Machine learning-based nonlinear Fourier transform for finite-genus solutions: implementation and application in fibre-optic communications

Stepan Aleksandrovich Bogdanov

Doctor of Philosophy

Aston University

September 2024

Aston University

Machine learning-based nonlinear Fourier transform for finite-genus solutions: implementation and application in fibre-optic communications

Stepan Aleksandrovich Bogdanov

Doctor of Philosophy, 2024

The constantly growing demand for fibre-optic communication traffic motivates researchers to develop new data transmission approaches. The nonlinear Fourier transform (NFT) technique effectively linearizes an information channel and has the potential to overcome the nonlinear capacity limit. However, this method has not been studied thoroughly, especially its counterpart – the periodic NFT. In the context of fibre-optic communication, the periodic NFT is closely related to the finite-genus solutions of the nonlinear Schrödinger equation (NLSE). Previously, analysis of data transmission systems with finite-genus solutions was performed, but the capacity was underestimated due to the restrictions of the periodic NFT. This thesis is devoted to developing the NFT for finite-genus solutions, avoiding any limitations, and providing a fair analysis of the corresponding communication systems.

The complete NFT framework for finite-genus solutions to the NLSE is developed in the thesis. The Riemann-Hilbert problem (RHP) parametrization of finite-genus solutions is exploited. Among the operations constituting the NFT, the inverse problem and the evolution of scattering data are defined in the RHP method, while solving the direct problem is limited. This transformation is performed with a convolutional neural network that lifts existing restrictions. With this neural network-based direct transform, the NFT framework for finite-genus solutions becomes complete.

Having such NFT tools in hand, fair performance estimations of fibre-optic communications with finite-genus solutions data carriers are performed. Numerical simulations of the near-real communication systems are implemented, but the computational complexity of the NFT algorithms is disregarded. In such a system, additional distortions are caused by deviation from the original NLSE model. Applying a convolutional neural network at the receiver to compensate for these impairments while simultaneously recovering the scattering data provides high spectral efficiency comparable to conventional NFT techniques.

Keywords: Convolutional neural network, Data transmission, Nonlinear waves dynamic, Periodic nonlinear Fourier transform, Riemann-Hilbert problem

To my wife and son, who have filled my life with meaning...

Acknowledgments

I would like to say many thanks to my supervisor, Professor Sergei Turitsyn, who brought this project to life. I appreciate the constant support I have received throughout this journey. The wide range of opportunities he provided deserves special thanks.

I am also grateful to my associate supervisor, Dr. Jaroslaw Prilepsky. His exceptional role in this project cannot be overstated. Jaroslaw encouraged me at every stage of the research and spent many hours discussing, sharing ideas, and engaging in friendly conversations.

I would also like to thank my colleagues and friends who accompanied me: Dr. Morteza Kamalian-Kopae for his patient tutoring and sharing of invaluable experience; Professor Dmitry Shepelsky for his indispensable contribution and knowledge; Dr. Anastasiia Vasylchenkova for her significant support; Dr. Egor Sedov for his helpfulness; Dr. Pedro Freire for his deep knowledge of machine learning and Dr. Iryna Karpenko for productive collaboration.

Finally, many thanks to my colleagues at AIPT for their daily support.

Contents

List of acronyms

AG algebro-geometric, 27, 34, 65

AIR achievable information rate, 77

AS auxiliary spectrum, 27

BER bit error ratio, 62, 64, 69, 70, 74-80

CL convolutional layer, 79

CNN convolutional neural network, 23-24, 66-76, 78-82

DBP digital backpropagation, 21

DSP digital signal processing, 18

EDFA erbium-doped fibre amplification, 17-19, 63, 72, 74

FCL fully connected layer, 79

FEC forward error correction, 62, 64, 69-70, 75-76, 80

FGS finite-genus solution, 27, 37-39, 56, 58, 70, 81-82, 84-85

FGS-GT finite-genus solution of a generic type, 41, 43, 54, 81

HD-FEC hard-decision forward error correction, 75-76, 78, 80

MS main spectrum, 27, 37

MSE mean squared error, 45

NFT nonlinear Fourier transform, 2, 14-17, 20-25, 27-28, 33, 36, 38, 40, 43-44, 48, 54-57, 63-66, 73, 78, 80-81, 83-86

NLSE nonlinear Schrödinger equation, 2, 13-17, 20-22, 24, 26-30, 32-34, 37-39, 43, 49, 56-57, 62, 64, 67, 70, 73, 81-83, 85

NN neural network, 22, 41, 43-45, 49, 54, 66, 73-74, 79-81, 84-85

NPDE nonlinear partial differential equation, 14

NS nonlinear spectrum, 14

PAPR peak-to-average power ratio, 56

PNFT periodic nonlinear Fourier transform, 27, 56-57, 62, 64, 66

PSK phase-shift keying, 49, 63, 65, 69, 74

QAM quadrature amplitude modulation, 18, 62-63

RHP Riemann-Hilbert problem, 2, 25-27, 32-36, 38-39, 41-43, 48, 63, 65-66, 68-72, 81, 83-85

SDM spatial division multiplexing, 18-19

SE spectral efficiency, 22, 62-63, 77-78

SSMF standard single mode fibre, 17, 62-63, 69, 74

UWB ultrawideband, 19

WDM wavelength division multiplexing, 17-18, 58

* denotes complex conjugation

$\boldsymbol{I}$ is the unit matrix

List of figures

Figure 3.8: The dependence of the data transmission quality in terms of BER as a function of signal power. The corresponding values of the imaginary part of the main spectrum, parameter a in eq. 3.2, are depicted on the upper x-axis. The FEC threshold is 3.8×10^{-3} (7% overhead). The inset provides the distribution of the phases at the receiver for the optimal power $\approx -6.3\,\mathrm{dBm}$. (Cited from [CP1].)

Figure 3.9: The genus-4 symbols at the transmitter generated with parameter $a = 0.6$ and random phases. The central symbol with duration $T = T_1$ encapsulates nonlinear harmonics with smaller periods T_2, T_3, T_4. The symbols are concatenated one by one without extension prefixes. The amplitude and time are provided in dimensionless units. (Cited from [J2].)

Figure 3.10: The scheme of the fibre-optic communication system. The RHP solver generates information-carrying symbols for the given main spectrum and phases at the transmitter. Then, signals propagate through N optical fibre spans with erbium-doped amplifiers. The described CNN processes the signals at the receiver. In some scenarios, optical bandpass filters and downsampling are applied. (Cited from [J2].)

Figure 3.11: BER estimated in dependence on the signal's launch power after propagation on $1040\,\mathrm{km}$ (13 spans). Two different sampling rates at the receiver $128\,\mathrm{sam/sym}$ (green) and $64\,\mathrm{sam/sym}$ (yellow) are considered. The HD-FEC threshold is 3.8×10^{-3}. The upper x-axis depicts the values of parameter a corresponding to the power on the low x-axis. (Cited from [J2].)

Figure 3.12: The distribution of the phases at the receiver at the optimal power for $64\,\mathrm{sam/sym}$ scenario, $a = 0.75$ (see Fig. 3.11). The shades of red depict the phases from different nonlinear harmonics, while the green points are referred values at the transmitter. (Cited from [J2].)

Figure 3.13: BER for different numbers of optical fibre spans (each $80\,\mathrm{km}$). The $128\,\mathrm{sam/sym}$ scenario is depicted in green, while $64\,\mathrm{sam/sym}$ is in yellow. The HD-FEC threshold equal to 3.8×10^{-3} (7% overhead) also shown. (Cited from [J2].)

Figure 3.14: The linear spectra for the signal at the optimal power $a = 0.75$ at the transmitter (left) and receiver (right). The grey depicts the spectrum before applying a narrow-band filter, and the green is after filtering. (Cited from [J2].)

Figure 3.15: BER provided with the clustered model in dependence on the degree of compression (the number of clusters). For comparison, the BER of the baseline model and the HD-FEC threshold of 3.8×10^{-3} are provided. (Cited from [CP2].)

List of tables

Chapter 1

Introduction

1.1 Nonlinear Fourier transform

The nonlinear Schrödinger equation (NLSE) is a mathematical model for a wide range of physical phenomena such as light propagation in an optical fibre [1, 2], fluid mechanics [3, 4], the Bose-Einstein condensate [5, 6], superconductivity [7, 8], and others [9, 10]. The NLSE describes the evolution of slowly varying envelopes of quasi-monochromatic wave packets[1] in a medium with chromatic dispersion and weak nonlinearity [1]. When the nonlinear term in the NLSE is absent, it simplifies to the standard, linear Schrödinger equation describing the evolution of a quantum-mechanical state of a physical system.

The NLSE is a nonlinear partial differential equation. It can take either a focusing or defocusing form depending on the sign of the nonlinearity and the type of dispersion, which is anomalous or normal. In the medium where self-focusing effects are observed, the focusing NLSE governs the propagation of signals. In such a scenario, energy tends to be concentrated in a localized area, allowing the formation of solitons, demonstrating modulation instability or other phenomena [11, 12]. At the same time, the defocusing NLSE describes dispersive waves propagation leading to energy spreading or generation of the so-called dark solitons [13]. For example, in the description of signal propagation through a silica fibre, the type of dispersion depends on a wavelength and a type of fibre. For the standard single-mode fibre, the dispersion changes its sign at the wavelength of $\lambda_D = 1312\,\mathrm{nm}$ [1]. The propagation of a signal with a carrier wavelength longer than λ_D is governed by the focusing NLSE. However, if the wavelength of a carrier is shorter than λ_D, the defocusing NLSE describes dynamics. The NLSE takes the following dimensionless form:

$$iq_z + \sigma\frac{1}{2}q_{tt} + |q|^2 q = 0, \tag{1.1}$$

[1] referred to as “signals” later in the manuscript.

where t is temporal variable, z is evolutionary variable, and $q(t, z)$ is an envelope of a wave packet. The parameter σ defines the type of the equation, $\sigma = 1$ for the focusing and $\sigma = -1$ for the defocusing NLSE.

The NLSE allows analytical solutions in a restricted number of cases [14, 15]. Among others, there is a fundamental analytical solution to the focusing NLSE – soliton. The soliton is a solitary wave that keeps its waveform due to the joint action of dispersion and nonlinearity while propagating through a media [16, 17]. The solitons were a subject of interest for a long time in such areas as optical communication [18] and water surface dynamics [4]. However, in many practical cases, the exploration of a signal dynamic is only available through numerical methods. There are a few approaches to solve the NLSE numerically: the split-step Fourier method [1, 19], the finite-difference method [1], and the nonlinear Fourier transform technique [20, 21].

The linear Fourier transform is a powerful tool for the analysis of linear differential equations. The approach allows the construction of an analytical solution in the linear Fourier domain as well as linear modes decomposition, providing an opportunity for the comprehensive examination of the solution. However, this approach doesn't work for nonlinear equations such as the NLSE (1.1). This is because the superposition principle is not satisfied in the presence of nonlinearity. In other words, the modes interact with each other and do not propagate independently. Nevertheless, some nonlinear partial differential equations (NPDE) allow for a generalization of the Fourier transform method through the decomposition of a solution into nonlinear modes. It is called *the nonlinear Fourier transform*[2] (NFT), obviously because it deals with nonlinear equations. The systems allowing such analysis are known as integrable. They are the Korteweg–de Vries equation, the mentioned nonlinear Schrödinger equation, the Sine-Gordon equation, and others [22, 23]. In this context, integrability means that for a given NPDE, there exists an ordinary linear differential equation or a system of equations for which the solution of NPDE serves as a potential. While the solution of NPDE experiences complicated dynamics, the spectrum of the operator corresponding to the ordinary linear differential equation has a trivial linear evolution property [20, 21]. Therefore, the nonlinear Fourier transform represents the solution of integrable NPDE as a set of nonlinear modes that have a linear evolution property despite their nonlinear nature. Such an approach enables solving and comprehensive analysis of integrable nonlinear equations. It found its application in fluid mechanics, nonlinear optics, and other areas.

In the frames of the NLSE, *the nonlinear Fourier transform* is an approach to convert the nonlinear evolution governed by the NLSE into a linear evolution of special parameters, so-called *nonlinear spectrum*[3] (NS), defined within the nonlinear Fourier domain [21]. It is possible because the NLSE belongs to the class of integrable systems, for

[2] referred to as *inverse scattering transform* in literature.
[3] also known as *spectral data* or *scattering data*.

which an effective linearization of the nonlinear evolution is possible [16, 20]. The NFT framework consists of solving *the direct problem* that is finding the spectral data for a solution to the NLSE given in the time domain $q(t, z)$, *the inverse problem* that is retrieving of $q(t, z)$ from the nonlinear spectrum, and *evolution lows* for the scattering data in the nonlinear Fourier domain, that are linear relations showing how the nonlinear spectrum changes with z. In other words, the NFT is a way to solve an initial value problem for the NLSE getting around the complex nonlinear dynamic of a signal: (i) a solution $q(t, z)$ is transformed by means of solving the direct problem to the spectral data $Q(k, z)$ (k is a spectral parameter), (ii) of which linear evolution properties are exploited to find them at any value of z, and then (iii) the solution in time domain is retrieved again at new z value of interest with the inverse problem. A scheme demonstrating the approach is depicted in Fig. 1.1.

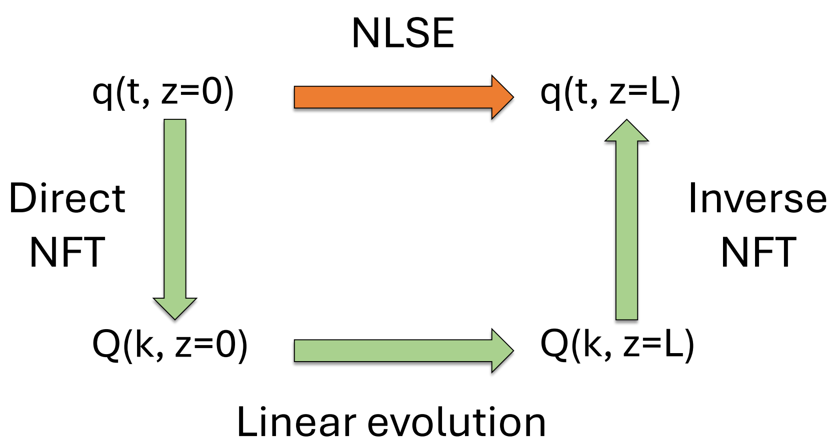

Figure 1.1: The scheme of the NFT technique to solve the NLSE. $q(t, z)$ is a solution in the time domain, $Q(k, z)$ is the spectral data.

Plenty of works dedicated to the conventional nonlinear Fourier transform are presented in the literature [16, 20, 21]. Initially, the NFT was developed for vanishing boundary signals $q(t, z)$, which decay exponentially as t tends to infinity. In this case, the nonlinear spectrum consists of two parts: a discrete spectrum and a continuous spectrum[4], which are defined by solving the appropriate generalized eigenvalue problem, also known as the Zakharov-Shabat problem with $q(t, z)$ playing the role of potential:

$$\mathbf{\Phi}_t(t, z, k) = \begin{pmatrix} -ik & q(t, z) \\ -\sigma q^*(t, z) & ik \end{pmatrix} \mathbf{\Phi}(t, z, k), \tag{1.2}$$

where k is a spectral parameter (eigenvalue) and $\mathbf{\Phi}(t, z, k)$ is an auxiliary vector eigen-

[4]The existence of the discrete spectrum, solitonic components, is possible exclusively for the focusing NLSE. In the defocusing scenario, only the continuous spectrum exists.

function. σ takes value "$+1$" and "-1" for the focusing and defocusing case correspondingly. Solving the system that is finding its eigenvalues and eigenfunctions gives the nonlinear spectrum[5] (a detailed procedure is given [20, 21]). *The discrete spectrum* consists of eigenvalues $\{k_i \in \mathbb{C}^+\}_{i=1}^N$ that are associated with *the complex amplitudes* $\{r(k_i) \in \mathbb{C}\}_{i=1}^N$, defined through the solution of the system (N is a number of discrete spectrum components). It represents solitonic components in the signal $q(t, z)$ that are either solitary soliton or multisoliton structures. The frequency and amplitude of a soliton are equal to the real and imaginary part of the k_i, while $r(k_i)$ determines a time shift and phase. *The continuous spectrum* can be found from the solution of the system with the spectral parameter k defined on the real axis $\{r(k), k \in \mathbb{R}\}$. It corresponds to radiative waves. In the quasilinear scenario, the limit of a small signal's amplitude, the nonlinear spectrum does not contain discrete components, while the continuous spectrum tends to the linear Fourier transform of the signal $q(t, z)$.

The discrete spectrum is invariant, while a wave packet propagates over a medium, and the focusing NLSE governs its evolution. Whereas the continuous spectrum and complex amplitudes have a linear evolution:

$$\begin{aligned} r(k, z) &= r(k, z_0) e^{2ik^2(z-z_0)}, \\ r(k_i, z) &= r(k_i, z_0) e^{2ik_i^2(z-z_0)}. \end{aligned} \tag{1.3}$$

The solution to the inverse problem is found through a couple of integral equations, the Gelfand-Levitan-Marchenko system, with the kernel expressed through the discrete and continuous spectra. When the spectral data are provided, they uniquely determine $q(t, z)$. A more detailed and mathematically rigorous description of the NFT framework can be found in corresponding literature [16, 17, 20, 21].

The nonlinear Fourier transform has been actively applied in fibre-optic communications, where the NLSE is a master model describing the propagation of the information signals through optical fibre [24]. These signals experience the joint action of chromatic dispersion and nonlinearity, which complicates their evolution. Initially, nonlinearity was considered a harmful property of an optical fibre that forces communication systems to operate in a linear regime. At the same time, dispersion introduces the channel memory and can be compensated easily in the linear Fourier domain. The advent of the NFT refined the role of nonlinearity: in the NFT framework, nonlinearity is considered an intrinsic property of signals but not a detrimental phenomenon. Using the linear evolution properties of the scattering data allows compensation for the action of nonlinearity as well as chromatic dispersion. Hasegawa's and Nyu's fundamental work proposed the idea of using the nonlinear spectrum with its properties, where the infor-

[5] Although the system (1.2) is linear, the term "nonlinear spectrum" is widely used in the literature, because related to the analysis of nonlinear partial differential equations.

mation was encoded to a discrete spectrum [25]. Later, this approach was implemented in distinct techniques such as continuous spectrum modulation, soliton communication, b-modulation, and periodic nonlinear Fourier transform [26, 27].

Alternatively, the nonlinear Fourier transform for periodic signals was developed [28, 29, 30]. However, this technique is mathematically involved and requires analysis with the algebro-geometric approach or the Riemann-Hilbert problem method [31]: behind this periodic NFT are finite-genus solutions to the NLSE [32]. Nevertheless, some data transmission systems have been constructed based on this method [33, 34]. A detailed description of the periodic NFT and communications with this technique is presented in this thesis.

1.2 Fibre-optic communications review

Two key technologies that established the era of fibre-optic communications were the advent of low-loss optical fibre in 1970 and the invention of AsGa semiconductor laser able to operate at room temperature [35]. The performance of the first commercial fibre-optic communication systems was extremely low due to periodic optoelectronic regeneration. The loss of optical fibre required signals to be received, regenerated in the electronic domain, and transmitted again. The distance between transceivers was about $50\,\mathrm{km}$. In the earlier 1990s, such systems provided capacity of $5\,\mathrm{Gb/s}$ over 2 wavelength channels ($2.5\,\mathrm{Gb/s}$ per channel) being restricted mainly by the interface rate of existing transponders [36].

For reference, the Internet connection speed I have in my university's office in 2024 is about 100 Mb/s. It allows me to browse the Internet, watch high-quality videos, and have online meetings. Therefore, 5 Gb/s systems available in the 1990s could satisfy only about 50 modern users.

The epoch of regenerating systems ended with the advent of optical fibre amplifiers [37]. The erbium-doped fibre amplifier (EDFA) was invented at the beginning of the 1990s and allowed direct signal amplification without converting into the electrical domain. Though an EDFA didn't provide a gain in systems' capacity, it enabled the development of wavelength-division multiplexing (WDM) technique through effective optical signal amplification all over the bandwidth that EDFA is able to operate without the necessity for channels separation. Commercial WDM systems available at the beginning of 1990s had a capacity of up to $20\,\mathrm{Gb/s}$ provided by 8 frequency channels [36]. The intrachannel data rate of $2.5\,\mathrm{Gb/s}$ was restricted by the chromatic dispersion of the standard single-mode fibre (SSMF) that is about $17\,\mathrm{ps/nm/km}$ at $1550\,\mathrm{nm}$. The further growth of capacity was achieved by means of special dispersion-shifted fibres providing reduced dispersion at the operation wavelength, and the dispersion management technique enabled compensation for an aggregated dispersion through mixing fibres of different types. By 2009, the commercial systems based on the dispersion management supported data transmission at $40\,\mathrm{Gb/s}$ per channel with aggregated capacity $3.2\,\mathrm{Tb/s}$ provided by 80 wavelengths channels [38]. This throughout was restricted by the bandwidth that an EDFA could operate with (C+L band ranging from $1530\,\mathrm{nm}$ to $1625\,\mathrm{nm}$) and fibre loss in the expanded range of wavelengths [35]. The further increase in capacity required new amplification systems to overcome these limits.

