Title
Computing with Femtosecond Laser Pulses

Author
ZHOU, TINGYI

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Computing with Femtosecond Laser Pulses

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Electrical and Computer Engineering

by

Tingyi Zhou

2022

ABSTRACT OF THE DISSERTATION

Computing with Femtosecond Laser Pulses

by

Tingyi Zhou
Doctor of Philosophy in Electrical and Computer Engineering
University of California, Los Angeles, 2022
Professor Bahram Jalali, Chair

Time stretch instruments have been exceptionally successful in discovering single-shot ultrafast phenomena such as optical rogue waves and have led to record-speed microscopy, spectroscopy, lidar, etc. These instruments encode the ultrafast events into the spectrum of a femtosecond pulse and then dilate the time scale of the data using group velocity dispersion. Generating as much as Tbit per second of data, they are ideal partners for deep learning networks which by their inherent complexity, require large datasets for training. However, the inference time scale of neural networks in the millisecond regime is orders of magnitude longer than the data acquisition rate of time stretch instruments. This underscores the need to explore means where some of the lower-level computational tasks can be done while the data is still in the optical domain. To address this predicament, we propose the Nonlinear Schrödinger Kernel computing. This real time computing framework utilizes optical nonlinearities to map the data onto a new domain in which classification accuracy is enhanced, without increasing the data dimensions. A novel training scheme for the kernel is developed by utilizing digital phase encoding of the input data.

The dissertation of Tingyi Zhou is approved.

Achuta Kadambi

Fabien Scalzo

Kang Lung Wang

Bahram Jalali, Committee Chair

University of California, Los Angeles

2022

To my family

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

It's been five years since I started my journey as a graduate researcher. While the work is never easy, the excitement of discovery and the sense of achievement always keep me up through the hard days. Apart from that, people around me also gave me the courage to face the challenges. Here, I would like to express my heartfelt gratitude to those who were with me, who are with me, and who will continue to be with me.

I would like to express my deepest appreciation to my advisor and mentor, Prof. Bahram Jalali, for the valuable opportunity to work with him. His passion for science and discovery inspired me and supported me to break the barriers. His guidance always comes in timely and helps me solves critical issues. He taught me how to think and act as a professional researcher. I benefit a lot from him, not only as a researcher but also as a person.

I would also like to extend my gratitude to my TPC member, Prof. Kang Lung Wang, Prof. Fabien Scalzo, Prof. Achuta Kadambi, as well as Prof. Vwani Roychowdhurry who used to serve on my committee, for their valuable opinions in my Ph.D. exams. Especially, I want to thank Prof. Kang Lung Wang for taking the urgent call being on my committee.

I would like to extend my sincere thanks to my UCLA labmates, Cejo K. Lonappan, Jacky Chan, Yunshan Jiang, Tianwei Jiang, Madhuri Suthar, Shuqian Sun, Zhuoya Bai, Sifeng He, Yutong Lu, Yiming Zhou, and Callen Mcphee for the great advice and inspiring discussions. And also my MX colleagues, Fran Mazdyasian, Ling Sha, Dave King, Keyhaneh Gharagozluyan, and Hiepp Luu for their help.

I would like to thank all of my publication co-authors, which our works consist of many chapters of this dissertation. Chapter 2 is a version of Journal of Lightwave Technology, Vol. 40, Issue 5, 1308-1319, 2022. Chapter 3 is a version of the manuscript we just submitted to the Journal of Physics: Photonics, 2022. The projects are funded by the Defense Advanced Research Projects Agency (DARPA) MTO PEACH program under contract number HR00111990050.

I would like to thank my friends, Jianhong He, Tianyi Wang, Chenlei Zhao, Chang Tian, Yihang Song, Yuening Li, Qiujing Lu, Yipeng Zhang, Hengjie Yang, Yunzhe Qiu, Weijia Yuan, Fangyao Liu, Haoxin Zheng, and all others for the joyful time we shared together.

I'm deeply indebted to my parents, Wei Zong and Chang Zhou, who raised me to become who I am, protected me all along the way, opened up my eyes to the world, and supported me with any available resources they can provide. Whenever I lose faith in myself, their trust always picks me up so I can get back on my feet. I'm grateful for all the support and encouragement.

Finally, I cannot begin to express my thanks to my partner, Zelin, for the love, support, and companionship that helped me through the darkest days. Thank you for accepting me, taking care of me, and making me a better person. All of this cannot happen without you. You complete me.

VITA

2017	B.S. School of Optoelectronics, Beijing Institute of Technology
2019	M.S. Department of Electrical and Computer Engineering, UCLA
2019	Teaching Assistant, Department of Electrical and Computer Engineering, UCLA
2020	Teaching Assistant, Department of Electrical and Computer Engineering, UCLA
2020	Teaching Assistant, Department of Electrical and Computer Engineering, UCLA
2021	Hardware Design Engineer Intern, MX Imaging

PUBLICATIONS

Zhou, T., Scalzo, F., and Jalali, B. (2022). Nonlinear Schrödinger Kernel for Hardware Acceleration of Machine Learning. Journal of Lightwave Technology, 40(5), 1308-1319. INVITED

Zhou, T. and Jalali B. Low Latency Computing for Time Stretch Instruments Journal of Physics: Photonics (Submitted) INVITED

Zhou, Y., Samiee, A., Zhou, T., and Jalali, B. (2020). Deep learning interference cancellation in wireless networks. arXiv preprint arXiv:2009.05533.

CHAPTER 1

Introduction

Photonic time stretch is a state-of-the-art sensing technique that measures single-shot ultrafast events by first encoding it onto the spectrum of an optical pulse and then dilating the time scale of the data using group velocity dispersion[1998 and Unified] [BCJ98][ZCJ22]. Time stretch instruments have been exceptionally successful in various applications, including optical rogue waves [SRK07], optical soliton dynamics [HKJ17], relativistic electron dynamics [REL15], chemical transients in combustion [MLM20], stock wave [HGI18], optical gyroscope [KSC22], and have led to record-speed microscopy [SKV16] and lidar [JKJ20]. While encouraging, these femtosecond instruments that acquire billions of frames of data per second also require ultrafast realtime processing of the acquired data.

Neural networks have emerged as a generic computational framework that achieves spectacular performance in image and speech recognition and synthesis. While revolutionary, these performance gains are not without cost. First, the exponential growth in the required computing power outpaces the semiconductor roadmap known as Moore's law by several orders of magnitude [XDH18][TGL20]. Second, the large, labeled datasets required for training these models are not available in many application domains. Third, the latency inherent in deep neural networks can be problematic in applications where sensors generate a torrent of data that must be classified in realtime.

These predicaments call for a fresh look into the design of machine learning and computing systems that can operate with small training datasets, offer low latency, and do not rely on ever-increasing computational performance and memory size. Among several alter-

native approaches to computing is the use of analog systems to efficiently perform specialized computations, an approach that echoes the early days of computing [Isa14]. An attractive attribute of analog systems is that they do not need to be trained; however, purely analog computing systems are susceptible to noise and do not scale. A modern approach to analog computing is one where analog does not replace the digital computer but rather serves as a hardware accelerator for it [SJ15]. For example, one can simulate the properties of an inaccessible and complex system, such as that underlying hydrodynamic phenomena, with a more user-friendly and compact proxy system, such as wave propagation in a nonlinear optical fiber, which is governed by the same differential equations. A natural computer based on such rapid phenomenon can serve as a surrogate for the computation of fluid dynamics phenomena [SJ15]. When considering the vast state space that needs to be explored, the benefits are even greater. A related framework could involve the use of wave propagation in a metamaterial to perform specialized computational tasks such as solving specific integral equations [EEE19]. With an instrument capable of capturing the output of such natural computers in real time, billions of scenarios can be readily acquired on an ultrashort timescale to map the vast space of complex outputs that emerge from the nonlinear responses to different inputs. A related technique is Reservoir Computing, which utilizes analog systems to transform the input data into a much higher dimensional space and perform pattern analysis with a simple readout [TYH19]. Other studies have been on the emulation of a neural network with optical components. The approach is to use lasers and spatial light modulators or integrated optical circuits to mimic the matrix multiplication function of a neural network [LRY18][GMA19][WOG20].

In chapter 2, a new concept in data representation and classification is reported [ZSJ22]. Specifically, we show that when seeded with data, nonlinear optics enable a linear learning algorithm to learn a nonlinear function or decision boundary that separates the data into the correct classes. The core of this system, which we call the Nonlinear Schrödinger Kernel, nonlinearly projects data that have been modulated onto the spectrum of a femtosecond

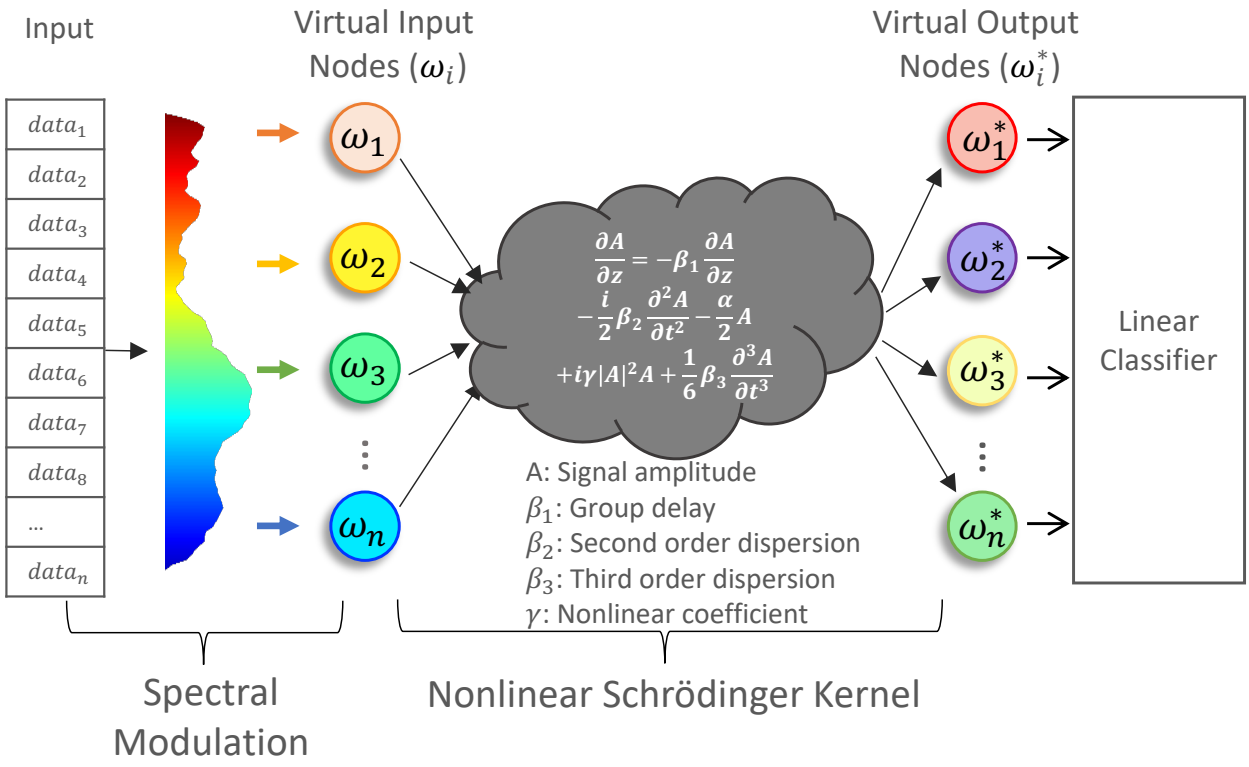

Figure 1.1: The Nonlinear Schrödinger Kernel computing framework. Samples of the input data are first assigned to virtual nodes through spectral modulation. The input data are nonlinearly projected to an output representation with a transformation governed by the Nonlinear Schrödinger Equation (NLSE). [ZSJ22]

laser pulse into a space in which the data that normally could not be linearly separated, can now be separated with a linear predictor.

Fig. 1.1 shows a qualitative interpretation of our approach. The input data are first mapped to spectrum nodes of a femtosecond laser pulse by a spectral modulator and then undergo a nonlinear transformation in an optical device causing complex energy and phase transfer between the nodes. The nonlinear transformation is governed by the Nonlinear Schrödinger Equation (NLSE) [Agr12]. The output spectrum is measured and then fed into a numerical machine learning algorithm for classification. In this approach, the spectral

modulation of data followed by nonlinearly transforming the data into an implicit high-dimensional representation enables certain nonlinear classification problems to be solved with a low-latency linear classifier. We emphasize that this dimensionality increase is implicit, not explicit (actual increase in dimensionality is not necessary although that may improve performance).

We verify this technique by performing classification on a variety of data. We show that the combination of the Nonlinear Schrödinger Kernel with a digital classifier provides improvements in accuracy and latency over a conventional, purely digital classifier. To provide insight into the operation of the kernel, we compare its performance to the radial basis function (RBF), the most common numerical kernel in machine learning, and show that both enable a linear classifier to perform nonlinear classification, while the physical kernel is orders of magnitude faster, with high resilience against quantization and additive noise [ZSJ22].

We also note that since this is an open-loop (untrained) system with a low degree of tunability, its performance is inevitably data-dependent. In chapter 3, we propose an approach to tune and optimize the nonlinear optical transfer function of the kernel. While the nonlinear optical system parameters cannot be tuned, the nonlinear process can be effectively engineered by varying the phase of the femtosecond laser pulse. This is based on the insight that most nonlinear interactions, such as self-phase modulation, four-wave mixing, etc. are coherent processes that depend on the input phase [Agr12]. Therefore, tuning the spectral phase of the input data will inevitably change the nonlinear interaction among spectral components [CBJ05]. Since the data is encoded onto the spectrum, this process changes the nonlinear interaction among data dimensions. A digital feedback loop is added to the system for optimization of the kernel.

In chapter 4, we provide additional studies on the Nonlinear Schrödinger Kernel, including the further acceleration of the kernel and its dimension effect. We show that the kernel can be further accelerated by orders of magnitude via (1) using the integrated silicon nitride

waveguide as the nonlinear optical element and (2) implementing the linear classifier on an edge device. We also show that the Nonlinear Schrödinger Kernel does not increase the dimension of the data. On the contrary, it works even when the dimension is reduced.

Since this technique operates with femtosecond pulses and spectrally modulated data, it is inherently compatible with a wide range of single-shot instruments enabled by time stretch data acquisition, including time stretch microscopy [CMT16] [LMC19] and spectroscopy [KCB99] for continuous realtime acquisition and classification of ultrafast events. To this end, successful operations with both a conventional grating spectrometer and the time stretch spectrometer are demonstrated.

CHAPTER 2

Nonlinear Schrödinger Kernel for hardware acceleration of machine learning

Alternative machine learning approaches that have extremely low latency and can work with only a small training dataset are needed for applications where the insatiable demands of deep learning methods for computing power and large training data cannot be met [TGL20]. In this chapter, we report a new optical accelerator for AI that exploits femtosecond laser pulses for both data acquisition and computing enabling classification at short time scales for fast optical imaging, sensing, and metrology [ZSJ22]. Modulation of data onto the spectrum of femtosecond optical pulses followed by projection into a new space using nonlinear optics reduces the latency in the nonlinear classification of certain data by orders of magnitude. The approach is validated by the classification of various datasets, including brain intracranial pressure [SLH12], cancer cell imaging [CMT16][LMC19], spoken digit recognition [JSF18], and the classic exclusive OR benchmark for nonlinear operation. The concept is demonstrated by seeding the nonlinear effect that is responsible for many fascinating natural phenomena, such as optical rogue waves [SRK07]. Stimulation of nonlinear optical interactions with spectrally modulated data transforms the data such that a computationally-light linear algorithm can learn a nonlinear decision boundary that separates the data into the correct classes. Since the optical kernel is not trained, its performance is inevitably data-dependent. Quantitative comparison with a popular numerical kernel offers insights into how this physical technique accelerates inference. Single-shot operation is demonstrated using time stretch data acquisition [BCJ98][ZCJ22].

