Tampere University

Lauri Salmela

PREDICTING EXTREME EVENTS IN FIBER-OPTIC INSTABILITIES USING MACHINE LEARNING

ABSTRACT

Lauri Salmela: Predicting extreme events in fiber-optic instabilities using machine learning
Master of Science Thesis
Tampere University
Science and Engineering
October 2019

The study of ultrafast instabilities in optics is one of the key research area in nonlinear science. Optical fibers constitute an excellent testbed for the study of complex dynamics and instabilities because of the controlled environment and large repetition rate of thousands or millions laser pulses per second. One of the best known and most studied example of these instabilities is modulation instability which describes the exponential amplification of a small periodic modulation on top of continuous background signal. Modulation instability is a nonlinear process governed by the nonlinear Schrödinger equation and it can be observed in various nonlinear systems such as in fluids dynamics, plasma physics and Bose-Einstein condensates. Modulation instability can also be seeded by noise. The noise-driven dynamics lead to the emergence of high intensity temporal structures with random statistics, and it has been suggested that modulation instability may lead to the emergence of extreme events or rogue waves with high temporal intensities. The initial spectral broadening associated with noise-seeded modulation instability can further seed supercontinuum generation dynamics by which an initially narrowband spectrum is transformed into a broad spectral continuum via various nonlinear effects. Because of the noise-seeded dynamics, the generated supercontinuum spectra show large shot-to-shot fluctuations between subsequential pulses. Under particular conditions the noise-driven dynamics may lead to the generation of a small number of extreme red-shifted rogue solitons associated with long-tailed "rogue wave" statistics.

The study of ultrafast noise-seeded dynamics must be conducted in real-time, i.e. on a single-shot basis, since average measurements will wash out any dynamical features of the system. In general, current real-time techniques often provide only partial information excluding information about the phase of the light field which precludes the possibility for directly relating the spectral and temporal properties of laser light. Methods for direct ultrafast time-domain measurements of the electric field do exists but the associated experimental setups are generally complex and remain limited in various ways. This thesis provides a new approach to the study of nonlinear instabilities and shows how machine learning can overcome these restrictions to yield time-domain information of noise-seeded dynamics based only on spectral intensity measurements.

Specifically, to correlate the spectral and temporal properties of modulation instability, numerical simulations of the generalized nonlinear Schrödinger equation are used for training a supervised neural network which is further applied to experimental modulation instability spectra to obtain the probability distribution for the maximum temporal intensity of the instability field. The maximum temporal intensity from the simulations is predicted with high accuracy, given that the spectral dynamic range is sufficiently high. For the experimental spectra, a real-time high dynamic range experimental setup based on spectral windowing and differential attenuation is introduced. The statistics of modulation instability are well reproduced by applying the simulation-trained neural network to the experimental spectra. Additionally, an unsupervised clustering analysis is used for classifying the spectra into classes with distinct temporal characteristics. Finally, the analysis is extended to the extreme case of supercontinuum generation and it is shown how a neural network model can be used for obtaining temporal information on rogue solitons. In particular, the temporal intensity and the Raman-induced temporal shift of the extreme red-shifted rogue solitons are predicted with excellent accuracy. The results of this thesis open up new perspectives for the study and prediction of time-domain characteristics of ultrafast noise-seeded dynamics.

Keywords: extreme events, modulation instability, rogue waves, machine learning

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

TIIVISTELMÄ

Lauri Salmela: Ääri-ilmiöiden ennustus kuituoptiikan epästabiiliuksissa koneoppimisella
Diplomityö
Tampereen yliopisto
Teknis-luonnontieteellinen
Lokakuu 2019

Ultranopeiden epästabiiliuksien tutkimus optiikassa on yksi keskeisimmistä epälineaarisen tieteen tutkimusalueista. Optiset kuidut tarjoavat erinomaisen mahdollisuuden tutkia monimutkaisia systeemejä sekä epästabiiliuksia laserien suuren toistonopeuden ja helposti kontrolloidun ympäristön johdosta. Yksi tunnetuin ja tutkituin esimerkki näistä epästabiiliuksista on modulaatioepästabiilius, joka kuvaa jatkuvan signaalin päällä olevan heikon jaksollisen moduloinnin eksponentiaalista vahvistusta. Modulaatioepästabiilius on epälineaarinen prosessi, jota voidaan kuvata epälineaarisella Schrödingerin yhtälöllä ja se voidaan havaita monissa epälineaarisissa systeemeissä, kuten hydrodynamiikassa, plasmafysiikassa ja Bose-Einstein kondensaateissa. Modulaatioepästabiilius voi myös syntyä kohinasta, mikä johtaa hetkellisiin korkean intensiteetin piikkeihin, jotka käyttäytyvät kaoottisesti. On myös ehdotettu, että modulaatioepästabiilius voi johtaa ääri-ilmiöiden tai roistoaaltojen syntymiseen, joiden hetkellinen voimakkuus on erityisen suuri. Modulaatioepästabiiliudesta johtuvaa spektrin levenemistä voidaan käyttää superjatkumon luomiseen, jossa alun perin kapeakaistainen laservalo muunnetaan laajaksi jatkuvaksi spektriksi epälineaaristen ilmiöiden kautta. Kohinalähtöisestä dynamiikasta johtuen generoidut superjatkumospektrit vaihtelevat paljon eri pulssien välillä. Erityistilanteissa kohina voi johtaa pieneen määrään äärimmäisen punasiirtyneitä roistosolitoneja.

Kohinasta syntyvien ilmiöiden tutkimus on suoritettava reaaliaikaisesti, toisin sanoen yksittäisten laserpulssien pohjalta, koska keskiarvoon perustuvat mittaukset eivät anna tietoa systeemin dynaamisista ominaisuuksista. Nykyiset reaaliaikaiset optiset tekniikat antavat kuitenkin usein vain osittaista tietoa jättämällä valokentän vaiheen huomioimatta, mikä estää mahdollisuuden linkittää suoraan laservalon spektrin ja ajalliset ominaisuudet. Menetelmiä suoriin sähkökentän aikatason ultranopeisiin mittauksiin on olemassa, mutta ne ovat kokeellisesti monimutkaisia sekä monin tavoin rajoitettuja. Tämä työ esittää uuden menetelmän epälineaaristen epästabiiliuksien tutkimiseen ja osoittaa, miten koneoppiminen voi voittaa kokeelliset rajoitukset ja antaa tietoa kohinalähtöisen dynamiikan aikatasosta ainoastaan taajuustason intensiteettimittauksien pohjalta.

Tarkemmin sanottuna modulaatioepästabiiliuden taajuus- ja aikaominaisuuksien korreloimiseksi käytetään ohjattuun oppimiseen perustuvaa neuroverkkoa, joka on koulutettu käyttäen numeerisia simulaatioita perustuen yleistettyyn epälineaariseen Schrödingerin yhtälöön. Neuroverkkoa käytetään ennustamaan todennäköisyysjakauma hetkellisen intensiteetin maksimille kokeellisista modulaatioepästabiiliusspektreistä. Korkein hetkellinen intensiteetti ennustetaan simulaatioista suurella tarkkuudella, kunhan spektrien dynaaminen alue pidetään riittävän korkeana. Kokeelliset spektrit mitataan reaaliaikaisella ja korkeaan dynaamiseen alueeseen kykenevällä laitteistolla, joka perustuu spektrien ikkunointiin ja differentiaaliseen vaimennukseen. Modulaatioepästabiiliuden tilastollinen käyttäytyminen toistetaan tarkasti soveltamalla simulaatiodatalla koulutettua neuroverkkoa kokeellisiin spektreihin. Lisäksi ohjaamattomaan oppimiseen perustuvaa ryhmittelyanalyysiä voidaan käyttää spektrien luokitteluun eri dynaamisiin luokkiin, jotka osoittavat aikatasossa tunnusomaisia piirteitä. Analyysi laajennetaan ääri-ilmiöihin superjatkumon generoinnissa, ja työssä näytetään, kuinka neuroverkko voi ennustaa roistosolitonien hetkellisen intensiteetin ja Ramanilmiön aiheuttaman ajallisen siirtymän erinomaisella tarkkuudella. Tämän työn tulokset avaavat uusia näkökulmia ja mahdollisuuksia ultranopean kohinasta syntyvän dynamiikan aikatason ominaisuuksien tutkimiseen ja ennustamiseen.

Avainsanat: ääri-ilmiöt, modulaatioepästabiilius, roistoaallot, koneoppiminen

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

PREFACE

This thesis done in the Photonics Laboratory of Tampere University under supervision of Prof. Goëry Genty. I am very grateful to him for the possibility to join the UKO group two years back and work on an interesting project. I truly appreciate the support I have been given, not to mention the freedom and possibility to travel to various conferences during this time.

Special thanks goes to rest of the UKO group: Piotr, Shanti, Caroline, Zahra, Roosa and all the former members that I have had the pleasure to know. All the chicken wings, movie nights, escape rooms, boat trips, ice skating and running together are things to remember. I would also like to thank Dr. Mikko Närhi for the guidance during the first months of my work and Prof. John M. Dudley for fruitful conversations and hosting me during my short visit in Besançon. Additionally, I would like to thank rest of the Optics corridor for a warm and friendly atmosphere.

Finally, a special thanks goes to my parents for all the support.

Tampere, 7 October 2019

Lauri Salmela

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

Symbols

$A(t)$	Electric field amplitude in time domain
$\tilde{A}(\omega)$	Electric field amplitude in frequency domain
A_{eff}	Effective mode area
$a.u.$	Arbitrary unit
$b_i^{(k)}$	Neuron bias
c	Speed of light in vacuum
$c.c.$	Complex conjugate
D	Dispersion parameter
$E/E(t)$	Complex electric field in time domain
$\tilde{E}/\tilde{E}(\omega)$	Complex electric field in frequency domain
$F(r)$	Modal amplitude distribution
f	Natural frequency
$f(x)$	Transfer function
f_R	Raman contribution
g_R	Raman gain
$\hbar$	Reduced Planck constant
$g_{12}^{(1)}$	Complex degree of first-order coherence
$h_i^{(k)}$	Neuron output
I	Intensity
$I_{1/3}$	Mean intensity of highest third
I_{RW}	Rogue wave threshold intensity
i	Imaginary unit
L_D	Dispersive length
L_{NL}	Nonlinear length
N	Soliton number
$n/n(\omega)$	Refractive index
n_2	Nonlinear refractive index
n_{eff}	Effective refractive index
$\mathbf{P}$	Material polarization
P	Optical power
P_0	Incident power
$R(T)$	Nonlinear response function

S	Spectrum
T	Co-moving time
T_0	Pulse width in time
T_{mod}	Modulation period
t	Time
v_g	Group velocity
v_p	Phase velocity
w	Mode width parameter
$w_{ij}^{(k)}$	Neuron weight
z	Propagation coordinate
α	Attenuation coefficient
β	Propagation constant
β_2	Group-velocity dispersion parameter
γ	Nonlinear coefficient
ϵ_0	Vacuum permittivity
ε	Error function
η	Learning rate
κ	Phase mismatch
λ	Wavelength
λ_0	Center wavelength
μ_0	Vacuum permittivity
ξ	Normalized propagation coordinate
ρ	Pearson correlation coefficient
τ	Normalized time coordinate
τ_{shock}	Optical shock characteristic time
φ	Phase shift
$\chi^{(j)}$	jth-order susceptibility
$\chi_{eff}^{(1)}$	Effective susceptibility
ψ	Normalized amplitude
Ω	Frequency shift
Ω_c	Limit frequency of modulation instability gain
Ω_{max}	Frequency of maximum modulation instability gain
ω	Angular frequency
ω_0	Center angular frequency
ω_a	Anti-Stokes (angular) frequency
ω_{DW}	Dispersive wave central frequency
ω_{mod}	Modulation frequency
ω_s	Stokes (angular) frequency

Acronyms

AB	Akhmediev breather
AI	Artificial intelligence
CW	Continuous wave
DFT	Dispersive Fourier transform
DL	Deep learning
FWHM	Full width at half maximum
FWM	Four wave mixing
GNLSE	Generalized nonlinear Schrödinger equation
GPU	Graphics processing unit
GVD	Group velocity dispersion
IST	Inverse scattering transform
KM	Kuznetsov-Ma (soliton)
MI	Modulation instability
ML	Machine learning
NLSE	Nonlinear Schrödinger equation
NN	Neural network
OSA	Optical spectrum analyser
PCF	Photonic crystal fiber
PDF	Probability density function
PS	Peregrine soliton
RS	Rogue soliton
RW	Rogue wave
SC	Supercontinuum
SFB	Soliton on finite background
SPM	Self-phase modulation
SRS	Stimulated Raman scattering
SSFS	Soliton self-frequency shift
THG	Third harmonic generation
XPM	Cross-phase modulation
ZDW	Zero-dispersion wavelength

1 INTRODUCTION

The increase in computational power has enabled the development of novel machine learning algorithms and their use in various scientific applications is a quickly emerging trend [1]. In this thesis, machine learning is applied to analyse and predict extreme events in ultrafast fiber instabilities.

In the linear regime of propagation, the phase and amplitude of the light field may be affected but its spectrum remains unchanged. However, as the intensity of the light field is increased, the propagation becomes nonlinear, leading to the creation of new spectral components through various nonlinear effects. Modulation instability (MI) is a nonlinear process whereby a weak periodic modulation on top a continuous wave (CW) signal is exponentially amplified [2,3]. In the spectral domain, this process can be described as a four wave mixing (FWM) process where energy is transferred to spectral sidebands symmetrically on both sides of the central frequency [4]. In addition to a periodic modulation, MI can be seeded by noise, leading to the formation of temporal breather structures that display chaotic behavior. Noise-seeded MI has been shown to lead to the formation of rare extreme events or rogue waves (RW) with high peak intensity compared to the average intensity of the temporal structures present in the chaotic MI field [5,6].

A highly incoherent noise-seeded MI field can act as a seed for supercontinuum (SC) generation, a broadband laser source generated via various nonlinear effects from narrowband laser light [7]. In this case, the temporal breather structures emerging from the MI stage lead to the formation of fundamental solitary waves or solitons that separate from the residual MI field [8] as a result of the Raman-induced soliton self-frequency shift (SSFS) that shifts the central wavelength of the solitons towards the longer wavelengths [9]. Correspondingly in the time-domain, the group-velocity dispersion separates the solitons from the initial MI field. The statistics of these red-shifted solitons in SC generation have been studied experimentally by using a long-pass spectral filter that captures only the most red-shifted spectral components [10]. Interestingly, large shot-to-shot fluctuations of the SC long-wavelength edge were observed highlighting the randomness of the soliton dynamics when noise triggers the SC generation process. A very small number of extremely red-shifted rogue solitons (RS) was observed, with a highly skewed "L-shaped" intensity distribution similar to that of rogue waves observed at the ocean surface.

A long-standing problem in ultrafast optics is to relate the spectral and temporal properties of light, which typically requires the knowledge if the complex field (intensity and phase) in either domain. While real-time spectral intensity measurements are straightforward to

conduct, measuring the phase on a single-shot basis is much more complex, such that directly relating the spectrum with the temporal properties is generally very difficult. Direct time-domain measurements of instabilities are however possible [11] but the measurements are typically complex and remain limited in various ways (power, bandwidth and resolution). The goal of this thesis is to overcome these limitations and show that machine learning can be used to relate the spectrum (no phase information) to the temporal characteristics of a chaotic field. This thesis concentrates on two highly chaotic cases: noise-seeded MI and rogue solitons in long pulse SC generation. For the MI case, the goal is to infer the maximum temporal intensity associated with a given single-shot spectrum. As it will be shown, concentrating on the maximum peak power only allows capturing the emerging extreme events or rogue waves with extremely high peak power from solely spectral intensity measurements. In the case of SC, the goal is to predict both the peak power and Raman-induced temporal shift of rogue solitons from the spectral intensity profile without any phase information.

Chapter 2 introduces the reader to propagation of light in a nonlinear medium and describes the most common nonlinear effects in silica fibers. A method for modeling the propagation is introduced as well. Chapter 3 describes the process of modulation instability in details along with the statistics of extreme events or rogue waves that may arise from the background noise. Chapter 4 gives an introduction to different dynamical regimes of supercontinuum generation. The statistics of extremely red-shifted rogue solitons are described in details. Chapter 5 introduces different machine learning techniques that are subsequently applied to the predictions of extreme events in MI and SC fiber-optic systems in Chapter 6. Finally, Chapter 7 gives a conclusion about the results obtained in this thesis along with thoughts on future work.

2 NONLINEAR PULSE PROPAGATION IN OPTICAL FIBERS

Similarly to all electromagnetic phenomena, the propagation of light in an optical fiber is governed by Maxwell's equations. To describe the electric field associated with a laser pulse, it is often convenient to use the complex form. In this formalism, the electric field E (or E-field) can be written in the complex form as

$$E(r, z, t) = F(r)A(z, t)\,\mathrm{e}^{-i\omega_0 t},\tag{2.1}$$

where $F(r)$ describes the modal amplitude distribution inside the fiber, $A(z, t)$ is the temporal envelope of the pulse and the exponential term represents the carrier wave with carrier frequency $\omega_0 = 2\pi c/\lambda_0$. The carrier wave is a rapidly oscillating field that modulates the (slowly-varying) envelope, such that one is generally only interested in the temporal envelope of the electric field. In single-mode fibers where the propagation of only one fundamental mode is supported, the modal distribution can be considered as constant along the propagation. The modal distribution of the fundamental mode can be approximated as a Gaussian distribution [2], and one can generally write the electric field as a function of time and propagation distance only. The propagation coordinate in this thesis is defined as z.

The temporal and spectral properties of a light field are related by the Fourier transform, such that in the spectral domain the electric field $\tilde{E}(z, \omega)$ can be obtained from the Fourier transform of the temporal field by [2]

$$\tilde{E}(z, \omega) = \mathcal{F}\{E(z, t)\} = \int_{-\infty}^{+\infty} E(z, t)\,\mathrm{e}^{i\omega t}\,\mathrm{d}t,\tag{2.2}$$

where $\mathcal{F}$ denotes the Fourier transform and ω the angular frequency. From an experimental point of view, one typically measures only the optical power P where the phase information of the field is lost. The optical power can be presented as the modulus squared of the electric field

$$P(z, t) = |E(z, t)|^2.\tag{2.3}$$

Similarly in the spectral domain, the spectral intensity or spectrum is defined as

$$S(z, \omega) = |\tilde{E}(z, \omega)|^2.\tag{2.4}$$

We stress that the fact that in general the phase information is lost is an important limitation since the spectral and temporal properties cannot be directly related anymore via simple Fourier transform and this thesis provides an alternative method for relating those two despite the lack of phase information.

2.1 Linear propagation regime

In linear regime, the propagation dynamics are independent from the intensity of the propagating light. A monochromatic wave propagates with a phase velocity of $v_p = c/n$, where c is the speed of light in vacuum and n is the refractive index of the propagation medium. In the case of a short laser pulse, however the field is not monochromatic and the spectrum consists of a continuous range of frequencies. Because the refractive index of the medium is dependent on the frequency $n(\omega)$ in general, the different spectral components travel with different speeds inside a medium. This phenomenon is known as *dispersion*.

Another linear effect is attenuation which leads to loss of energy. The *attenuation coefficient* α measures the total power losses per unit distance, given by

$$P(z) = P_0 \, \mathrm{e}^{-\alpha z}, \tag{2.5}$$

where P_0 is the input power, z is the propagation coordinate and α is the (linear) attenuation coefficient. It is common to use unit of dB/km where decibels (dB) are defined as

$$R \text{ (in dB)} = 10 \log_{10} R. \tag{2.6}$$

The fibers used in telecommunications, exhibit total loss of as small as 0.2 dB/km [12]. After propagation of 100 km, the total loss is 20 dB which corresponds to 1% power at the output compared to the initial power.

2.1.1 Dispersion

Dispersion is a general feature of wave propagation. It describes the frequency dependence of propagation speed for optical waves, water waves or acoustic waves. In optics, the frequency dependence of the refractive index of a medium is known as *chromatic dispersion*, or material dispersion. Fundamentally, the chromatic dispersion originates from the characteristic resonance frequencies of the medium [2]. The refractive index of a bulk medium can be expressed using the Sellmeier equation [13]

$$n(\omega)^2 = 1 + \sum_{j=1}^{m} \frac{B_j \omega_j^2}{\omega_j^2 - \omega^2}, \tag{2.7}$$
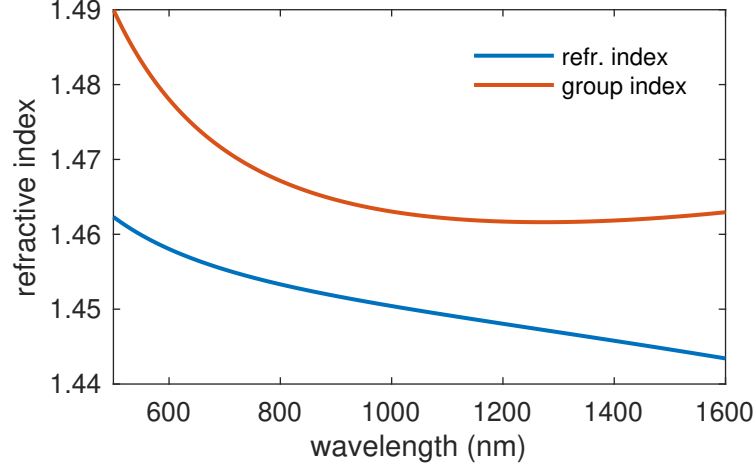
Figure 2.1. *Refractive index in blue and group index in orange for fused silica.*

where ω_j is the *j*th resonance frequency of the material with strength B_j. Figure 2.1 shows the refractive index for fused silica [14] commonly used as the material for optical fibers.

However, material dispersion is not the only factor affecting the refractive index "seen" by the light field. The fiber dimensions and refractive index difference between the core and cladding also affect the frequency-dependence of the effective refractive index. Indeed, short wavelengths are more confined into the core than longer wavelengths such that they effectively experience a higher refractive index. Dispersion plays a significant role in the pulse propagation dynamics.

An optical pulse propagates in a fiber at the *group velocity* given by $v_g = \partial\omega/\partial\beta$, where β is the propagation constant. Due to the frequency dependent nature of the refractive index, the propagation constant varies across the frequency band of the pulse, and is therefore given by $\beta(\omega) = n(\omega)\omega/c$. The propagation constant can be expanded around the central frequency ω_0 using a Taylor series expansion

$$\beta(\omega) = \beta(\omega_0) + \beta_1(\omega - \omega_0) + \frac{1}{2}\beta_2(\omega - \omega_0)^2 + \frac{1}{6}\beta_3(\omega)(\omega - \omega_0)^3 + \cdots, \qquad (2.8)$$

where $\beta_n = \partial^n\beta/\partial\omega^n|_{\omega_0}$.

The parameters β_1 and β_2 have a particular significance. The parameter β_1 corresponds to the inverse of the group velocity that describes the propagation velocity of an optical pulse or wave packet. It is given by

$$\beta_1 = \frac{1}{v_g} = \frac{n_g}{c} = \frac{1}{c}\left(n + \omega\frac{\partial n}{\partial\omega}\right) \qquad (2.9)$$

where n_g is the *group index*. Similarly to the refractive index for bulk medium, the group index determines the propagation velocity of different spectral components. The group index for fused silica is shown in Figure 2.1. The parameter β_2 represents the frequency dependence of β_1 and it is referred to as the group-velocity dispersion (GVD). For short laser pulses consisting of a large band of spectral components, dispersion induces
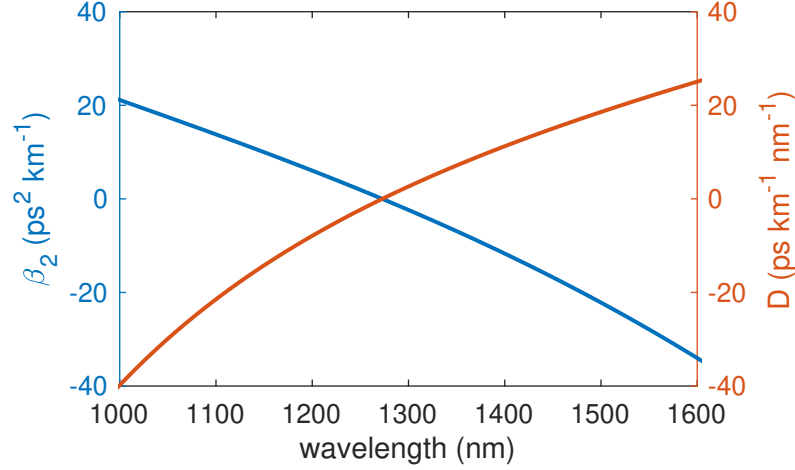
Figure 2.2. *Group velocity dispersion parameter β_2 in blue and dispersion parameter D in orange for bulk-fused silica.*

temporal broadening as the different wavelengths travel at a different speed [15].

The parameter β_2 determines the dispersion of the group velocity, and it is given by

$$\beta_2 = \frac{\partial \beta_1}{\partial \omega} = \frac{1}{c}\left(2\frac{\partial n}{\partial \omega} + \omega\frac{\partial^2 n}{\partial \omega^2}\right). \tag{2.10}$$

β_2 describes the rate of temporal broadening experienced by the optical pulse. The group velocity dispersion for fused silica is shown in Figure 2.2. In the same figure, the dispersion parameter D, defined as

$$D = \frac{\partial \beta_1}{\partial \lambda} = -\frac{2\pi c}{\lambda^2}\beta_2 \tag{2.11}$$

is also shown. D describes how much the different spectral components are separated in time during the propagation along an fiber and it is often used by fiber manufacturers to quantify the dispersion properties. Higher-order dispersion coefficients ($\beta_3, \beta_4, \cdots$) can also cause distortions to the optical pulse but in general the effect is less profound except in the case of pulses with vary large bandwidth where they must be included for better fit the Taylor series approximation [15]. Figure 2.2 plots the dispersion for bulk silica. For optical fibers, the dispersion slightly deviates from the bulk due to the doping of the fiber core and the confinement of light inside the fiber core.

