T}_\ell}{\partial d_\ell}\,\mathbf{T}_{(\ell-1,0)} \tag{6.19}$$

$$= -ik\tilde{n}_\ell \mathbf{T}_{(n,\ell)}\,\boldsymbol{\sigma}_3\,\mathbf{T}_{(\ell,0)}, \tag{6.20}$$

where we have used (6.2) to compute the derivative.

Inspection of (6.20) immediately suggests the general method for computing the gradients: If we precompute all the front matrices $\mathbf{T}_{(\ell,0)}$ as well as the back matrices $\mathbf{T}_{(n,\ell)}$ for each layer ℓ, then the gradients can be computed trivially in n matrix multiplications. (The multiplication by the Pauli matrix $\boldsymbol{\sigma}_3$ is not counted as it is computationally equivalent to a scalar multiplication, as explained in the previous section.) More importantly, the front matrices can be computed in n matrix multiplications by simply computing them inductively as shown in Section 6.2. The same is true of the back matrices, though there are some complications that will be covered in Section 6.4.3. The entire gradient can thus be computed in $O[n]$ matrix multiplications, a significant improvement over the $O[n^2]$ operations required for a naïve finite difference gradient.

6.4 Gradients of $\mathbf{T}'(k)$

6.4.1 General Method

The scheme outlined in the previous section can be applied in a straightforward way to find gradients of any order wavenumber derivative, albeit with significant growth in complexity as higher derivatives are used. As justified earlier in Chapter 5, we will simply demonstrate the method for the first wavenumber derivative used to compute GD. Taking the k derivative of (6.18), the matrix product rule gives us the following decomposition:

$$\mathbf{T}'(k) = \mathbf{T}'_{(n,\ell)}\,\mathbf{T}_\ell\,\mathbf{T}_{(\ell-1,0)} + \mathbf{T}_{(n,\ell)}\,\mathbf{T}'_\ell\,\mathbf{T}_{(\ell-1,0)} + \mathbf{T}_{(n,\ell)}\,\mathbf{T}_\ell\,\mathbf{T}'_{(\ell-1,0)}. \tag{6.21}$$

The ℓth gradient element is then

$$\frac{\partial \mathbf{T}'(k)}{\partial d_\ell} = -ik\tilde{n}_\ell \mathbf{T}'_{(n,\ell)}\,\boldsymbol{\sigma}_3\,\mathbf{T}_{(\ell,0)} + \mathbf{T}_{(n,\ell)}\,\frac{\partial \mathbf{T}'_\ell}{\partial d_\ell}\,\mathbf{T}_{(\ell-1,0)} - ik\tilde{n}_\ell \mathbf{T}_{(n,\ell)}\,\boldsymbol{\sigma}_3\,\mathbf{T}'_{(\ell,0)}, \tag{6.22}$$

where we've applied the result in (6.20) to simplify the outer two terms. From the definition of T_ℓ in (6.2) we obtain

$$\frac{\partial \mathbf{T}'(k)}{\partial d_\ell} = -ik\tilde{n}_\ell \left[\mathbf{T}'_{(n,\ell)}\boldsymbol{\sigma}_3\mathbf{T}_{(\ell,0)} + \mathbf{T}_{(n,\ell)}\boldsymbol{\sigma}_3\mathbf{T}'_{(\ell,0)}\right] + \frac{1}{2p}\mathbf{T}_{(n,\ell)}\begin{pmatrix} 2p(d_\ell k\tilde{n}+i)(\tilde{n}+k\tilde{n}') - ik\tilde{n}p' & -ik\tilde{n}p'e^{-i2d_\ell k\tilde{n}} \\ ik\tilde{n}p'e^{i2d_\ell k\tilde{n}} & 2p(d_\ell k\tilde{n}-i)(\tilde{n}+k\tilde{n}') + ik\tilde{n}p' \end{pmatrix}\mathbf{T}_{(\ell,0)}, \quad (6.23)$$

where all primes refer to wavenumber derivatives and we've dropped some of the unambiguous ℓ subscripts for convenience. In arriving at the above, we solved for the matrix which takes the simultaneous wavenumber and layer thickness derivative of $\mathbf{T}_\ell$.

The front and back derivative matrices, $\mathbf{T}'_{(\ell,0)}$ and $\mathbf{T}'_{(n,\ell)}$ respectively, are found using the exact methods of Section 6.2. The front matrices are available directly as they are the intermediate results of computing $\mathbf{T}'$. The back derivative matrices require extra computation, however, and can be found by proceeding through the stack in reverse, doing right multiplications in lieu of left multiplications. (In the case where the constant coupling approximation is used, there is a more efficient way to compute the back matrices, discussed in Section 6.4.3.)

Higher order dispersion terms beyond what we've shown here can be computed similarly, and the computation of the front and back matrices will scale as $O[nm^2]$, as per Section 6.2. However, the number of matrix multiplications required for the final matrices [e.g. equation (6.23)] grows exponentially, as $O[3^m]$, and the complexity of the *elements* in each matrix increases considerably. Thus, higher order dispersion quickly becomes infeasible with this method, and even the $m = 1$ case (for group delay) is rather complex, as can be seen from (6.23). In the next section, we will show how to use the constant coupling approximation to greatly simplify the gradient computation, significantly speeding up low order dispersion and enabling the gradient computation of higher order dispersion.

6.4.2 Constant Coupling Approximation

We have already seen how the assumption that $\mathbf{D}'_\ell(k) \to 0$ greatly speeds up the computation of the front and back matrices, as needed for (6.23). However, it also greatly simplifies the final terms in (6.23). Under the constant coupling approximation, the off diagonal terms in the last product term vanish, leaving a trivial scalar multiplication in lieu of a matrix product. Thus, equation (6.23) becomes

$$\frac{\partial \mathbf{T}'(k)}{\partial d_\ell} = -ik\tilde{n}_\ell \left[\mathbf{T}'_{(n,\ell)}\boldsymbol{\sigma}_3\mathbf{T}_{(\ell,0)} + \mathbf{T}_{(n,\ell)}\boldsymbol{\sigma}_3\mathbf{T}'_{(\ell,0)}\right] + \frac{i(\tilde{n}+k\tilde{n}')}{2p}\mathbf{T}_{(n,\ell)}\boldsymbol{\sigma}_3\mathbf{T}_{(\ell,0)} + d_\ell k\tilde{n}(\tilde{n}+k\tilde{n}')\mathbf{T}, \quad (6.24)$$

where the third term is a scalar multiplication of the zeroth-order gradient element from (6.23), and the last term is just a scaling of the total transfer matrix.

In general, the simplification afforded by the constant coupling approximation not

only takes the front and back matrix computation from $O[nm^2]$ to $O[nm]$, but also makes the final matrix expression [e.g. (6.24)] scale as $O[2^m]$ instead of $O[3^m]$, as all terms consist of only two full matrices. In the specific case of group delay, the total speedup in practice is roughly a factor of two, assuming n is large enough such that the computation of front and back matrices dominate.[66]

Finally, the approximation makes it reasonable to compute GDD gradients, affording a speed-up of roughly a factor of four. Taking another k derivative of (6.24) and combining like terms gives

$$\begin{aligned}\frac{\partial \mathbf{T}''(k)}{\partial d_\ell} = &-i[(n+kn')/2p - n]\left[\mathbf{T}'_{(n,\ell)}\boldsymbol{\sigma}_3\mathbf{T}_{(\ell,0)} + \mathbf{T}_{(n,\ell)}\boldsymbol{\sigma}_3\mathbf{T}'_{(\ell,0)}\right] - \\ &2ikn\left[\mathbf{T}'_{(n,\ell)}\boldsymbol{\sigma}_3\mathbf{T}'_{(\ell,0)} + \mathbf{T}''_{(n,\ell)}\boldsymbol{\sigma}_3\mathbf{T}_{(\ell,0)} + \mathbf{T}_{(n,\ell)}\boldsymbol{\sigma}_3\mathbf{T}''_{(\ell,0)}\right] + \\ &\frac{in'}{2p}\mathbf{T}_{(n,\ell)}\boldsymbol{\sigma}_3\mathbf{T}_{(\ell,0)} + (d_\ell\tilde{n}^2 + 2d_\ell nn'k)\mathbf{T} + d_\ell nk(\tilde{n}+\tilde{n}'k)\mathbf{T}'. \quad (6.25)\end{aligned}$$

The individual terms above are all derived in Ref. [66]. While it is certainly possible to compute GDD gradients without the constant coupling approximation, the final gradient terms become extremely cumbersome. Fortunately, given the accuracy of the constant coupling approximation for GDD, as demonstrated in Fig. 6-3, there is little reason to use exact GDD computations except perhaps as a final refinement step.

6.4.3 Efficient Computation of Back Derivative Matrices

An element of extending the constant coupling method of Section 6.2.2 to gradients that is not straightforward is the issue of efficiently handling the back matrices, $\mathbf{T}_{(n,\ell)}$, which take the fields from the interior of the stack to the end. The efficiency of the constant coupling approximation hinges on the fact that we build the full matrix from successive left multiplications of **PD** layers, as in (6.2). Were we to simply compute the matrices by using right multiplications and working our way backwards from the end, therefore, the constant coupling approximation would not yield any advantage.

The way around this is to actually compute the back matrices as the "front" matrices for the reversed stack. This can be done without having to recompute any of the individual transfer matrices by using the reversal theorem of transfer matrices[70]

$$\mathbf{T}^R = \frac{\mathbf{T}^\dagger}{|\mathbf{T}|}, \quad (6.26)$$

where $\mathbf{T}^R$ denotes the transfer matrix for the reversed stack. In terms of a specific layer, we must also take into account the fact that the propagation must occur after the boundary. With this in mind, we can write a single layer of the reversed stack in terms of the components of the original stack,

$$\mathbf{T}^R_{n-\ell+1} = \mathbf{P}^\dagger_\ell \frac{\mathbf{D}^\dagger_{\ell-1}}{|\mathbf{D}_{\ell-1}|}. \quad (6.27)$$

Note that the propagation matrix is once again exposed on the left side as in (6.9).

If we were to have to compute all of the determinants arising from (6.27), the extra complexity would mitigate any advantage of the constant coupling approximation. Fortunately, however, we can safely ignore the determinants as they cancel in the end. To see how, consider the computation for $\mathbf{T}_{(n,\ell)}$ in terms of the reverse stack,

$$\mathbf{T}_{(n,\ell)} = \frac{\left(\mathbf{T}^R_{(\ell,0)}\right)^\dagger}{\left|\mathbf{T}^R_{(\ell,0)}\right|} \tag{6.28}$$

$$= \frac{\left[\frac{\mathbf{D}^\dagger_{\ell-1}}{|\mathbf{D}_{\ell-1}|}\left(\mathbf{P}^\dagger_{\ell-1}\frac{\mathbf{D}^\dagger_\ell}{|\mathbf{D}_\ell|}\right)\cdots\left(\mathbf{P}^\dagger_{n-1}\frac{\mathbf{D}^\dagger_n}{|\mathbf{D}_n|}\right)\right]^\dagger}{\left|\frac{\mathbf{D}^\dagger_{\ell-1}}{|\mathbf{D}_{\ell-1}|}\left(\mathbf{P}^\dagger_{\ell-1}\frac{\mathbf{D}^\dagger_\ell}{|\mathbf{D}_\ell|}\right)\cdots\left(\mathbf{P}^\dagger_{n-1}\frac{\mathbf{D}^\dagger_n}{|\mathbf{D}_n|}\right)\right|}. \tag{6.29}$$

The groups in parentheses represent the individual layer matrices of the reversed stack. Moving the determinants in the denominator outside the surrounding determinant yields the product of the squared determinants (since we are dealing with 2×2 matrices), giving us

$$\mathbf{T}_{(n,\ell)} = \frac{\left[\mathbf{D}^\dagger_{\ell-1}\left(\mathbf{P}^\dagger_{\ell-1}\mathbf{D}^\dagger_\ell\right)\cdots\left(\mathbf{P}^\dagger_{n-1}\mathbf{D}^\dagger_n\right)\right]^\dagger}{\frac{\left|\mathbf{D}^\dagger_{\ell-1}\left(\mathbf{P}^\dagger_{\ell-1}\mathbf{D}^\dagger_\ell\right)\cdots\left(\mathbf{P}^\dagger_{n-1}\mathbf{D}^\dagger_n\right)\right|}{(|\mathbf{D}_{\ell-1}||\mathbf{D}_\ell||\mathbf{D}_n|)^2}}. \tag{6.30}$$

The determinants then all cancel (the determinant of a propagation matrix is one) yielding a very simple and direct way to go from the individual layer matrices to the back matrices,

$$\mathbf{T}_{(n,\ell)} = \left[\mathbf{D}^\dagger_{\ell-1}\left(\mathbf{P}^\dagger_{\ell-1}\mathbf{D}^\dagger_\ell\right)\cdots\left(\mathbf{P}^\dagger_{n-1}\mathbf{D}^\dagger_n\right)\right]^\dagger \tag{6.31}$$

This expression has two advantages. First, and most importantly, it allows us to us to build up the back matrices using successive left multiplications of **PD** matrix pairs, enabling the use of the fast approximate algorithm discussed in Section 6.2.2 to find the k derivatives of (6.31). Second, everything on the right hand side of (6.31) has already been computed in finding the front matrices, $\mathbf{T}_{(\ell,0)}$. Consult the code referenced in Section 6.8 for further details and a demonstration.

6.5 Dispersion Gradients from Matrix Gradients

Having computed gradients of the full transfer matrix and its k derivatives, the final step in any optimization will be the translation of those values into gradients of dispersion for use in the merit function gradient computation. For reference, we provide formulas

for the case of reflection group delay:

$$\Gamma'(k) = \frac{T_{12}T'_{11} - T'_{12}T_{11}}{T_{11}^2}, \tag{6.32}$$

$$\phi'(k) = \frac{\Im[\Gamma']\Re[\Gamma] - \Re[\Gamma']\Im[\Gamma]}{|\Gamma|^2}, \tag{6.33}$$

$$\nabla\Gamma(k) = -\frac{1}{T_{11}}(\Gamma\nabla T_{11} + \nabla T_{12}), \tag{6.34}$$

$$\nabla|\Gamma(k)|^2 = 2\left(\Re[\nabla\Gamma]\Re[\Gamma] + \Im[\nabla\Gamma]\Im[\Gamma]\right), \tag{6.35}$$

$$\nabla\Gamma'(k) = \frac{1}{T_{12}^2}\left[\nabla T_{12}T'_{11} + \nabla T_{11}T'_{12} + \Gamma(2\nabla T_{11}T'_{11} - T_{11}\nabla T'_{11} - T_{11}\nabla T'_{12})\right], \tag{6.36}$$

$$\nabla\phi(k) = \frac{\Im[\nabla\Gamma]\Re[\Gamma] - \Re[\nabla\Gamma]\Im[\Gamma]}{|\Gamma|^2}, \tag{6.37}$$

$$\nabla\phi'(k) = \frac{1}{|\Gamma|^2}\left[\Im[\nabla\Gamma']\Re[\Gamma] - \Re[\nabla\Gamma']\Im[\Gamma] - \nabla\phi\big(\Re[\Gamma']\Re[\Gamma] - \Im[\Gamma']\Im[\Gamma]\big) + \frac{\phi'\nabla|\Gamma|^2}{2}\right]. \tag{6.38}$$

In the above, all gradients are with respect to the layer thicknesses, and all primes refer to k derivatives.

6.6 Algorithm Overview

Here we summarize the overall dispersion gradient algorithm. The general steps are:

Algorithm 2 *Analytic Dispersion Gradient Computation*

1. *Precompute all individual layer transfer matrices* $\mathbf{P}_\ell$ *and* $\mathbf{D}_\ell$ *and their derivatives with respect to k, .*

2. *Iteratively compute front transfer matrices* $\mathbf{T}_{(\ell,0)}$ *and their derivatives for* $n-1>\ell>1$*, as demonstrated in Section 6.2 and Ref. [66].*

3. *Compute back transfer matrices* $\mathbf{T}_{(n,\ell)}$ *and derivatives by repeating Step 2 for the reversed stack and then taking the hermitian transpose, as shown in* (6.31) *and explained in Section 6.4.3.*

4. *Loop through all layers, computing matrix gradient elements of* $\nabla\mathbf{T}$ *and* $\nabla\mathbf{T}'$ *as per the exact expression* (6.20) *or the approximate expression* (6.24)*.*

5. *Translate matrix gradients into reflectivity and group delay gradients as done in Section 6.5.*

6. *Compute total gradients from the matrix gradients found in step 5.*

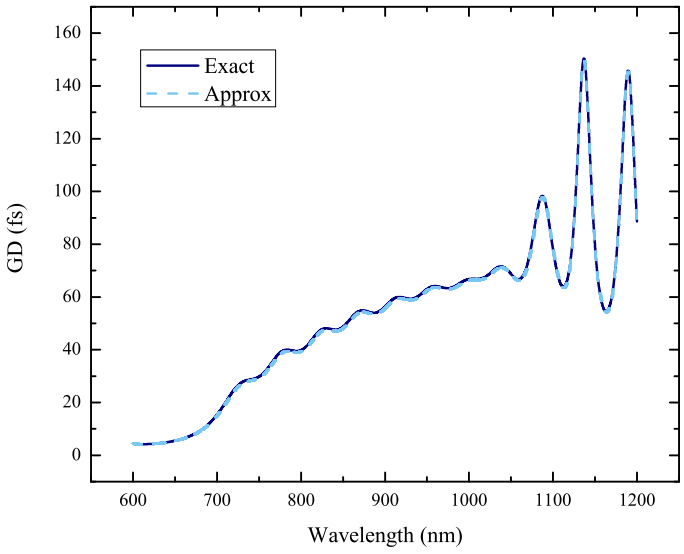

Figure 6-2: Spectral group delay of example chirped mirror. A portion of the response past the high reflectivity region (wavelengths greater than about 1050 nm) is shown to demonstrate that the approximation even holds when the group delay is rapidly varying.

6.7 Example Gradient Computation

To demonstrate the efficacy of the constant coupling approximate gradient algorithm, we have computed the gradient of group delay at a central wavelength for a typical chirped mirror. For reference, the chirped mirror spectral GD and GDD are plotted in Figures 6-2 and 6-3, respectively.

When computing the gradient at 256 wavelengths, the constant coupling gradient was empirically found to be 63% faster than the exact gradient algorithm, demonstrating that computational savings in the gradient extend from the algorithms given in Ref. [66]. We thus estimate that our GD gradient method is roughly four times faster than computing gradients using GDD.

Note that the gradient itself is exact (i.e. it is an exact analytic gradient of an approximate group delay) so using the approximation should not affect the convergence rate, but may have a small effect on the final solution. We expect that, on average, optimizations will be made correspondingly faster. Should final group delay error be so small that the constant coupling approximation error is no longer negligible, a few refinement steps can be performed at the end with the exact gradient to converge at the exact solution.

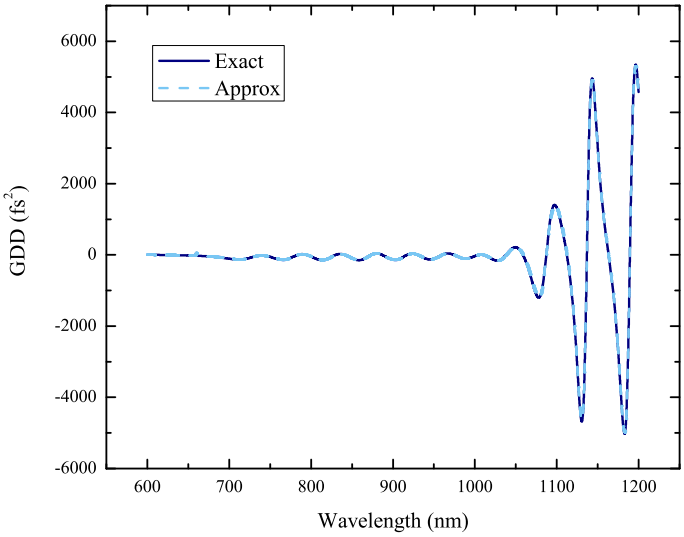

Figure 6-3: Spectral group delay dispersion of chirped mirror shown in Fig. 6-2.

6.8 MATLAB and C Code

To simplify use of these algorithms, we have provided MATLAB and C code of both the GD gradient algorithms discussed in this thesis. The C codes are written as MATLAB MEX functions to enable their use in optimization routines within MATLAB, though each function is available as a stand-alone for use in user programs. Several ancillary MATLAB functions that facilitate mirror optimization using the Optimization Toolbox are also included. The code is available in electronic form at *http://www.mit.edu/~birge/dispersion*.

Chapter 7

Robust Mirror Optimization

Currently the limiting factor to dispersion compensating mirror performance is the extent to which manufacturing errors cause the fabricated mirrors to deviate from their designed specification. In this chapter, we develop a design method which seeks to produce designs which sacrifice some amount of nominal performance for an increased tolerance to manufacturing errors. This work was done jointly with Omid Nohadani and Demetris Bertsimas.

7.1 Introduction

The dispersion compensating mirror, first proposed in 1994 [44], has become an enabling technology for modern ultrafast lasers. Solid-state mode-locked lasers can only operate at or below few-cycle pulse widths when the total cavity dispersion is reduced to nearly zero, with only a small amount (on the order of a few fs^2) of residual second-order dispersion. While prisms can be used to compensate for second- and third-order cavity dispersion, their relatively high loss and inability to compensate for arbitrary dispersion limits their use; pulse durations below ten femtoseconds were not possible directly from oscillators until the development of high performance double-chirped mirror pairs [16, 69, 17].

As bandwidths increase, so do the number of layers required to produce a mirror with the high reflectivity needed for an intra-cavity mirror. For bandwidths exceeding an octave, mirror pairs with over 200 total layers are generally required. The sensitivity of a dielectric stack to manufacturing errors increases with the number of layers, and dispersion compensating mirrors push the limits of manufacturing tolerances, requiring layer precisions on the order of a nanometer. Currently, this challenges even the best manufacturers.

While the nominal optimization of layer thickness has lead to successful design of dispersion compensating dielectric mirrors allowing dispersion and reflectivity control over nearly an octave bandwidth, in practice the performance for such complicated mirrors is limited by the manufacturing tolerances of the mirrors. Small perturbations in layer thickness not only result in suboptimal designs, but due to the nonlinear nature of mode-locking, such perturbation may completely destroy the phenomenon.

Despite the fact that manufacturing errors often limit the performance of thin-film devices [71], there has been little work on optimizing thin-film designs to mitigate the effects of errors. Some previous work in designing fault-tolerant mirrors has focused on optimizing first-order tolerances, a method readily available in commercial thin film design codes. One approach to tolerance optimization is performed by adding terms to the overall cost function (see [72], for example) of the form

$$\sum_i \sum_k w(\lambda_k) \left(\frac{\partial g(\lambda_k; \mathbf{x})}{\partial x_i} \right)^2, \tag{7.1}$$

where x_i is the thickness of the ith layer, $g(\lambda; \mathbf{x})$ is a relevant characteristic of the mirror (such as reflectivity or dispersion), and $w(\lambda)$ is a weighting vector. It can be shown that as the merit function approaches zero, the sum in (7.1) converges to the trace of the Hessian of the merit function, a second-order estimate of the robustness. However, to rely on tolerances of this form assumes that the merit function can be well approximated by a quadratic form. Given symmetry, the robust minimum and the nominal minimum of a purely quadratic function are equivalent. Thus, if robust optimization is needed, it implies that the merit function cannot be effectively modelled by a second-order series over the expected error range.

Our approach to robust optimization probes the exact merit function in a bounded space of potential thickness errors. While this results in a much more computationally involved optimization, the result is arguably more robust to significant perturbation as the full structure of the merit function is considered in a neighborhood around a nominal solution. Furthermore, the robustness is guaranteed to be equal or better than that obtained with nominal optimization, and in the case where it is equal, no sacrifice in nominal optimality will be made.

Other prior work was done by Yakovlev and Tempea [73], who employed stochastic global optimization to achieve robustness of the final solution by virtue of the fact that they optimized a Monte Carlo computed integral over a neighborhood around a nominal design. Their method does not suffer the limitations of first-order tolerances, and was able to produce mirrors with significant improvement over nominally optimized designs, demonstrating conclusively that robustness can be greatly improved at the design level by proper optimization.

Ben-Tal and Nemirovski provided a first robust optimization approach based on an application in antenna design [74]. Recent works have been devoted to problems with convex objectives and constraints (e.g. linear) [75, 76, 77]. These works have shown that a convex optimization problem with parameter uncertainty can be transformed to another convex optimization problem. Despite significant advancements, all these results are limited to convex problems. But modern engineering design often involves problems with objectives and constraints that are not explicitly given and highly nonconvex. Thus, no internal structure can be exploited.

We present a new deterministic robust optimization method that provides for designs which are intrinsically protected against potentially significant layer thickness perturbations occurring during manufacturing. The presented method is generic and can be applied to many problems that are solved through numerical simulations. Here, we

introduce the algorithm specifically for double chirped mirrors and tailor the parameters to this particular problem. First, we discuss the optical properties of these mirrors and define a cost function based on reflectivity and group delay. We continue with the introduction of the concept of the uncertainty set as well as a novel method to identify worst-case designs within this set. Once these configurations are found, we show how an update direction can be found which eliminates these worst cases. Furthermore, we demonstrate the performance of the nominal and robust solutions for a large range of perturbation and propose a technique to increase the manufacturing yield.

7.2 Computation of Cost Function

The merit function for a chirped mirror is typically composed of two terms, one representing the performance of the reflectivity (which is ideally one) and another which quantifies the deviation of the dispersion from ideal. We employ a cost or merit function than is given as

$$f(\mathbf{x}) = \sum_k w_r(\lambda_k) \left[R(\lambda_k;\mathbf{x}) - 1\right]^4 + \sum_k w_d(\lambda_k) \left[\tau_g(\lambda_k;\mathbf{x}) - \hat{\tau}_g(\lambda_k) + \tau_0(\mathbf{x})\right]^2, \tag{7.2}$$

where $R(\lambda;\mathbf{x})$ is the wavelength domain reflectivity of the total mirror pair described by layer thicknesses $\mathbf{x}$, $\tau_g(\lambda;\mathbf{x})$ is the group delay (GD) of the pair, $\hat{\tau}_g(\lambda)$ is the ideal GD, and the $w_{r,d}(\lambda)$ are weighting functions. To account for an irrelevant offset between the computed and ideal group delay curves, we include a constant offset, $\tau_0(\mathbf{x})$, that minimizes the error. For the reflectivity errors, we use the fourth power of the error to approximate a Chebychev norm, though a standard squared error can also be used.