2.1 Technical Approach

The effectiveness of the proposed approach is evaluated through a number of experimental protocols. An interesting property of the Nonlinear Schrödinger Kernel is that it projects the original signal into a modified space that emphasizes certain nonlinear properties of the data. In terms of functionality, similarities exist between the Nonlinear Schrödinger Kernel and the concept of "kernel projection" or "the kernel trick" in the machine learning literature. The utility of this processing method is that it transforms nonlinearly separable data to become linearly separable in the new, modified space.

2.1.1 Demonstrated Tasks

We illustrate the effectiveness and versatility of the proposed technique on several examples, including XOR logic operation followed by several classification tasks, including the detection of intracranial hypertension [SLH12], cell image classification for cancer detection via time stretch microscopy [CMT16][LMC19], and recognition of spoken digits[JSF18]. All the datasets include ground-truth class labels that are used to train a numerical machine learning model in a supervised manner. The model is placed after the physical Nonlinear Schrödinger Kernel. Additional details on these datasets are provided in Materials and Methods (section 2.7).

2.1.2 Experiment Description

As shown in Fig. 2.1, the prototype Nonlinear Schrø"dinger Kernel computing system consists of a fiber-based supercontinuum laser source, a spectral modulator, a nonlinear optical element, and a spectrum readout followed by a machine learning classifier operating in the digital domain. The data are first modulated onto the laser spectrum through the spectral modulator. The modulated waveform then passes through a nonlinear optical element where the data are nonlinearly transformed. We define the combination of the spectral

modulation and the nonlinear transformation as the Nonlinear Schrödinger Kernel. Because of the spectral nature of this kernel, it can also be called the Lambda Kernel, in homage to the convention of representing wavelengths with the symbol lambda (λ). The optical spectrum output from the Nonlinear Schrödinger Kernel is read out and fed into the digital signal processor (DSP) for XOR logic and machine learning classification tasks. For classification, the spectrum is sampled such that it has the same dimensionality as the input data. The nonlinear element in the experiment is a highly nonlinear fiber. For capturing the output spectrum in the experiments, both the traditional grating-based spectrometer and time stretch spectrometer for fast single-shot operation [KCB99] are demonstrated. To maximize the nonlinearities, the experimental system works in anomalous dispersion regime where extreme nonlinear processes such as optical rogue waves occur [SRK07]. Despite the highly nonlinear nature of the pulse propagation, the system is made stable by keeping the optical power below the threshold for the onset of modulation instability.

In addition to the experimental setup, we employ mathematical simulations to help design the experimental system and gain insight into its operation. A detailed description of the simulation model and experimental setup can be found in Materials and Methods (2.7).

2.1.3 Machine Learning Models

As part of our experiments, we evaluate the performance of popular machine learning algorithms with and without preprocessing using the Nonlinear Schrödinger Kernel. These algorithms include ridge regression [HK70], random forests [Ho95], spectral regression for kernel discriminant analysis (SR-KDA) [CHH11], neural networks, and support vector machines (SVMs) [CV95]. For all these methods, it is assumed that the input data are sampled as $X \in R^{n \times d}$ and the output class label as $Y \in R^{n \times c}$, where c is the number of classes. The dimensionality d and the number of samples n of the input data vary with respect to the dataset used. A more detailed description of these algorithms is presented in Materials and Methods (section 2.7).

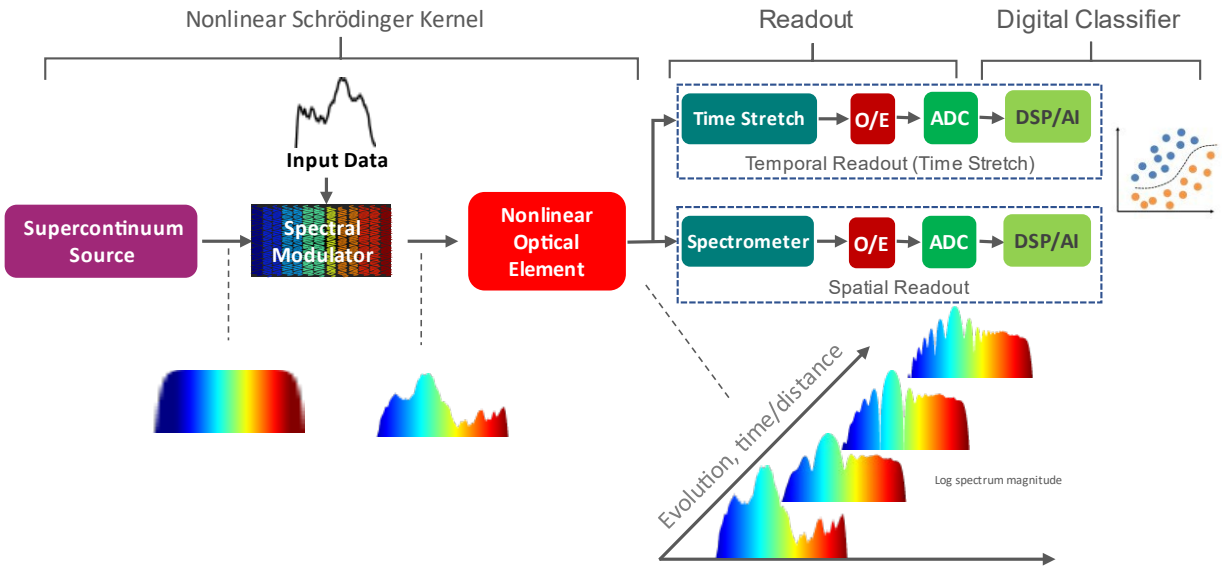

Figure 2.1: Simplified block diagram showing the basic building blocks. A broadband femtosecond mode-locked laser is used as the spectral source. The data are modulated onto the optical spectrum and transformed by nonlinear propagation in an optical medium. The inset shows the evolution of the data-modulated spectrum (cell image data). The spectrum at the output is read by a conventional spectrometer based on diffraction gratings or by time stretch spectroscopy for fast single-shot performance (examples of each are shown). Classification is performed using machine learning algorithm in the digital domain. [ZSJ22]

2.1.4 Performance Evaluation

The performance of each machine learning model is evaluated using 3-fold cross-validation, and the accuracy is reported using the area under the receiver operating characteristic (ROC) curve (AUC). The AUC represents the probability that the model will be able to predict the label correctly for a new pair of positive and negative samples. With the same cross-validation split, we compare the AUC of each model with the data before and after processing with the Nonlinear Schrödinger Kernel, and we repeat the experiment protocol on each dataset. A detailed description of these steps can be found in the Materials and Methods (section 2.7).

2.2 Demonstration of Exclusive OR Operation

We first examine the Exclusive OR (XOR) task, a classic problem that cannot be solved with a linear model and that serves as a benchmark for nonlinear operation. When four points representing the binary input pairs ((0,0), (0,1), (1,0), (1,1)) and their corresponding XOR output (0,1,1,0) are used as the training data, a linear predictor will inevitably result in an error when attempting to reproduce the results of the XOR operation (as illustrated in Fig. 2.2) because the data points are linearly nonseparable in their original space. On the other hand, after processing the data with the Nonlinear Schrödinger Kernel, a linear predictor can be trained to successfully perform the XOR operation without error.

To gain insight into how the Nonlinear Schrödinger Kernel projects the input data into a linearly separable representation, the output probability of '0' is represented by a grayscale map with values ranging from 0 (black) to 1 (white), as shown in Fig. 2.2. The boundaries of the grayscale map are depicted as dashed lines, which clearly show the accurate separation of the four pairs of inputs. This example reveals that with the use of the Nonlinear Schrödinger Kernel, the linearly nonseparable data points are projected to a space where they become linearly separable. An important property of the Nonlinear Schrödinger Kernel projection is that it is untrained; in other words, the output labels of the pairs of data points are not used to obtain the projected values in the new space.

For further intuition into the transformation performed by the combined operation of spectral modulation and nonlinear evolution, we examine the scores and the predictions obtained by a linear SVM model with and without the application of the Nonlinear Schrödinger Kernel (Fig. 2.3). The scores, calculated as the inner products of the model weights and the input data, show that the accuracy improvement is caused by the modification of the score into a linearly separable form.

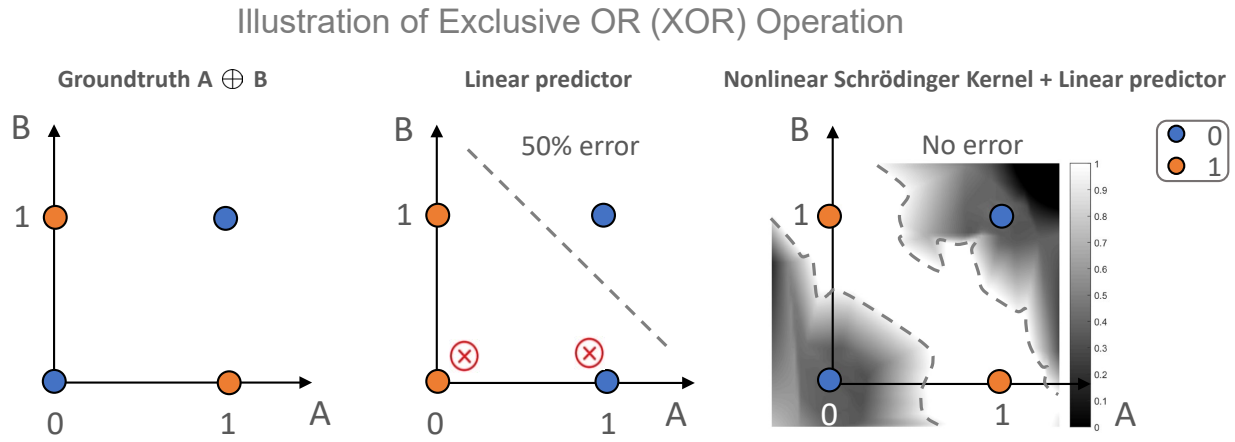

Figure 2.2: Demonstration of nonlinear operation using a linear predictor. The data points are not linearly separable, but the Nonlinear Schrödinger Kernel maps them into a space where they are. Left pane shows the XOR ground truth. Middle pane shows that a linear predictor cannot produce the correct output on its own. Right pane is the output of a linear predictor preceded by the Nonlinear Schrödinger Kernel, showing an error-free XOR operation. The output probability of '0' is mapped in grayscale over a range from 0 (black) to 1 (white). The boundary of the prediction is outlined with dashed gray lines, which can be seen to separate the expected output values of the 4 possible pairs. The details of the experimental setup are described in the Materials and Methods (section 2.7). [ZSJ22]

2.3 Demonstration of Classification Tasks

Having demonstrated the ability of the Nonlinear Schrödinger Kernel to perform nonlinear projection, we then test the hypothesis that the kernel may act as a nonlinear feature extractor that highlights certain properties of the signal and makes challenging signal classification problems solvable with a simple machine learning model. In particular, we focus on several challenging tasks with varying degrees of complexity, including the detection of brain intracranial hypertension, classification of blood cell images for cancer detection, and recognition of spoken digits. Both grating-based spectrometer and time stretch spectrometer

XOR		without Nonlinear Schrödinger Kernel		with Nonlinear Schrödinger Kernel	
Input	Ground Truth	Score	Output	Score	Output
00	0	0.99	1	**-1.03**	**0**
01	1	1.05	1	**1.03**	**1**
10	1	-1.05	0	**0.99**	**1**
11	0	-0.99	0	**-0.99**	**0**

Figure 2.3: The table shows the scores and predicted output by a linear support vector machine (SVM) model with and without the Nonlinear Schrödinger Kernel. These results show that without the Nonlinear Schrödinger Kernel, the operation fails, while implementation of the physical kernel results in successful operation. The results show that the scores are transformed into a linearly separable array. [ZSJ22]

are demonstrated. For grating based spectrometer, the spectral range varies among different datasets. For Brain data (Fig. 2.4e), the spectral range is $1550nm$ to $1630nm$ with 0.1 nm resolution. For the cell images with grating-based spectrometer (Fig. 2.4), it is $1560nm$ to $1620nm$.

Fig. 2.4 provides the classification results for four cases: cancer cell image data with time stretch spectrometer readout (2.4a), cancer cell image data with grating-based spectrometer readout (2.4b), brain ICP (2.4c), and spoken digit (2.4d). All results are experimental except the spoken digit which is based on simulation because the length of which (5000 samples) exceeds pixels numbers in our spectral modulator. The accuracy is calculated using the AUC after 3-fold cross-validation (details provided in the Materials and Methods, section 2.7). As a benchmark, we compare the performance of the Nonlinear Schrödinger Kernel with the commonly used numerical kernel, the RBF, for performing nonlinear classification with a linear classifier. When the same data are applied to both models, for most cases, the

Nonlinear Schrödinger Kernel reaches an accuracy similar to the numerical RBF kernel. One exception is the spoken digit, where the Nonlinear Schrödinger Kernel has much higher AUC (accuracy) than the RBF kernel, showing that the performance of this technique is data-dependent, which also happens to the conventional numerical kernel. These results suggest that the physical processes of spectral modulation and nonlinear evolution perform, with orders of magnitude lower latency (Fig. 2.4.e), a conceptually similar task as the so-called "kernel trick" in machine learning literature.

While Fig. 2.4 highlights the benefit of the Nonlinear Schrödinger Kernel when implemented with a linear classifier, we now turn to more advanced machine learning classifiers that already include nonlinearity in their models, including random forests, SR-KDA, SVM (with RBF kernel), and neural network. Specifically, we aim to test whether the Nonlinear Schrödinger Kernel can further improve the result obtained with these already nonlinear classifiers. The bar chart in Fig. 2.5 summarizes the experimentally measured performance of the Nonlinear Schrödinger Kernel with five different classification algorithms. The AUC is reported using the data before (gray) and after processing with the Nonlinear Schrödinger Kernel (blue). Among all datasets, consistent improvement in the AUC can be observed across most classification methods; for example, an average of 5.8% improvement in AUC was observed for the brain ICP dataset. The table shows that similar improvement is observed on other datasets. These results reveal that the Nonlinear Schrödinger Kernel can benefit the classification, and the performance is data-dependent. One exception is cell image dataset with random forest and SVM classifiers. There the Nonlinear Schrödinger Kernel leads to less accuracy (-0.1% to -0.3%). This is addressed in the discussion section (section 2.5).