From Figure 2.2 one can observe that, for a specific wavelength, the dispersion crosses the zero line. This is known as the *zero dispersion wavelength* (ZDW). Around this wavelength, the temporal broadening of the pulse is minimum. For bulk silica it is around 1270 nm. However, for optical fibers the ZDW is shifted towards longer values due to the doping and confinement of light. Typical values range from 1300 nm to 1580 nm [12]. Using a special fiber design known as *photonic crystal fiber* (PCF) [16] the ZDW can be lowered significantly. PCF consists of a small core of fused silica or air surrounded by a lattice of air holes. Ranka *et al.* [17] obtained ZDW of approximately 780 nm. A schematic of a photonic crystal fiber is shown in Figure 2.3.
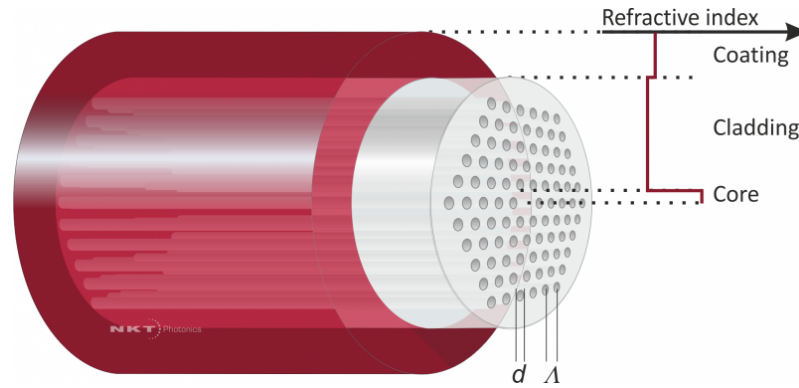
Figure 2.3. *Schematic of a photonic crystal fiber. Variable d describes the diameter of the holes in the crystal while Λ describes the period of the structure. Image by NKT Photonics [18].*

The value of the central wavelength respect to the ZDW of the fiber defines an important feature of the light propagation dynamics. Depending on the sign of β_2 the dynamics of the propagation vary significantly. In the case of positive β_2, the dispersion is *normal*. For negative β_2, on the other hand, the dispersion is *anomalous.* In the case of normal dispersion, the longer wavelength components propagate with higher speed, such that the leading edge of a propagating pulse consists of the low frequencies (or long wavelengths) while the frequency increases towards the trailing edge of the pulse. This is known as *up-chirp* or positive chirp. On the contrary, anomalous dispersion causes short wavelengths to travel faster, such that the frequency decreases towards the trailing edge of the pulse causing a *down-chirp* or negative chirp.

2.2 Nonlinear effects

In the case of extremely high intensity light, the propagation medium (atoms, molecules) does not response to the electric field in a linear manner anymore. In the linear regime (i.e. when the intensity of the light field is small), light can experience dispersion, attenuation, delay or deflection but no new frequencies are created during the propagation [15]. However, when the intensity of the light is increased, the material response becomes nonlinear in similar manner to a spring that is stretched too much and the restoring force required is not proportional to the displacement anymore.

For optical fields, the source of nonlinearity is the nonlinear polarization. It originates from the anharmonic oscillation of the bound electrons in the dielectric material. In the linear propagation regime, the material polarization is given by

$$\mathbf{P_L}(t) = \epsilon_0 \chi^{(1)} \mathbf{E}(t), \tag{2.12}$$

where $\mathbf{P_L}(t)$ is the linear polarization of the material, ϵ_0 is the vacuum permittivity and $\chi^{(1)}$ is the linear susceptibility [4]. Therefore, the material polarization scales with the electric field $\mathbf{E}(t)$. But in the nonlinear regime, the oscillation of the bound electrons does not

follow the sinusoidal oscillation of the E-field and higher-order susceptibility components lead to creation of new spectral components through various nonlinear processes.

2.2.1 Nonlinear polarization

In the presence of a high intensity light field, the polarization response of a dielectric material is not proportional to the electric field acting on it. The induced polarization consists of both linear and nonlinear components given by

$$\begin{aligned} \mathbf{P}(t) &= \mathbf{P}_L(t) + \mathbf{P}_{NL}(t) \\ &= \epsilon_0 \chi^{(1)} \mathbf{E}(t) + \epsilon_0 \chi^{(2)} \mathbf{E}^2(t) + \epsilon_0 \chi^{(3)} \mathbf{E}^3(t) + \cdots, \end{aligned} \tag{2.13}$$

where $\chi^{(k)}(k = 1, 2, ...)$ is the kth-order susceptibility. It must be noted that the value of the susceptibility is dependent on the wavelength ($\chi^{(k)} = \chi^{(k)}(\omega)$). Also, susceptibilities are tensors in general, as the E-field is a vector quantity [4]. The linear polarization accounts for linear propagation processes such as dispersion and attenuation described earlier. The linear susceptibility is related to the refractive index n by [4]

$$n^2(\omega) = 1 + \chi^{(1)}. \tag{2.14}$$

Note that attenuation can be explained from the complex value of the susceptibility.

The higher-order terms in Equation 2.13 are significant only in the presence of an intense electric field. This is because the higher-order susceptibilities are many orders of magnitude smaller compared to the lower orders, which must be compensated for by the electric field amplitude. The term $\chi^{(2)}$ vanishes in the case of centrosymmetric material, such as silica. And as noted, the susceptibilities decrease quickly in magnitude when the order is increased, and thus extremely high intensity of light would be required to observe those components such as in general $\chi^{(3)}$ is the most important nonlinear term when considering nonlinear propagation in silica fibers [2].

2.2.2 Nonlinear processes

The nonlinear polarization seeds the emergence of various nonlinear processes. The magnitude or strength of the nonlinear effects is determined by the *nonlinear coefficient* (in units of W^{-1}m^{-1}) given by [2]

$$\gamma = \frac{\omega_0 n_2}{c A_{eff}}, \tag{2.15}$$

where n_2 is the *nonlinear refractive index* discussed in the following section and A_{eff} is the *effective mode area* given by

$$A_{eff} = \frac{\left(\int\int_{-\infty}^{\infty} |F(x,y)|^2 \, \mathrm{d}x \, \mathrm{d}y\right)^2}{\int\int_{-\infty}^{\infty} |F(x,y)|^4 \, \mathrm{d}x \, \mathrm{d}y}, \tag{2.16}$$

where $F(x,y)$ is the modal distribution introduced earlier. The effective area describes the area of the mode confined inside the core. The effective area depends on the core diameter and the refractive index difference between the core and cladding. For a Gaussian intensity distribution, the area can be approximated as $A_{eff} = \pi w^2$, where w is the mode width parameter. Typical values for effective area range from 1 to $100\,\mu m^2$ for single-mode fibers [2]. The following sections describe the most typical effects in nonlinear fiber optics associated with the third-order nonlinear susceptibility $\chi^{(3)}$.

Optical Kerr effect

The optical Kerr effect arises from the intensity dependence of the refractive index. Consider an electric field $E = A\,\mathrm{e}^{-i\omega_0 t} + A^*\,\mathrm{e}^{i\omega_0 t} = A\,\mathrm{e}^{-i\omega_0 t} + c.c.$. The notation $c.c.$ refers to the complex conjugate. Neglecting the vectorial nature of the E-field, the polarization induced in a third-order nonlinear medium is given by [15]

$$\begin{aligned}
P_{total} &= P_L + P_{NL} \\
&= \epsilon_0 \left[\chi^{(1)} E(t) + \chi^{(3)} E^3(t)\right] \\
&= \epsilon_0 E(t) \left[\chi^{(1)} + \chi^{(3)} \left(A^2\,\mathrm{e}^{-2i\omega_0 t} + A^{*2}\,\mathrm{e}^{2i\omega_0 t} + 2|A|^2\right)\right] \\
&= \epsilon_0 \chi^{(1)} \left(A\,\mathrm{e}^{-i\omega_0 t} + c.c.\right) + \epsilon_0 \chi^{(3)} \left(A^3\,\mathrm{e}^{-3i\omega_0 t} + |A|^2 A\,\mathrm{e}^{-i\omega_0 t} + 2|A|^2 A\,\mathrm{e}^{-i\omega_0 t} + c.c.\right) \\
&= \epsilon_0 \chi^{(1)} \left(A\,\mathrm{e}^{-i\omega_0 t} + c.c.\right) + \epsilon_0 \chi^{(3)} \left(3|A|^2 A\,\mathrm{e}^{-i\omega_0 t} + c.c.\right) + \epsilon_0 \chi^{(3)} \left(A^3\,\mathrm{e}^{-3i\omega_0 t} + c.c.\right).
\end{aligned} \tag{2.17}$$

The first term on the last line of Equation 2.17 is the linear polarization P_L. The last two terms originate from the nonlinear polarization due to third-order nonlinear susceptibility. The last term oscillating with (angular) frequency of $3\omega_0$ is known as the *third harmonic generation* (THG). Three photons are absorbed from the fundamental field by the medium and one photon is generated with three times the fundamental frequency. However, the THG field is not sustained in general. To effectively sustain the THG field, the refractive indices of the fundamental and the THG fields must be equal (i.e. $n(\omega_0) = n(3\omega_0)$). This is known as the *phase matching condition* [4]. This is not the case in general since the refractive index is dependent on the wavelength.

One can notice that the second term on the last line of the Equation 2.17 oscillates at the fundamental frequency ω_0. Therefore, the total polarization oscillating at the fundamental

frequency is given by

$$
\begin{aligned}
P(\omega_0) &= P_L(\omega_0) + P_{NL}(\omega_0) \\
&= \epsilon_0 \chi^{(1)} A \, \mathrm{e}^{-i\omega_0 t} + 3\epsilon_0 \chi^{(3)} |A|^2 A \, \mathrm{e}^{-i\omega_0 t} \\
&= \epsilon_0 A \, \mathrm{e}^{-i\omega_0 t} \left(\chi^{(1)} + 3\chi^{(3)} |A|^2 \right) \\
&= \epsilon_0 A \, \mathrm{e}^{-i\omega_0 t} \, \chi_{eff}^{(1)}.
\end{aligned}
\tag{2.18}
$$

Here, we have defined the *effective linear susceptibility* $\chi_{eff}^{(1)} = \chi^{(1)} + 3\chi^{(3)} |A|^2$. Since the refractive index is related to the linear susceptibility through Equation 2.14, we can define a similar relation for the effective susceptibility. Assuming the effective susceptibility induced a *small* change Δn in the refractive index, we can derive the effective refractive index

$$
\begin{aligned}
n_{eff}^2 &= (n + \Delta n)^2 \\
&= n^2 + 2n\Delta n + \Delta n^2 \\
&\approx n^2 + 2n\Delta n,
\end{aligned}
\tag{2.19}
$$

where we have neglected Δn^2. By noting $n_{eff}^2 = 1 + \chi_{eff}^{(1)}$, we get

$$
\begin{aligned}
n_{eff}^2 &= 1 + \chi_{eff}^{(1)} \\
&= 1 + \chi^{(1)} + 3\chi^{(3)} |A|^2 \\
&= n^2 + 3\chi^{(3)} |A|^2,
\end{aligned}
\tag{2.20}
$$

where we have utilized Equation 2.14. By comparing Equations 2.19 and 2.20, we can extract the refractive index change caused by the E-field:

$$
\Delta n = \frac{3\chi^{(3)} |A|^2}{2n} = \frac{3\chi^{(3)} I}{4\epsilon_0 n^2 c} = n_2 I,
\tag{2.21}
$$

where $n_2 = \frac{3\chi^{(3)}}{4\epsilon_0 n^2 c}$ and we have used the relation $I = 2n\epsilon_0 c |A|^2$ for intensity. Using the first two terms from Taylor series expansion of Equation 2.19, we get the effective refractive index

$$
n_{eff} = \sqrt{n^2 + 2nn_2 I} = n\sqrt{1 + 2n_2 I/n} \approx n\left(1 + n_2 I/n\right) = n + n_2 I.
\tag{2.22}
$$

From Equation 2.22, we see that the refractive index of a third-order medium is modified by the intensity of an incoming (intense) light field. The intensity dependent refractive index is responsible for most of the nonlinear effects introduced in this Chapter. The physical process of light altering the refractive index of a medium is known as the optical Kerr effect.

Self-phase modulation

Due to the Kerr effect, an optical field can modulate its own phase. This process is know as *self-phase modulation* (SPM). A high intensity optical pulse induces a change to the refractive index which in turn induces a nonlinear phase-shift φ_{NL} [4]. Consider a (high intensity) monochromatic propagating wave with associated E-field $E(z,t) = A\,e^{i(\beta z - \omega_0 t)} + c.c.$. The wave "sees" the effective refractive index of the medium, and therefore, the propagation constant can be expressed as $\beta = n_{eff}\frac{2\pi}{\lambda}$. Inserting the effective index into the E-field, we can clearly notice the induced nonlinear phase-shift:

$$
\begin{aligned}
E(z,t) &= A\,e^{i(\beta z - \omega_0 t)} + c.c. \\
&= A\exp\left[i\left(n_{eff}\frac{2\pi}{\lambda}z - \omega_0 t\right)\right] + c.c. \\
&= A\exp\left[i\left(n\frac{2\pi}{\lambda}z + n_2 I\frac{2\pi}{\lambda}z - \omega_0 t\right)\right] + c.c. \\
&= A\exp\left[i\left(\varphi_L + \varphi_{NL} - \omega_0 t\right)\right] + c.c.,
\end{aligned}
\tag{2.23}
$$

where $\varphi_L = n\frac{2\pi}{\lambda}z = n\frac{\omega_0}{c}z$ is the linear phase-shift and $\varphi_{NL} = n_2 I\frac{2\pi}{\lambda}z = n_2\frac{|A|^2}{A_{eff}}\frac{\omega_0}{c}z$ is the nonlinear phase-shift. As we can see, the phase of the pulse is dependent on its own intensity.

Since the intensity of a pulse is time-dependent (for example Gaussian or hyperbolic secant), the induced phase-shift also depends on time. The time-varying phase causes changes to the pulse spectrum, typically making it broader [4]. The temporal intensity however remains unchanged.

To further describe the temporal properties of different spectral components, it is convenient to introduce the concept of *instantaneous frequency* $\omega(t)$ [4]. The frequency component arriving at time t is given by

$$
\omega(t) = \omega_0 + \delta\omega(t) = \omega_0 - \frac{\partial\varphi_{NL}}{\partial t}
\tag{2.24}
$$

where $\delta\omega(t)$ describes the variation to the instantaneous frequency. The frequency chirp caused by SPM is positive and quasilinear for the central (temporal) region of the pulse [2]. Self-phase modulation is especially important for ultra-short pulses [4]. If the maximum value of the induced frequency shift exceeds the spectral width of the pulse, the pulse is significantly broadened in the spectral domain.

Similarly to SPM, the phase of an optical field may be modulated by a strong external field. This is known as *cross-phase modulation* (XPM) highlighting the fact that it is another field modulating the phase of a given field [2]. XPM adds an additional term to the nonlinear polarization given in Equation 2.17, and thus modifies the nonlinear phase-shift in Equation 2.23.

Four wave mixing

Four wave mixing (FWM) is a parametric process that results from the mixing of four different optical fields with distinct frequencies [4]. Four photons are destroyed or annihilated during this process. Similarly to the case of self-phase modulation, the origin of FWM is the nonlinear polarization induced by the third-order nonlinear susceptibility and which includes a term that oscillates with a frequency given by the summation (or subtraction) of the incident fields.

To efficiently transfer energy at a specific frequency, two conditions must be fulfilled [2]. Similarly to self-phase modulation, the total energy of the light fields with frequencies ω_1, ω_2, ω_3 and ω_4 is conserved in the FWM process such that $\omega_1 + \omega_2 = \omega_3 + \omega_4$. In this case, photons at frequency ω_1 and ω_2 would be destroyed and photons at frequency ω_3 and ω_4 would be created, or vice versa. Limitation to this process comes from the phase-matching. To efficiently amplify the creation of photons at ω_3 and ω_4, we must have $\Delta k = \beta(\omega_1) + \beta(\omega_2) - \beta(\omega_3) - \beta(\omega_4) = 0$, where $\beta(\omega_i) = n(\omega_i)\omega_i/c$. The phase-matching insures the conservation of momentum.

Modulation instability

Modulation instability (MI) describes a nonlinear process where a weak periodic perturbation on top of a continuous signal is exponentially amplified. It is present in many nonlinear systems that show dispersive and nonlinear effects such as fiber optics, water waves and plasma physics. In fiber optics, MI leads to the break-up of a continuous time-domain signal, resulting into a periodic train of short high-intensity sub-pulses [2].

Modulation instability is one of the main topics of this thesis. A more comprehensive analysis of MI is given in Chapter 3.

Solitons

In the case of anomalous dispersion, the nonlinear and dispersive effects may cancel out each other [2]. The specific solutions are referred as solitary waves or *solitons*. They propagate undistorted in a fiber, or follow specific periodic structures in time and remain undistorted due to the down-chirp induced by the anomalous dispersion and up-chirp induced by SPM and XPM. This is the case for, so called, *fundamental* solitons. There are also higher-order solitons that are very sensitive to any perturbations and may thus break up into fundamental solitons. Solitary waves were first observed for water waves in 1884 by Scott Russel [19]. Solitons will be discussed in more details in Chapters 3 and 4.
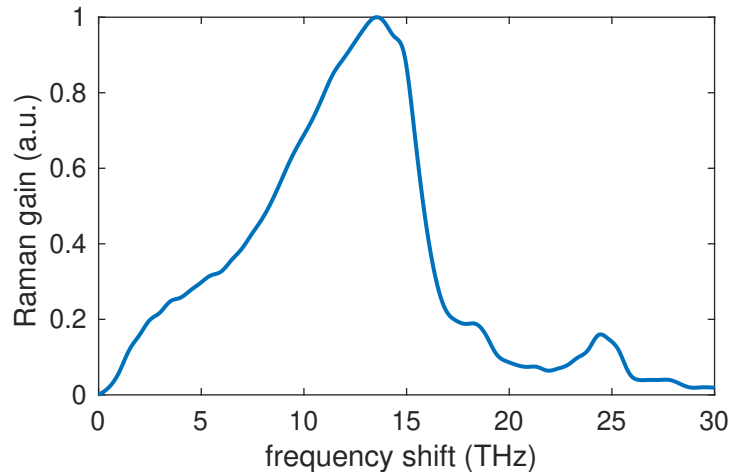
Figure 2.4. *Normalized Raman gain for silica. After Ref. [21].*

Stimulated Raman scattering

Raman scattering results from the interaction between the vibrational states of a molecule and light, effectively leading to energy transfer between two optical fields by which the frequency of the incident light is typically down-shifted towards longer wavelengths. The transfer is caused by inelastic scattering of photons from the molecules, and it is known as the *Raman effect* [20]. However, the spontaneous Raman scattering is a rather weak process since it is triggered by noise and the emission is in the form of dipole radiation [4]. Therefore, the Raman effect is often observed in the form of stimulated Raman scattering (SRS), such that the scattering process is stimulated by a weak co-propagating light field with the same frequency as that of the scattered photon. The stimulated process is typically much stronger than the spontaneous one and the emission is in a narrow cone shape [4].

Higher-frequency photons (with energy $\hbar\omega_p$) interact with molecules exciting them to a higher vibrational state. As a result, the photon is converted to a lower-frequency photon with energy $\hbar\omega_s$. The optical field at frequency ω_s is known as the *Stokes field*. In the process, an excited vibrational state of a molecule can also be released. In this case, a photon with higher frequency ω_a compared to the fundamental field is emitted. This field is known as the *anti-Stokes field* but it is typically much weaker than the Stokes field as the probability of thermal excitation to higher energy state is low [4].

The strength of the Raman effect is given by the Raman gain. The Raman gain (g_R) for silica is shown in Figure 2.4. The maximum gain is achieved for a frequency detuning about 13.2 THz from the pump. However, the Raman gain spans up to approximately 30 THz, such that in principle fields with detuning within the Raman gain bandwidth can be amplified. Given that pump intensity is sufficiently large, the resulting Stokes wave is exponentially amplified [2].

2.3 Propagation equation

The starting point for a general propagation equation is the Maxwell's equations. From Maxwell's equation, one can obtain the wave equation [2]

$$\nabla \times \nabla \times \mathbf{E} = -\frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} - \mu_0 \frac{\partial^2 \mathbf{P}}{\partial t^2}$$

$$= -\frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} - \mu_0 \frac{\partial^2 \mathbf{P}_L}{\partial t^2} - \mu_0 \frac{\partial^2 \mathbf{P}_{NL}}{\partial t^2} \tag{2.25}$$

where μ_0 is the vacuum permittivity and $\mathbf{P}$ is the total induced polarization introduced in Chapter 2.2.1 that can be separated into linear and nonlinear terms according to Equation 2.13. Using the vector calculus identity $\nabla \times \nabla \times \mathbf{E} = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2\mathbf{E}$ and relation $\nabla \cdot \mathbf{D} = \nabla \cdot \mathbf{E} = 0$ (due to the spatial independence of refractive index [2]) one obtains

$$\nabla^2\mathbf{E} - \frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} = \mu_0 \frac{\partial^2 \mathbf{P}_L}{\partial t^2} + \mu_0 \frac{\partial^2 \mathbf{P}_{NL}}{\partial t^2}. \tag{2.26}$$

By noticing that the linear polarization is related to the E-field through Equation 2.13, one can write

$$\nabla^2\mathbf{E} - \frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} = \mu_0 \frac{\partial^2 \epsilon_0 \chi^{(1)}\mathbf{E}}{\partial t^2} + \mu_0 \frac{\partial^2 \mathbf{P}_{NL}}{\partial t^2}. \tag{2.27}$$

To make the derivation easier, it is convenient to continue in the spectral domain. Since the electric field is expressed in the form $E(z,t) = A(z,t)\,\mathrm{e}^{-i\omega t}$ (the modal distribution neglected here), the Fourier transform of the second time derivative yields an additional coefficient of $(-i\omega)^2 = -\omega^2$. Therefore, Equation 2.27 can be written in the form

$$\nabla^2\tilde{\mathbf{E}} + \frac{\omega^2}{c^2}\tilde{\mathbf{E}} = -\mu_0 \epsilon_0 \omega^2 \tilde{\chi}^{(1)}(\omega)\tilde{\mathbf{E}} - \mu_0 \omega^2 \tilde{\mathbf{P}}_{NL}$$

$$\nabla^2\tilde{\mathbf{E}} = -\frac{\omega^2}{c^2}\tilde{\mathbf{E}} - \frac{\omega^2}{c^2}\tilde{\chi}^{(1)}(\omega)\tilde{\mathbf{E}} - \frac{\omega^2}{\epsilon_0 c^2}\tilde{\mathbf{P}}_{NL}$$

$$\nabla^2\tilde{\mathbf{E}} = -\frac{\omega^2}{c^2}\left(1 + \tilde{\chi}^{(1)}(\omega)\right)\tilde{\mathbf{E}} - \frac{\omega^2}{\epsilon_0 c^2}\tilde{\mathbf{P}}_{NL}, \tag{2.28}$$

where the relation $\epsilon_0\mu_0 = 1/c^2$ is used. The notation for electric field in the spectral domain is relative to the central frequency, i.e. $\tilde{\mathbf{E}} = \tilde{\mathbf{E}}(\omega - \omega_0)$. Notation $\tilde{\mathbf{P}}_{NL}$ refers to the Fourier transform of the nonlinear polarization. Now we can use Equation 2.14, leading to

$$\nabla^2\tilde{\mathbf{E}} + n^2(\omega)\frac{\omega^2}{c^2}\tilde{\mathbf{E}} = -\frac{\omega^2}{\epsilon_0 c^2}\tilde{\mathbf{P}}_{NL}$$

$$\nabla^2\tilde{\mathbf{E}} + \beta^2(\omega)\tilde{\mathbf{E}} = -\frac{\omega^2}{\epsilon_0 c^2}\tilde{\mathbf{P}}_{NL}. \tag{2.29}$$

Equation 2.29 is still very general in nature. In the case of no nonlinearities, it is written in the form

$$\nabla^2\tilde{\mathbf{E}} + \beta^2(\omega)\tilde{\mathbf{E}} = 0, \tag{2.30}$$

which is known as the Helmholtz equation.

The propagation equation can be used for modeling the propagation of optical field in a fiber. It includes all the linear and nonlinear processes that take place during the propagation. Some of the processes can be solved analytically but to include all the nonlinear processes numerical methods are needed due to high complexity. Those will be explained in Section 2.3.2.

2.3.1 Generalized nonlinear Schrödinger equation

Equation 2.29 can be further developed provided that several assumptions are made. First, let us assume that the transverse modal distribution of the field remains constant along propagation. It can be assumed to be a Gaussian function for the fundamental mode as mentioned before [2]. We can then write the E-field as a scalar variable ($\tilde{E} = F(x,y)\tilde{A}(z, \omega - \omega_0)\,\mathrm{e}^{i\beta_0 z} + c.c.$) and solve the equation in 1+1D (propagation + time). Due to the tensor nature of $\chi^{(3)}$, it is possible to have up to 81 nonzero elements describing the polarization response of different E-field components [4]. Yet, typically only few of the components are independent, and thus is it reasonable to make a scalar approximation with only one tensorial element with a maintained single polarization state.