The computation of reflectivity from a thin-film stack is done using transfer matrix methods [68]. In a standard nominal optimization, the merit function and its gradient must be evaluated thousands of times over hundreds of wavelengths. In a robust optimization, the computational burden is even greater, with the merit function typically computed on the order of a million times. Any discrepancy in the gradient will hinder the convergence rate. Thus, it is imperative that the merit function be computed efficiently and accurately. We employ the methods described in [66, 49], where the group delay is computed in an approximate analytic form that allows for a significant reduction in computational complexity. The approximation simply neglects the local change in wavelength of the Fresnel reflections between each layer. For chirped mirrors, the approximation error is negligible, as demonstrated and explained in [66]. The gradient of the group delay is computed analytically in a self-consistent manner with the approximation, resulting in an optimization that converges quickly, both in terms of iterations and total processing time.

7.3 Problem Statement

Our design problem consists of a double chirped mirror pair with 208 layers for use in a few-cycle Titanium:sapphire mode-locked laser [61]. The initial design was computed

using the analytic method of [17]. The materials used were SiO_2 and TaO_5, with the dispersion of each modelled using Sellmeier coefficients obtained from fits to manufacturer's index data. The total reflection dispersion of the pair is specified to compensate for 2.2 mm of Titanium:sapphire, 2 meters of air, and 8 mm of Barium Fluoride in a cavity containing six mirrors. The group delay and reflectivity are optimized over 156 wavelengths, uniformly spaced from 650 to 1200 nanometers. This discretization was empirically found to be sufficient to avoid narrow resonances "leaking" through the grid. The angle of incidence is taken to be six degrees, and the polarization is assumed to be transverse in the magnetic field (TM). The reflectivity and group delay are optimized as in Eq. (7.2), with constant weightings $w_r = 1$ and $w_d = 10^{-8}\,\mathrm{fs}^{-2}$.

7.4 Implementation Errors

7.4.1 General Model

For the purposes of developing this algorithm, we model manufacturing errors as independent random sources of additive noise, since any known systematic errors, such as miscalibration, can be best addressed in the actual production. As empirically supported, the layer-thickness errors can be regarded as not correlated. Therefore, we assume that when manufacturing a mirror with layer thicknesses given by $\mathbf{x}$, statistically independent additive implementation errors $\Delta\mathbf{x} \in \mathbb{R}^n$ may be introduced due to variation in the coating process, resulting in an actual thicknesses $\mathbf{x}+\Delta\mathbf{x}$. We assume a mean of zero and a variance on each layer that is motivated by actual manufacturing errors. Here, $\Delta\mathbf{x}$ resides within an uncertainty set

$$\mathscr{U} \quad := \quad \left\{\Delta\mathbf{x} \in \mathbb{R}^n \mid \|\Delta\mathbf{x}\|_2 \leq \Gamma\right\}. \tag{7.3}$$

Note that $\Gamma > 0$ is a scalar describing the size of perturbation against which the design needs to be protected. For this paper, we took the manufacturing uncertainty to be normally distributed with a standard deviation of $\sigma = 0.5$ nm. To maintain 95% cumulative confidence to capture all errors within $\mathscr{U}$ for this 208-dimensional problem, we chose $\Gamma = 0.0075\ \mu$m. We seek a robust design $\mathbf{x}$ by minimizing the *worst case cost*

$$g(\mathbf{x}) \quad := \quad \max_{\Delta\mathbf{x}\in\mathscr{U}} f(\mathbf{x}+\Delta\mathbf{x}). \tag{7.4}$$

The worst case cost $g(\mathbf{x})$ is the maximum possible cost of implementing $\mathbf{x}$ due to an error $\Delta\mathbf{x} \in \mathscr{U}$. Thus, the *robust optimization problem* is given through

$$\min_{\mathbf{x}} \quad g(\mathbf{x}) \quad \equiv \quad \min_{\mathbf{x}} \quad \max_{\Delta\mathbf{x}\in\mathscr{U}} \quad f(\mathbf{x}+\Delta\mathbf{x}). \tag{7.5}$$

In other words, the robust optimization method seeks to minimize the worst case cost. When implementing a certain design $\mathbf{x} = \hat{\mathbf{x}}$, the possible realization due to implementation errors $\Delta\mathbf{x} \in \mathscr{U}$ lies in the set

$$\mathscr{N} \quad := \quad \{\mathbf{x} \mid \|\mathbf{x}-\hat{\mathbf{x}}\|_2 \leq \Gamma\}. \tag{7.6}$$

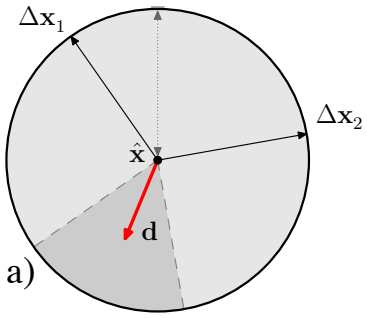

Figure 7-1: A two-dimensional illustration of the neighborhood. For a design $\hat{\mathbf{x}}$, all possible implementation errors $\Delta\mathbf{x} \in \mathcal{U}$ are contained in the shaded circle. The bold arrow $\mathbf{d}$ shows a possible descent direction and thin arrows $\Delta\mathbf{x}_i^*$ represent worst errors.

We call $\mathcal{N}$ the *neighborhood* of $\hat{\mathbf{x}}$; such a neighborhood is illustrated in Fig. 7-1. A design $\mathbf{x}$ is a *neighbor* of $\hat{\mathbf{x}}$ if it is in $\mathcal{N}$. Therefore, $g(\hat{\mathbf{x}})$, is the maximum cost attained within $\mathcal{N}$. Let $\Delta\mathbf{x}^*$ be one of the worst implementation error at $\hat{\mathbf{x}}$, $\Delta\mathbf{x}^* = \arg\max_{\Delta\mathbf{x}\in\mathcal{U}} f(\hat{\mathbf{x}}+\Delta\mathbf{x})$. Then, $g(\hat{\mathbf{x}})$ is given by $f(\hat{\mathbf{x}}+\Delta\mathbf{x}^*)$.

Since we seek to navigate away from all the worst implementation errors, the inner maximization problem needs to be solved first. Given that f is nonconvex and provided through numerical calculations, we cannot exploit any possible internal structure to compute g. Therefore, we conduct local searches to determine worst configurations within $\mathcal{N}$.

Previously, it was shown that all worst-case scenarios reside on the shell of $\mathcal{N}$ [76]. Thus, to improve the speed of the inner maximization, we can restrict ourselves to only considering error vectors $\Delta\mathbf{x}$ such that the $\|\Delta\mathbf{x}\|_2 = \Gamma$. Problem (7.5) then transforms into a constrained maximization over the shell $\|\Delta\mathbf{x}\|_2 = \Gamma$, which makes the search computationally more efficient.

7.4.2 Restricted Search Space

To protect a design against errors, it is helpful to utilize available understanding of possible errors. For example, if there are worst-case scenarios in the respective neighborhood that are very rare according to our assumed layer perturbation distribution, there is no need for them to be considered during the inner maximization problem (7.4). By excluding these rare events from $\mathcal{U}$, we are able protect the design against realistic and statistically relevant errors only, without needlessly sacrificing nominal performance to guard against rare errors. Moreover, this approach leads to a reduction of the size of the respective search space and, thus, to an increase of the computational efficiency.

It is well known that the reflection coefficients of thin-film stacks are closely related to the Fourier transform of the layer thicknesses [78]. Thus, one promising class of rare perturbations to eliminate from consideration are those which have strong correlations between the layers. These errors involve, for example, shifting of all the thicknesses in one direction, which results in a spectral shift regardless of the design. Even though

such errors may occur in actual manufacturing due to systematic issues, there is little or nothing that can be done to deal with them by design optimization, and to attempt to do so will only result in a highly compromised design. We thus restrict ourselves to considering only statistically independent random perturbations to the layers. In this context, the probability of errors occurring with high correlation between the layers is negligible, and thus we should not concern ourselves with protecting against them. We therefore seek a class of errors which restrict the allowable correlation between layers, i.e., we restrict the maximum variation in the amplitude of the Fourier components of the error vector. A straightforward way to do this, is to restrict the search to the class of error vectors with minimum coherence, requiring all Fourier components to have a uniform amplitude.

In addition to the above, this choice of subset is justified empirically. Monte Carlo simulations reveal that the set of perturbations with uniform amplitude in the Fourier domain with uniformly distributed phases have virtually identical statistics to the general uncertainty set $\mathscr{U}$ defined in Eq. (7.3). The cumulative probability distribution of the reduced set never exceeds the full set by more than 4%. This confirms that our worst case search over the reduced subset will not miss anything statistically relevant in the full set, and thus robustness is not compromised by using this set.

In the restricted space, the components of $\Delta\mathbf{x}$ can be written as

$$\Delta x_j = \frac{\Gamma}{\lfloor N/2 \rfloor} \sum_{k=1}^{\lfloor N/2 \rfloor} \cos\left(\frac{2\pi k j}{N} + \phi_k\right), \tag{7.7}$$

where ϕ_k is the phase of the kth Fourier component of $\Delta\mathbf{x}$ and N is the number of layers. We furthermore assume the constant (zero frequency) component is zero, which corresponds to the aforementioned pathological case of all layers shifting a similar amount. Using *Parseval's theorem*, i.e., the sum of the square of a function is equal to the sum of the square of its transform, we can verify that the magnitude of the errors remains on the shell of the original uncertainty set $\mathscr{U}$,

$$\|\Delta\mathbf{x}\|_2^2 = \sum_{k=1}^{N} |\Delta x_k|^2 = \Gamma^2. \tag{7.8}$$

Using this transformation, we search over the phases ϕ_k for worst-case neighbors. Therefore, the search space dimensionality is reduced to $\lfloor N/2 \rfloor$, hence, the efficiency of this algorithm increases by N^2. Most importantly, since the maximization problem is over the free phase-space on the shell and the magnitude of these vectors are constant, the advantages of an unconstrained search can be exploited. Consequently, we obtain the set of local maxima in the phase space using standard gradient-based optimization. Furthermore, we obtain a set of true bad neighbors, which is significantly smaller in size ($\ll 500$) than had we left the search space more general. Since this size determines the number of constraints in the Problem, we experience a significant speed up in this part of the algorithm as well.

7.5 Robust Optimization

Once worst-case neighbors are identified, a direction is sought along which an updated neighborhood would not include these worst-case scenarios any longer. This direction is a vector that spans the largest angle $\Theta \geq 90^\circ$ to all worst implementation errors at $\hat{\mathbf{x}}$ in the *set of worst implementation errors*

$$\mathscr{U}^*(\hat{\mathbf{x}}) \quad := \quad \left\{ \Delta\mathbf{x}^* \mid \Delta\mathbf{x}^* = \arg\max_{\Delta\mathbf{x}\in\mathscr{U}} f(\hat{\mathbf{x}} + \Delta\mathbf{x}) \right\}. \tag{7.9}$$

To navigate away from the elements in $\mathscr{U}^*(\hat{\mathbf{x}})$, a descent direction $\mathbf{d}^*$ can be found efficiently by solving the following second-order cone problem (SOCP):

$$\begin{array}{ll} \underset{\mathbf{d},\beta}{\text{minimize}} & \beta \\ \text{subject to} & \|\mathbf{d}\|_2 \leq 1 \\ & \mathbf{d}\cdot\Delta\mathbf{x}^* \leq \beta \qquad \forall \Delta\mathbf{x}^* \in \mathscr{U}^*(\hat{\mathbf{x}}) \\ & \beta \leq -\epsilon, \end{array} \tag{7.10}$$

where ϵ is a small positive scalar, β is an auxiliary variable, and $||\cdot||_2$ is the Euclidean norm. A feasible solution to Problem (7.10), $\mathbf{d}^*$, forms the maximum possible angle $\theta_{\max}$ with all $\Delta\mathbf{x}^*$, as illustrated in Fig. 7-1. This angle is always greater than 90° due to the constraint $\beta \leq -\epsilon < 0$. This constraint guarantees that d^* will provide an updated design neighborhood that excludes all known $\Delta\mathbf{x}^*$. The value of ϵ is chosen heuristically such that when Problem (7.10) is infeasible, then $\hat{\mathbf{x}}$ is a robust local minimum. Note, that the constraint $\|\mathbf{d}^*\|_2 = 1$ is automatically satisfied if the problem is feasible. Such an SOCP can be solved efficiently using both commercial and noncommercial solvers. Because $\Delta\mathbf{x}^*$ usually reside among designs with nominal costs higher than the rest of the neighborhood, the following algorithm summarizes a heuristic strategy for the robust local search: [79]

Algorithm 3

Step 0. Initialization: Let $\mathbf{x}^1$ *be an arbitrarily chosen initial decision vector. Set* $k := 1$.

Step 1. <u>*Neighborhood Exploration*</u> *:*

Find a set of implementation errors $\Delta\mathbf{x}_i$ *with the highest cost within the neighborhood of* $\mathbf{x}^k$*. For this, we conduct multiple unconstrained maximization searches over the shell of the uncertainty set starting from random initial configurations. The results of all function evaluations* $\big(\mathbf{x}, f(\mathbf{x})\big)$ *are recorded in a bad-neighbors set.*

Step 2. <u>*Robust Local Move*</u> *:*

(i) Solve the SOCP (7.10); terminate if the problem is infeasible.

(ii) Set $\mathbf{x}^{k+1} := \mathbf{x}^k + t^k\mathbf{d}^*$*, where* $\mathbf{d}^*$ *is the optimal solution to the SOCP.*

(iii) Set $k := k+1$*. Go to Step 1.*

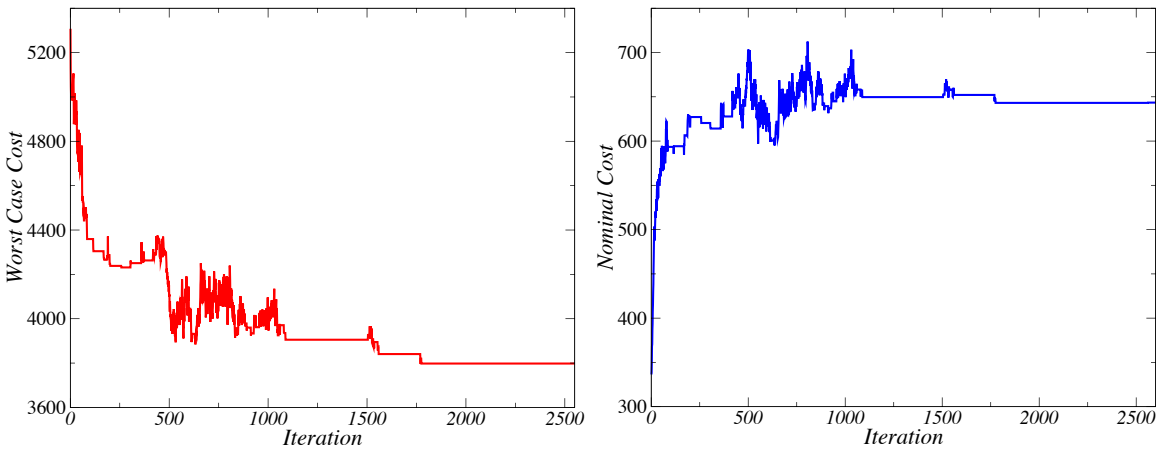

Figure 7-2: The robust optimization algorithm improves (left) the worst-case cost in the neighborhood of the current design. Discoveries of new bad neighbors cause the small peaks. (Right) The Price of Robustness is an increase in the nominal cost.

The step size t^k is computed as the shortest step size that eliminates all bad neighbors from $\mathbf{x}^k$. Reference [79] provides a detailed discussion on the actual implementation.

7.6 Results

Starting from a nominally optimized solution, our robust optimization algorithm successively decreased the worst-case cost, as in Eq. (7.4). This performance is shown in Fig. 7-2. The significant improvement of robustness comes at the price of a small increase in the nominal cost. The algorithm converges to the robust local minimum, at which point no descent direction can be found.

The reflectivity and group delay of the robust and nominal optimum are shown in Fig. 7-3. While both solution satisfy the design objectives, the robust design is significantly more protected against possible errors. The unavoidable "Price of Robustness" through a decrease on the nominal performance of the robust solution is apparent, with increased ripple in the group delay and reflectivity. This price is especially apparent in the bottom plot of Fig. 7-3, which compares the total group delay error for the robust and nominally optimized mirror pairs. However, as will be shown in Figs. 7-5 and 7-6, the robust solution performs better when the layer perturbations are taken into account. Even though the nominally optimized design is able to achieve GD errors of less than one femtosecond, it turns out that the half nanometer layer perturbations we took as our assumed manufacturing tolerances result in GD errors on the order of plus or minus five In a sense

In Fig. 7-4, we show the layer thicknesses for the mirror pair both after nominal optimization and robust optimization. The general structure of the mirror is preserved in the robust optimum solution, in keeping with the observation that its nominal performance is not degraded significantly. The larger variations are found in the first several

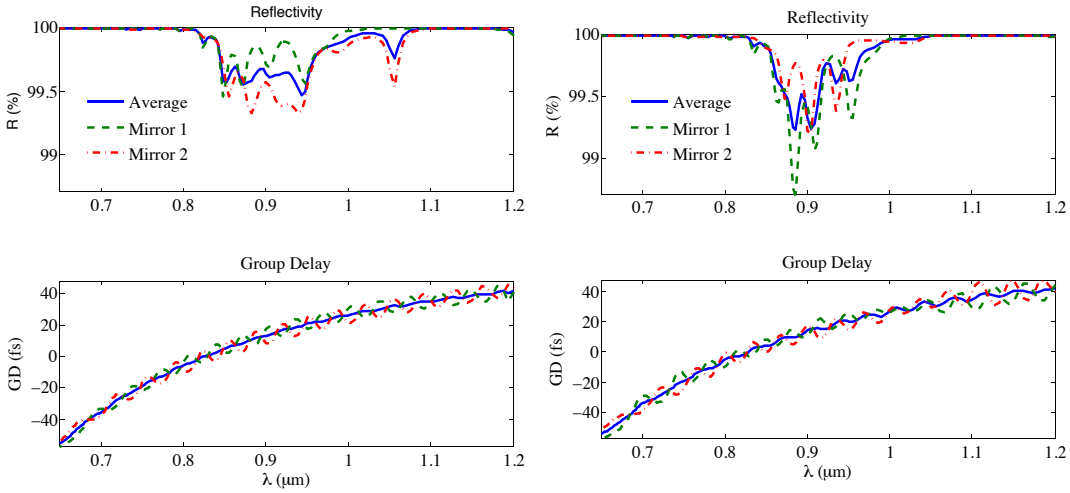

Figure 7-3: Reflectivity and group delay for each chirped mirror in the pair: (Left) nominally optimal design; (Right) robustly optimal configuration.

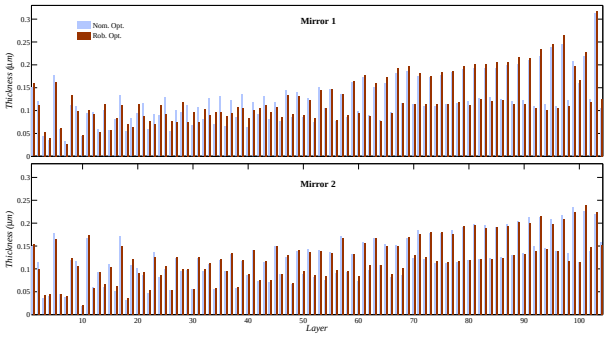

Figure 7-4: Layer thicknesses of nominal optimum and robust optimum of the mirror pair.

layers, which perform impedance matching into the chirped stack, suggesting that they are the most sensitive to perturbation. This is consistent with the fact that any spurious reflections off of the front surface of the mirror will significantly degrade the GD performance.

While we intended to match the size of the uncertainty to the reported manufacturing and measurement errors, the value of Γ might not fully reflect the actual errors. Therefore, our algorithm seeks to find robust solutions with stable performance even beyond predicted errors. To illustrate these effects, we varied the size of the uncertainty set and evaluated the worst possible neighbor within this neighborhood. The worst-case scenarios of the nominal optimum and robust optimum, both in cost as well as the optical properties are compared for increasing neighborhood size in Fig. 7-5.

The worst-case performance of both the nominal and robust designs behave fairly similarly within a small range of perturbations, which is in fact comparable to Γ. However, once the size of possible errors increases, the worst-case cost of the nominal design drastically rises, showing that this design would lose its phenomena completely.

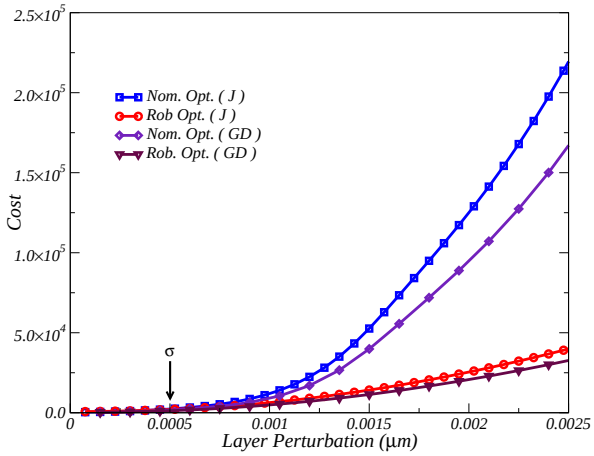

Figure 7-5: Comparison of worst case cost and worst case GD cost of two designs, the nominal and robust optimum, for increasing size of possible perturbations or errors.

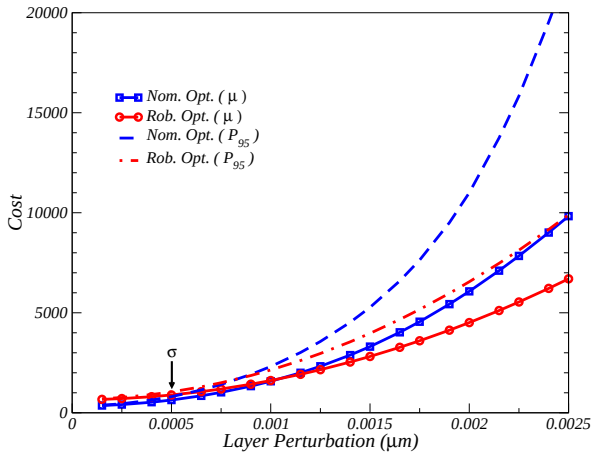

Figure 7-6: Comparison of the nominal and robust design: mean and 95th percentile of the cost distribution of 10^6 randomly sampled designs for varying perturbation sizes.

Since any manufacturing process is to some extend statistical, it is essential for a design to yield a high manufacturing yield. Our robust optimization method not only minimizes the worst-case performance, but also addresses these statistical effects. This is demonstrated in Fig. 7-6. A series of Monte Carlo simulations, each with 10^6 randomly sampled designs with normally distributed layer perturbations, were performed with varying standard deviations at the robust and the nominal optimum. The mean μ and the 95th percentile P_{95} of the distribution for each perturbation size are plotted to illustrate the center and the actual width of this statistical process. While both designs are similarly distributed within the expected errors σ, they deviate significantly beyond this mark. In fact, the mean and more importantly the spread of the distribution for the nominal optimum design increases rapidly beyond σ, while the robust optimum is more moderate. Moreover, the mean of the nominal optimum at all perturbation sizes is within the distribution (P_{95}) of the robust optimum, demonstrating that the manufacturing yield of the robust solution remains high and provides performances comparable

to the nominal design, even beyond the assumed errors. Since the notion of the actual manufacturing errors are often somewhat uncertain, our method can provide a robust solution despite these uncertainties.

7.7 Summary

We have developed a new robust optimization technique specifically tailored to the problem of thin-film filter optimization. Our method obtains robust solutions by performing a series of deterministic gradient ascent searches around a given trial solution for worst-case errors. To avoid taking into account extremely rare potential errors, we perform this search over the space of all errors on the shell of our neighborhood whose components are minimally coherent. This avoids taking into account rare but highly significant errors, such as those associated with certain types of systematic manufacturing errors, that would otherwise dominate the optimization. This modification allows an unconstrained inner maximization over a reduced search space, thus, improves the efficiency. Once, a set of worst-case designs are identified within an uncertainty set, our method provides an updated design, that has reduced worst-case performance. After a number of iterations, we obtain a robust optimum that has the lowest worst-case performance.

We apply the method to a demanding optimization of a 208 layer chirped mirror pair with nearly an octave of bandwidth. The robust solution is compared with that obtained using standard optimization techniques, and is found to achieve improved statistical performance for layer errors of half a nanometer. Furthermore, the fault tolerance of the robust solution increases significantly relative to the nominally optimized mirror as the error variance increases, demonstrating that the robust solution is not tied to the particular manufacturing error variance assumed during optimization (see Fig. 7-6). Therefore, our robust design warrants for a high manufacturing yield even when errors occur that are larger than originally assumed.

In this initial demonstration, we performed the optimization on a fixed number of layers. However, the robust optimization problem can be viewed as providing a new cost function which takes into account robustness and, thus, can be used within other refinement algorithms, such as needle optimization [80], that allow for changing layer counts.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 8

Systematic Errors in Spectral Shearing Methods for Few-Cycle Pulses

One of the prominent characterization methods for very short pulses is spectral phase interferometry for direct electric-field reconstruction (SPIDER). Due to its unique bandwidth advantages and noise properties, spectral shearing interferometry is well suited to the measurement of few- and single-cycle pulses. However, there is a calibration sensitivity with standard SPIDER (and many of its variants) that needs to be taken into consideration for pulses on the order of about two-cycles and below.