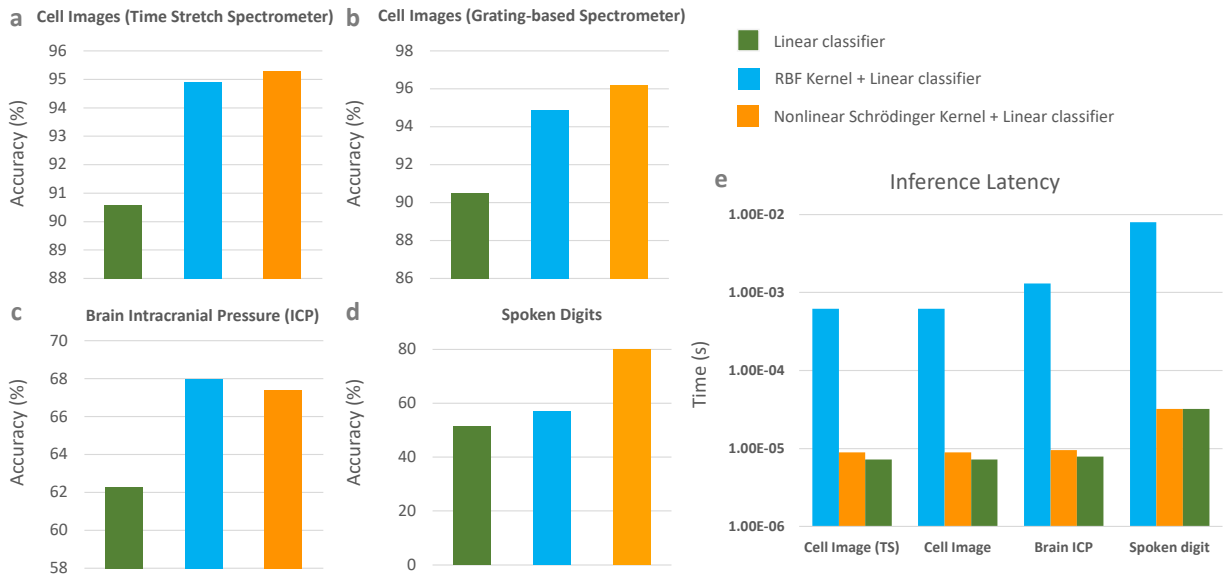

Figure 2.4: Results for four different cases: (a) microscope-derived cancer cell images with a time stretch (TS) spectrometer Kernel readout, (b) microscope-derived cancer cell images with a grating-based spectrometer Kernel readout, (c) brain intracranial pressure (ICP), and (d) spoken digits. For each dataset, the area under the curve (AUC) is reported after threefold cross-validation. For all four datasets, compared with the linear classifier alone, the (physical) Nonlinear Schrödinger Kernel shows improvement that is similar to or better than that obtained by the (numerical) radial basic function (RBF) kernel. (e) A bar chart shows the latency in the processing of one sample through the RBF kernel and the Nonlinear Schrödinger Kernel. The results are grouped by dataset, including cell data using the grating-based spectrometer readout, cell data using the TS spectrometer readout, brain ICP, and the audio signals recorded from spoken digits. The bar chart shows a latency (on a Xeon 3 GHz CPU with 64 GB of RAM) on the order of 1E-3 to 1E-2 seconds for the RBF kernel with linear classifier, while the Nonlinear Schrödinger Kernel achieved a substantially reduced latency, on the order of 1E-5, almost negligible compared to linear classifier itself.

	Brain Intracranial Pressure (ICP)			Spoken Digits		
	No Kernel	Nonlinear Schrödinger Kernel	Δ AUC	No Kernel	Nonlinear Schrödinger Kernel	Δ AUC
Ridge Regression	62.3	67.4	5.1	51.3	77.5	26.2
SR-KDA	68.0	77.1	9.1	57.0	79.9	22.9
Random Forest	51.3	58.8	7.5	78.2	87.4	9.2
SVM	76.3	81.1	4.8	54.1	89.9	35.8
Neural Network	65.2	67.7	2.5	51.9	84.7	32.8

	Cell Images (Time Stretch Spectrometer)			Cell Images (Grating-based Spectrometer)		
	No Kernel	Nonlinear Schrödinger Kernel	Δ AUC	No Kernel	Nonlinear Schrödinger Kernel	Δ AUC
Ridge Regression	90.6	95.3	4.7	90.5	96.2	5.7
SR-KDA	94.9	95.8	0.9	94.9	97.0	2.1
Random Forest	99.4	99.1	-0.3	99.5	99.4	-0.1
SVM	98.9	98.7	-0.2	98.9	98.7	-0.2
Neural Network	96.0	96.4	0.4	96.0	96.8	0.8

Figure 2.5: Classification results with different machine learning algorithms. The performance of the Nonlinear Schrödinger Kernel in the classification of various datasets (tasks) is shown: brain intracranial pressure (ICP), spoken digit, cell image data with time stretch spectrometer readout, and cell image data with grating-based spectrometer readout. For each dataset, we show the performance with and without the Nonlinear Schrödinger Kernel for five different classifiers: ridge regression, spectral regression for kernel discriminant analysis (SR-KDA), random forests, support vector machine (SVM) with RBF kernel, and neural network. The area under the curve (AUC) is reported using the original data with no kernel as input before (gray) and after processing with the Nonlinear Schrödinger Kernel (blue). The improvement achieved following processing with the Nonlinear Schrödinger Kernel is consistent for most algorithms and on all datasets, showing that the Nonlinear Schrödinger Kernel can improve the performance of a wide range of machine learning algorithms, both linear and nonlinear.

2.4 Practical Considerations: Impacts of Additive and Quantization Noise

We next investigate the limitations imposed by signal detection. When dealing with physical systems that are inherently analog in nature, the quantization noise of the analog to digital converter, as measured by its number of bits, can act as a performance bottleneck. This is particularly true for real-time systems that operate at high bandwidth [MCB17]. In addition, the number of bits used to store the model weights also needs to be considered because despite the improvement in accuracy, a higher number of bits requires larger memory.

To this end, Fig. 2.6 shows the robustness of the Nonlinear Schrödinger Kernel-projected representations across the various quantization resolutions applied to the input and internal weights of the model. For this purpose, we utilize the ridge regression model and quantize the model weights with different number of bits while simulating additive white Gaussian noise of 0, 5, and 10% of the data. The noise is added to the data after it is processed with the Nonlinear Schrödinger Kernel, and the error is evaluated for each combination of noise and quantization precision using the relative mean squared error (MSE), which is calculated between the predicted output of the ridge regression model and the ground-truth label and reported for various levels of quantization precision for the model weights. Finally, a continuous line that best fits the error values is plotted. For all four cases, the MSEs stay steady until the number of bits reduces below certain levels. For brain ICP and cell image, the MSE sharply increases when the number of bits reaches 2 or 3. For spoken digits data, the MSE significantly rises when the number of bits is less than 6. This difference can be explained by the property of data itself. As spoken digit datasets are recordings of sound waves, it is more dynamic compared to brain ICP or cell image waveform (Fig. 2.7 in appendices section 2.8). Therefore, it requires a higher resolution during quantization. It can also be observed that compared to quantization, the additive noise has much less influence on the prediction. Overall, the results in Fig. 2.6 are notable because they demonstrate the

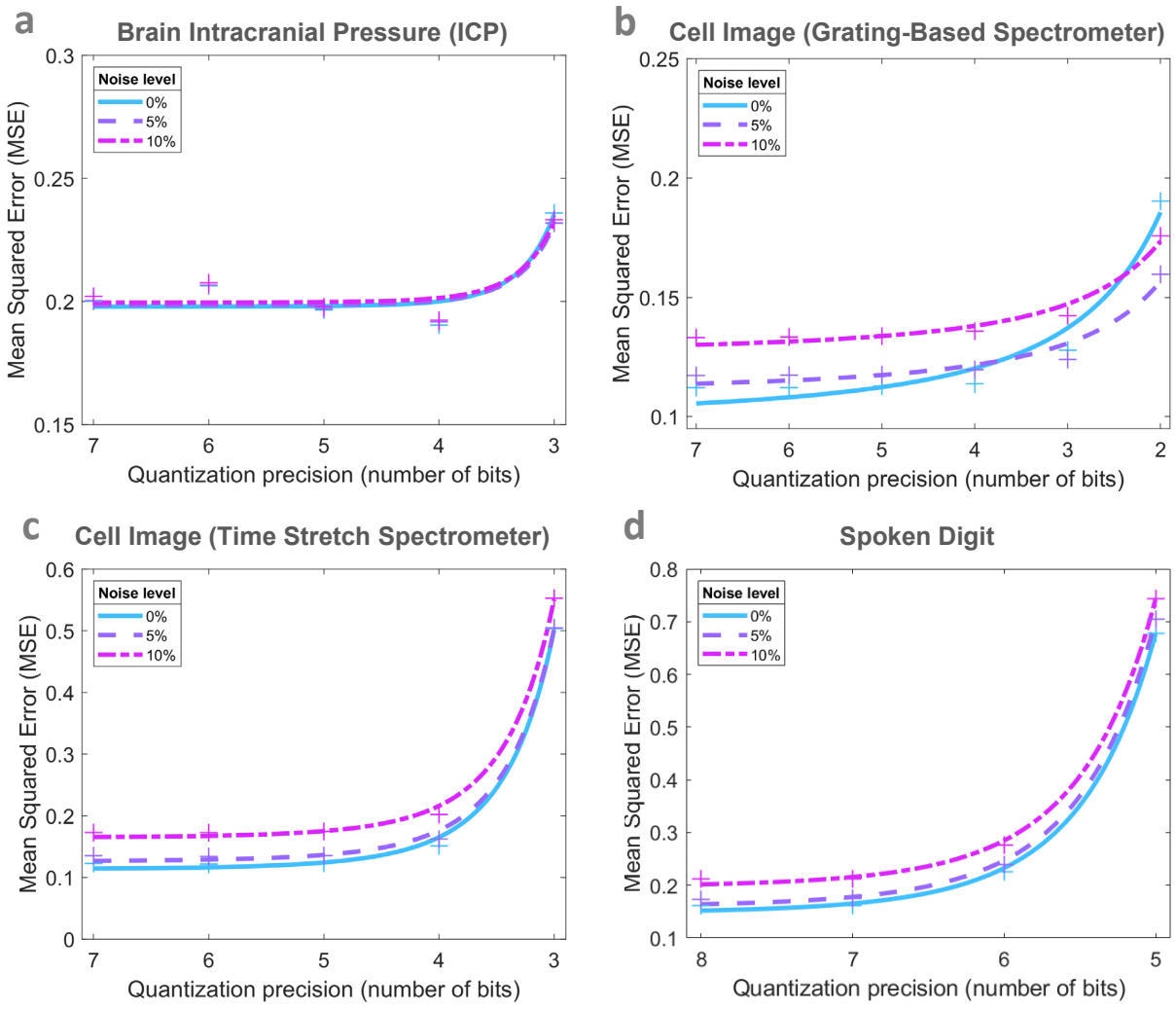

Figure 2.6: The robustness of the Nonlinear Schrödinger Kernel-based model against additive and quantization noises. A ridge regression model was trained on 4 datasets processed after the data were projected into an implicit high-dimensional space with the Nonlinear Schrödinger Kernel. The mean squared error (MSE) between the predicted output of the model and the ground truth is reported for various levels of quantization precision. Three levels of additive noise are reported: no noise, 5% and 10%. The lines are fits to the data points.

robustness of the Nonlinear Schrödinger Kernel technique. Additional results for nonlinear XOR operation can be found in appendices section 2.8

2.5 Discussion

In the present implementation of the Nonlinear Schrödinger Kernel, the input data is mapped onto the optical spectrum, in contrast to the commonly used temporal modulation. We choose spectral modulation, instead of temporal modulation, because the time scale of femtosecond pulse ($50fs$) is too short to allow modulation of complex waveforms. Expressed differently, there are no electro-optic modulators with such bandwidth. On the other hand, these short optical pulses have broad spectral bandwidth, which allows modulation of 100's or even 1000's of data samples onto the spectrum. We point out that many time stretch instruments, such as the time stretch microscope, operate based on this principle of spectral modulation of data onto a femtosecond pulse [CMT16][GAG12].

Fig. 2.5 shows that the performance of this technology depends on the dataset and in a small subset, the performance slightly decreases. This is expected because the optical kernel is an open-loop system, which means it is not trained for optimized performance on every single dataset. Furthermore, in the present system, the control of nonlinearity can only be achieved by varying the laser power and the length of the nonlinear element, so there are only two degrees of tuning. Also, for classification tasks, we don't rely on increases of dimensionality – the output spectrum is subsampled such that it has the same dimension as the input data. As a result, it is less flexible compared to trainable high-dimensional models such as deep neural networks. On certain datasets, such as spoken digit and brain ICP, the Nonlinear Schrödinger Kernel provides significant improvement. While the low degrees of tuning of the physical kernel imposes such limitations, it does have an important benefit. By allowing a simple linear classifier to perform nonlinear classification tasks, it eliminates the need for large training datasets and reduces latency by orders of magnitude.

2.6 Summary

The past decade has seen blazing advances in artificial intelligence for image and speech recognition and synthesis. Despite the increasing speed of digital processors, the execution time of these algorithms is still orders of magnitude slower than the time scales in ultrafast optical imaging, sensing, and metrology. In this chapter, we reported a new concept in hardware acceleration of AI that exploits femtosecond pulses for both data acquisition and computing. In the experiments, data is first modulated onto the spectrum of a supercontinuum laser. Then, a nonlinear optical element performs a data transformation analogous to a kernel operation projecting the data into an intermediate space in which data classification accuracy is enhanced. The output spectrum is sampled by a spectrometer and is sent into a digital classifier that is lightly trained. We showed that the nonlinear optical kernel can improve the linear classification results just like a traditional numerical kernel (such as the radial-basis-function) but with orders of magnitude lower latency. We further showed that this technique can work with various other digital classifiers. Finally, we demonstrated that the technique is resilient to nonidealities such as additive and quantization noise in the system. The performance is data-dependent due to the absence of training in the optical part of the system.

2.7 Matherials and Methods

We provide in this section (1) a description of the datasets, (2) the mathematical description of the Nonlinear Schrödinger Kernel, (3) details of the simulations model, (4) the experimental setup, (5) the machine learning algorithms, and (6) the machine learning evaluation methodology. The datasets are chosen to cover various classification applications and signal modalities including imaging, intracranial pressure, and audio recordings

2.7.1 Datasets

The dataset used in this study include the detection of intracranial hypertension, cell image classification for cancer detection via time stretch microscopy, and recognition of spoken digits. The details are as follows:

Continuous Intracranial Pressure (ICP) Waveforms: Traumatic brain injury (TBI) can lead to secondary brain injury due to pressure effects created by a hematoma and disrupted cerebrospinal fluid circulation. The immediate goal of TBI treatment is minimizing this risk by mitigating elevations in intracranial pressure (ICP). However, due to the wide heterogeneity of the degree of severity of TBI and the difficulty to assess them, the medical management of TBI patients is particularly challenging. This dataset originates from 154 patients [SLH12] admitted for various conditions related to the known risk of intracranial hypertension (IH). The majority of the patients (108 patients) were treated for brain injuries (TBI, subarachnoid hemorrhage (SAH), and intracerebral hemorrhage (ICH)). A total of 63,954 ICP alarms were recorded from bedside monitors. ICP signals were recorded continuously at a sampling rate of 240 Hz using ventriculostomy systems. In our experiments, a random sample of 434 alarms is used. Expert annotation was obtained by a biomedical engineer using dedicated annotation software created in our research laboratory and was asked to label them using the following criterion: an alarm is a false positive if there was no drainage to treat ICP elevation within 15 minutes following the alarm, a true alarm otherwise. This anonymized dataset of ICP signals originates from the University of California, Los Angeles (UCLA) Medical Center, and its usage was approved by the local institutional review board committee (IRB).