The development of optical coherent detection made it possible to operate with the whole signal at the receiver, allowing applying advanced data transmission techniques [39]. Modulation methods developed to use the amplitude and phase of a signal, as well as both polarizations, provided higher spectral efficiency. Such modulation formats as binary/quadrature phase shift keying or 16-QAM (quadrature amplitude modulation), 64-QAM, or even 1024-QAM significantly increased the capacity of systems. $400\,\mathrm{Gb/s}$ per WDM channel transmission based on dual-polarization dual-carrier $32\,\mathrm{Gboud}$ and 16-QAM over $550\,\mathrm{km}$ was available in 2015 [40]. Later, in 2020 $1.6\,\mathrm{Tb/s}$ per channel system was reported which operated at $95.6\,\mathrm{Gboud}$ with 64-QAM signals providing data transmission up to $1000\,\mathrm{km}$ [41]. At the same time, access to a full signal enabled by coherent detectors gave rise to digital signal processing (DSP) methods. They include signal modulation, pulse shaping, and precompensation techniques at the transmitter, as well as equalization (chromatic dispersion and nonlinearity compensation), filtering, demodulation, and decoding at the receiver [42]. Modern digital signal processing systems have become an important part of communications, consuming a significant portion of total energy [43].

By the moment the fibre-optic communication systems had reached their limit with the WDM and the coherent detection technologies, the further increase of capacity was able by the parallel deployment of new systems. This meant a linear growth of the systems' throughput with cost. The spatial-division multiplexing (SDM) technology provides access to a spatial degree of freedom to transmit information [44]. It can be implemented by deploying special multicore fibres (containing a few cores under the cladding) operating with signals propagated over many cores simultaneously. Another approach is using multimode fibres, allowing many spacial modes to be propagated over one core [45]. The different spacial channels in SDM systems can share common elements, including light sources, amplifiers, and DSP processors, saving energy and cost-per-bit. The SDM technology being proposed in the 1990s became available only after 2010 because of progress in fibre fabrication and the advent of techniques allowing for compensation inter cores/modes interactions [46]. The report published in 2018 demonstrated data transmission over more than 100 spacial channels in a single fibre that consisted of 19 cores, each supporting 6 modes. The transmission exploited 739 WDM channels in C+L band and increased the aggregated capacity from $100\,\mathrm{Tb/s}$ provided by single-mode single core fibre up to $10.16\,\mathrm{Pb/s}$ over the distance $11.3\,\mathrm{km}$

[47]. Therefore, applying the SDM technology increased capacity proportional to the number of special channels.

A new, fast-developing area of research is ultrawideband (UWB) communications. Initially, communication systems operated at wavelengths from $1530\,\mathrm{nm}$ to $1565\,\mathrm{nm}$ (C-band), corresponding to the range where silica fibre has a minimal loss, and an EDFA is able to provide amplification. By 1998 modified EDFA also covered the L-band ($1565\,\mathrm{nm}$ to $1625\,\mathrm{nm}$), providing in total transmission over 80 frequency channels in aggregated C+L-band [35]. Functioning behind this range is challenging due to the increased fibre loss and dispersion. Moreover, amplification technologies based on EDFA don't support operation in the extended wavelength range, and new approaches were required [48]. Later, hybrid amplification schemes made it possible to transmit data through a single-mode fibre beyond the C+L band. The $190\,\mathrm{Tb/s}$ transmission over $54\,\mathrm{km}$ was demonstrated with standard EDFA amplification in C+L band and additional Raman/thulium-doped fibre amplifier in S-band [49]. Also, the record occupation of a single-mode fibre bandwidth was shown using all bands (O-E-S-C-L-U-band) in a silica fibre transparency window. The transmission over the range from $1260\,\mathrm{nm}$ to $1675\,\mathrm{nm}$ was implemented with a complex combined amplification scheme. In this experiment, the capacity achieved $402.2\,\mathrm{Tb/s}$ over $50\,\mathrm{km}$ [50].

Finally, it is worth mentioning the increasing role of machine learning and artificial neural networks in modern fibre-optic communications. Such techniques accompany all stages of the transmission system's functioning. Machine learning applications in optical communications can be categorized as offline and online approaches. Offline applications are used at all stages before the system exploitation. They are system design, parameters optimization, and performance estimations. Prediction of traffic, as well as quality estimation of the future systems, are critical to prevent overusing resources and saving them for other requests [51]. At the same time, the adoption of modern approaches in communications makes systems to be attributed with many parameters and complex for optimization. Neural networks can effectively reveal hidden relations between parameters (bandwidth, modulation format, power, and others) and provide optimal configurations [52]. On the other hand, the operation of modern communication systems requires permanent online analysis of their state and real-time processing of transmitting/receiving data. Optical performance monitoring provides information for the system self-configuration as well as for the computation of an optimal transmission path [53]. Many researches are devoted to the application of neural networks for digital signal processing, including equalization, nonlinear distortion compensation, and others [54, 55]. These are just a few machine learning applications in optical communication; a broader list can be found in [51, 52].

1.3 NFT-based communications

The nonlinear Schrödinger equation, eq. (1.1), is a master model for information signals propagation in a fibre-optic communication system [1]. It describes the intricate interplay between dispersion and nonlinear effects, resulting in a complicated signal dynamic. At the same time, the NLSE belongs to the so-called integrable systems for which nonlinear Fourier transform can be developed [21]. The NFT effectively linearizes an optical channel due to the linear property of the scattering data. While an information signal propagates through an optical fibre and experiences joint action of dispersion and nonlinearity, its nonlinear spectrum demonstrates linear evolution that can be compensated in the nonlinear Fourier domain with a trivial phase rotation, eq. (1.3), [26, 27] (see Fig. 1.1). A constantly growing need for data transmission capacity leads to an increase in the power of information signals to use high-modulation formats. However, such systems suffer from nonlinear effects that can not be neglected. It, in turn, restricts the performance of communications and causes a so-called nonlinear limit [24, 56]. Therefore, the NFT approach is an attractable tool for simultaneous dispersion and nonlinearity management.

One of the first applications of the NFT in fibre-optic data transmission systems relates to the idea of eigenvalue communication [25]. It is proposed that information is encoded into invariant eigenvalues of solitonic signals. In the following research, different approaches to NFT communication were developed. They are based on encoding information into continuous spectrum [57, 58, 59] or discrete eigenvalues (solitonic components) [60, 61, 62, 63] as well as both these components of the nonlinear spectrum [64, 65, 66]. In these scenarios, the evolution of scattering data is trivial; see eq. (1.3). Later, an advanced technique known as b-modulation was proposed. This approach provides higher system effectiveness and full control over signal duration in contrast to the conventional method based on continuous spectrum modulation [67, 68, 69, 70]. Finally, data transmission systems exploiting the periodic NFT were proposed [33, 34, 71, 72, 73]. The last approach is described in detail in the corresponding chapter below.

One approach to apply the NFT in fibre-optic communication is the direct modulation of the nonlinear spectrum [58, 61, 64]. The information is encoded into the nonlinear spectrum (continuous, discrete, or both simultaneously) in this scenario. Then, the inverse problem is solved at the transmitter, and the signal $q(t, z = 0)$ corresponding to this nonlinear spectrum is propagated through the optical fibre. At the receiver side, the direct problem for the detected signal $q(t, z = L)$ (where L is propagation distance) is solved to get its scattering data. The evolution of this nonlinear spectrum is compensated with eq. (1.3) to get the scattering data at the transmitter ($z = 0$), and it is decoded. The scheme of this procedure is depicted in Fig. 1.2.

Alternatively, the NFT may be applied at the receiver at the digital backpropagation (DBP) stage [74, 75]. In such an approach, the data is transmitted with traditional data carriers, such as sinc pulses or their root-raised cosine modifications [24]. At the receiver, after propagation, signals undergo the direct NFT, the evolution of their scattering data is compensated, and the inverse NFT operation is performed. The procedure's output is information signals with compensated accumulated dispersion and nonlinearity corresponding to the signals at the transmitter (see Fig. 1.2). A detailed description of both approaches is given in [26].

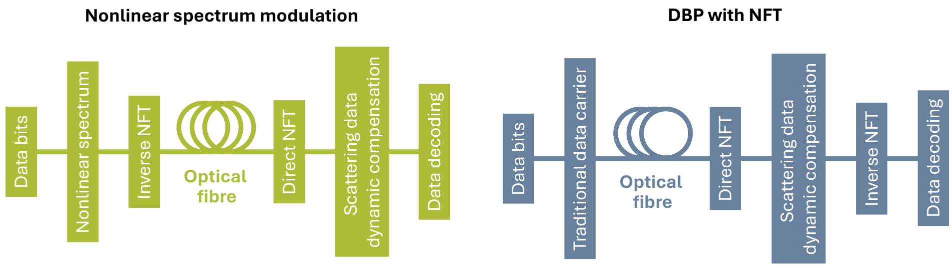

Figure 1.2: Two applications of the NFT in fibre-optic communications: encoding information through the nonlinear spectrum modulation (left) and using the NFT for digital backpropagation (right).

The lossless NLSE in the form of eq. (1.1) contains only nonlinear and dispersion terms. However, in real communication systems, information-bearing signals undergo periodic attenuation and amplification due to the loss of optical fibre and the action of amplifiers. The model involving the last two phenomena is not integrable, and the NFT does not work in such systems. In other words, the scattering data no longer undergo the evolution equations (1.3). Even though the dynamics of the signals are not governed by the lossless NLSE, the propagation of the signals averaged over many fibre spans is still described by the NLSE in the form of eq. (1.1), the so-called lossless path-averaged model [58].

Though the NFT methods provide a solution to effectively compensate for the coupled action of dispersion and nonlinearity in the nonlinear Fourier domain, their practical implementation is limited due to the high numerical complexity of the direct and inverse transforms [26, 76]. The most computationally effective algorithms are known as "superfast NFT". They provide the direct NFT transformation with complexity $O(M\log^2 M)$ for signals containing the continuous spectrum only and $O(M^2)$ for discrete spectrum (M is the number of samples in signal $q(t)$) [77, 78]. At the same time, the "superfast inverse NFT" for an arbitrary spectrum can be implemented with total complexity $O(MN + M\log^2 M)$, where N is the number of solitonic components [79, 80]. Despite these efforts in developing the NFT algorithms, they didn't find applications in real systems because their complexity is still too high, and their implementation is not justified in contrast to the conventional transmission systems. However, it was demonstrated that the performance of the NFT systems can outperform their linear counterparts [69, 81]. Therefore, NFT systems can find practical applications if they provide significantly higher capacity to compensate for the excessive costs related to their complexity [82].

To finish this section, a few remarkable communications results with the NFT should be mentioned. In the first example, data transmission was realized with the b-modulation method [83]. The authors investigated how the noise in the system correlates with information signals, enabling them to improve performance. The transmission parameters were the following: the bandwidth was $5\,\mathrm{GHz}$, and the propagation distance was $960\,\mathrm{km}$. The maximum SE of $5.51\,\mathrm{bits/s/Hz}$ was reported for such settings. Another example is a dual-polarization NFT scheme with nonlinear frequency division multiplexing [69]. The Hermite-Gaussian-based wave carriers, which provide exceptionally narrow bandwidth, were exploited in this system. SE of $12\,\mathrm{bits/s/Hz}$ over both polarizations was demonstrated with the bandwidth of $4.75\,\mathrm{GHz}$ and the transmission distance of $800\,\mathrm{km}$. This work provides a direct comparison with linear transmission in terms of spectral efficiency and demonstrates the outperformance of the NFT technique.

Moreover, some noteworthy applications of neural networks (NN) in NFT-based communications have been developed. A nonlinear frequency division multiplexing system with NN-based symbols decision was proposed in [84]. Information was encoded into solitons that were processed directly in the time domain at the receiver, avoiding the NFT transformation. In the non-ideal system, where the signals were distorted with periodic gain/loss and noise, the approach demonstrated significantly higher performance than the conventional numerical NFT. In another research, two convolutional neural networks, a small-size serial NN and a high-speed parallel NN, for application in nonlinear frequency division multiplexing systems are proposed [85]. The authors provided a comprehensive analysis of the models and gave specific recommendations for the hardware implementation of each NN. In one more example, the direct spectral problem to the NLSE was solved with a convolutional neural network, coined later NFT-Net [86]. The NN retrieved the continuous nonlinear spectrum and demonstrated advantages in comparison with the conventional numerical NFT when the processed signals were distorted by noise. This feature of the method is especially important in fibre-optic communications applications. For more research on this topic, see [87, 88, 89, 90].

1.4 Motivation and thesis contribution

Many methods of data transmission based on the NFT are covered in the literature. They are mainly devoted to the NFT with decaying boundary signals. However, practical implementation of these approaches still requires significant efforts to address existing problems. At the same time, it was reported that the periodic NFT has some advantages over its conventional counterpart. However, only a few works mention communications based on the periodic NFT. It is explained by the incomplete theory behind the approach. Therefore, the performance of such data transmission systems was underestimated.

For the mentioned reasons, the direct problem solver for finite-genus solutions that are based on convolutional neural networks (CNN) was implemented in this work. The development of this approach provided the complete NFT framework for finite-genus solutions. The method and its analysis are the contribution of the thesis. Chapter 2 covers that study. Another research was devoted to the numerical realization of a fibre-optic communication system based on the complete NFT framework for finite-genus solutions. Investigation of data transmission with a proposed CNN-based receiver, simulations of such systems, and their performance analysis are the second contributions of the thesis. It is described in Chapter 3.

The following works have been published in journals (J) and as conference papers (CP):

[J1] S. Bogdanov, D. Shepelsky, A. Vasylchenkova, E. Sedov, P. J. Freire, S. K. Turitsyn, and J. E. Prilepsky, "Phase computation for the finite-genus solutions to the focusing nonlinear Schrödinger equation using convolutional neural networks." *Communications in Nonlinear Science and Numerical Simulation*, vol. 125: 107311, 2023.

[J2] S. Bogdanov, D. Shepelsky, M. Kamalian-Kopae, A. Vasylchenkova, and J. E. Prilepsky, "Finite-genus solutions-based optical communication with the Riemann-Hilbert problem transmitter and a convolutional neural network receiver." *Journal of Lightwave Technology*, vol. 42, no. 16, pp. 5529-5536, 2024.

[J3] D. Shepelsky, I. Karpenko, S. Bogdanov, and J. E. Prilepsky, "Periodic finite-band solutions to the focusing nonlinear Schrödinger equation by the Fokas method: inverse and direct problems." *Proceedings of the Royal Society A*, vol. 480 (2286): 20230828, 2024.

[CP1] S. Bogdanov, J. E. Prilepsky, D. Shepelsky, M. Kamalian-Kopae, A. Vasylchenkova, E. Sedov, and S. K. Turitsyn, "Fiber-optic communications based on finite-genus solutions to the NLS with a convolutional neural network receiver." *49th European Conference on Optical Communications (ECOC 2023)*, Hybrid Conference, Glasgow, UK, pp. 467-470, 2023.

[CP2] S. Bogdanov, P. J. Freire, and J. E. Prilepsky, "Complexity reduction of neural networks for nonlinear Fourier transform in optical transmission systems." *International Conference on Machine Learning for Communication and Networking (ICMLCN)*, Stockholm, Sweden, pp. 537-542, 2024.

[CP3] S. Bogdanov, “Nonlinear Fourier transform for finite-genus solutions of a generic type: application in fibre-optic communications.” *XI-th International Conference “SOLITONS, COLLAPSES AND TURBULENCE: Achievements, Developments and Perspectives”*, Belgrade, Serbia, 2024.

[CP4] S. Bogdanov, “Fibre-optic communication based on the inverse scattering transform for finite-genus solutions.” *4th IMA Conference on Inverse Problems from Theory to Application*, Bath, UK, 2024.

1.5 This thesis

- In the introduction, a short description of the NLSE and the corresponding NFT is provided. It followed with a brief history of fibre-optic communication systems, showing how the advent of different technologies enhanced data transmission capacity. This ends with a discussion of the communications based on the NFT.
- Chapter 2 starts with the theory behind the periodic NFT and finite-genus solutions. Two frameworks, the algebro-geometric approach and the Riemann-Hilbert problem method, are described. Then, I define finite-genus solutions of a generic type and specify the problems limiting the applicability of the periodic NFT. Later in the chapter, a description of a convolutional neural network-based method to solve the direct problem for finite-genus solutions of a generic type is provided. This approach is able to process arbitrary finite-genus solutions, lifting the restriction to operate in a quasi-linear limit. It follows with details of the method implementation and performance estimation. Finally, the result of the CNN model compression with the weight clustering technique is outlined.
- In Chapter 3, the principles of communications with the periodic NFT and its existing implementations are reviewed. Then, the data transmission using finite-genus solutions of a generic type and the method of phase recovery described in Chapter 2 is introduced and illustrated with a few numerical experiments. First, the idealized fibre-optic communication system with minimal deviation from the NLSE model is presented. At the receiver, the CNN-based detector is exploited to retrieve the phases of information symbols. Second, a comprehensive analysis of a more practical transmission scheme is provided. The configuration with a non-zero gain/loss profile is realized while the CNN-based receiver processes non-periodic signals. It is implemented to achieve higher capacity. Estimations of efficiency and comparison with other NFT transmissions are provided, Finally, the weight clustering technique is applied to compress the CNN model.
- The thesis concludes with a brief summary.

Chapter 2

NFT for finite-genus solutions

2.1 The Riemann-Hilbert problem approach

An idea behind the Riemann-Hilbert problem (RHP) formalism is constructing an analytic function in the complex plane, having known its behavior on the boundaries of analyticity domain[1]. Let's consider an oriented contour Γ in the complex plane and a mapping $\boldsymbol{G}$ from this contour to an invertible $N \times N$ matrix[2]. The Riemann-Hilbert problem consists of constructing the matrix function $\boldsymbol{\Psi}(\lambda)$ such that:

- $\boldsymbol{\Psi}(\lambda)$ is analytic everywhere in $\mathbb{C} \setminus \Gamma$;
- $\boldsymbol{\Psi}(\lambda)$ satisfies to the jump condition $\boldsymbol{\Psi}^{+}(\lambda) = \boldsymbol{\Psi}^{-}(\lambda)\boldsymbol{G}(\lambda)$, with $\boldsymbol{\Psi}^{+}(\lambda)$ and $\boldsymbol{\Psi}^{-}(\lambda)$ being the limiting values as λ approaches Γ from the left and right side correspondingly;
- as λ tends to infinity $\boldsymbol{\Psi}(\lambda) \to \boldsymbol{I}$ (identity matrix).

The conditions provided above are enough to uniquely determine the analytic function $\boldsymbol{\Psi}(\lambda)$. It is worth noting that contour Γ might be nontrivial, consisting of a few components, being closed, and having self-intersections. This contour, along with the jump matrix $\boldsymbol{G}(\lambda)$, are defined by the problem under consideration. The detailed explanation and rigorous mathematical description of the Riemann-Hilbert problem can be found in [91, 92].

In a simple scenario, when the problem of interest is reduced to a scalar RHP, i.e., $N = 1$, the solution $\Psi(\lambda)$ can be found explicitly in terms of the contour integrals (Cauchy-type integrals). Such a case takes place in the analysis of the special functions and other applications; see for examples [92]. However, in general, the RHP can be

[1] In this section, I am not chasing for mathematical rigor. Here is just a sketch to provide insight for reading the following sections.

[2] The general case of the matrix RHP is considered. However, a scalar problem also appears in many applications.

reduced to a linear singular integral equation. It is a typical situation when the original problem that has been formulated in terms of RHP is nonlinear. Therefore, the RHP formalism allows for an effective linearization of the initially nonlinear problem [91].

An important application of the Riemann-Hilbert problem approach is the construction of solutions to nonlinear partial differential equations: the Korteweg–de Vries equation, the nonlinear Schrödinger equation, and others [93, 94]. The benefits of using the RHP come from the following: a solution to the nonlinear equation expressed in the term of the RHP has an explicit linear dependence from its temporal/spatial and evolutionary variables. The approach developed for numerical solving of the Riemann-Hilbert problem made it possible to construct solutions to the nonlinear equations [92, 95]. To illustrate, let's consider schematically a solution to the NLSE with vanishing boundary initial conditions. It is constructed with the RHP approach using the continuous spectrum of the solution as input data for the problem [96]. The matrix-valued function $\mathbf{\Psi}(\lambda)$ satisfies:

$$\mathbf{\Psi}^{+}(\lambda, t, z) = \mathbf{\Psi}^{-}(\lambda, t, z)\boldsymbol{G}(\lambda, t, z), \tag{2.1}$$

where t and z, temporal and evolutionary variables of the NLSE, eq. (1.1), are parameters of the problem now. The jump matrix $\boldsymbol{G}(\lambda, t, z)$:

$$\boldsymbol{G}(\lambda, t, z) = e^{-i\lambda t\boldsymbol{\sigma}_3} \begin{pmatrix} 1 & r^*(\lambda^*) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ r(\lambda) & 1 \end{pmatrix} e^{i\lambda t\boldsymbol{\sigma}_3}. \tag{2.2}$$

$r(\lambda)$ is the continuous spectrum, eq. (1.3), and $\boldsymbol{\sigma}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ is the Pauli matrix. The contour Γ is the real axis. However, just because $r(\lambda)$ decays to zero as $\lambda \to \pm\infty$, Γ can be truncated at some values. Then, the solution to the NLSE can be expressed from the asymptotic behavior:

$$\mathbf{\Psi}(\lambda, t, z) = \boldsymbol{I} + \frac{\mathbf{\Psi^1}(t, z)}{\lambda} + \dots, \quad \lambda \to \infty \tag{2.3}$$

as (1,2) entrance of the matrix $\mathbf{\Psi^1}(t, z)$:

$$q(t, z) = 2i\mathbf{\Psi^1}_{1,2}(t, z). \tag{2.4}$$

As mentioned, $\mathbf{\Psi}(\lambda, t, z)$ has the explicit parametric dependence on t and z, providing the solution $q(t, z)$ for any values of these variables.