In this chapter, we begin by providing a brief overview of spectral shearing, including a derivation showing that SPIDER involves a calibration whose error effects scale with the square of the bandwidth. We then discuss experimental techniques to mitigate these effects, and conclude with a survey of various alternative approaches to spectral shearing which avoid some or all of the issues discussed.

8.1 Introduction

Steady progress in ultrafast laser sources over the past several decades has led to the recent development of robust sources of few-cycle laser pulses. Sub-two cycle pulses can now be produced directly from oscillators, and sources of single-cycle pulses are under development [60, 81]. Furthermore, as few-cycle lasers are increasingly used to drive attosecond extreme-UV and X-ray pulses, these applications will require extremely accurate and precise characterization of the few- and single-cycle pulse envelopes used to drive the high harmonic generation process [82].

The technology for measuring ultrashort optical pulses must, of course, keep pace with the lasers themselves, and few- and single-cycle pulses present unique difficulties in this regard. The most obvious difficultly in few-cycle pulse measurement stems from the tremendous bandwidths involved. All self-referenced pulse characterization methods involve nonlinear operations of some sort [83], and in the case of few-cycle pulses one essentially has to implement a specialized analog optical switch capable of operating with hundreds of terahertz of bandwidth. Any bandwidth filtering is especially relevant for techniques where the *amplitude* of the trace is a critical parameter, such as

interferometric autocorrelation (IAC) [84] and frequency resolved optical gating (FROG) [85].

As pulses become shorter, the time scales of nonidealities do not always scale with them. For example, the relative delays of satellite pulses due to secondary reflections off dispersion compensating mirrors do not scale with the main pulse width. In fact, if anything they scale inversely to the pulse width as mirrors become thicker to accommodate wider spectral range. The spectral phase oscillation periods caused by such delays are usually on the order of 5–10 THz. Thus, as bandwidths approach 200 THz and beyond, the time-bandwidth product required for a full characterization on even a well-compressed pulse can exceed 20.

In addition, the extreme bandwidths involved result in higher-order material dispersion playing a significant role in pulse shaping, yielding pulses which are typically highly asymmetric. As such, the commonly used technique of IAC, which is relatively insensitive to pulse asymmetries [86], can miss details. The squared intensity operation inherent to IAC greatly suppresses the effect of satellite pulses, for instance. As such, iterative reconstructions based on IAC (e.g. [84]) may fail to properly converge in the presence of noise when higher-order dispersive effects are significant.

As one of the dominant characterization methods for very short pulses is spectral shearing inteferometry, typified by SPIDER. In many ways, SPIDER is uniquely suited to the task of measuring few- and single-cycle pulses. However, there is a calibration sensitivity with standard SPIDER (and many of its variants) that needs to be taken into consideration for pulses on the order of a single cycle.

We begin by briefly explaining the principle behind spectral shearing and its advantages in the few-cycle regime. Due to its importance to both accuracy and sensitivity, we discuss the importance of choosing a proper shear frequency, and the considerations for doing so. Next, we analyze the sensitivity of the measurement to the delay reference inherent to spectral shearing, deriving an analytic result and simple scaling law. We show that relative measurement errors scale quadratically with pulse bandwidth, leading to extreme sensitivity to calibration errors as pulse widths decrease past a few optical cycles. We then illustrate this principle on experimental data from a sub-two-cycle oscillator, showing that a standard SPIDER apparatus would require mechanical stability far exceeding the scale of the wavelengths involved. Finally, we conclude by suggesting ways to mitigate the sensitivity, and survey some recent alternative spectral shearing methods which attempt to eliminate it.

Our focus on spectral shearing is not meant to imply that we feel it is alone in presenting challenges to the characterization of pulses approaching a single cycle. In fact, our feeling is that spectral shearing is otherwise so well-suited to handling such pulses that it is worth examining these issues in some detail.

8.2 Spectral Shearing Interferometry

8.2.1 Theory

Spectral phase interferometry for direct electric-field reconstruction, or SPIDER, was developed by Iaconis and Walmsley in 1999[18]. It is a modification of spectral interferometry, where both interfering components are obtained from the input pulse and are slightly shifted in frequency. This obviates the need for the reference pulse that is required in standard spectral interferometry. Called spectral shearing interferometry, this general idea represents a fundamentally unique mode of pulse measurement, in that it directly observes interference between two adjacent frequency components. As such, it can be an extremely robust and direct method that avoids the need for iterative inversion algorithms.

Figure 8-1 shows a conceptual schematic of SPIDER. While the practical details vary greatly, all spectral shearing methods involve a nonlinear interferometer whereby one arm experiences a different frequency shift than the other. Interference is observed between the two by putting some sort of phase shift on one of the arms. In the case of standard SPIDER, a linear phase shift is realized by interferometrically delaying the two pulse copies by τ.

By upconverting the pulse to be measured with a pair of frequencies separated by Ω, any original wavelengths in the pulse which are separated by this "shear" frequency are mapped to the same wavelength in the output. One may thus directly observe the phase delay between two nearby wavelengths, and thereby the spectral dispersion. In standard SPIDER, the spectrally resolved output is given by [18]

$$D(\omega) = |E(\omega-\Omega)|^2 + |E(\omega)|^2 + 2\,|E(\omega-\Omega)E(\omega)|\cos[\underbrace{\phi(\omega-\Omega)-\phi(\omega)}_{\phi'(\omega-\Omega/2)\Omega+O[\Omega^2]}+\tau\omega], \tag{8.1}$$

where $E(\omega) = |E(\omega)|e^{-i\phi(\omega)}$ is the upconverted spectrum of the pulse and τ is the delay between the two upconverted copies. (In practice $E(\omega)$ will be roughly doubled in frequency from the fundamental, but its phase will be identical and its phase is all we are concerned with.) The oscillating cosine "carrier" fringe is the only element of interest, and its phase encodes a finite difference of the pulse spectral phase, approximately proportional to the spectral group delay. The method for isolating this phase is beyond the scope of this paper, and is not important to our end here. It suffices to mention that so long as τ is sufficiently large, the fringe phase may be reliably extracted by standard signal processing techniques. Since the initial development of SPIDER, many variants of spectral shearing interferometry have been invented, but all share the same fundamental property of generating a carrier fringe in some domain (perhaps in space or time, if not frequency) which is shifted in proportion to the finite difference of the spectral phase, $\phi(\omega-\Omega)-\phi(\omega)$. As such, spectral shearing interferometry essentially "samples" the spectral phase (up to a constant phase) with a discretization of Ω.

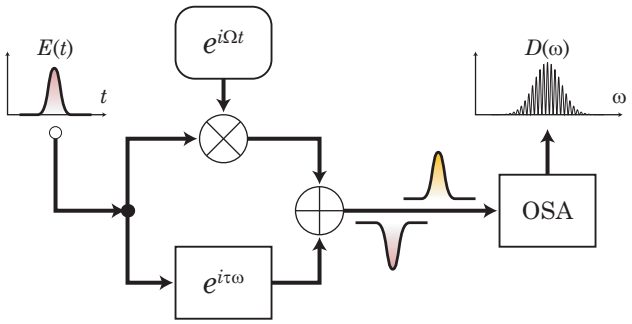

Figure 8-1: Conceptual schematic of SPIDER. This model intentionally abstracts away several practical details, such as the fact that, in practice, both pulse copies are upconverted by slightly different optical frequencies. However, these details are not important to an analysis of SPIDER.

8.2.2 Application to Few-cycle Pulses

The spectral shearing approach has three salient features relevant in the context of few-cycle pulses. First, the upconversion of the short pulse with a monochromatic field is fundamentally easier to perform than the full second-harmonic operation inherent to IAC and FROG. To begin with, the output relative bandwidth is roughly half that of the input as the spectrum is simply shifted, as the spectra is not convolved with itself as in the case with second-harmonic-based methods. Most importantly, however, only one of the field components in the nonlinear operation contains the full bandwidth, which greatly facilitates phase matching; Type II upconversion can be engineered to have significant bandwidth in one of the input fields with a narrow bandwidth in the other, a perfect match for spectral shearing interferometry (see [87] for an illustration of this).

A second advantage comes from the use of phase to encode the spectral group delay. The spectral signal (8.1) produced by SPIDER can be viewed as a carrier wave that is frequency modulated by a signal proportional to the spectral group delay. Much as frequency modulation is more robust to interference than amplitude modulation for a given signal power [88], this modulation scheme renders SPIDER methods relatively impervious to phase matching bandwidth effects, as well as highly immune to experimental noise. This noise tolerance was observed by Gallman and others in [89], and by Jensen in [90]. Robustness to noise is especially important given the relatively low efficiency of most spectral shearing embodiments, wherein much of the measured light is thrown away to create the chirped upconverting signal. The phase modulation scheme also makes spectral shearing tolerant to the presence of unwanted signals (such as the fundamental pulse or higher diffraction orders from a grating spectrometer), which become increasingly difficult to suppress as bandwidths exceed an octave.

Lastly, spectral shearing directly measures spectral phase rather than the effects of it on the pulse envelope. Together with the aforementioned noise immunity of the encoding scheme, this makes spectral shearing methods extremely sensitive to the kind of pulse asymmetries and secondary pulses that are common in few-cycle lasers.

However, one issue that the original SPIDER does share with its correlation-based cousins (e.g. IAC) is that it requires the measured pulse to be split and delayed. For

pulses approaching an octave of bandwidth, it is not yet possible to implement a dispersionless beamsplitter and the dispersion of the beamsplitter is imprinted on the SPIDER measurement. Fortunately, this is not fundamental to spectral shearing; a few methods, to be discussed later, have been developed which involve nothing but reflections for the measured pulse.

8.2.3 Choice of Shear Frequency

The shear frequency Ω plays a critical role in both determining the sensitivity, as well as the accuracy of a SPIDER measurement. Since Ω determines the frequency spacing at which we concatenate the spectral phase, this ratio is equal to the number of points at which we sample our spectrum over its bandwidth. According to the Shannon sampling theorem, the temporal window which we can handle without aliasing is the reciprocal of twice the shear. The time-bandwidth product that can be resolved is therefore $\Delta\omega/2\Omega$.

Since the modulation of the SPIDER fringe (8.1) is proportional to the shear, one maximizes the signal to noise ratio of a spectral shearing measurement by choosing the largest shear that will avoid aliasing. As will be shown later, this also results in the least sensitivity to calibration errors. However, how can one determine, a priori, what that is? In theory, there is no way to know without actually making a measurement. Normally when one samples a signal, the bandwidth is known. However, in this case the "bandwidth" is the temporal extent of the pulse, and there is no reliable way to know that without having already done a pulse characterization. In practice, however, one generally knows the range of dispersion expected. Furthermore, one can fairly assume that structures in the power spectral density will coincide with oscillations in the spectral phase. In most cases, especially with few-cycle pulses, it is the latter that determines the required spectral sampling resolution. Thus, picking a shear that is sufficient to resolve the features of the amplitude spectrum will usually suffice. If in doubt, a sequence of shears can be used to effectively verify sufficient sampling.

Most lasers produce pulses with satellite structures and pedestals, to some extend or another. As such, in a properly performed measurement of a well-compressed pulse, with the spectral features sufficiently sampled, the vast majority of the energy is contained in a relatively small region of the resolvable temporal window. Given such a result, it is tempting for the user to assume that a larger shear can be safely used, with the argument being that if the power outside the main pulse region is negligible, it will not hurt to ignore its effects when aliased. However, what may appear negligible in intensity when well-separated may have a significant effect when added coherently to the main pulse. Consider the case of a measurement where a pulse has distant satellite pulses that are no more than 1% of the intensity of the main pulse. Increasing the shear enough to alias the pedestal onto the main pulse can potentially result in 20% relative changes in the main pulse on an intensity basis. In general, the only way to verify sufficient sampling is to take another measurement at a different resolution.

8.3 Spectral Shearing Delay Sensitivity

8.3.1 Pulse Width Error Scaling

Unfortunately, the advantages enumerated above come at a certain cost. In order to gauge the phase of the fringe in (8.1) we must know the nominal period of the fringe, given by $1/\tau$. Of course, all ultrafast measurement techniques contain inherent length references which must be calibrated (such as the distance traveled by a delay stage in an autocorrelation or the spectrometer grating period in any spectrally resolved method). In most methods, the calibrations affect the measurement in a proportional way. However, it turns out that an error $\delta\tau$ in the interpulse delay τ will result in an *additive* measurement error, and thus this calibration becomes increasingly sensitive as pulses become shorter.

In this section we derive a rough scaling law for the worst case error δt in the measured pulse width as a function of $\delta\tau$ by considering the characterization of a Gaussian pulse with a spectral $1/e^2$ half-width of $\Delta\omega$ (a figure which we use for mathematical simplicity, and which is within ten percent of the commonly used full-width-half-maximum, or FWHM). We assume that the pulse we're measuring is dispersed solely by an amount of second-order dispersion given by ϕ'', and we are concerned with the error in estimating the pulse width in the presence of a given uncertainly $\delta\tau$ in the interpulse delay. The complex spectrum of the pulse is given by

$$E(\omega) = e^{-(\omega/\Delta\omega)^2 + \frac{1}{2}i\phi''\omega^2}. \tag{8.2}$$

From (8.1) we can see that any unaccounted delay $\delta\tau$ occurring between the two upconverted pulses will be associated, to first-order, with the group delay and thus be interpreted by the reconstruction as an erroneous linear group delay,

$$\delta\phi'(\omega) = \frac{\delta\tau\,\omega}{\Omega}. \tag{8.3}$$

Taking the derivative of both sides of this equation with respect to ω gives us an expression for the erroneous dispersion contributing to the measurement as a result of the delay error,

$$\delta\phi'' = \frac{\delta\tau}{\Omega}. \tag{8.4}$$

To consider the effect of this extra dispersion we begin with the well-known result (see [91], for example) for the temporal width (where we have translated the formula so that it is in terms of $1/e^2$ width) of a pulse broadened by second-order dispersion,

$$T = T_0\sqrt{1 + \left(\frac{\phi''\Delta\omega}{2}\right)^2}, \tag{8.5}$$

with T and T_0 the dispersed and Fourier limited widths, respectively. The actual measured pulse width, $T + \delta T$, can be written by replacing the dispersion in (8.5) with the measured dispersion $\delta\phi'' + \phi''$. Using the expression for $\delta\phi''$ given in (8.4), this means

the pulse width actually measured is simply

$$T + \delta T = T_0 \sqrt{1 + \left(\frac{\delta\tau \Delta\omega}{2\Omega} + \frac{\phi'' \Delta\omega}{2} \right)^2}. \tag{8.6}$$

We consider the scenario where we erroneously measure a chirped pulse to be shorter than its true width. This implies we consider the situation where ϕ'' (the true dispersion) is enough to significantly broaden the pulse, and $\delta\tau/\Omega$ (the erroneous dispersion added by the measurement) is of opposite sign so as to diminish our estimate. Because of the second-order nature of broadening in (8.6), the measured pulse width can be very close to transform limited, and yet still include enough dispersion such that the effects of extra delay are well approximated by a linear treatment of the dispersion curve. To derive a rough scaling law of the sensitivity of the measurement, then, we consider the first-order change in (8.6) to $\delta\tau$. Performing a series expansion of (8.6) with respect to $\delta\tau$ and then solving for δT gives

$$\delta T = \left(\frac{\phi''}{\sqrt{4/\Delta\omega^4 + (\phi'')^2}} \right) \left(\frac{\delta\tau \, \Delta\omega}{\Omega} \right) + O[\delta\tau^2]. \tag{8.7}$$

Strictly speaking, this first-order expression is only accurate for $\phi'' \gg -\delta\tau/\Omega$, but it turns out to be off by no more than 25% as long as $|\phi''| > |\delta\tau/\Omega|$. However, this expression is only an accurate prediction if we can estimate the actual dispersion of the pulse (as in the case of an intentionally chirped pulse) and we know our delay error is small relative to it. Nonetheless, the prefactor is close to one for any significant actual chirp (the case we must worry about). For example, it is already about 0.45 for a pulse that is 12% wider than its transform limit. Therefore, we have a worst case error that roughly scales as

$$\delta T \approx \frac{\Delta\omega}{\Omega} \delta\tau. \tag{8.8}$$

The final result is rather intuitive, and simply states that the absolute measurement error is approximately the uncertainty in the interpulse delay times the dimensionless quantity $\Delta\omega/\Omega$, found earlier to be proportional to the number of spectral samples, and thus also the time-bandwidth product. Again, this formula is an overestimate in the case where our measured pulse is transform limited (since $\phi'' = -\delta\tau/\Omega$). However, in reality this is somewhat offset by the fact that very short pulses usually have residual higher-order dispersion which cannot be perfectly compensated. In practice, this formula is thus fairly accurate, as illustrated later in Section 8.5.

A corollary of (8.8) is that the relative error in measurement scales with the square of the spectrum,

$$\frac{\delta T}{T + \delta T} \approx \frac{\Delta\omega^2}{2\Omega} \delta\tau, \tag{8.9}$$

which was obtained by multiplying both sides of (8.8) by the bandwidth and using the Fourier uncertainty relation, where we have assumed that the pulse *as measured* is close to the Fourier limit.

One might hope that as pulses become shorter, the number of sampling points could

be kept constant by increasing Ω in proportion to the bandwidth. Unfortunately, this is generally not the case for few-cycle pulses, as explained in Section 8.4. As bandwidths increase in an optical system, the temporal window which we must resolve becomes limited by the pulse pedestal and secondary pulses, and at this point the shear must remain fixed and the number of sampling points must grow with the spectrum. Thus, while (8.9) would indicate a linear scaling with bandwidth if everything in a laser scaled in unison, in practice this is not the case, and we must conservatively assume that the scaling is square in the bandwidth.

8.3.2 Tolerance on Delay Uncertainty for Compressed Pulses

In the preceding section, a rough intuitive scaling law was derived where several approximations were made, most notably that some residual measured dispersion remains. We now address the situation where the pulse has been perfectly compressed according to the measurement. In this case, we can exactly determine how much delay uncertainty can be allowed while still being assured of having a compressed pulse to within some tolerance. Taking 5% as our tolerance for deviation from the transform limited width, the allowable delay uncertainty $\delta\tau_{5\%}$ can be found without approximation by taking $T \to 1.05T_0$ and $\phi'' \to \tau_{5\%}/\Omega$ in (8.5). Solving for $\delta\tau_{5\%}$ gives

$$\delta\tau_{5\%} = 0.64\frac{\Omega}{\Delta\omega^2} = 0.32\frac{T_0}{N}. \tag{8.10}$$

where N is the number of spectral sampling points within the bandwidth (twice the time-bandwidth product).

8.4 Calibration of SPIDER in Practice

8.4.1 Required Precision

By way of example, we consider the prospect of measuring a single-cycle gaussian pulse whose full width $1/e^2$ bandwidth is $\Delta f = 282$ THz. Using a shear of $\Omega = 2\pi \cdot 5$ THz will result in a resolvable time-bandwidth product of roughly 30 (using the FWHM values). Based on our experience with few-cycle lasers, this would be a conservative resolution requirement for a single-cycle laser. In practice, for few-cycle pulses and below, the spectra tend to be highly structured, with the number of spectral samples required on the order of 20–100. See, for example, Refs. [92, 81, 93]. For a standard SPIDER configuration, this shear implies a delay of around $\tau = 200$ fs. (This is to ensure that sufficient chirp is used such that the upconversion can be considered a pure shift.)

According to (8.10), in order to limit our maximum error to within five percent (roughly 0.14 fs), the interpulse delay error must be measured and maintained to within 25 attoseconds, corresponding to a delay of 7.5 nm. Recently, the shortest isolated pulses ever published [81] were measured using a modified SPIDER (modified to amplify the chirped pulses) with a shear of $\Omega = 2\pi \cdot 4.11$ THz, which implies a tolerance of about 21 as, or 6.3 nm.

8.4.2 Sources of Delay Error

There are several avenues through which unaccounted for interpulse delays can arise in practice with a standard SPIDER setup. It is our hope that by enumerating them, researchers can mitigate their effects simply by keeping aware of them during construction and operation of a SPIDER apparatus.

1. **Delay calibration.** The most obvious source of delay error is simple error in the calibration measurement. In our example, a relative measurement precision of 0.025% is required, over four orders of magnitude. This is not exactly trivial, but certainly achievable using interferometric means. As pointed out by Dorrer in [94], however, errors in the calibration of the spectrometer will translate into errors in the effective τ used. Thus, the spectrometer used in the measurement must be free of relative errors (over the pulse bandwidth) to within the same precision unless the errors can be canceled out by self-calibration (see below).

2. **Thermal drift.** A perhaps more worrisome source of delay error is thermal drift in the setup over time. Taking, for instance, the thermal expansion coefficient of aluminum (2.5×10^{-5}), and considering a relatively small Michelson interferometer with arm lengths of 2.5 cm, a temperature differential of 0.006 degrees Celsius between the two arms will cause a problematic change in delay. Uniform temperature shifts should not pose a problem; an interferometer of any size will be able to withstand a temperature shift of up to 5 degrees Celsius before a noticeable delay occurs.

3. **Alignment drift.** Another source of delay is alignment drift of the incoming laser beam. Any misalignment in the beam will change the delay by τ times the cosine of the angle error. Assuming perfect alignment to begin with, this means that 15 milliradians of change in laser pointing will cause noticeable errors for a single-cycle pulse. This is not an issue for passive stability of the laser, but suggests that any tuning of the laser itself will require a recalibration of the SPIDER for few-cycle pulses.

4. **Unmeasured path difference.** Calibrating the interpulse delay often involves changing the experimental configuration somewhat (such as rotating the nonlinear crystal to produce Type I second-harmonic generation). Thus, care must be taken such that the delay measured is identical to the delay actually used. For example, if the SHG interference fringe is to be used to calibrate out the delay phase, it is even important that the total distance travelled by the beam not change from the calibration configuration to the measurement configuration. This could be a potential issue with noncollinear arrangements where the SHG geometry will be fundamentally different from the SFG geometry. Assuming the two pulse copies are only known to be collinear to within 1 milliradian, changing the total propagation length by more than two centimeters would put the unknown delay out of tolerance for a single cycle pulse.

8.4.3 Avoiding Delay Error

The most obvious lesson from the above is that for few-cycle pulse characterization with standard SPIDER, a new calibration should be performed immediately preceding each measurement to avoid issues of delay stability, leaving only the matter of delay measurement.

Fortunately, the issue of the delay calibration becoming enmeshed with the spectrometer calibration has been previously addressed. In [94], Dorrer shows in that if the delay is calibrated using the interference of the individual second harmonic of each pulse copy using the same spectrometer which will be used for the SPIDER measurement, then any spectrometer error will cancel out. When making measurements of few-cycle pulses using a standard SPIDER, it is thus imperative that the delay phase be removed in such a self-calibrated way. Otherwise, the delay calibration sensitivity translates into the more difficult issue of calibrating the spectrometer to within at least four orders of magnitude relative precision and measuring a fringe period to the same degree. One potential issue with this is that the SHG signals must cover the same bandwidth as the sheared upconverted signals. Given the difference in phasematching between the SHG and SFG signal, this could present a difficulty, and may explain why a self-calibrating SPIDER measurement has not been demonstrated (to our knowledge) for a few-cycle pulse.

Ideally, for the sake of avoiding any possibility of delay drift, the calibration of the delay would be done simultaneously with the measurement. Dorrer has developed a method [95] to do just that by taking advantage of multiple diffraction orders in a grating spectrometer. His method is also self-calibrating in the sense described above, since both the calibration fringe at the fundamental frequency and the SPIDER fringe at the upconverted frequency share the same wavelength range on the detector.

When standard SPIDER is used, and when the pulse lengths are not so short that dispersion is an issue, beamsplitting may be done using an etalon (as in [96]) to eliminate issues of thermal sensitivity.

Lastly, the effects of beam pointing can be greatly mitigated by ensuring that the interferometer generating the pulse copies is well-aligned, such that pointing errors introduce only second-order delay errors. Collinear SPIDER implementations, which rely on pulse shaping to create the pulse copies [97], and those which use no delay (see 8.6.3 below), have an advantage in this regard.

8.5 Numerical Simulations

In the derivation of Section 8.3.1, we relied on Gaussian analysis. However, the spectra in real lasers tend to have more complicated spectra that are often closer to rectangular than Gaussian. To test the validity of our analytic results, we simulated the effect of a spurious delay $\delta\tau$ on a standard SPIDER measurement of a sub-two-cycle pulse, using spectral data from an actual Ti:sapphire laser.

We recently constructed a sub-two-cycle ring laser [98] and characterized it using 2DSI [20], the spectrum and phase of which are shown for reference in Figure 8-2. The

Table 8.1: Summary of simulated Ti:sapphire measurements

		$\delta\tau$		
	Meas.	15 nm/c	30 nm/c	60 nm/c
FWHM	4.94 fs	5.032 fs	5.26 fs	12.8 fs
RMS	10.95 fs	11.31 fs	11.85 fs	13.31 fs

power and phase spectrum are both rather oscillatory, caused by SPM and extra reflections from the intracavity dispersion compensating mirrors. In order to resolve the finest oscillations, a shear of 4.5 THz was required. Using the spectral phase measured by 2DSI, we simulated the case where a standard SPIDER measurement shows a very slightly chirped pulse of 4.94 fs FWHM. This is nearly as compressed as can be achieved by bulk material compensation. We then computed what the actual pulse was assuming a worst-case interpulse delay error due to several lengths of extra path length. The envelopes were computed using a padded FFT of the complex spectrum, neglecting any carrier offset (since we are only interested in the pulse envelope). The pulse RMS width (over a 40 fs window) and the FWHM were computed. The latter was calculated by using a Newton method to solve for the intersection of a cubic spline with the 50% point. The resulting pulses are shown in Figure 8-3, with results tabulated in Table 8.1. (The RMS widths are much larger due to significant satellite pulses and pedestal.)