Spoken Digits: The Free-Spoken Digit Dataset (FSDD) [JSF18] is an audio/speech dataset consisting of 2000 recordings of spoken digits (0-9) in wav files at 8kHz. Each recording contains a spoken digit that was recorded under different conditions and from four different individuals. The recordings were trimmed so that they have near minimal silence

at the beginnings and ends. Our classification experiments are posed by considering the classification between digits "zero" and "one" (total number of 400 data examples). Each recording is described by 5,000 dimensions.

Time Stretch Cell Images: This dataset corresponding to cell images was from time stretch flow through microscope described in [CMT16][GAG12][LMC19]. Using a pair of diffraction gratings, a femtosecond laser pulse is converted into a collimated 1-D spatial rainbow. When this spatially and spectrally dispersed pulse illuminates the sample, the spatial information is modulated onto the optical spectrum. Using a low-loss dispersive fiber, the spectrum is then mapped into time and stretched such that it is slow enough to be digitized by a realtime analog to digital converter. In order to achieve high sensitivity under single-shot operation, the dispersive fiber is pumped to create simultaneous Raman amplification and time stretching. The sample was blood cells flowing serially through a microfluidic channel, and the details of which have been described in our earlier publications [CMT16]. The classification task amounts to differentiating between three cases: the background, the white blood cells (OT-II hybridoma T-cells), and the colon cancer cells (SW-480 epithelial) [CMT16]. This dataset corresponds to a random sample of 600 observations equally distributed between the three classes (i.e. background, normal, and cancer). Each data sample is described by 128 dimensions.

2.7.2 Mathematical Description

Here we provide a mathematical description of the Nonlinear Schrödinger Kernel. As illustrated in Fig. 2.1, the data is first modulated onto the E-field of the optical pulse in spectrum domain with a mapping function, M:

$$M : data(n) \rightarrow X(\omega) \tag{2.1}$$

where $data(n)$ is the nth sample of the input dataset, $X(\omega)$ is the laser optical spectrum after its modulation with the data. If the data is unipolar, it is modulated on the power

spectrum; if the data is bipolar, it is modulated on the field spectrum.

The data residing on the modulated laser pulse is projected onto another space when the pulse undergoes a nonlinear transformation characterized by the Nonlinear Schrödinger Equation (NLSE):

$$\frac{\partial X}{\partial z} = -\beta_1 \frac{\partial X}{\partial z} - \frac{j}{2}\beta_2 \frac{\partial^2 X}{\partial t^2} + \frac{\beta_3}{6}\frac{\partial^3 X}{\partial t^3} + j\gamma|X|^2 X - \frac{\alpha}{2}X \tag{2.2}$$

here z is the propagation distance, β_1 is the group delay, β_2 and β_3 are dispersion parameters, γ is nonlinear coefficient, and γ is the attenuation coefficient.

The output is obtained by solving the NLSE with an initial condition,

$$X(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} X(\omega)e^{j\omega t}d\omega \tag{2.3}$$

This partial differential equation can be solved numerically via Split-Step Fourier Method (SSFM). The SSFM assumes that for optical field propagating a small distance h, the dispersive and nonlinear effects act independently. More specifically, propagation from z to $z + h$ can be split into two steps. In the first step, the nonlinearity acts alone in the time domain. And in the second step, dispersion act alone in the spectrum domain [Agr12]. Mathematically:

$$SSFM(X(t)) = F^-1\{e^{h\hat{D}} \cdot \{F(X(t)) \cdot e^{h\hat{N}X(t)}\}\} \tag{2.4}$$

here, F is the Fourier Transform, $\hat{D}$ and $\hat{N}$ is the dispersion and nonlinear operator respectively, which can be written as:

$$\hat{D} = \frac{j}{2}\beta_2\omega^2 + \frac{j}{6}\beta_3\omega^3 - \frac{\alpha}{2} \tag{2.5}$$

$$\hat{N} = j\gamma[|X|^2 + \frac{2j}{\omega_0}\frac{\partial}{\partial T}(|X|^2 X) - T_R\frac{\partial|X|^2}{\partial T}] \tag{2.6}$$

here T_R is related to the slope of Raman gain spectrum.

The output of the nonlinear propagation is converted into the frequency domain by a Fourier transform performed by the diffraction grating or the time stretch spectrometer,

$$Y(\omega) = \int_{-\infty}^{+\infty} Y(t)e^{-j\omega t}dt \tag{2.7}$$

Signal detection is achieved by the optical to electrical converter, such as photodetectors, which measures the optical power:

$$I_{PD} = \eta |Y(\omega)|^2 \tag{2.8}$$

eta is the photosensitivity, I_PD is the photocurrent. The generated photocurrent is then quantized and send into a classifier in the digital domain:

$$C_x = H(\sum wI_{PD}) \tag{2.9}$$

here, C_x is the predicted class, H is the activation function, w is the trained weights.

2.7.3 Simulations

We first developed an optical simulation model in order to design the experiments and gain insight into the operation of the Nonlinear Schrödinger Kernel computing. It was further used to study the dependence of the classification accuracy with respect to signal to noise ratio and quantization noise of the digitizer which plays a paramount role in single-shot operations. Two simulations are reported in this chapter: classification on spoken digit dataset and XOR operation. The simulation parameters are set differently.

Since the spoken digit dataset has a high dimensionality (5000) that exceeds the limit of the spectral modulator in the experiment setup. So, we perform this task via simulation. In this simulation, the laser is modeled with a Gaussian pulse centered at $1550nm$ with $28nm$ bandwidth. The peak power of the input laser is set to $8200W$. The spectral modulator is modeled by multiplying the amplitude of the spectrum with the data. The propagation of the laser pulse inside the highly nonlinear fiber is modeled by solving the Nonlinear Schrödinger Equation (NLSE) using the split-step Fourier method (SSFM). In this system, the fiber model was a $50m$ highly nonlinear fiber with nonlinear index $\gamma = 15/(W \cdot km)$, dispersion D $= -4ps/(nm \cdot km)$, and dispersion slope $S = -0.019ps/(nm^2 \cdot km)$ at $1550nm$. The optical loss is $0.02dB/km$. The kernel output is resampled to maintain the same dimensionality as the input data (5000).

For XOR operation, the laser is centered at $1589nm$ with a super-Gaussian spectrum. The spectral bandwidth is $39.1nm$ to match the bandwidth of the spectral modulator. The peak power is $20W$. As for the nonlinear optical element, its length is set to $500m$ with $0.02dB/km$ optical loss. The dispersion is $1.43ps/(nm^2 \cdot km)$. And the dispersion slope is $0.04ps/(nm^2 \cdot km)$. The linear model is trained using the output spectrum of the kernel with four pairs of input: $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$. The model is then tested with the same data to show the nonlinear operation the kernel can perform. To further evaluate this technology, 400 pairs of input (x,y) are generated where $x,y \in [0,1]$. The output is tested with the trained model and the probability of predicting '0' can be calculated.

The software model is implemented in MATLAB and was executed on a server equipped with 64 GB memory and an Nvidia RTX TITAN GPU with 24 GB memory.

2.7.4 Experiment

The experimental setup is illustrated in Fig. 2.1. Its main components are a mode-locked laser source, a spectral modulator, a highly nonlinear fiber, and an optical spectrum readout mechanism. The data is first linearly interpolated to match the total bandwidth of the spectrum modulator, then used as an input to shape the amplitude of the spectrum filter applied on the broadband laser source by the spectral modulator. The spectrally modulated pulse propagates through the highly nonlinear fiber where the nonlinear evolution happens. The output is measured with the optical spectrum readout mechanisms such as traditional grating-based spectrometer or time stretch spectrometer [KCB99]. For every input data observation, a corresponding spectrum output is measured. The measured spectrum is used as an input for machine learning classification.

The optical source is a femtosecond mode-locked fiber laser followed by an Erbium-Doped Fiber Amplifier (EDFA) (ELMO from Menlo Systems). The source produces $< 90fs$ pulses with $90nm$ optical bandwidth centered at $1560nm$ with $30W$ optical peak power. The spectral modulator is a Waveshaper model 1000 S/L by Finisar which operates in

the L-band from $1567nmto1609nm$ with a $42nm$ bandwidth. It can control the amplitude and phase of an optical pulse by applying a reconfigurable spectrum filter with a 500-pixel resolution. The highly nonlinear fiber, manufactured by Corning, is a $500m$ fiber with nonlinear index $\gamma = 11/(W \cdot km)$, dispersion $D = -0.089ps/(nm \cdot km)$ and dispersion slope $S = 0.04ps/(nm^2 \cdot km)$ at 1550 nm. With this information, the dispersion at the center of the modulated pulse can be approximated:

$$D_{1588nm} = D_{1550nm} + S \cdot (1588 - 1550) = 1.43ps/(nm \cdot km) \tag{2.10}$$

Despite it's working in the anomalous dispersion regime, the system is still stable because its optical power is below the threshold for modulation instability. The grating-based spectrometer is Ando AQ6317B optical spectrum analyzer which is set to work at the range from $1550nm$ to $1630nm$ for Brain data (Fig. 2.4c) with $0.1nm$ resolution. For the cell images, the grating-based spectrometer (Fig. 2.4b) ranges from $1560nm$ to $1620nm$. The time stretch spectrometer is a $-101ps/(nm \cdot km)$ dispersion compensation fiber followed by a $10GHz$ Discovery DSC-R402 photodetector and a Tektronix DPO 71604 Oscilloscope with $16GHz$ electric bandwidth and $50GSample/s$ sampling rate. For classification tasks, the output of the Nonlinear Schrödinger Kernel is sampled so that it has the same dimensions as the input data.

2.7.5 Machine Learning Algorithms

Ridge Regression: Ridge regression [HK70] is used in our experiments as a baseline machine learning model. In the context of linear regression, the goal is to fit a function $f(x_i) = a^t x + b$ to the pairs of training data samples (x_i, y_i) that minimizes the residual sum of square (RSS):

$$RSS = \sum_{i=1}^{n} (f(x_i) - y_i)^2 \tag{2.11}$$

By concatenating '1' to each input vector x_i, the RSS can be written as a vector product:

$$RSS = (X^T a - Y)^T (X^T a - Y) \tag{2.12}$$

In ridge regression, a regularization term is added based on the norm of the model a:

$$RSS_{ridge} = \sum_{i=1}^{n}((a^T x_i - y_i)^2 + \alpha|a|^2) = RSS + \alpha|a|^2 \tag{2.13}$$

where model vector a is obtained as follows:

$$a = (XX^T + \alpha I)^{-1} Xy \tag{2.14}$$

and I is the identity matrix.

Random forest (decision tree): A random forest [Ho95] is an ensemble learning method that relies on a set of independently trained random forests. Random forests have successfully been used for classification and regression problems. They operate by constructing multiple random forests at the training phase, followed by aggregating their results by a majority vote (classification) or averaging (regression). In most applications, the random forest model tends to provide higher accuracy than single random forest approaches.

In a random forest, each tree is constructed according to a randomly sampled subset of observations and a randomly sampled set of variables. This allows a diverse set of learners that are then averaged, thus reduces the variance associated with a single tree, and decreasing the generalization error. In this work, the trees are constructed by the CART tree fitting method [BFS84]. At each iteration, a set of observations is split into disjoint subsets, such that a loss criterion is minimized. The resulting representation is a tree data structure in which each internal (non-leaf) node is labeled with a variable name and a corresponding split value, while a leaf is labeled with a fitted value, the class label for classification problems, or a numerical value for regression problems.

Spectral Regression for Kernel Discriminant Analysis (SR-KDA): Spectral regression (SR-DA) [CHH07] is a method developed to solve discriminant analysis as formulated in ridge regression RSS_{ridge} but using linear graph embedding (LGE). SR-DA proposes to can first find eigenvector y by solving $Wy = \xi Dy$ and then solve a regularized linear least square problem:

$$\underset{a}{\operatorname{argmin}} \|y - X^T a\|_2 + \xi\|a\| \tag{2.15}$$

This two-step process is termed spectral regression (SR) that enjoys two key benefits: (1) W and D are usually sparse and hence an efficient algorithm Lanczos is available to obtain eigenvectors; (2) Various matured techniques are available to solve the linear least squares (e.g., ridge regression). SR-DA is an application of SR in solving a supervised learning problem where y contains class labels for X. In this case, W becomes a block-diagonal with W_{ij} equal to $1/nC$ when both x_i and x_j belong to class C and nC is the number samples belonging to class C. SR-KDA [CHH11] generalizes SR-DA by projecting the input data onto a high-dimensional space via a kernel K, and class labels y to obtain vectors α. This is achieved using a Cholesky decomposition:

$$r = chol(K + \delta I) \tag{2.16}$$

$$\alpha = r \setminus (r^T \setminus y) \tag{2.17}$$

Support Vector Machines (SVM): A support vector machine (SVM) [CV95] is a supervised machine learning technique for classification problems where each sample $x_i \in X$ is labeled by $y_i \in \{-1, 1\}$. SVM aims at finding the optimal separating hyperplane that minimizes the misclassification rate on the training set while maximizing the sum of distances of the training samples from this hyperplane. Formally, this problem accounts for finding the parameter α, such that:

$$\underset{\alpha}{\operatorname{argmin}} \frac{1}{2}\alpha^T Q\alpha - e^T subjecttoy^T\alpha = 0, 0 \leq \alpha_i \leq C, i = 1, ..., n \tag{2.18}$$

where C is a constant that controls the amount of penalty on the error term during the minimization process, e is the vector of all ones, and Q is a matrix defined as:

$$Q_{ij} = y_i y_j K(x_1, x_i) \tag{2.19}$$

This minimization provides a solution for the following function:

$$f(x) = \sum_{i=1}^{n} (y_i \alpha_i K(x, x_i) + b) \tag{2.20}$$

Neural Network: Feedforward networks consist of a series of layers. The first layer has a connection from the network input. Each subsequent layer has a connection from the previous layer. The final layer produces the network's output. Feedforward networks can be used for any kind of input to output mapping. A feedforward network with one hidden layer and enough neurons in the hidden layers can fit any finite input-output mapping problem. By using stochastic gradient descent with momentum (SGDM) optimizer together with cross-entropy loss function, we train a 7-layer feedforward network.

2.7.6 Machine Learning Evaluation Methodology

Through our experiments, we aim at evaluating the effectiveness of the Nonlinear Schrödinger Kernel on various datasets and machine learning algorithms. The original input is processed through the Nonlinear Schrödinger Kernel to produce a modified signal that is used as input to a classifier. Standardization of the data is performed by subtracting the median for each feature. After standardization, the total range is rescaled in [0, 1].

All our classifications are performed using three-fold cross-validation. For a trained classifier, the receiver operating characteristics (ROC) curve is used to describe its sensitivity and specificity and summarized as the area under ROC (AUC) which provides a quantitative robust measure of classifier performance. The AUC parameter serves as an effective analysis metric for finding the best-performing classifier algorithms. The hyperparameters of each model are optimized independently using nested cross-validation, which is applied to the training dataset at each iteration. For cell image dataset, since it has three classes, we calculate the average AUC of three one vs. rest classifiers [PD03]. When studying the effect of nonlinearity, we calculate the classification error via 3-fold cross-validation using a linear SVM classifier.