Neglecting the transverse modal distribution, the Laplace operator ∇^2 can be replaced by a second derivative along the propagation coordinate $\partial^2/\partial z^2$, and Equation 2.29 can be written as

$$\frac{\partial^2 \tilde{E}}{\partial z^2} + \beta^2(\omega)\tilde{E} = -\frac{\omega^2}{\epsilon_0 c^2}\tilde{P}_{NL}$$

$$\left(\frac{\partial^2 \tilde{A}}{\partial z^2} + 2i\beta_0\frac{\partial \tilde{A}}{\partial z} - \beta_0^2\tilde{A}\right)\mathrm{e}^{i\beta_0 z} + \beta^2(\omega)\tilde{A}\,\mathrm{e}^{i\beta_0 z} + c.c. = -\frac{\omega^2}{\epsilon_0 c^2}\tilde{P}_{NL}. \tag{2.31}$$

To further simplify the expression above, a slowly-varying envelope approximation can be made [2]. The approximation states that the second derivative is negligible compared to the second term in Equation 2.31. Physically, it means that the relative change of E-field envelope amplitude over a propagation distance comparable to the wavelength is small:

$$\frac{\partial^2 \tilde{A}}{\partial z^2} \ll \beta_0\frac{\partial \tilde{A}}{\partial z} \implies \frac{\partial \tilde{A}}{\tilde{A}} \ll \frac{2\pi}{\lambda}\partial z. \tag{2.32}$$

Equation 2.31 can now be written in the more convenient form:

$$\left(2i\beta_0\frac{\partial \tilde{A}}{\partial z} + (\beta^2(\omega) - \beta_0^2)\tilde{A}\right)\mathrm{e}^{i\beta_0 z} + c.c. = -\frac{\omega^2}{\epsilon_0 c^2}\tilde{P}_{NL}. \tag{2.33}$$

Assuming that the refractive index of the fiber varies slowly vs. wavelength, the term $\beta^2(\omega) - \beta_0^2$ can be further approximated as $2\beta_0(\beta(\omega) - \beta_0)$ [2]. By writing the propagation constant as a Taylor series expansion as done in Equation 2.8 and including the three first

terms, one obtains

$$\left(2i\beta_0\frac{\partial\tilde{A}}{\partial z} + 2\beta_0\beta_1(\omega-\omega_0)\tilde{A} + 2\beta_0\frac{\beta_2}{2}(\omega-\omega_0)^2\tilde{A}\right)e^{i\beta_0 z} + c.c. = -\frac{\omega^2}{\epsilon_0 c^2}\tilde{P}_{NL}. \quad (2.34)$$

Switching back to the time-domain yields another convenient form for the propagation equation:

$$\left[2i\beta_0\frac{\partial A}{\partial z}e^{-i\omega_0 t} + 2\beta_0\beta_1\left(-\frac{1}{i}\frac{\partial A}{\partial t}e^{-i\omega_0 t}\right) + 2\beta_0\frac{\beta_2}{2}\left(-\frac{\partial^2 A}{\partial t^2}e^{-i\omega_0 t}\right)\right]e^{i\beta_0 z} + c.c.$$
$$= \frac{1}{\epsilon_0 c^2}P_{NL}, \quad (2.35)$$

which simplifies to

$$\left(\frac{\partial A}{\partial z} + \beta_1\frac{\partial A}{\partial t} + i\frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2}\right)e^{i(\beta_0 z - \omega_0 t)} + c.c. = -\frac{i}{2\beta_0\epsilon_0 c^2}P_{NL}. \quad (2.36)$$

The only unknown term left in the equation above is the nonlinear polarization. This term depends on the nonlinear medium where light propagates. The nonlinear polarization for a third-order nonlinear medium (such as silica) has been derived in Equation 2.17, and the second time-derivative can be written as

$$\frac{\partial^2 P_{NL}}{\partial t^2} = \frac{\partial^2}{\partial t^2}(\overbrace{\epsilon_0\chi^{(3)}3|A|^2 A}^{P(t)}e^{i(\beta_0 z - \omega_0 t)} + c.c.)$$
$$= \left(\frac{\partial^2 P(t)}{\partial t^2} - 2i\omega_0\frac{\partial P(t)}{\partial t} - \omega_0^2 P(t)\right)e^{i(\beta_0 z - \omega_0 t)} + c.c.$$
$$\approx -\omega_0^2 P(t)e^{i(\beta_0 z - \omega_0 t)} + c.c., \quad (2.37)$$

where the time derivatives of the polarization have been neglected (slowly-varying envelope approximation). Equation 2.36 can now be separated into two separate equations that are complex conjugates of each other. The exponential term can be also simplified, resulting in

$$\frac{\partial A}{\partial z} + \beta_1\frac{\partial A}{\partial t} + i\frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2} = i\frac{\omega_0^2}{2\beta_0\epsilon_0 c^2}P(t)$$
$$= i\frac{\omega_0^2}{2\beta_0\epsilon_0 c^2}\epsilon_0\chi^{(3)}3|A|^2 A$$
$$= i\frac{3\chi^{(3)}\omega_0}{2n\omega_0 c}|A|^2 A$$
$$\Rightarrow \frac{\partial A}{\partial z} + \beta_1\frac{\partial A}{\partial t} + i\frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2} = i\gamma|A|^2 A, \quad (2.38)$$

where γ is the nonlinear coefficient introduced earlier in Equation 2.15.

Equation 2.38 is known as the *nonlinear Schrödinger equation* (NLSE). To further simplify

the notation, is it convenient to introduce the retarded or co-moving frame of reference. This means that the time coordinate is chosen in a way that it moves along the propagation coordinate at the group velocity of the pulse envelope. The new co-moving time coordinate is given by $T = t - \beta_1 z$. The nonlinear Schrödinger equation in the co-moving frame of reference is given by [2]

$$\frac{\partial A}{\partial z} + i\frac{\beta_2}{2}\frac{\partial^2 A}{\partial T^2} = i\gamma|A|^2 A. \tag{2.39}$$

The unit of the electric field amplitude is normally V/m but however it is convenient to normalize the amplitude in a way that $|A|^2$ represents the optical power. Therefore, the unit of A is typically given as $W^{1/2}$. The nonlinear coefficient has unit of $W^{-1}m^{-1}$.

Equation 2.39 is typically sufficient to model the propagation of an optical pulse. It includes the nonlinearities arising from self-phase modulation, cross-phase modulation and four wave mixing. However, in the general case it may be necessary to include higher-order dispersion as well as stimulated Raman scattering. The Raman gain amplifies the low-frequency components of a pulse at the cost of higher frequencies shifting the pulse towards low frequencies [2]. This phenomenon is known as *intrapulse Raman scattering* and it can play important role in the propagation of solitons, for example, causing the soliton self-frequency shift [9].

Including the higher-order nonlinear terms, the *generalized nonlinear Schrödinger equation* (GNLSE) governing the evolution of the electric field amplitude $A(z,T)$ (in units of $W^{1/2}$ and co-moving frame of reference) can be written as [2,7,22,23]

$$\frac{\partial A}{\partial z} + \frac{\alpha}{2}A - \sum_{k \geq 2}\frac{i^{k+1}}{k!}\beta_k\frac{\partial^k A}{\partial T^k} = i\gamma\left(1 + i\tau_{shock}\frac{\partial}{\partial T}\right) \times$$
$$\left(A\int_{-\infty}^{+\infty} R(T')|A(z, T - T')|^2 \,\mathrm{d}T'\right). \tag{2.40}$$

Left side of the GNLSE accounts for the linear effects, including attenuation (α) and dispersion. Attenuation arises from the imaginary part of the linear susceptibility by $\alpha(\omega) = \frac{\omega}{nc}\mathrm{Im}\{\chi^{(1)}\}$ [2]. Higher-order dispersion terms are also included. On the right side, the various nonlinear effects are included. The nonlinear coefficient γ describes the strength of the nonlinear effects in the medium. The term τ_{shock} describes the frequency dependence of the nonlinear coefficient. It results from the frequency dependence of both the nonlinear refractive index n_2 and the effective mode area A_{eff}. In practice however, a reasonable and often used approximation is $\tau_{shock} = 1/\omega_0$ [2]. The time derivative following τ_{shock} models the dispersion of the nonlinearities, which is associated with the phenomena of *self-steepening* and *shock formation* [7].

The nonlinear response function $R(T)$ includes the both instantaneous electronic contribution (Kerr effect) and delayed Raman response h_R (Raman scattering) [7].

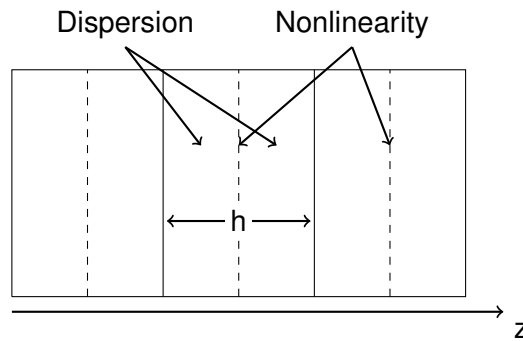$$R(T) = 1 - f_R + f_R h_R \tag{2.41}$$

Figure 2.5. *Schematic of split-step Fourier method with step size h. After Ref. [2].*

The Raman contribution f_R is measured to be 0.18 for silica [24].

2.3.2 Numerical modeling

The GNLSE (Eq 2.40) is a nonlinear differential equation that is not generally solvable in analytic form except for some special cases. Therefore, the equation needs to be solved numerically. An often used method for this purpose is the *split-step Fourier method* (SSFM) [2]. A schematic of the this method is shown in Figure 2.5. A more complete description of the method can be read from References [2,7], for example. In this work, the attenuation of the pulse is not included but it can be added in a straightforward manner. Other methods such as finite-element methods exist but the SSFM is the most simple and computationally fast for pulse propagation modeling.

The split-step Fourier scheme utilizes the fact the temporal and spectral domain are connected to each other by Fourier transforms. The basic idea is the following: over a small distance h, the linear terms on the left hand side of Equation 2.40 ($\hat{D}$) and nonlinear terms on the right hand side ($\hat{N}$) can be assumed to be independent of each other and estimated separately. The GNLSE can then be simply written as $\partial A/\partial z = (\hat{D} + \hat{N})A$. The fiber is thus divided into elementary steps of size h and the two steps of linearity and nonlinearity are sequentially solved along the propagation.

Solving the linear part is trivial in the frequency domain [2]. The spectral field after a small step h can be calculated from exponential integration

$$A(z_0 + h, \omega) = e^{h\hat{D}} A(z_0, \omega). \tag{2.42}$$

The nonlinear term is more challenging. For that, a second-order Runge-Kutta algorithm is applied but other algorithms are also possible. In the second-order Runge-Kutta scheme, the derivative of the spectral field amplitude is evaluated at initial point $z = z_0$. The Runge-Kutta algorithm is mostly carried out in the frequency domain as the convolution term can be simply calculated as a multiplication. After multiplication with the $A(z, \omega)$, shock-term and the nonlinear coefficient, the derivative $\partial A/\partial z$ at z_0 is obtained, from which
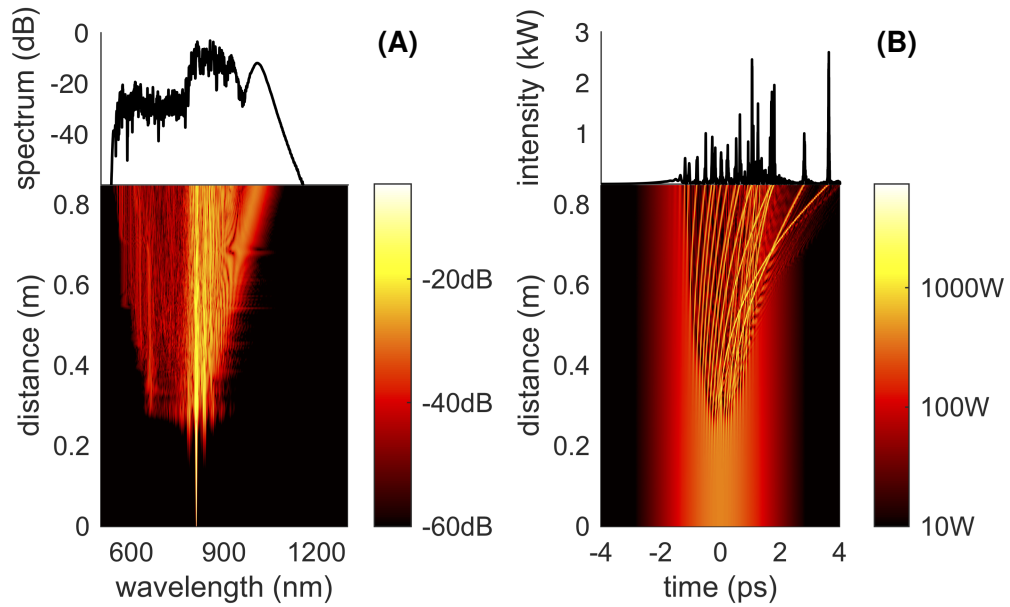
Figure 2.6. *Example of a simulated 2 ps pulse evolution over a propagation distance of 85 cm. The initial pulse has a peak power of 400 W and it is injected into the anomalous dispersion regime of a photonic crystal fiber. (A) shows the evolution of the spectral intensity (spectrum) of the pulse in dBs while (B) shows the evolution of the temporal intensity. The insets on top of the evolution show the output of the fiber.*

$A(z_0 + h/2, \omega)$ is calculated. Again, the derivative is evaluated but now at the mid-point $z = z_0 + h/2$, using the mid-point amplitude calculated at the first step Runge-Kutta step. Derivative at the mid-point (k_2) is then used to estimate the E-field amplitude at point $z = z_0 + h$ using

$$A(z_0 + h, \omega) = A(z_0, \omega) + k_2 h. \tag{2.43}$$

The linear and nonlinear terms are then combined, i.e. dispersion is let to affect the field calculated from the nonlinear interactions. The corresponding temporal field amplitude is calculated by the inverse Fourier transform of the spectral field

$$A(z, t) = \mathcal{F}^{-1}\{A(z, \omega)\} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} A(z, t) e^{-i\omega t}\, d\omega. \tag{2.44}$$

An example of simulated evolution of a single pulse propagating in a nonlinear optical fiber is shown in Figure 2.6.

3 MODULATION INSTABILITY

Modulation instability (MI) is a nonlinear process governed by the nonlinear Schrödinger equation that results from the interplay between linear and nonlinear effects. MI is a process by which a small initial periodic perturbation on top of a (quasi-) continuous wave (CW) signal is exponentially amplified. MI manifests itself as the break-up of the continuous time domain signal into a train of short temporal sub-pulses. Correspondingly in the spectral domain, the growth of frequency sidebands are observed symmetrically around the CW component. Anomalous dispersion is required for observing MI [2]. In addition to fiber optics, MI has been observed in water waves or fluids [25–27], plasma physics [28] and for Bose-Einstein condensates [29], to name a few.

Apart from a coherent periodic modulation, MI can also be seeded by noise. This results into highly complex and chaotic dynamics. It has been shown that noise-seeded or spontaneous MI can be associated with the formation of extreme events or rogue waves [5,6].

3.1 Stability analysis

The stability analysis done in this Chapter assumes a CW signal but the analysis is valid for quasi-CW signals, i.e. also in the long pulse regime ($\gtrsim$ 1 ps). Let us consider the pure NLSE derived in Section 2.3.1

$$i\frac{\partial A}{\partial z} = \frac{\beta_2}{2}\frac{\partial^2 A}{\partial T^2} - \gamma|A|^2 A, \tag{3.1}$$

which includes group velocity dispersion and Kerr nonlinearity. The CW corresponds to field $A(0,T) = \sqrt{P_0}$, where P_0 is the incident power. Since A does not depend on time, Equation 3.1 can be readily solved as

$$A(z,T) = \sqrt{P_0}\,\mathrm{e}^{i\phi_{NL}}, \tag{3.2}$$

where $\phi_{NL} = \gamma|A|^2 z = \gamma P_0 z$ is the nonlinear phase shift induced by self-phase modulation.

To analyse the stability of the propagation dynamics, let us add a small perturbation to the continuous wave signal such that

$$A(z,T) = \sqrt{P_0 + a(z,T)}\,\mathrm{e}^{i\phi_{NL}}, \tag{3.3}$$

where $a(z, T)$ is a *small* perturbation to the steady-state, i.e. $|a(z, T)|^2 \ll P_0$. Injecting A to the NLSE and neglecting all the small second-order terms (a^2), one obtains the equation governing the evolution of the small perturbation:

$$i\frac{\partial a}{\partial z} = \frac{\beta_2}{2}\frac{\partial^2 a}{\partial T^2} - \gamma P_0(a + a^*). \tag{3.4}$$

The equation can be solved using an Anzatz solution of the form

$$a(z, T) = a_1 e^{i(Kz - \Omega T)} + a_2 e^{-i(Kz - \Omega T)}, \tag{3.5}$$

where a_1 and a_2 are coefficients, K and Ω are the wavenumber and frequency of the perturbation, respectively. It must be noted that both the wave number and frequency are relative to the CW field, i.e. the actual wavenumber and frequency of the perturbation are $\beta_0 \pm K$ and $\omega_0 \pm \Omega$, respectively. Injecting a into the Equation 3.4, one obtains

$$iKa_1\,e^{i\phi} - iKa_2\,e^{-i\phi} - i\Omega^2\frac{\beta_2}{2}a_1\,e^{i\phi} - i\Omega^2\frac{\beta_2}{2}a_2\,e^{-i\phi} = i\gamma P_0\left(Ka_1\,e^{i\phi} + Ka_2\,e^{i\phi} + c.c.\right), \tag{3.6}$$

where $\phi = Kz - \Omega T$. Since Equation 3.6 must hold for all z and T, one can separate the equation into two term (with $e^{i\phi}$ and $e^{-i\phi}$ canceling out)

$$Ka_1 - \Omega^2\frac{\beta_2}{2}a_1 = \gamma P_0(Ka_1 + Ka_2)$$
$$-Ka_1 - \Omega^2\frac{\beta_2}{2}a_1 = \gamma P_0(Ka_1 + Ka_2), \tag{3.7}$$

which can be written in a matrix form

$$\begin{bmatrix} K - \Omega^2\beta_2/2 - \gamma P_0 & -\gamma P_0 \\ -\gamma P_0 & K - \Omega^2\beta_2/2 - \gamma P_0 \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = 0. \tag{3.8}$$

A nontrivial solution exists when the determinant of the matrix in Equation 3.8 is equal to zero. Thus we can solve for K, leading to [2]

$$K = \pm\frac{|\beta_2|\Omega}{2}\sqrt{\Omega^2 + \frac{4\gamma P_0}{\beta_2}}$$
$$= \pm\frac{|\beta_2|\Omega}{2}\sqrt{\Omega^2 + \text{sgn}(\beta_2)\Omega_c^2}, \tag{3.9}$$

where $\text{sgn}(\beta_2)$ is the sign of β_2 ($= \pm 1$) and $\Omega_c^2 = 4\gamma P_0/|\beta_2| = 4/(|\beta_2|L_{NL})$, with $L_{NL} = 1/(\gamma P_0)$ the *nonlinear length*. Physically, L_{NL} describes the propagation distance over which the nonlinear effects cannot be neglected anymore [2].

From Equation 3.8 one can observe that depending on the sign of the group velocity dispersion, K can be either real or complex. In the case of positive β_2 (normal dispersion), $\Omega^2 + \text{sgn}(\beta_2)\Omega_c^2$ is always positive and thus K is real valued. Inserting real valued K to the equation of the perturbation (3.5), the exponential terms will not grow in amplitude.
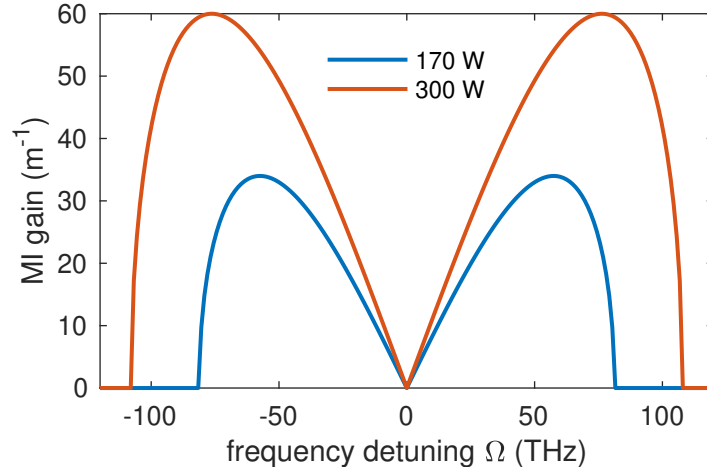
Figure 3.1. *MI power gain curve for NL-PM-750 fiber (NKT Photonics) at 825 nm. Blue and orange curves are with peak powers of 170 and 300 W, respectively. $\beta_2 = 1.03 \times 10^{-26} \ s^2 m^{-1}$, $\gamma = 0.1 \ W^{-1} m^{-1}$.*

Therefore, the steady state is stable against small perturbations in the normal dispersion regime. However, in the anomalous dispersion regime this is not the case. If $|\Omega| < \Omega_c$, the perturbation is form of

$$
\begin{aligned}
a(z, T) &= a_1 \exp\left(-\frac{|\beta_2|\Omega}{2}\sqrt{\Omega_c^2 - \Omega^2}z\right) \exp\left(-i\Omega T\right) \\
&\quad + a_2 \exp\left(\frac{|\beta_2|\Omega}{2}\sqrt{\Omega_c^2 - \Omega^2}z\right) \exp\left(i\Omega T\right) \\
&= a_1 \exp\left(-\frac{1}{2}g(\Omega)z\right) \exp\left(-i\Omega T\right) + a_2 \exp\left(\frac{1}{2}g(\Omega)z\right) \exp\left(i\Omega T\right).
\end{aligned} \tag{3.10}
$$

We observe exponential growth with *power* gain

$$
g(\Omega) = |\beta_2|\Omega\sqrt{\Omega_c^2 - \Omega^2} = |\beta_2|\Omega\sqrt{\frac{4\gamma P_0}{|\beta_2|} - \Omega^2}. \tag{3.11}
$$

Therefore, the steady state is not stable in the anomalous dispersion regime but the weak perturbations are exponentially amplified. This is process is known as modulation instability. The maximum gain is achieved for a frequency separation (from pump) of

$$
\Omega_{max} = \pm\frac{\Omega_c}{\sqrt{2}} = \pm\sqrt{\frac{2\gamma P_0}{|\beta_2|}}, \tag{3.12}
$$

resulting in the maximum value for the power gain

$$
g_{max} = g(\Omega_{max}) = \frac{|\beta_2|}{2}\Omega_c^2 = 2\gamma P_0. \tag{3.13}
$$

However, not only the frequencies with maximum gain are amplified, but all of the frequencies that fall inside the gain curve will be amplified. Few examples of MI (power) gain curves with different pump powers are shown in Figure 3.1.
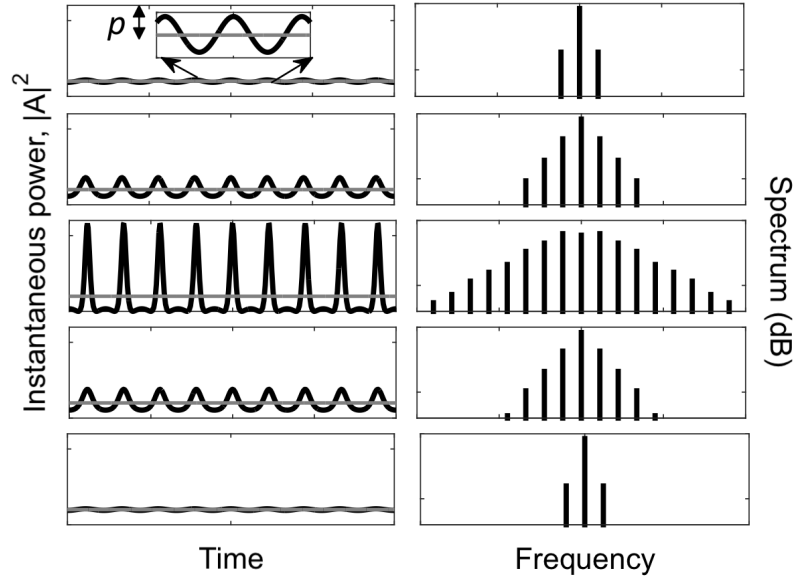
Figure 3.2. *Schematic of MI evolution. A continuous wave signal with a small sinusoidal modulation of amplitude p is broken up into high intensity structures in time, while the growth of multiple sidebands is observed in the spectral domain. The gray line illustrates the continuous power. Adapted from Ref. [30].*

In the frequency domain, MI can be interpreted as a degenerate four wave mixing process for frequencies that fall under the MI gain curve, i.e. $2\omega_p = \omega_1 + \omega_2$, where ω_p is the frequency of the pump and $\omega_1 = \omega_p - \omega_{mod}$ and $\omega_2 = \omega_p + \omega_{mod}$ are the generated frequencies [2]. An example of MI evolution for periodically modulated CW signal is shown in Figure 3.2. One can observe that the time domain signal breaks up into a train of high intensity sub-pulses. The period of the structures is determined by the inverse of the modulation frequency ω_{mod} as $T_{mod} = 2\pi/\omega_{mod}$. As the modulation is further amplified, one can observe an increasing number of sidebands in the frequency domain, each separated by the modulation frequency ω_{mod}. These sidebands are the result of a degenerate four wave mixing process between the pump and/or other sidebands, known as *cascaded FWM*. For a single-mode fiber, the phase mismatch for MI process is given by

$$\kappa = 2\gamma P_0 + 2 \sum_{m=1}^{\infty} \frac{\beta_{2m}}{(2m)!} \Omega^{2m}, \qquad (3.14)$$

where β_{2m} are the even-order dispersion terms [7]. Therefore, perfect phase matching ($\kappa = 0$) in a NLSE system is achieved with the frequency of maximum growth $\Omega_{max} = \sqrt{2\gamma P_0/|\beta_2|}$. The phase matching is achieved due to the combination of anomalous dispersion (negative GVD) and nonlinear phase shift induced by the SPM.