Note that it only took an extra delay of 30 nm, or 100 as, to cause an error over 5%. Furthermore, this example was actually conservative in that the measured pulse was well compressed and thus the nonlinear relation of the FWHM width to dispersion helped; the same data also imply that had the actual pulse been 12.8 fs long, only 30 nm of spurious delay would've appeared to be only 5.25 fs long. The point here is that if spectral shearing is used to measure pulses that are intentionally chirped (as in the case of pulses used in coherent control or those precompensated for material dispersion) the measurement will be maximally sensitive, such that (8.8) is an accurate estimate.

To test the applicability of the analytic results from Section 8.3.1, the relative error was simulated for a range of $\delta\tau$ between zero and 60 nanometers, and compared with that predicted by (8.9), taking the $1/e^2$ half width $\Delta\omega$ to be $2\pi\cdot 138$ THz (about half of the full range of the measured spectrum). The results are shown in Figure 8-4. After enough dispersion, the FWHM behaves severely nonlinearly as subpulses grow past 50%. As anticipated by the fact that this pulse is nearly transform limited as measured, the linear scaling law overestimates the errors. However, the error is not large, and it is generally within a factor of two of the RMS width. By comparison, the exact 5% tolerance predicted by (8.10) is $\delta\tau_{5\%} = 81$ nm, a significant underestimate of that actually achieved (due to the residual higher order dispersion in this measured pulse). This is because (8.10) assumes a smooth spectrum with no high order dispersion. Thus, if our laser is representative, the sensitivity estimates given in Section 8.4.1 for a single-cycle laser may be conservative.

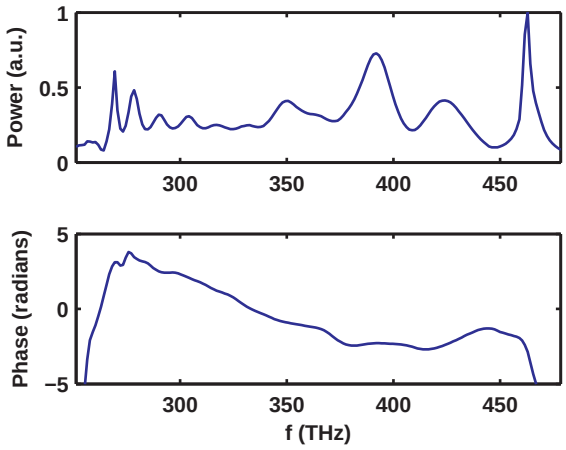

Figure 8-2: The spectral power density and spectral phase for an actual sub-two-cycle Ti:sapphire oscillator measured with 2DSI.

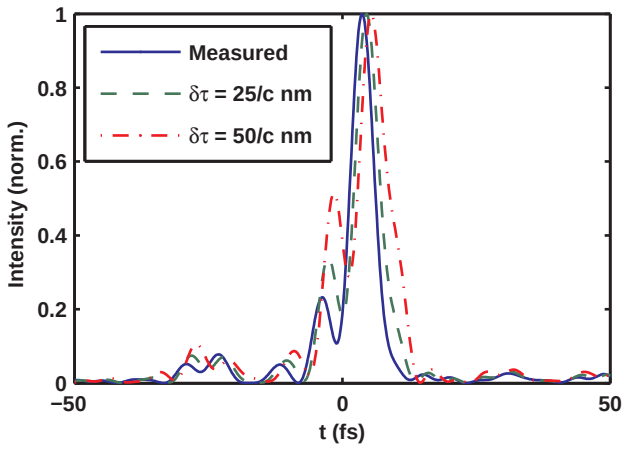

Figure 8-3: Simulated pulse intensity as measured (solid) and in truth (various hashed) for the Ti:sapphire laser whose spectra is shown in Figure 8-2, assuming delay errors of 25 and 50 nm.

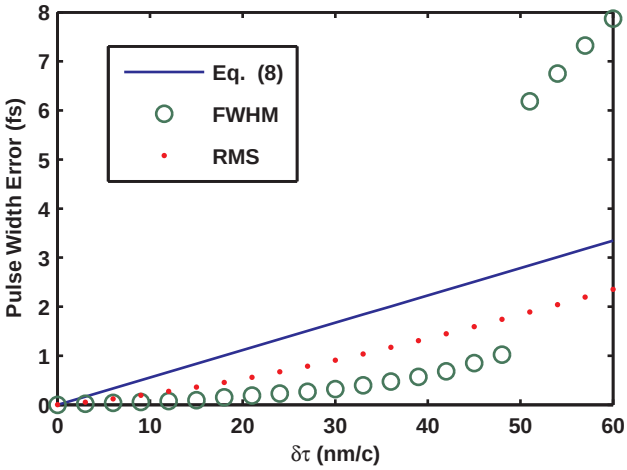

Figure 8-4: The measured pulse FWHM and RMS widths of the pulse in Fig. 8-3 for a range of $\delta\tau$ values, compared with that predicted by (8.8) using the half-width of the spectrum as $\Delta\omega$.

8.6 Alternative Spectral Shearing Methods

We have already presented an outline of various ways to mitigate systematic errors related to spurious delays in spectral shearing. For cases where the sensitivity or pulse splitting of SPIDER cannot be tolerated, several alternative modes of spectral shearing have been developed in the past several years which address these issues.

8.6.1 Arbitrary Shear Methods

As discussed before, the shear plays a crucial role in the sensitivity of the measurement. This applies to any spectral shearing method. Since the error δT is proportional to the absolute uncertainty $\delta\tau$, and not the relative uncertainty, the effect of calibration errors can be minimized by choosing as small a τ as possible while still allowing decoding of the fringe phase. Similarly, we want to choose as *large* an Ω as is consistent with the Nyquist criterion of the spectral phase.

In standard SPIDER, unfortunately, the delay and shear are linked through the dispersion used to produce the monochromatic signal used for upconversion. From [18] this relation is $\tau = -\Omega\phi''_{\text{chirp}}$. It can be shown that the amount of chirp needed to avoid artifacts is proportional to $1/\Omega^2$. However, the minimum requirement on τ is actually quite complicated, and it is best found empirically. Thus, SPIDER does not, in general, offer sufficient degrees of freedom to optimally choose τ and Ω.

However, there are several spectral shearing variants that do allow for independent selection of τ and Ω. The first to do so was HOT SPIDER [99], which uses a homodyne technique to allow two measurements against a local oscillator to be combined to yield a normal SPIDER trace. While this method requires a second source that covers the upconversion wavelengths, it also has the benefit of implicitly calibrating the delay. However, this method may be of limited use for very short pulses because of the requirement

for a separate source with the same bandwidth but at twice the frequency (the main source can also be used, but this would be highly inefficient and against the point of homodyning).

8.6.2 Zero-dispersion Methods

Another approach that is capable of arbitrary shears is ZAP-SPIDER [100], developed by Baum, which introduced the idea of using dual chirped pulses to upconvert a single short pulse. This means that the pulse to be measured never has to pass through any material (other than the thin nonlinear crystal before it is upconverted) and hence this method adds Zero Additional Phase. To our knowledge, ZAP-SPIDER was the first demonstration of a self-referencing pulse characterization method that involved no added dispersion to the measured pulse. One potentially issue is that the noncollinear nature of ZAP-SPIDER may present difficulties in measuring τ, at least in a self-calibrating way. SEA-SPIDER [101] and 2DSI [20] also use two chirped pulses to avoid dispersion on the measured pulse, and furthermore both set the delay to zero. This brings us to another way to address the delay calibration: avoid it altogether.

8.6.3 Zero-delay Methods

In the SPIDER interferogram, described in (8.1), the dense fringe created by the delay phase $\tau\omega$ allows for robust and unambiguous extraction of the $\phi(\omega) - \phi(\omega - \Omega)$ term in which we are interested. However, this "carrier" fringe need not be in the spectral domain; the phase shown in the lower arm of the SPIDER schematic in Figure 8-1 does not have to be a function of optical frequency. In fact, having any component of it in the spectral domain is really the origin of the entire calibration sensitivity issue discussed in this paper.

To this end, SEA-SPIDER, developed by Kosik [101] and demonstrated for sub-10 fs pulses by Wyatt [102], is a version of SPIDER that creates a fringe in the spatial domain on an imaging spectrometer. A related method, 2DSI [20], developed by the authors, uses a collinear output and creates a fringe in the time domain.

In theory, these two methods should be immune to the delay uncertainty errors discussed in this paper. However, in practice things may not be so simple. The error given in equation (8.8) is true for any spectral shearing method, even those with nominally zero delay between the two pulse copies. Thus, any incidental path length difference that occurs will contribute to measurement errors in exactly the same way as with standard SPIDER.

In the case of SEA-SPIDER, the spatial fringe is created by sending the two upconverted pulses along separate routes in a plane before meeting at the spectrometer. This creates a spatially dependent delay in an axis perpendicular to the axis over which the spectrum is resolved. Any deviation of these pulses out of the plane, or delays incurred during their separate travels, will create a $\delta\tau$ that must be either calibrated or avoided to the same precision as for standard SPIDER. Furthermore, if the spectrometer grating axis is rotated with respect to the nominal spatial fringe, this will have the same effect as a delay, and may have to be calibrated.

In 2DSI, the fringe is produced by scanning the phase of one of the chirped pulses, and a two-dimensional fringe is produced as a function of wavelength and this phase. The only thing that matters in the 2DSI fringe is the absolute phase of the fringe at a given wavelength, and thus no calibration is needed. The cost of this is that 2DSI is incapable of single-shot measurements.

The two upconverted pulses in 2DSI originate from the same point and are collinear, so it should not be possible for a delay to occur between the pulse copies. However, misalignment and nonidealities in the imaging of the pulses into the spectrometer could potentially introduce an unwanted delay. Nonetheless, we have not seen any evidence of this after several measurements of few-cycle pulses [19, 103]. In fact, we have recently measured a 4.9 fs pulse directly from an oscillator, and verified it against both an IAC and a simulation of the laser [98, 93, 24].

8.6.4 Multiple Shearing

Finally, for those methods where multiple shears can be produced–and where changing the shear can be guaranteed to have no affect on the delay–the issue of delay calibration can be eliminated by making two or more measurements with different shear frequencies. By subtracting the phase of two spectral shearing measurements made with different shears, the phase of the delay drops out and one is essentially left with a SPIDER measurement performed with the difference in the shears.

As long as one of the measurements is done with a shear that is consistent with the sampling theorem, the subsequent "calibration" measurements can be made with much larger shears with the only requirement being that they are integer multiples of the shear used for the final measurement. A novel version of SPIDER which relies on this principle for a continuum of shears was recently presented by Gorza [104].

8.7 Conclusion

We have shown, though analysis and numerical simulation, that as pulses approach the single-cycle limit, the SPIDER technique involves a calibration that is exceedingly difficult. However, given the unique position of SPIDER as the only direct method of phase measurement, and given its inherent bandwidth advantages over other methods, it is worthwhile to search for mitigation strategies. We conclude that for most cases, awareness of the calibration sensitivity and careful adherence to the principle that calibrations must be done before every measurement are sufficient to yield accurate results. For extremely short pulses, however, it may best to employ one of the variants of SPIDER discussed in Section 8.6 that remove the delay calibration issue.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 9

Two-Dimensional Spectral Shearing Interferometry

9.1 Introduction

Spectral shearing interferometry (pioneered by the SPIDER method of Iaconis and Walmsley [18]) is unique among pulse measurement techniques in that it unambiguously measures the spectral phase of an optical signal by directly interfering neighboring frequency components. Furthermore, by encoding group delay information on the phase of a spectral domain fringe, SPIDER produces a frequency modulated signal and thus enjoys excellent immunity from noise and nonlinear phasematching bandwidth effects. This gives spectral shearing an advantage for measuring very wide bandwidths, relative to the "amplitude modulated" signals characteristic of other methods, differences in efficiency notwithstanding. Consequently, it has become one of the principal methods used to measure few-cycle pulses [89], alongside FROG [85], and the many variants of both. As is the case for any method, however, challenges arise with standard SPIDER as bandwidths approach the single-cycle. Our method seeks to address these issues by trading single-shot ability (at least in the present version) for calibration stability.

All spectral shearing methods, by their nature, involve measuring spectral group delay by observing the interference of two spectrally shifted copies of the pulse being measured. As will be explained in more detail in Section 9.2, any linear phase (delay) that occurs between the two components will be interpreted as a quadratic phase (dispersion) in the reconstruction. It turns out that the resulting error in measured pulse width scales linearly with the unaccounted for delay, multiplied by a factor proportional to the resolution of the measurement. At the very least, this implies that great care must be taken in a SPIDER measurement to ensure that any linear phase is calibrated out.

Here, we present a method recently we developed [19] that seeks to ensure that there is no possibility for uncalibrated phase, by eliminating the possibility for a delay by robustly encoding the spectral phase measurement in a series of spectra of a single output pulse. Our technique requires only the non-critical calibration of the shear frequency and does not perturb the pulse before up-conversion. Rather than encode the spectral group delay in a fringe in the spectral domain, 2DSI encodes phase in a pure cosine

fringe along a completely independent dimension, by scanning the relative phase of the two spectrally sheared components. This reduces the resolution demands on the spectrometer to that required for proper sampling of the pulse itself, allowing for complex phase spectra to be measured with high accuracy over extremely large bandwidths, potentially exceeding an octave. We believe that 2DSI is thus a cost effective and efficient method for accurately and reliably measuring few- and even single-cycle pulses. While the method is relatively recent, it is well tested and has been successfully demonstrated on several different lasers, including one producing 4.9 fs pulses with 4.3 fs transform limited bandwidth [20, 21, 22].

9.2 Background and Motivation for Delay-Free Method

In any spectral shearing method, the net result is the creation of two spectral copies of the pulse under measure, with a small frequency shift Ω disposed between them. For sake of simplicity, in this paper we will ignore the overall offset in frequency caused by the nonlinear upconversion, and will express everything in terms of the lowest upconverted pulse spectrum $A(\omega) = |A(\omega)|e^{i\phi(\omega)}$. In the conventional SPIDER method, the two output pulses are delayed with respect to one another by a time τ, leading to a dense fringe in the spectral domain upon which the group delay spectrum is encoded,

$$I(\omega) = |A(\omega)|^2 + |A(\omega-\Omega)|^2 + 2|A(\omega)A(\omega-\Omega)|\cos[\tau\omega + \phi(\omega) - \phi(\omega-\Omega)]. \tag{9.1}$$

If the delay is large enough, the argument to the cosine can be isolated by signal processing of the fringe in the Fourier domain. The contribution of the delay is then subtracted out, leaving the finite difference of the spectral phase. More details are provided in [18] and [1].

As explained in more detail in Chapter 8, any unaccounted for delay is lumped in with the phase finite difference, and mistaken for a quadratic phase term, potentially one which results in underestimation of the true pulse width. Any error in estimating the interpulse delay results in an absolute error in the extracted pulse width [105]. The worst-case measurement error is equal to the delay error multiplied by the ratio of the shear to the bandwidth; this ratio is essentially the number of points over which we sample our spectral phase. Because this error is absolute, the relative measurement error caused by a given interpulse delay uncertainty is proportional to the square of the pulse bandwidth, as shown in equation (8.9). It should be kept in mind that this is a worst case error and an approximate analytic result; we are not claiming all SPIDER measurements are uncertain to the degree given in (8.9). That notwithstanding, the scaling of the error with the square of the bandwidth applies regardless and implies that calibration requirements increase dramatically as the single-cycle is approached.

The shear is the sampling interval in the spectral domain. By the sampling theorem, it is inversely proportional to the temporal window over which we are measuring. As we will argue further in Section 9.8, the minimum time window we must resolve (i.e. the spectral sampling density) ceases to scale inversely to bandwidth once pulse widths become shorter than the characteristic time scale of intracavity dispersion oscillations

in the source laser (such as those caused by reflections off the surface of dispersion compensating mirrors). The scaling of the delay uncertainty is more complicated, but by similar arguments, it either scales with the temporal measurement window (in the case of standard SPIDER) or is a constant (in the case of zero delay variants). As such, beyond a certain bandwidth the measurement uncertainty scales as the square of the bandwidth.

As the single cycle limit is approached, the effect of even tens of attoseconds of unknown delay becomes a concern. Calibrating and maintaining a delay on the order of hundreds of femtoseconds to within tens of attoseconds is not trivial, especially if the experimental configuration must be changed to measure the delay, as in those systems where the delay is measured by the spectral interference between the second harmonics of the pulses. At the very least, it is certainly necessary to recalibrate the device before each measurement.

The preferred way of calibrating a standard SPIDER fringe, introduced by Dorrer [94], does not involve explicitly computing a delay. In this approach, two SPIDER measurements with different shears are subtracted from each other to obtain the SPIDER phase net of any common phase due to the pulse delay and spectrometer nonlinearity. (This assumes that it can be done so such that the delay is invariant between the two measurements. One can ensure this is the case by taking them simultaneously, as in [95].) When the spectrometer and delay calibrations are mixed together in this way, one then does not speak of "delay" calibration but rather a general baseline phase calibration that handles simultaneous calibration of the delay and any spectrometer nonlinearity. However, whether explicit or not, a measurement of delay is inherent, and as the main component of the SPIDER fringe and the leading term in any calibration error, it makes sense to cast any discussion of sensitivity in terms of the effective delay uncertainty. Even in the absence of systematic delay miscalibration, simple measurement noise will contribute some effective $\delta\tau$ that will be the dominant source of pulse width uncertainty. This is borne out in the noise simulations discussed in Section 9.9.

A second difficulty with standard SPIDER is the need to split the measured pulse into two copies. In practice, doing so without the introduction of additional dispersion is extremely difficult, especially for pulses with spectra exceeding an octave. The splitting issue can be avoided by the use of two chirped pulses, as pioneered by Zero Additional Phase SPIDER (ZAP-SPIDER) [100]. ZAP-SPIDER produces a standard SPIDER fringe, but does so without requiring the interferometric splitting of the short pulse, adding no additional dispersion from the beamsplitting operation, hence the name. However, the calibration issue still remains in ZAP-SPIDER, as the two chirped pulses are disposed along separate paths to allow for the upconverted components to be delayed with respect to one another.

Our solution to the above issues is to use two collinear chirped beams such that the output is a single beam, in essence a single pulse with a complicated spectrum. This arrangement should eliminate the potential for a delay to occur between the two output components. That this is the case is suggested, at the least, by our results in Section 9.10.2. However, we must admit that the precise extent to which this is true cannot be proven conclusively down to the level of attoseconds. In the absence of a perfectly known "reference pulse," a numerical spatiotemporal simulation of the noncollinear up-

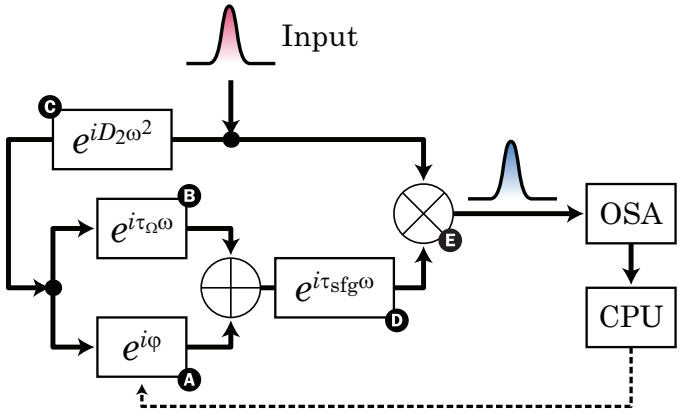

Figure 9-1: Frequency domain block diagram of 2DSI process.

conversion would have to be performed to determine the effects of phase matching on the output beam properties.

9.3 Principle of 2DSI Operation

9.3.1 Spectral Shearing Approach

In what follows, refer to Figure 9-1 for a conceptual frequency domain block diagram of the conceptual process, and Figure 9-2 for a prototypical experimental layout.

As with any shearing method, the method makes use of two CW beams separated in frequency by Ω, which will be used to modulate the pulse under test in a nonlinear crystal. The only requirement on these CW beams is that they are coherent to each other and their frequency difference is stable. As such, they may originate from a separate mode-locked laser from the one being measured, or even from two locked CW lasers. In the case of measuring weak laser pulse it may be helpful to amplify a narrowband portion of the original pulse to be used for the CW beams. In the majority of cases, however, it is most convenient to use the short pulse as the source of the CW signals, and this is accomplished by dispersion and splitting of the pulse into two chirped pulses that act as quasi-CW sources (element C in Figure 9-1).

In any case, because the short pulse is mixed with two essentially pure frequencies the output spectrum consists of two spectrally shifted copies of the original spectrum, each with a slightly different shift. If the chirped pulses are collinear, the two frequency components will have the same propagation vector and thus we can regard the output as a single pulse. This neglects the small difference in transverse photon momentum between the two chirped pulses. However, for small deviations such that nonidealities in the imaging optics can be neglected, there should be no possibility for delay to arise.

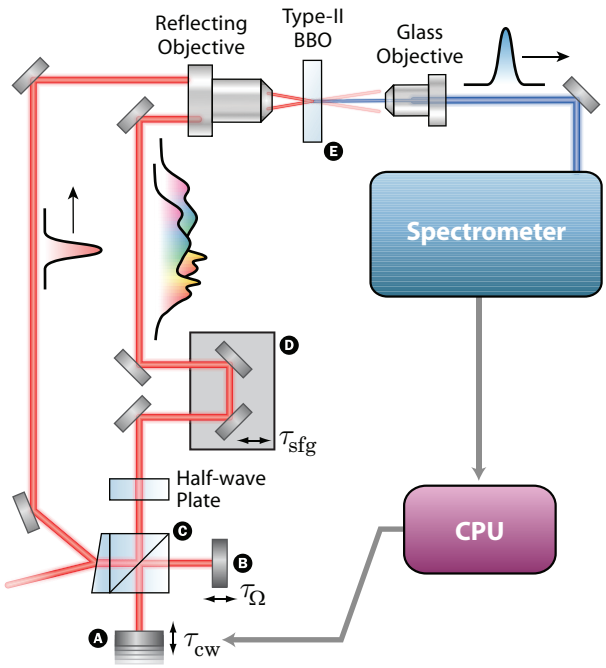

Figure 9-2: Experimental schematic of 2DSI setup.

A given wavelength is imaged to the same pixel regardless of output angle, and thus no phase difference should arise by Fermat's principle.

The upconversion frequencies can be independently and arbitrarily chosen to suit the pulse characteristics and maximize the overall signal to noise ratio. A discussion of how best to do so is taken up in Section 9.8. This ability also allows for a wide range of pulse widths and complexities to be measured by the same setup. Given that a large range is likely to be encountered in the course of optimizing a laser, it is useful to have this flexibility. In contrast, the shear in standard SPIDER is constrained by the dispersive element and the required delay, limiting its versatility. The collinearity of the chirped beams in 2DSI also means that the shear can be changed without affecting the alignment of the system.

The pulse under test only experiences a few reflections, no transmission through bulk materials, and is therefore relatively unperturbed before measurement. The Michelson interferometer that splits the chirped pulses, on the other hand, can be highly dispersive and even unbalanced, allowing for the use of simple splitting optics. The low distortion experienced by the short pulse is extremely important for single-cycle pulses, as any spurious dispersion inherent to the apparatus will invariably end up (in the opposite sign) in the pulse after optimization.

The fact that the output is essentially one pulse, the spectrum of which encodes the spectral group delay, is the key to the stability and reliability of 2DSI. In theory, all the information needed to recover the spectral phase is present in the spectrum of one of these output pulses if we were willing to trust the spectral amplitude information. However, such operation would negate one of the primary strengths of SPIDER methods: the encoding of group delay as a fringe phase, lending immunity to phasematching band-

width effects. It would also introduce a time direction ambiguity.

Thus, to observe the spectral phase difference between the two pulse components at a given wavelength (which we recall is proportional to the local group delay) in a robust way, the phase of one of the quasi-CW beams is scanned over a few optical cycles prior to mixing (block A in Figure 9-1), which results in a phase φ being added to the upconversion. The net effect is to produce a variable differential phase between the upconverted spectral components. Scanning this phase allows one to directly observe interference between the two spectral components as a function of wavelength, providing a direct observation of the spectral group delay of the original pulse.

This phase scan is typically achieved by vibrating one of the mirrors in the Michelson interferometer, though it could also be achieved with some other kind of phase modulator, such as a liquid crystal or electro-optic device. Assuming the quasi-CW beam is monochromatic over a time scale of a few optical cycles, a delay τ_{cw} simply manifests as a pure phase $\varphi = \tau_{\mathrm{cw}}\omega_{\mathrm{cw}}$, and doesn't materially shift the frequency of upconversion.

9.3.2 Form of Interferogram

The spectrum of the up-converted signal is then recorded as a function of the phase φ, yielding a 2-D intensity function that is given by

$$\begin{aligned} I(\omega,\varphi) &= |A(\omega) + A(\omega-\Omega)e^{i\varphi}|^2 \\ &= 2|A(\omega)A(\omega-\Omega)|\cos[\varphi + \underbrace{\phi(\omega)-\phi(\omega-\Omega)}_{\tau_g(\omega-\Omega/2)\Omega+O[\Omega^2]}] + \mathrm{D.C.}, \end{aligned} \tag{9.2}$$

where, as before, $A(\omega)$ is the low upconverted pulse spectrum, and $\phi(\omega)$ is its spectral phase. (For simplicity we refer to everything in terms of the upconverted fields.) The under-bracketed expression can be viewed as a finite difference approximation of the group delay scaled by the shear. This term is what all SPIDER variants measure, though in 2DSI it is available directly without any filtering, as discussed in the next section, since there are no other terms dependent on the optical frequency. A simple two-dimensional raster plot of $I(\omega,\varphi)$ reveals the up-shifted pulse spectrum along the ω-axis, with fringes along the φ-axis that are locally shifted in proportion to the group delay at the given frequency (as illustrated in Figure 9-3). The user can thus immediately ascertain salient properties of the complex spectrum simply by looking at the raw data: the cosine fringe at each wavelength is vertically shifted in proportion to its actual delay in time, with the fringe amplitude roughly proportional to the power spectral density (neglecting bandwidth effects). The ability to use the raw spectrometer data when optimizing a laser yields information not available from processed data from an inversion algorithm alone, such as measurement noise and laser stability.