Another example is the finite-genus solutions to the Korteweg–de Vries equation, which were recovered using the asymptotics of the Baker–Akhiezer functions. The lasts, in turn, were represented as a solution to the corresponding Riemann-Hilbert problem [93]. Later, this approach was developed to calculate the large-genus solutions to the Korteweg–de Vries equation thanks to employing weighted Chebyshev basis and weighted Cauchy transforms [97].

2.2 Periodic nonlinear Fourier transform and finite-genus solutions

The nonlinear Fourier transform has been developed initially for the vanishing boundary signals: $q(t)$ decays to 0 as t tends to $\pm\infty$ [16, 20, 21]. Alternatively, integration of the nonlinear Schrödinger equation (1.1) can be implemented with the periodic initial condition:

$$q(t) = q(t+T),\ T \in \mathbb{R}. \tag{2.5}$$

The periodic nonlinear Fourier transform (PNFT) is a framework for parametrizing periodic signals in the nonlinear Fourier domain. The PNFT is closely linked with *finite-genus solutions (FGS)*, also referred to as finite-band or finite-gap solutions in the literature.

The finite-genus solutions were initially introduced in the works devoted to integrating the Korteweg-de Vries equation with a periodic Cauchy initial problem [28]. The initial value problem for NLSE with periodic signals was solved by means of *the algebro-geometric (AG) approach* in Kotlyarov and Its works [15, 29] as well as Ma and Ablowitz [98]. A detailed description of the nonlinear Schrödinger equation integration, as well as others, with the AG method, is given in [31]. Alongside the algebro-geometric approach, a solution to the NLSE can be constructed with *the Riemann-Hilbert problem (RHP) method*, where a solution to the NLSE is defined from the asymptotic behavior of the analytic function constructed through solving the corresponding Riemann-Hilbert problem in the complex plane [30, 92].

The finite-genus solutions to the focusing NLSE[3] allow parametrization with a discrete spectrum in the nonlinear Fourier domain. This spectrum consists of *the main spectrum (MS)* and *the auxiliary spectrum (AS)*. The main spectrum is invariant: it remains constant while a signal evolves according to the NLSE. Meanwhile, the auxiliary spectrum describes the signal's dynamic and depends on t and z variables in eq. (1.1). The main spectrum is the same for the algebro-geometric approach and the Riemann-Hilbert problem method. The term auxiliary spectrum was introduced in the AG approach, while for the RHP method, it is replaced by *the phases.* In this context, the nonlinear Fourier transform framework for finite-genus solutions consists of the direct problem, which is finding the main and auxiliary spectra from a given $q(t, z_0)$ at a specific value of z_0, and the inverse problem, which involves retrieving the signal in the time domain from its nonlinear spectrum. Together with these transformations, the

[3] Later in the text, I talk about the finite-genus solutions to the focusing NLSE equation if other is not specified. However, the same framework can be developed for the defocusing NLSE.

evolution low of the auxiliary spectrum constitutes the NFT framework for finite-genus solutions (see Fig. 1.1).

The following sections in this chapter describe two approaches to parameterize finite-genus solutions in the nonlinear Fourier domain: the algebro-geometric approach and the Riemann-Hilbert problem method. It is followed by a description of the numerical implementation of the methods and the peculiarities of their use. Then, I provide a generalization of the NFT for finite-genus solutions of a generic type with the support of machine learning techniques. The section 2.3 is devoted to the method developed in the frame of this research and published in [J1]. It solves the direct problem of the NFT for finite-genus solutions of a generic type with a convolutional neural network. A description of the model is provided, as well as details of the training process and estimation of the performance. Moreover, the weight clustering compression technique is applied to reduce the computational complexity of the neural network. This research was presented at a conference and published [CP2]. The chapter ends with a conclusion summarizing the research.

2.2.1 Algebro-geometric approach

Without loss of generality, consider the parametrization of finite-genus solutions with the algebro-geometric approach in the terms provided by Kotlyarov and Its [15, 29]. According to the formalism, a genus-N solution to the focusing NLSE $q(t,z)$ can be described with the following set of parameters:

- *Main spectrum:* $N+1$ pairs of complex conjugated number $\{\lambda_j, \lambda_j^*\}_{j=0}^N \in \mathbb{C}$;
- *Auxiliary spectrum:* N values $\{\mu_j\}_{j=1}^N \in \mathbb{C}$;
- *Riemann surface sheet indices:* $\sigma_j \in \{-1, 1\}$ defined for each μ_j;
- *Initial amplitude:* $q(t_0, z_0)$.

The main spectrum of a genus-N solution comes as $N+1$ pairs of complex conjugated numbers[4] [99]. An example of the main spectrum of a genus-4 solution is given in Fig. 2.1, where the complex conjugated values of the main spectrum are joined with arcs. If all $\{\lambda_j\}_{j=0}^N$ are different, they provide a non-degenerated solution [100]. The given main spectrum generates two-sheeted Riemann surface Γ:

$$\Gamma : \left\{ (\lambda, P), P^2 = \prod_{j=0}^{N} (\lambda - \lambda_j)(\lambda - \lambda_j^*),\ \lambda, P \in \mathbb{C} \right\}. \tag{2.6}$$

[4] for simplicity denoted as $\{\lambda_j\}_{j=0}^N$ later in the text, where all λ_j lie in the upper half of the complex plane.

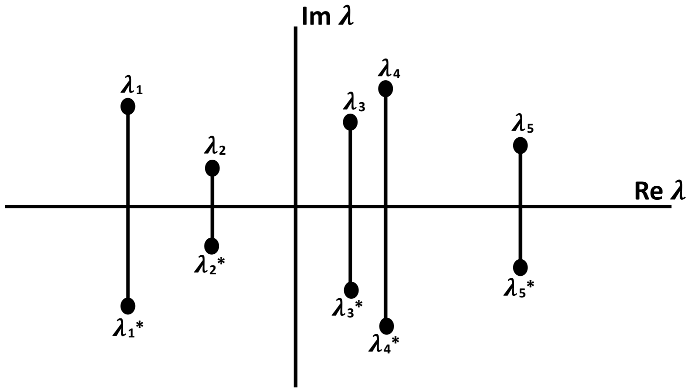

Figure 2.1: An example of the main spectrum of a genus-4 solution.

A topological genus of this Riemann surface coincides with the genus of the solution corresponding to the main spectrum.

The auxiliary spectrum $\{\mu_j = \mu_j(t_0, z_0)\}_{j=1}^N$ is initial values of corresponding *hyperelliptic modes* $\{\mu_j(t, z)\}_{j=1}^N$. These modes are a dynamic part of the spectrum; they are functions of t and z variables. Therefore, the auxiliary spectrum changes while the solution $q(t, z)$ evolves according to the NLSE. These hyperelliptic modes are defined on the two-sheeted Riemann surface (2.6), unique for the given main spectrum, and generate closed trajectories on this surface when $q(t, z)$ is periodic. Each μ_j is attributed with a sheet index $\sigma_j = \sigma_j(t_0, z_0)$ to specify a sheet of the Riemann surface where μ_j is located initially. As μ_j moves over the Riemann surface, it possibly switches from one sheet of the surface to another. In this case, the corresponding index σ_j changes its sign. The discrete functions $\{\sigma_j(t, z)\}_{j=1}^N$ determine the evolution of these parameters. Therefore, the sets $\{\mu_j(t, z)\}_{j=1}^N$ and $\{\sigma_j(t, z)\}_{j=1}^N$ completely describe the dynamic of the solution. It is worth noting that not each set of complex numbers can be an auxiliary spectrum to provide a finite-genus solution (see more details in [100, 101]).

For the given values $\{\mu_j\}_{j=1}^N$ and $\{\sigma_j\}_{j=1}^N$ the functions $\{\mu_j(t, z)\}_{j=1}^N$, $\{\sigma_j(t, z)\}_{j=1}^N$ are retrieved from the system of equations:

$$\partial_t \mu_j = -2i\sigma_j \frac{\sqrt{\prod_{k=0}^{2N+1} (\mu_j - \lambda_k)}}{\prod_{\substack{m=1 \\ m\neq j}}^{N} (\mu_j - \mu_m)}, \tag{2.7}$$

$$\partial_z \mu_j = -2 \left(\sum_{\substack{m=1 \\ m\neq j}}^{N} \mu_m - \frac{1}{2} \sum_{k=0}^{2N+1} \lambda_k \right) \partial_t \mu_j, \tag{2.8}$$

where specific $\sigma_j(t, z)$ switches a sign when $\mu_j(t, z)$ reaches an arc joining any pair from $\{\lambda_j, \lambda_j^*\}_{j=0}^N$. It corresponds to the move from one sheet of the Riemann surface to another. Therefore, the functions $\{\sigma_j(t,z)\}_{j=1}^N$ can be recovered from tracking if the corresponding μ_j transferred to another sheet on the current step of the solving procedure or not. Another couple of equations provides the solution $q(t,z)$ with the initial value $q(t_0,z_0)$:

$$\partial_t \ln q = 2i\left(\sum_{j=1}^{N}\mu_j - \frac{1}{2}\sum_{k=0}^{2N+1}\lambda_k\right), \tag{2.9}$$

$$\partial_z \ln q = 2i\left(\sum_{\substack{j,k=0\\ j>k}}^{2N+1}\lambda_j\lambda_k - \frac{3}{4}\left(\sum_{k=0}^{2N+1}\lambda_k\right)^2\right)+4i\left(\frac{1}{2}\left(\sum_{k=0}^{2N+1}\lambda_k\right)\left(\sum_{j=1}^{N}\mu_j\right) - \sum_{\substack{j,k=1\\ j>k}}^{N}\mu_j\mu_k\right). \tag{2.10}$$

Therefore, if the main spectrum is provided, it generates the Riemann surface. With known $\{\mu_j\}_{j=1}^N$ and $\{\sigma_j\}_{j=1}^N$, the hyperelliptic modes can be found from eq. (2.7) and (2.8). Finally, eq. (2.9) and (2.10) give the solution $q(t,z)$ from initial value $q(t_0,z_0)$. However, because the hyperelliptic modes involved in these equations are defined on the two-sheeted Riemann surface providing alternating eq. (2.7), the analysis of the system is significantly complicated.

Inverse problem and the Riemann-theta function formalism

Alternatively, an approach based on the Riemann theta function can be applied to construct a finite-genus solution. The Riemann theta function is:

$$\Theta(\boldsymbol{z}|\boldsymbol{\tau}) = \sum_{\boldsymbol{m}\in\mathbb{Z}^N} e^{\pi i\boldsymbol{m}^T\boldsymbol{\tau}\boldsymbol{m}+2\pi i\boldsymbol{m}^T\boldsymbol{z}},\ \boldsymbol{z}\in\mathbb{C}^N. \tag{2.11}$$

It is a reduction of a multidimensional Fourier series because, in the general case, Fourier coefficients are elements of a tensor of rank N, but $\boldsymbol{\tau}$ is just a matrix [4]. $\boldsymbol{\tau}$ is a $N\times N$ symmetric matrix of periods; its imaginary part is positively defined to provide convergence of the series. The Riemann theta function is periodic in all components of the vector-valued argument $\boldsymbol{z}$. Summation is performed over integer-valued parameters $\boldsymbol{m} = \{m_1, m_2, \dots, m_N\}$ in N dimensional space. The series is truncated after some m_j^{max} to provide predefined calculation accuracy. The Riemann theta function is an important tool in the analysis of nonlinear differential equations and the periodic nonlinear Fourier transform [4, 102, 103].

Finite-genus solutions to the focusing NLSE can be expressed through the Riemann theta function:

$$q(t,z) = A_0\frac{\Theta(\frac{\pi}{2}(\boldsymbol{\omega}t+\boldsymbol{k}z+\boldsymbol{\delta}^-)|\boldsymbol{\tau})}{\Theta(\frac{\pi}{2}(\boldsymbol{\omega}t+\boldsymbol{k}z+\boldsymbol{\delta}^+)|\boldsymbol{\tau})}e^{i\omega_0 t+ik_0 z}. \tag{2.12}$$

The matrix of periods $\boldsymbol{\tau}$, as well as the vectors of temporal and spatial frequencies $\boldsymbol{\omega}, \boldsymbol{k}$ are a result of the integration of holomorphic differentials over a canonical homology basis defined on the Riemann surface. Therefore, this matrix and vectors are unique for the given main spectrum. $\boldsymbol{\delta}^+, \boldsymbol{\delta}^-$ depend on the auxiliary spectrum also and are calculated as integrals on the Riemann surface. Finally, the parameters A_0, ω_0, k_0 are determined from the asymptotic expansion of the Abel integrals. The matrix $\boldsymbol{\tau}$ has the size of $N \times N$ while vectors $\boldsymbol{\omega}, \boldsymbol{k}, \boldsymbol{\delta}^+, \boldsymbol{\delta}^-$ have the size of N, where N is a genus of the solution. For this reason, high-genus solutions are associated with multidimensional Riemann theta functions that are inapplicable for numerical analysis. A detailed procedure and explanation of the approach to construct finite-genus solutions with the Riemann theta function are given in [4, 31].

Calculating the parameters in eq. (2.12) is mathematically involved and is beyond my review purpose. However, because the temporal frequencies $\boldsymbol{\omega}$ have an exceptional role in the following analysis, their detailed calculation is considered. The given main spectrum $\{\lambda_j\}_{j=0}^{N}$ produces the Riemann surface (2.6) with the basis of holomorphic differentials:

$$dU_j = \frac{\lambda^{j-1} d\lambda}{P(\lambda)}, \quad j = 1, \ldots, N, \tag{2.13}$$

with $P(\lambda)$ being values of (2.6) with a corresponding sign of the squared root. At the same time, a canonical basis of oriented contours on the Riemann surface can be introduced $\{a_j, b_j\}_{j=1}^{N}$. The basis is canonical if all $\{a_j\}_{j=1}^{N}$ do not intersect each other and $\{b_j\}_{j=1}^{N}$ also do not intersect each other; a_j and b_j cross only once and, finally, no one of the contours can be continuously deformed to other one or contracted to zero. Any closed curve on the Riemann surface is decomposed into that canonical basis. Then, the temporal frequencies are:

$$\omega_j = -\frac{4i}{\pi} \boldsymbol{A}^{-1}_{j,N}, \tag{2.14}$$

where matrix $\boldsymbol{A}$ is combined from integrals over the basis contours $\{a_j\}_{j=1}^{N}$:

$$A_{j,k} = \int_{a_k} dU_j, \quad j, k = 1, \ldots, N. \tag{2.15}$$

Being calculated the frequencies (2.14) constitute the vector $\boldsymbol{\omega}$ in eq. (2.12). For more details about this calculation and also definitions of the other parameters in eq. (2.12) see [31, 33, 99].

Definition 1. *The closed-form expression for a finite-genus solution (2.12) is convenient to introduce a **nonlinear harmonic** or **nonlinear mode**. A genus-N solution is attributed with the vectors $\boldsymbol{\omega}$ and $\boldsymbol{k}$ with N elements each. These vectors' components ω_j, k_j are temporal and spatial frequencies. The contribution associated with each pair ω_j, k_j to the overall genus-N solution is a nonlinear harmonic. An additional $(N+1)$th mode must be associated with $\omega_0 = 0, k_0 = 0$ that do not contribute to (2.12). The rea-*

son for introducing this harmonic is obvious when considering a genus-0 solution: while $N = 0$, it consists of a single nonlinear mode[5]. *In this sense, a genus-N solution $q(t, z)$ consists of $N + 1$ nonlinear harmonics or nonlinear modes. This definition is useful for describing finite-genus solutions and will be actively used further in the text.*

2.2.2 The Riemann-Hilbert problem approach

An alternative approach to construct finite-genus solutions is the Riemann-Hilbert problem (RHP) formalism [104]. In this method, a genus-N solution is parametrized with:

- *Main spectrum:* $N + 1$ pairs of complex conjugated number $\{\lambda_j, \lambda_j^*\}_{j=0}^N \in \mathbb{C}$; the same main spectrum as defined in the algebro-geometric approach;
- *Phases:* $N + 1$ values $\{\phi_j\}_{j=0}^N \in [0, 2\pi)$.

This set of parameters uniquely determines a genus-N solution to the NLSE. To calculate the solution, the following algorithm can be applied:

1. The given main spectrum $\{\lambda_j, \lambda_j^*\}_{j=0}^N$ forms the oriented contour $\Gamma = \cup_{j=0}^N \Gamma_j$, where $\Gamma_j = (\lambda_j, \lambda_j^*)$ is the arc connecting λ_j and λ_j^*.
2. For the given contour Γ a jump matrix is introduced:
$$\boldsymbol{J}(t, z, \lambda) = \begin{pmatrix} 0 & ie^{-i(\phi_j + 2\lambda t + 4\lambda^2 z)} \\ ie^{i(\phi_j + 2\lambda t + 4\lambda^2 z)} & 0 \end{pmatrix}, \tag{2.16}$$
where $\lambda \in \Gamma_j$, $j = 0, \dots, N$. This matrix contains the phases $\{\phi_j\}_{j=0}^N$ as parameters.
3. For the given oriented contour Γ and $\boldsymbol{J}(t, z, \lambda)$ an analytic 2×2 matrix function $\boldsymbol{\Psi}(t, z, \lambda)$ of variable $\lambda \in \mathbb{C} \setminus \Gamma$ and real-valued parameters t, z such that:
    - the limiting values $\boldsymbol{\Psi}_\pm(t, z, \lambda)$ as λ approaches the contour Γ from both sides hold $\boldsymbol{\Psi}_+(t, z, \lambda) = \boldsymbol{\Psi}_-(t, z, \lambda)\boldsymbol{J}(t, z, \lambda),\ \lambda \in \Gamma$;
    - $\boldsymbol{\Psi}(t, z, \lambda)$ has singularities no higher than the inverse fourth root at λ_j and λ_j^*;
    - and $\boldsymbol{\Psi}(t, z, \lambda) \to \boldsymbol{I}$ as $\lambda \to \infty$;

    is a solution to the Riemann-Hilbert problem.

[5] It becomes clearer when considering FGS's parametrization with the RHP provided in the following section.

Being evaluated $\boldsymbol{\Psi}(t,z,\lambda)$ provides a finite-genus solution to the NLSE from its asymptotic behavior as $\lambda \to \infty$:

$$\boldsymbol{\Psi}(t,z,\lambda) = \boldsymbol{I} + \frac{\boldsymbol{\Psi_1}(t,z)}{\lambda} + \dots, \tag{2.17}$$

and, finally, with $[\boldsymbol{\Psi_1}(t,z)]_{1,2}$ being the $1,2$ entry of the matrix-valued function $\boldsymbol{\Psi_1}(t,z)$:

$$q(t,z) = 2i[\boldsymbol{\Psi_1}(t,z)]_{1,2}. \tag{2.18}$$

This algorithm constitute *the inverse problem* of the NFT for finite-genus solutions: starting from $\{\lambda_j\}_{j=0}^N$ and $\{\phi_j\}_{j=0}^N$ the corresponding finite-genus solution $q(t,z)$ can be constructed. A detailed description of the procedure is provided in [30, 34, 99, 105].

Alternatively, a finite-genus solution can be expressed through the solution of another RHP with the jump matrix that contains the temporal and special frequencies in an explicit way as in eq. (2.12). Let's introduce an analytic function in $\mathbb{C} \setminus \Gamma$:

$$\omega(\lambda) = \prod_{j=0}^{N} \sqrt{(\lambda - \lambda_j)(\lambda - \lambda_j^*)}. \tag{2.19}$$

Each contour Γ_j oriented from λ_j^* to λ_j (all λ_j are such that $\operatorname{Im}\lambda_j > 0$), and $\omega^+(\lambda)$ is a liming value of $\omega(\lambda)$ as λ reaches Γ from the right side. The following $N \times N$ matrix $\boldsymbol{K}$ with elements:

$$K_{m,j} = \int_{\Gamma_j} \frac{\lambda^{m-1} d\lambda}{\omega^+(\lambda)}, \quad m,j = 1,\dots,N, \tag{2.20}$$

determines vectors of temporal $\boldsymbol{C^f} = (C_1^f, C_2^f, \dots, C_N^f)^T$ and spatial frequencies $\boldsymbol{C^g} = (C_1^g, C_2^g, \dots, C_N^g)^T$ through the systems of equations:

$$\boldsymbol{K} \cdot \boldsymbol{C^f} = [0,0,\dots,-2\pi i]^T, \tag{2.21}$$

$$\boldsymbol{K} \cdot \boldsymbol{C^g} = [0,0,\dots,-8\pi i, -4\pi i \sum_{j=0}^{N} (\lambda_j + \lambda_j^*)]^T. \tag{2.22}$$

Calculating values f_0 and g_0 from asymptotic behaviour of the following analytic functions in $\mathbb{C} \setminus \Gamma$:

$$f(\lambda) = \frac{\omega(\lambda)}{2\pi i} \sum_{j=1}^{N} \int_{\Gamma_j} \frac{C_j^f d\xi}{\omega^+(\xi)(\xi - \lambda)} = \lambda + f_0 + O(1/\lambda), \tag{2.23}$$

$$g(\lambda) = \frac{\omega(\lambda)}{2\pi i} \sum_{j=1}^{N} \int_{\Gamma_j} \frac{C_j^g d\xi}{\omega^+(\xi)(\xi - \lambda)} = 2\lambda^2 + g_0 + O(1/\lambda) \tag{2.24}$$

the finite-genus solution can be constructed as:

$$q(t,z) = 2i[\mathbf{\Phi_1}(t,z)]_{1,2} e^{2if_0 t + 2ig_0 z}, \tag{2.25}$$

where $\mathbf{\Phi_1}(t,z)$ is defined by the solution of the Riemann-Hilbert problem described above (from the decomposition of asymptotic behavior as in eq. (2.17)) on the same contour Γ with jump matrix:

$$\hat{\boldsymbol{J}}(t,z,\lambda) = \begin{pmatrix} 0 & ie^{-i(\phi_j + C_j^f t + C_j^g z)} \\ ie^{i(\phi_j + C_j^f t + C_j^g z)} & 0 \end{pmatrix}, \quad \lambda \in \Gamma_j, \tag{2.26}$$

with parameters $C_0^f = 0$ and $C_0^g = 0$.