Memory and compute resources may impose restrictions on the precision of the models and their input, and often limit deployment of the machine learning models on the edge devices. As part of our experiments, we tested the sensitivity of the model to both degree of

quantization and level of noise, as measured by the relative mean squared error (MSE). The quantization is applied to both the data (after Nonlinear Schrödinger Kernel) and model weights. In addition to that, additive noise (from 0% to 10%) is also used for further distorting the data.

2.8 Appendices

In this section, we present (1) Example output of the Nonlinear Schrödinger Kernel. (2) additional simulations showing the immunity to noise,

2.8.1 Example Output of the Nonlinear Schrödinger Kernel

Here we provide example output of the Nonlinear Schrödinger Kernel. Fig. 2.7 shows the example output of Nonlinear Schrödinger Kernel for all datasets. Compared to cell image and brain ICP, spoken digit dataset gives a much more complicated signal with high-frequency component and large dynamic range. Therefore, it requires higher quantization precision.

2.8.2 XOR Immunity to Noise

Here we demonstrate the robustness of the Nonlinear Schrödinger Kernel in nonlinear XOR operation. Fig. 2.8 demonstrates the robustness of the exclusive or (XOR) logic operation (which is the standard benchmark) by Nonlinear Schrödinger Kernel in presence of (a) quantization noise and (b) additive noise. In these simulations, the nonlinear element was the highly nonlinear fiber used in our experiments and described in the Materials and Methods section. The system can operate with a digitizer resolution as low as 2 bits and large additive noise. These results supplement the dependence of classification accuracy with bit depth and signal to noise ratio presented in the main text.

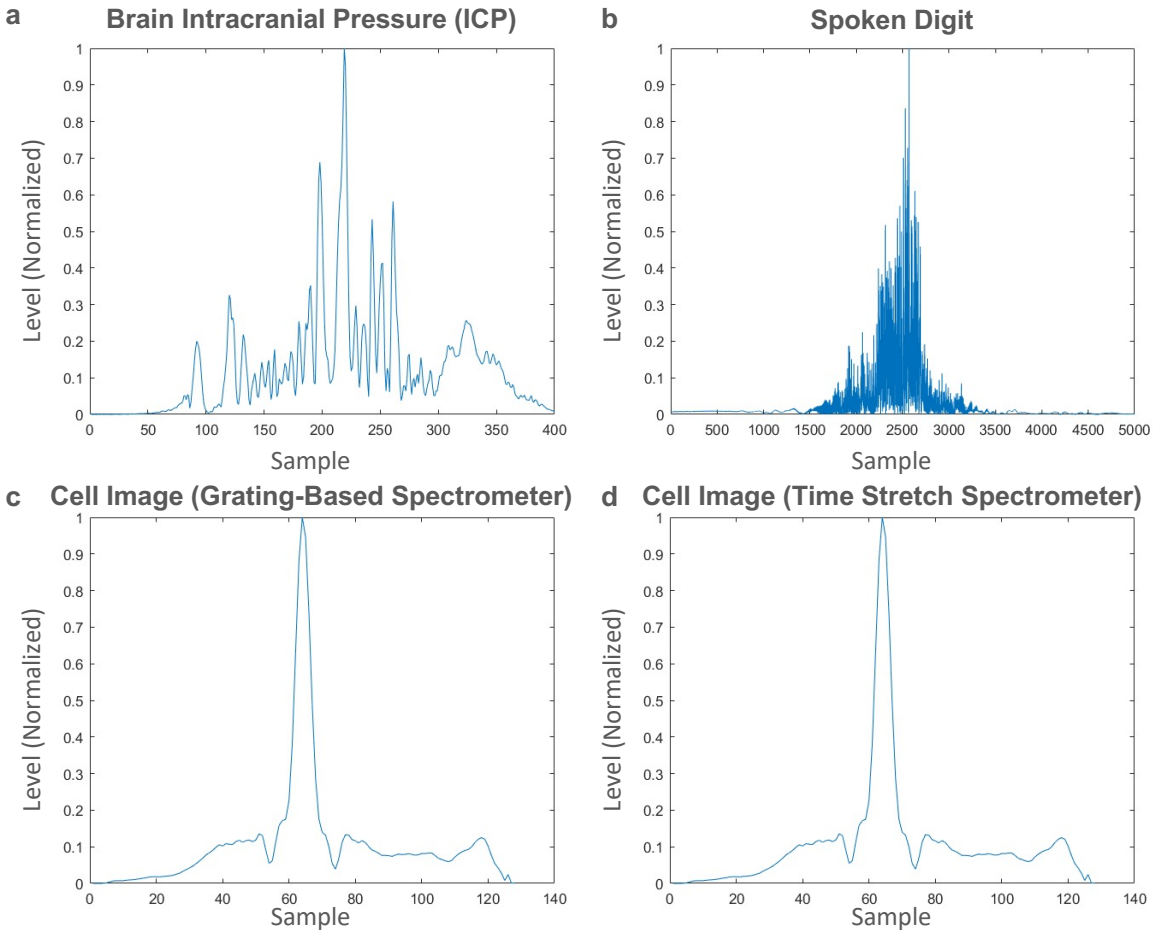

Figure 2.7: Example output of the Nonlinear Schrödinger Kernel for (a) Brain intracranial pressure (ICP) (b) Spoken digit (c) Cell image with grating-based readout (d) Cell image (Time stretch readout). The horizontal axis is the sample number, the vertical axis is the normalized amplitude level. The (c) and (d) are similar because the data are the same. They're only different in spectrum readout

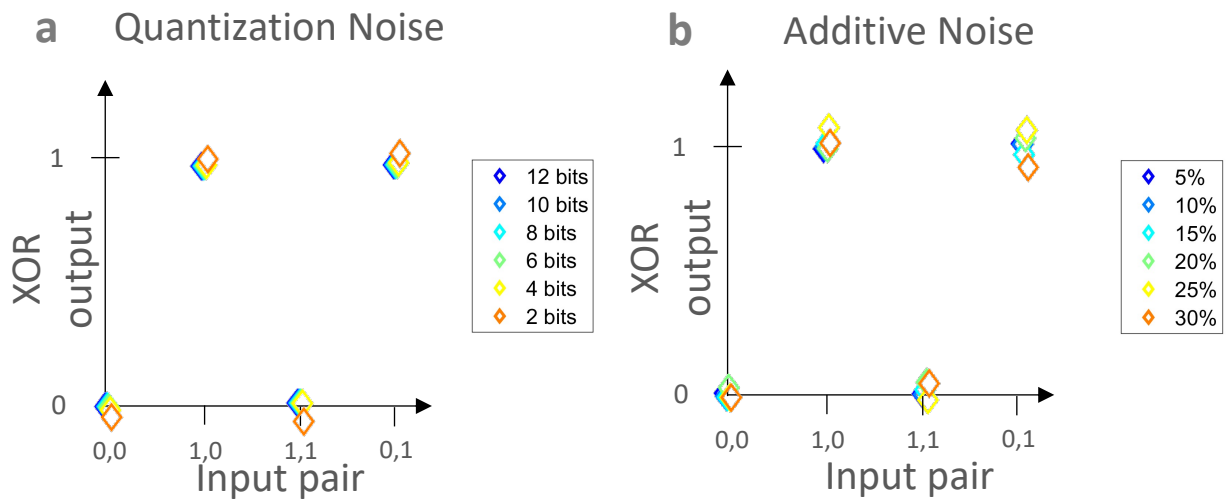

Figure 2.8: Exclusive OR (XOR) logic gate simulation with linear predictor and Nonlinear Schrödinger Kernel showing immunity to (a) quantization noise, and (b) additive noise. Since the output of the Nonlinear Schrödinger Kernel is analog, the resolution and sampling rate of the digitizer can impact the prediction. This figure shows the impact of quantization and noise on the classification when performing the XOR task. The experimentally measured optical spectrum after the Nonlinear Schrödinger Kernel is used to train a linear regression model. Then quantization and additive noises are added to the data. Inference is performed using the trained model on the noisy data and the values of the model output are recorded. In (a), we show the effect of quantization on the inference output. In (b), we show the effect of noise on the quantization output. Even at 2-bit quantization or with 30% additive noise, the linear regression output still very close to the ground truth. These results show that the Nonlinear Schrödinger Kernel offers strong robustness against noise

CHAPTER 3

Low Latency Computing for Time Stretch Instruments

In the last chapter, we introduced a novel computing framework, the Nonlinear Schrödinger Kernel. We demonstrate that this technique utilizes optical nonlinearities to accelerate machine learning data classification by projecting the data onto an implicit high-dimensional space in which classification accuracy is enhanced. A major limitation of this technique is the fixed optical transfer function, which prevents training and generalizability. Here we show that the optical kernel can be effectively tuned and trained by utilizing digital phase encoding of the femtosecond laser pulse leading to a reduction of the error rate in data classification.

3.1 Experimental System

Fig. 3.1 shows the system block diagram featuring closed-loop optimization of the optical kernel. This system contains a highly tunable Nonlinear Schrödinger Kernel and a digital feedback loop. In the tunable kernel, the input $data(n)$ is mapped into the spectrum domain ($data(\omega)$) and amplitude modulated onto a supercontinuum femtosecond laser pulse. The data then travels through a nonlinear optical element where it is nonlinearly transformed. The transformed data is finally captured by a spectrometer and sent to a backend digital classifier which adopts a light machine learning model. To avoid increasing data dimensions, the output spectrum of the kernel is sampled to match the dimension of the input data. The tuning is achieved by modulating a spectral phase $\varphi(\omega)$ onto the laser pulse during spectral modulation, equivalent to encoding the phase of the input data. This 'phase code' modifies

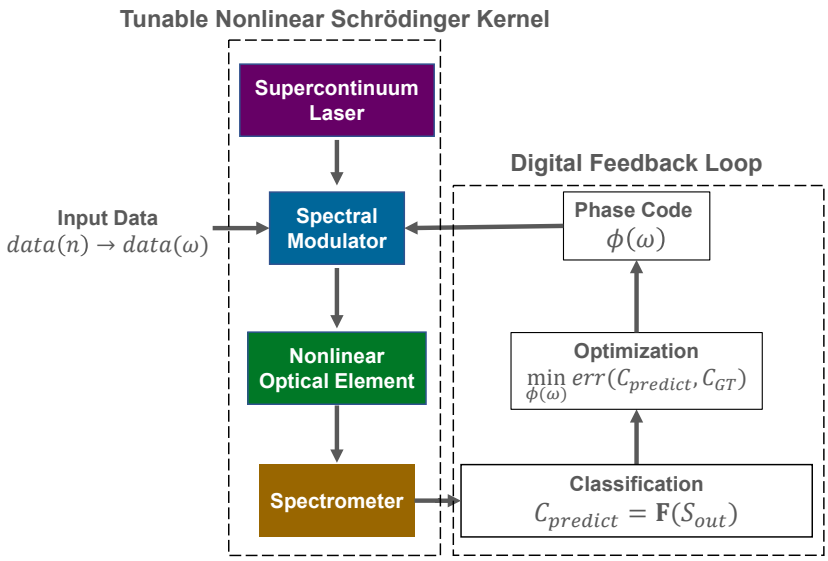

Figure 3.1: Optimization of a tunable Nonlinear Schrödinger Kernel. The system contains a tunable Nonlinear Schrödinger Kernel and a digital feedback loop. In the tunable Nonlinear Schrödinger Kernel, the phase-encoded input $data(n)$ is mapped onto the spectrum of a supercontinuum laser via spectral modulation. The modulated laser propagates through a nonlinear optical element, where the nonlinear process is engineered by the phase code $\varphi(\omega)$. The output spectrum of the nonlinear optical element S_{out} is then acquired using a spectrometer and sent to a classifier F. The classification error is calculated by comparing the predicted class $C_{predict}$ and the ground truth C_{GT}. This error is used as an input to an optimization algorithm to update the phase code for achieving lower classification errors. The details can be found in the Materials and Methods (section 3.6).

the nonlinear interaction in the nonlinear element and hence engineers the optical kernel. In the digital feedback loop, the algorithm compares the predictions with the ground truth and calculates the average classification error for the whole dataset. Subsequently, a new phase code is generated by an optimizer who aims at minimizing this error. The system runs iteratively to obtain the optimal phase code, through which the optical kernel is trained for optimal performance. The details of the experiment system can be found in the Materials and Methods (section 3.6).

3.2 The Crucial Role of Optical Nonlinearity

In Fig. 3.2 we show via simulation the evolution of a femtosecond pulse, that has been spectrally modulated with data, through the nonlinear optical element. The output spectrum is classified using a simple machine learning algorithm. Also shown in Fig. 3.2 is the error produced by using a linear support vector machine (SVM) classifier as the digital backend. Here the nonlinear kernel is fixed, i.e. not tuned. As described in the Methods section, the data is the 1-D linescan images of biological cells flowing through a microfluidic channel. The images are captured by a time stretch microscope [CMT16][LMC19]. The dataset contains three types of images: (1) no cell is present, (2) a normal cell, and (3) a cancer cell. The classification task is to distinguish three different types.

In Fig. 3.2a, a linear kernel is simulated by setting the nonlinear coefficient to zero. As expected, the optical spectrum remains unchanged. As shown in Fig. Fig. 3.2c the classification error remains almost unchanged at 14.9% compared to the 14.7% baseline error obtained by feeding the data directly into the backend digital classifier. Fig. Fig. 3.2b shows the propagation of the same spectrally modulated pulse in the nonlinear element. To avoid an increase in data dimension, we operate in the spectrum narrowing regime. This occurs when the pulse undergoing self-phase modulation has a negative chirp [Agr12]. The output spectrum is sampled such that it has the same dimensions as the input data (128) as

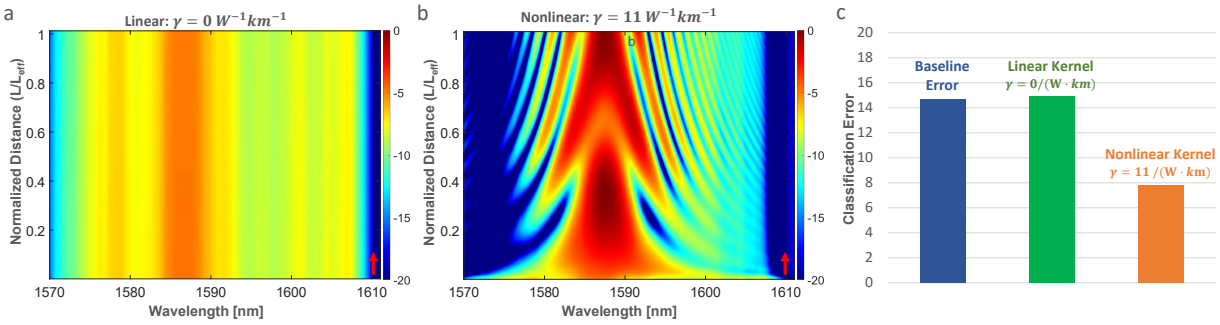

Figure 3.2: (a) The evolution of the optical spectrum in the linear optical kernel where nonlinear coefficient $\gamma = 0/(W \cdot km)$(b) The evolution of the optical spectrum in the nonlinear optical kernel where $\gamma = 11/(W \cdot km)$. (c) Bar chart comparing the classification error for three cases: the baseline error calculated without kernel (blue, 14.7%), the linear optical kernel (green, 14.9%), and the nonlinear optical kernel (orange, 7.8%). The baseline error is calculated by directly feeding the input data to the digital backend, which in this case is a linear support vector machine (SVM) classifier. For (a) and (b), the horizontal axis is the wavelength, the vertical axis is the propagation distance (normalized to the effective length of the optical element). The color indicates the optical intensity in the log scale with color bar on the side. The red arrow point to the propagation direction.

explained in the Methods section. The nonlinear transformation reduces the classification error to 7.8%, confirming the utility of the optical kernel in enhancing machine learning without sacrificing (ie. increasing) the data dimensionality.