As the sidebands grow in amplitude, the undepleted pump approximation done in the stability analysis does not hold after the initial exponential growth stage, and the gain derived in Equation 3.11 does not lead to exponential growth anymore. One can see from Figure 3.2 that after the point of 100% contrast in the temporal intensity, energy starts flowing back to the pump component, essentially leading to the restoration of the

initial state. This kind of periodic energy exchange between the pump and the modulation sidebands is a common behavior for many nonlinear systems and it is known as the Fermi-Pasta-Ulam recurrence [31].

3.2 Spontaneous modulation instability

In the absence of a weak periodic modulation on top the continuous wave signal considered so far, MI can still develop as the result of noise amplification. Noise can be considered as a small perturbation to the CW signal instead of a well-defined periodic modulation. In general, the noise is broadband, spanning a large continuous range of frequency components. The noise can then seed MI and all the noise components within the MI gain are exponentially amplified simultaneously. This process of exponential growth of broadband noise is known as *spontaneous* or *noise-seeded* modulation instability.

The maximum MI gain is achieved for a frequency $\omega_0 \pm \Omega_c/\sqrt{2} = \omega_0 \pm \sqrt{2\gamma P_0/|\beta_2|}$ as we saw earlier, and this means that the time domain will be mainly modulated with a frequency corresponding to the maximum gain, i.e. $\sqrt{2\gamma P_0/|\beta_2|}$. Since all of the frequencies under the MI gain curve are amplified, however, the time domain modulation is not perfectly periodic but rather consists of irregularly spaced high intensity peaks. In the frequency domain, one observes sidebands that are not at located at discrete frequencies (compare to Figure 3.2) but that fills the entire MI gain curve. Cascaded FWM subsequently transfers energy to higher-order sidebands. The resulting triangular-shaped spectrum (in logarithmic scale) is the result of the cascaded energy transfer to frequencies further away from the pump [32]. Part of the energy is also eventually flowing back to the pump but this process is not identical to the case of periodic modulation, as the phase coherence is not maintained and the initial state is then never restored. A numerical NLSE simulation in Figure 3.3 illustrates the dynamics.

When seeded by noise, MI shows complex dynamics and the process is chaotic in nature such that the output spectrum after propagation in a fiber is highly dependent on the initial conditions [5,33]. Without noise (or periodic modulation), MI does not occur. Therefore, the initial random noise essentially determines the intensity profile at the output of the fiber. From an experimental point of view, noise can also manifests along the propagation due to inhomogeneities in the fiber for example, but in typical numerical simulations the initial noise condition determines the output in a deterministic way. The output spectrum shows large shot-to-shot variation between different pulses that have the same initial CW background but different noise [33]. In numerical simulations, the background noise is typically implemented by adding one photon with random phase in each frequency bin [7].
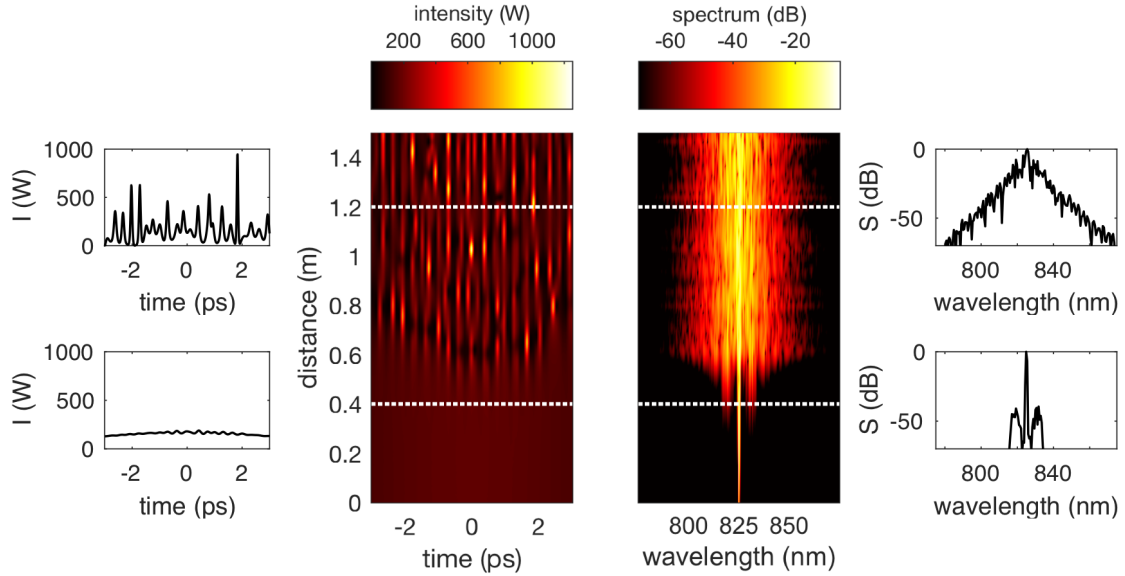
Figure 3.3. *Chaotic MI field. The noise on top of a long pulse (10 ps, 170 W peakpower) is exponentially amplified to create sidebands that further develop into a triangular shaped spectrum due to cascaded FWM. The left and right panels show the temporal intensity (I) and spectrum (S) for propagation distances of 0.4 and 1.2 m. respectively. λ_0 = 825 nm, β_2 = 1.03 × 10^{-25} s^{2}m^{-1}, γ = 0.1 W^{-1}m^{-1}.*

3.3 Breather formalism

The analytic form of the breather structures emerging from MI can be derived using the *inverse scattering transform* (IST) [34–36]. This is more convenient to be done in the *dimensionless* form of the NLSE using the following notations [2]

$$U = \frac{A}{\sqrt{P_0}}, \quad \xi = \frac{z}{L_D} = \frac{z|\beta_2|}{T_0^2}, \quad \tau = \frac{T}{T_0}, \quad N^2 = \frac{L_D}{N_{NL}} = \frac{\gamma P_0 T_0^2}{|\beta_2|}, \quad (3.15)$$

where U, ξ and τ are the normalized amplitude, propagation coordinate and time, respectively. The variable $L_D = T_0^2/|\beta_2|$ is the *dispersive length* that describes the propagation length over which dispersion becomes significant for the pulse propagation (similarly to nonlinear length L_{NL} for the nonlinear effects). The variable T_0 is the incident pulse width (related to the full width at half maximum (FWHM) of the pulse) and N is the soliton number. The NLSE can be thus written as

$$i\frac{\partial U}{\partial \xi} = \text{sgn}(\beta_2)\frac{1}{2}\frac{\partial^2 U}{\partial \tau^2} - N^2|U|^2 U, \quad (3.16)$$

which can be further simplified by assuming anomalous dispersion and defining new unitless amplitude $\psi = NU$. The dimensionless canonical form the the NLSE in the anomalous dispersion regime can thus be written as [2]

$$i\frac{\partial \psi}{\partial \xi} + \frac{1}{2}\frac{\partial^2 \psi}{\partial \tau^2} + |\psi|^2 \psi = 0. \quad (3.17)$$
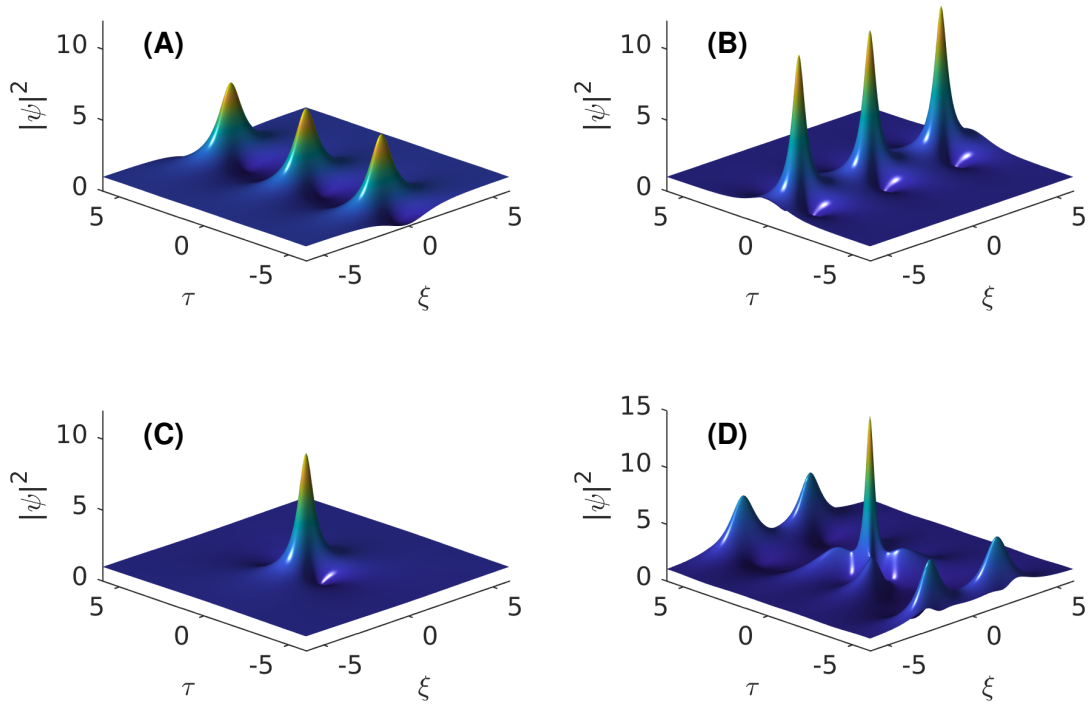
Figure 3.4. *Analytical breather solutions. (A) shows the Akhmediev breather with $a = 0.25$, (B) shows the Kuznetzov-Ma soliton with $a = 0.7$, (C) shows the Peregrine soliton with $a = 0.5$, and (D) shows a collision of two ABs with $a_1 = 0.2$ and $a_2 = 0.3$.*

There are an infinite number of solutions to the NLSE. The periodic evolution of a CW signal with small periodic modulation of top of it (as it is the case with MI) is called the *Akhmediev breather* (AB) [37]. Other important solutions include the *Peregrin soliton* (PS) [38] and the *Kuznetsov-Ma soliton* (KM soliton) [39]. All of these solutions belong to the family of solutions that evolve from a CW field, and they are known as *solitons on finite background* (SFB). These SFB solutions are shown in Figure 3.4.

The generic solution of the SFBs can be written in the dimensionless form as

$$\psi(\xi, \tau) = \left[1 + \frac{2(1 - 2a)\cosh(b\xi) + ib\sinh(b\xi)}{\sqrt{2a}\cos(\omega_a \tau) - \cosh(b\xi)}\right] e^{i\xi}, \tag{3.18}$$

and it is governed by a single variable a. Other variables are related to a as $b = \sqrt{8a(1 - 2a)}$ and $\omega_a = 2\sqrt{1 - 2a}$. Depending on the magnitude of a different SFBs are observed.

For $0 < a < 0.5$, the solution is the Akhmediev breather (AB), where maximum localization (highest amplitude and minimum temporal width) is observed for $a = 0.25$. As seen from Figure 3.4A, an AB manifests as a localized pulse train along ξ, first growing and then decaying back to the initial state. Maximum localization is observed at $\xi = 0$. For $a = 0.25$, the maximum peak intensity along propagation is approximately $|\psi_{AB}|^2_{max} = 5.8$.

For $a > 0.5$, the NLSE yields the Kuznetsov-Ma soliton that is, unlike AB, localized in time and periodic along propagation.

In the limit $a \to 0.5$, one obtains the Peregrine soliton (PS) which can be written as [38]

$$\psi_{PS}(\xi, \tau) = \left[1 - \frac{4(1 - 2i\xi)}{1 + 4\tau^2 + 4\xi^2} \right] e^{i\xi} . \tag{3.19}$$

As seen from the Figure 3.4C, the Peregrine soliton is localized both in time and the space dimensions. It also possesses the highest peak intensity among all the SFB solitons with a maximum normalized intensity of $|\psi_{PS}|^2_{max} = 9$. It must be noted that the KM soliton can exhibit higher instantaneous peak power using this formalism but after proper normalization, the peak intensities of KM do not reach 9 [36]. Peak intensities over 9 can be obtained only as the result of collision of two SFBs. An example of collision of two ABs is illustrated in Figure 3.4D.

Similar structures to these analytical solutions can also be seen in Figure 3.3 where the MI process is seeded from noise. One can see structures that are localized in time but grow and decay periodically over the propagation, similar to the dynamics of KM soliton. Akhmediev breather -like structures can be observed as well, and single high intensity peaks can be seen with localization characteristics resembling that of the Peregrine soliton. The observed structures are not of course perfectly matching to the SFB solutions but comparing the analytical solutions to the case of noise-seeded dynamics gives an insight to the physical processes that occur during MI, and during nonlinear pulse propagation in general. We can see as well that breathers may interact with each other, leading to the collision of two or more breathers.

3.4 Extreme events in modulation instability

The chaotic MI-field seeded by noise exhibits a large number of localized temporal structures (as seen in Figure 3.3, for example). Similarly, the spectral bandwidth of the field fluctuates significantly along the propagation. On rare occasions, the chaotic evolution may also lead to temporal structures with extremely high instantaneous power. These extreme events, known as *rogue waves*, seem to emerge from nowhere and disappear fast [40]. Since under specific conditions deep ocean waves are also governed by the NLSE, these extreme waves have been suggested to be the optical counterpart of the giant waves that can occur in an oceanic context. Due to the chaotic evolution of oceanic or optical waves, the study of rogue waves is extremely difficult and one can only often resort to the statistics to study these events. The physics behind rogue wave formation is still not well understood but it is suggested that it is a combination of linear and nonlinear mechanisms that drive the emergence of rogue waves [41]. Their prediction is even more difficult, and it is one of the topic of this thesis.

The study of extreme events in optics is rapidly developing due to the recent developments in real-time experimental approaches [42,43]. Measurements in real-time are crucial in order the understand the fundamental physical processes that occur during ultrafast dynamics since average measurements typically wash out any dynamical feature of

the system. The development of the dispersive Fourier transform (DFT) technique has enabled the measurements of breather spectra on a single-shot basis [10,33]. The method relies on stretching the field though a dispersive element (such as a fiber) that separates the wavelength components enough to be measured with an oscilloscope in real-time. Another significant development in real-time measurements is the time-lens system [43] that has enabled the study of picosecond breathers directly in the time domain [11]. Nevertheless, the study of ultrafast instabilities in real-time (i.e. tens of femtosecond duration) is still very challenging, and the measurements remain limited in several aspects (power and bandwidth for example). It is important to bear in mind that conventional measurements are limited to the spectral intensity only (no phase information), and therefore inferring the spectral properties from temporal properties is impossible, or vice versa, is generally not possible. Additionally, the dynamic range that describes the ratio of the minimum measurable signal compared to the maximum measurable signal, is limited to approximately 20–25 dB for single-shot spectral measurements in a fiber DFT system [44]. This is a major limitation to the study of breathers in the spectral domain since the extreme events in MI are associated with an increased spectral bandwidth that is only visible below -40 dB [45,46].

The study of the breathers in this work is limited to a fixed distance after propagation in a fiber. Figure 3.5 shows the dynamics of early MI stage. In A and B, a typical evolution of a single realization (pulse) is shown. As seen before, the evolution begins with the growth of distinct first-order sidebands and small temporal envelope fluctuations, evolving with the emergence of higher-order sidebands and the complete breakdown of the pulse envelope into various breather structures. Figures 3.5C and D show examples of corresponding spectral and temporal intensity profiles where we can observe large variation in both spectral and temporal profiles due to the random noise background. For some of the realizations we observe one extremely high intensity temporal peak with broadened spectral profile. The maximum temporal intensity is highlighted with a blue circle. For an "average" case however, the temporal profile consists of multiple breathers with relative small intensities for each of them. The top panels show the average spectrum and temporal intensity calculated over 50,000 realization. We can see that the averaged temporal profile is essentially smooth, highlighting the randomness of the breather dynamics. All of the dynamic features are washed away when averaging such dynamics, and therefore real-time measurements are extremely important when trying to understand the underlying physics.

Figure 3.5E shows the calculated probability density function (PDF) of the breather intensities from an ensemble 50,000 distinct simulations with different random noise seeds. The solid blue line shows the peak power distribution of all breathers that appear inside a 1.5 ps time window centered at zero time. The red dashed curve shows the PDF of the largest breather intensity for each realization. For both cases, we can observe the long extended tail towards the high intensities that is typical behavior for MI [11,47]. As we are interested in the extreme events with the highest intensities, it should be clear from Figure 3.5E that these correspond to the tail of the PDF which is identical to the PDF of
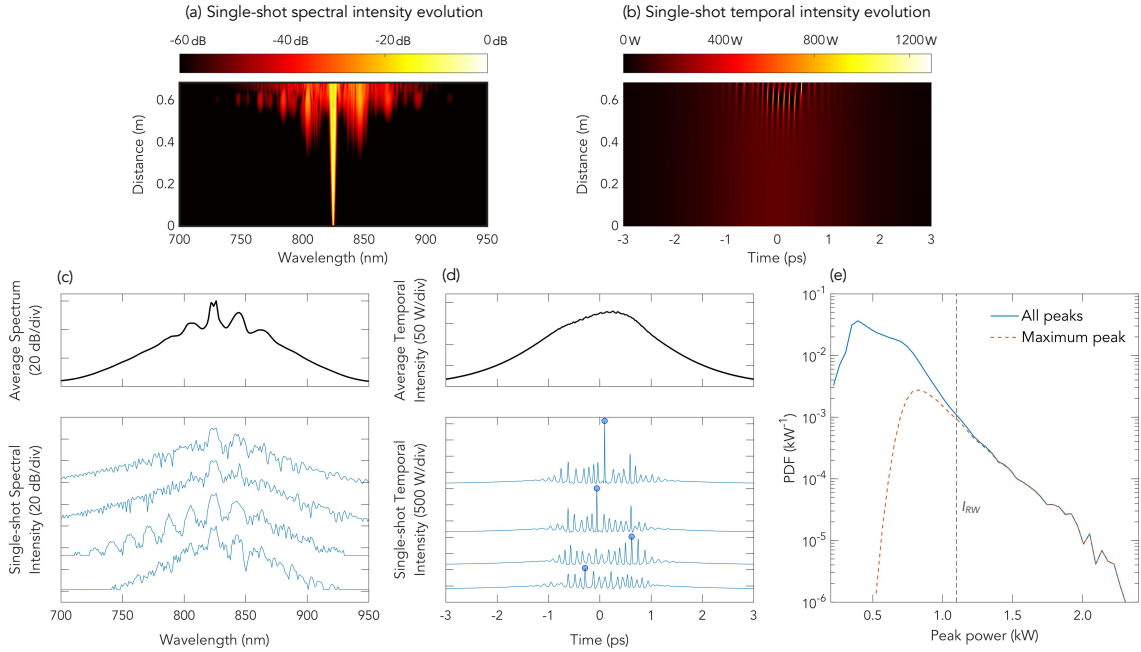
Figure 3.5. *Simulated dynamics of picosecond pulse propagation. (a) and (b) shows the evolution of a single realization in the spectral and temporal domain, respectively. (c) and (d) show a few selected examples of corresponding single-shot spectra and temporal intensity profiles, respectively. The top panels show the average spectra and temporal intensity calculated over 50,000 distinct realizations. The maximum temporal intensity for each realization is highlighted with blue circles. (e) shows the calculated probability density function of all breather intensities (within 1.5 ps time window) in blue and the maximum intensity in each realization in red. I_{RW} is the rogue wave threshold defined as twice the mean intensity of the highest third of the breather intensities. Taken from Ref. [48].*

the maximum intensity in each realization. The vertical dashed black line in Figure 3.5E indicates the rogue wave threshold intensity defined as $I_{RW} = 2I_{1/3}$ where $I_{1/3}$ is the mean intensity of highest third of the all breather intensities.

As noted in Section 3.3, the highest intensity of the elementary breather solutions is obtained with the Peregrine soliton. Event with higher intensities can only be obtained as the result of breather collisions. Therefore, we can anticipate that the extremely high intensities observed in MI would be the result of an emergence of Peregrine soliton or higher-order ABs emerging from the chaotic dynamics.

4 SUPERCONTINUUM GENERATION

Supercontinuum (SC) refers to the generation of ultra-broadband light from narrowband laser light via various nonlinear processes [7]. Since the first report of generation of SC in bulk glass in 1970 [49,50], studies of SC generation in various media have been reported [51,52]. In general, SC generation in bulk media involves complex coupling between spatial and temporal effects [7]. As intriguing this type of physics is, from an application view point it is common to use nonlinear optical fibers delivering a single Gaussian mode and where the spatial effects can be neglected. In multimode fibers, however, the spatial effects cannot be neglected as higher-order modes can contribute significantly to the observed dynamics and spectral broadening [53]. Over the last few decades, the use of photonic crystal fibers (PCFs) has gained large interest due to the possibility of dispersion engineering [7]. Large effort has been put into the development of SC sources in the ultraviolet or infra-red. Applications of supercontinuum sources range from spectroscopy and remote sensing to imaging applications.

The properties of SC are highly sensitive on the input field parameters, the dispersion regime and the nonlinear coefficient. This Chapter gives a brief introduction to the different dynamic regimes of SC generation. The main focus is on the highly incoherent long pulse and anomalous dispersion regime where the SC generation is seeded by an initial stage of chaotic MI as described in the previous Chapter.

4.1 Supercontinuum dynamics

Supercontinua can be generated in various ways. The methods can range from the use of ultrashort mode-locked laser pulses to continuous wave lasers at various wavelengths [7]. In addition to the pump source, various fibers can be used from soft glass and telecommunication fibers to PCFs which all have different dispersion and nonlinear properties.

The different regimes of SC dynamics can be coarsely divided into four categories depending on the pump pulse duration (short/long) and the dispersion regime of the pump (normal/anomalous). The dominant effects for each regime are show in Table 4.1. The limit between short and long pulses is somewhat arbitrary because the dynamics are dependent on other variables as well in principle, but 500 fs can be considered as a good approximation [30].

Table 4.1. *Different dynamical regimes of supercontinuum generation with dominating nonlinear effects. After Ref. [15].*

	Short pulses	Long pulses
Anomalous	Soliton dynamics	Modulation instability
		Soliton dynamics
Normal	Self-phase modulation	Raman scattering
	Four wave mixing	Four wave mixing

4.1.1 Short pulse regime

In the short pulse regime, it is the pulse spectral components that seed the SC generation. We begin by describing the dynamics in the **anomalous dispersion** regime. The dominating effects include soliton dynamics and dispersive wave generation. As mentioned before, optical solitons are solutions of the Schrödinger equation where the SPM-induced nonlinear phase-shift is compensated by the anomalous dispersion induced down-chirp. They can propagate undistorted over long propagation distances [54]. A *fundamental* sech-type soliton can be analytically expressed as

$$A(z, T) = \sqrt{P_0} \operatorname{sech}\left(\frac{T}{T_0}\right),$$ (4.1)

with $T_0 = T_{FWHM}/1.763$, where T_{FWHM} is the FWHM of the intensity profile [2]. The requirement to excite a fundamental soliton is that the input parameters should fulfill the condition

$$N = \sqrt{\frac{\gamma P_0 T_0^2}{|\beta_2|}} = 1 \quad \Leftrightarrow \quad P_0 = \frac{|\beta_2|}{\gamma T_0^2}.$$ (4.2)

The variable N is the soliton number or soliton order used in Equation 3.15.

Fundamental solitons are not, however, the only solitonic solution of the NLSE, and higher-order solutions also exist where the soliton number N can take any positive integer value. These higher-order solitons do not propagate undistorted during propagation but they show periodic variations in both the spectral and temporal domain [2].

Higher-order solitons are highly sensitive to perturbations such as higher-order dispersion and Raman scattering [55]. Since the bandwidth of ultrashort pulses is relative broad by nature, higher-order effects cannot be neglected (GNLSE regime). A Nth-order soliton will eventually break up into N fundamental solitons who will propagate independently and separate temporally due to different group velocities. The peak powers and durations of the ejected fundamental solitons $A_n(z, T) = \sqrt{P_n} \operatorname{sech}(T/T_n)$ are given by [7,55]

$$P_n = P_0 \frac{(2N - 2n + 1)^2}{N^2}$$ (4.3)

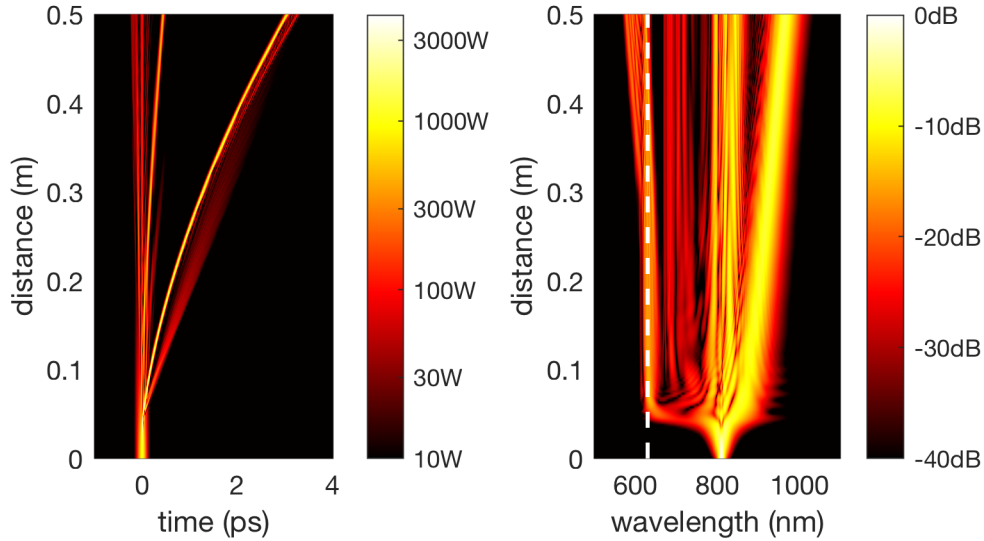$$T_n = \frac{T_0}{2N - 2n + 1},$$ (4.4)

Figure 4.1. *Short pulse (100 fs duration, 1 kW peak power) SC in the anomalous dispersion regime. Left panel shows the SC evolution in time, and right panel in spectrum over the propagation distance of 50 cm. The ZDW of the fiber is 750 nm, while the pump wavelength is 810 nm. The white dashed line indicates the position of the dispersive wave in the normal dispersion regime.*

where $n = 1, 2, ..., N$. This process whereby a higher-order soliton breaks up into fundamental solitons is known as *soliton fission*. The propagation distance after which soliton fission typically occurs is approximately given by $z_{fiss} \approx L_D/N = NL_{NL}$ [7], where L_D and L_{NL} are the dispersive and nonlinear length, respectively. Figure 4.1 shows simulated (GNLSE) evolution of a $N = 5$, 100 fs pulse injected into the anomalous dispersion regime. One can see how soliton fission occurs approximately after a distance of 6 cm following the maximum temporal compression (and spectral expansion) of the pump.