The real-valued entries in eq. (2.26) $\{C_j^f\}_{j=1}^N$ and $\{C_j^g\}_{j=1}^N$ are the temporal and spacial frequencies of nonlinear modes. Together with constants f_0 and g_0, they dictate the finite-genus solution's dependence on t and z. It is worth noticing that f_0 can be set to zero by shifting the main spectrum along the real axis [34]. The main spectrum $\{\lambda_j\}_{j=0}^N$ fully defines all these parameters.

Note: *Similar to the Definition 1, in the RHP approach the nonlinear harmonics can be associated with particular values C_j^f, C_j^g and ϕ_j. Moreover, these parameters are the entries of a jump matrix, defined independently for each gap formed by a pair $\{\lambda_j, \lambda_j^*\}$. Therefore, a finite-genus solution consists of $N+1$ nonlinear modes, each associated with a particular pair $\{\lambda_j, \lambda_j^*\}$*, frequencies C_j^f, C_j^g and phase ϕ_j.

Finally, the scattering data characterizing a genus-N solution through the RHP approach have a trivial evolution. The main spectrum $\{\lambda_j\}_{j=0}^N$ is invariant while a signal $q(t,z)$ evolves according to the NLSE. At the same time, the phase of each nonlinear harmonic:

$$\phi_j(z) = \phi_j(0) + (C_j^g - 2g_0)z, \tag{2.27}$$

where z is a normalized propagation distance, C_j^g and g_0 are the corresponding spatial frequencies as specified above. A rigorous mathematical description of the described procedure is given in [J1]. Also, more details about calculating the finite-genus solutions to the focusing NLSE with the RHP approach see in [30, 34, 105, 96].

2.2.3 The direct problem and monodromy matrix

The scattering data for both methods, the AG approach 2.2.1 and the RHP 2.2.2, can be found through monodromy matrix formalism. However, this method is applicable only to periodic solutions (2.5) for which the period T is defined. It, in turn, requires commensurability of the frequencies of nonlinear harmonics that are elements of the vector $\boldsymbol{\omega}$ in eq. (2.12) in the AG approach or $\boldsymbol{C^f}$ in the RHP method.

Definition 2. *For a genus-N solution the elements of the vector $\boldsymbol{\omega} = [\omega_1, \omega_2, \dots, \omega_N]$ are frequencies of nonlinear harmonics. We say that these frequencies are* ***commensurable*** *if there exists a set of arbitrary integer numbers $\{k_1, k_2, \dots, k_N\}$ such that $k_1\omega_1 = k_2\omega_2 = \dots = k_N\omega_N = \omega_{com}$. In other words, the frequencies are commensurable if a common multiple ω_{com} exists. The same is true for the periods of nonlinear harmonics $\{T_j = 2\pi/\omega_j\}_{j=1}^{N}$. The periods are commensurable if a common multiple T_{com} is defined as provided in Fig. 2.2.*

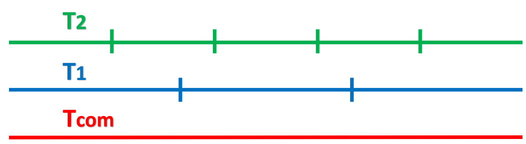

Figure 2.2: The schematic of commensurable periods with T_{com} being a common multiple for all others.

Monodromy matrix can be applied to calculate the scattering data for the subset of finite-genus solutions of which nonlinear modes have commensurable frequencies, Definition 2. Let's denote such a periodic solution with a period T as $q_p(t,z)$. The solution $\boldsymbol{\Phi}(t,z,\lambda)$ of the Zakharov-Shabat system (1.2) on the period T with potential $q_p(t,z)$ and the initial condition $\boldsymbol{\Phi}(0,z,\lambda) = \boldsymbol{I}$ (the unit matrix) provides the monodromy matrix:

$$\boldsymbol{M}(z,\lambda) = \boldsymbol{\Phi}(T,z,\lambda). \tag{2.28}$$

The main spectrum $\{\lambda_j, \lambda_j^*\}_{j=0}^{N}$ is simple zeros of the equation:

$$(\boldsymbol{M}_{1,1} + \boldsymbol{M}_{2,2})^2 - 4 = 0. \tag{2.29}$$

At the same time, the auxiliary spectrum $\{\mu_j\}_{j=1}^{N}$ is the double zeros of

$$\boldsymbol{M}_{1,2} = 0 \tag{2.30}$$

or simple zeros of eq. (2.30) that do not coincide with double zeros of eq. (2.29) [34, 100, 105]. To calculate the sheet indexes that take values from $\{-1, 1\}$, the following expression can be used [4]:

$$\sigma_j = \left. \frac{i\,\mathrm{Im}(\boldsymbol{M}_{1,1}(\lambda))}{\left(\boldsymbol{M}_{1,2}^*(\lambda)\boldsymbol{M}_{2,1}(\lambda) - \mathrm{Im}^2(\boldsymbol{M}_{1,1}(\lambda))\right)^{1/2}} \right|_{\lambda=\mu_j} \tag{2.31}$$

To find the phases $\{\phi_j\}_{j=0}^{N}$ from the RHP approach, the following system must be solved:

$$\boldsymbol{K} \cdot \boldsymbol{\phi} = \boldsymbol{B}, \tag{2.32}$$

where the $(N+1) \times (N+1)$ matrix $\boldsymbol{K}$ is calculated as (the same as eq. (2.20) but including $j, m = 0$):

$$K_{mj} = \int_{\Gamma_j} \frac{\lambda^m d\lambda}{\omega^+(\lambda)}, \quad j, m = 0, \ldots, N. \tag{2.33}$$

The elements of the vector $\boldsymbol{B}$ are:

$$B_m = -i \sum_{j=0}^{N} \int_{\Gamma_j} \frac{P(\lambda)\lambda^m d\lambda}{\omega^+(\lambda)}, \quad m = 0, \ldots, N, \tag{2.34}$$

with the function:

$$P(\lambda) = \log\left(\sqrt{-\frac{\boldsymbol{M}_{1,2}(\lambda)}{\boldsymbol{M}_{2,1}(\lambda)}} \prod_{j=1}^{N} \frac{\lambda - \mu_j^*}{\lambda - \mu_j}\right). \tag{2.35}$$

Equations (2.29), (2.30), and (2.31) provide the main spectrum, auxiliary spectrum, and sheet indexes that together with $q(t_0, z_0)$ constitute the scattering data for finite-genus solutions in the algebro-geometric approach and, therefore, the solution of the direct problem of the periodic NFT. At the same time, the solution of the system (2.32) is the phases. The main spectrum (eq. (2.29)) and the phases are the full set of spectral parameters for the RHP approach. However, the application of the monodromy matrix formalism is restricted by the finite-genus solutions to those that meet the commensurability condition (see Definition 2). More details are provided in the following section. A comprehensive explanation of how to solve the direct problem can be found in [34, 100, 105, 96].

2.2.4 Implementation of the periodic NFT

Two methods to implement the periodic NFT, the algebro-geometric and the Riemann-Hilbert problem approaches, have been described above. Despite their rigorous mathematical definitions, both have peculiarities that complicate practical implementation if unknown. Their limitations and nuances of numerical realizations are described in this section.

Formulation of the algebro-geometric approach 2.2.1 allows a straightforward implementation of the inverse problem (finding a signal $q(t, z)$) through solving the system of equations (2.7) – (2.10). In practice, such calculation is complicated because the auxiliary spectrum $\{\mu_j\}_{j=1}^{N}$ is defined on the Riemann surface that leads to the alternating coefficient σ_j in eq. (2.7). This change of sign corresponds to the transfer of auxiliary spectrum points from one sheet of the Riemann surface to another.

Alternatively, the inverse problem can be solved by using the Riemann theta function. For given scattering data 2.2.1, the parameters of the Riemann theta function to calculate a finite-genus solution can be found from the Abel map [31, 99]. The evaluation of the Riemann theta function (2.11) provides a finite-genus solution (2.12). However, the function (2.11) is presented as a multi-dimensional Fourier series that complicates its numerical evaluation [4, 103]. The problem becomes more valuable for a numerical calculation of high-genus solutions: computational complexity grows exponentially with the number of dimensions, making straightforward analysis completely impractical. The method to calculate such a high-dimensional Riemann theta function with a tensor-train-based algorithm is given in [102]. The numerical approaches to the Riemann surfaces and available software are described in [106].

Solving the direct problem with the monodromy matrix approach bounds the class of finite-genus solutions for which the scattering data can be calculated. In the set of FGS to the focusing NLSE, the subset of ones that have commensurable frequencies of their nonlinear modes can be processed with the monodromy matrix, Fig. 2.3. Such limita-

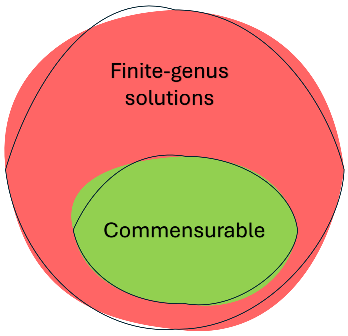

Figure 2.3: The subset of finite-genus solutions with commensurable frequencies of their nonlinear harmonics (see Definition 2).

tions on frequencies, in turn, impose restrictions on the main spectrum configurations. In the general case, the vector of frequencies, $\boldsymbol{\omega}$ in eq. (2.12), is a function of the real and imaginary parts of the main spectrum. FGS shows a nonlinear nature, the degree of which grows with an increase of $\operatorname{Im}\lambda_j$. On the other hand, in the linear limit, when the imaginary part of the MS points tends to be zero, the modes demonstrate linear behavior, and their frequencies and amplitudes do not depend on each other. Moreover, in such a limit, the main spectrum approximates amplitudes of the linear Fourier spectrum $|q(\omega)|$ [Appendix, 34]. Therefore, constructing a periodic finite-genus solution is possible in a quasilinear regime when $\{\operatorname{Im}\lambda_j\}_{j=0}^{N}$ is close to zero. Using MS with a small enough imaginary part and adjusting it for each nonlinear mode independently, it is possible to provide frequency commensurability and, finally, a finite-genus solution to be periodic. This approach was used in fibre-optic communication applications of FGS [33, 34]. At the same time, commensurable finite-genus solutions (having commensurable frequencies of nonlinear harmonics) that are not restricted by quasilinearity conditions were found only for particular configurations of the main spectrum [107, 108]. In the general case of an arbitrary MS, the commensurability is not guaranteed. A trick to build a commensurable finite-genus solution with an arbitrarily high imaginary part of the main spectrum, where it is calculated for an existing periodic function, is given in [Appendix, 34]. However, in this approach, the frequencies and amplitude of separate harmonics are uncontrollable. To summarize, applying the monodromy matrix formalism reduces the set of finite-genus solutions that can be processed with the periodic NFT framework. The inverse problem and evolution can still be solved for an arbitrary main spectrum, but the direct problem is only for the finite-genus solutions with commensurable frequencies.

Regarding the RHP approach, the inverse problem can be solved numerically for any main spectrum configuration. Having a contour on the complex plane and a jump matrix, one can solve the corresponding Riemann-Hilbert problem (see section 2.2.2). The numerical approach and its practical implementation are given in [92, 95, 109, 110].

The limitation on the main spectrum also is applicable to the direct problem in the RHP framework. This is because the direct problem in the approach also exploits the monodromy matrix formalism. The main spectrum in the RHP and the algebro-geometric method is the same and generates the same frequencies of nonlinear harmonics. Hence, the values $\{C_j^f\}_{j=1}^N$ must be commensurable as well. Again, using the full periodic NFT framework is possible for the restricted subset of finite-genus solutions (as depicted in Fig. 2.3), while the inverse problem itself and the phase evolution can be solved for an arbitrary main spectrum.

Specifically, implementing the periodic NFT with the RHP approach suffers from additional restrictions. Among the finite-genus solutions that have commensurable frequencies and for which monodromy matrix formalism can be applied, the signals with the phases $\{\phi_j < \pi\}_{j=0}^N$ can be processed only. An algorithm with such a restriction was used in [34], where the phases of FGS were used to carry information in a fibre-optic communication system. Using the phases by modulo π reduced the system's capacity twice. This limitation on the phase originates from the formula (2.35) that is fair for the case when all auxiliary spectrum points are located on one sheet of the Riemann surface. Below an approach to resolve this restriction is described.

2.2.5 The analytical solving of the direct problem via the RHP method

In section 2.2.2, the approach to construct the finite-genus solutions to the NLSE through the RHP method is described. For the given scattering data, the main spectrum, and the phases, the procedure provides signal $q(t, z)$. At the same time, the opposite transformation, which is retrieving the spectral data, can be implemented by means of monodromy matrix formalism. However, this method was restricted to operate with the phases defined on the interval $[0, \pi]$. To lift this additional limitation and use the full range of phases that is especially important in practice, the described approach to solve the direct problem for FGS was expanded in [J3]. It was reported that retrieving the phases of finite-genus solutions is possible by constructing another Riemann-Hilbert problem with a given signal $q(t, z)$ and the main spectrum as initial data.

To build this new RHP the following needs to be considered. First, the initial boundary value problem to the NLSE can be solved for the periodic initial condition (eq. (2.5)) through an RHP [111, 112]. To build that solution, the periodic signal $q(t, z)$ is exploited as initial data. The contour in this RHP formulation is the union of real and imaginary axes with a number of finite arcs in the complex plane. The jump matrix, in turn, is constructed using the scattering matrix of the Zahkarov-Shabat system. Second, the RHP for the initial boundary value problem, if transformed to the problem in the form 2.2.2, provides the desired phases as entries of a jump matrix. More details about that approach can be found in [J3].

The following examples demonstrate the method's feasibility. First, the genus-1 solution with the main spectrum $\lambda = \{0.2780 + i, 1.2780 + i\}$ and different initial phases was exploited. The inverse problem was solved for these scattering data to provide the corresponding solutions $q(t)$, as given in section 2.2.2. Then, the phases were calculated using the developed method. The retrieved phases for different initial values are in Table 2.1. The error of recovery was smaller than 10^{-3}. For these examples, the original and recovered phases, as well as the main and auxiliary spectra and signals $q(t)$, are depicted in Fig. 2.4.

Ex	Original phases		$q(0,0)$	Aux. spectrum	Recovered phases	
	ϕ_0	ϕ_1	q	μ	ϕ_0	ϕ_1
1	0.4	0.8	$1.5844 - 1.0839i$	$0.7780 + 0.3163i$	0.4005	0.7995
2	3.5416	3.9416	$-1.5844 + 1.0839i$	$-0.7780 - 0.3163i$	3.5420	3.9411
3	0.4	3.9416	$0.1463 + 0.2139i$	$0.7780 - 3.9526i$	0.4000	3.9416

Table 2.1: The original phases, values of $q(0, 0)$, auxiliary spectra and recovered phases of three different genus-1 solutions used to demonstrate the method (cited from [J3]).

Second, a genus-2 solution was also considered. The following configuration of the main spectrum $\lambda = \{-1 + 3i, 5i, 1 + 3i\}$ provides a periodic finite-genus solution [107]. The initial phases were $\phi = \{0.1, 0.2, 0.3\}$ while the auxiliary spectrum $\mu = \{2.1061 + 0.4161i, 2.1061 + 0.4161i$. The recovered phases were $\phi = \{0.1007, 0.2000, 0.2993\}$, agreeing with the original values at 10^{-3}. The main and auxiliary spectra, initial and recovered phases, as well as $q(t)$ are shown in Fig. 2.5.

In conclusion, the method briefly described here was developed to solve the direct problem of finite-genus solution in the RHP framework. In contrast to the previous approach, this one is able to process the FGS parametrized with a full range of phases. It was demonstrated with a few simple examples, genus-1 and genus-2 solutions. However, the expanding of the approach for a case of higher genera can be conducted if required. The full description is provided in [J3].

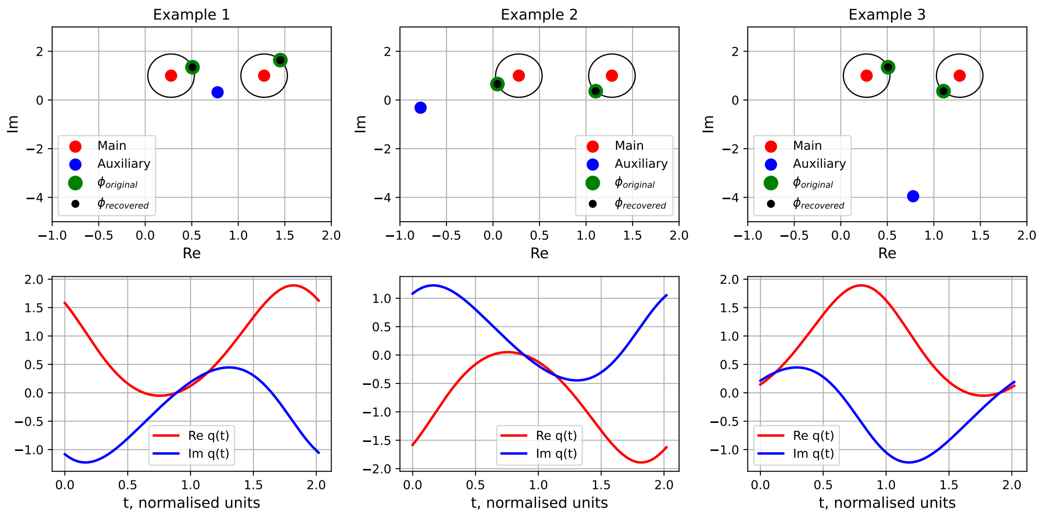

Figure 2.4: Three genus-1 signals used to demonstrate the method. The original and recovered phases, main and auxiliary spectra, and signals $q(t)$ are provided; see also Table 2.1. (Cited from [J3].)

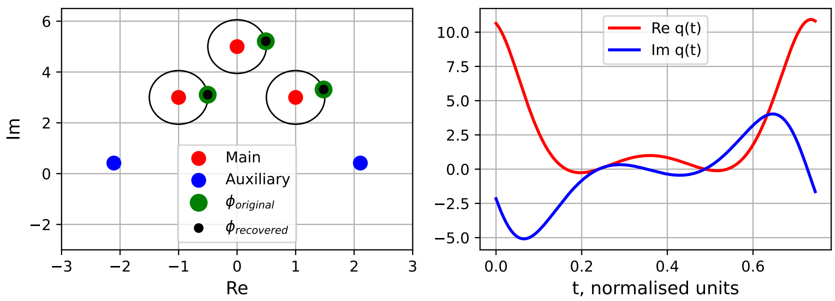

Figure 2.5: The main and auxiliary spectra, original and recovered phases, as well as the signal $q(t)$, for an example genus-2 solution. (Cited from [J3].)

2.3 Nonlinear Fourier transform for finite-genus solutions of a generic type with a convolutional neural network

Consider an arbitrary main spectrum $\{\lambda_j, \lambda_j^*\}_{j=0}^N$ providing a genus-N solution. The solution consists of $N+1$ nonlinear modes (Definition 1), each with a period T_j[6] and periods $\{T_j\}_{j=1}^N$ are not commensurable (Definition 2) or a common period T_{com} is too big to be processed with any existed algorithm. At the same time, there is no need for a solution $q(t, z_0)$[7] to be defined over the entire interval T_{com} to solve the direct problem of the periodic NFT and retrieve the scattering data from the signal. Among all periods $\{T_j\}_{j=1}^N$, the maximum $T_{max} = \max_j T_j$ can be chosen as an interval on which the signal is processed: $q(t, z_0)$ such that $t \in [0, T_{max}]$. Because T_{max} is the maximum

[6]the period corresponding to $\omega_0 = 0$ is not defined and is not considered.
[7]$q(t, z_0)$ is taken at a specific value of z_0.

period among all nonlinear modes, it guarantees that the signal $q(t, z_0)$ contains at least one period for any nonlinear harmonic. It, in turn, means that all necessary information about each nonlinear mode is contained in the signal $q(t, z_0)$ defined on T_{max}.

Definition 3. *Let's call a finite-genus solution* ***periodic*** *if it has commensurable frequencies of nonlinear harmonics and is defined on T_{com}. Other words, $q(t, z_0) = q(t + T_{com}, z_0)$ and $t \in [0, T_{com}]$. Therefore, the direct problem for such solutions can be solved with the monodromy matrix.*

Definition 4. *The set of* ***finite-genus solutions of a generic type (FGS-GT)*** *includes:*

- *the finite-genus solutions that have non-commensurable frequencies of nonlinear harmonics, which means that T_{com} does not exist;*
- *the finite-genus solutions defined on the interval smaller than T_{com} (more practical scenario for an arbitrary main spectrum);*
- *the periodic finite-genus solutions (Definition 3).*

Therefore, the finite-genus solutions of the generic type are generalizations of the periodic finite-genus solutions.

To the best of my knowledge, there is still no theoretical approach to retrieve the scattering data of finite-genus solutions of a generic type. The only existing method is to apply monodromy matrix formalism (described in the previous sections), which works only with periodic signals. At the same time, machine learning and artificial neural networks (NN) offer a powerful tool to approximate any transformation despite its unknown explicit form. The universal approximation theorem guarantees the mapping of any function with a neural network [113]. Thus, this property of NN can be used to solve the direct problem for a finite-genus solution of a generic type.