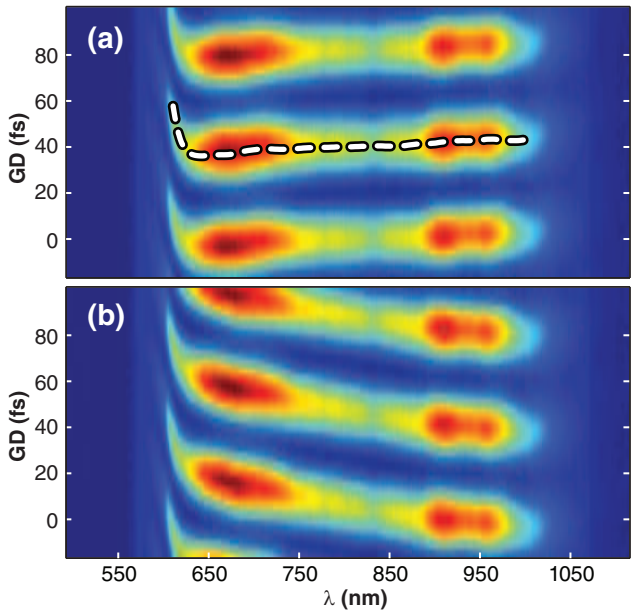

Figure 9-3: Raw 2DSI traces from 5 fs laser, with (a) extracted spectral group delay overlaid to demonstrate the interpration of fringe offset, and (b) the same pulse with one mm of fused silica. The presence of extra dispersion is evident in the raw trace without any need for reconstruction.

9.4 Relation to Other Spectral Shearing Methods

9.4.1 Relation to SPIDER

Comparison of equations (9.1) and (9.2) shows that the fringes produced by 2DSI and conventional SPIDER are mathematically identical, except for the fact that the fringe in 2DSI is produced by a phase occurring in a separate domain (the φ domain), whereas in SPIDER the fringe oscillates in the spectral domain. In either case, the fringe can be viewed as creating sidebands in the respective Fourier domain of the fringe; with SPIDER we get sidebands in the pseudo-time domain and in 2DSI the sidebands are pulled into the "pseudo-frequency" domain relative to the phase delay. This is illustrated in Figure 9-4. In both, the purpose of the fringe is to pull the sidebands away from the central DC term so that they don't interfere, rendering the phase extraction insensitive to the amplitude of the fringe. In SPIDER, the fact that the sidebands are in the optical frequency domain results in a significant increase in required spectrometer resolution over that needed to simply resolve the fundamental pulse spectrum. In 2DSI, the extra dimension means that the spectrometer resolution required is simply that required by the Nyquist limit for the pulse being measured (i.e. determined by the time window that must be resolved), enabling larger time bandwidth products to be measured than with standard SPIDER.

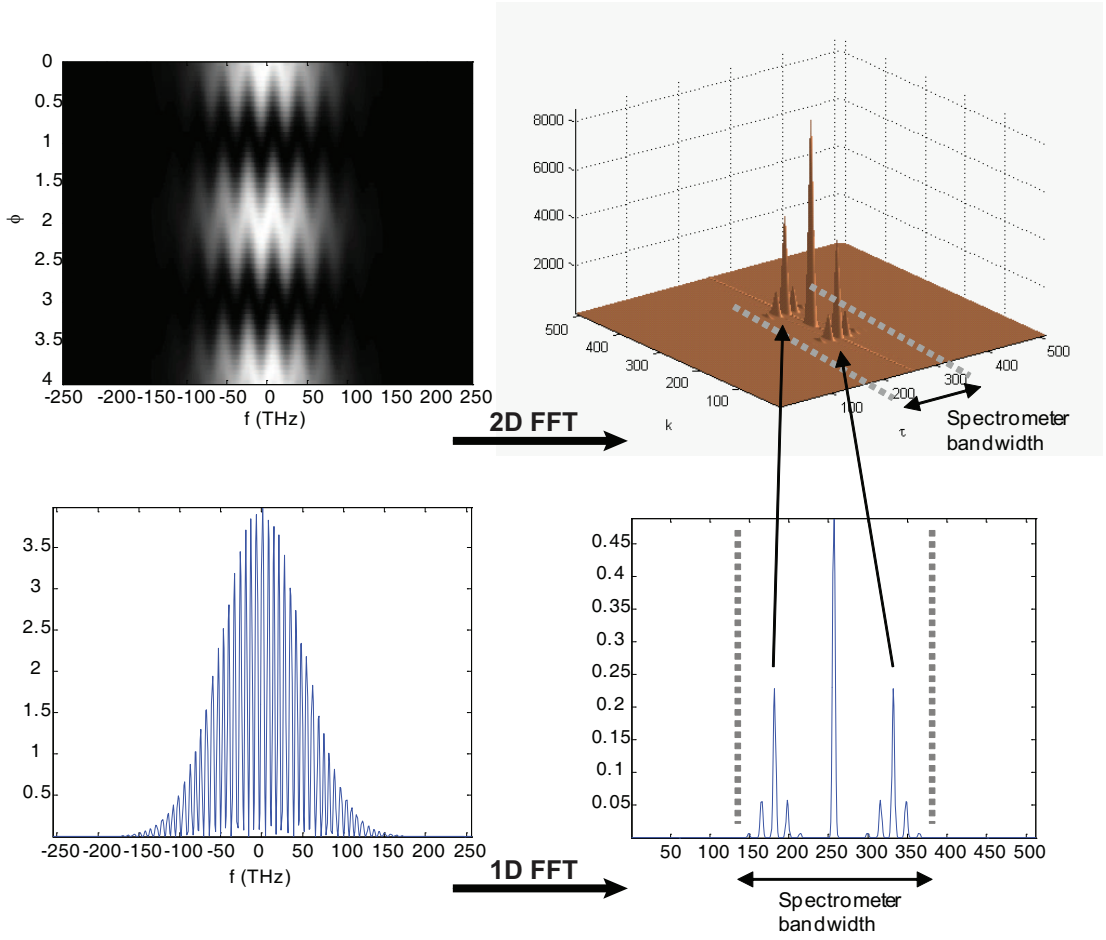

Figure 9-4: Illustration of the different sideband schemes between 2DSI (top) and SPIDER (bottom), showing the two schemes for pulling the information containing sidebands out of the DC term for a gaussian pulse. In SPIDER, the sidebands are created in the frequency domain, whereas in 2DSI the same sidebands remain stationary in spectral domain and are shifted in a second dimension in the 2D interferogram.

9.4.2 SEA-SPIDER

SEA-SPIDER, a method developed by Ellen Kosik [101] and successfully adapted for few-cycle pulses by Adam Wyatt [102], follows ZAP-SPIDER in using two chirped pulses. By putting a tilt between the two upconverted beams and measuring the output with an imaging spectrometer, a spatial fringe is observed in one axis while spectrally resolving the other,

$$I(\omega, x) = 2|A(\omega, x)\,A(\omega - \Omega, x)|\cos[\omega \cos\theta\, x/c + \phi(\omega) - \phi(\omega - \Omega)] + \text{D.C.}, \tag{9.3}$$

where x is the transverse spatial dimension and θ is the half angle of intersection between the two upconverted beams.

SEA-SPIDER's use of spatial encoding allows for single-shot pulse characterization, including some kinds of spatiotemporal measurement along one axis. 2DSI is similar to SEA-SPIDER in many respects, with the fringe encoded in terms of an upconversion phase, rather than mixed with the pulse profile and encoded in space. However, the use of separate output paths from the crystal in SEA-SPIDER makes the method more susceptible to a delay occurring between the point of upconversion and interference on the CCD detector. It is necessary to calibrate both the intersection angle of the two beams in SEA-SPIDER (the nominal fringe spacing) as well as the angle of the spectrometer grating axis (i.e. the nominal fringe angle) and inspection of (4) and comparison to (1) shows that miscalibration of either angle is equivalent to an erroneous delay in SPIDER, manifesting as discussed in Section 9.2.

Because nonlinear mixing does not commute with a linear delay, the fact that a 2DSI system involves a scanning of the chirped pulse delay *prior* to upconversion renders it fundamentally different from SEA-SPIDER, which involves delays applied after. The result is that in SEA-SPIDER there is a varying time delay between pulse copies (with an associated frequency dependence), whereas in 2DSI there is a changing pure phase that is constant in frequency.

We believe varying the chirped pulse phase before upconversion results in a more robust measurement than changing a delay after upconversion, as our approach results in a simpler interferogram that requires less calibration. Furthermore, because the output of 2DSI is a single pulse, it is impossible for the two components to experience further differential phase shifts.

9.4.3 CAR-SPIDER

More recently, Ian Walmsley et al introduced a method which simultaneously records multiple SPIDER fringes over a range of different shears [104]. This allows for the calibration of the delay through a consistency requirement on all of the fringes, and was recently demonstrated for a 70 fs pulse. Because it implicitly relies on a thick crystal, however, applying this method to few- or single-cycle pulses may not be possible.

The orthogonality of the temporal fringe to the spectral domain is also why the fringe period does not need to be known in 2DSI. At each wavelength, the fringe encodes only one piece of information (the group delay), and we are only concerned with the relative

shift between wavelength. We thus do not care about the frequency of the fringe or even whether or not it is constant, only needing to know its relative phase. This eliminates many potential avenues for measurement error in a real system, such as scan linearity and calibration. In fact, the only calibration needed by 2DSI is for the up-conversion frequencies which produce the shear. This is a relatively non-critical calibration, as the relative pulse measurement uncertainty is proportional to the relative uncertainty in the shear [18].

9.4.4 Single-shot potential

There is a cost for the present version of 2DSI's simplicity in reconstruction and experiment, which is the loss of single-shot capability. An alternative, spatially-resolved embodiment of 2DSI (which might be termed 3DSI) should be capable of single-shot operation (as discussed in Section 9.11). Nonetheless, while the requirement for scanning in 2DSI renders single-shot operation impossible, it is capable of video-rate operation. This helpful for laser tuning or automated pulse compression. In fact, the scanning mirror only needs to move a few microns, at most, and thus the system is theoretically capable of operation at kilohertz rates with sufficient laser power.

In practice, the issue of single-shot measurement is often moot; signal to noise generally puts a lower bound on the required integration time to yield sufficient accuracy. This requirement will usually exceed the pulse repetition rate for unamplified pulses, especially given the low efficiency of spectral shearing techniques (save for modified SPIDER, which uses amplification of the ancillary chirped pulses to increase signal [106]).

9.5 Physical Layout and Operation

As with any method, there are several ways to implement the optical operations required for 2DSI. In Figure 9-2, we provide a schematic of one approach, to which we will refer here. This arrangement has proven to be robust and cost effective.

9.5.1 Pulse Chirping and Splitting

To begin with, a highly chirped pulse is created by picking off a portion of the pulse to be measured, and dispersing it. Roughly four percent of the short pulse under test is picked off by the Fresnel reflection from the glass cube beamsplitter (C) used in the Michelson interferometer. This can be done with a wedge bonded to the beamsplitter, or by simply operating the interferometer at slightly shallower than a right angle geometry. The remainder of the pulse is split in the interferometer, where a one inch glass beamsplitter provides sufficient chirping to measure a pulse in the few-cycle regime (see Section 9.8.2). If longer pulses need to be measured, an additional glass block can be placed in the output path of the interferometer. The arms of the interfeometer are significantly delayed relative to one another (on the order of picoseconds) such that at any given instant in time there is a difference of Ω between the instantaneous angular frequencies of the chirped beams.

9.5.2 Upconversion

After the interferometer, the polarization of each chirped beam is rotated by a half-wave plate. (The polarization must be rotated since we are using Type II upconversion.) A simple low-order half-wave plate is sufficient since the rotation only needs to occur at two frequencies that are separated by the shear, which is typically on the order of 5 to 10 THz.

The short pulse to be measured and the chirped pulses are made parallel and then focused by an off-axis parabolic mirror into a thin (roughly 30 μm) Type II BBO crystal. Provided they are sufficiently dispersed, when the two chirped pulses are then mixed with the original short pulse in a Type-II $\chi^{(2)}$ crystal in a noncollinear geometry, the short pulse effectively sees a single frequency of each chirped beam, presuming they were sufficiently dispersed.

As pointed out by Walmsley in [1], BBO is rather fortuitous as an upconversion medium for spectral shearing methods, as its Type II phase matching curve can be engineered to have octave-spanning bandwidths in one polarization, with narrow bandwidths in the other. In Figure 9-5 we show a phasematching curve for a typical 2DSI geometry. Numerical simulations show that the Type II bandwidth is maximized when the narrow band beams (and therefore the output) is collinear. 2DSI thus preserves the favorable phase-matching of standard SPIDER. Coupled with the fact that bandwidth effects don't impair spectral shearing measurements beyond reducing the signal to noise ratio, this renders 2DSI theoretically capable of precise self-referenced measurements of pulses down to a single-cycle.

The net effect of each CW signal to shift the frequency of the pulse under test. The absolute upconversion frequencies are set by adjusting the total delay through each path of the interferometer, with the difference between them (the shear Ω) thus determined by the relative delay between each arm. Inevitably, there will also be second harmonic generation (SHG) that occurs in the crystal alongside the desired sum frequency generation (SFG). Since the short pulse and chirped pulses were focused into the crystal at opposite incoming angles, the SFG signal can be separated from the fundamental and second harmonic terms the output by use of apertures. Lastly, the upconverted beam is focused into a spectrometer.

9.6 Control

Mirror (B) is controlled manually by a translation stage with a relatively long travel (at least millimeters). This mirror is used to control the delay between the chirped pulse copies that determines the shear. The other mirror (A) is controlled by a short throw piezoelectric translation stage. Since this scanning mirror will only need to be translated a few microns, at most, a flexure stage can be used to maximize stability and reduce noise during the scan. (Obviously, which mirror is used for which function is arbitrary, and both the large manual adjustment and the small piezo oscillation can be performed on one mirror.)

A third delay stage (D) is used to adjust the temporal overlap between the short pulse

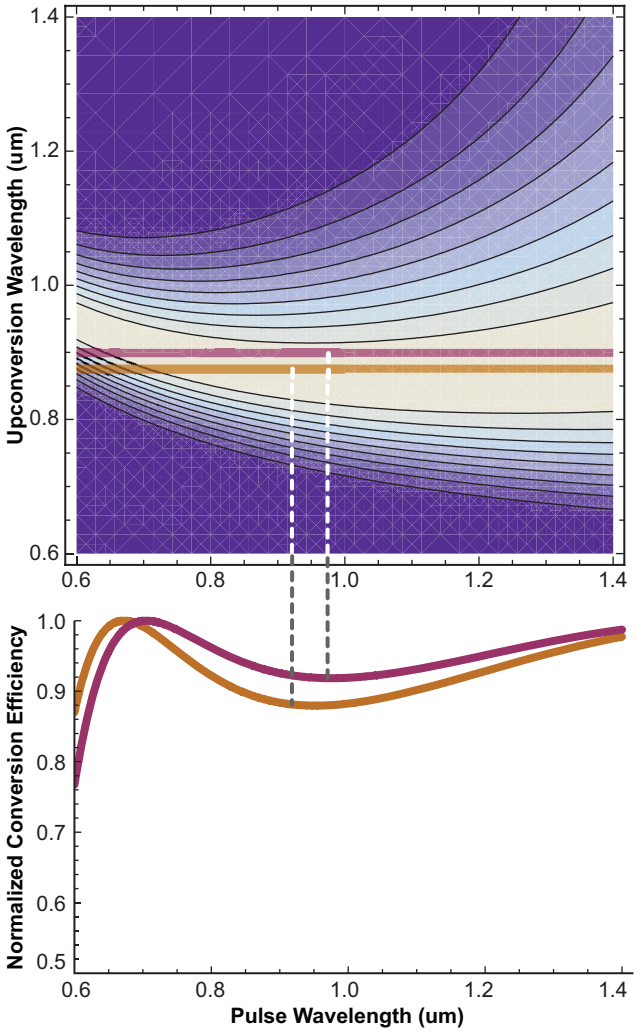

Figure 9-5: top: 2DSI Phase matching plot for Type II sum frequency generation for BBO cut to measure a typical few-cycle Ti:sapphire laser. The lined areas denote the phase-matched regions, with each line denoting increased efficiency by 10 percent. bottom: Slices of the phase matching curves for two upconversion wavelengths separated by 6 THz, showing the efficiency of upconversion for the two spectrally sheared components.

and the chirped pulses, which determines the overall frequency of the upconversion. This stage should be rather long, on the order of a centimeter, at least, to provide for ease in alignment. Together, stages (B) and (D) provide sufficient degrees of freedom that both upconversion frequencies can be independently chosen.

While 2DSI appears on the surface to have an experimental complexity similar to IAC or FROG, due to the scanning, this is not the case given the short distance of the scan and the lack of any calibration needed for it. The scan itself may be handled in an open loop fashion, and a basic function generator may be used to ramp the mirror while spectra are taken at regular intervals. (This is the method we employed to provide the experimental demonstrations shown in the following section, in fact.)

Our simulations show that nonlinearity of the fringe has no effect on the accuracy, though it does affect the reconstruction in the presence of noise by broadening the fringe in the frequency domain. Empirically, we found that even 50% nonlinearity (defined as the maximum relative deviation from the nominal scan rate) results in only a halving of the signal to noise ratio. Nonlinearities below 10% were not found to have an appreciable effect on the noise performance. The linearity of the scan will thus not be an issue for most commercial stages.

Before a reconstruction is performed, the upconversion frequencies must be determined. This is done by alternately blocking one of the arms of the Michelson and recording the spectrum of each individual upconverted component. By cross-correlating the upconverted spectrum with the fundamental pulse spectrum (taken by a separate OSA), the individual upconversion frequencies can be computed, and from this the shear, providing all information necessary for the reconstruction of the pulse. The calibration of the upconversion frequencies is not particularly sensitive; once this calibration is done, it does not need to be repeated for subsequent measurements unless the configuration is changed.

To take a measurement, a computer controlling the piezoelectric stage moves mirror (A) over a total range of roughly 1 micron, recording a dozen or so spectra during the scan. Due to the effective monochromaticity of the chirped pulses, this is equivalent to scanning a pure zeroth-order phase of the corresponding upconverted spectral component. As discussed in Section 9.7, only four spectra are actually needed for reconstruction, but taking more results in a more intuitively understandable trace and one that allows for a simplified reconstruction algorithm.

9.6.1 Alignment

Aligning 2DSI is no more difficult than for any SFG arrangement. Nonetheless, getting ultrashort pulses to arrive at the same time in a micron-sized volume is not an entirely trivial task and there are a few techniques that will make the initial setup much easier.

During the inital building, one must make every effort to ensure that the optical pathlengths seen by the short pulse and by the chirped pulse are the same, such that they arrive at the same time in the crystl. Of course, this can only be done to rough order. It is thus wise to choose a long travel delay stage (one with at least a few cm of travel) for stage (D). It is also helpful to only work with one chirped beam during initial alignment, to avoid the possibility of destructive interference attenuating the output signal

(i.e. during alignment you have no idea if you're sitting on the null of a fringe).

The first alignment task is to ensure that the short pulse beam and the chirped beam are parallel, and that they are aligned in time such that the short pulse will arrive during the broad peak of the chirped pulse. These two goals can be simultaneously achieved by observing a spatial interference pattern, preferably using a CCD camera. To widen the coherence length and increase fringe visibility, a narrowband pass filter should be used to limit the incoming bandwidth to something on the order of 10 nm. Once stage (D) is adjusted to the point of observing strong interference (indicating good overlap of the short with the chirped pulse) they made be made parallel to high precision by attempting to null the fringes in the overlap of the spatial mode tails. Stopping down irises to diffract the beams can be helpful to widen the spatial mode and increase visibility.

The next step is to make sure the focus of the beams fall in the interior of the crystal. This can be achieved by optimizing the SHG signal of each beam independently. At this point, allowing both chirped and short pulses to focus in the crystal should yield an SHG signal visible on a detector. A blue glass filter may be useful to isolate the SHG from the large fundamental scattered light. It is then a simple matter of further aligning the beams and tuning the temporal overlap to maximize the signal. Lastly, the unaligned chirped beam can be brought into alignment by observing spatial interference with it and the aligned chirped beam (again, with a bandlimiting filter in place).

9.7 Reconstruction Algorithm

9.7.1 FFT Harmonic Inversion

The inversion process for 2DSI is significantly simpler than that required for FROG, or even SPIDER. The only information we need to extract from the 2D interferogram is the phase of the fringes along the direction of the scan delay (i.e. the vertical direction in Figure 9-3). Precise quantitative determination of the fringe phase, and thus the group delay, at each wavelength, can be obtained in one of several ways. The most direct method of extracting the fringe phases is to simply fit a cosine to the fringe at each wavelength. While this approach is potentially accurate, it is relatively slow and will not always converge in the presence of noise.

If the fringe period is known exactly, the phase can be very effectively recovered using a 2D variation of the Takada algorithm [107], analogously to what is done in SEA-SPIDER [101]. However, this would require having a calibrated scan. It would complicate the experiment to require the scan to be known and linear, necessitating either feedback control or accurate measurement of the mirror displacement and linearization via signal processing.

Fortunately, the fact that no information is encoded in each fringe other than the spectral group delay means that we can use simpler reconstruction algorithms. In essence, 2DSI reconstruction is an issue of very simple single-term harmonic inversion [108] (i.e. characterization of an unknown sinusoid from a time series) rather than the decoding of a phase modulated known carrier typical of other shearing methods. In fact, any harmonic inversion algorithm should work well to decode a 2DSI signal, especially one

with interference from the fundamental or SHG signals, though in general this would be overkill.

An simple, computationally efficient and direct way to access the fringe phase is to compute a series of 1D FFTs along the fringe axis. According to (2), the fringes generate well-separated sidebands in the continuous Fourier domain (see Figure 9-4) and their phase angle represents the spectral group delay term in (1). However, depending on the length of the scan and how close it is to a multiple of the fringe period, windowing effects may broaden the sidebands such that they and the central DC term interfere, perturbing the phase. This can be effectively dealt with by applying a windowing function to eliminate the broadening. We have empirically found a hamming window to perform quite well if the number of fringes initially visible is at least three. This method comes with the disadvantage of throwing away information at the edges of the scan, however, which will necessarily reduce the SNR of the measurement. Nonetheless, we mention it as it is by far the simplest approach to take, and will work in virtually any situation.

In practice, fortunately, windowing isn't necessary so long as a sufficient number of fringes are observed. Because we are actually measuring group delay, any constant offset is meaningless, which mitigates the effects of windowing. The worst-case relative error in the recovered pulse spectral phase ϕ_{rec} can be shown (see Appendix A) to be approximately limited by

$$|\epsilon_r(\omega)| \leq \frac{\Delta\nu}{n+\Delta\nu}, \tag{9.4}$$

where n is the index of the FT component (the harmonic number), $\Delta\nu$ is the difference in frequency between the fringe and the harmonic, normalized relative to the fundamental, such that the worst case scenario is $\Delta\nu = 0.5$. As can be seen, the error will be relatively small if the fringe frequency is close to the FFT harmonic, or if n is large. As a rule of thumb, so long as three or more fringes are visible, or the scan length is within 10% of the fringe period, the worst possible error will be no more than a few percent and windowing will not be necessary.

It is often advisable to filter in the wavelength domain to suppress noise outside the measurement time window. This filtering can be efficiently combined with the reconstruction by computing a 2D FFT of the interferogram, and selecting a subset of the single line of points representing the 2D sideband within the shear Nyquist limit. Wavelength domain filtering is especially important when the detector is significantly oversampling the spectrum.

9.7.2 Finite Difference Inversion

Having determined the phase of the fringe at each wavelength, multiplying by the shear yields the finite difference of the spectral phase, as shown in (9.2). The final step, then, is to compute the spectral phase of the measured pulse from these finite differences. This step is common to all spectral shearing methods, and has been discussed elsewhere in the literature.

The most straightforward way is to simply treat the finite differences as proportional to the spectral group delay, numerically integrating them using standard methods and accepting the $O(\Omega^2)$ error. However, this does not result in the most accurate measure-

ment possible, as the truncation error due to the finite difference is actually reversible. One can regard a finite difference as approximating the continuous derivative operator with a two term discrete time FIR filter. The amplitude transfer function of this effective FIR filter is sinusoidal in the pseudotime domain, peaking at half the inverse period of the finite difference (half the shear Nyquist rate). Thus, it underestimates the magnitude of phase oscillations, with the underestimation increasing as the oscillation period decreases. In terms of pulse reconstruction, this means that the amplitude of satellite pulses will be underestimated the further away they are from the main pulse.

A rigorous approach to the phase reconstruction is to compute the inverse of the finite difference operation, yielding the phase without any errors other than those caused by noise. To do so, the data must first be antialiased by filtering out all ω-domain oscillations with a period below Ω. Then, a sinc interpolation can be used to compute the fringe phase ϕ on a regular grid of points spaced by Ω. A simple cumulative sum will then yield the spectral phase of the pulse. It may seem that this approach is not optimal in terms of noise performance, as the data will be concatenated at a lower resolution than that provided by the spectrometer, seemingly throwing out data points that are being skipped over. However, the antialiasing filter step provides the averaging in this case. By suppressing all noise beyond the shear's Nyquist limit, the final filtered phase data will be made internally consistent such that the concatenation operation will yield the same result regardless of the starting point.

9.8 Design Considerations

The construction and alignment of 2DSI presents no challenges beyond that required for any Michelson interferometer and sum frequency generation. The experimental setup is essentially the same as a conventional SPIDER, with the main differences being the location of the dispersion and the addition of motion control to one delay stage. Care should be taken to ensure relative collinearity of the chirped beams with respect to each other. Fortunately, any deviation from this should not result in a spurious spectral phase, as discussed in Section 9.3. In fact, all conceivable misalignments, to the best of our knowledge, will simply result in attenuation of signal, not errors.