In the presence of higher-order dispersion, soliton fission is modified in two ways [7]. Firstly, the fundamental solitons are shifted towards longer wavelengths due to the Raman effect. This is known as the soliton self-frequency shift (SSFS) as mentioned previously. One can see this from Figure 4.1 where the ejected high intensity solitons start red-shifting towards longer wavelengths, and in time, the trajectory of the solitons start to curve, separating from the pump residue at increasing rate. Secondly, one often sees that part of the spectrum is extended into the normal dispersion regime. This part of the spectrum is known as the *dispersive wave*. The higher-order soliton radiates part of its energy as Charenkov radiation. The frequency of the dispersive wave can obtained from the phase-matching condition between the soliton and the dispersive wave, given by

$$\omega_{DW} = \omega_0 + 3\frac{|\beta_2|}{\beta_3} - \frac{\gamma P_0}{2}, \tag{4.5}$$

where, respectively, ω_0 and P_0 are the central frequency and peak power of the soliton [56]. Figure 4.1 shows the position of the dispersive wave marked as a dashed white line.

In the **normal dispersion** pumping, soliton dynamics do not occur. The dominating effects
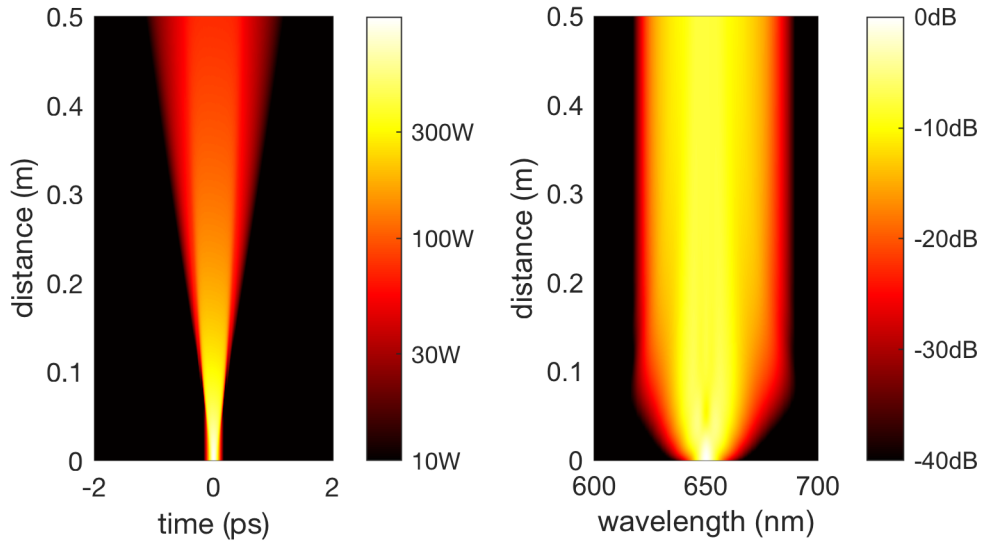
Figure 4.2. *Short pulse (100 fs duration, 1 kW peak power) SC in normal dispersion regime. Left panel shows the SC evolution in time, and right panel in spectrum over the propagation distance of 50 cm. The ZDW of the fiber is 750 nm, while the pump wavelength is 650 nm.*

are thus self-phase modulation and four wave mixing of the pump spectral components. Figure 4.2 shows these dynamics. The figure shows how the initial broadening by SPM stabilizes as the pulse spreads in time due to dispersion. This is because of the nonlinear nature of SPM. The nonlinear phase-shift, and thus the instantaneous frequency, induced by SPM are dependent on the peak power. As the peak power decreases along the propagation due to dispersion, the spectral broadening decreases rapidly. Since the spectral broadening occur during the first centimeters, there is no need for long fibers, unlike in the anomalous dispersion regime where the soliton self-frequency shift broadens the spectrum along the propagation. In general, the spectra generated in the normal dispersion regime are very flat in contrast with that obtained in the anomalous regime [15].

4.1.2 Long pulse regime

The dynamics governing long pulse SC generation differ from that in the short pulse regime. This is because the input spectrum is relatively narrow and the dynamics in this case are not seeded by the pump itself but rather by the noise outside the pump bandwidth.

In the **normal dispersion** regime, the dominating nonlinear mechanisms are Raman scattering and four wave mixing. Essentially all dynamics in the long pulse regime are seeded by noise. The noise can be amplified by stimulated Raman scattering causing a broadband peaks on the long-wavelength side of the spectrum. All the noise components under the Raman gain curve shown in Figure 2.4 are amplified. FWM with the pump and cascaded Raman peaks can cause additional spectral components, leading to cascaded FWM.
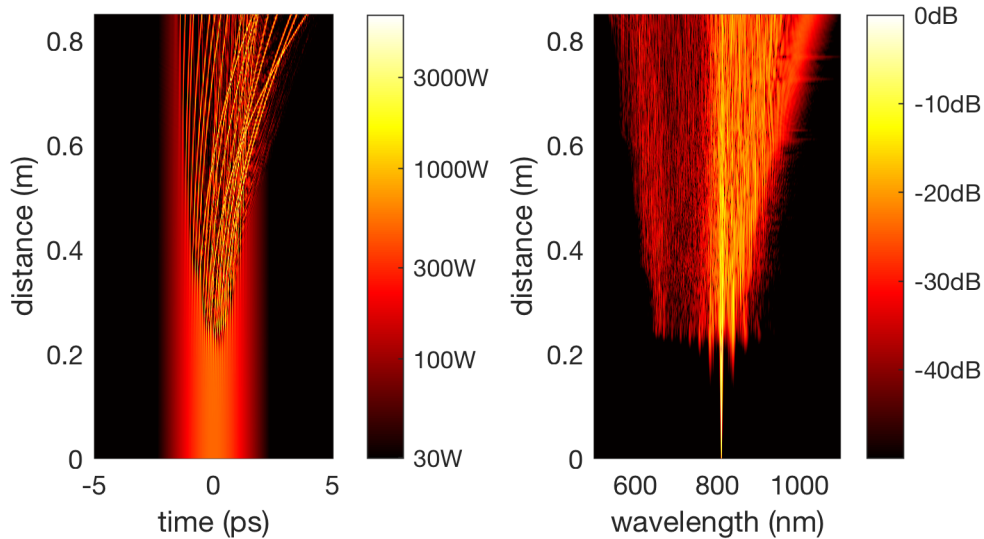
Figure 4.3. *Long pulse (2 ps duration, 500 W peak power) SC in anomalous dispersion regime. Left panel shows the SC evolution in time, and right panel in spectrum over the propagation distance of 85 cm. The ZDW of the fiber is 750 nm, while the pump wavelength is 810 nm. The soliton number for this case is 72, compared to 5 in the case of fs-SC in Figure 4.1.*

In the **anomalous dispersion** regime, long pulse SC generation is generally seeded by spontaneous modulation instability described in details in Section 3.2. Spontaneous MI leads to the emergence of chaotic high-intensity temporal breathers along with the broadened triangular-shaped spectrum. This chaotic process triggers the SC generation. As spontaneous MI exhibits large shot-to-shot variation, the generated SC spectra also show significant fluctuations. This mechanism is different compared to short pulse SC, where the soliton fission occurs in a deterministic way from the pump spectral components rather than from the noise-seeded dynamics. An example of SC generated from long pulses is shown in Figure 4.3.

Due to perturbations to the pure NLSE regime, such as the Raman effect and higher-order dispersion, the AB-like breathers generated in the MI phase subsequently evolve into fundamental solitons that separate from the pump residues [8]. Similarly to SC generated in the short pulse regime, the fundamental solitons experience the Raman induced self-frequency shift and dispersive wave generation. The solitons that are created close to the center of the pulse envelope typically have the largest intensity. These high intensity solitons undergo the largest rate of soliton self-frequency shift [57], and are thus highly red-shifted. Soliton collisions may increase the rate of frequency-shift [58,59].

In the spectral domain, one can see the spectral wings of MI after propagation of approximately 20 cm which initiate the spectral broadening. The formation of solitons (at around 25 cm of propagation) further increases the spectral bandwidth, leading to multiple dispersive waves on the short-wavelength side and extension to longer wavelengths due to the soliton self-frequency shift. Different initial noise, however, would lead to different output spectrum as we will see in the next section.

4.2 Coherence of supercontinuum

An important feature of SC for many applications is its coherence. Coherence essentially describes how much the output fields of the fiber resemble each other, both in amplitude and phase. This is known as temporal coherence, describing the coherence properties at a given spatial location but at different time instances. Similarly, spatial coherence measures the correlation of electric field at a given time but at different spatial locations. For single-mode fibers, however, the spatial coherence is perfect and, thus is not discussed further in this work.

A common approach to analyse the stability of SC spectra is to use the (modulus of) complex degree of first-order coherence at each wavelength defined as

$$|g_{12}^{(1)}(\lambda)| = \left| \frac{\langle E_1^*(\lambda) E_2^*(\lambda) \rangle}{\sqrt{\langle |E_1(\lambda)|^2 \rangle \langle |E_2(\lambda)|^2 \rangle}} \right|, \tag{4.6}$$

where the angular brackets denote an ensemble average over many pairs of generated SC spectra [60]. The maximum value is 1 and minimum 0, corresponding to completely coherent and incoherent cases, respectively. The complex degree of first-order coherence measures the stability of phase only, and does not directly give information about intensity fluctuations of the E-field. However, the phase and intensity fluctuations are closely related, and therefore, the complex degree of first-order coherence is a good parameter for describing the stability of supercontinua [7]. In the case of identical output complex spectra, the complex degree of first-order coherence would be one over the whole SC bandwidth. Conversely, if the SC spectra are generated with fully random phase, coherence would be zero.

The stability of SC highly depends on the pump and fiber parameters. In the short pulse regime, SC generated in the normal dispersion are generally coherent if the pump wavelength is far from the ZDW since the chaotic MI and soliton dynamics are excluded [7]. Secondly, the effect of pulse duration is demonstrated in Figure 4.4. As mentioned previously, SC generation in the anomalous dispersion regime and with long pulses is seeded by the chaotic MI field. The nature of these random dynamics manifests in the coherence of the generated supercontinua as well.

In the anomalous dispersion regime, the soliton number can also be used to provide an estimate of the expected coherence of the generated SC [7]. In the case of low soliton number ($N < 10$), the supercontinuum exhibits high coherence close to unity. Other the other hand, large soliton number ($N > 30$) leads to incoherence in general. For the cases in Figure 4.4, the soliton number corresponding to the coherent femtosecond SC in (A) is low ($N = 5$) while the incoherent picosecond SC in (C) is generated with a high soliton number ($N = 72$).
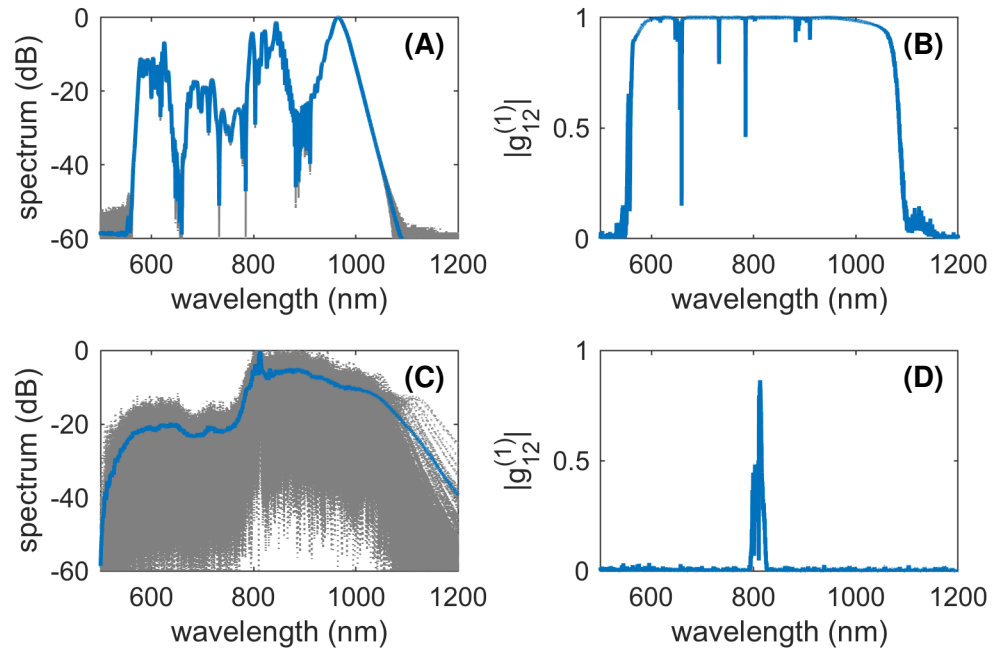
Figure 4.4. *(A) and (C) show the mean spectrum over an ensemble of 200 simulated realizations in blue for femtosecond and picosecond supercontinua, respectively. The individual realizations are shown as gray dashed lines. The complex degree of first-order coherence $|g_{12}^{(1)}|$ corresponding to femtosecond (A) and picosecond SC (C) is shown in (B) and (D), respectively. We can observe the high coherence for the short pulse SC while the SC generated in the long pulse regime shows high shot-to-shot fluctuations, leading to low coherence.*

4.3 Optical rogue waves

The term *optical rogue wave* was first introduced in 2007 by Solli *et al.* [10], who studied shot-to-shot characteristics of incoherent long-pulse SC spectra. They observed that on rare occasions, the SC spectra was extended far to the long wavelengths. Interestingly, similar dynamics have also been observed with relatively short (200 fs) pulses as well [61]. A conventional method to study the statistics of optical rogue waves is to use fast detectors to measure the energy of a limited bandwidth of the supercontinuum, which is typically chosen as the long-wavelength edge that contains the highly redshifted solitonic components. However, recent developments in real-time measurement techniques (as we will see in Chapter 6) have made it possible to measure relatively broadband spectra in real-time, which enables new possibilities in the characterization of rogue waves.

A characteristic feature of optical rogue waves (in SC generation context) is the highly skewed statistical distribution of the long-wavelength edge energy of an incoherent super-continuum. The so-called "L-shaped" distribution has been observed in both experiments and numerical simulations [5,10,62–64]. Figure 4.5 shows an example of this distribution. The figure shows the energy distribution over 30,000 SC realizations filtered with a long-pass filter inserted at the long-wavelength edge of the supercontinua. One can
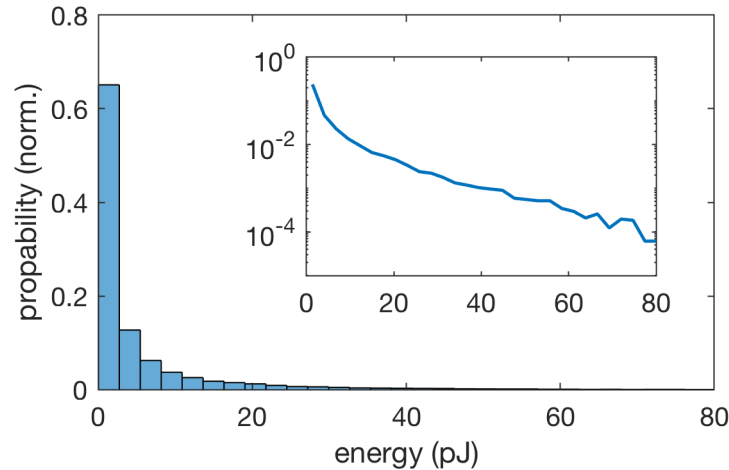
Figure 4.5. *"L-shaped" supercontinuum long-wavelength edge energy distribution. The histogram shows the energy distribution over 30,000 simulated long pulse supercontinua using a long-pass filter. The distribution extends far to large energies indicating emergence of rogue waves. The inset shows the histogram in logarithmic scale.*

observe that the large majority of the realizations exhibit nearly zero energy after addition of the filter. Occasionally, however, a relatively large portion of the pulse energy passes the long-pass filter. These rare events can be observed in Figure 4.4C as well. For some of the realizations, the long-wavelength edge is significantly extended. These events of extremely broad spectrum are referred as *optical rogue waves*. The fluctuations on the long-wavelength edge are also reflected on the short-wavelength edge, through the coupling between solitons and dispersive waves [61]. It must be noted that the selection of the cut-on wavelength of the long-pass filter alters the energy distribution. By setting the cut-on wavelength closer to the pump, the distribution is no longer "L-shaped" but rather a Gaussian [61].

The use of a long-pass filter to analyse rogue waves isolates solitons that are red-shifted more compared to the average case. This is well illustrated in Figure 4.6, where examples of a rogue wave and an average case are shown using the spectrogram representation (which is obtain by X-FROG [65,66], for example). The spectrogram representation reveals how the different spectral components are spread in time, and vice versa. The solitons appear as distinct localized structures, both in time and wavelength, on the the long-wavelength side, while the dispersive waves appear in the short-wavelength side in the spectrogram. The residual of the pump can be seen as a vertical line at the pump wavelength. The rogue wave case corresponds to emergence of an extremely red-shifted soliton, referred to as *rogue soliton* (RS), while for the average case none of the solitons are red-shifted significantly (i.e. on average basis).

Since SC generation with long pulses is highly chaotic, there is possibility for another spectral broadening mechanism in addition to the soliton self-frequency shift. As seen in Figures 3.3 and 4.3, the collision of solitons leads to significant instantaneous spectral broadening. Extending the analysis from extremely red-shifted rogue solitons to the "full
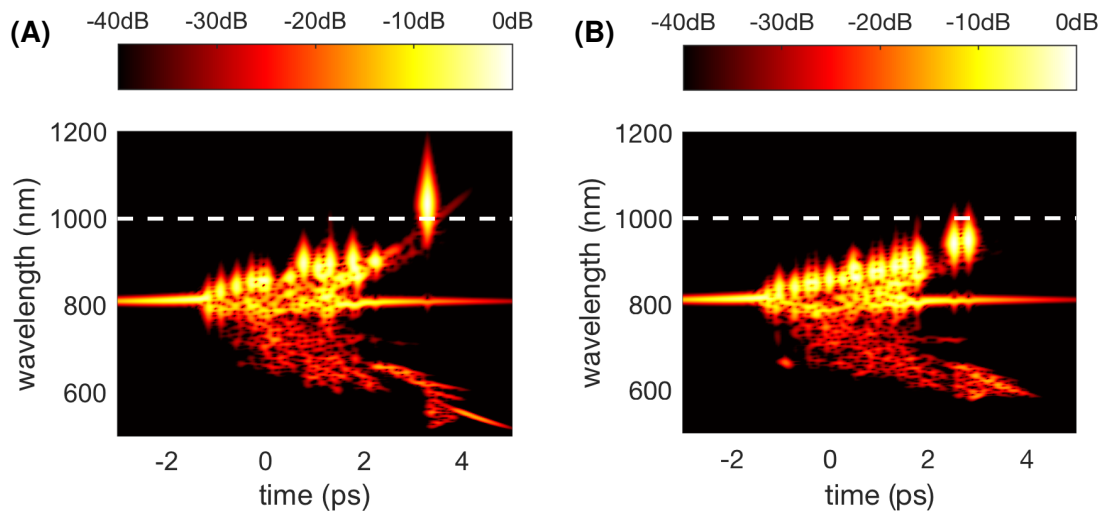
Figure 4.6. *Spectrogram representation of noisy supercontinua generated in the anomalous dispersion regime with long pulses. In (A), one can notice a highly red-shifted rogue soliton that escapes the long-pass filter indicated with white dashed line. (B) shows an average realization where none of the solitons pass the virtual long-pass filter.*

field" reveals interesting insight to the chaotic SC dynamics [63]. In fact, the highest instantaneous intensities in the long pulse SC do not typically result from rogue solitons but from the collision of two solitons that have not red-shifted in a rogue soliton manner. The "full field analysis" here refers to identifying the maximum temporal intensities over the full temporal intensity window. The term rogue wave has also been used for describing extremely high peak powers in general, similarly to the cases seen in spontaneous MI. The difficulty of predicting such extreme events lies in the multiple contributions from rogue solitons and soliton collisions taking place in the chaotic MI field, or even from collisions of two already red-shifted solitons. Extremely large bandwidths associated with soliton collisions may contribute to the energy distribution in Figure 4.5, as well, depending on the location of the long-pass filter. Nevertheless, soliton collisions are essentially always involved in the formation of rogue waves, was it purely about extremely high instantaneous intensities or the emergence of rogue solitons.

The "full field" peak power distribution of 30,000 SC simulations (approximately 550,000 peaks retrieved in total) is shown in Figure 4.7 along with the distribution obtained after applying the long-pass filter (rogue solitons). The distribution of the highest peak power associated with each realization is also shown, and one can clearly see that in fact rogue solitons are actually not the events with the largest intensity. Instead, the rogue events correspond to the collisions of two (or even three) solitons. Figure 4.7B shows the spectrogram of such event exhibiting extremely high intensity and significant spectral broadening. It must be noted that the spectral filter used for capturing the rogue solitons may decrease their actual peak power but in any case their intensity is below that of soliton collisions in general.

One needs to keeps in mind that the term *rogue wave* refers to a wave with unusually
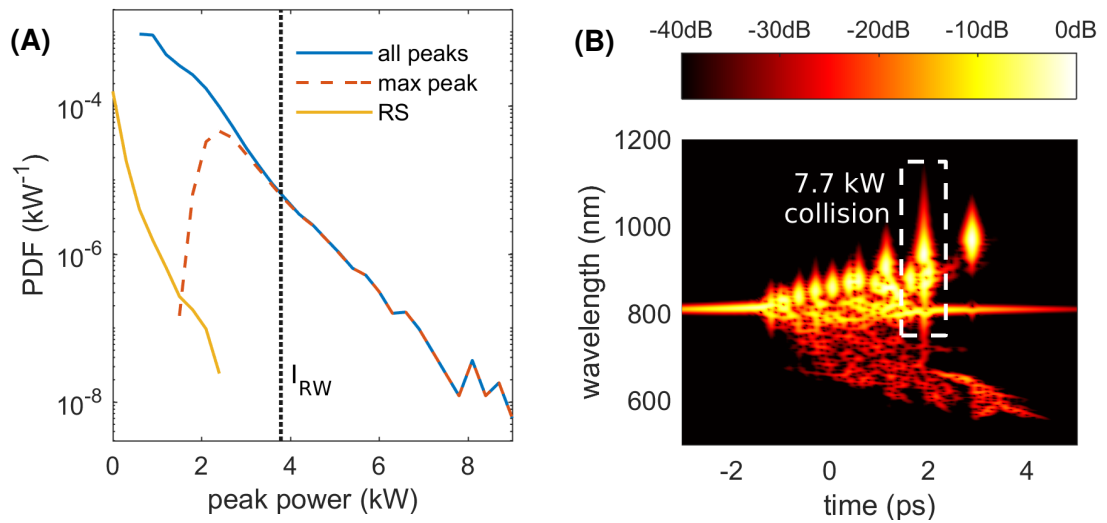
Figure 4.7. *Full field analysis of rogue waves in supercontinuum generation. (A) shows the probability density function (PDF) calculated from the full temporal intensity trace in blue and its maximum for each realization in red, along with the peak powers of the rogue solitons (RS) in yellow where the long-pass filter at 1000 nm is applied. The rogue wave threshold I_{RW} for all of the temporal peaks is indicated with the dashed vertical line. (B) shows a spectrogram of a soliton collision associated with extremely high peak power (7.7 kW). In fact the extreme end of the distribution in (A) is populated with such soliton collisions only.*

high amplitude. This does not necessarily mean that the amplitude or intensity of rogue waves would be extremely large. Yet, compared to an average wave (i.e. typical to specific conditions), the rogue waves are enormous.

5 MACHINE LEARNING

Machine learning (ML) is currently one of fastest growing fields of science. It describes an ensemble of computational and statistical methods that can generate predictive models or find patterns from a set of data without relying on predetermined physical models. The applications of ML are vast, ranging from computer vision and speech recognition, to language processing and data-driven diagnostics in medicine, to name a few [1]. Perhaps one of the most striking application of ML is the AlphaGo Zero -algorithm [67] that has allowed to beat the world champion in an ancient Chinese game of Go, a task that was thought to be impossible for a machine. The massive increase in the use of ML can be explained by two factors. Firstly, the computational power, especially using the graphics processing units (GPUs), has massively increased during the last decade. This in turn has enabled the development of new algorithms for machine learning such as convolutional [68] and recurrent neural networks [69].