Later in this chapter, the method to solve the direct problem for FGS-GT (see Definition 4) through neural networks is described. The RHP approach is exploited to parameterize finite-genus solutions for which the scattering data are the main spectrum and the phases. To provide better performance, the neural network solves a simplified task: for the given main spectrum, it predicts the phases only. In other words, the NN model's input is a discretized signal $q(t, z_0)$ defined on the interval $[0, T_{max}]$ and with a fixed main spectrum. The output of the neural network is $N + 1$ phases.

It is worth remembering how the main spectrum configuration defines the property of the corresponding finite-genus solution. In the linear limit, when the imaginary part of the main spectrum is close to zero, the frequencies of nonlinear harmonics are defined by the real part of the main spectrum, while the imaginary part provides the amplitudes of the modes. With the growth of $\mathrm{Im}[\lambda_j]$, the nonlinear nature of signals becomes significant, and the imaginary part of the main spectrum also impacts the frequencies. In the last scenario, more nonlinearity-involved behavior is expected; the nonlinear modes are not independent; they influence each other, determining mutual properties. The investigation of signals' frequencies and amplitudes in dependence on the main spectrum structure is given in [34].

2.3.1 Training data

A genus-N solution can be parameterized by the $N+1$ points of the main spectrum and $N+1$ phases. As mentioned, a neural network was trained to predict the phases of finite-genus solutions for a fixed main spectrum. One of the configurations was a genus-4 solution with the profile and the main spectrum provided in Fig. 2.6. Such a choice is explained by its intermediate properties: a genus-4 solution contains 5 nonlinear modes and, therefore, is not a trivial signal among finite-genus solutions, but it is still not complex to be quickly and effectively generated with the inverse problem algorithm in the frame of the RHP approach. At the same time, the imaginary part $\mathrm{Im}[\lambda_j] = 1$ used in this example ensures nonlinear properties.

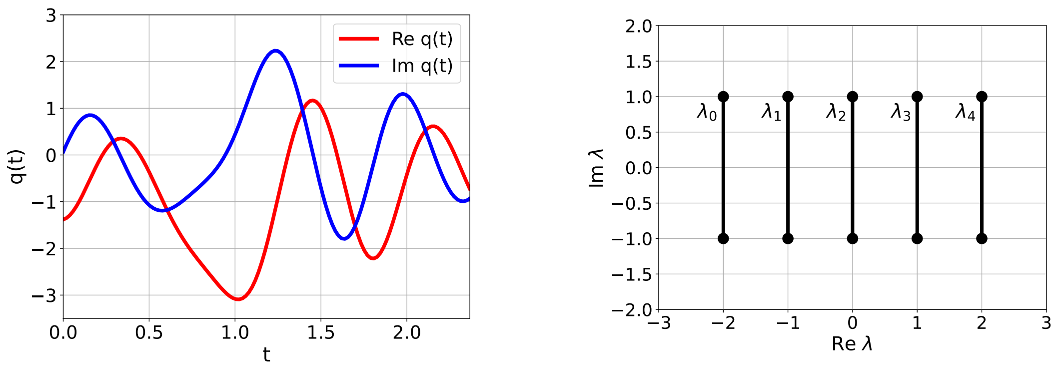

Figure 2.6: (left) The genus-4 solution in the time domain with the phases $\{\pi, \frac{\pi}{6}, \frac{3\pi}{2}, \frac{2\pi}{3}, \frac{7\pi}{6}\}$ (t and $q(t)$ are in dimensionless units) and (right) the corresponding main spectrum (cited from [J1]).

The other configurations of the main spectrum were used to demonstrate the feasibility and flexibility of the method. Among them was the genus-4 solution with the imaginary part of the main spectrum $\mathrm{Im}[\lambda_j] = 5$. A larger imaginary part provides more nonlinearity and a stronger dependence of nonlinear harmonics on each other. Also, to estimate the method's performance with higher genus solutions, genus-8 was exploited. The three configurations of the main spectrum for which a neural network was trained independently are provided in Table 2.2.

The signals $q(t, z_0)$ were generated with the Riemann-Hilbert problem approach for all main spectra under the study with random phases from 0 to 2π distributed uniformly. The details of the methods are provided in the section 2.2.2. The pairs of the main spectrum points $\{\lambda_j, \lambda_j^*\}_{j=0}^N$ generate the contour Γ; see Fig. 2.6. The phases $\{\phi_j\}_{j=0}^N$

Genus	Main spectrum
4-genus	-2+i, -1+i, i, 1+i, 2+i
4-genus	-2+5i, -1+5i, 5i, 1+5i, 2+5i
8-genus	-4+i, -3+i, -2+i, -1+i, i, 1+i, 2+i, 3+i, 4+i

Table 2.2: The main spectrum configurations for which neural networks were trained. They were genus-4 solutions such that $\mathrm{Im}[\lambda_j] = 1$ and $\mathrm{Im}[\lambda_j] = 5$, and genus-$8$ solution such that $\mathrm{Im}[\lambda_j] = 1$.

and predefined temporal and special frequencies $\{C_j^f\}_{j=1}^N$, $\{C_j^g\}_{j=1}^N$ produce the jump matrices according to eq. 2.26. These contour and jump matrices are the input data for the numerical routine to solve the RHP [92, 95]. The output is a discretized function $\{q(t_k, z_0)\}_{k=1}^K$, a finite-genus solution to the NLSE, calculated at given values $\{t_k\}_{k=1}^K$ that are usually equidistant (K is the number of samples) and some fixed z_0. Finally, the phases $\{\phi_j\}_{j=0}^N$ and corresponding to them solution $\{q(t_k, z_0)\}_{k=1}^K$ constitute the train data for supervised learning.

The considered main spectra provided non-periodic signals interpreted as FGS-GT. A neural network-based solver was implemented to retrieve the phases of such signals. The NN took the discrete values of the signal $\{q(t_k, z_0)\}_{k=1}^K$ and returned its phases. As the processing window at the input of the neural network, the interval $[0, T_{max}]$ was chosen where T_{max} is the maximum period among all nonlinear modes. Each nonlinear harmonic is guaranteed to contain at least one period in this interval, along with all the information needed to retrieve the phases. To provide a high level of discretization and some universality of the approach, each finite-genus solution was sampled with 128 values per interval.

Finally, finite-genus solutions are periodic in the phases. Therefore, the data labels used for neural network training must also be periodic. A straightforward approach of taking just phases in the interval $[0, 2\pi]$ does not meet this condition. The finite-genus solution with $\phi_j = 0$ and $\phi_j = 2\pi$ is the same signal, while labels 0 and 2π are located at distinct edges of the interval. To avoid this problem, the following labels were used: each phase ϕ_j is replaced with two numbers $\mathrm{Re}[e^{i\phi_j}]$ and $\mathrm{Im}[e^{i\phi_j}]$. In other words, the real-valued phases are transformed to the points on the unit circle with coordinates $\left(\mathrm{Re}[e^{i\phi_j}], \mathrm{Im}[e^{i\phi_j}]\right)$, that, are periodic in phase.

2.3.2 The neural network model

The direct problem for the finite-genus solutions of a generic type has been implemented with a convolutional neural network. A similar model was used for the conventional NFT (when signals decay exponentially fast with $t \to \pm\infty$) and named as NFT-Net neural network [86]. Such architecture demonstrated a high tolerance for the noise of input signals, making it a promising tool for signal processing in fibre-optic communications. The NFT-Net is an example of an encoder-decoder model: the first part of NN, the encoder, transforms the input signal into intermediate states while the second part, the decoder, generates the output. The convolutional layers make the NN noise-resistant. Such neural networks also have other applications, for example, audio signal processing [114].

In the current study, the NFT-Net was an initial model in the Bayesian optimizer, which adjusted the model's hyperparameters. The Bayesian optimization usually performs better than the straightforward approaches such as grid or random search [115] and is used to adjust neural network models. This is because the algorithm exploits the result of the previous optimization steps to select an area of search on the next one. For the model applied in this research, the optimizer operated with the following parameters: a number and a type of layers, convolutional or fully connected, and an activation function after each. The procedure selected the best number of filters for convolutional layers as well as their size, stride, padding, and dilation. At the same time, the number of neurons was optimized for fully connected layers.

The search for the best hyperparameters with the Bayesian optimizer starts from the declaration of the range for adjustable parameters and the objective function, that is, loss function or another metric, to evaluate model performance. The objective function is a function of optimized parameters (or NN hyperparameters in the particular task considered here); this function is computationally difficult to estimate. The Bayesian optimization is an iterative process. In the first step, based on prior knowledge about the objective function, for example, its continuity, a surrogate model is built that is just an approximation of the objective function. Given the subsequent evaluations of the objective function, the posterior, the surrogate model is updated according to the Bayesian rule (this makes the method "Bayesian"). A goal at each iteration is to find such values of parameters that provide the optimal value, maximum or minimum, of the objective function with the smallest possible number of iterations. For this, a special auxiliary acquisition function based on the surrogate model is constructed. This function can be the probability of improvement or expected improvement of the surrogate model. Therefore, parameters providing the maximum of the acquisition function are the best candidates for the next estimation of the objective function. The choice of the acquisition function and its settings is a question of "exploration-exploitation" balance, the optimization process either exploiting a new area that it is not aware of or refining the position of the estimated optimum. Therefore, step by step, looking for the best candidate with the acquisition function, evaluating the objective one, and improving the surrogate model, the Bayesian optimizer provided the optimal parameters of the model.

The resulting model architecture and hyperparameters received after optimization with genus-4 and $\mathrm{Im}[\lambda_j] = 1$ signals at the input of the NN are presented in Fig. 2.7, (while padding $p = 0$ and dilation $d = 1$ for all convolutional layers). Independent optimization was performed for other signals (genus-4 with $\mathrm{Im}[\lambda_j] = 5$ and genus-8), with the initial guess being the optimized model for the genus-4 with $\mathrm{Im}[\lambda_j] = 1$. No improvement was observed while optimizing the two last models. Therefore, the initial architecture was applied for all types of finite-genus solutions under the study.

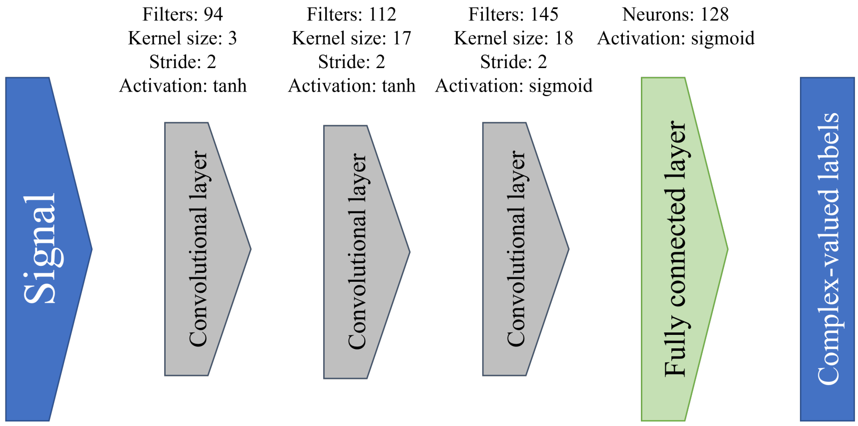

Figure 2.7: The model's architecture and hyperparameters after applying the Bayesian optimization (cited from [J1]).

As mentioned before, the generated model has an encode-decoder architecture. The convolutional part of the NN produces a set of intermediate states from a finite-genus solution sampled in the time domain. These states, then, are transformed into the phases with the fully connected layer. The role of Bayesian optimization is to provide the NN hyperparameters with the maximum possible performance on the particular task of phase extraction. The Bayesian optimizer was run to select an NN with a limited number of trainable parameters (trainable weights of NN), delivering the best model among ones with restricted complexity.

To train the neural networks (three different main spectra), 4×10^5 signals and corresponding labels were generated for each scenario. The data were split into training, validation, and test data in proportion 80%, 17.5%, and 2.5% correspondingly. The training dataset was exploited to adjust the neural network weights, while the validation data were used at the model optimization stage. Finally, the test data were saved to estimate the final performance of the neural network with the data that the model didn't see.

The points on the complex plane with coordinates $\left(\mathrm{Re}[e^{i\phi_j}], \mathrm{Im}[e^{i\phi_j}]\right)$ were used as labels. The measure of error for the phase of each nonlinear mode was just the distance between the predicted and true point on the complex plane:

$$\Delta_j = e^{i\phi_j^{\text{pred}}} - e^{i\phi_j^{\text{true}}}. \tag{2.36}$$

Therefore, the loss function was just the mean squared error (MSE) across all nonlinear modes and all signals in the training dataset:

$$\text{Loss} = \sum_j \left[(\text{Re}[\Delta_j])^2 + (\text{Im}[\Delta_j])^2 \right], \tag{2.37}$$

where $\text{Re}[\Delta_j]$ and $\text{Im}[\Delta_j]$ are real and imaginary parts of the error, (2.36).

The neural networks were implemented with the TensorFlow 2.0 framework. Each model was trained with the Adam optimizer, learning rare 10^{-4}. The training was stopped after 5000 epochs when no improvement was observed. An example of training and validation loss for the model trained with genus-4 and $\text{Im}[\lambda_j] = 1$ data is depicted in Fig. 2.8.

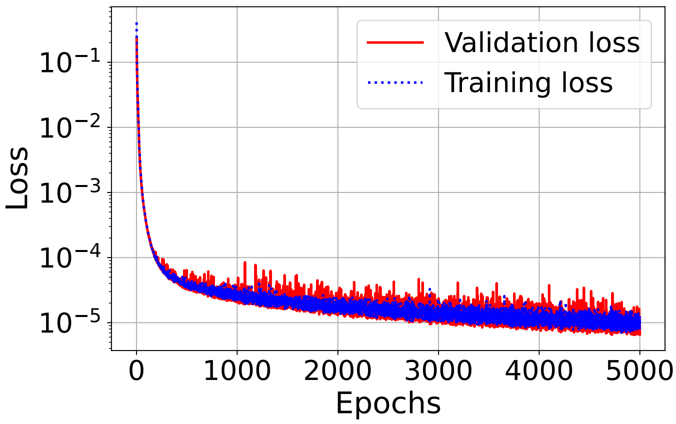

Figure 2.8: Training and validation loss as a function of the number of epochs for the scenario: genus-4 solutions with $\text{Im}[\lambda_j] = 1$ (cited from [J1]).

2.3.3 Results

To estimate the performance of the method across all considered scenarios, the following measure of error was introduced:

$$\phi_{err} = |\phi_{true} - \phi_{pred}|, \tag{2.38}$$

that is the absolute value of the difference between true and predicted phases. The loss function in the neural network models had a different form to provide the required periodicity in phase (2.37). However, the method was developed to predict the phases of finite-genus solutions and the absolute value of the error (2.38) was used for that reason.

The homogeneous distribution of the errors (2.38) was registered across the phases of all signals. The distribution of individual predictions for all three scenarios (genus-4 $\text{Im}[\lambda_j] = 1$, genus-4 $\text{Im}[\lambda_j] = 5$, and genus-8) is provided in Fig. 2.9. To calculate the dependence of the mean error, the interval $[0, 2\pi]$ was divided into 100 subintervals for which the error was averaged across the phases in the subinterval and all nonlinear modes. The mean value of error also demonstrated homogeneous distribution in dependence on the phase and amounted to $1.9 \times 10^{-3}, 9.7 \times 10^{-3}$ and 1.5×10^{-2} for three considered main spectra correspondingly.

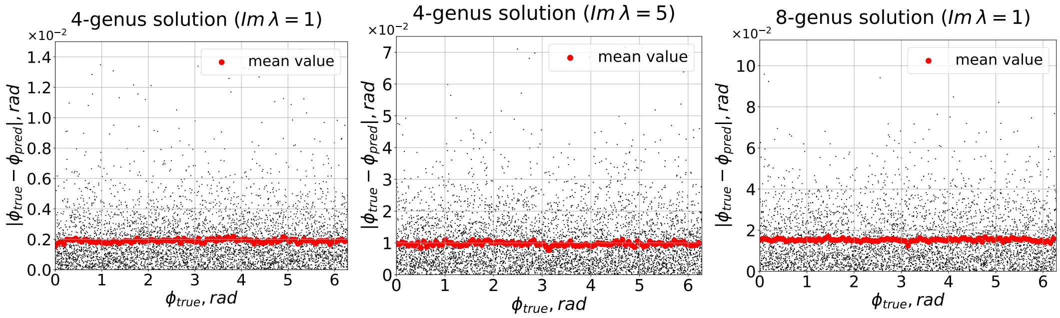

Figure 2.9: The error distributions of the individual phase predictions and their mean values for three different configurations of the main spectrum (cited from [J1]).

The result of applying trained neural network models to retrieve the phases of specific signals is presented in Fig. 2.10. The points with the coordinates $(\mathrm{Re}[e^{i\phi_j}], \mathrm{Im}[e^{i\phi_j}])$ for predicted and true phases are plotted. The figure provides results for all three scenarios: genus-4 solutions $\mathrm{Im}[\lambda_j] = 1$ with the phases $\{0, \frac{5\pi}{3}, \frac{19\pi}{12}, \frac{7\pi}{6}, \pi\}$, genus-$4$ $\mathrm{Im}[\lambda_j] = 5$ with $\{\frac{7\pi}{6}, \frac{2\pi}{3}, \frac{\pi}{12}, \frac{3\pi}{2}, \frac{\pi}{2}\}$ and, finally, genus-8 with $\{\frac{7\pi}{12}, \frac{7\pi}{4}, \frac{3\pi}{2}, \pi, \frac{4\pi}{3}, \frac{11\pi}{6}, \frac{\pi}{12}, \frac{5\pi}{3}, \frac{13\pi}{12}\}$.

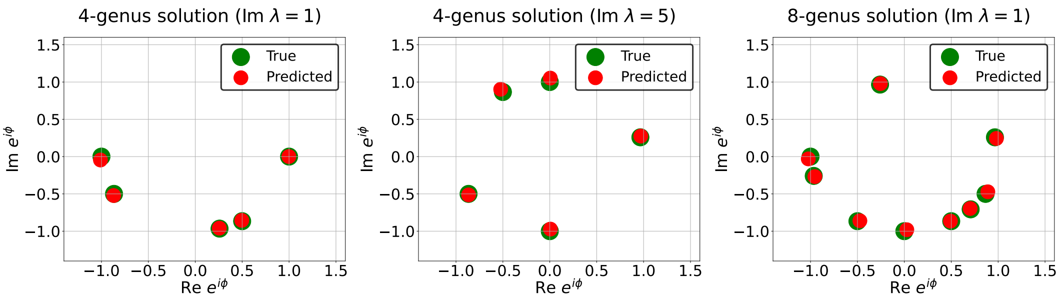

Figure 2.10: The values of the true (green) and predicted (red) phases (cited from [J1]).

The prediction performance was analyzed for the phase corresponding to each nonlinear harmonic. The phase errors were calculated independently and averaged across all signals, Fig. 2.11. In the figure, the error value corresponding to each nonlinear mode is placed above the gap in the complex plane formed by the main spectrum points.

The error behavior in Fig. 2.11 demonstrates the patterns that require explanation. First, the models provided higher accuracy of prediction for the scenario genus-4 and $\mathrm{Im}[\lambda_j] = 1$ (blue line in the left figure) in comparison with genus-4 and $\mathrm{Im}[\lambda_j] = 5$ (red line). This is because the higher imaginary part of the main spectrum corresponds to a stronger nonlinear interplay between nonlinear modes. Moreover, solving the Riemann-

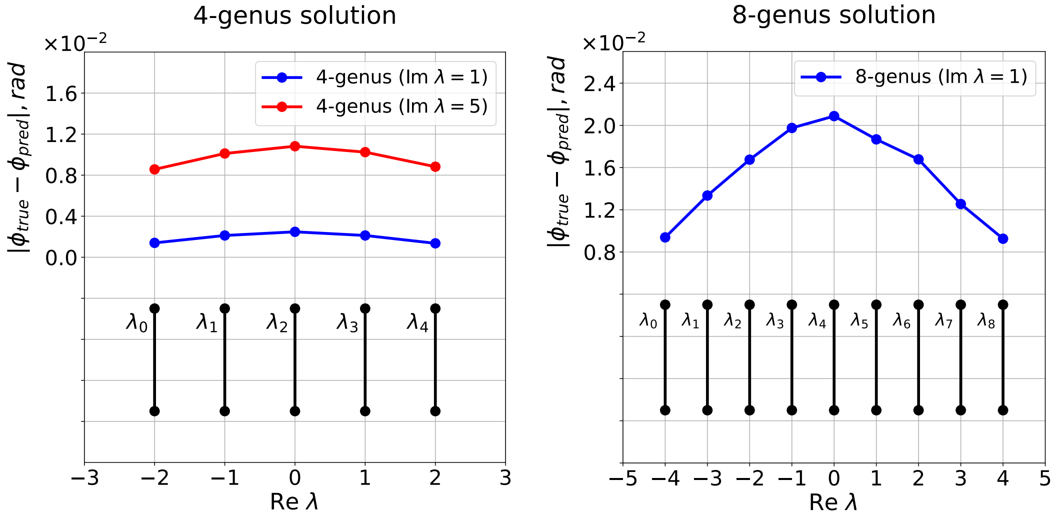

Figure 2.11: The distribution of the phase prediction error for different nonlinear harmonics. Three scenarios are considered: two genus-4 solutions (left) and genus-8 (right). Plots are attributed with the gaps provided by the main spectrum and corresponding to each phase. (Cited from [J1].)