When engineering a 2DSI setup, there are essentially six design issues that must be decided:

1. How much shear Ω to use
2. How far to scan and how linear it must be made
3. How much dispersion D_2 to use for chirping
4. How often to measure the spectrum during the scan
5. The required spectrometer resolution
6. The optimal nonlinear crystal thickness

These six issues are not independent, and are all determined by the bandwidth and temporal extent T of the measurement. All but 2 apply to all other spectral shearing methods, as well. Each is discussed individually in the following subsection.

9.8.1 Shear Frequency

The shear frequency ultimately becomes the "sampling period" of the phase in the spectral domain, and most of the other parameters can be derived from it. By the Shannon sampling theorem, this determines the temporal window over which we can reliably measure,

$$T \leq \frac{2\pi}{\Omega} \tag{9.5}$$

We will refer to this time period as the shear's Nyquist duration. While it may be tempting to simply pick a shear sufficient to include a region of interest, such as the main pulse, this is not sufficient. Any satellite pulses or pedestal structure will still be measured, but will be aliased into our chosen temporal window, resulting in errors. Thus, the shear must be chosen so as to allow resolution of all spectral features. Moreover, satellite pulses which may seem negligible can have significant effects when aliased onto the main pulse, due to the fact that aliases add in field. Our numerical simulations have shown that a satellite pulse with an intensity of only 2% of the main pulse can change the FWHM of the main pulse by up to 5.5% should it be aliased on top of it.

The required temporal measurement window T is not known a priori, by definition, in a pulse that we are seeking to characterize. It is tempting to simply conclude the sampling was sufficient by taking a measurement and verifying that the reconstruction is well contained with the time window. However, the nature of aliasing is that once sampling occurs, one cannot tell the difference between correctly or insufficiently sampled signals. The only reliable way to assure that the sampling rate is sufficient is to take measurements at a series of decreasing shears, and verify that the measurements converge to required precision.

In practice, for sub-two-cycle pulses produced by oscillators with dispersion compensating mirrors, we have found that a shear of around 4-âĂŞ5 THz is required to sufficiently resolve the satellite pulses and pedestal. Unfortunately, this appears to be independent of the final pulse bandwidth, as one would expect given that the mechanisms for creating them are independent of bandwidth.

9.8.2 Chirping Dispersion

The signal of the final measurement is inversely proportional to the amount of dispersion used to create the quasi-CW beams, and thus the chirp of the ancillary pulses should be the minimal amount required to result in an accurate measurement. Over-chirping results in an unnecessarily weak signal, and under-chirping results in a complicated blurring of the measurement in the spectral domain as the upconversion occurs with a range of interacting wavelengths. Having determined the required spectral resolution with the shear, the chirping should be selected such that the associated blur is smaller than this resolution.

In the limit of large chirping (which is a prerequisite, at any rate) such that the dispersed pulse width is much larger than the original short pulse, the change $\Delta\omega_{\text{cw}}$ in the chirped pulse's instantaneous frequency over the temporal measurement window T is given by $\Delta\omega_{\text{cw}} \approx T/D_2$. We require that the chirped pulse's instantaneous frequency does not change by more than the shear over the entire temporal window, so $\Delta\omega_{\text{cw}} < \Omega$. In light of (9.5), this implies that we must choose

$$D_2 \geq \frac{2\pi}{\Omega^2}. \tag{9.6}$$

For example, if 5 THz of shear is used, the dispersion required is at least 80,000 fs^2. To give a sense of scale, this is the GDD provided by roughly 40 cm of SF11 glass. This is quite a bit of dispersion, and will significantly impair the signal power.

However, equation (9.6) presumes that we care about properly measuring the field over the entire temporal window. In most cases, we only care about accurately measuring a main pulse. (This is consistent with our previous statement that all satellite structure must be resolved; while we may not be interested in the accurate measurement of satellite pulses, we still must ensure that they do not alias.) If a well-separated satellite pulse sees shifted local frequencies of the chirped pulses, it will simply result in a local error in the reconstruction of that feature in proportion. Thus, a more reasonable criterion is simply for the frequency of the CW beam to change by no more than the shear over the temporal extent of the main pulse. Taking this width to be T_p, this gives

$$D_2 \geq \frac{T_p}{\Omega}. \tag{9.7}$$

Since the optimal shear cannot be known a priori (as explained above) neither can the dispersion be determined without knowing the temporal extent of the pulse, exactly the thing we seek to measure. Ideally, one would iteratively increase the chirping as with the shear, until the measurement converged to some satisfactory precision. However, this is generally not feasible, as variable sources of dispersion with wide variability are not easily found and would be expensive regardless. As such, the chirp will have to be chosen somewhat conservatively when the 2DSI system is built, with the worst-case pulse width in mind and a conservative estimate for the smallest shear likely to be used. A safe configuration for a few-cycle laser can be found by assuming a shear of 5 THz, and a pulse width of no more than 25 fs, yielding a dispersion of 5000 fs^2. This is a small enough amount that it can be provided by the cube beamsplitter used in the chirped beam interferometer (as in Figure 9-2).

9.8.3 Delay Scan Length

The scan must be long enough such that the sidebands illustrated in Figure 9-4 are well separated from the central DC peak. This distance will vary depending on whether or not we are performing windowing, and how accurately the scan can be matched to the fringe period. In general, one will be fine as long as at least three fringes are visible, as derived in Section 9.7. Beyond this consideration, the length of the scan actually does

not matter. This implies scan lengths on the order of a micron or two, at most, allowing very stable short-throw piezoelectric stages to be used, and enabling high scan rates limited only by signal levels. In our setup, we have been able to achieve scan rates of several Hertz, limited by the interface speed of our piezo controller.

9.8.4 Scan Sample Rate

The number of points over which the scan is sampled is largely irrelevant so long as they are sufficiently sampled. Beyond that, doubling the number of samples but halving the integration time of each will result in identical measurements. The only thing that matters is the total measurement time. The exception to this is if the measurement is limited by readout noise, in which case keeping the samples to a minimum is advised. In such a situation, it is actually possible to get by with only four measurements .

Otherwise, a rate high enough to yield a human-readable spectrogram can be useful, such as that shown in Figure 9-3. This allows the user to visually gauge the functioning of both the laser and the 2DSI system.

9.8.5 Spectrometer Resolution

The required spectrometer resolution is simply that consistent with the sampling rate set by the shear. Any more will only result in excess measurement noise. Of course, the spectrometer resolution is not a readily changed variable, so we recommend choosing a spectrometer with a resolution of at least 2 THz, to ensure the apparatus can be used to measure any reasonable time-bandwidth product.

9.8.6 Nonlinear Crystal Thickness

Thanks to the phase encoding inherent to spectral shearing, the bandwidth of the nonlinear crystal only affects the measurement from a signal-to-noise standpoint. The uniformity of the upconversion is thus not an issue, and the ideal crystal is such that the minimum conversion efficiency is maximized. In most cases, this entails choosing the crystal angle that maximizes the minimum conversion over the expected bandwidth, regardless of the resulting variation of efficiency across the spectrum.

As mentioned earlier, a unique aspect of spectral shearing interferometry is that the nonlinear operation is between a signal with high bandwidth (the pulse under test) and one which is nearly monochromatic (the chirped beams). This allows one to take advantage of the inherent asymmetry in Type II upconversion, wherein one polarization will have greater bandwidth than the other. In the case of BBO, the dispersion works especially favorably for Type II upconversion in the NIR, such that bandwidths exceeding an octave can be efficiently upconverted with crystals of reasonable thickness. This, combined with the natural noise resilience of spectral shearing, helps to make it up for the relatively low optical efficiency of SPIDER methods. To optimize the conversion efficiency, a full non-collinear phase matching curve must be considered for the specific wavelength ranges to be used. An example of such a curve is shown in Figure 9-5.

9.9 Sensitivity to Noise

As illustrated in Figure 9-4, the sidebands in 2DSI are spectrally compact. The fringe is simply an impulsive line in the 2D Fourier domain, the sharpness of which is limited by the scan length. This spectral compactness means that the bulk of any noise will not interfere with the signal, maximizing SNR. A significant amount of noise immunity is gained by isolating the sideband assuming the noise is uniformly distributed over all frequencies. Furthermore, the information is contained in the phase of the sideband, not its intensity, further reducing the sensitivity to noise.

An important question to answer is whether or not 2DSI's need to scan over a delay renders it more or less sensitive to noise than standard SPIDER. Intuitively, one might expect that for a given total integration time, they would be comparable in light of our previous assertion that 2DSI simply takes the SPIDER sidebands and moves them in a new direction. To illustrate the noise immunity of 2DSI (and in fact spectral shearing in general) we simulated the measurement of a pulse with a rectangular spectrum of 300 THz bandwidth, whose transform-limited pulse width is roughly 3 fs. Additive gaussian noise was simulated, with a SNR of 0.5. The SPIDER method was simulated using the same noise source, with an integration time equal to the total measurement time of the 2DSI trace. The SPIDER calibration method used was that of [94], where a calibration measurement with exactly zero shear was assumed to be available that exactly preserved the delay.

The results are shown in Figure 9-6. For a given measurement time, 2DSI has half the variance of SPIDER. The difference between the two is entirely due to uncertainty introduced into the SPIDER calibration by the detector noise. Furthermore, it is apparent from the middle plot in Figure 9-6 that the SPIDER error is predominantly composed of a second-order term, validating our consideration of the calibration as largely manifesting as an effective delay uncertainty. Both methods recover the spectral phase well, despite the low signal to noise. Note, also, that the 2DSI reconstruction involved no knowledge of the fringe spacing by the algorithm.

If the SPIDER calibration noise issue is ignored, the two methods perform identically, regardless of the type of noise used (i.e. additive or shot). This is to be expected, given that 2DSI simply takes the SPIDER sidebands and moves them into another dimension. From a signal processing point of view, a SPIDER measurement is equivalent to a 2DSI measurement where the upconversion phase is constant and the sidebands are created by a temporal delay. It is thus to be expected that the two methods would perform similarly outside of calibration issues.

9.10 Experimental Demonstration

9.10.1 Precision Test

To gauge the relative precision of the method, a few-cycle (5 fs FWHM) pulse from a prismless Ti:sa laser was measured using a 2DSI setup similar to that shown in Figure 1, using a shear of 18 THz. The pulse was measured both before and after dispersion by a

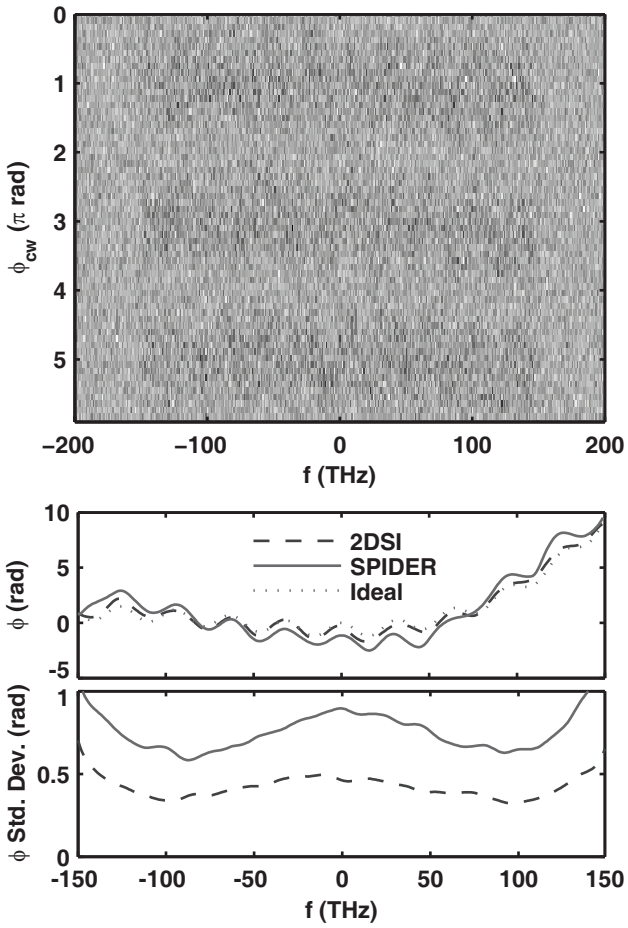

Figure 9-6: Simulated 2DSI spectrogram (top) measured with 64 phase steps for a sinc pulse with second- and third-order dispersion and a satellite, measured in the presence of additive and shot noise, such that the resulting SNR per spectrum is 0.5. A sample reconstruction, including comparison with SPIDER is shown in the middle frame. The bottom frame shows the standard deviation of the phase measurement for both SPIDER and 2DSI, showing that the lack of delay calibration in 2DSI yields a factor of two improvement in noise performance.

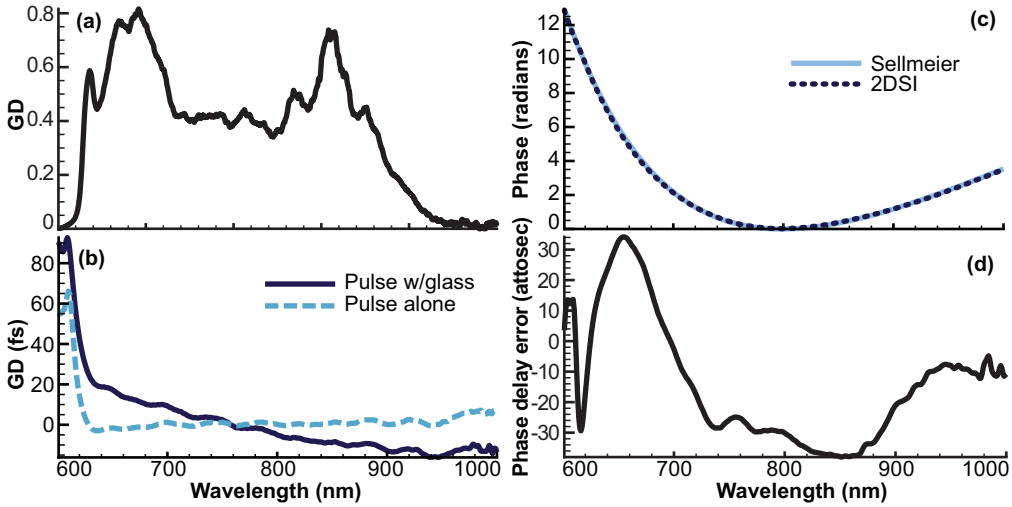

Figure 9-7: (a) Spectrum of 5 fs laser used in test; (b) extracted group delay both with and without glass slide; (c) Phase of glass slide as measured by 2DSI and as predicted by known glass dispersion; (d) Net phase delay error in glass dispersion measurement.

one mm fused silica plate, Figure 9-7(b). It is apparent from the spectral GD curves that the pulse is initially slightly negatively chirped, and the positive dispersion introduced by the glass plate is evident. The sharp roll-off in the GD below 650 nm is genuine, and caused by phase distortion from the output coupler. Oscillations in the spectral group delay, caused by the chirped mirrors, are also clearly visible. Despite these perturbations in the individual spectra, the oscillations completely cancel in the difference between the phase of the two measurements, which matches well to that predicted by the known Sellmeier equations for fused silica, as shown in Figure 9-7(c). In fact, in terms of phase delay, the 2DSI system measured the glass dispersion to within 30 attoseconds of phase delay over a bandwidth from 600 to 1000 nm, Figure 9-7(d). This precision was achieved despite the absolute phase delay of each measurement ranging over more than 40000 as.

While these results suggest that the 2DSI apparatus is capable of precise measurements, they do not rule out the possibility of constant errors occurring that are consistent between measurements. For example, if there were an unknown linear phase creeping in the measurement somehow such a systematic error would not be evident from these dispersion measurements.

9.10.2 Accuracy Test

To qualitatively demonstrate the absolute accuracy of the system and rule out the existence of systematic errors, we recently performed a measurement on an octave spanning sub-two-cycle pulse [93, 98] from another unamplified Ti:sapphire oscillator, and compared the 2DSI measurement to that obtained with a standard IAC (Figure 9-8).

In Figure 9-8(d), we show the reconstructed pulse envelope and phase for the pulse, measured to have a full width half maximum of 4.9 fs. To our knowledge, this is the

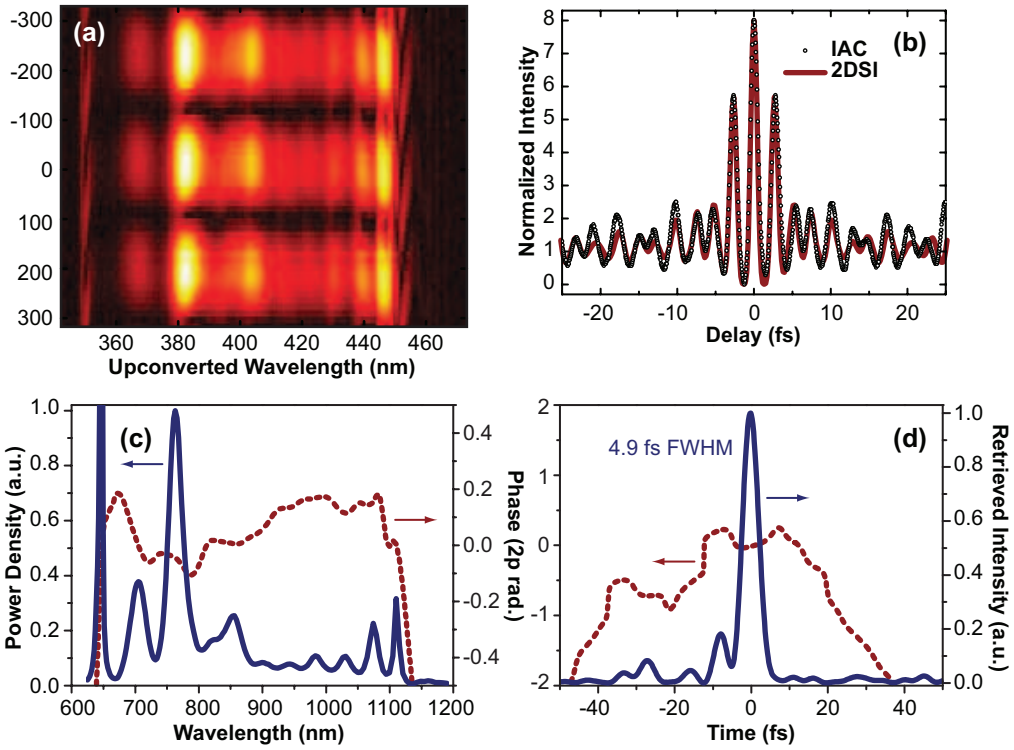

Figure 9-8: (a) Raw 2DSI data; (b) comparison of IAC and that predicted from the 2DSI measurement; (c) Extracted spectral phase (dashed); (d) Reconstructed pulse (solid), simulated pulse (dotted) and temporal phase (dotted).

shortest pulse measured with a spectral shearing method directly from an oscillator (i.e. without use of an external amplifier). As shown in Figure 9-8(b), the measured IAC and that predicted by the 2DSI measurement show fairly close agreement. We attribute most of the difference between the two to band-limiting effects on the IAC, which is not well suited to measuring a 4.9 fs pulse (note, for example, the lack of symmetry in the IAC trace). Thus, we do not present this measurement as further evidence of the precision of 2DSI, as an IAC is not a particularly reliable measurement of fine detail. However, the fact that the overall pulse widths predicted by both are consistent does suggest that we are correct in assuming that no appreciable hidden linear phase can occur in 2DSI. This validates the assertion that 2DSI does not require a separate calibration step. This has also been corroborated by detailed simulations of the laser cavity in question, which also predict a 4.9 fs pulse with a large sub-pulse 9 fs away from the peak (see [52] for details).

9.11 Future Work

The geometry of 2DSI is unique in that the pulse to be measured never encounters a dispersive element, and yet the arrangement is still collinear. Were the spectrometer replaced by an imaging spectrometer, a spatially resolved 2DSI measurement could be taken along one axis. Moreover, if a full 2D imaging spectrometer were used, such as that available with imaging Fourier transform spectroscopy or grisms, one could make a full 2DSI measurement at a 2D array of points along both transverse axes. By spatially filtering the chirped beams so that they were spatially coherent, they would provide a constant phase reference across the beam profile, enabling a full spatiotemporal reconstruction up to a trivial constant and quadratic spatial phase (focusing). This would allow for the first self-referenced 3D measurement of few-cycle pulses.

By using a nanostructured stepped mirror, a single-shot version of 2DSI might be implemented. The differential upconversion phase would then be encoded as a function of space, and the fringe could be read using an imaging spectrometer. This arrangement would result in the mixing of the spatial profile with the fringe, essentially creating a collinear variation of SEA-SPIDER with a spatially varying zeroth-order phase as opposed to linear phase. This could provide similar advantages to SEA-SPIDER, but with potentially simpler calibration and alignment.

9.12 Summary

Two-dimensional spectral shearing interferometry involves a relatively simple optical setup with little calibration required, and yet is capable of spectral phase measurements accurate to within tens of attoseconds of phase delay over octaves of bandwidth. The lack of dispersion of the pulse to be measured, the absence of delay between the sheared pulses, and the relaxed spectrometer resolution requirements make 2DSI extremely well suited for the measurement of wide-bandwidth pulses, including those with potentially complicated spectral phase.

Appendix A

Derivation of Worst-Case FFT Harmonic Inversion Error

We derive the relative error in recovering the phase of a pure cosine fringe by taking the phase of the dominant FFT component, as discussed in Section 9.7. The FFT will be approximated by a continuous integral over the space $[0, 2\pi]$, as it is the windowing, not the discretization, that we are concerned with. We consider a fringe with a frequency of $n + \Delta\nu$, where n is an integer, and $\Delta\nu$ is the deviation from that integer frequency. In other words, n is the approximate number of complete fringes contained within the scan, and corresponds to the index of the dominant FFT component we are to consider. Without loss of generality we will consider a cosine fringe with phase offset ϕ. The complex value f_n of the nth FFT component will be approximately given by the continuous integral

$$f_n \approx \int_0^{2\pi} dx \cos[(n+\Delta\nu)x + \phi] e^{-inx}. \tag{A1}$$

Performing this integral, computing the argument of the complex value and simplifying the expression gives the extracted fringe phase

$$\begin{aligned} \phi_{\text{ext}} &= \arctan \frac{\Im f_n}{\Re f_n} && \text{(A2)} \\ &= \arctan \left[\frac{n \tan[\pi\Delta\nu + \phi]}{n + \delta\nu} \right] && \text{(A3)} \end{aligned}$$

The above expression gives us the absolute phase recovered from the FFT, which is proportional to the spectral group delay of the pulse we're measuring. Since constant offsets to GD are irrelevent, we are really only concerned with the slope of the relation between phi_{rec} and ϕ. This is the factor by which fringe phase changes will be magnified. It is thus the relative error in our extracted dispersion at a given wavelength,

$$\begin{aligned} \epsilon_r(\omega) &= 1 - \left. \frac{\partial \phi_{\text{ext}}}{\partial \phi} \right|_\omega && \text{(A4)} \\ &= 1 - \frac{n(n+\Delta\nu)\sec[\pi\Delta\nu + \phi]^2}{(n+\Delta\nu)^2 + n^2 \tan[\pi\Delta\nu + \phi(\omega)]^2}. && \text{(A5)} \end{aligned}$$

Since we cannot control the offset of ϕ, we must consider the worst-case error over all ϕ. We thus solve for the stationary point of (A5). This yields the equation

$$\frac{2n\Delta\nu(n+\Delta\nu)(2n+\Delta\nu)\sec[\pi\Delta\nu+\phi]^2\tan[\pi\Delta\nu+\phi]}{\left((n+\Delta\nu)^2+n^2\tan[\pi\Delta\nu+\phi]^2\right)^2}=0. \tag{A6}$$

There are multiple roots of (A6), but the one that applies for $\Delta\nu$ near zero is $\phi=-\pi\Delta\nu$. Plugging this solution in (A5) and simplifying gives us a final expression for the worst case relative measurement error due to the FFT harmonic inversion,

$$|\epsilon(\omega)|\leq\frac{\Delta\nu}{n+\Delta\nu}. \tag{A7}$$

Appendix B

Newton-Krylov Solver Code

For reference, a copy of the most important parts of the Newton-Krylov MATLAB code are detailed below.