One can find in the literature different terms to describe machine learning including deep learning (DL) and artificial intelligence (AI). The difference between these terms has somewhat become vague and the main focus in machine learning has shifted to artificial neural networks (NN) introduced in Section 5.1.4. But roughly speaking, AI describes all algorithms that are "intelligent". It is the broadest class of these three. ML is therefore a subclass of AI. Essentially ML provides to the computational and statistical tools or methods to realize AI. On the other hand, DL involves the use of large neural networks, and is thus a special subclass of ML.

The use of machine learning in physics has also gained significant interest during the last few years, including in optics. The use of ML for medical imaging applications [70,71] is common nowadays but other unconventional imaging and spectroscopy applications have been demonstrated such as imaging through scattering media [72] or ML-based ghost imaging [73]. Also, various ultrafast optical applications have been demonstrated including laser optimization [74] and ultrashort pulse characterization [75].

The tasks of machine learning are typically separated into two categories: *supervised* and *unsupervised* learning tasks. In supervised learning, one has the knowledge of the inputs and outputs of a specific system. The goal is to build a predictive model that can be utilized for *classification* or *regression* tasks. In unsupervised learning, the task is to find natural patterns from an ensemble of data, typically for classification purposes. These two main classes will be explained in details in the following sections with emphasis on the neural networks that fall into the supervised learning category.

5.1 Supervised learning

The supervised learning methods are the most used nowadays [1]. The goal is to create a model that can be used for mapping outputs (Y) from input variables $(\mathbf{X})$. The training of the model is realized by introducing corresponding pairs of inputs and output training data $(\mathbf{X}_n, Y_n)$. The predictions from the derived model, noted by $\hat{Y}$, are compared to the actual values (Y) corresponding the input and the model is adjusted in a way that the prediction error over the training ensemble is minimized.

Input data $(\mathbf{X})$ may be in a form of vectors, or they may be more complex such as images or text. Outputs (Y) can take various forms as well, from a single floating number that corresponds to a regression problem to an integer value for a classification problem. Similarly to the input, the output of the model may be complex in the form text or image.

5.1.1 Least squares and nearest neighbors

Least squares is a regression method used to fit a data set. For simplicity, let us consider a linear fit $y_n = ax_n + b + e_n$ to two-dimensional data (x_n, y_n). Here, a is the first order coefficient, b is the constant term and e_n is error of the fit for individual data points. For a set of data, the problem can be written in a matrix form

$$
\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix} \quad \Leftrightarrow \quad \mathbf{y} = \mathbf{X}\theta + \mathbf{e}. \tag{5.1}
$$

Given that $\mathbf{X}^T\mathbf{X}$ is not singular, the solution for coefficients a and b that minimizes the squared error $\mathbf{e}^T\mathbf{e}$ is given by [76]

$$
\hat{\theta}_{LS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \tag{5.2}
$$

Similar methods can be readily applied for arbitrarily shaped fits by changing the matrix $\mathbf{X}$.

Nearest neighbor is an extremely simple and fast classification method. The method is based on defining a set of different classes that depend on the nearby samples in the input space. Nearest neighbor simply extracts the class label of the most similar training sample. For K-nearest neighbor method, the class is selected as a "majority vote" among K nearest neighbors. The class label for point x in a two-class case (0 or 1) can be thus simply calculated from

$$
\hat{y}(x) = \frac{1}{k} \sum_{x_i \in N_k(x)} y_i, \tag{5.3}
$$

where $N_k(x)$ is the neighborhood of x defined by k closest points (x_i) in the training data

set [76]. The neighborhood is typically determined by the Euclidean distance.

5.1.2 Linear classifiers

Linear classifiers have been widely used before the development of neural networks and are still often used. The classification rule for a two-class case can be written as

$$F(\mathbf{x}) = \begin{cases} \text{Class 1,} & \text{if } \mathbf{w}^T\mathbf{x} < b \\ \text{Class 2,} & \text{if } \mathbf{w}^T\mathbf{x} \geq b, \end{cases} \tag{5.4}$$

where the weights $\mathbf{w}$ are determined from the training data and b is the class threshold.

Linear discriminant analysis finds the best possible linear projection for high dimensional data [76]. In other words, one is looking for a projection $\mathbb{R}^N \to \mathbb{R}$ that maximizes the separation of the classes.

One of the most common methods during early 2000s was the **support vector machine**. The goal is to find *support vectors* that separate two classes in an optimal way, i.e. maximize the margin M between two parallel planes separating the classes. The optimization problem can be written as

$$\begin{aligned} &\underset{\mathbf{w},b,\|\mathbf{w}\|}{\text{maximize }} M \\ &\text{subject to } y_i(\mathbf{x}_i^T\mathbf{w}) + b \geq M, \quad i = 1, \cdots, N, \end{aligned} \tag{5.5}$$

where $y_i \in \{-1, 1\}$ are the class labels [76]. This process is illustrated in Figure 5.1. In the case of non-separable classes, similar method can be used but the samples on the wrong side of the decision border need to be penalized. Using the *kernel trick*, the margins may be nonlinear as well.

Another commonly used linear classifiers is the **logistic regression**, a probabilistic model that gives the probability for a sample to belong to a specific class. Essentially, the logistic regression maps a linear projection through a logistic function to give a probability for the prediction [76].

5.1.3 Ensemble methods

Ensemble methods are different from the methods introduced so far. They consist of an ensemble of weak classifiers and the classification is done as a "majority vote" from the predictions of the weak classifiers [77]. The advantage is that, even though the individual classifiers are not as accurate as many other methods, the averaged predictions decrease the error.

Perhaps the most popular of the ensemble methods is the **random forest**. As the name
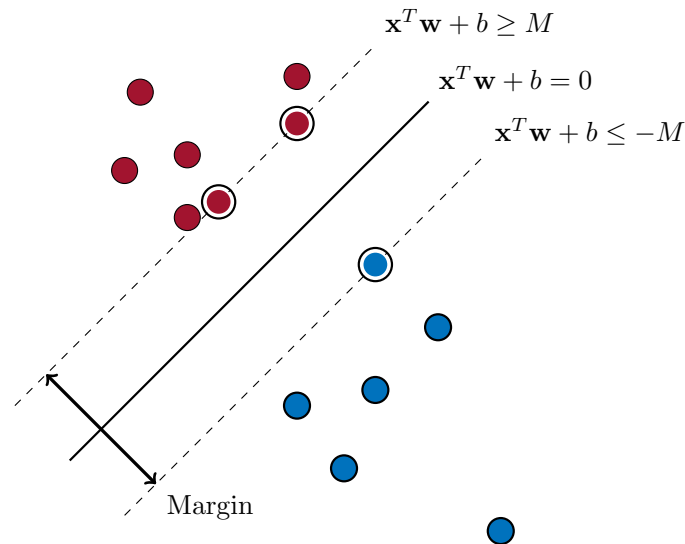
$$\mathbf{x}^T\mathbf{w} + b \geq M$$
$$\mathbf{x}^T\mathbf{w} + b = 0$$
$$\mathbf{x}^T\mathbf{w} + b \leq -M$$

Margin

Figure 5.1. *Schematic of support vector machine. The two classes are shown with different colors and the support vectors have a white circle around them. The decision boundary is shown with solid line and the margins are shown with dashed line.*

suggests, random forest classifiers consist of many decorrelated decision trees [76]. This is because the tree structures often overlearn the training data. Therefore, the trees are trained with hiding some of the information from the training data. Most of the other ensemble methods are based on variation of random forests, such as AdaBoost and gradient boosted trees.

5.1.4 Neural networks

Neural networks (NNs) are widely used nowadays for various machine learning applications. In particular, the use of convolutional neural networks has revolutionized image recognition where they are now widely used. The more traditional feed-forward neural networks that are used in this work still constitute an essential aspect of the machine learning field. This section gives the theoretical background about the training and predictions made by feed-forward NNs.

NNs can be considered as highly nonlinear transfer functions that relate the inputs and outputs of the network. They consist of one or more computing layers that are connected to the previous and following layers with varying importance between the computational units, called nodes or neurons, which create each layer. The prediction of the NN (*forward prediction*) is obtained by feeding the input to the network, which is the spectral intensity profile in our case. Training of the network is done by a process called *backpropagation*.

Forward prediction

The machine learning problem in our case is essentially a regression problem. A given vector input is mapped into a scalar output that describes the input vector. The task is to find a transfer function that maps the inputs and outputs over a set of input and output pairs.

As mentioned before, a feed-forward neural network (also known as the vanilla network) is used in this work. The NN consists of nonlinear computational units called nodes or neurons that are arranged into layer structures. The layers include an input layer, hidden layer(s) and an output layer. The input layer is not involved in the training of the network. Its task is to preprocess the inputs of the network, for example by normalizing the data. The data are then passed to the hidden layers. Each of the hidden nodes are connected to the nodes in the previous and the following layers with adjustable importance or *weights*. In case of a fully connected network, i.e. where all of the hidden nodes are connected to all of the nodes in the previous and following layers, the network is called *dense*. The last layer is the output layer that produces the response of the network to a given input, i.e. the prediction of the network. A schematic illustration of a feed-forward NN is shown in Figure 5.2. The input is the spectral intensity profile in this case and the output is the corresponding temporal intensity maximum. Figure 5.2B shows the operation of a single hidden neuron $n_i^{(k)}$, where i indicates the position of the neuron in layer k.

The output of a generic kth layer $\mathbf{h}^{(k)} \in \mathbb{R}^M$ is calculated as the weighted sum of the outputs from the previous layer $\mathbf{h}^{(k-1)} \in \mathbb{R}^D$. Here, k denotes the index of the layer, and M and D are the dimensions of the output vectors for layers k and $k-1$, respectively. The weighted sum for layer k is calculated by

$$\mathbf{g}^{(k)} = \mathbf{W}^{(k)}\mathbf{h}^{(k-1)} + \mathbf{b}^{(k)}, \tag{5.6}$$

where $\mathbf{W}^{(k)} \in \mathbb{R}^{M \times D}$ is a matrix of weights between the layers $k-1$ and k. The vector $\mathbf{b}^{(k)} \in \mathbb{R}^M$ contains the bias terms for each node in layer k. The weighted sum is then followed by a nonlinear activation function $f()$

$$\mathbf{h}^{(k)} = f(\mathbf{g}^{(k)}), \tag{5.7}$$

producing the output of the layer k that is passed onto the next layer or out from the network as the output. In this work, hidden layers were connected by hyperbolic tangent sigmoid activation function $f(x) = 2/[1 + \exp(-2x)] - 1$. A single node with linear activation function was used for the output layer. Therefore, the output of a generic node in layer k can be written as

$$h_i^{(k)} = f\left(\sum_{j=1}^{N_{k-1}} w_{ij}^{(k)} h_j^{(k-1)} + b_i^{(k)}\right). \tag{5.8}$$

Here $w_{ij}^{(k)}$ are the weights between nodes i and j in layers k and $k-1$, respectively.
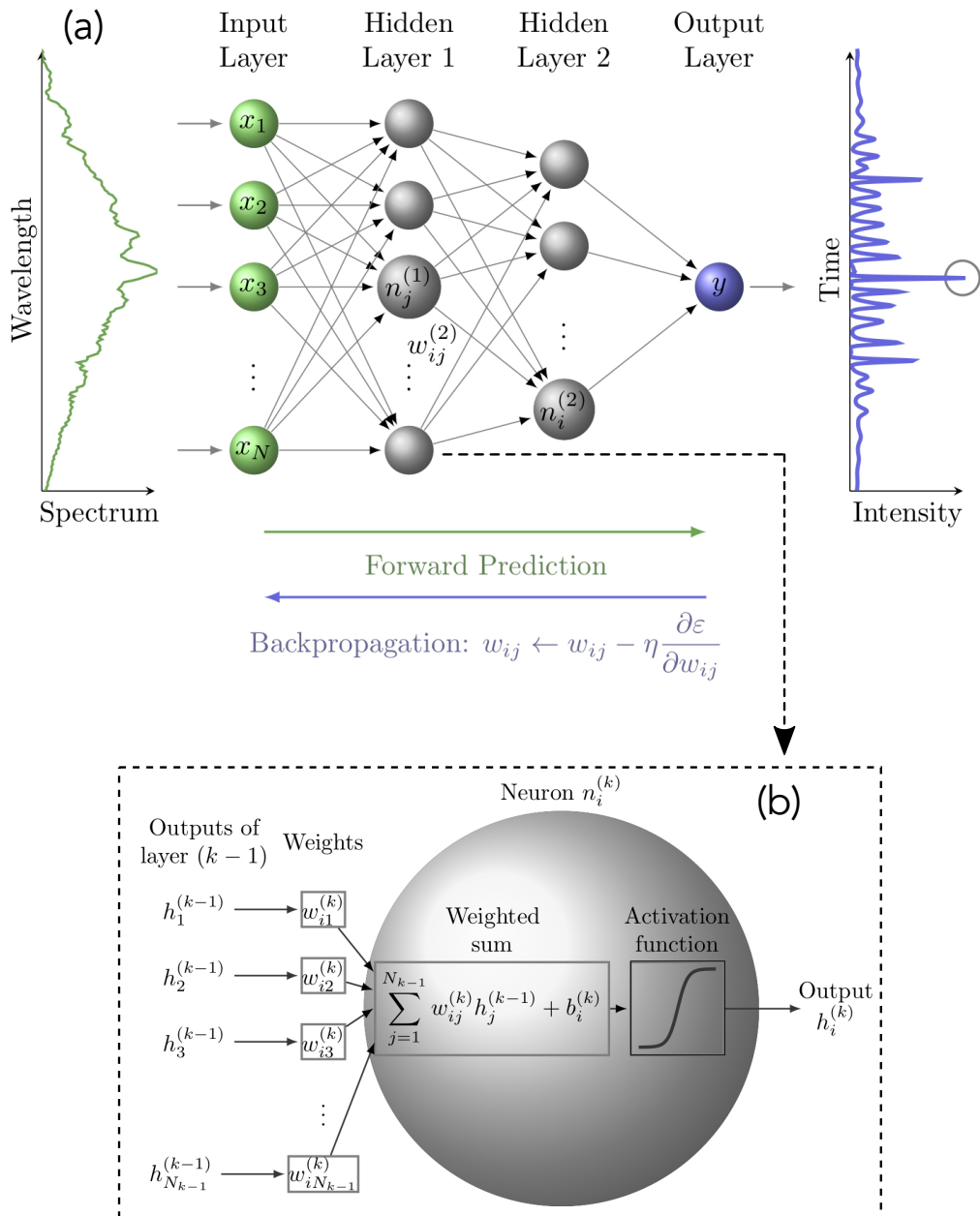
Figure 5.2. *Schematic of a feed-forward neural network. (A) Spectral intensity vector $\mathbf{X}_n = [x_1, x_2, \ldots, x_N]$ is fed to a feed-forward neural network consisting of 2 hidden layers and a single output neuron. Output of the network Y_n is the maximum instantaneous intensity. The arrows between the neurons correspond to the weights $w_{ij}^{(k)}$ that connect the neuron $n_i^{(k)}$ in layer k and neuron $n_j^{(k-1)}$ in layer $k-1$. During backpropagation, the weights are adjusted towards the negative gradient of the error function ε by a step size η. (B) Operation of a single neuron $n_i^{(k)}$ in layer k that is connected to N_{k-1} neurons in layer $k-1$. Taken from Ref. [48].*

Variable $b_i^{(k)}$ is the bias term associated with node $n_i^{(k)}$. The summation includes all the N_{k-1} nodes in layer $k-1$.

Backpropagation

The NN is trained in a process called backpropagation [78]. The goal of training is to minimize the prediction errors of the network. In this work, the error function (also called cost function) that is used for determining the accuracy of the predictions is the mean squared error function defined as

$$\varepsilon = \frac{1}{N}\sum_{n=1}^{N}(Y_n - \hat{Y}_n)^2, \tag{5.9}$$

where Y_n is the target value, $\hat{Y}_n$ is the prediction of the NN and N is the number of training samples.

The error of the prediction builds up from all of the weights and biases in the network. Both weights and biases are initially set randomly, and therefore large prediction errors are expected in the beginning of the training phase. The network variables are typically adjusted by a conjugate backpropagation algorithms such as stochastic gradient descent. The algorithm used in this work is the conjugate gradient backpropagation with Fletcher-Reeves updates [79,80]. In order to minimize the prediction error, a partial derivative respect to each weight and bias must be calculated, $\partial\varepsilon/\partial w_{ij}^{(k)}$ and $\partial\varepsilon/\partial b_i^{(k)}$. Let us consider the case of a single prediction for simplicity. The backpropagation algorithm is performed by averaging predictions from multiple samples, and the network variables are then adjusted. The predicted value is denoted as $\hat{y}_i$ and the correct, desired value as y_i. The error function is then $\varepsilon = (y_i - \hat{y}_i)^2$. The output of the network (layer K) is calculated as a weighted sum $g_i^{(K)} = w_{i1}^{(K)}h_1^{(K-1)} + w_{i2}^{(K)}h_2^{(K-1)} + \cdots + w_{iN_{K-1}}^{(K-1)}h_{N_{K-1}}^{(K-1)} + b_i^{(K)}$, followed by an activation function, as explained earlier, which produces the output (prediction) of the network $\hat{y}_i = h_i^{(K)} = f(g_i^{(K)})$.

To calculate the partial derivative of the error function respect to weight $w_{ij}^{(K)}$, the chain rule for derivatives is applied. This can be written as:

$$\frac{\partial\varepsilon}{\partial w_{ij}^{(K)}} = \frac{\partial g_i^{(K)}}{\partial w_{ij}^{(K)}}\frac{\partial h_i^{(K)}}{\partial g_i^{(K)}}\frac{\partial\varepsilon}{\partial h_i^{(K)}} = h_j^{(K-1)}f'(g_i^{(K)})\frac{\partial\varepsilon}{\partial h_i^{(K)}}. \tag{5.10}$$

The first term on the right-hand side, $h_j^{(K-1)}$, is derived from the derivative of the weighted sum $g_i^{(K)}$ respect to the weight $w_{ij}^{(K)}$, resulting output of the jth node in layer $K-1$. This value is known. The second term on the right-hand side, $f'(g_i^{(K)})$, is simply the derivative of the of the activation function $(\partial f(g_i^{(K)})/\partial g_i^{(K)})$, and it can be easily calculated since we know the equation for the activation function and the weighted sum. The last term of the chain rule depends on the layer we are interested in. For the output layer, $\partial\varepsilon/\partial h_i^{(K)} = 2(y_i - \hat{y}_i)$ in case of a mean squared error function. For a hidden layer, the chain rule must be extended further inside the network until the output layer is included to the chain, i.e. the error is backpropagated to each of the weights inside the network. Similar procedure is used for updating the bias terms.

The weights and biases are adjusted relative to their current state by taking partial derivatives from the error function. The "speed" of the adjustment can be chosen by the step size η, also known as the learning rate. The adjustment of the weight $w_{ij}^{(k)}$ is always towards the negative gradient of the error function, and the change is given by $\Delta w_{ij}^{(k)} = -\eta \partial \varepsilon / \partial w_{ij}^{(k)}$. After adjustment, more samples are fed to the network, and the weights and biases are again adjusted. Once all of the input and output pairs $(\mathbf{X}_n, Y_n)$ in the training set are passed through the network (forward prediction and backpropagation) one *epoch* has passed. Depending on the problem at hand, the network may be trained for hundreds or thousands of epochs until convergence of the error function is achieved. The network is then tested with a separate test or validation set that is not used in the training phase to get a realistic evaluation of the performance of the network. Potential issues in the training phase such as overlearning will be discussed further in the Section 5.3.

5.2 Unsupervised learning

Unlike supervised learning used for the neural network, unsupervised learning refers to label free machine learning methods where the data is not categorized or labeled by any additional information, for example, the maximum instantaneous intensity. The learning is thus more exploratory and the goal is to find inherent patterns, similarities or dissimilarities from the input data. Therefore, no "correct" labels or predictions exist.

An important, and actually one of the most popular subclass of unsupervised learning is the **cluster analysis**, or data clustering, where the goal is to find similarities (or dissimilarities) and natural grouping in data [81]. The goal in this work is to use a cluster analysis to classify the spectral intensity profiles into different clusters depending on their overall shape. Here, a variation of a well-known data clustering algorithm *k-means* [82], called *k-means++* [83], is used. The principle of k-means algorithm is to find k cluster means, referred as centroids, in a such manner that it minimizes the squared Euclidean distance between the data points and the closest centroid [81]. The main steps in the k-means algorithm are:

1. to (randomly) initialize the centroids
2. to classify the data points to the closest centroid
3. to compute the new k centroids as a mean of the samples classified to the cluster

Steps 2 and 3 are iterated until the centroids are stabilized, i.e. repeating steps 2 and 3 do not change the cluster populations.

5.3 Generalization and regularization

Generalization is one of the most important features of machine learning. It refers to a process where the machine is not only copying the predictions from the training data but it

can make accurate predictions from previously unseen data. The ability of performing well on new, unobserved data is referred as generalization, and the strategies to achieve it are collectively known as *regularization* [84].

Overlearning is a common problem in machine learning. One may notice that there is a large difference between the accuracy of the predictions for the training and test data sets. Poor generalization is often caused by overlearning where the ML model learns the training data too well and loses the ability to adapt to new data. On the other hand, the ML model may sometimes be too simple for complex problems and even the training data is not well learnt. This is known as *underlearning*. The tendency to over- or underlearning is determined by the *capacity* of the model [84]. The larger learning capacity, the bigger the chance of overlearning is. An easy way the understand the meaning of capacity for ML model is to consider a polynomial fit. The larger learning capacity, the higher the order of the polynomial is.

To tackle difficult ML problems, the network should have high capacity. *Dropout* [85] is an easy way to avoid overlearning. It is a process where some of the nodes are randomly dropped out during the forward prediction and backpropagation, resulting to a "thinned" network created from the original network. Training of the network is done by training large amount "thinned" networks, and in this way, overlearning can be avoided.

Perhaps an even more simple, yet not always easy, approach is to increase the amount of training data [84]. Large amount of training data increases the optimal capacity of ML model which can therefore learn more complex tasks. However, it must be noted that the increase of the amount of training data does not automatically increase the accuracy of the predictions. Equally important is the quality of the training data; it should represent the data correctly, including also extreme cases in the data, not average cases only.

Lastly, one may regulate the ML model weights $\mathbf{w}$. One of the most popular methods is to add a squared penalty for the weights. This promotes the use of small coefficients and reduces the probability for overlearning. Let us consider a linear regression problem

$$y_n = \mathbf{w}^T \mathbf{x}_n + e_n, \tag{5.11}$$

where $y_n \in \mathbb{R}$ are the outputs, $\mathbf{x}_n \in \mathbb{R}^P$ are the inputs, $\mathbf{w} \in \mathbb{R}^P$ are the weights and $e_n \in \mathbb{R}$ are the errors to the linear model. Instead of finding the least squares solution for the regression problem as done in Section 5.1.1, a squared penalty is added

$$\underset{\mathbf{w}}{\text{minimize}} \left(\sum_{n=1}^{N} (y_n - \mathbf{w}^T \mathbf{x}_n)^2 + \lambda \mathbf{w}^T \mathbf{w} \right), \tag{5.12}$$

where λ is a positive regularization constant and N is the number of samples. This regularization method is known as the Tikhonov or L$_2$ regularization. Similarly, one may

use an absolute penalty

$$\underset{\mathbf{w}}{\text{minimize}} \left(\sum_{n=1}^{N} (y_n - \mathbf{w}^T \mathbf{x}_n)^2 + \lambda \left\| \mathbf{w} \right\|_1 \right), \tag{5.13}$$

where $\left\| \mathbf{w} \right\|_1 = \sum_{p=1}^{P} |w_p|$. This is known as the LASSO or L_1 regularization.

6 APPLICATION OF MACHINE LEARNING TO MODULATION INSTABILITY & SUPERCONTINUUM GENERATION

In this work, machine learning was applied to analyse temporal properties of modulation instability and rogue solitons in supercontinuum generation from solely spectral intensity data. For the modulation instability case, the simulation-trained neural network was also applied to single-shot experimental spectra.

The following sections describe the main results and parameters used in the simulations and experiments. Section 6.1 consists of results on the maximum temporal intensity predictions from MI numerical simulations, and Section 6.2 focuses on the application of a numerically trained neural network to experimental MI spectra. Section 6.3 shows results from highly incoherent SC simulations and predictions of the associated rogue soliton peak power and temporal shift. Finally, predictions of the maximum breather or soliton peak power in the entire SC field are also analysed.

6.1 Predicting extreme events from modulation instability spectra

The numerical simulations of MI are performed using the generalized NLSE, including the Raman effect, self-steening and higher-order dispersion. Hyperbolic secant pulses with 3 ps pulse duration (FWHM), 175 W peak power and 825 nm central wavelength are injected into the anomalous dispersion regime of a 68 cm long photonic crystal fiber (PCF). The dispersion coefficients are shown in Table 6.1 and they are extracted from the data provided by the manufacturer (NKT Photonics NL-PM-750). The nonlinear coefficient and zero-dispersion wavelength of the fiber are γ = 0.1 W^{-1}m^{-1} and 750 nm, respectively. These parameters correspond to those of the experiments to which the trained NN will be applied.