Comparing the effect of linear and nonlinear kernels, it can be observed that the enhancement in classification accuracy cannot be achieved without optical nonlinearity. However, as a proper machine learning technique, the nonlinearity must be tunable so it can be optimized. Previous research has demonstrated the successful control of optical nonlinearity using the spectral phase modulation of the input light [CBJ05]. Here we apply to same technique to tuning and optimization of the nonlinear optical kernel where the optimization is guided by the classification error.

3.3 Training of Optical Nonlinearities for Machine Learning

In this section, we demonstrate that the nonlinear optical kernel can be tuned by applying phase encoding to the input data. As mentioned in the introduction, the intuition behind this approach is as follows. Nonlinear optical interactions such as self-phase modulation, Four Wave Mixing (FWM), etc. are coherent in nature, i.e. they are sensitive to the phase of the input pulse. It then follows that manipulating the input phase influences the output produced by the optical nonlinearity.

The experimental implementation is shown in Fig. 3.1. Spectral phase modulation within a digital feedback loop controls the nonlinear optical interactions and hence tunes the optical kernel. A genetic algorithm arrives at the optimal phase code that minimizes the error of the digital classifier. The results for three datasets are experimentally demonstrated. The datasets described in the Materials and Methods (section 3.6) include cell images (Fig. 3.3a) [CMT16][LMC19], phalanges bones outline (Fig. 3.3b)[BLB17], and EEG (Fig. 3.3c) [DG17].

Fig. 3.3 shows that, in all three datasets, both the trained (orange), as well as the fixed

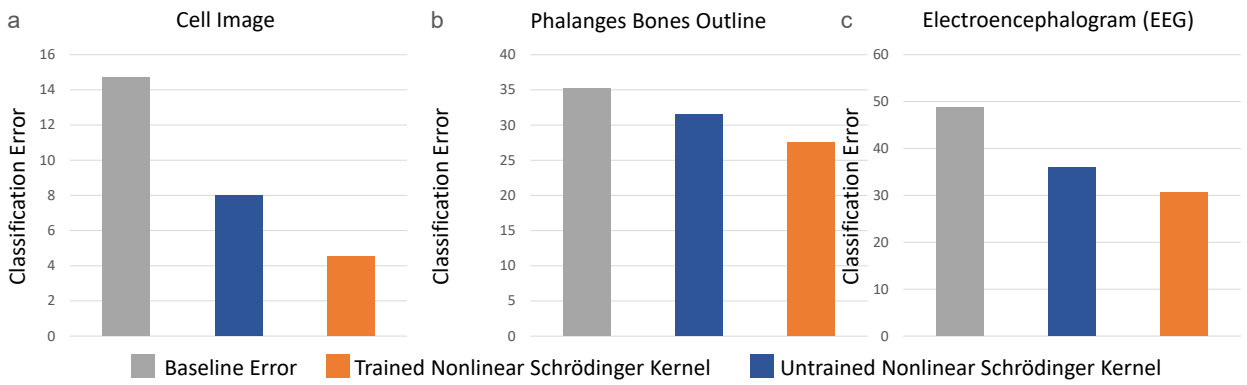

Figure 3.3: Optimization of Nonlinear Schrödinger Kernel on three datasets: (a) Time stretch biological cell image (b) Phalanges bones outline (c) Electroencephalogram (EEG). In each bar chart, the classification error for three cases is compared: Baseline error (gray), untrained Nonlinear Schrödinger Kernel (blue), and trained Nonlinear Schrödinger Kernel (orange). The baseline error is calculated by feeding the input data directly into the digital backend – a linear support vector machine (SVM) classifier. All the results are calculated via 3-fold cross validation.

(untrained) kernel (blue), produce a lower error rate compared to the baseline case where no optical kernel is used (gray). However, the trained optical kernel leads to a lower error than the fixed kernel. The baseline error is obtained by feeding the input data directly to the digital backend. The results prove that the phase encoding can effectively tune the performance of the Nonlinear Schrödinger Kernel, and combined with the digital feedback loop renders the optical kernel trainable.

In these experiments, the phase code is generated using a polynomial with two tunable parameters for the second-order and the third-order coefficients. This scheme can be easily expanded to more parameters to provide additional degrees of freedom. Details of the experiments are also in the Materials and Methods (section 3.6).

3.4 Limitations

Some of the limitations of this technique and potential future research are as follows. First, the maximum allowed dimension of the input data is dependent on spectral modulation. In our experiments, we use a commercial waveshaper (details in the Methods section) with 500 pixels. This limits the maximum dimension of the input data to 500. Second, even though the kernel can be trained, the performance in terms of classification error is still data-dependent, as seen in 3.3. Such is the case with all machine learning techniques because they are statistical in nature (as opposed to deterministic). One possible direction for future research is to correlate the classification performance with the properties of the input data to identify the type of data for which optical kernel computing is most effective.

3.5 Summary

The recently introduced optical kernel computing utilizes optical nonlinearities to transform data such that nonlinear classification can be done with a computationally light linear digital

classifier. The so-called Nonlinear Schrödinger Kernel computing is ideal for low latency classification of data that is modulated onto the spectrum of femtosecond lasers. Such is the case with time stretch imaging and spectroscopy instruments [MCB17][ZCJ22]. In the previous implementation of this technique, the property of the kernel was entirely governed by the nonlinear coefficient of the optical medium. Hence the kernel could not be trained or optimized as required in machine learning tasks. In this chapter, we presented a solution to this predicament by introducing spectral phase modulation of the input pulse within a digital feedback loop. Phase modulation influences how data is transformed by nonlinear optical interactions and allows the optical kernel to be trained. The training is shown to reduce the classification error on three diverse datasets.

3.6 Materials and Methods

In this section, we provide details of (1) experimental implementation for kernel optimization, (2) simulation study showing the critical role of optical nonlinearities, (3) mathematical formulation of the optical kernel, and (4) the datasets and machine learning model.

3.6.1 Experiments

The physical implementation of the closed loop optical kernel computing system follows Fig. 3.1. The supercontinuum laser is a mode-locked Erbium-doped fiber laser (ELMO) followed by an Erbium-doped fiber amplifier (ELMA), both from Menlo Systems. It produces $< 90fs$ laser pulses centered at $1560nm$ with $90nm$ bandwidth. The approximate pulse peak power is 30 W after the amplifier. The data modulates the optical spectrum using a Finisar (now II-VI / Coherent) Waveshaper model 1000 S/L. It operates in the L band ($1567nm$ $1609nm$) with a 500-pixel resolution. The Waveshaper performs both amplitude and phase modulation simultaneousl. The amplitude is the input data (scaled between 0 to 1), while the phase is the (phase) code that is applied to tune the nonlinear effects. The nonlinear

optical element is a $500m$ highly nonlinear fiber (HNLF) from Corning. It has a nonlinear coefficient $\gamma = 11/(W \cdot km)$, a low dispersion of $D = 1.43ps/(nm \cdot km)$, and dispersion slope $S = 0.04ps/(nm^2 \cdot km)$ at $1588nm$. An Ando AQ6317B optical spectrum analyzer with $0.1nm$ resolution measures the output spectrum. The sampling range of the spectrum analyzer is set differently for each dataset depending on the actual bandwidth of the output. For the cell image and phalanges bones outline dataset, it is $1545nm \quad 1635nm$. For the EEG dataset, it is $1550nm \quad 1630nm$. The measured spectrum is resampled so that it has the same dimension as the input. It is then min-max standardized and sent to a linear SVM (support vector machine) classifier.

The digital feedback loop is implemented in MATLAB on a computer with 16 GB RAM memory. It runs iteratively to optimize the performance of the Nonlinear Schrödinger Kernel. In each iteration, the classification error rate is the average over the entire dataset and is calculated via 3-fold cross validation. It is sent to the optimizer which is a genetic algorithm [Hol75] from MATLAB Global Optimization Toolbox. To minimize the classification error, it compares the result from the current iteration with the previous ones and generates the phase code for the next iteration. In this paper, the phase code is calculated by a third-order polynomial with adjustable coefficients,

$$\varphi(\omega) = a\omega^2 + b\omega^3 \tag{3.1}$$

here, a and b are the tunable coefficients generated by the optimizer. The first-order coefficient is not taken into consideration as it simply causes a constant delay and does not affect the nonlinear optical process. The optimization results are shown in Fig. 3.3. The optimal coefficients for the cell image dataset (Fig. 3.3a) are shown in Table 3.6.1. The optimized classification error, $Err_{trained}$ are also attached in comparison with the unoptimized classification error $Err_{untrained}$ and the baseline error $Err_{baseline}$.

	Cell Imgage	Phalanges Bones	EEG
$a(\times 10^{-24})$	1.353	0.854	1.022
$b(\times 10^{-37})$	-3.628	0.925	-4.742
$Err_{trained}$	4.5%	27.5%	30.7%
$Err_{untrained}$	8.0%	31.5%	36.04%
$Err_{baseline}$	15.0%	35.2%	30.7%

Table 3.1: Optimal Phase Code

3.6.2 Insight into the Critical Role of Optical Nonlinearity

To assess the role of optical nonlinearities in the operation of the Nonlinear Schrödinger Kernel, a computer model is created in MATLAB. The supercontinuum laser source is modeled as a transform-limited pulse with a supper-Gaussian spectrum centered at $1588nm$. It has a $40nm$ bandwidth consistent with the passband of the spectral modulator (waveshaper). The spectral modulation is modeled by multiplying the input data (time stretch cell image data, scaled between 0 to 1) by the laser spectrum. For these simulations, the phase code is set to 0 to simulate the unoptimized (open loop) system. The modulated laser pulse is then sent through a nonlinear optical element, in this case, an HNLF. The complex propagation in the fiber is modeled by solving the time domain Nonlinear Schrödinger Equation (NLSE) using the split-step Fourier method (SSFM) implemented in MATLAB. This algorithm divides the fiber into short segments (steps) to separate different effects, computing iteratively to obtain an approximate solution [Agr12]. In each step, the spectrum of the optical pulse is recorded to track the evolution of the input, as shown in Figure 2. The output of the HNLF is measured using a spectrometer, which is modeled by a fast Fourier transform (FFT) and absolute square. Finally, the collected spectrum is resampled so that it has the same dimension as the input, mean-std standardized, and sent to a linear SVM classifier.

For the simulation shown in Fig. 3.2, the length of the fiber is set to $500m$. The optical loss is $0.02dB/km$, the dispersion coefficient (D) is $1.43ps/(nm \cdot km)$, the dispersion slope

(S) is $0.04ps/(nm^2 \cdot km)$ at $1588nm$, and the optical power P is $20W$. For the linear kernel, the nonlinear coefficient (γ) is $0rad/(W \cdot km)$. While for the nonlinear kernel, $\gamma = 11rad/(W \cdot km)$. The spectrum is sampled from $1570nm$ to $1610nm$ with 128 sampling points to match the dimension of the input data.

The MATLAB model runs on a server equipped with 64 GB memory and an NVIDIA RTX TITAN GPU with 24 GB memory. The SSFM algorithm is modified from the open-source function ssprop and accelerated by implementation on GPU using CUDA.

3.6.3 Mathematical Model

The mathematical description of the Nonlinear Schrödinger Kernel computing is provided in this section. As shown in 3.1, the data is first mapped into the optical spectrum by assigning each entry of the data to a frequency. In the current experiments, the data resides in a file and is modulated onto the femtosecond pulse spectrum using a waveshaper. The waveshaper performs the following mapping M:

$$M : data(n) \rightarrow data(\omega) \tag{3.2}$$

$data(n)$ is the input where n is its n-th entry, and ω is the corresponding optical frequency. This filter is then applied to a supercontinuum laser pulse:

$$E_{in}(\omega) = \sqrt{data(\omega)} \cdot E_{laser} e^{\varphi(\omega)} \tag{3.3}$$

here, E_{laser} is the supercontinuum laser source, $\varphi(\omega)$ is the phase code which is also applied by the waveshaper for the purpose of training. E_{in} is the modulated pulse which will then enter the nonlinear stage. The data is modulated onto the power spectrum, which is equivalent to multiplying its square root by the E-field. The modulated laser is sent into a nonlinear optical element, where complex nonlinear transformation is governed by the NLSE:

$$\frac{\partial E(t)}{\partial z} = (\hat{D} + \hat{N})E(t) \tag{3.4}$$

here E is the electric field of the optical pulse, z is the propagation distance, $\hat{D}$ is the dispersion operator, and $\hat{N}$ is the nonlinear operator, which can be calculated from the parameters of the nonlinear optical element [11]. By solving NLSE using the initial condition:

$$E_i nt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i\omega t} E_{in}(\omega) d\omega \tag{3.5}$$

The output $E_{out}(t)$ is thus obtained. The nonlinearly trasnformed input data is subsequently acquired along with the laser spectrum $S_{out}(\omega)$. According to the Fourier transform, the spectrum is:

$$E_{out}(\omega) = \int_{-\infty}^{+\infty} e^{-i\omega t} E_o ut(t) dt \tag{3.6}$$

$$S_{out}(\omega) = |E_{out}(\omega)|^2 \tag{3.7}$$

The classification is then performed on the measured output spectrum:

$$C_{predict} = F(S_{out}) \tag{3.8}$$

here, C_{redict} is the predicted class, and F is a linear SVM classifier. We chose the linear SVM for its ubiquity and simplicity. Other classifiers can also be used.

In the feedback loop, the classification error $err(C_{predict}, C_{GT})$ is calculated by comparing the prediction and the ground truth. Since C_{predic} is a function of phase code $\varphi(\omega)$, we can obtain the optimal $\varphi(\omega)$ by minimizing the error:

$$\min_{\varphi(\omega)} err(C_{predict}, C_{GT}) \tag{3.9}$$

In this paper, the optimizer is a genetic algorithm, as mentioned in the experiments section.

3.6.4 Datasets and Machine Learning Model

The performance of the trained Nonlinear Schrödinger Kernel is evaluated with three datasets: time stretch cell image [CMT16][LMC19], phalanges bones outline [BLB17], and EEG [DG17]. The details are provided as follows.