B.1 Solver

```
function [Ufinal, converged, Fnorms] = ...
  cavitysolverls(U0, cav, cavfun, tol, maxiter, ngcr, diags)
% CAVITYSOLVERLS Computer steady state solution to NL cavity.
% BFGS. Experiment with Jacobian computation in step function, perhaps averaging
% before and after parameters. Try using striped preconditioner, assuming
% decomposition of such a matrix is O[n^2]. Base
% termination of GCR on direction stabilizing, not length.
% Is there a way to make this
% parallel? (I suppose not, since Krylov spaces are defined in terms of
% matrix powers.)                                                          10

% Initialization and parameters.
a = mean(abs(U0))/10000; % perturbation used in finite difference
s = 1000; % ORTHOMIN restart?
f = cav.f;
n = length(f);
if nargin < 5
  maxiter = 10;
elseif nargin < 6 || ngcr == 0
  ngcr = 64; % max GCR iterations                                          20
  autogcr = true;
else
  autogcr = false;
end
if nargin < 7
  diags = false;
end
```

```
if nargout > 2
  Fnorms = zeros(1,2);
end
k = 0;
fn = 0;
p = zeros(n, ngcr);
Jp = zeros(n, ngcr);


% *** Newton Iterations. ***
U = U0;
[Ucav, Uout, Jdiag] = cavfun(f, U, cav, 'norm'); fn = fn + 1;
Binv = diag(1./(1 - Jdiag));
BinvF = Binv*U - Binv*Ucav; % what we want to set to zero
F = U - Ucav;
Fnorm = norm(F)/norm(U);
if nargout > 2
  Fnorms(1) = Fnorm;
end
kiter = 1;
converg = false;
done = false;

% Diagnostics.
if diags
  nl = 45;
  fprintf([repmat('=', 1, nl) '\n'])
  fprintf('iter\tevals\t|du| (log)\t|F| (log)\n')
  fprintf([repmat('-', 1, nl) '\n'])
  fprintf('%d\t%d\t%f\t%f\n', k, fn-1, 0, log10(Fnorm))
end

while ~done % *** Newton ***
  m = 0;
  r = -BinvF; % null starting vector
  dU = zeros(n,1);
  gcrdone = false;
  while (m < ngcr) && ~gcrdone, % *** GCR ***
    m = m + 1;
    kiter = kiter + 1;

    p(:,m) = r; % use residual as search direction

    % Compute Approximate Binv*J*p using finite differences.
    d = a/norm(p(:,m));
    dUp = d*p(:,m);
    Ucavdp = cavfun(f, U + dUp, cav, 'norm'); fn = fn + 1;
```

```
  Jp(:,m) = Binv*(p(:,m) + (Ucav - Ucavdp)/d);

  % Make the new Jp vector orthogonal to the s most recent Jp vectors.
  for j = max(1,m-1-s):m-1,
    beta = real(Jp(:,m)' * Jp(:,j)); % projection
    p(:,m) = p(:,m) - beta*p(:,j); % subtract out orthogonal parts
    Jp(:,m) = Jp(:,m) - beta*Jp(:,j); % "
  end

  % Make the orthogonal Jp vector of unit length.
  Jpnorm = norm(Jp(:,m),2);
  Jp(:,m) = Jp(:,m)/Jpnorm;
  p(:,m) = p(:,m)/Jpnorm;

  % Determine the optimal amount to change x in the p direction
  % by projecting r onto Mp.
  alpha = real(r' * Jp(:,m));

  % Update x and r
  dU = dU + alpha*p(:,m);
  r = r - alpha*Jp(:,m);

  % Automatically terminate.
  if autogcr
    if abs(alpha)/norm(dU) < Fnorm
      gcrdone = true;
    end
  end

  % These don't count, since it's just for diagnostics.
  if nargout > 2
    Utest = cavfun(f, U + dU, cav, 'norm');
    Fnorms(kiter) = norm(U + dU - Utest)/norm(U);
  end
end % *** GCR ***

k = k + 1;
kiter = kiter + 1;

% Do quadratic line search with first three Newton iterates.
b = 2; % distance along line of second point (2 works well)
nF = norm(F);
[Ucav1, Uout, J1] = cavfun(f, U + dU, cav, 'norm');
Fnorm1 = norm(U + dU - Ucav1);
[Ucav2, Uout, J2] = cavfun(f, U + b*dU, cav, 'norm');
Fnorm2 = norm(U + b*dU - Ucav2);
p2 = polyfit([0 1 b], [nF Fnorm1 Fnorm2], 2);
```

```
  rmin = roots(polyder(p2));
  [Ucavls, Uout, Jls] = cavfun(f, U + rmin*dU, cav, 'norm');
  Fnormls = norm(U + rmin*dU - Ucavls);
  [Fmin, kmin] = min([Fnorm1 Fnorm2 Fnormls]);
  dU = take([dU b*dU rmin*dU], kmin);
  Ucav = take([Ucav1 Ucav2 Ucavls], kmin);
  Jdiag = take([J1, J2, Jls], kmin);
  if diags
    fprintf('line search: option %d\n', kmin)                    130
  end

  % Handle update.
  fn = fn + 1; % two could be done completely in parallel, one is diagnostic
  U = U + dU;
  dUnorm = norm(dU);
  %[Ucav, Uout, Jdiag] = cavfun(f, U, cav, 'norm'); fn = fn + 1;
  Binv = diag(1./(1 - Jdiag));
  BinvF = Binv*U - Binv*Ucav;
  F = (U - Ucav);                                                140
  Fnorm = norm(F)/norm(U);

  converg = Fnorm < tol;
  done = converg || (k > maxiter);

  if nargout > 2
    Fnorms(kiter) = Fnorm;
  end

  % Output                                                       150
  if diags
    fprintf('%d\t%d\t%f\t%f\n', k, fn-1, log10(dUnorm), log10(Fnorm))
  end
end % *** Newton ***
if diags
  fprintf([repmat('=', 1, nl) '\n'])
end

if nargout > 1
  if converg                                                     160
    converged = fn;
  else
    converged = 0;
  end
end

Ufinal = U;
```

B.2 Cavity round trip function

```
function [Ucav, Uout, Jdiag] = ringcavitystep(f, Uin, cavparams, norm, diag)
% RINGCAVITYSTEP Propagate soliton-like laser cavity for m steps.
% Assumes length(ts) is an even number (best if a power of 2).
% Optionally returns the diagonal of the Jacobian in Jdiag, for use in
% preconditioning solvers.
%
% This is a fast version of the solitoncavityfd function that is meant to
% be called from other functions, notably cavnewtonfd. It uses column
% vectors and only computes one step.

% Cavity parameters. %
Harm1 = cavparams.Harm1; % Linear arm transfer function
Harm2 = cavparams.Harm2; % ""
g0 = cavparams.g0re; % Small signal gain amplitude
gainphi = cavparams.g0im; % Small signal gain phase
TR = cavparams.TR; % cavity roundtrip time (fs from MHz)
Psat = cavparams.Psat; % gain saturation power (W)
q = cavparams.q; % saturable absorber gain per RT
Isat = cavparams.Isat; % saturation intensity of absorber (W)
oc = cavparams.l; % output coupler gain spectrum
if ~isfield(cavparams, 'nstep')
  nstep = 32;
else
  nstep = cavparams.nstep;
end
if nargin > 3
  if strcmp(norm, 'norm')
    normed = true;
  else
    normed = false;
  end
else
  normed = false;
end
if nargin < 5
  diag = false;
end

% Memory allocation and precalculations. %
n = length(f);
df = abs(f(2) - f(1));
dt = 1/df/n;
A = dt/TR; % average power integral scaling
nlparams.d = cavparams.d; % SPM coefficient of nl material (1/W)
nlparams.phi = gainphi;
nlparams.Isat = cavparams.Isat;
```

```
nlparams.q = cavparams.q;

% Calculation. %

% Arm 1
U = Harm1.*Uin;

% Nonlinear gain material with SA
P = real(U'*U)*A/n;
nlparams.g0 = g0/(1 + P/Psat); % ss gain for RT
U = nlpropgen(U, nlparams, floor(nstep), diag);

% Arm 2
U = Harm2.*U;

% OC
if nargout > 1
  Uout = (1 - abs(oc).^2).*U;
end
U = oc.*U;

% Phase normalization and output.
if normed
  [Ucav, dphi] = altphasenorm(f, Uin, U);
else
  Ucav = U;
end

% Compute approximate Jacobian diagonal.
if nargout >2
  u = ifft(U);
  Iu = real(u.*conj(u)); % intensity (W)
  dnl = mean(exp(-1j*cavparams.d*Iu));
  sa = mean(exp(q./(1 + Iu/Isat)));
  Jdiag = dnl .* sa .* oc .* Harm1 .* Harm2 .* ...
    exp(g0/(1 + P/Psat) + 1j*gainphi);
  if normed
    Jdiag = Jdiag .* exp(1j*dphi);
  end
end
```

B.2.1 Nonlinear propagation

```
function Uout = nlpropgen(Uin, nlparams, m, diag)
% nlpropfd Propagate NLSE through material broken into m steps.

% Input %
```

```
if nargin < 4
  diag = false;
end

% Cavity parameters %
g0 = nlparams.g0; % total small signal gain
phi = nlparams.phi;
d = nlparams.d; % SPM coefficient (1/W)
q = nlparams.q; % SA
Isat = nlparams.Isat;

% Memory allocation and precalculations %
%w = 2*pi*f;
H = exp((g0 + 1j*phi)/m);
H2 = exp((g0 + 1j*phi)/m/2);

% Calculate m steps. %
% Initial frequency domain:
U = H2.*Uin;
for k = 1:m,
  % Time domain:
  u = ifft(U);
  Iu = real(u .* conj(u)); % intensity (W)
  u = exp(q./(1 + Iu/Isat)/m + 1j*d*Iu/m) .* u;
  U = fft(u);

  % Output:
  if diag
    fprintf('nlprop: %d/%d, SPM: %f, cumSPM: %f\n', k, m, max(d/m*Iu), max(d*Iu));
  end

  % Frequency domain:
  if k == m
    Uout = H2.*U;
  else
    U = H.*U;
  end
end
```

B.2.2 Phase normalization

```
function [Uout, dphiout] = altphasenorm(f, Uin, U)
% Minimize the change in phase between Uin and Uout by adding trivial
% phases to U. The idea is that this forces the cavity Jacobian to be as
% close to the identity as possible, enhancing convergence.

phiUin = unwrap(angle(Uin));
```

```
phiU = unwrap(angle(U));
weff = abs(Uin).*abs(U);
Weff = sum(weff);
omegaWeff = sum(weff.*f);                                          10
omega2Weff = sum(weff.*f.^2);
M = [Weff omegaWeff; omegaWeff omega2Weff]; % always a constant
phierr0 = phiU - phiUin;
b = [sum(phierr0.*weff); sum(phierr0.*weff.*f)];
x = M\b;
phi0 = x(1);
phi1 = x(2);
dphi = phi0 + phi1*f;
Uout = U.*exp(-1j*dphi);
                                                                   20
if nargout > 1
  dphiout = -dphi;
end
```

B.3 Preconditioner

```
function d = cavityprecond(f, U, cav)
% CAVITYPRECOND  Compute diagonal preconditioner
%   Compute diagonal preconditioner for the cavity given by cav, around the
%   point given by the Fourier component vector U, at the frequencies given
%   in f.
%
%   It would be slightly more efficient to compute this in
%   solitoncavitystep, but is put here for the sake of

n = length(U);                                                     10
dt = 1/f(2)/n;
w = 2*pi*f;
u = ifft(U);
oc = sqrt(cav.l);
P = (u,*u)*dt/cav.TR;
Iu = real(u .* conj(u));
dgsp = cav.g0*lorentznorm(f, cav.Wg)/(1 + P/cav.Psat);
phinet = -j*(cav.D2net*w.^2/2 + cav.D3net*w.^3/6);
d = oc.*exp(mean(j*cav.d*Iu + cav.q./(1 + Iu/cav.Isat)) + phinet + dgsp);
```

Appendix C

PDR Gradient Code

Here, we provide sample code for computing the analytic gradient of the PDR from gradients of group delay, as discussed in Chapter 4. We believe computing the GD and then converting to PDR is the most reliable way to implement PDR-based optimization.

C.1 Single mirror PDR Error

```
function [Z, Zgrad] = mirrorpdrerr(...
  ds, ks, n0, ns, dns, theta, pol, gdgoal0)
% Computes error energy for reflection for mirror ds is the *square
% root* of the thickness vector. This ensures that we cannot have a
% negative layer.

c = .2997924580; % um/fs

nk = length(ks);
n1 = length(ds)/2;

% Compute matrix which solves for floating phases
weff = ones(size(ks));
Weff = sum(weff);
omega = c*ks;
omegaWeff = sum(weff.*omega);
omega2Weff = sum(weff.*omega.^2);
M = [Weff omegaWeff; omegaWeff omega2Weff]; % always a constant
Minvt = inv(M).';

gdgoal = gdgoal0 - sum(weff.*gdgoal0)/Weff;
c = .2997924580; % um/fs
phigoal = c*cumtrapz(ks, gdgoal);

% Compute GD and GD gradients
[r, gd, Jr, Jgd] = stackgdgradmex(ks, ds.^2, n0, ns, dns, theta, pol);
```

```
gd = gd - mean(gd);

rserr = 1 - r;

phi = cumtrapz(omega, gd);
phierr0 = phi - phigoal;
b = [sum(phierr0.*weff), sum(phierr0.*weff.*omega)];
x = b*Minvt;
phi0 = x(1);
phi1 = x(2);
phierr = phierr0 - phi0 - phi1*omega;

Z = sum(weff.*phierr.^2) + sum(rserr);

Jphi = cumtrapz(omega, Jgd, 2);
bgrad = [Jphi*weff.', Jphi*(weff.*omega).'];
xgrad = bgrad*Minvt;
phi0grad = xgrad(:,1);
phi1grad = xgrad(:,2);
Jphierr = Jphi - phi0grad*ones(1,nk) - phi1grad*omega;

gradphi = Jphierr*(weff.*phierr).'
gradr = sum(Jr,2);
Zgrad = 2*ds.*(gradphi + gradr);
```

Appendix D

Thin Film Gradient Code

This is the code which is central to dispersion compensating mirror optimization, where the group delay and group delay gradient is computed at a series of wavelengths.

D.1 Stack GD Gradient MEX

```
/*==========================================================================
 *
 * stackgdgradmex.c
 *
 * Requires a C99 compiler.
 *
 *========================================================================*/


#include <math.h>
#include <complex.h>
#include <string.h>
#include "mex.h"


/* Input Arguments */
#define KS_IN          prhs[0]
#define DS_IN          prhs[1]
#define N0_IN          prhs[2]
#define NS_IN          prhs[3]
#define DNS_IN         prhs[4]
#define THETA_IN       prhs[5]
#define POL_IN         prhs[6]

/* Output Arguments */
#define R2_OUT         plhs[0]
#define GD_OUT         plhs[1]
```

```
#define R2GRAD_OUT plhs[2]
#define GDGRAD_OUT plhs[3]

/* Definitions */
#define C 0.2997924580


/* Core computation routine */
static void stackgdgrad(
    const int nk,
    const int n,
    const double ks[],
    const double ds[],
    const double n0,
    const double ns[],
    const double dns[],
    const double theta,
    const int isTM,

    double r2[],
    double gd[],
    double r2grad[],
    double gdgrad[])
{
    /* Variables */
    int dx, kx;
    double neffs[3], dneffs[3], dterms[3], kneffs[3];
    double knefflay[n+1];
    complex dTdkgradterm, dTdkgradmat1[n+1], dTdkgradmat2[n+1];
    double pps[4], pms[4], dps[4];
    const int lastx = (n+1) & 1; /* last material index index */
    complex ephi;
    complex Dlay[n+1], Tlay1[n+1], Tlay2[n+1], dTlay1[n+1], dTlay2[n+1];
    complex Tfor1[n+1], Tfor2[n+1], dTfor1[n+1], dTfor2[n+1];
    complex Trev1[n+1], Trev2[n+1], dTrev1[n+1], dTrev2[n+1];
    complex dTgradfor1[n], dTgradfor2[n];
    complex T1, T2;
    complex R, dR;
    double r, dr, dphi;
    complex Rgrad, dRgrad;

    /* Common terms and offset vectors. */
    const double n0sintheta2 = (n0*sin(theta))*(n0*sin(theta));
    const double *nsofk = ns, *dnsofk = dns; /* offset index pointers */
    double *r2gradofk = r2grad, *gdgradofk = gdgrad;

    /* Initialization. */
```

```
dTlay1[n] = 0; dTlay2[n] = 0;
dTrev1[n] = 0; dTrev2[n] = 0;

/*
 * Loop over all wavenumbers.
 */
for (kx = 0; kx < nk; kx++)
{
   /*
    * Precalculate all material parameters.
    * index variables: [n1, n2, nsub]
    * p variables: [p01, p12, p21, p2sub]
    */
   {
      register int j;
      double pTEs[4], dpTEs[4];

      /* Calculate effective indices and common expressions. */
      for (j = 0; j < 3; j++)
      {
         neffs[j] = sqrt(nsofk[j]*nsofk[j] - n0sintheta2);
         dneffs[j] = nsofk[j]*dnsofk[j]/neffs[j];
         dterms[j] = neffs[j] + ks[kx]*dneffs[j];
         kneffs[j] = ks[kx]*neffs[j];
      }
      /* Calculate pTEs (needed regardless of polarization). */
      pTEs[0] = n0*cos(theta)/neffs[0];
      pTEs[1] = neffs[0]/neffs[1];
      pTEs[2] = 1/pTEs[1];
      pTEs[3] = neffs[lastx]/neffs[2];
      dpTEs[0] = -pTEs[0]*dneffs[0]/neffs[0];
      dpTEs[1] = (dneffs[0] - pTEs[1]*dneffs[1])/neffs[1];
      dpTEs[2] = (dneffs[1] - pTEs[2]*dneffs[0])/neffs[0];
      dpTEs[3] = (dneffs[lastx] - pTEs[3]*dneffs[2])/neffs[2];

      /* Calculate TM or TE as needed. */
      if (isTM)  /* TM */
      {
         double ps, p0s[4], dp0s[4];

         /* Calculate pTM from p0 (normal incidence) and pTE. */
         p0s[0] = n0/nsofk[0];
         p0s[1] = nsofk[0]/nsofk[1];
         p0s[2] = 1/p0s[1];
         p0s[3] = nsofk[lastx]/nsofk[2];
         dp0s[0] = -p0s[0]*dnsofk[0]/nsofk[0];
         dp0s[1] = (dnsofk[0] - p0s[1]*dnsofk[1])/nsofk[1];
```

```
        dp0s[2] = (dnsofk[1] - p0s[2]*dnsofk[0])/nsofk[0];
        dp0s[3] = (dnsofk[lastx] - p0s[3]*dnsofk[2])/nsofk[2];
        for (j = 0; j < 4; j++)
        {
            ps = pTEs[j]/(p0s[j]*p0s[j]);
            dps[j] = (p0s[j]*dpTEs[j]-2*pTEs[j]*dp0s[j])/(p0s[j]*p0s[j]*p0s[j]);
            pps[j] = 1 + ps;
            pms[j] = 1 - ps;
        }
    }
    else  // TE
    {
        for (j = 0; j < 4; j++)
        {
            dps[j] = dpTEs[j];
            pps[j] = 1 + pTEs[j];
            pms[j] = 1 - pTEs[j];
        }
    }
}

/*
 * Precalculate all layer matrices.
 */
{
    int nx, px;  /* material data indices */

    for (dx = 0; dx < n; dx++)
    {
        /* Select appropriate material parameter index */
        if (dx & 1) {  /* even layer (mod(dx,2) == 1) */
            nx = 1;
            px = 1; }
        else if (dx == 0) { /* first layer */
            nx = 0;
            px = 0; }
        else {   /* odd layer */
            nx = 0;
            px = 2;
        }

        /* ephi = exp(I*d*k*neff)/2 */
        ephi = (cos(ds[dx]*kneffs[nx]) - I*sin(ds[dx]*kneffs[nx]))/2;
        Dlay[dx] = -I*ds[dx]*dterms[nx];
        Tlay1[dx] = ephi*pps[px];
        Tlay2[dx] = ephi*pms[px];
        dTlay1[dx] = Dlay[dx]*Tlay1[dx] + ephi*dps[px];
```

```
        dTlay2[dx] = Dlay[dx]*Tlay2[dx] - ephi*dps[px];
        knefflay[dx] = -kneffs[nx];
        dTdkgradterm = -(ds[dx]*kneffs[nx]+I)*(neffs[nx]+ks[kx]*dneffs[nx]);
        dTdkgradmat1[dx] = dTdkgradterm - I*kneffs[nx]*dps[px]/pps[px];
        dTdkgradmat2[dx] = dTdkgradterm + I*kneffs[nx]*dps[px]/pms[px];
    }
}
/* Propagation into substrate matrix. */
Tlay1[n] = pps[3]/2;
Tlay2[n] = pms[3]/2;
dTlay1[n] = dps[3]/2;
dTlay2[n] = -dps[3]/2;

/*
 * Step forward through structure, calculating forward matrices.
 */
Tfor1[0] = Tlay1[0]; Tfor2[0] = Tlay2[0];
dTfor1[0] = dTlay1[0]; dTfor2[0] = dTlay2[0];
dTgradfor1[0] = (1/ds[0] + I*knefflay[0])*dTfor1[0];
dTgradfor2[0] = (1/ds[0] + I*knefflay[0])*dTfor2[0];
for (dx = 1; dx < n+1; dx++)
{
    /* Calculate Lth forward matrix. */
    Tfor1[dx] = Tlay1[dx]*Tfor1[dx-1] + Tlay2[dx]*conj(Tfor2[dx-1]);
    Tfor2[dx] = Tlay1[dx]*Tfor2[dx-1] + Tlay2[dx]*conj(Tfor1[dx-1]);

    __builtin_prefetch(Tlay1+dx+1);
    __builtin_prefetch(Tlay2+dx+1);

    /* Calculate total k derivative of the Lth forward T matrix. */
    dTfor1[dx] = dTlay1[dx]*Tfor1[dx-1] + dTlay2[dx]*conj(Tfor2[dx-1]) +
        Tlay1[dx]*dTfor1[dx-1] + Tlay2[dx]*conj(dTfor2[dx-1]);
    dTfor2[dx] = dTlay1[dx]*Tfor2[dx-1] + dTlay2[dx]*conj(Tfor1[dx-1]) +
        Tlay1[dx]*dTfor2[dx-1] + Tlay2[dx]*conj(dTfor1[dx-1]);
}

/*
 * Step backward through structure, calculating reverse matrices.
 */
Trev1[n] = Tlay1[n]; Trev2[n] = Tlay2[n];
dTrev1[n] = dTlay1[n]; dTrev2[n] = dTlay2[n];
for (dx = n-1; dx > 0; dx--)
{
    __builtin_prefetch(Trev1+dx-1,1);
    __builtin_prefetch(Trev2+dx-1,1);
    __builtin_prefetch(Tlay1+dx-1);
    __builtin_prefetch(Tlay2+dx-1);
```

```
    Trev1[dx] = Trev1[dx+1]*Tlay1[dx] + Trev2[dx+1]*conj(Tlay2[dx]);
    Trev2[dx] = Trev1[dx+1]*Tlay2[dx] + Trev2[dx+1]*conj(Tlay1[dx]);

    dTrev1[dx] = dTrev1[dx+1]*Tlay1[dx] + dTrev2[dx+1]*conj(Tlay2[dx]) +
        Trev1[dx+1]*dTlay1[dx] + Trev2[dx+1]*conj(dTlay2[dx]);
    dTrev2[dx] = dTrev1[dx+1]*Tlay2[dx] + dTrev2[dx+1]*conj(Tlay1[dx]) +
        Trev1[dx+1]*dTlay2[dx] + Trev2[dx+1]*conj(dTlay1[dx]);
}

/*
 * Calculate full stack scalars at the current wavelength.
 */
T1 = Tfor1[n];
T2 = Tfor2[n];
R = -T2/T1;
r = cabs(R);
r2[kx] = r*r;
dR = (T2*dTfor1[n] - T1*dTfor2[n])/(T1*T1);
dr = (creal(dR)*creal(R) + cimag(dR)*cimag(R))/r;
dphi = (cimag(dR)*creal(R) - creal(dR)*cimag(R))/r2[kx];
gd[kx] = dphi/C;

/*
 * Step through structure, calculating gradients and output.
 */
{
    complex dTdkgradfor1, dTdkgradfor2;
    complex Tgrad1, Tgrad2, dTgrad1, dTgrad2;
    double rgrad, phigrad, dphigrad;

    for (dx = 0; dx < n; dx++)
    {
        Tgrad1 = I*knefflay[dx]*
            (Trev1[dx+1]*Tfor1[dx] - Trev2[dx+1]*conj(Tfor2[dx]));
        Tgrad2 = I*knefflay[dx]*
            (Trev1[dx+1]*Tfor2[dx] - Trev2[dx+1]*conj(Tfor1[dx]));

        /* if (dx > 0) */
        if (__builtin_expect(!!(dx > 0), 1))
        {
            dTdkgradfor1 = dTdkgradmat1[dx]*Tlay1[dx]*Tfor1[dx-1] +
                dTdkgradmat2[dx]*Tlay2[dx]*conj(Tfor2[dx-1]);
            dTdkgradfor1 += I*knefflay[dx]*(Tlay1[dx]*dTfor1[dx-1] +
            Tlay2[dx]*conj(dTfor2[dx-1]));
            dTdkgradfor2 = dTdkgradmat1[dx]*Tlay1[dx]*Tfor2[dx-1] +
                dTdkgradmat2[dx]*Tlay2[dx]*conj(Tfor1[dx-1]);
```

```
            dTdkgradfor2 += I*knefflay[dx]*(Tlay1[dx]*dTfor2[dx-1] +
            Tlay2[dx]*conj(dTfor1[dx-1]));
        }
        else
        {
            dTdkgradfor1 = dTdkgradmat1[dx]*Tlay1[dx];
            dTdkgradfor2 = dTdkgradmat2[dx]*Tlay2[dx];
        }                                                              270

        __builtin_prefetch(r2gradofk+dx, 1, 1);
        __builtin_prefetch(gdgradofk+dx, 1, 1);

        dTgrad1 = Trev1[dx+1]*dTdkgradfor1 + Trev2[dx+1]*conj(dTdkgradfor2);
        dTgrad2 = Trev1[dx+1]*dTdkgradfor2 + Trev2[dx+1]*conj(dTdkgradfor1);
        dTgrad1 += I*knefflay[dx]*(dTrev1[dx+1]*Tfor1[dx] -
        dTrev2[dx+1]*conj(Tfor2[dx]));
        dTgrad2 += I*knefflay[dx]*(dTrev1[dx+1]*Tfor2[dx] -
        dTrev2[dx+1]*conj(Tfor1[dx]));                                 280

        /* Final gradients */
        Rgrad = -(R*Tgrad1 + Tgrad2)/T1;
        rgrad = (creal(Rgrad)*creal(R) + cimag(Rgrad)*cimag(R))/r;
        r2gradofk[dx] = 2*r*rgrad;
        dRgrad = (Tgrad2*dTfor1[n] + Tgrad1*dTfor2[n] +
        R*(2*Tgrad1*dTfor1[n] - T1*dTgrad1) - T1*dTgrad2)/(T1*T1);
        phigrad = (cimag(Rgrad)*creal(R) - creal(Rgrad)*cimag(R))/r2[kx];
        dphigrad = (cimag(dRgrad)*creal(R) - creal(dRgrad)*cimag(R))/r2[kx] -
            (phigrad*dr + rgrad*dphi)/r;                               290
        gdgradofk[dx] = dphigrad/C;
      } /* gradient for loop */
    } /* gradient scope */

    /* Update offset vectors for next wavenumber. */
    nsofk += 3;
    dnsofk += 3;
    r2gradofk += n;
    gdgradofk += n;
  }                                                                    300

  return;
}


/* MEX function gateway routine. */
void mexFunction(
  int nlhs, mxArray* plhs[],
  int nrhs, const mxArray* prhs[])
```

```
{                                                                          310
   /* Check dimensions. */
   int nk = mxGetN(KS_IN);
   int n = mxGetM(DS_IN);

#ifdef DEBUG
   /* Check for proper number of arguments. */
   if (nrhs < 7)
   {
      mexErrMsgTxt("stackgdfast2mex: 7 input arguments required.");
   }                                                                       320
   else if (nlhs != 4)
   {
      mexErrMsgTxt("stackgdfast2mex: 4 output arguments required.");
   }

   /* Check for correct input dimensions. */
   int mk = mxGetM(KS_IN);
   int m = mxGetN(DS_IN);
   if ((mk != 1) || (m != 1))
   {                                                                       330
      mexErrMsgTxt("StackGDFast2 called with incorrect dimensions.");
   }
#endif

   /* Create matrices for return arguments */
   R2_OUT = mxCreateDoubleMatrix(1, nk, mxREAL);
   GD_OUT = mxCreateDoubleMatrix(1, nk, mxREAL);
   R2GRAD_OUT = mxCreateDoubleMatrix(n, nk, mxREAL);
   GDGRAD_OUT = mxCreateDoubleMatrix(n, nk, mxREAL);
                                                                           340
   /* Assign pointers and values to the parameters */
   double* r2 = mxGetPr(R2_OUT);
   double* gd = mxGetPr(GD_OUT);
   double* r2grad = mxGetPr(R2GRAD_OUT);
   double* gdgrad = mxGetPr(GDGRAD_OUT);

   double* ks = mxGetPr(KS_IN);
   double* ds = mxGetPr(DS_IN);
   double n0 = mxGetScalar(N0_IN);
   double* ns = mxGetPr(NS_IN);                                            350
   double* dns = mxGetPr(DNS_IN);
   double theta = mxGetScalar(THETA_IN);
   int polstrlen = mxGetN(POL_IN);
   char polstr[8];
   mxGetString(POL_IN, polstr, polstrlen + 1);
   int isTM = (strncmp("TM", polstr, 2) == 0);
```

```
#ifdef DEBUG
    /* Check for negative thicknesses. */
    for (int dx = 0; dx < n; dx++)                                    360
    {
        if (ds[dx] < 0.0)
        {
            mexWarnMsgTxt("Negative layer thickness.\n");
            break;
        }
    }
#endif

    /* Do the actual computation */                                   370
    stackgdgrad(nk, n, ks, ds, n0, ns, dns, theta, isTM,
                r2, gd, r2grad, gdgrad);

    return;
}
```

THIS PAGE INTENTIONALLY LEFT BLANK

Bibliography

[1] F. X. Kärtner, Ed., *Few-Cycle Laser Pulse Generation and its Applications*. Springer, 2004.