Hilbert problem required numerical integration over gaps formed by the main spectrum points and their complex-conjugated counterparts. In the case of $\mathrm{Im}[\lambda_j] = 5$, the gaps are 5 times longer, and the same discretization level provides low performance. Second, a stable pattern is observed in all scenarios: the neural network models demonstrate a higher prediction error for the phases corresponding to the central gaps. Such a behavior can be explained by the influence of nonlinear modes on each other: the harmonics corresponding to the central gaps experience nonlinear impacts from both sides' modes compared to the edge ones.

2.3.4 Accuracy analysis

A few factors determine the accuracy of the inverse problem and the direct problem algorithms in the developed NFT for finite-genus solutions.

Inverse problem. Accuracy of the inverse problem is associated with two stages: (i) the calculation of the jump matrix $\hat{\boldsymbol{J}}(t, z, \lambda)$ (2.26) for a given main spectrum and (ii) solving the Riemann-Hilbert problem for given contour Γ and $\hat{\boldsymbol{J}}(t, z, \lambda)$ that involves numerical integration over that Γ. In the first stage, the calculation of the temporal and spatial frequencies $\{C_j^f\}_{j=1}^N$ and $\{C_j^g\}_{j=1}^N$ is required. This step is associated with the numerical solving of the linear equation systems (2.21) and (2.22). It was observed that matrix $\boldsymbol{K}$ (2.20) is ill-conditioned when the genus of a solution reaches 15 or more [96]. This problem also restricts the maximum genus achievable with the RHP method. The second stage is integration over the contour Γ to solve the Riemann-Hilbert problem and calculate $\boldsymbol{\Psi}(t, z, \lambda)$ (2.17) [104]. It was implemented with RHPackage software

[95, 110]. On each arc connecting the complex conjugated points of the main spectrum, the integral is estimated at n points. Therefore, n determines the accuracy of solving the Riemann-Hilbert problem and calculating $\mathbf{\Psi}(t, z, \lambda)$.

Direct problem. The data to train the neural network to solve the direct problem were generated with the inverse problem. Consequently, the accuracy of signals in the dataset is defined by the factors described above. One more aspect is the size of that dataset. The value 4×10^5 was estimated experimentally. At smaller values, performance degraded, while the increase in the dataset's size did not improve.

Obviously, the accuracy of the proposed NN-based method for the direct problem can be increased by calculating the input signals both for training and predictions with higher precision. At the same time, to extract more information from these precise signals, a more complex NN model might be required, and a bigger dataset to train it. However, in this study, the error of the phase predictions, about $10^{-3} - 10^{-2}\,\mathrm{rad}$, was accepted (see Fig. 2.9). In the following chapter, this method is applied at the receiver of a fibre-optic communication system. The information signals experience a lot of distortions before they reach the receiver; they are the noise of amplification, deviation of the channel model from the exact NLSE, eq. (1.1), interchannel interference, and other effects. Therefore, I need to care only that the NN-based direct problem solver is not the main contributor to the final error of the retrieved phases. Based on the standard phase modulation formats, such as 16-PSK, 32-PSK, and 64-PSK, a typical final error of the received phases can be estimated. For example, the order of magnitude of phase error in the 32-PSK system is $\frac{2\pi}{32\times 2} \approx 0.1\,\mathrm{rad}$, where the extra factor 2 in the denominator is because the maximum possible error is the distance from the center of one of 32 intervals to its edge. This is a rough order-of-magnitude estimate because does not consider the actual error distribution of the received phases. However, it is enough to see that the NN-based solver has an accuracy of one or two orders of magnitude higher and, therefore, can be accepted for this application.

2.4 Complexity reduction

2.4.1 Neural network complexity reduction with weight clustering

Artificial neural networks can approximate any nonlinear transformation and have found their applications to many practical problems. However, implementation with real hardware systems requires simplification of the neural networks and reducing computational costs. Popular complexity reduction methods include weight clustering [116, 117], pruning [118, 119, 120], and quantization [121, 122]. In this section, the performance of the neural network models solving the direct problem for finite-genus solutions described above is estimated for different degrees of complexity reduction with the weight clustering technique. Application of this method to a convolutional neural network demonstrated remarkable effectiveness with examples such as VGG-16 and ResNet, which showed prediction performance comparable with a baseline model while compressed up to ten times [117].

The weight clustering technique[8] consists of sharing the weight values of a model, effectively reducing the number of multiplications and memory demand for weight storage [116, 123]. Instead of a continuous weight distribution, the procedure selects only a few values and assigns indexes to them. An initialized centroid contains the indexed values of weights, providing operation only with the indexes instead of the weights. The principle of the technique is presented schematically in Fig. 2.12. An initial matrix contains a variety of weight values received from a continuous distribution after training. To compress these weights are shared between clusters, four in my example, and only one value, centroid, is chosen to represent each cluster. The centroids can be selected by the K-means clustering method, for example. The centroids of all clusters are indexed and saved. Having such a correspondence between indexes and centroids, the matrix of weights can be represented as a matrix of indexes. Therefore, in the example in Fig. 2.12, instead of 16 floats, only 16 integers and 4 floats are used to represent weights. The memory saving is bigger for larger models and smaller numbers of clusters.

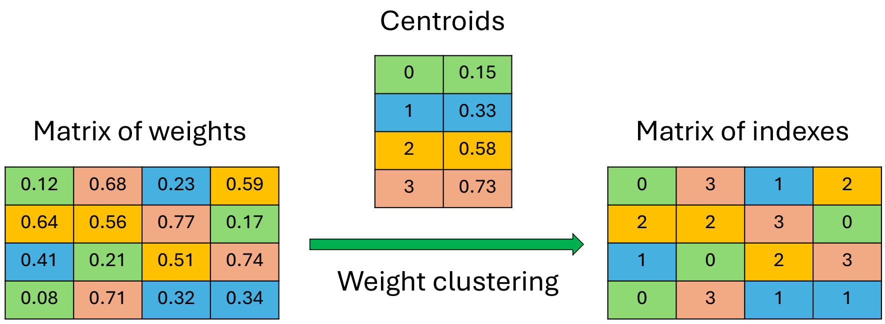

Figure 2.12: Memory saving with the weight clustering technique: instead of a continuous weight distribution, they are shared between a few clusters, and one value, centroid, is used to represent the whole cluster. The centroids are indexed, allowing keeping in memory only the integer values of indexes.

At the same time, the weight clustering technique allows for an effective reduction in the number of multiplications. Let's consider the action of a feed-forward layer implemented as the product of the input vector to the weight matrix. The weights with the same centroid that are from the same cluster can be put out of brackets as a common multiplier; see Fig. 2.13, where such weights are marked with the same color. In this toy example, with 4 elements in the input layer, 16 weights, and 3 clusters, the initial 16 multiplications in the uncompressed model turn to 10 after applying the technique. A

[8] or weight-sharing compression.

significant reduction in the number of multiplications can be achieved for real models. The degree of compression depends on the input vector size, the number of clusters, and the weight matrix.

$$\begin{bmatrix} o_1 \\ o_2 \\ o_3 \\ o_4 \end{bmatrix} = \begin{bmatrix} c_1 & c_2 & c_3 & c_2 \\ c_3 & c_3 & c_2 & c_3 \\ c_2 & c_1 & c_3 & c_1 \\ c_1 & c_2 & c_1 & c_2 \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \\ i_3 \\ i_4 \end{bmatrix} \xrightarrow{\text{Weight clustering}} \begin{bmatrix} i_1 c_1 + (i_2 + i_4) c_2 + i_3 c_3 \\ (i_1 + i_2 + i_4) c_3 + i_3 c_2 \\ i_1 c_2 + (i_2 + i_4) c_1 + i_3 c_3 \\ (i_1 + i_3) c_1 + (i_2 + i_4) c_2 \end{bmatrix}$$

Figure 2.13: Mechanism of the weight clustering techniques: the weights with shared values are put out of brackets in matrix multiplication.

Despite the fact that the illustrations above are provided for a simple feed-forward layer, the same approach can be applied to a convolution layer. In this case, the simplification is provided by sharing the values between the kernels' elements of different filters on each layer [117]. Implementation of the weight clustering technique includes initial training of the model with continuous weight distribution, the definition of the number of clusters and initialized centroids, and finally, fine-tuning the model with new clustered weights to achieve maximum performance.

The weight clustering approach, if applied to a 1D convolutional layer, provides the following computational complexity in terms of the number of multiplications:

$$C_{CNN} = \left(\frac{n_{in} + 2p - d(n_k - 1) - 1}{s} + 1 \right) \times n_f \times n_{cl}. \tag{2.39}$$

Here n_{in} and n_k are the size of an input vector and kernel, n_f and n_{cl} are the number of filters and clusters. The parameters of a convolutional layer, such as padding, dilation, and stride, are denoted with p, d, and s, respectively. The expression in the brackets in (2.39) corresponds to the number of steps each filter makes, while n_{cl} is the number of multiplication on each step. More details are given in [123, 124].

2.4.2 Complexity reduction implementation and results

To implement the weight clustering technique and estimate the performance of the compressed model, the neural network from the previous section was chosen, Fig. 2.7. Among three different finite-genus solutions considered before, genus-4 $\mathrm{Im}[\lambda_j] = 1$ signals were used to operate with (the first line in Table 2.2). Initially, the baseline model was trained over 5000 epochs with standard continuous distributions of the weights. After that stage, the weight clustering technique was applied with the number of clusters varying from 2 to 32. Finally, the fine-tuning was performed with another 3000 epochs. The training and validation loss from the number of epochs for training the baseline model and fine-tuning training with four clusters is depicted in Fig. 2.14. The weight clustering procedure was implemented using the TensorFlow 2.0 framework. The op-

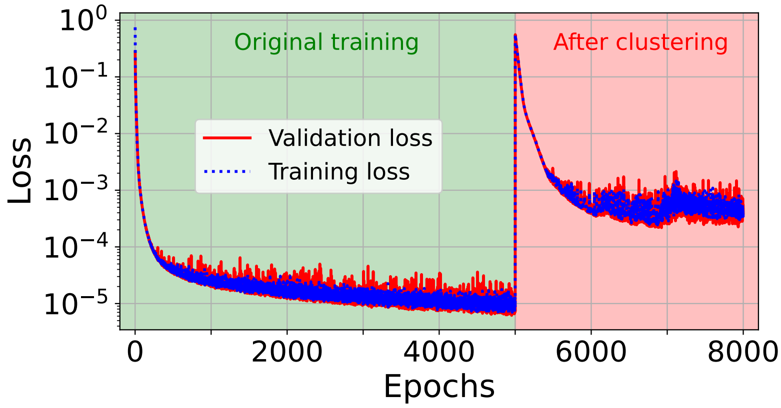

Figure 2.14: Training and validation loss before (green shading) and after (red shading) weight-sharing compression with four clusters (cited form [CP1]).

tion `kmeans_plus_plus` was chosen for the centroid initialization. The neural network was trained with a learning rate 10^{-4} for the baseline model and 10^{-6} at the fine-tuning stage, all with the Adam optimizer.

The continuous weight distributions related to different layers of the original model after its training are plotted in Fig. 2.15. The figure also depicts the weights after applying the weight-sharing compression with four clusters and consecutive fine-tuning. These weights were received in the training shown in Fig. 2.14. All layers demonstrated four different weight values, including the “Conv. 3” layer, for which the fourth cluster is unseen because it is too small.

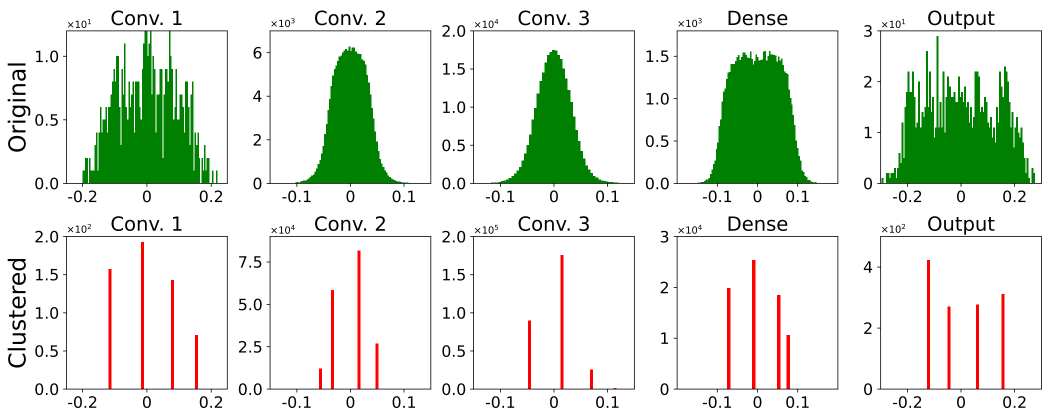

Figure 2.15: Weights for the baseline model with continuous distribution and after applying the weights clustering technique (cited from [CP1]).

The performance of the compressed model in dependence on the number of clusters is depicted in Fig. 2.16. The model's weights corresponding to the lowest value of the loss function (2.37) estimated during the training on the validation data were saved. For these weights, the value of the loss function was estimated using test data. The

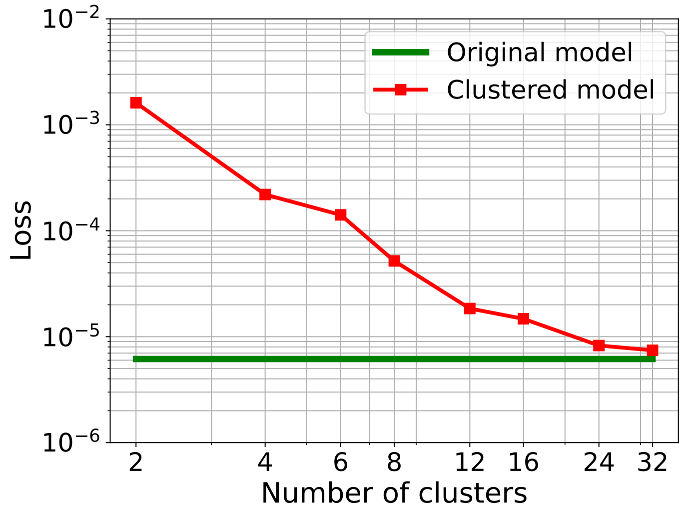

Figure 2.16: Loss function estimated on test data for the different numbers of clusters and loss of the baseline model (cited from [CP1]).

performance dependence demonstrates the expected trend: the model's effectiveness grows with the number of clusters and approaches that of the baseline model. This is because more clusters allow a better approximation of the original model's weight distribution.

The complexity of the clustered model calculated in terms of the number of multiplications in dependence on the number of clusters is provided in Fig. 2.17. It is calculated with (2.39) and related to the complexity of the uncompressed model. With the number of clusters 8 and less, the model provides compression degree 99% and more.

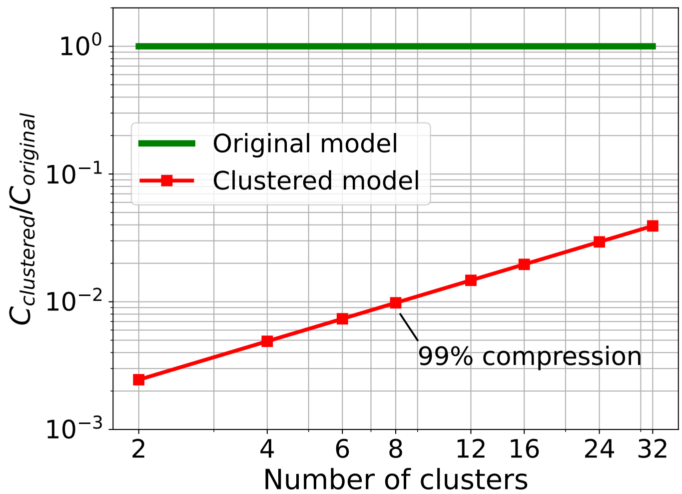

Figure 2.17: The ratio of the clustered model $C_{clustered}$ and baseline model $C_{baseline}$ complexity in terms of the number of multiplications (cited from [CP1]).

2.5 Chapter conclusion

Existing analytical approaches to solve the direct problem of the NFT for finite-genus solutions can process only periodic signals defined over the entire interval T_{com} (Definition 3). However, solutions for which T_{com} can be introduced substitute only a small subset among all finite-genus solutions. It happens because an arbitrary main spectrum does not guarantee that the periods of nonlinear harmonics are commensurable (see Definition 2). Generally, the periods may be non-commensurable, or their common period T_{com} may be too large to be processed effectively with existing algorithms.

The method based on a neural network was developed to avoid these limitations. The key idea is that if a signal $q(t, z_0)$ is defined on the interval $[0, T_{max}]$, where T_{max} is the maximum period among all nonlinear modes, then, the signal contains all necessary information to retrieve the scattering data. The signals meeting this condition were coined finite-genus solutions of a generic type (Definition 4). The method is based on a convolutional neural network trained to solve the direct problem to the NFT for FGS-GT. The Riemann-Hilbert problem parametrization was chosen; consequently, the scattering data comprised the main spectrum and phases 2.2.2. Three different configurations of the main spectrum were used to demonstrate the method's feasibility and flexibility. They were genus-4 $\mathrm{Im}[\lambda_j] = 1$, genus-$4$ $\mathrm{Im}[\lambda_j] = 5$, and genus-$8$ $\mathrm{Im}[\lambda_j] = 1$. In all scenarios, the error of the phase prediction was $10^{-3} - 10^{-2}\,\mathrm{rad}$, which, however, can be decreased with more precise input signals, a bigger dataset, and a more complex NN model. Therefore, the approach was tested with the finite-genus solutions, demonstrating moderate nonlinearity, stronger nonlinearity, and a higher genus regime. While the main spectrum was fixed, the neural network model predicted the phases of signals. A detailed description of the method and estimations of its performance are given in this chapter. Also, it was published in [J1] and presented at conferences [CP3], [CP4]. The simplification of the neural networks in use was studied independently to gain practical value of the approach. The weight clustering technique demonstrated a high level of complexity reduction (up to 99%) with the accuracy of a compressed model comparable with the baseline neural network. This research was presented at a conference [CP2].

This method implemented for finite-genus solutions of a generic type completes the NFT for FGS. As mentioned, the inverse problem and the evolution of scattering data can be realized for any configuration of the main spectrum. In contrast, the direct problem can be solved with analytical methods for the restricted subset of the periodic finite-genus solutions (Definition 3). The proposed approach, based on a neural network, implements the direct problem for FGS-GT. This, in turn, broadens the applicability of the periodic NFT. For example, a few researches were devoted to data transmission with finite-genus solutions in fibre-optic communication systems [33, 34]. However, the incompleteness of the NFT for periodic signals was the cause of the performance underestimation in such systems. The next chapter provides the implementation of a data transmission system based on the finite-genus solution of a generic type.

Finally, a few comments about the potential development of the method must be made. For some tasks, retrieving the main spectrum along with the phases of finite-genus solutions of a generic type may be of interest. It can find an application in the analysis of general waveforms. For example, in the signal processing methods applied in data transmission systems or investigating the dynamics of ocean waves. The proposed approach implemented to get the phases can be expanded for the case of the main spectrum retrieving. The same idea still works: a signal defined on the maximum period among nonlinear modes contains all information about scattering data. Consequently, using the main spectrum along with the phases as labels of a neural network makes it possible to train it to predict the complete set of scattering data. Here, I restricted myself computing the phases only because the main goal of the research was to develop a fibre-optic communication system based on finite-genus solutions of a generic type. In such a task, the main spectrum is fixed: it defines signal parameters, including bandwidth, duration, and power. Information is encoded to the phases only, and they have to be retrieved. This approach is described in detail in the next chapter.

Another possible research direction is to expand the method to the case of Manakov's system, which is also integrable and for which the periodic NFT can be developed [125]. Such an update can be applied in fibre-optic communications where data are transmitted over both polarizations [126]. Some investigations on constructing finite-genus solutions for Manakov's system were done before [127, 128]. However, there is no developed theoretical approach to link the finite-genus solutions to Manakov's system and a solution to the Riemann-Hilbert problem associated with a piece-wise constant jump matrix. Another question is how to express the nonlinear spectrum of such solutions in terms of the main spectrum and the phases. Therefore, the NFT for the finite-genus solutions of Manakov's system remains a subject for further research.

Chapter 3

Finite-genus solutions-based communications

3.1 Communications with the periodic NFT

Alternatively to the conventional NFT, the schemes of data transmission based on the periodic NFT (PNFT) were proposed [26, 100]. In the general case, the PNFT assumes that any smooth enough periodic solution to the NLSE can be approximated with a finite-genus solution (FGS) [32]. This gives rise to the development of the PNFT with two described approaches, the algebro-geometric approach 2.2.1 and the Riemann-Hilbert problem method 2.2.2. The PNFT lacks some shortcomings of the conventional NFT developed for vanishing boundary signals [71, 72]. First, the PNFT functions with non-decaying signals, avoiding burst mode operation and high peak-to-average power ratio (PAPR). Second, signals are periodically continued with cyclic extension prefixes. Therefore, it is enough to process them in one-period intervals at the receiver (smaller than for the conventional NFT). That, in turn, reduces computational complexity and noise-related effects. Finally, the PNFT framework provides full control over the signal's parameters, such as duration, bandwidth, and power. However, the cost for this advantage is the theoretical complexity of the methods to operate with finite-genus solutions and, consequently, a lack of fast numerical routines to solve the direct and inverse problem of the periodic NFT.