The simulations use 4096 spectral/temporal grid points with a temporal window of 12 ps. The noise is added in the frequency domain by adding one photon with random phase to each frequency bin to trigger the MI noise-seeding. To compensate the input pulse shot-to-shot fluctuations due to the use of acousto-optic modulator in the experimental part (see Section 6.2), $\pm$ 5% variation was added to the peak power. A total number of

Table 6.1. *Fiber dispersion coefficients for modulation instability at 825 nm for NKT Photonics NL-PM-750.*

$$\beta_2 = \text{-}1.03 \times 10^{-26} \text{ s}^2 \text{ m}^{-1}$$
$$\beta_3 = 4.74 \times 10^{-41} \text{ s}^3 \text{ m}^{-1}$$
$$\beta_4 = 2.35 \times 10^{-56} \text{ s}^4 \text{ m}^{-1}$$
$$\beta_5 = \text{-}1.17 \times 10^{-70} \text{ s}^5 \text{ m}^{-1}$$
$$\beta_6 = \text{-}9.07 \times 10^{-85} \text{ s}^6 \text{ m}^{-1}$$

50,000 simulations were carried out. An example of simulated evolution and few selected realizations along with the peak power distribution are shown in Figure 3.5. Since the goal is to apply the NN to experimental MI data, the simulation data is processed to match the spectral resolution of 1 nm and dynamic range of 60 dB obtained in the experiments introduced in the next section. The input for the neural network (NN), i.e. the single-shot spectral intensity, consists of 121 uniformly distributed intensity bins. The NN consists of two hidden layers, with 30 and 10 hidden nodes each, and a single output node yielding the maximum temporal intensity associated with the input spectrum fed to the NN.

The results of the NN predictions for the simulation data are shown in Figure 6.1. The NN is trained with a training set of 30,000 distinct realizations, and the results are obtained from a testing set of 20,000 realizations that were not used in the training step. Figure 6.1A shows the NN predictions for the maximum temporal intensity when single-shot spectra are fed with a dynamic range of 60 dB. Notice the logarithmic scale for better visualization. We can see excellent agreement between the predicted and actual values from the simulations around the one-to-one relation ($x = y$) shown as the white dashed line. The variable ρ is the Pearson correlation coefficient, and a correlation of $\rho = 0.92$ is achieved by the network confirming the good visual agreement between the predicted and desired values. Figure 6.1B shows the probability density function of the ML predictions (red) and that from the actual values taken from the simulations (blue). The PDFs show excellent agreement, including for the long-extending tail that contains temporal structures of extreme intensity. It is rather remarkable that the statistics that span over two orders of magnitude are well predicted.

To examine the effect of a reduced dynamic range for the spectral intensity used as the network input, simulations with a dynamic range of 25 dB were used for the predictions, which corresponds to a typical fiber dispersive Fourier transform setup (see inset in Figure 6.4). The results are shown in Figure 6.1C and D similarly to the higher dynamic range in A and B. One can observe a clear drop in the accuracy of the predictions. The correlation is now reduced to $\rho = 0.69$, and the PDF of the predictions fails to reproduce the slope of the distribution tail, such that a high dynamic range is essential for accurate predictions and capturing the rare high-intensity events in MI.

In addition to the supervised NN predictions, unsupervised k-means clustering analysis was applied to further study the MI process. The goal in this case was to classify single-shot MI spectra into various dynamic classes that can be related to the emergence of
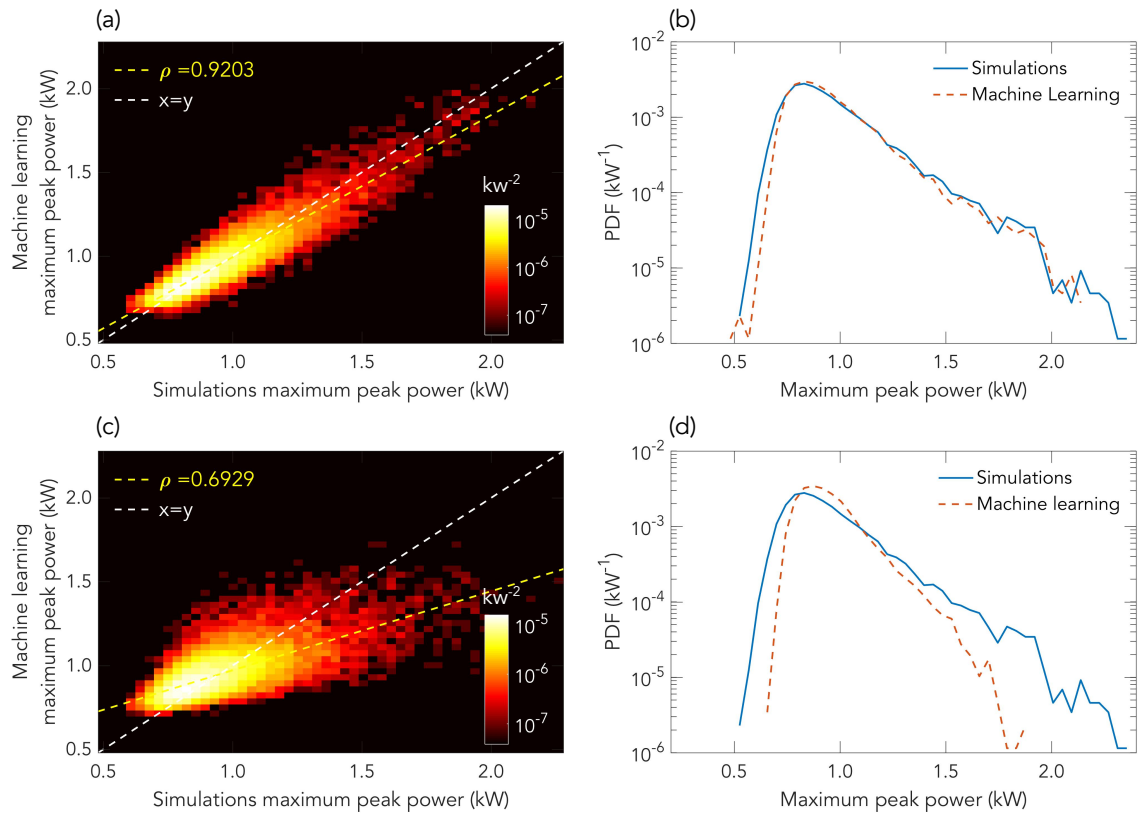
Figure 6.1. *Results of the supervised learning on simulated modulation instability spectra. (A) and (C) show the predicted peak powers respect to the actual values from the simulations for dynamic ranges of 60 dB and 25 dB, respectively. The figures show the predictions as a logarithmic histogram. Variable ρ is the Pearson correlation coefficient, and the yellow dashed line shows the linear fit to the predictions along with a perfect correlation $x = y$ in white. (B) and (D) show the probability density function (PDF) of the machine learning predictions in red and the actual values from the simulations in blue for 60 dB and 25 dB dynamic ranges, respectively. Taken from Ref. [48].*

specific breather structures in the time-domain. It is expected that the spectral clusters close to the mean spectrum would correspond in the time-domain to the emergence of Akhmediev breathers at the maximum MI gain, while breathers with spectral clusters with the broadest spectral bandwidth would correspond to extreme confinement in the time-domain (i.e. Peregrine solitons and/or breather collisions). Despite the complex and low-amplitude fine structures, the clustering analysis can very well separate the spectral profiles with respect to the emergence of specific breather subclasses in the time-domain, as shown in Figure 6.2. An ensemble of 50,000 simulated MI spectra were used for the k-means algorithm with 9 clusters (use of 5-30 clusters yielded similar results). Figures 6.2A and B show the spectral cluster with average spectral shape and the corresponding highest intensity breather structure in the time-domain, respectively. The blue curves show the individual spectra and time-domain profiles associated with the cluster, and the black curve is the cluster mean (both in spectrum and time). As expected, the extracted time-domain maxima match well with the analytical Akhmediev breather plotted in yellow in Figure 6.2B. In fact, the clusters close to the mean spectrum have the largest population and the
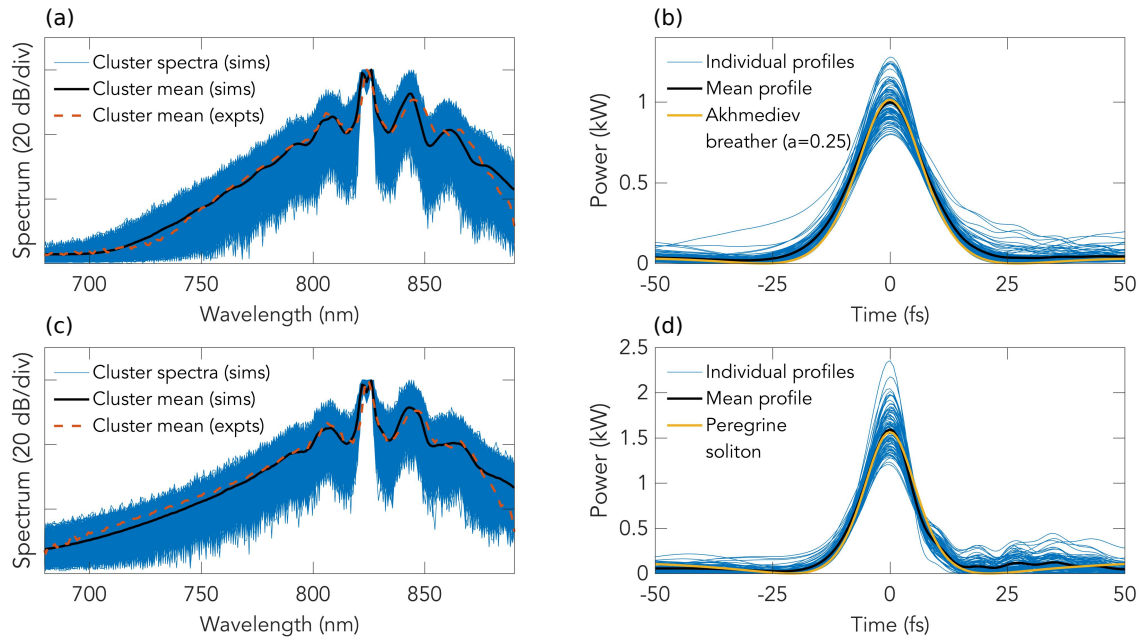
Figure 6.2. *Results of the unsupervised clustering analysis of 50,000 modulation instability spectra. (A) shows the cluster closest to the mean spectrum. The blue curves show the individual simulated spectra in the "average" cluster along with the mean in black and mean of the "average" experimental cluster. The corresponding temporal intensity profiles (blue) of the maximum breather is shown in (B) along with the mean (black) and analytical Akhmediev breather (yellow). (C) shows the cluster with the broadest spectral bandwidth and (D) the corresponding temporal profiles along with the analytical Peregrine soliton (yellow). Taken from Ref. [48].*

time-domain maxima are close to the maximum of the peak power PDF (see Figure 3.5). In Figures 6.2A, the "average" cluster of the experimental spectra (see Section 6.2) is also shown by a red dashed line and found to be in good agreement with the one from the numerical simulations.

Similarly, Figures 6.2C and D show the spectral cluster and associated time-domain maximum peaks for the broadest spectral bandwidth. The yellow curve in Figure 6.2D plots the analytical Peregrine soliton and we can observe good agreement with the extracted cluster mean. One can also see few extremely high peak powers that are the result of first-order breather collisions. Even though the clustering analysis does not directly yield the temporal intensity maximum, it offers an easy tool for identifying the spectra associated with the emergence of time-domain extreme events in MI.

6.2 Application to experimental modulation instability spectra

Time-domain information about ultra-fast instabilities is generally difficult obtain experimentally due to the short time-scale of emerging structures, and as shown in this section, a simulation-trained neural network may offer a convenient solution to obtain time-domain information from only single-shot spectra. To apply the NN to the experimental noise-
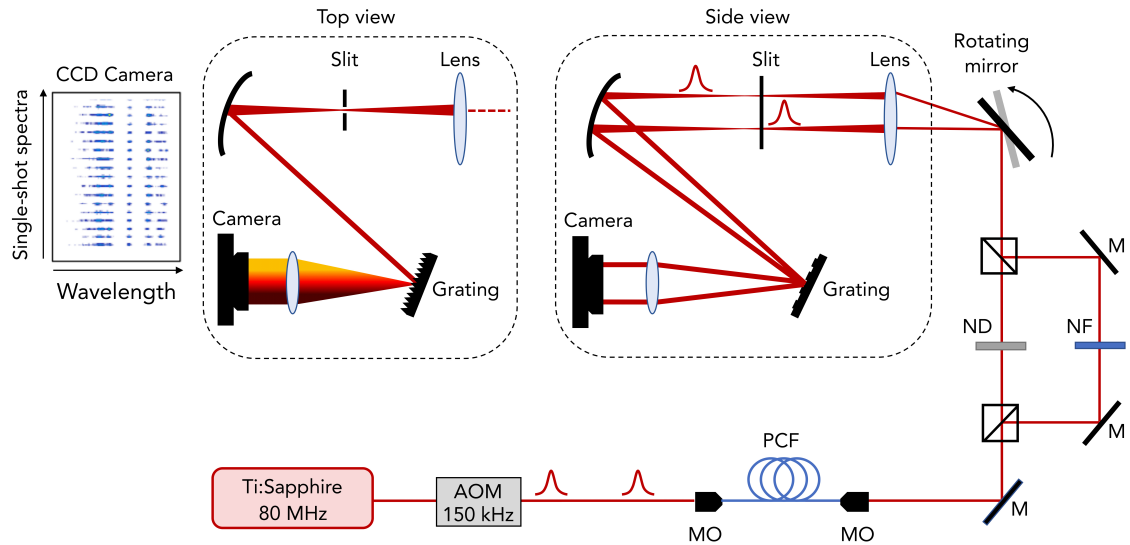
Figure 6.3. *Experimental setup. Ti:Sapphire: Titanium-sapphire mode-locked laser, AOM: acousto-optic modulator, MO: microscope objective, PCF: photonic crystal fiber, M: mirror, ND: neutral density filter, NF: notch filter, CCD: charge-coupled device. Top and side views of a Czerny–Turner spectrograph are shown. Taken from Ref. [48].*

seeded MI spectra, high dynamic range spectral intensity measurements are required, as discussed in the previous section.

The experimental setup used in this work is shown in Figure 6.3. Similarly to the numerical simulations, the noisy MI field is generated by injecting 3 ps pulses (FWHM) centered at 825 nm from Titanium-sapphire 80 MHz mode-locked laser into the anomalous dispersion regime of a PCF (NKT Photonics NL-PM-750) with the ZDW at 750 nm. The repetition rate of the laser is reduced to 150 kHz using an acousto-optic modulator. A rapidly-rotating mirror (200 rev/min) is used for focusing (lens with focal length of $f = 150$ mm) consecutive pulses to different vertical positions of a Czerny-Turner spectrograph using a grating with 300 lines/mm and 500 nm blaze (ThorLabs GR25-0305). The spectra are captured using highly sensitive electron-multiplying charged-coupled device (EMCCD) camera (Andor Ixon 3) with a spectral resolution of 1.1 nm.

This setup can capture single-shot MI spectra with a dynamic range of ca. 48 dB. To further increase the dynamic range, spectral windowing and differential attenuation are used. The beam path is divided into two arms where in one the high intensity central spectral region (20 nm bandwidth) is attenuated by a 40 dB notch filter and in the other a neutral density filter is used. Essentially, one arm is used for measuring the low intensity spectral wings and the other arm measures the high intensity central region, respectively. The arms are
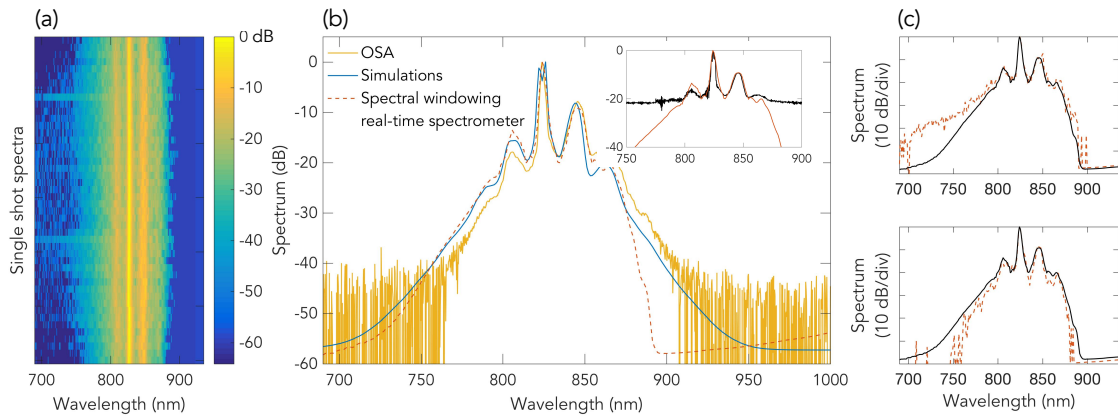
Figure 6.4. *Recorded experimental single-shot modulation instability spectra. (A) shows one sweep of the mirror on the CCD camera with about 50 recorded spectra. (B) shows the comparison of the average spectra from optical spectrum analyser (yellow), simulations (blue) and the real-time technique (red). The inset shows comparison between fiber-DFT (black) and the real-time technique (red). (C) shows two selected single-shot spectra in red compared with the average spectrum of the real-time measurements (black). Taken from Ref. [48].*

combined with a delay of 200 ps to avoid spectral interference and the full spectrum is obtained by post-processing. This method yields an effective dynamic range of more than 50 dB, a 3 orders of magnitude improvement compared to conventional fiber dispersive Fourier transform method. Figure 6.4 shows examples of the recorded spectra. In (A), one sweep of the rotating mirror is shown with around 50 recorded spectra where one can clearly see the large shot-to-shot fluctuations associated with noisy MI. Figure 6.4B shows the comparison between average measurements using OSA (yellow), the real-time spectral windowing method introduced above from 3,000 measured spectra (red) and that from 50,000 simulations (blue). The inset shows comparison between the real-time spectral windowing method (red) and convention fiber dispersive Fourier transform method (black). One can see how the dynamic range of the real-time measurement approach is indeed exceeding 50 dB. Figure 6.4C highlights the large spectral fluctuations of the MI field from selected realizations (red) against the mean spectrum (black), both measured using the real-time spectral windowing method.

We can see that the measured spectra are in good agreement with the simulations, excluding the discrepancy on the long-wavelength edge due to the reduced throughput efficiency of the measurement setup. The same simulation-trained neural network was then applied to the experimental MI spectra, and the results are shown Figure 6.5. The PDF of the maximum peak power predictions from 3,000 experimental MI spectra are plotted in red while that from the simulations is plotted in blue. One can see excellent agreement between the simulations and the predicted PDF from the experiments. Especially significant is the good agreement on the long tail of the distribution that contains the emergence of structures with high intensity. This shows that the NN trained on the simulation data can therefore be used for predicting the temporal domain properties using only spectral intensity measurements that are relative easy to perform compared to real-time time
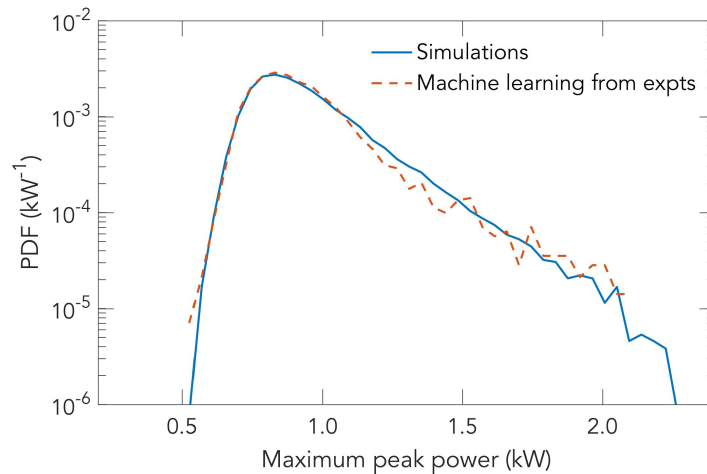
Figure 6.5. *Probability density function (PDF) of the peak power predictions from experimental modulation instability spectra in red compared with the PDF extracted from the numerical simulations in blue. Taken from Ref. [48].*

domain measurements or full-field measurements where the phase of the electric field is measured as well.

In addition to the supervised NN-based predictions, k-means clustering analysis with the same amounts of 9 clusters was applied to an ensemble of 3,000 experimental noisy MI spectra. The results are plotted in Figure 6.2 along with the results obtained from the numerical simulations. Figure 6.2A shows the cluster closest to the average spectrum while (C) shows the cluster associated with the broadest spectral bandwidth. The experimental clusters are shown with red dashed lines and one can observe the close resemblance to the simulation results. Again, the majority of the clusters are found to have a spectral profile close to the mean spectrum while the population of the clusters with increased spectral bandwidth is decreased. Similarly to the results from the simulations, the clustering analysis can be used for fast classification to identify the emergence of extreme events in MI.

6.3 Analysis of rogue solitons in supercontinuum generation

To study extreme events in the supercontinuum generation, an ensemble of 30,000 simulations was generated from the generalized NLSE. Hyperbolic secant pulses with 2 ps pulse duration (FWHM), 400 W peak power and 810 nm central wavelength are injected into the anomalous dispersion regime of an 85 cm long PCF. The fiber that is used here is the same as that used for the MI simulations. The dispersion coefficients at the pump wavelength are shown in Table 6.2. The nonlinear coefficient of the fiber is $\gamma = 0.1$ $W^{-1}m^{-1}$. Compared to the simulations of MI, the pulse duration is slightly reduced and the peak power is increased. The pump wavelength is also shifted towards the ZDW

Table 6.2. *Fiber dispersion coefficients for supercontinuum generation at 810 nm for NKT Photonics NL-PM-750.*

$$\beta_2 = -1.24 \times 10^{-26} \text{ s}^2 \text{ m}^{-1}$$
$$\beta_3 = 8.94 \times 10^{-41} \text{ s}^3 \text{ m}^{-1}$$
$$\beta_4 = -2.54 \times 10^{-56} \text{ s}^4 \text{ m}^{-1}$$
$$\beta_5 = -7.01 \times 10^{-70} \text{ s}^5 \text{ m}^{-1}$$
$$\beta_6 = 2.28 \times 10^{-84} \text{ s}^6 \text{ m}^{-1}$$
$$\beta_7 = -2.21 \times 10^{-99} \text{ s}^7 \text{ m}^{-1}$$

of the fiber to enhance soliton dynamics after the initial stage of MI and efficiently seed the SC generation process.

The simulations use 16384 spectral/temporal grid points with a temporal window of 20 ps. The noise is added in a similar manner to the MI simulations to trigger the initial MI state of the SC generation. An example of simulated propagation dynamics can be seen in Figure 2.6. The simulation data is processed to a spectral resolution of 1 nm and dynamic range around 60 dB (even though no experiments have yet been conducted in this case). To estimate the rogue soliton peak power, a super Gaussian long-pass filter at around 950 nm is used to capture the highly red-shifted spectral components. The rogue soliton peak power is then obtained by inverse Fourier transform of the highly red-shifted spectral field components. The input for the neural network is the single-shot spectral intensity of the entire SC in this case, consisting of 801 uniformly distributed intensity bins. The size of the NN is increased to compensate the increased number of spectral points to 80 and 20 hidden nodes for the first and second hidden layer, respectively. A single output node is used to yield the rogue soliton peak power and temporal shift associated with a given SC spectrum. Two independently trained NNs are used for the peak power and Raman-induced temporal shift.

The results of the rogue soliton predictions are shown in Figure 6.6. In (A), the predicted RS peak power is plotted against the actual value from the simulations. We can see excellent visual clustering around the perfect $x = y$ relation. Also the Pearson correlation coefficient gives nearly perfect correlation of $\rho = 0.99$ confirming the very high accuracy of the NN predictions. It must be noted here that there is some correlation between the RS peak power and the long-wavelength edge of the supercontinuum. As the long-pass filter is shifted towards longer wavelengths, the spectral bandwidth of the long-wavelength edge correlates more with the RS peak power. However, when the long-pass filter is placed at around 950 nm, the spectral bandwidth is only weakly correlated with the rogue soliton peak power. Altogether, the accuracy of the neural network predictions always clearly exceeds the accuracy of the predictions based on a single parameter such as spectral bandwidth regardless of the filter position. The overall shape and the fine details of the spectrum carry significant information about the temporal properties that cannot be estimated by conventional methods. Figure 6.6B shows the predictions of the temporal shift of the rogue soliton. One can see that the temporal shift is also well predicted with a
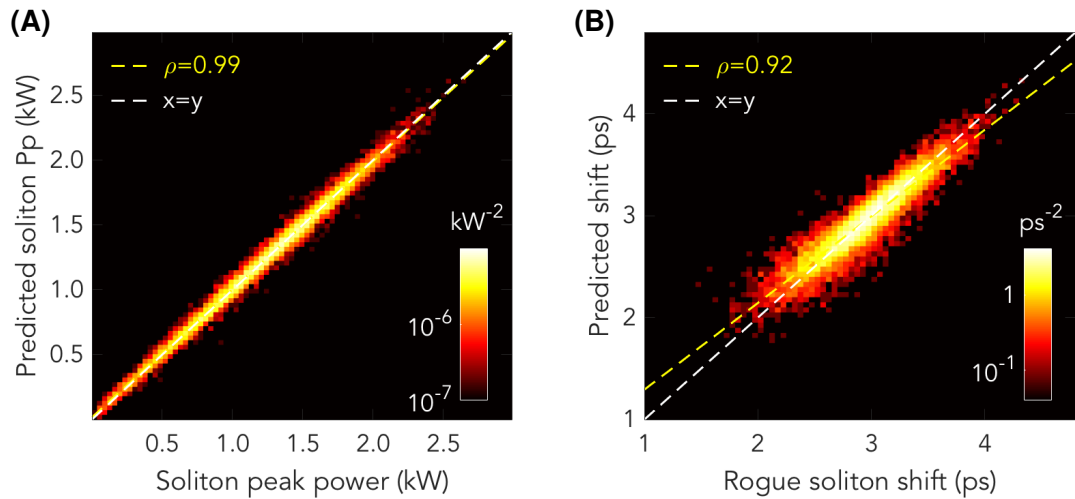
Figure 6.6. *Results of the rogue soliton peak power (Pp) and temporal shift predictions. (A) and (B) show the predicted soliton peak powers and temporal shifts respect to the actual values from the simulations, respectively.*
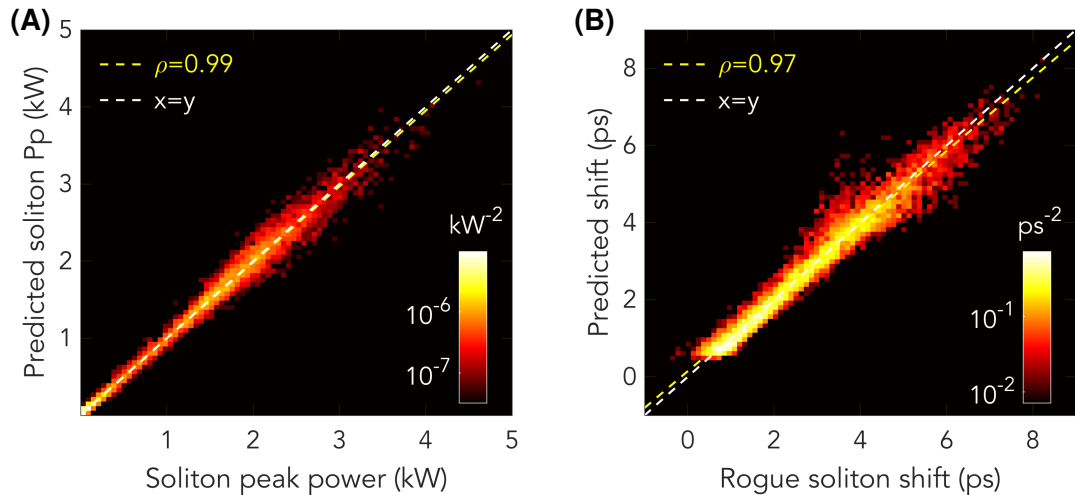

Figure 6.7. *Results of the rogue soliton peak power (Pp) and temporal shift predictions with ± 50% variation to the input pulse duration and peak power. (A) and (B) show the predicted soliton peak powers and temporal shifts respect to the actual values from the simulations, respectively.*

correlation of ρ = 0.92.