The time stretch cell image dataset is the image of cells acquired in time stretch microscopic flow cytometry [CMT16][MCB17]. In this setup, a femtosecond laser pulse is spatially dispersed into a collimated 1-D rainbow using a pair of diffraction gratings. The rainbow illuminates a microfluidic channel while sample blood cells pass by. Through this, the spatial features of the illuminated cells are modulated onto the spectrum of the femtosecond laser pulse. Those features are then read out in realtime using a time stretch spectrometer, where a low-loss dispersive fiber maps the spectrum in time. The temporal waveforms that mimic the optical spectrum are digitized by a realtime analog to digital converter. Each frame is one linescan of the cells moving along with the microfluidic flow. The details of the setup are described in our earlier publications [MCB17]. The dataset contains three types of linescans: the background, the colon cancer blood cells (SW-480 epithelial), and the white blood cells (OT-II hybridoma 497 T-cells). Each class includes 200 observations (waveforms), and each waveform has 128 dimensions (spatial features). The details can also found in 2.7.1

Phalanges bones outline is an open-source dataset from the UCR Time Series Classification Archive [BLB17]. It studies the correctness of an automatic phalanges bone outline extraction algorithm. The algorithm is applied to X-ray images for extracting the outlines of three bones of the middle finger (phalanges). The extraction is then evaluated by three human evaluators. This paper uses a selection of 400 observations evenly distributed in two classes: correct outlines and incorrect outlines. The details of the dataset can be found in [Dav13].

The EEG dataset is an open-source dataset from UCI Machine Learning Repository [DG17]. It comes from the test that measures the brain's electrical activity. The data is collected using the Emotiv EEG Neuroheadset for studying the correspondence between EEG signals and human eye motions. This paper uses a subsample of 400 observations from the dataset with two eye motion states (200 for each): eye open and eye closure. Each observation has 14 dimensions.

The backend digital classifier in this study adopts a linear SVM model [Vap63]. The

classification errors are calculated via three-fold cross validation.

CHAPTER 4

Additional Study on the Nonlinear Schrödinger Kernel

In previous chapters, we introduced a novel low-latency computing technology for ultrafast femtosecond instruments. It utilizes the complex nonlinear process in the optical components to project the data onto an implicit high dimensional space, where the classification can be achieved using simple lightweight digital classifiers. Since this projection is implicit, there's no actual increase in the data dimension. More importantly, the inference latency of this Nonlinear Schrödinger Kernel is in microsecond timescale, which is orders of magnitude lower than the traditional numerical kernel.

In this chapter, we report additional studies on the characteristics of the Nonlinear Schrödinger Kernel. We first explore the further acceleration of the optical kernel, and then investigate its dimension effect. Finally, we describe some future works that can be done to help understand and deploy this optical kernel.

4.1 Accelerating Nonlinear Schrödinger Kernel

The Nonlinear Schrödinger Kernel has shown impressive acceleration of machine learning data classification. It outperforms the traditional numerical method by orders of magnitude (Section 2.3). To further accelerate classification, we perform a theoretical study on a simulated Nonlinear Schödinger Kernel using the model built in chapter (3). We divide the system into two parts, the physical processing in the optical system, and the digital computing in the computer. In this section, we first demonstrate that we can accelerate

the physical processing by substituting the lengthy HNLF with an integrated waveguide, then demonstrate that we can accelerate the digital computing by implementing the linear classifier onto an edge device.

4.1.1 Acceleration in the Optical System

In previous chapters, we report a Nonlinear Schrödinger Kernel using a highly nonlinear fiber (HNLF). Even with fascinating results, it can hardly deny that meters - even hundreds of meters of fibers are required for the system to properly function. On the other hand, the same 3rd-order optical nonlinearities are also present in integrated waveguides composed of materials such as silicon and silicon nitride. Since the scale of the waveguides is usually in cm or mm, if implemented into the kernel, can easily achieve picosecond latency. We demonstrate the successful operation of an optical kernel with a waveguide on the XOR task (Fig. 4.1). Two different waveguides are used in this simulation: silicon [BJ04][LEP17] and silicon nitride [TON18]. The silicon waveguide has a dimension of $220nm \times 480nm$ with $10mm$ in length. Such a waveguide has $-850ps/(nm \cdot km)$ dispersion and $16000/(W \cdot km)$ nonlinear coefficient at $1588nm$ (center wavelength of the modulated pulse). The silicon nitride has a dimension of $550nm \times 300nm$ with a length of $0.2mm$. The dispersion at $1588nm$ is $223.68ps/(nm \cdot km)$ and nonlinear coefficient is $500,000/(W \cdot km)$. In both cases, we use a laser source with $50W$ peak power and $45nm$ bandwidth. The optical spectrum after Nonlinear Schrödinger Kernel is read out by a spectrometer with $200nm$ bandwidth ($1500nm$ to $1700nm$) and $0.1nm$ resolution.

4.1.2 Acceleration in the Digital Classifier

The successful operation demonstrated in Fig. 4.1 shows that the integrated nonlinear waveguide can be used in place of the lengthy HNLF. This reduces the delay caused by the propagation inside the nonlinear optics of several orders of magnitude. Another bottleneck

XOR		Prediction before Nonlinear Schrödinger Kernel		Prediction after Nonlinear Schrödinger Kernel (Silicon)		Prediction after Nonlinear Schrödinger Kernel (Silicon Nitride)	
Input	Ground Truth	Score	Output	Score	Output	Score	Output
00	0	0.99	1	-0.88	0	-0.99	0
01	1	1.05	1	0.95	1	1.00	1
10	1	-1.05	0	1.00	1	1.00	1
11	0	-0.99	0	-1.07	0	-1.00	0

Figure 4.1: The latency can be greatly reduced by using an integrated waveguide instead of an optical fiber to induce 3rd-order nonlinearities. Exclusive OR (XOR) operation results of the Nonlinear Schrödinger Kernel using an integrated waveguide as nonlinear optical components. Results are from numerical simulations. Two different waveguides are used in this simulation using the model built in chapter 3: silicon [BJ04][LEP17] and silicon nitride [TON18]. The silicon waveguide has a dimension of $220nm \times 480nm$ with $10mm$ in length. Such a waveguide has $-850ps/(nm \cdot km)$ dispersion and $16000/(W \cdot km)$ nonlinear coefficient at $1588nm$ (center wavelength of the modulated pulse). The silicon nitride has a dimension of $550nm \times 300nm$ with a length of $0.2mm$. The dispersion at $1588nm$ is $223.68ps/(nm \cdot km)$ and nonlinear coefficient is $500,000/(W \cdot km)$. We use a laser source with $50W$ peak power and $45nm$ bandwidth in both cases. The optical spectrum after Nonlinear Schrödinger Kernel is read out by a simulated spectrometer with $100nm$ bandwidth and $0.1nm$ resolution. A linear support vector machine (SVM) is trained using the spectrum. The table shows the predictions of the trained model as well as the corresponding scores. It demonstrates that the waveguide can also be used as a nonlinear optical component and could greatly reduce the latency because of the compact size.

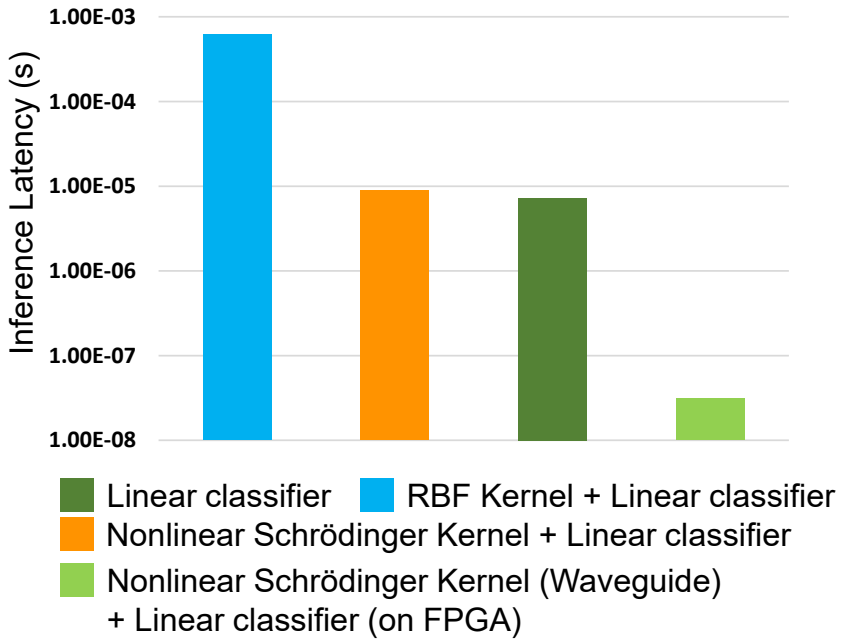

Figure 4.2: Inference latency of a Nonlinear Schrödinger Kernel using a waveguide as the nonlinear element and Fied Programmable Gated Array (FPGA) as digital backend. The waveguide used in this system is a $1.3mm$ silicon nitride with a dispersion of $-233ps/(nm \cdot km)$, and a nonlinear coefficient of $500000/(W \cdot km)$. The FPGA parameters are taken from a Xilinx Zync UltraScale+ RFSOC ZU42DR.

is the speed of the digital classifier, as shown in Fig. 2.4, where the inference latency of the optical kernel is almost the same as a linear classifier. This can be resolved by simply implementing the digital linear classifier onto an edge device, such as a Field Programmable Gated Array (FPGA). Here we analyze the inference latency of a kernel built on an integrated silicon nitride waveguide and an FPGA as a digital backend. The results are shown in Fig. 4.2. The data used in this example is the time stretch cell image data [CMT16][LMC19].

In this system, the length of the silicon nitride waveguide is $1.3mm$, so the propagation latency is:

$$t_{prop} = n\frac{L_{Si_3N_4}}{c} = 2 * 1.3e-3/299792458 = 8.67e-12s = 8.67ps \tag{4.1}$$

here t_{prop} is the pulse propagation time inside the waveguide, n is the refractive index, c is the speed of light.

The latency of the digital classifier is determined by the computation speed of the FPGA. For a Xilinx Zync UltraScale+ RFSOC ZU42DR, the computation of a linear classifier requires 25 clocks. Using an 800 MHz clock, the latency is $25 \times 1.25ns = 31.25ns$

As a result, the total inference delay is 31.258 ns, which is another three orders of magnitude lower than the original Nonlinear Schrödinger Kernel. Since the FPGA can only perform simple computations, highly nonlinear models such as radial basis function (RBF) cannot be implemented. Therefore, the acceleration brought by the optical kernel becomes 5 orders of magnitude compared to the numerical RBF Kernel.

4.2 Dimension Effect of the Nonlinear Schrödinger Kernel

Next, We turn to the dimension study. In this system, the data is mapped onto the optical spectrum, and the nonlinear transformation is achieved by shaping this spectrum. This is a highly dynamic process, which can be affected by various parameters, including the value of the 2nd-order dispersion [Agr12]. Fig. 4.3 shows examples of nonlinear processes inside the nonlinear optical component under different conditions. Here, the nonlinear processes are illustrated by plotting the evolution of the optical spectrum of the laser (modulated with time stretch cell image data[CMT16]). Three cases: normal dispersion, anomalous dispersion, and zero dispersion are shown here. For normal dispersion, there're two stages of evolution. In the first stage, when the peak power of the optical pulse is high, self-phase modulation (SPM) dominates, which results in the broadening of the laser spectrum. During this, however, group velocity dispersion also gradually brings down the peak power which is essential for SPM. This led to the second stage, where the dispersion dominates the process and the optical spectrum remains unchanged, as shown in Fig. 4.3a. When it comes to anomalous dispersion 4.3b, the situation becomes more complicated. As a result

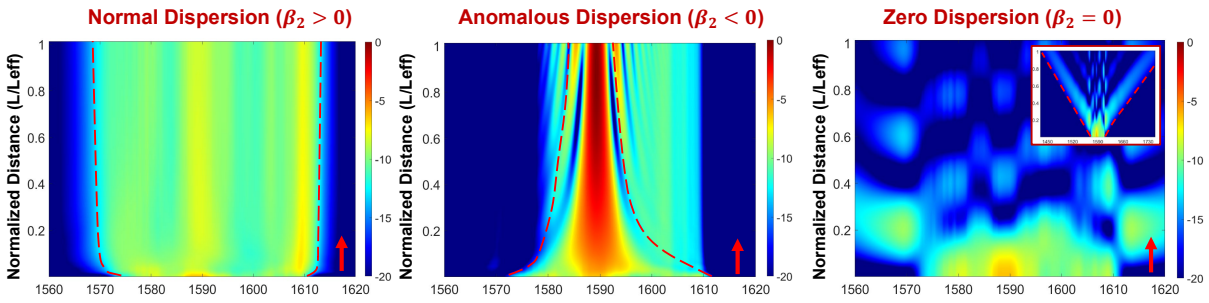

Figure 4.3: The evolution of the laser spectrum during nonlinear propagation in Nonlinear Schrödinger Kernel working in (a) Normal dispersion (b) Anomalous dispersion and (c) Zero dispersion regimes using time stretch cell image dataset as input. This nonlinear process is illustrated using 2-D heatmaps, with the horizontal axis as the wavelength, and the vertical axis as the pulse propagation distance (normalized to the effective length of the optical medium). The horizontal axis for all three cases is set to the same range to show the input spectrum is the same. For the zero dispersion case in (c), however, the output spectrum is far broader than the other two cases, thus an inset is placed to show the evolution of the full spectrum. The color indicates the optical intensity in the log scale with a color bar next to each figure. The red arrow point to the propagation direction. The dashed red line shows the trend of spectral evolution. Both spectral broadening and spectral narrowing can be observed.

of fiber dispersion, the laser is positively chirped during the initial stage of propagation. The chirping then interacts with fiber nonlinearity, compressing the optical spectrum. For zero dispersion, since the peak power of the temporal pulse is kept as is all the way through the evolution with negligible loss, it accumulates huge nonlinearity. As a result, significant spectral broadening can be observed in Fig. 4.3c.

Since the data is mapped onto the spectrum of the femtosecond laser pulse, spectral broadening and compression also affect the dimension of the data. The spectral broadening increase the data dimension, and the compression reduces it. To avoid putting an extra

burden on the digital classifier, we chose to work in dimension reduction, in other words, an anomalous dispersion regime.

We further evaluate the dimension effect of the kernel. Fig. 4.4 shows the analysis of dimensions' influence on the classification error. We control the output data dimension by varying the sampling resolution of the optical spectrum. We sample the output spectrum with a various number of sampling points, train a linear support vector machine (SVM) classifier with the resampled data and then calculate the classification error via 3-fold cross validation. The plot has two parts, the dimension reduction ($\frac{output}{input} < 1$, $ln\frac{output}{input} < 0$, colored in red), and the dimension increase ($\frac{output}{input} > 1$, $ln\frac{output}{input} > 0$, colored in yellow). We observe that the classification error drops below the baseline error even before the output reaches the same dimension as the input. From this, we can conclude that the Nonlinear Schrödinger Kernel can effectively reduce the data dimension while improving the classification accuracy.

4.3 Further Studies

The Nonlinear Schrödinger Kernel has shown to be an effective technique for ultrafast data processing in femtosecond instruments. However, there're a few missing pieces. First of all, the effectiveness of this technology, though experimentally demonstrated, is not mathematically proven. Also, the prototype system is bulky, making it difficult to be deployed in real instruments. Here we provide some potential studies that can be performed in the future.

4.3.1 Mathematical Explanation Behind the Optical Kernel

In previous chapters, to explain the working principle of the Nonlinear Schrödinger Kernel, we drew an analogy to a numerical kernel function, and name it after the term 'kernel'. However, this is rather metaphorical. In the next step, we can continue in this direction, and examine if this is an exact kernel.