A better understanding of the developed methods, results, and unsolved problems in the context of communications with the period NFT requires demonstrating how the scattering data of finite-genus solutions determine the signal's parameters. They depend on the main spectrum only, whereas the auxiliary spectrum/phases do not contribute. Let's consider the main spectrum configuration with equidistant along the real axis points, the structure that is of interest from the communications point of view, Fig. 3.1. This configuration provides commensurability of nonlinear modes' frequen-

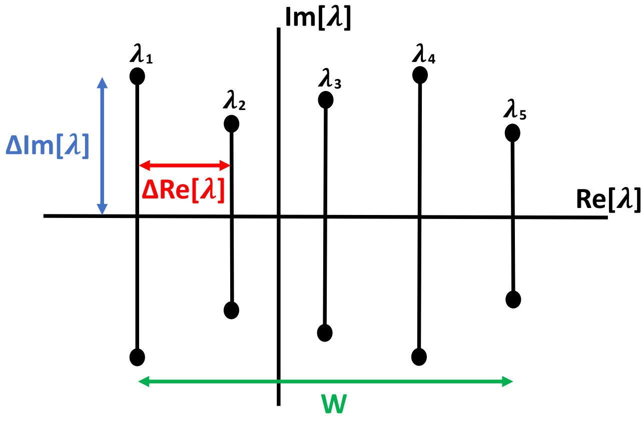

Figure 3.1: The main spectrum of a genus-4 solution with equidistantly placed cuts. The value $\Delta\mathrm{Im}[\lambda_j]$ determines the amplitude of the corresponding nonlinear mode. $\Delta\mathrm{Re}[\lambda]$ is the same for each pair of neighboring cuts; it provides the period of the solution, while W, the range of values $\mathrm{Re}[\lambda_j]$ is the bandwidth of the signal.

cies (Definition 2) in the quasi-linear limit. For the given example with equidistant main spectrum points, the quasi-linearity conditions are defined by the ratio ${}^{\Delta\mathrm{Im}[\lambda]}/_{\Delta\mathrm{Re}[\lambda]}$. The nonlinear properties of a particular finite-genus solution (its deviation from being simply the sum of nonlinear modes and, consequently, the impact of nonlinear harmonics on each other) are dictated by this ratio, not by the imaginary parts of the main spectrum only. It is also useful to consider ${}^{\Delta\mathrm{Im}[\lambda]}/_{\Delta\mathrm{Re}[\lambda]}$ to estimate the degree of nonlinearity because it is possible to increase the imaginary part of the main spectrum with linear scaling of both $\Delta\mathrm{Im}[\lambda]$ and $\Delta\mathrm{Re}[\lambda]$, thereby saving the property of the solution. For more details about the scaling properties of the NLSE, see [27]. In the quasi-linear limit, the amplitude of each nonlinear mode is defined by its $\Delta\mathrm{Im}[\lambda_j]$, and the total amplitude of the signal depends on all values $\Delta\mathrm{Im}[\lambda_j]$ simultaneously. In turn, the duration of the signal, its period, equals $T_0 = {}^{\pi}/_{\Delta\mathrm{Re}[\lambda]}$. Finally, the bandwidth of the signal is specified by the total range $W = N \times \Delta\mathrm{Re}[\lambda]$ (N is the genus of solution). However, as soon as the ratio ${}^{\Delta\mathrm{Im}[\lambda]}/_{\Delta\mathrm{Re}[\lambda]}$ provides no quasi-linear regime, the imaginary part of the main spectrum starts to affect the signals' parameters: the bandwidth broadens nonlinearly, and the frequencies of nonlinear modes are no longer commensurable. This fact restricts the applicability of the PNFT in data transmission systems because the previously developed methods are limited to operating with the periodic signals only (Definition 3), which requires commensurability (a detailed explanation of these peculiarities is provided in the previous chapter). A comprehensive description of how the signals' parameters depend on the main spectrum is given in [34].

These properties are demonstrated below with finite-genus solutions of two different genera. The genus-4 solution with the main spectrum $\lambda = \{-2+ai, -1+ai, ai, 1+ai, 2+ai\}$, with a taking values $\{0.1, 1, 10\}$ is depicted in Fig. 3.2, as well as their linear Fourier spectra. The solutions were calculated using the Riemann-Hilbert problem approach with the phases $\phi = \{5.24, 0.98, 2.24, 4.44, 0.63\}$ for all a. It is seen from the figure that the solution is almost periodic at $a = 0.1$ and non-periodic at $a = 10$. The deviation from commensurability with the growth of the imaginary part of the main spectrum is demonstrated by the distribution of the frequencies provided in Table 3.1. The corresponding linear Fourier spectra demonstrate nonlinear bandwidth increase for higher values of a. It is also can be observed that in the quasi-linear regime ($a = 0.1$), the absolute value of the linear Fourier spectrum $|q(\nu)|$ coincides with the main spectrum (a similar analysis is provided in [34]). However, that is not the case for higher a.

$a = \mathrm{Im}[\lambda]$	The frequencies of nonlinear modes of the genus-4 solution
0.1	{0, 2.01, 4.02, 6.03, 8.04}
1	{0, 2.65, 5.02, 7.38, 10.03}
10	{0, 7.59, 11.94, 16.29, 23.88}

Table 3.1: The frequencies of the nonlinear harmonics for different imaginary parts of the main spectrum $\mathrm{Im}[\lambda]$ of the genus-4 solution.

The same is true for another example of the genus-10 solution, Fig. 3.3 with the main spectrum $\lambda = \{-5+ai, -4+ai, -3+ai, -2+ai, -1+ai, ai, 1+ai, 2+ai, 3+ai, 4+ai, 5+ai\}$. The phases used in constructing the solutions were $\phi = \{1.43, 5.74, 2.03, 0.76, 3.65, 2.78, 4.12, 1.89, 5.11, 0.23, 2.37\}$. In the scenarios, the deviation from commensurability starts from smaller values of the imaginary part and can be seen clearly for $a = 1$ (see also the corresponding distribution of frequencies in Table 3.2). A higher number of nonlinear modes in the signal explains this behavior.

$a = \mathrm{Im}[\lambda]$	The frequencies of nonlinear modes of the genus-10 solution
0.1	{0, 2.01, 4.01, 6.02, 8.03, 10.03, 12.03, 14.04, 16.04, 18.05, 20.06}
1	{0, 2.58, 4.84, 7.05, 9.23, 11.39, 13.56, 15.74, 17.95, 20.21, 22.79}
10	{0, 7.26, 10.37, 14.54, 17.80, 21.32, 24.84, 28.10, 32.27, 35.38, 42.64}

Table 3.2: The frequencies of the nonlinear harmonics for different $\mathrm{Im}[\lambda]$ of the genus-10 solution.

At the same time, the spatial-temporal dynamics of FGS must be considered. When propagating through an optical fibre, signals can exhibit peaks in their waveforms associated with energy localization. This can produce undesirable nonlinear effects, such as soliton formation and others. Moreover, the corresponding spectral broadening can cause additional interchannel distortions in the data transmission scenario with many WDM channels. Fig. 3.4 and Fig. 3.5 show genus-4 and genus-10 signals as a function of both variables t and z. The main spectra and the phases from the examples above were used (with the particular value $a = 1.0$ for both solutions).

Finally, the finite-genus solutions with the configuration of the main spectrum presented in Fig. 3.1 are convenient for data transmission systems. The parameter $\Delta\mathrm{Re}[\lambda]$ defines the signal's duration and bandwidth (last together with a genus N). $\Delta\mathrm{Im}[\lambda_j]$ being independent for each gap, for each nonlinear harmonic, is suitable for nonlinear amplitude modulation. Moreover, each genus-N solution is attributed with N auxiliary spectrum points or $N+1$ phases that can be used for data encoding.

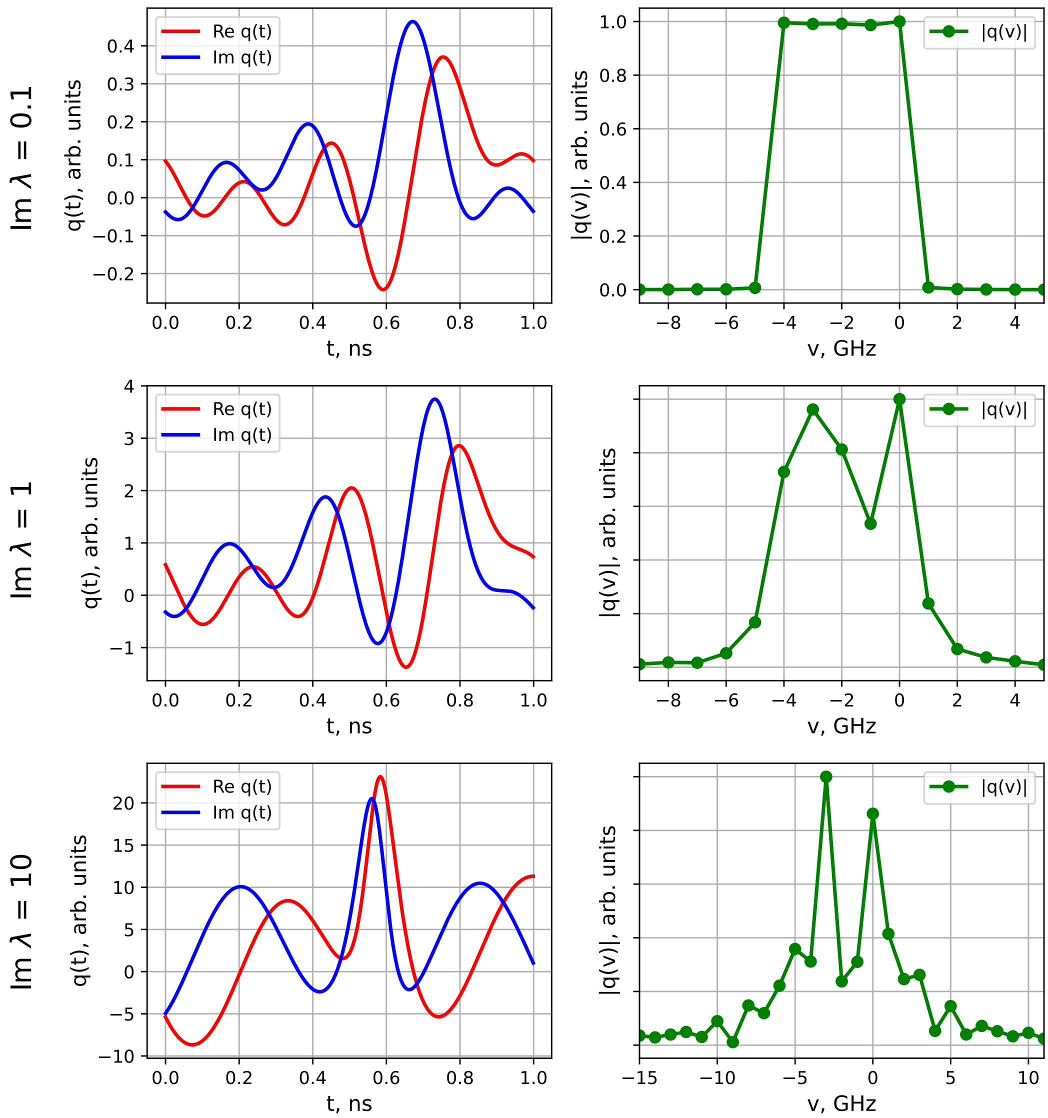

Figure 3.2: Three genus-4 solutions for different values of $\mathrm{Im}[\lambda] = \{0.1, 1, 10\}$ and corresponding to them linear Fourier spectra.

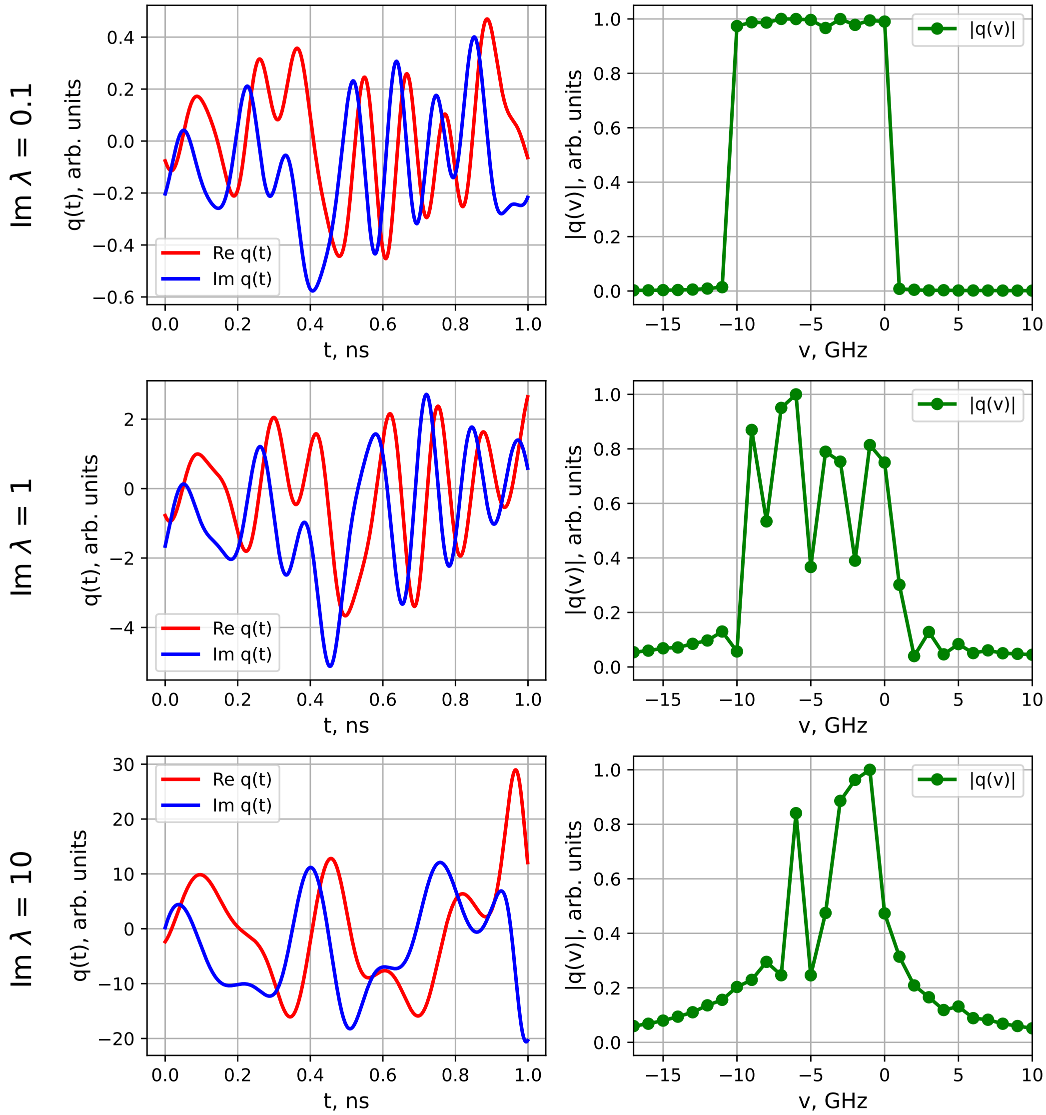

Figure 3.3: Three genus-10 solutions for different values of $\mathrm{Im}[\lambda] = \{0.1, 1, 10\}$ and corresponding to them linear Fourier spectra.

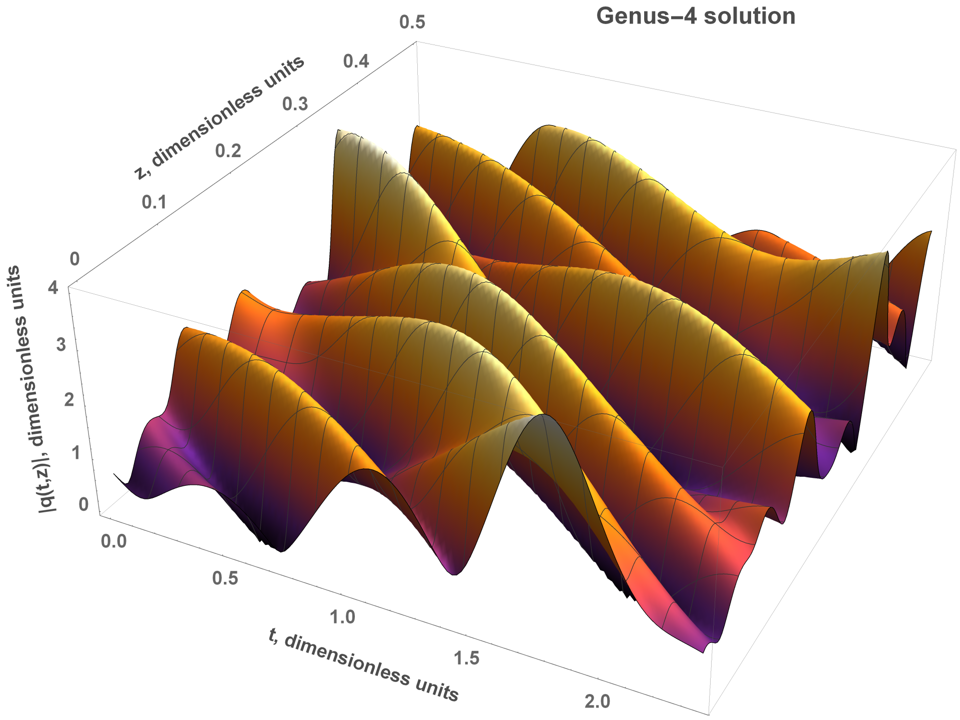

Figure 3.4: The spatial-temporal dynamic of the genus-4 solution with the main spectrum and phases from the example above, with $a = 1.0$.

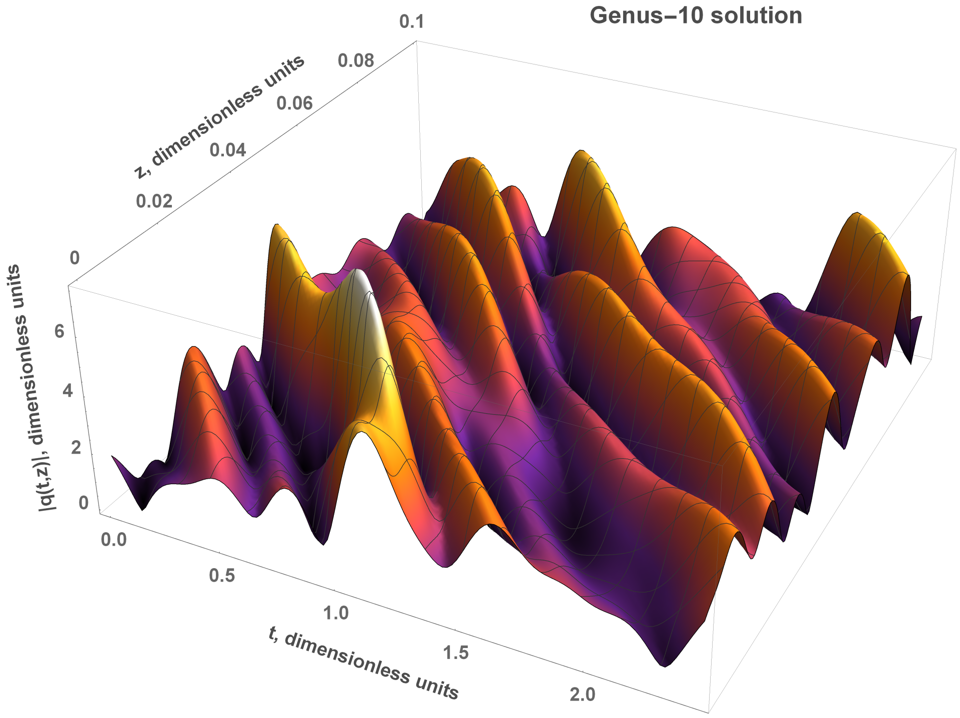

Figure 3.5: The genus-10 solution as a function of t and z, calculated using the particular main spectrum and phases from the example above, with $a = 1.0$.

3.2 The previously developed transmission systems based on the PNFT

A careful consideration of the data transmission system with the algebro-geometric approach is given in the works [33, 73, 129]. First, a numerical algorithm is provided to build finite-genus solutions from their main and auxiliary spectra. Based on that algorithm, the data transmission system was simulated using nonlinear frequency amplitude modulation [33]. The author operates with genus-3 solutions modulating independently the amplitudes of four nonlinear modes. The spectrum of perturbed plane waves was used to construct quasi-periodic finite-genus solutions. The resulting main spectrum was $\lambda = \{-30 + g_1 i, -10 + g_2 i, 10 + g_3 i, 30 + g_4 i\}$, where parameters g_1, g_2, g_3, g_4 were modulated independently with four levels $\{5, 7, 9, 11\}$. Therefore, the ratio ${}^{\Delta \mathrm{Im}[\lambda]}/_{\Delta \mathrm{Re}[\lambda]} = 0.25 - 0.55$ in this research that corresponds to quasi-linear regime. The simulation was performed using the part-averaged NLSE as a model with propagation over the standard single-mode fibre (SSFM) and one polarization. The approach provided data transmission below the forward error correction (FEC) threshold (chosen as 10^{-3}) with a spectral efficiency (SE) of $0.67\,\mathrm{bits/s/Hz}$ at a distance of $1575\,\mathrm{km}$.

Another part of this research was devoted to the experimental demonstration of the data transmission with genus-2 solutions [73]. A single polarization communication system with SSMF and erbium-doped fibre amplification (EDFA) scheme was implemented. Information was encoded to the main spectrum: symbols had four configurations of the main spectrum and carried 2 bits each. Every symbol was attributed with a half-period cyclic prefix while the total duration was $1\,\mathrm{ns}$. Therefore, the system provided a data rate of $2\,\mathrm{Gbits/s}$ with a spectral efficiency of $0.45\,\mathrm{bits/s/Hz}$ at a transmission distance of $2000\,\mathrm{km}$. For more details, see the original papers.