As the rogue soliton peak power and the temporal shift are very well predicted, the next step is towards more general case. To this end, a large ± 50% variation to the input pulse duration of 2 ps and peak power of 400 W was induced. This results into diverse dynamics from pure MI to broad octave-spanning SC. Other simulation parameters are kept identical. The results shown in Figure 6.7 demonstrate the ability of the network to still predict very well the most red-shifted soliton peak power with a correlation of ρ = 0.99 between the predicted and actual values. In the case of the temporal shift, the network also performs well, and the accuracy of the predictions is actually increased compared to Figure 6.6B where no variation is added to the input pulse parameters. The Pearson
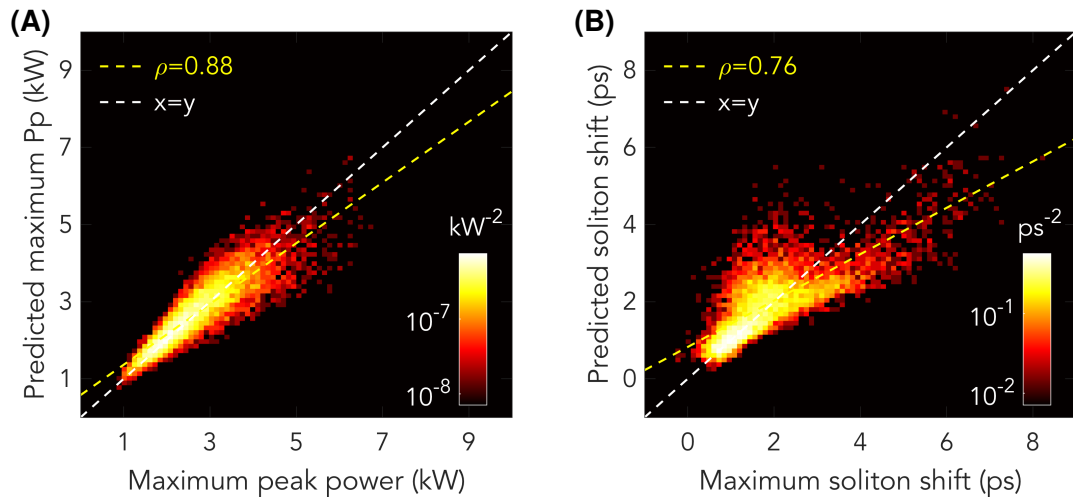
Figure 6.8. *Results of the maximum soliton peak power (Pp) and temporal shift predictions with ± 50% variation to the input pulse duration and peak power. (A) and (B) show the predicted soliton peak powers and temporal shifts respect to the actual values from the simulations, respectively.*

correlation coefficient is increased up to $\rho = 0.97$. In this case, some of the "rogue solitons" are observed in the pure MI field (to close zero temporal shift) despite the use of the super Gaussian long-pass filter. This highlights the fact that for lower input peak powers no solitons emerge from the MI stage and the Raman-induced temporal shift is then very weak. On the other hand, some realizations (corresponding to high input peak power) exhibit extreme red-shift and large temporal shift (up to 10 ps). Despite the large variety of the underlying dynamics, the statistics of rogue solitons are predicted with excellent accuracy.

To further generalize the analysis of extreme events in supercontinuum generation, the long-pass filter was removed. Now the contribution from the MI field becomes more significant as it contains high intensity structures which are mostly filtered out. Figure 6.8 shows the predictions of the total temporal intensity maximum. The breather or soliton with the highest peak power can be found either in the noisy MI field or in a fully developed SC a highly red-shifted rogue soliton. In Figure 6.8A, the predictions for the maximum peak power are shown. One can see how the accuracy is reduced with a correlation of $\rho = 0.88$. The majority of the peak powers are predicted fairly well but the extremely high peak powers (that are the result of a breather collision in the MI field) are not well captured by the neural network. Similarly, Figure 6.8B shows the predictions of the temporal shift of the soliton associated with the maximum peak power. The contribution from the MI field also significantly reduces the prediction accuracy of the temporal location. Perhaps a more complex NN architecture or a combination of different ML algorithms are required to accurately estimate the full temporal characteristics of an incoherent supercontinuum.

7 CONCLUSION

Machine learning is becoming an essential part of our daily life and the use of neural networks has become a central in many commercial products for data analysis, in particular for image and voice recognition. The use of different machine learning techniques has also started to spread in many research fields. This thesis belongs to that trend. The use of machine learning in ultrafast optics is still at its infancy but an increasing number of papers have recently appeared showing the potential of machine learning techniques for obtaining additional information from only partial measurements.

In this thesis, the basic concepts of linear and nonlinear optics were explained in Chapter 2 along with the numerical methods used for modeling the propagation of light. The linear propagation of light is the starting point of all optical phenomena. The increase of the light field intensity may however affect the response of the propagation medium leading to the generation of new spectral components. This thesis explained the most common nonlinear effects that take place in a third-order nonlinear material such as silica that is commonly used to manufacture optical fibers. The combination of linear and nonlinear effects may lead to interesting phenomena such as modulation instability and the generation of optical solitons. Modulation instability was described in details in Chapter 3. Modulation instability may be used as a seed for a extreme broadband light sources known as supercontinuum, and due to the high sensitivity of MI to changes in the initial conditions, i.e. background noise, it may lead to the (rare) formation of extreme events. The different dynamical regimes of supercontinuum generation were explained in Chapter 4 with a focus on the highly incoherent long-pulse supercontinua where soliton dynamics play a dominant role. The emergence and statistics of extremely red-shifted rogue solitons were discussed. Chapter 5 described the most common machine learning techniques with a focus on feed-forward neural networks that were used in this thesis.

The first part of this thesis concentrated on modulation instability dynamics. The noise-seeded MI leads to the emergence of high intensity temporal structures that exhibit highly chaotic dynamics. The extreme temporal confinement of light may be harmful for sensitive applications and estimating the peak intensity of these structures is thus of significant interest. However, conventional measurement techniques cannot currently measure these ultrafast dynamics due their extremely short duration. In this thesis, the method for estimating the temporal intensity maximum of MI utilized a feed-forward neural network to establish a nonlinear transfer function that relates the spectral intensity of MI with the time-domain intensity maximum. The neural network is able to accurately predict the maximum

temporal intensity based only on spectral intensity profile without any phase information which is a significant advantage from an experimental point of view since the single-shot spectral intensity measurements can be routinely conducted. Remarkably, the rare and high intensity rogue wave events that occur in the chaotic MI field are well predicted by the neural network. It was also noticed that the dynamic range of the measurements plays a crucial role in the accuracy of the predictions. Another significant contribution of this work is the demonstration of the use of unsupervised clustering analysis to identify specific subclasses of temporal structures associated with a chaotic MI field.

Experimental MI measurements were subsequently conducted using a specifically developed approach allowing for capturing single-shot spectra with a 60 dB dynamic range, exceeding the dynamic range of a typical fiber-based dispersive Fourier transform method by nearly four orders of magnitude. The high dynamic range was critical since the low intensity spectral wings carry important information about the temporal characteristics of the noisy MI field. The experimental approach introduced in this thesis is straightforward and can be applied to all wavelength ranges where suitable spectrometers exist, and it can be applied to study various ultrafast optical phenomena. The high dynamic range then allows to use the simulation-trained neural network to yield accurate temporal statistics from the experimental MI spectra.

The last part of this thesis focuses on supercontinuum generation in the long pulse regime. Supercontinuum developing from an initial stage of chaotic MI which may lead to the formation of optical solitons that are shifted towards long wavelengths due to the Raman effect and separate in time from the MI field. Occasionally, extremely red-shifted rogue solitons are created as a result of a collision between two solitons. In order to estimate the temporal characteristics of these rogue solitons, a neural network was used to relate the SC spectrum with the rogue soliton peak power and associated temporal shift. Again, no phase information was used. The predictions made by the neural network are very accurate, and even adding large variation to the input pulse duration and peak power the accuracy of the predictions was found to be preserved.

In addition to rogue solitons, this thesis includes predictions about the temporal intensity maximum in long pulse SC. Estimating the temporal intensity maximum of full SC field includes the contributions from both the noisy MI field and the rogue solitons which makes it a complex task. However, the results show potential for more general temporal characterization of SC based only on spectral intensity measurements and may allow for more comprehensive control of both spectral and temporal properties of SC. More generally, machine learning can bring new insight into the study and prediction of nonlinear dynamics where direct time-domain measurements are difficult.

REFERENCES

[1] M. I. Jordan and T. M. Mitchell. Machine learning: trends, perspectives, and prospects. *Science* 349.6245 (2015), 255–260.

[2] G. Agrawal. *Nonlinear fiber optics*. 5th. Academic press, 2013.

[3] A. Hasegawa. Generation of a train of soliton pulses by induced modulational instability in optical fibers. *Optics letters* 9.7 (1984), 288–290.

[4] R. W. Boyd. *Nonlinear optics*. 3rd. Academic press, 2008.

[5] J. M. Dudley, F. Dias, M. Erkintalo and G. Genty. Instabilities, breathers and rogue waves in optics. *Nature Photonics* 8.10 (2014), 755.

[6] V. E. Zakharov, A. I. Dyachenko and A. O. Prokofiev. Freak waves as nonlinear stage of Stokes wave modulation instability. *European Journal of Mechanics - B/Fluids* 25.5 (2006), 677–692.

[7] J. M. Dudley, G. Genty and S. Coen. Supercontinuum generation in photonic crystal fiber. *Reviews of Modern Physics* 78.4 (2006), 1135.

[8] C. Mahnke and F. Mitschke. Possibility of an Akhmediev breather decaying into solitons. *Physical Review A* 85.3 (2012), 033808.

[9] F. M. Mitschke and L. F. Mollenauer. Discovery of the soliton self-frequency shift. *Optics Letters* 11.10 (1986), 659–661.

[10] D. R. Solli, C. Ropers, P. Koonath and B. Jalali. Optical rogue waves. *Nature* 450.7172 (2007), 1054.

[11] M. Närhi, B. Wetzel, C. Billet, S. Toenger, T. Sylvestre, J.-M. Merolla, R. Morandotti, F. Dias, G. Genty and J. M. Dudley. Real-time measurements of spontaneous breathers and rogue wave events in optical fibre modulation instability. *Nature Communications* 7 (2016), 13675.

[12] G. Agrawal. *Fiber-optic communication systems*. 3rd. John Wiley & Sons, Inc., 2002.

[13] D. Marcuse. *Light transmission optics*. 2nd. Van Nostrand Reinhold New York, 1972.

[14] I. H. Malitson. Interspecimen comparison of the refractive index of fused silica. *Josa* 55.10 (1965), 1205–1209.

[15] G. Genty, M. Närhi, C. Amiot and M. Jacquet. Supercontinuum generation in optical fibers. *Frontiers in Modern Optics* 190 (2016), 233.

[16] J. C. Knight, T. A. Birks, P. S. J. Russell and D. M. Atkin. All-silica single-mode optical fiber with photonic crystal cladding. *Optics Letters* 21.19 (1996), 1547–1549.

[17] J. K. Ranka, R. S. Windeler and A. J. Stentz. Visible continuum generation in air–silica microstructure optical fibers with anomalous dispersion at 800 nm. *Optics Letters* 25.1 (2000), 25–27.

[18] NKT Photonics. *"Photonic crystal fibers"*. `https://www.nktphotonics.com`. Online; accessed 28 April 2019. n.d.

[19] J. S. Russell. Report on waves. *14th meeting of the British Association for the Advancement of Science*. Vol. 311. 390. 1844, 1844.

[20] C. V. Raman. A new radiation. *Indian Journal of Physics* 2 (1928), 387–398.

[21] R. H. Stolen and E. P. Ippen. Raman gain in glass optical waveguides. *Applied Physics Letters* 22.6 (1973), 276–278.

[22] K. J. Blow and D. Wood. Theoretical description of transient stimulated Raman scattering in optical fibers. *IEEE Journal of Quantum Electronics* 25.12 (1989), 2665–2673.

[23] T. Brabec and F. Krausz. Nonlinear optical pulse propagation in the single-cycle regime. *Physical Review Letters* 78.17 (1997), 3282.

[24] R. H. Stolen, J. P. Gordon, W. Tomlinson and H. A. Haus. Raman response function of silica-core fibers. *JOSA B* 6.6 (1989), 1159–1166.

[25] V. E. Zakharov and L. A. Ostrovsky. Modulation instability: the beginning. *Physica D: Nonlinear Phenomena* 238.5 (2009), 540–548.

[26] V. E. Zakharov. Stability of periodic waves of finite amplitude on the surface of a deep fluid. *Journal of Applied Mechanics and Technical Physics* 9.2 (1968), 190–194.

[27] T. B. Benjamin and J. E. Feir. The disintegration of wave trains on deep water Part 1. Theory. *Journal of Fluid Mechanics* 27.3 (1967), 417–430.

[28] T. Taniuti and H. Washimi. Self-trapping and instability of hydromagnetic waves along the magnetic field in a cold plasma. *Physical Review Letters* 21.4 (1968), 209.

[29] W.-M. Liu, B. Wu and Q. Niu. Nonlinear effects in interference of Bose-Einstein condensates. *Physical Review Letters* 84.11 (2000), 2294.

[30] M. Närhi. *Measurements of noise-seeded dynamics in nonlinear fiber optics*. English. Tampere University of Technology. Publication. Tampere University of Technology, Nov. 2017. ISBN: 978-952-15-4051-6.

[31] E. Fermi, P. Pasta, S. Ulam and M. Tsingou. *Studies of the nonlinear problems*. Tech. rep. Los Alamos Scientific Lab., N. Mex., 1955.

[32] N. Akhmediev, A. Ankiewicz, J. M. Soto-Crespo and J. M. Dudley. Universal tri-angular spectra in parametrically-driven systems. *Physics Letters A* 375.3 (2011), 775–779.

[33] D. R. Solli, G. Herink, B. Jalali and C. Ropers. Fluctuations and correlations in modulation instability. *Nature Photonics* 6.7 (2012), 463.

[34] A. Shabat and V. Zakharov. Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media. *Soviet Physics JETP* 34.1 (1972), 62.

[35] M. J. Ablowitz and H. Segur. *Solitons and the inverse scattering transform*. Vol. 4. Siam, 1981.

[36] N. N. Akhmediev and A. Ankiewicz. *Solitons: nonlinear pulses and beams*. Chapman & Hall, 1997.

[37] N. N. Akhmediev and V. I. Korneev. Modulation instability and periodic solutions of the nonlinear Schrödinger equation. *Theoretical and Mathematical Physics* 69.2 (1986), 1089–1093.

[38] D. H. Peregrine. Water waves, nonlinear Schrödinger equations and their solutions. *The ANZIAM Journal* 25.1 (1983), 16–43.

[39] Y.-C. Ma. The perturbed plane-wave solutions of the cubic Schrödinger equation. *Studies in Applied Mathematics* 60.1 (1979), 43–58.

[40] N. Akhmediev, J. M. Soto-Crespo and A. Ankiewicz. Extreme waves that appear from nowhere: on the nature of rogue waves. *Physics Letters A* 373.25 (2009), 2137–2145.

[41] S. Toenger. Linear and nonlinear rogue waves in optical systems. PhD thesis. Université de Franche-Comté, 2016.

[42] K. Goda and B. Jalali. Dispersive Fourier transformation for fast continuous single-shot measurements. *Nature Photonics* 7.2 (2013), 102.

[43] R. Salem, M. A. Foster and A. L. Gaeta. Application of space–time duality to ultrahigh-speed optical signal processing. *Advances in Optics and Photonics* 5.3 (2013), 274–317.

[44] A. Mahjoubfar, D. V. Churkin, S. Barland, N. Broderick, S. K. Turitsyn and B. Jalali. Time stretch and its applications. *Nature Photonics* 11.6 (2017), 341.

[45] N. Akhmediev, A. Ankiewicz, J. M. Soto-Crespo and J. M. Dudley. Rogue wave early warning through spectral measurements?: *Physics Letters A* 375.3 (2011), 541–544.

[46] N. Akhmediev, J. M. Soto-Crespo, A. Ankiewicz and N. Devine. Early detection of rogue waves in a chaotic wave field. *Physics Letters A* 375.33 (2011), 2999–3001.

[47] S. Toenger, T. Godin, C. Billet, F. Dias, M. Erkintalo, G. Genty and J. M. Dudley. Emergent rogue wave structures and statistics in spontaneous modulation instability. *Scientific Reports* 5 (2015), 10380.

[48] M. Närhi, L. Salmela, J. Toivonen, C. Billet, J. M. Dudley and G. Genty. Machine learning analysis of extreme events in optical fibre modulation instability. *Nature communications* 9.1 (2018), 4923.

[49] R. R. Alfano and S. L. Shapiro. Emission in the region 4000 to 7000 Å via four-photon coupling in glass. *Physical Review Letters* 24.11 (1970), 584.

[50] R. R. Alfano and S. L. Shapiro. Observation of self-phase modulation and small-scale filaments in crystals and glasses. *Physical Review Letters* 24.11 (1970), 592.

[51] R. Dorsinville, P. P. Ho, J. T. Manassah and R. R. Alfano. Applications of super-continuum: present and future. *The supercontinuum laser source*. Springer, 1989, 377–398.

[52] J. M. Dudley and J. R. Taylor. *Supercontinuum generation in optical fibers*. Cambridge University Press, 2010.

[53] Z. Eslami, P. Ryczkowski, C. Amiot, L. Salmela and G. Genty. High-power short-wavelength infrared supercontinuum generation in multimode fluoride fiber. *JOSA B* 36.2 (2019), A72–A78.

[54] A. Hasegawa and F. Tappert. Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers. I. Anomalous dispersion. *Applied Physics Letters* 23.3 (1973), 142–144.

[55] Y. Kodama and A. Hasegawa. Nonlinear pulse propagation in a monomode dielectric guide. *IEEE Journal of Quantum Electronics* 23.5 (1987), 510–524.

[56] N. Akhmediev and M. Karlsson. Cherenkov radiation emitted by solitons in optical fibers. *Physical Review A* 51.3 (1995), 2602.

[57] M. Erkintalo. *Nonlinear instabilities and extreme wave localization in fiber optics*. English. Tampereen teknillinen yliopisto. Julkaisu. Awarding institution:Tampere University of Technology. Tampere University of Technology, 2011. ISBN: 978-952-15-2691-6.

[58] M. H. Frosz, O. Bang and A. Bjarklev. Soliton collision and Raman gain regimes in continuous-wave pumped supercontinuum generation. *Optics Express* 14.20 (2006), 9391–9407.

[59] N. Korneev, E. A. Kuzin, B. Ibarra-Escamilla, M. Bello-Jiménez and A. Flores-Rosas. Initial development of supercontinuum in fibers with anomalous dispersion pumped by nanosecond-long pulses. *Optics Express* 16.4 (2008), 2636–2645.

[60] J. M. Dudley and S. Coen. Coherence properties of supercontinuum spectra generated in photonic crystal and tapered optical fibers. *Optics Letters* 27.13 (2002), 1180–1182.

[61] M. Erkintalo, G. Genty and J. M. Dudley. Rogue-wave-like characteristics in femtosecond supercontinuum generation. *Optics Letters* 34.16 (2009), 2468–2470.

[62] B. Wetzel, A. Stefani, L. Larger, P.-A. Lacourt, J.-M. Merolla, T. Sylvestre, A. Kudlinski, A. Mussot, G. Genty, F. Dias et al. Real-time full bandwidth measurement of spectral noise in supercontinuum generation. *Scientific Reports* 2 (2012), 882.

[63] M. Erkintalo, G. Genty and J. M. Dudley. On the statistical interpretation of optical rogue waves. *The European Physical Journal Special Topics* 185.1 (2010), 135–144.

[64] N. Akhmediev, B. Kibler, F. Baronio, M. Belić, W.-P. Zhong, Y. Zhang, W. Chang, J. M. Soto-Crespo, P. Vouzas, P. Grelu et al. Roadmap on optical rogue waves and extreme events. *Journal of Optics* 18.6 (2016), 063001.

[65] R. Trebino, K. W. DeLong, D. N. Fittinghoff, J. N. Sweetser, M. A. Krumbügel, B. A. Richman and D. J. Kane. Measuring ultrashort laser pulses in the time-frequency domain using frequency-resolved optical gating. *Review of Scientific Instruments* 68.9 (1997), 3277–3295.

[66] S. Linden, H. Gießen and J. Kuhl. XFROG–a new method for amplitude and phase characterization of weak ultrashort pulses. *Physica Status Solidi B* 206.1 (1998), 119–124.

[67] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science* 362.6419 (2018), 1140–1144.

[68] A. Krizhevsky, I. Sutskever and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*. 2012, 1097–1105.

[69] T. Mikolov, M. Karafiát, L. Burget, J. Černocký and S. Khudanpur. Recurrent neural network based language model. *Eleventh annual conference of the international speech communication association*. 2010.

[70] M. N. Wernick, Y. Yang, J. G. Brankov, G. Yourganov and S. C. Strother. Machine learning in medical imaging. *IEEE Signal Processing Magazine* 27.4 (2010), 25–38.

[71] C. L. Chen, A. Mahjoubfar, L.-C. Tai, I. K. Blaby, A. Huang, K. R. Niazi and B. Jalali. Deep learning in label-free cell classification. *Scientific Reports* 6 (2016), 21471.

[72] R. Horisaki, R. Takagi and J. Tanida. Learning-based imaging through scattering media. *Optics Express* 24.13 (2016), 13738–13743.

[73] M. Lyu, W. Wang, H. Wang, H. Wang, G. Li, N. Chen and G. Situ. Deep-learning-based ghost imaging. *Scientific Reports* 7.1 (2017), 17865.

[74] T. Baumeister, S. L. Brunton and J. N. Kutz. Deep learning and model predictive control for self-tuning mode-locked lasers. *JOSA B* 35.3 (2018), 617–626.

[75] T. Zahavy, A. Dikopoltsev, D. Moss, G. I. Haham, O. Cohen, S. Mannor and M. Segev. Deep learning reconstruction of ultrashort pulses. *Optica* 5.5 (2018), 666–673.

[76] J. Friedman, T. Hastie and R. Tibshirani. *The elements of statistical learning*. Vol. 1. 10. Springer series in statistics New York, 2001.

[77] T. G. Dietterich. Ensemble methods in machine learning. *International workshop on multiple classifier systems*. Springer. 2000, 1–15.

[78] D. E. Rumelhart, G. E. Hinton, R. J. Williams et al. Learning representations by back-propagating errors. *Cognitive Modeling* 5.3 (1988), 1.

[79] M. T. Hagan, H. B. Demuth, M. H. Beale et al. *Neural network design*. Vol. 20. Pws Pub. Boston, 1996.

[80] R. Fletcher and C. M. Reeves. Function minimization by conjugate gradients. *The Computer Journal* 7.2 (1964), 149–154.

[81] A. K. Jain. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters* 31.8 (2010), 651–666.

[82] J. A. Hartigan and M. A. Wong. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28.1 (1979), 100–108.

[83] D. Arthur and S. Vassilvitskii. k-means++: The advantages of careful seeding. *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics. 2007, 1027–1035.

[84] I. Goodfellow, Y. Bengio and A. Courville. *Deep Learning*. `http://www.deeplearningbook.org`. MIT Press, 2016.

[85] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research* 15.1 (2014), 1929–1958.