In the numerical kernel trick, the kernel function obeys Mercer's theorem, also known as

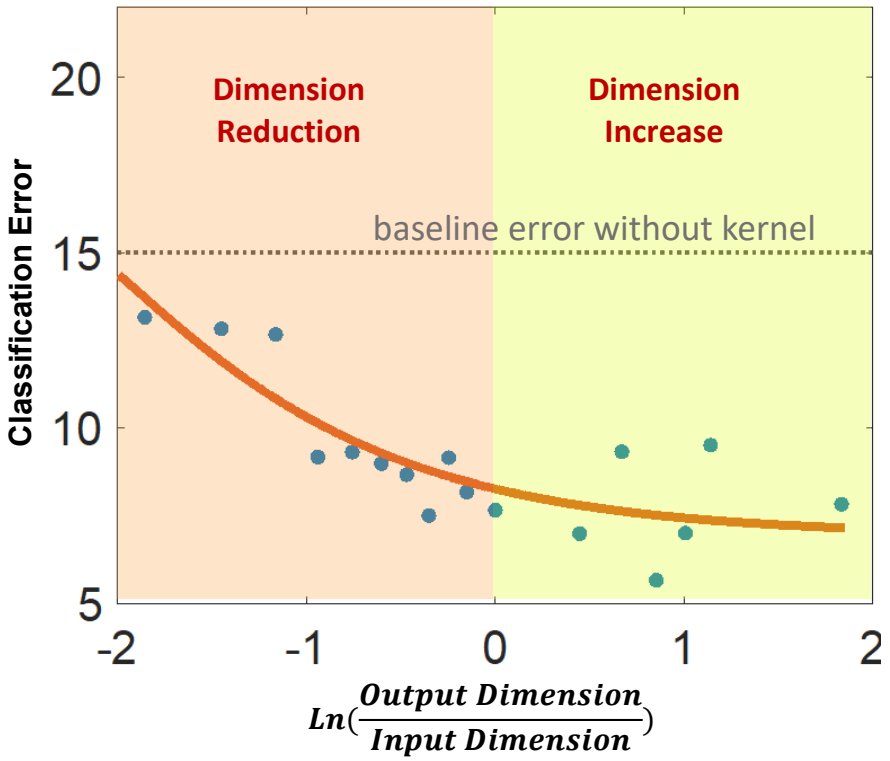

Figure 4.4: Classification error vs the data dimension. The horizontal axis is the natural log of the ratio of the output data dimension to the input data dimension of the optical kernel, and the vertical axis is the classification error. The blue dots are the classification error calculated using data with corresponding dimensions, and the red curve is the fit. The gray dashed line is the baseline classification error calculated by feeding the input data directly into the digital backend. The output data with different dimensions are obtained by sampling the original output spectrum with different dimensions. The resampled data and used to train a linear Support Vector Machine (SVM) classifier. The classification error is calculated via 3-fold cross validation. This simulation uses the time stretch cell image dataset.

Mercer's conditions. It states that an effective kernel function must be symmetric, continuous, and, when written in matrix form, takes the form of a positive semi-definite matrix [Bar][Wei]. We can build on the mathematical model reported in previous chapters and examine whether it satisfies Mercer's conditions.

This is one direction of explaining the working principle of the Nonlinear Schrödinger Kernel. Note that our results have shown the effectiveness of this technology regardless it constitutes a real mathematical kernel or not.

4.3.2 Deployment of the Optical Kernel

Despite the effectiveness of the reported system, it's a bulky prototype that is difficult to be implemented in ultrafast instruments. For deployment, certain parts need to be switched to integrated components. For example, the nonlinear optical media in the current system is a 500m highly nonlinear fiber, which can be replaced by a millimeter silicon nitride waveguide. The time stretch spectrometer uses a spool of dispersion compensation fiber, which can be replaced by a millimeter highly dispersive waveguide as well. Also, we can focus on the study of the new integrated optical sources to replace the mode-locked laser used in the current system.

CHAPTER 5

Conclusion

The fast-evolving neural networks have become powerful tools for various applications. However, the inference latency is orders of magnitude slower than the time scales of the femtosecond instruments [MCB17][ZCJ22]. To solve this problem, we demonstrated a new concept in AI hardware acceleration that exploits femtosecond laser pulses for computing. This technology, known as the nonlinear optical kernel, is analogous to a traditional numerical kernel, functioning by projecting the data into an implicit high-dimensional space and reading it out using a linear classifier in micro or even nanosecond timescale. It is robust against system nonidealities such as quantization and additive noise and requires no increase in data dimensions. The optimization of the kernel can be achieved by introducing spectral phase modulation of the femtosecond laser pulse within a digital feedback loop

This technology is a great example of using optical physics for accelerating computing in Artificial Intelligence (AI), which is part of a bigger emerging paradigm known as physics-AI symbiosis, where physics and AI accomplish each other. Another example is AI surrogate models, where AI is used to accelerate the solution of physics equations [JZK22]. This is the result of the evolution of science. In the early days, scientific advancement was purely empirical and data-driven. After the establishment of physics, this process became deterministic. In recent years, with AI being introduced into scientific discovery, the trend is shifting again. Even though the deterministic approach won't be completely replaced, this physics-AI symbiosis will exist for a long time.

REFERENCES

[Agr12] G.P. Agrawal. *Nonlinear Fiber Optics*. Optics and Photonics. Elsevier Science, 2012.

[Bar] Peter Bartlett. "Reproducing Kernel Hilbert Spaces.".

[BCJ98] Asuri S Bhushan, F Coppinger, and Bahram Jalali. "Time-stretched analogue-to-digital conversion." *Electronics Letters*, **34**(11):1081–1083, 1998.

[BFS84] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and regression trees*. CRC press, 1984.

[BJ04] Ozdal Boyraz and Bahram Jalali. "Demonstration of a silicon Raman laser." *Optics express*, **12**(21):5269–5273, 2004.

[BLB17] A. Bagnall, J. Lines, A. Bostrom, J. Large, and E. Keogh. "The Great Time Series Classification Bake Off: a Review and Experimental Evaluation of Recent Algorithmic Advances." *Data Mining and Knowledge Discovery*, **31**:606–660, 2017.

[CBJ05] Jason Chou, Ozdal Boyraz, and Bahram Jalali. "Adaptive optical post distortion linearization." *Optics Express*, **13**(15):5711–5718, 2005.

[CHH07] Deng Cai, Xiaofei He, and Jiawei Han. "Spectral regression for efficient regularized subspace learning." In *2007 IEEE 11th international conference on computer vision*, pp. 1–8. IEEE, 2007.

[CHH11] Deng Cai, Xiaofei He, and Jiawei Han. "Speed up kernel discriminant analysis." *The VLDB Journal*, **20**(1):21–33, 2011.

[CMT16] Claire Lifan Chen, Ata Mahjoubfar, Li-Chia Tai, Ian K Blaby, Allen Huang, Kayvan Reza Niazi, and Bahram Jalali. "Deep learning in label-free cell classification." *Scientific reports*, **6**(1):1–16, 2016.

[CV95] Corinna Cortes and Vladimir Vapnik. "Support-vector networks." *Machine learning*, **20**(3):273–297, 1995.

[Dav13] Luke M Davis. *Predictive modelling of bone ageing*. PhD thesis, University of East Anglia, 2013.

[DG17] Dheeru Dua and Casey Graff. "UCI Machine Learning Repository.", 2017.

[EEE19] Nasim Mohammadi Estakhri, Brian Edwards, and Nader Engheta. "Inverse-designed metastructures that solve equations." *Science*, **363**(6433):1333–1338, 2019.

[GAG12] Keisuke Goda, Ali Ayazi, Daniel R Gossett, Jagannath Sadasivam, Cejo K Lonappan, Elodie Sollier, Ali M Fard, Soojung Claire Hur, Jost Adam, Coleman Murray, et al. "High-throughput single-microparticle imaging flow analyzer." *Proceedings of the National Academy of Sciences*, **109**(29):11630–11635, 2012.

[GMA19] Jonathan K George, Armin Mehrabian, Rubab Amin, Jiawei Meng, Thomas Ferreira De Lima, Alexander N Tait, Bhavin J Shastri, Tarek El-Ghazawi, Paul R Prucnal, and Volker J Sorger. "Neuromorphic photonics with electro-absorption modulators." *Optics express*, **27**(4):5181–5191, 2019.

[HGI18] Pierre-Henry Hanzard, Thomas Godin, Saïd Idlahcen, Claude Rozé, and Ammar Hideur. "Real-time tracking of single shockwaves via amplified time-stretch imaging." *Applied Physics Letters*, **112**(16):161106, 2018.

[HK70] Arthur E Hoerl and Robert W Kennard. "Ridge regression: Biased estimation for nonorthogonal problems." *Technometrics*, **12**(1):55–67, 1970.

[HKJ17] Georg Herink, Felix Kurtz, Bahram Jalali, Daniel R Solli, and Claus Ropers. "Real-time spectral interferometry probes the internal dynamics of femtosecond soliton molecules." *Science*, **356**(6333):50–54, 2017.

[Ho95] Tin Kam Ho. "Random decision forests." In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pp. 278–282. IEEE, 1995.

[Hol75] John H. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI, 1975. second edition, 1992.

[Isa14] Walter Isaacson. *The innovators: How a group of inventors, hackers, geniuses and geeks created the digital revolution*. Simon and Schuster, 2014.

[JKJ20] Yunshan Jiang, Sebastian Karpf, and Bahram Jalali. "Time-stretch LiDAR as a spectrally scanned time-of-flight ranging camera." *Nature photonics*, **14**(1):14–18, 2020.

[JSF18] Zohar Jackson, César Souza, Jason Flaks, Yuxin Pan, Hereman Nicolas, and Adhish Thite. "Jakobovski/free-spoken-digit-dataset: v1.0.8.", August 2018.

[JZK22] Bahram Jalali, Yiming Zhou, Achuta Kadambi, and Vwani Roychowdhury. "Physics-AI symbiosis." *Machine Learning: Science and Technology*, **3**(4):041001, 2022.

[KCB99] PV Kelkar, F Coppinger, AS Bhushan, and B Jalali. "Time domain optical sensing." In *1999 IEEE LEOS Annual Meeting Conference Proceedings. LEOS'99. 12th Annual Meeting. IEEE Lasers and Electro-Optics Society 1999 Annual Meeting (Cat. No. 99CH37009)*, volume 1, pp. 381–382. IEEE, 1999.

[KSC22] Igor Kudelin, Srikanth Sugavanam, and Maria Chernysheva. "Ultrafast Gyroscopic Measurements in a Passive All-Fiber Mach–Zehnder Interferometer via Time-Stretch Technique." *Advanced Photonics Research*, **3**(8):2200092, 2022.

[LEP17] Cosimo Lacava, Mohamed A Ettabib, and Periklis Petropoulos. "Nonlinear silicon photonic signal processing devices for future optical networks." *Applied Sciences*, **7**(1):103, 2017.

[LMC19] Yueqin Li, Ata Mahjoubfar, Claire Lifan Chen, Kayvan Reza Niazi, Li Pei, and Bahram Jalali. "Deep cytometry: deep learning with real-time inference in cell sorting and flow cytometry." *Scientific reports*, **9**(1):1–12, 2019.

[LRY18] Xing Lin, Yair Rivenson, Nezih T Yardimci, Muhammed Veli, Yi Luo, Mona Jarrahi, and Aydogan Ozcan. "All-optical machine learning using diffractive deep neural networks." *Science*, **361**(6406):1004–1008, 2018.

[MCB17] Ata Mahjoubfar, Dmitry V Churkin, Stéphane Barland, Neil Broderick, Sergei K Turitsyn, and Bahram Jalali. "Time stretch and its applications." *Nature Photonics*, **11**(6):341–351, 2017.

[MLM20] JG Mance, BM La Lone, JA Madajian, WD Turley, and LR Veeser. "Time-stretch spectroscopy for fast infrared absorption spectra of acetylene and hydroxyl radicals during combustion." *Optics Express*, **28**(20):29004–29015, 2020.

[PD03] Foster Provost and Pedro Domingos. "Tree induction for probability-based ranking." *Machine learning*, **52**(3):199–215, 2003.

[REL15] Eleonore Roussel, Clément Evain, Marc Le Parquier, Christophe Szwaj, Serge Bielawski, Laurent Manceron, J-B Brubach, M-A Tordeux, J-P Ricaud, Lodovico Cassinari, et al. "Observing microscopic structures of a relativistic object using a time-stretch strategy." *Scientific reports*, **5**(1):1–8, 2015.

[SJ15] Daniel R Solli and Bahram Jalali. "Analog optical computing." *Nature Photonics*, **9**(11):704–706, 2015.

[SKV16] Francesco Saltarelli, Vikas Kumar, Daniele Viola, Francesco Crisafi, Fabrizio Preda, Giulio Cerullo, and Dario Polli. "Broadband stimulated Raman scattering spectroscopy by a photonic time stretcher." *Optics express*, **24**(19):21264–21275, 2016.

[SLH12] Fabien Scalzo, David Liebeskind, and Xiao Hu. "Reducing false intracranial pressure alarms using morphological waveform features." *IEEE Transactions on biomedical engineering*, **60**(1):235–239, 2012.

[SRK07] Daniel R Solli, Claus Ropers, Prakash Koonath, and Bahram Jalali. "Optical rogue waves." *Nature*, **450**(7172):1054–1057, 2007.

[TGL20] Neil C Thompson, Kristjan Greenewald, Keeheon Lee, and Gabriel F Manso. "The computational limits of deep learning." *arXiv preprint arXiv:2007.05558*, 2020.

[TON18] DTH Tan, KJA Ooi, and DKT Ng. "Nonlinear optics on silicon-rich nitride—a high nonlinear figure of merit CMOS platform." *Photonics Research*, **6**(5):B50–B66, 2018.

[TYH19] Gouhei Tanaka, Toshiyuki Yamane, Jean Benoit Héroux, Ryosho Nakane, Naoki Kanazawa, Seiji Takeda, Hidetoshi Numata, Daiju Nakano, and Akira Hirose. "Recent advances in physical reservoir computing: A review." *Neural Networks*, **115**:100–123, 2019.

[Vap63] Vladimir Vapnik. "Pattern recognition using generalized portrait method." *Automation and remote control*, **24**:774–780, 1963.

[Wei] Kilian Weinberger. "Chpter 13: Kernels.".

[WOG20] Gordon Wetzstein, Aydogan Ozcan, Sylvain Gigan, Shanhui Fan, Dirk Englund, Marin Soljačić, Cornelia Denz, David AB Miller, and Demetri Psaltis. "Inference in artificial intelligence with deep optics and photonics." *Nature*, **588**(7836):39–47, 2020.

[XDH18] Xiaowei Xu, Yukun Ding, Sharon Xiaobo Hu, Michael Niemier, Jason Cong, Yu Hu, and Yiyu Shi. "Scaling for edge inference of deep neural networks." *Nature Electronics*, **1**(4):216–222, 2018.

[ZCJ22] Yiming Zhou, Jacky CK Chan, and Bahram Jalali. "A unified framework for photonic time-stretch systems." *Laser & Photonics Reviews*, **16**(8):2100524, 2022.

[ZSJ22] Tingyi Zhou, Fabien Scalzo, and Bahram Jalali. "Nonlinear Schrödinger Kernel for Hardware Acceleration of Machine Learning." *Journal of Lightwave Technology*, **40**(5):1308–1319, 2022.