[2] S. M. J. Kelly, "Characteristic sideband instability of periodically amplified average soliton," *Electron. Lett.*, vol. 28, pp. 806–807, 1992.

[3] H. A. Haus, "Mode-locking of lasers," *IEEE J. Sel. Top. Quant. Electron.*, vol. 6, pp. 1173–1185, 2000.

[4] T. H. Maimon, "Stimulated optical radiation in ruby," *Nature*, vol. 187, pp. 493–494, 1960.

[5] J. Ouellette, "Femtosecond lasers prepare to break out of the laboratory," *Physics Today*, p. 36, January 2008.

[6] M. Dantus, M. J. Rosker, and A. H. Zewail, "Real-time femtosecond probing of "transition states" in chemical reactions," *J. Chem. Phys.*, vol. 87, no. 4, pp. 2395–2397, 1987.

[7] A. Assion, T. Baumert, M. Bergt, T. Brixner, B. Kiefer, V. Seyfried, M. Strehle, and G. Gerber, "Control of chemical reactions by feedback-optimized phase-shaped femtosecond laser pulses," *Science*, vol. 282, pp. 919–922, 1998.

[8] T. Udem, R. Holzwarth, and T. W. Hänsch, "Optical frequency metrology," *Nature*, vol. 416, pp. 233–237, 2002.

[9] J. W. Chan, T. Huser, S. Risbud, and D. M. Krol, "Structual changes in fused silica after exposure to focused femtosecond laser pulses," *Opt. Lett.*, vol. 26, pp. 1726–1728, 2001.

[10] P. M. Paul, E. S. Toma, P. Breger, G. Mullot, F. Augé, P. Balcou, H. G. Muller, and P. Agostini, "Observation of a train of attosecond pulse from high harmonic generation," *Science*, vol. 292, pp. 1689–1692, 2001.

[11] G. Sansone, E. Benedetti, F. Calegari, L. Avaldi, R. Flammini, L. Poletto, P. Villoresi, C. Altucci, R. Velotta, S. Stagira, S. D. Silvestri, and M. Nisoli, "Isolated single-cycle attosecond pulses," *Science*, vol. 314, pp. 443–446, 2006.

[12] G. Agrawal, *Nonlinear Fiber Optics*. Academic Press, 2001.

[13] J. R. Birge and F. X. Kärtner, "A preconditioned newton-krylov method for computing stationary pulse solutions of mode-locked lasers," in *Conference of Lasers and Electro-Optics (CLEO)*. Baltimore: OSA, 2007, p. CTuC7.

[14] R. Telichevesky, K. Kundert, and J. White, "Efficient steady-state analysis based on matrix-free krylov-subspace methods," in *Proc. ACM IEEE Des. Auto. Conf.*, Santa Clara, June 1995, pp. 480–484.

[15] O. Nastov, R. Telichevesky, K. Kundert, and J. White, "Fundamentals of fast simulation algorithms for rf circuits," *Proc. IEEE*, vol. 95, no. 3, pp. 600–620, March 2007.

[16] F. X. Kärtner, N. Matuschek, T. Schibli, U. Keller, H. A. Haus, C. Heine, R. Morf, V. Scheuer, M. Tilsch, and T. Tschudi, "Design and fabrication of double-chirped mirrors," *Opt. Lett.*, vol. 22, pp. 831–833, 1997.

[17] F. X. Kärtner, U. Morgner, T. R. Schibli, E. P. Ippen, J. G. Fujimoto, V. Scheuer, G. Angelow, and T. Tschudi, "Ultrabroadband double-chirped mirror pairs for generation of octave spectra," *J. Opt. Soc. Am. B*, vol. 18, pp. 882–885, 2001.

[18] C. Iaconis and I. Walmsley, "Self-referencing spectral interferometry for measuring ultrashort pulses," *J. Quant. Elec.*, vol. 35, pp. 501–509, 1999.

[19] J. R. Birge, R. Ell, and F. X. Kärtner, "Two-dimensional spectral shearing interferometry (2dsi) for ultrashort pulse characterization," in *Conference of Lasers and Electro-Optics (CLEO)*. Long Beach: OSA, 2006.

[20] ——, "Two-dimensional spectral shearing interferometry for few-cycle pulse characterization," *Opt. Lett.*, vol. 31, no. 13, pp. 2063–2065, July 2006.

[21] H. M. Crespo, J. R. Birge, , M. Y. Sander, E. L. F. cao Filho, A. Benedick, and F. X. Kärtner, "Phase stabilization of sub-two-cycle pulses from prismless octave-spanning ti:sapphire lasers," *J. Opt. Soc. Am. B*, vol. 25, pp. B147–B154, 2008.

[22] K. F. Lee, K. J. Kubarych, A. Bonvalet, and M. Joffre, "Characterization of mid-infrared femtosecond pulses [invited]," *J. Opt. Soc. Am. B*, vol. 25, pp. A54–A62, 2008.

[23] R. W. Boyd, *Nonlinear Optics*, 2nd ed. San Diego: Academic Press, 2003.

[24] M. Y. Sander, H. M. Crespo, J. R. Birge, and F. X. Kärtner, "Modeling of octave-spanning sub-two-cycle titanium:sapphie lasers: simulation and experiment," in *Conference on Ultrafast Phenomena*, 2008, submitted.

[25] H. A. Haus, *Waves and fields in optoelectronics*. Prentice-Hall, 1984.

[26] A. Gordon and B. Fischer, "Phase transition theory of many-mode ordering and pulse formation in lasers," *Phys. Rev. Lett.*, vol. 89, no. 10, p. 103901, 2002.

[27] H. A. Haus, "Theory of mode-locking with a slow saturable absorber," *IEEE J. Quant. Electron.*, vol. QE-11, pp. 736–746, 1975.

[28] A. J. DeMaria, D. A. Stetser, and H. Heynau, "Self mode-locking of lasers with saturable absorbers," *Applied Physics Letters*, vol. 8, no. 7, pp. 174–176, 1966.

[29] D. E. Spence, P. N. Kean, and W. Sibbett, "60-fsec pulse generation from a self-mode-locked ti:sapphire laser," *Opt. Lett.*, vol. 16, pp. 42–44, 1991.

[30] H. A. Haus, "Theory of mode-locking with a fast saturable absorber," *J. Appl. Phys.*, vol. 46, pp. 3049–3058, 1975.

[31] L. F. Mollenauer and R. H. Stolen, "The soliton laser," *Opt. Lett.*, vol. 9, pp. 13–15, 1984.

[32] F. X. Kärtner, J. A. der Au, and U. Keller, "Mode-locking with slow and fast saturable absorbers—what's the difference?" *IEEE J. Select. Top. Quant. Electron.*, vol. 4, pp. 159–168, 1998.

[33] Y. Chen, F. X. Kärtner, U. Morgner, S. H. Cho, H. A. Haus, H. G. Fujimoto, and E. P. Ippen, "Dispersion managed mode locking," *J. Opt. Soc. Am. B*, vol. 16, no. 11, pp. 1999–2004, 1999.

[34] U. Morgner, F. X. Kärtner, S. H. Cho, Y. Chen, H. A. Haus, J. G. Fujimoto, and E. P. Ippen, "Sub-two-cycle pulses from a kerr-lens mode-locked ti:sapphire laser," *Opt. Lett.*, vol. 24, pp. 411–413, 1999.

[35] O. E. Martinez, R. L. Fork, and J. P. Gordon, "Theory of passively mode-locked lasers for the case of a nonlinear complex-propagation coefficient," *JOSA B*, vol. 2, pp. 753–760, 1985.

[36] H. A. Haus, J. G. Fujimoto, and E. P. Ippen, "Structures for additive pulse mode locking," *JOSA B*, vol. 8, pp. 2068–2076, 1991.

[37] H. A. Haus, W. S. Wong, and F. I. Khatri, "Continuum generation by perturbation of solitons," *JOSA B*, vol. 14, pp. 304–313, 1997.

[38] A. Arvanitoyeorgos, *An Introduction to Lie Groups and the Geometry of Homogeneous Spaces*. Americal Mathematical Society, 2003.

[39] Y. Saad and M. H. Schultz, "Gmres: A generalized minimal residual algorithm for solving nonsymmetric linear systems," *SIAM J. Sci. Stat. Comput.*, vol. 7, pp. 856–869, 1986.

[40] S. C. Eisenstat, H. C. Elman, and M. H. Schultz, "Variational iterative methods for nonsymmetric systems of linear equations," *SIAM Sci. Stat. Comput.*, vol. 20, no. 2, pp. 345–357, 1983.

[41] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed. Philadelphia: SIAM, 2003.

[42] K. Jea and D. M. Young, "Generalized conjugate gradient acceleration of nonsymmetrizable iterative methods," *Lin. Alg. Appl.*, vol. 34, pp. 159–194, 1980.

[43] L. N. Trefethen and D. Bau, *Numerical linear algebra.* SIAM, 1997.

[44] R. Szipöcs, K. Ferencz, C. Spielmann, and F. Karusz, "Chirped multilayer coatings for broadband dispersion control in femtosecond lasers," *Opt. Lett.*, vol. 19, pp. 201–203, 1994.

[45] V. Pervak, S. Naumov, F. Krausz, and A. Apolonski, "Chirped mirrors with low dispersion ripple," *Opt. Ex.*, vol. 15, pp. 13 768–13 772, 2007.

[46] G. Steinmeyer, "Dispersion oscillations in ultrafast phase-correction devices," *J. Quant. Elec.*, vol. 39, pp. 1027–1034, 2003.

[47] P. Dombi, V. S. Yakovlev, K. O'Keeffe, T. Fuji, M. Lezius, and G. Tempea, "Pulse compression with time-domain optimized chirped mirrors," *Opt. Express*, vol. 13, pp. 10 888–10 894, 2005.

[48] M. Trubetskov, A. Tikhonravov, and V. Pervak, "Time-domain approach for designing dispersive mirrors based on the needle optimization technique. theory." *Opt. Ex.*, vol. 16, pp. 20 637–20 647, 2008.

[49] J. R. Birge and F. X. Kärtner, "Efficient optimization of multilayer coatings for ultrafast optics using analytic gradients of dispersion," *Appl. Opt.*, vol. 46, no. 14, pp. 2656–2662, 2007.

[50] R. J. Jones, K. Moll, M. Thorpe, and J. Ye, "Phase-coherent frequency combs in the vacuum ultraviolet via high-harmonic generation inside a femtosecond enhancement cavity," *Phys. Rev. Lett.*, vol. 94, p. 193201, 2005.

[51] C.-H. Li, A. Benedick, P. Fendel, A. Glenday, F. X. Kärtner, D. Phillips, D. Sasselov, A. Szentgyorgyil, and R. Walsworth, "A laser frequency comb that enables radial velocity measurements with a precision on 1 cm s^{-1}," *Nature*, vol. 452, pp. 610–612, 2008.

[52] M. Y. Sander, J. R. Birge, A. Benedick, H. M. Crespo, and F. X. Kärtner, "Dynamics of dispersion managed octave-spanning titanium:sapphire lasers," *J. Opt. Soc. Am. B*, vol. 26, pp. 743–749, 2009.

[53] R. Ell, U. Morgner, F. X. Kärtner, J. G. Fujimoto, E. P. Ippen, V. Scheuer, G. Angelow, T. Tschudi, M. J. Lederer, A. Boiko, and B. Luther, "Generation of 5-fs pulses and octave-spanning spectra directly from a ti:sapphire laser," *Opt. Lett.*, vol. 26, pp. 373–375, 2001.

[54] P.-A. Bèlanger, "Stable operation of mode-locked fiber lasers: similariton regime," *Opt. Ex.*, vol. 15, pp. 11 033–11 041, 2007.

[55] V. L. Kalashnikov, E. Podivilov, A. Chernykh, S. Naumov, A. Fernandez, R. Graf, and A. Apolonski, "Approach the microjoule frontier with femtosecond laser oscillators: theory and comparison with experiment," *New J. of Phys.*, vol. 7, pp. doi:10.1088/1367–2630/7/1/217, 2005.

[56] A. Chong, W. H. Renninger, and F. W. Wise, "Properties of normal-dispersion femtosecond fiber lasers," *JOSA B*, vol. 25, pp. 140–148, 2008.

[57] H. A. Haus and Y. Silberberg, "Laser modelocking with addition of nonlinear index," *IEEE J. Quant. Electron.*, vol. 22, pp. 325–331, 1986.

[58] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability*. Athena Scientific, 2002.

[59] A. Stingl, M. Menzner, C. Spielmann, F. Krausz, and R. Szipöcs, "Sub-10-fs mirror-dispersion-controlled ti:sapphire laser," *Opt. Lett.*, vol. 20, pp. 602–604, 1995.

[60] T. R. Schibli, O. Kuzucu, J. Kim, E. P. Ippen, J. G. Fujimoto, F. X. Kärtner, V. Scheuer, and G. Angelow, "Towards single-cycle laser systems," *IEEE J. Sel. Top. Quant. Elec.*, vol. 9, pp. 990–1001, 2003.

[61] O. D. Mücke, R. Ell, A. Winter, J. Kim, J. R. Birge, L. Matos, and F. X. Kärtner, "Self-referenced 200 mhz octave-spanning ti:saphhire laser with 50 attosecond carrier-envelope phase jitter," *Opt. Express*, vol. 13, no. 5163, 2005.

[62] A. V. Tikhonravov, "Some theoretical apsects of thin-film optics and their applications," *Appl. Opt.*, vol. 32, pp. 5417–5426, 1993.

[63] A. V. Tikhonravov, P. W. Baumeister, and K. V. Popov, "Phase properties of multilayers," *Appl. Opt.*, vol. 36, pp. 4382–4392, 1997.

[64] C. J. v. d. Laan and H. J. Frankena, "Fast computation method for derivatives of multilayer stack reflectance," *Appl. Opt.*, vol. 17, pp. 538–541, 1978.

[65] J. R. Birge, C. Jirauschek, and F. X. Kärtner, "Efficient analytic computation of group delay dispersion from optical interference coatings," in *OSA Opt. Interference Coatings Top. Mtg.*, Tucson, 2004, p. ThA6.

[66] J. R. Birge and F. X. Kärtner, "Efficient analytic computation of dispersion from multilayer structures," *Appl. Opt.*, vol. 45, no. 7, pp. 1478–1483, 2006.

[67] K. Atkinson, *An Introduction to Numerical Analysis*. Wiley, New York, 1989.

[68] J. A. Kong, *Electromagnetic Wave Theory*. EMW, 2000.

[69] N. Matuschek, F. X. Kärtner, and U. Keller, "Theory of double-chirped mirrors," *IEEE J. Sel. Top. Quant. Elec.*, vol. 4, pp. 197–208, 1998.

[70] A. A. Tovar and L. W. Casperson, "Generalized reverse theorems for multipass applications in matrix optics," *J. Opt. Soc. Am. A*, vol. 11, pp. 2633–2642, 1994.

[71] B. Sullivan and J. Dobrowolski, "Deposition error compensation for optical multilayer coating: I. theoretical description," *Appl. Opt.*, vol. 31, pp. 3821–3835, 1992.

[72] V. Pervak, A. Tikhonravov, M. Trubetskov, S. Naumov, F. Krausz, and A. Apolonski, "1.5-octave chirped mirror for pulse compression down to sub-3 fs," *Appl. Opt.*, vol. 87, pp. 5–12, 2007.

[73] V. Yakovlev and G. Tempea, "Optimization of chirped mirrors," *Appl. Opt.*, vol. 41, pp. 6514–6520, 2002.

[74] A. Ben-Tal and A. Nemirovski, "Robust convex optimization," *Mathematics of Operations Research*, vol. 23, pp. 769–806, 1998.

[75] ——, "Robust optimization — methodology and applications," *Mathematical Programming*, vol. 92, no. 3, pp. 453–480, 2002.

[76] D. Bertsimas and M. Sim, "Robust discrete optimization and network flows," *Mathematical Programming*, vol. 98, pp. 49–71, 2003.

[77] ——, "Tractable approximations to robust conic optimization problems," *Mathematical Programming*, vol. 107, no. 1, pp. 5–36, 2006.

[78] P. Verly, "Fourier transform technique with refinement in the frequency domain for the synthesis of optical thin films," *Appl. Opt.*, vol. 35, pp. 5148–5154, 1996.

[79] D. Bertsimas, O. Nohadani, and K. M. Teo, "Robust nonconvex optimization for simulation-based problems," *Operations Research*, 2007, to Appear.

[80] A. Tikhonravov, M. Trubetskov, and G. DeBell, "Applictions of the needle optimization technique to the design of optical coatings," *Appl. Opt.*, vol. 35, pp. 5493–5508, 1996.

[81] E. Matsubara, K. Yamane, T. Sekikawa, and M. Yamashita, "Generation of 2.6 fs optical pulses using induced-phase modulation in a gas-filled hollow fiber," *J. Opt. Soc. Am. B*, vol. 24, pp. 985–989, 2007.

[82] P. B. Corkum and F. Krausz, "Attosecond science," *Nature Phys.*, vol. 3, p. 381, 2007.

[83] I. A. Walmsley and V. Wong, "Characterization of the electric field of ultrashort optical pulses," *J. Opt. Soc. Am. B*, vol. 13, no. 11, pp. 2453–2563, November 1996.

[84] J. Nicholson, J. Jasapara, W. Rudolph, F. Omenetto, and A. Taylor, "Full-field characterization of femtosecond pulses by spectrum and cross-correlation measurements," *Opt. Lett.*, vol. 24, pp. 1774–1776, 1999.

[85] D. Kane and R. Trebino, "Characterization of arbitrary femtosecond pulses using frequency-resolved optical gating," *J. Quant. Elec.*, vol. 29, pp. 571–579, 1993.

[86] J. Chung and A. Weiner, "Ambiguity of ultrashort pulse shapes retrieved from the intensity autocorrelation and the power spectrum," *J. Spec. Top. Quant. Elec.*, vol. 7, pp. 656–666, 2001.

[87] I. A. Walmsley, "Characterization of ultrashort optical pusles in the few-cycle regime using spectral phase interferometry for direct electric-field reconstruction," in *Few-Cycle Laser Pulse Generation and its Applications*, Kärtner, Ed. Springer, 2004, pp. 265–291.

[88] E. H. Armstrong, "A method of reducing disturbances in radio signaling by a system of frequency modulation," *Proc. of the IRE*, vol. 24, no. 5, pp. 689–740, 1936.

[89] L. Gallmann, D. Sutter, N. Matuschek, G. Steinmeyer, and U. Keller, "Techniques for the characterization of sub-10-fs optical pulses: a comparision," *Appl. Phys. B*, vol. 70, pp. S67–S75, 2000.

[90] S. Jensen and M. Anderson, "Measuring ultrashort optical pulses in the presence of noise: an empirical study of the performance of spectral phase interferometry for direct electric field reconstruction," *Appl. Opt.*, vol. 43, pp. 883–893, 2004.

[91] A. Yariv, *Optical Electronics in Modern Communications*, 5th ed. Oxford University Press, 1997.

[92] K. Yamane, Z. Zhang, K. Oka, R. Morita, and M. Yamashita, "Optical pulse compression to 3.4 fs in the monocycle region by feedback phase compensation," *Opt. Lett.*, vol. 28, pp. 2258–2260, 2003.

[93] H. M. Crespo, J. R. Birge, E. L. F. cao Filho, M. Y. Sander, A. Benedick, and F. X. Kärtner, "Non-intrusive phase-stabilization of sub-two-cycle pulses from a prismless octave-spanning ti:sapphire laser," *Opt. Lett.*, vol. 33, pp. 833–835, 2008.

[94] C. Dorrer, "Influence of the calibration of the detector on spectral interferometry," *J. Opt. Soc. Am. B*, vol. 16, no. 7, pp. 1160–1168, 1999.

[95] ——, "Implementation of spectral phase interferometry for direct electric-field reconstruction with a simultaneously recorded reference interferogram," *Opt. Lett.*, vol. 24, no. 21, pp. 1532–1534, 1999.

[96] T. Shuman, I. A. Walmsley, L. Waxer, M. Anderson, C. Iaconis, , and J. Bromage, "Real-time spider: ultrashort pulse characterization at 20 hz," *Opt. Express*, vol. 5, no. 134–143, 1999.

[97] B. von Vacano, T. Buckup, and M. Motzkus, "Shaper-assisted collinear spider: fast and simple broadband pulse compression in nonlinear microscopy," *J. Opt. Soc. Am. B*, vol. 24, no. 5, pp. 1091–1100, 2007.

[98] J. R. Birge, H. M. Crespo, M. Sander, and F. X. Kärtner, "Non-instrusive sub-two-cycle carrier-envelope stabilized pulses using engineered chirped mirrors," in *Conference of Lasers and Electro-Optics (CLEO)*. San Jose: OSA, 2008.

[99] C. Dorrer, P. Londero, and I. A. Walmsley, "Homodyne detection in spectral phase interferometry for direct electric-field reconstruction," *Opt. Lett.*, vol. 26, pp. 1510–1512, 2001.

[100] P. Baum, S. Lockbrunner, and E. Riedle, "Zero-additional-phase spider: full characterization of visible and sub-20-fs ultraviolet pulses," *Opt. Lett.*, vol. 29, pp. 210–212, 2004.

[101] E. Kosik, A. Radunksy, I. Walmsley, and C. Dorrer, "Interferometric technique for measuring broadband ultrashort pulses at the sampling limit," *Opt. Lett.*, vol. 30, no. 3, pp. 326–328, 2005.

[102] A. S. Wyatt, I. A. Walmsley, G. Stibenz, and G. Steinmeyer, "Sub-10 fs pulse characterization using spatially encoded arrangement for spectral phase interferometry for direct electric field reconstruction," *Opt. Lett.*, vol. 31, no. 12, pp. 1914–1916, 2006.

[103] J. R. Birge, R. Ell, and F. X. Kärtner, "Two-dimensional spectral shearing interferometry for few-cycle pulse characterization and optimization," in *Ultrafast Phenomena XV*, Corkum, Jonas, Miller, and Weiner, Eds. Springer, 2006, pp. 160–162.

[104] S.-P. Gorza, P. Wasylczyk, and I. A. Walmsley, "Spectral shearing interferometry with spatially chirped replicas for measuring ultrashort pulses," *Opt. Express*, vol. 15, pp. 15 168–15 174, 2007.

[105] J. R. Birge and F. X. Kärtner, "Analysis and mitigation of systematic errors in spectral shearing interferometry of pulses approaching the single-cycle limit [invited]," *J. Opt. Soc. Am. B*, vol. 25, pp. A111–A119, 2008.

[106] R. Morita, M. Hirasawa, N. Karasawa, S. Kusaka, N. Nakagawa, K. Yamane, L. Li, A. Suguro, and M. Yamashita, "Sub-5 fs optical pulse characterization," *Meas. Sci. and Tech.*, vol. 13, pp. 1710–1720, 2002.

[107] M. Takeda, H. Ina, and S. Kobayashi, "Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry," *J. Opt. Soc. Am.*, vol. 72, pp. 156–160, 1982.

[108] V. A. Mandelshtam and H. S. Taylor, "Harmonic inversion of time signals and its applications," *J. of Chem. Phys.*, vol. 107, no. 17, pp. 6756–6769, 1997. [Online]. Available: http://dx.doi.org/10.1063/1.475324