The algebro-geometric approach was also applied in the frame of the so-called “reduced” method [130, 131]. In this scenario, the finite-genus solutions with a special symmetry of the main spectrum and with the closed-form expression in terms of the Jacobi elliptic functions were used. Genus-2 solutions with such a symmetry have three parameters: two were exploited to build a 2-dimensional constellation and another to adjust the signal’s duration. Simulation of information symbols propagation over SSMF with an ideal distributed amplification provided data transmission with a bit error ratio (BER) below a threshold of 10^{-3} at a distance up to $1200\,\mathrm{km}$ (64-QAM modulation). It was also demonstrated that applying probabilistic constellation shaping improves the Q^2-factor by $2\,\mathrm{dB}$.

Some research was devoted to the data transmission with perturbed plane waves as information carriers [72, 132, 133]. For a perturbed plane wave, while a nondegenerate main spectrum point defines the signal parameters, a degenerate can be split, providing two-dimensional space for modulation. The data transmission with high modulation formats, including 128-QAM, 256-QAM, and even 512-QAM, was implemented in the work. A high degree of modulation provided significant spectral efficiency among such systems. SE of $3.27\,\mathrm{bits/s/Hz}$ at a distance of $500\,\mathrm{km}$ and $2.75\,\mathrm{bits/s/Hz}$ at $1000\,\mathrm{km}$ was reported. However, these results were obtained in a system with ideally distributed amplification, which limits the practical value of the method.

Communication systems based on the periodic NFT were also developed using the Riemann-Hilbert problem framework. The RHP approach solved the inverse problem at the transmitter and generated information-carrying signals. In the first works, genus-1 solutions were exploited [105, 131]. Such simple structures have two main spectrum points, λ_0 and λ_1: while $\mathrm{Im}[\lambda_0]$ and $\mathrm{Im}[\lambda_1]$ were used to construct 4-QAM symbols, the $\mathrm{Re}[\lambda_0]$ and $\mathrm{Re}[\lambda_1]$ were adjusted to achieve the desired signal duration. The numerical simulation of the communication system was performed with SSMF and both amplification schemes: ideal Raman and EDFA. Cyclic extension prefixes were incorporated to overcome intersymbol interference due to dispersion broadening, constituting a total symbol's duration of $1\,\mathrm{ns}$ and a bandwidth of $5\,\mathrm{GHz}$. For an EDFA scheme and a propagation distance of $880\,\mathrm{km}$, an optimal power of $-5\,\mathrm{dBm}$ was reported. It provided a data transmission rate of $2\,\mathrm{Gbits/s}$ up to $1200\,\mathrm{km}$. However, it was also demonstrated that the capacity of such a system can be increased by using higher modulation formats. At the same time, a significant drawback of the approach is a high oversampling factor: it was revealed through a numerical analysis that 128 samples per symbol are required. Moreover, this factor grows with the increase in the power of transmitted signals.

The further development of the approach based on the Riemann-Hilbert problem made it possible to exploit not only the main spectrum of finite-genus solutions but also their phases. The data transmission system with a modulation of the phases of finite-genus solutions was implemented, while the main spectrum provided desirable signals' parameters, such as duration, bandwidth, and power [34]. The authors performed a numerical simulation of the data transmission over SSMF with single-polarization and ideal distributed amplification. As information symbols, they exploited genus-14 solutions containing 15 phases each to encode information (in [134] genus-8 solutions are considered). The inverse problem solver based on the RHP method was used at the transmitter to generate a symbol stream for a given scattering data 2.2.2. The direct problem was solved at the receiver with the procedure 2.2.3 and evolution of the phases compensated with eq. (2.27). However, the developed algorithm for retrieving the phases operated only with values defined on the shortened interval $[0, \pi]$ but not the full $[0, 2\pi]$, thereby effectively halving the system's capacity. Nevertheless, $23.7\,\mathrm{Gb/s}$ with 8-PSK modulation format at a distance of $1040\,\mathrm{km}$ (13 spans of $80\,\mathrm{km}$ each) was reported. The symbol duration was $1.9\,\mathrm{ns}$, including 50% cyclic prefix on each side and an average bandwidth of $9.8\,\mathrm{GHz}$. Analysis of the system performance in dependence on the power of signals provided an optimal value of $-17\,\mathrm{dBm}$. The power-dependent error of the periodic NFT algorithms explains such a low value: accuracy decreases with power growth. The approach also suffers from oversampling: the 128 samples per symbol were used.

Finally, let's mention research devoted to applying machine learning in transmission systems with the periodic NFT. Different models, including support vector machine, k-nearest neighbors, k-means clustering, and Gaussian mixture model, are exploited to equalize/detect signals at the receiver side [135]. The authors applied their methods to the communication system developed in [130]. The best gain in BER was achieved when the support vector machine and k-nearest neighbors techniques were employed. They provided data transmission below a hard decision forward error correction (FEC) threshold of 3.8×10^{-3} in a wide range of signal's powers (with a propagation distance of $924\,\mathrm{km}$) while the direct detection method failed.

It is worth mentioning how periodic signals reduce the processing window at the receiver. Consider as an example the data transmission case described above: encoding information to the phases of the genus-14 solutions. The authors reported that an information symbol had 128 samples. At the same time, each symbol consisted of two periods because it had half-period cyclic extension prefixes on each side. However, the cyclic extension serves to avoid intersymbols interference due to chromatic dispersion, and all information about the signal is contained only in one-period interval. Therefore, in the example, only 64 samples can be processed. In comparison, for the conventional NFT (with decaying boundary signals), a full signal must be processed at the receiver because all parts of the signal after its propagation contain original information. Anyway, even with this reduction, the methods based on the PNFT still require significant oversampling. This makes them non-competitive compared to conventional communication approaches.

A significant drawback of both data transmission approaches is the operation in the quasi-linear regime. Whether the algebro-geometric approach or the Riemann-Hilbert problem method is used, constructing a finite-genus solution starts by choosing the main spectrum, which defines commensurability and periodicity (see Definition 2 and Definition 3). However, the only known approach to finding the main spectrum providing commensurability is to start construction from a linear plane wave, which, if perturbed, corresponds to the finite-genus solution with a desired period. Application of this technique is limited: with the growth of the ratio ${}^{\Delta \mathrm{Im}[\lambda]}/_{\Delta \mathrm{Re}[\lambda]}$ the resulting main spectrum does not provide a periodic solution. In the context of data transmission systems, the low value of ${}^{\Delta \mathrm{Im}[\lambda]}/_{\Delta \mathrm{Re}[\lambda]}$ restricts the power of signals or broadens a bandwidth.

The finite-genus solutions considered in the works described above are for the NLSE. Therefore, the developed systems use only one polarization for data transmission, which halves their capacity. To harness dual-polarizations in such systems, the finite-genus solutions of Manakov's system must be considered. However, the mathematical theory behind this method is still poorly developed, and the construction of such solutions is not described. I know a few studies touching on the topic [127, 128]. Consequently, developing dual-polarization data transmission systems with finite-genus solutions of Manakov's system is an area for future research.

3.3 Finite-genus solutions of a generic type for fibre-optic communications

As described in the previous chapter, the method to solve the direct problem for finite-genus solutions of a generic type based on neural networks (NN) provides opportunities to implement new communication systems. These systems can avoid the drawbacks of the data transmission approaches based on the periodic NFT developed early (see review in the previous section). Therefore, the following principal method was proposed: (i) information is encoded to the phases of finite-genus solutions of a generic type, while the main spectrum controls the signal's parameters; (ii) the Riemann-Hilbert problem approach solves the inverse problem and calculates signal $q(t, z_0)$ for a given scattering data at the transmitter; (iii) at the receiver a neural network retrieves the phases of finite-genus solutions and compensates for their evolution if required.

Actually, a genus-N solution contains $3\times(N+1)$ real-valued parameters: $2\times(N+1)$ from the main spectrum $\{\mathrm{Re}[\lambda_j], \mathrm{Im}[\lambda_j]\}_{j=0}^{N}$ and $N+1$ phases $\{\phi_j\}_{j=0}^{N}$[1]. However, in the proposed approach, information is encoded into the phases using a phase-shift keying (PSK) modulation format. The real values of the main spectrum $\{\mathrm{Re}[\lambda_j]\}_{j=0}^{N}$ define the frequencies of nonlinear harmonics and a total signal's duration and bandwidth. Finally, $\{\mathrm{Im}[\lambda_j]\}_{j=0}^{N}$ determine the amplitudes of nonlinear modes. However, when the ratio ${}^{\Delta\mathrm{Im}[\lambda]}/{}_{\Delta\mathrm{Re}[\lambda]}$ is out of the quasi-linear regime, the imaginary part of the main spectrum also affects signal's duration and bandwidth. It is worth noting that $\{\mathrm{Im}[\lambda_j]\}_{j=0}^{N}$ still can be used for data encoding in addition to the phases: the independent amplitudes can be modulated while the average value produces a desired power of the signal. But in such a scenario, the power variety from symbol to symbol increases, and the peak-to-average power ratio grows. That, in turn, reduces transmission quality. Therefore, involving amplitudes in modulation is not straightforward and requires investigation of the properties of such signals. This question is not considered in the current study.

From a practical point of view, the RHP framework has a principal advantage over the algebro-geometric approach. While the Riemann theta function's computational complexity grows exponentially with the genus of a solution (in the AG method), the RHP demonstrates significantly lower computational demands [105]. Moreover, the RHP evaluates each signal value $q(t)$ independently at every time t and, therefore,

[1] everywhere below the RHP framework is used, 2.2.2.

can be parallelized. Also, calculating the parameters δ^{+} and δ^{-} in eq. (2.12) from the auxiliary spectrum requires additional integration over the Riemann surface, whereas the phases in the RHP approach are utilized directly from their original form. All these factors influenced the choice of the RHP method in this study.

It was mentioned before that the monodromy matrix framework is the only existing method to solve the direct problem for finite-genus solutions. This limits the main spectrum configurations with which the periodic NFT can operate: the solutions must be periodic (Definition 3). The neural network-based direct problem solver can process any finite-genus solution, lifting the periodicity requirement. It, in turn, allows the use of an arbitrary value of ${}^{\Delta\mathrm{Im}[\lambda]}/_{\Delta\mathrm{Re}[\lambda]}$ harnessing the nonlinear nature of finite-genus solutions. Moreover, using a neural network removes the restriction to operate with the phases on the limited interval $[0, \pi]$ [34]. Instead, the full range can be processed, doubling the system's capacity. At the same time, neural networks at the receiver can compensate for some deterministic distortions caused by deviating from the ideal model. For example, applying the analytic PNFT requires cyclic prefixes to avoid interference between the neighboring symbols. However, neural networks can retrieve information from the symbols stream without cyclic extension, shortening a symbol's duration and increasing the total capacity. This will be demonstrated later in one of the numerical experiments. Also, NNs reduce the impact of other effects in fibre-optic communication systems, including periodic signal amplification and attenuation, as well as noise-induced distortions.

The following sections describe the numerical simulations of the different data transmission systems based on the proposed NFT for finite-genus solutions of a generic type with a convolutional neural network. In the first experiment, the NN-based direct problem solver was used only to retrieve the phases at the receiver while their evolution was compensated analytically with eq. (2.27); for details see [CP1]. In the second scenario, the NN solved the direct problem and compensated for the phases evolution, reducing the action of other deterministic effects [J2]. This is followed by applying a compression technique to the neural network to decrease complexity [CP2]. These research pieces constitute this chapter's results and are described below.

3.3.1 Data transmission with the CNN-based phase detector

In the first study, the fire-optic communication system based on finite-genus solutions of a generic type was implemented numerically. These results were also published in [CP1]. At the transmitter, the phases of finite-genus solutions were exploited to encode information, while the main spectrum provided control over signals' duration, power, and bandwidth. The sequence of symbols was generated with the RHP approach and fed into the optical fibre with ideally distributed amplification (lossless model). At the receiver, a convolutional neural network (CNN) calculated the phases, and their evolution was compensated analytically with eq. (2.27). In this research, it was demonstrated that the approach based on the CNN outperforms the analytical method of solving the direct problem reported in the work [34].

Channel model

The NLSE governs signal propagation over an optical fibre. In the lossless scenario, it takes the form:

$$iq_z - \frac{\beta_2}{2} q_{tt} + \gamma |q|^2 q = n(t, z), \tag{3.1}$$

where $q(t, z)$ is the signal's envelope in the time domain, t and z are the temporal and spatial variables correspondingly, β_2 and γ are the group velocity dispersion and nonlinearity parameters of the optical fibre, and $n(t, z)$ is the noise of amplification.

In the proposed system, the genus-4 solutions were data carriers: each information symbol consisted of the genus-4 solution, defined at some interval T (symbol's duration). Such signals have a complex structure that demonstrates intricate dynamics but, at the same time, can be generated effectively with existing algorithms. Every genus-4 solution contains 5 nonlinear modes providing 5 phases to encode information. The main spectrum had the following form:

$$\lambda = \{-2 + ai, -1 + ai, ai, 1 + ai, 2 + ai\}, \tag{3.2}$$

with the parameter a controlling the power of the signal. Meanwhile, the real part of the main spectrum (fixed in this study) and the imaginary part, parameter a, defined the symbol's duration and bandwidth. The value of the ratio $\Delta\mathrm{Im}[\lambda]/\Delta\mathrm{Re}[\lambda]$ provided substantial deviation from quasi-linearity: the corresponding frequencies of nonlinear harmonics were non-commensurable (see, Table 3.1 for a typical value $\mathrm{Im}[\lambda] = 1$). The duration of the symbol T was the longest period among all nonlinear modes. Each symbol was attributed with extension prefixes to prevent overlapping due to dispersion broadening. Because the signals were not periodic, the extension prefixes were not cyclic; they were just a continuation of the solution behind interval $[0, T]$. To guarantee the absence of interference, the total duration of the extended symbol was $5 \times T$. In other words, the extended symbol was defined on the interval $[-2T, 3T]$. This extension is significantly larger than required according to the linear signal broadening estimation [136]. Therefore, the distortions caused by dispersion were excluded from the analysis, although it reduced the system's efficiency. Finally, the extended symbols were concatenated into a single sequence and transmitted over an optical fibre.

Convolutional neural network-based receiver

The key component of the method was a convolutional neural network at the receiver exploited to retrieve the phases of information signals [J1]. The input of the CNN was a propagated symbol with the removed extension prefixes incorporated to avoid overlapping with the neighboring symbols due to the dispersion of a fibre. The neural network took 128 samples of the signal and returned 5 complex-valued parameters, which are points on the unit circle. This trick was applied to guarantee the periodicity with respect to the labels: the solutions are periodic in the phases[2]. Then, the phases are calculated trivially for the given points on the unit circle. Finally, the evolution of the phases was compensated with eq. (2.27).

The details of the neural network architecture are given in the work [J1]. The CNN consisted of three convolutional layers, one fully connected layer, and an output layer (see Fig. 3.6). Such a model has an encoder-decoder structure: while the convolutional part generates intermediate states, the feed-forward layer produces the output. Moreover, convolutional layers provide effective noise filtering (see more details in the previous chapter). The Bayesian optimization procedure delivered the optimal values of hyperparameters; they are summarized in Table 3.3 (in addition, stride $s = 2$, dilation $d = 1$, padding $p = 0$ for all convolutional layers) [115]; see the previous chapter as well. The optimal values of the hyperparameters were the same for all powers and remained unchanged in the simulations.

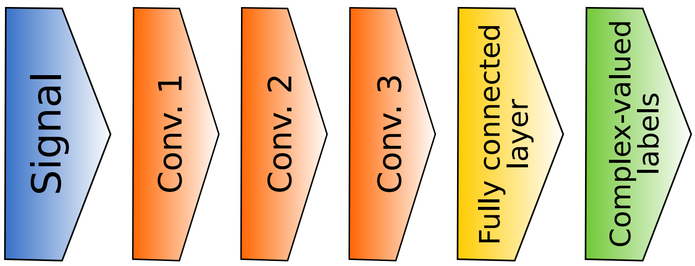

Figure 3.6: The convolutional neural network used at the receiver to retrieve the phases of information symbols (genus-4 solutions). It consists of three convolutional layers, one fully connected and an output layer. (Cited from [J2].)

Results and discussion

Data transmission with the genus-4 solutions as data carriers was numerically simulated. The RHP solver calculated the information signals at the transmitter, while the described CNN was used at the receiver. The scheme of the system is depicted in Fig. 3.7. The propagation path consisted of 15 spans of standard single-mode op-

[2]It is not possible if the interval $[0, 2\pi]$ is used. The labels for 0 and 2π are on opposite edges of the interval while corresponding to the same solution.

	Filters	Kernel size	Activation
1 conv.	94	3	tanh
2 conv.	112	17	tanh
3 conv.	145	18	sigmoid
Fully-con.	128 neurons		sigmoid

Table 3.3: The CNN's hyperparameters as a result of Bayesian optimization, see Fig. 3.6 (cited from [CP1]).

tical fibre. Each span was $80\,\mathrm{km}$ in length, totaling an overall propagation distance of $1200\,\mathrm{km}$. The parameters of the SSMF were the following: the group velocity dispersion was $\beta_2 = -21.7\,\mathrm{ps}^2/\mathrm{km}$, while the nonlinear factor was $\gamma = 1.3\,\mathrm{W}^{-1}\mathrm{km}^{-1}$. Amplification distributed ideally (lossless propagation) was introduced (see eq. (3.1)) with noise added after each span. The power of the noise was estimated as $N_{ASE} = \alpha L \hbar \nu_s K_T NF$ [24] with parameters: $\alpha = 0.2\,\mathrm{dBm/km}$ being the loss of the optical fibre, $L = 80\,\mathrm{km}$ the span length, $\hbar \nu_s$ energy of photon at the carrier's frequency ν_s, $K_T = 1.13$, and noise figure $NF = 4.5\,\mathrm{dB}$.

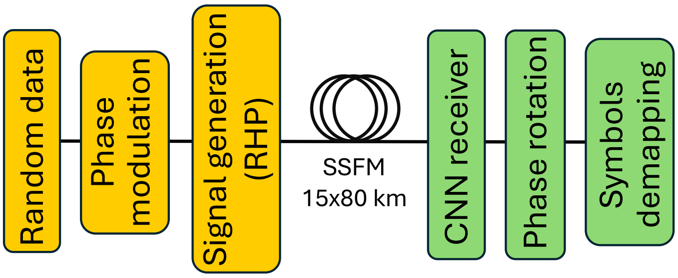

Figure 3.7: The principal scheme of the communication system under the study. It consists of the phase modulation, the RHP signal generation, 15 spans of optical fibre, the CNN-based phases detection and rotation, and, finally, symbols demapping (cited from [CP1]).

Each phase of a genus-4 solution was modulated with 16-PSK format. Therefore, 5 phases provided 5×4 bits of information per symbol. The symbols were normalized to have the duration of the central part of $1\,\mathrm{ns}$ and the total duration with extension prefixes of $5\,\mathrm{ns}$. To estimate the system's performance, BER was calculated by directly counting error bits (over 2×10^4 symbols). BER values for different signal's powers are shown in Fig. 3.8. The successful data transmission below the FEC threshold of 3.8×10^{-3} and a 7% overhead [137] was demonstrated for two power levels $\approx -6.3\,\mathrm{dBm}$ and $\approx -5\,\mathrm{dBm}$.

This research aimed to demonstrate the data transmission system with the CNN-based detector at the receiver. The communication systems with finite-genus solutions were implemented before, but their performance suffers from the restriction imposed by analytical solving of the direct problem, which is the periodicity of the signals. The method described here relieves these limitations and adopts the nonlinear nature of

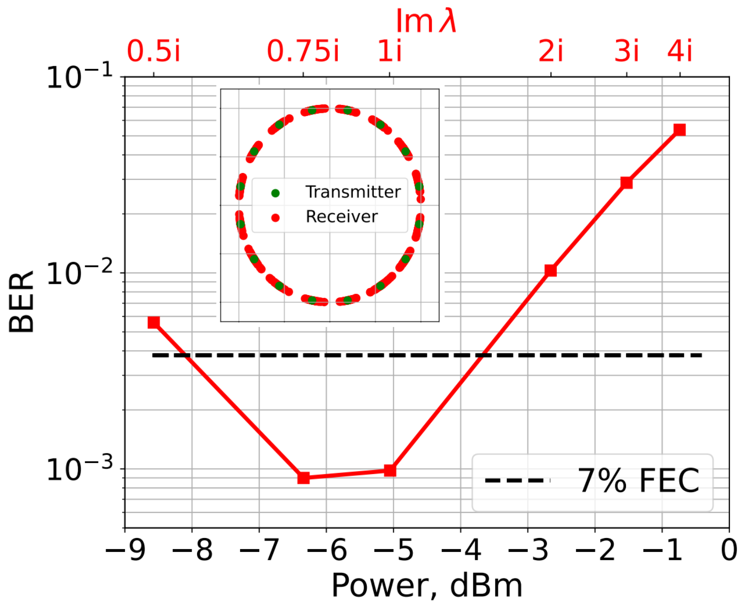

Figure 3.8: The dependence of the data transmission quality in terms of BER as a function of signal power. The corresponding values of the imaginary part of the main spectrum, parameter a in eq. (3.2), are depicted on the upper x-axis. The FEC threshold is 3.8×10^{-3} (7% overhead). The inset provides the distribution of the phases at the receiver for the optimal power $\approx -6.3\,\mathrm{dBm}$. (Cited from [CP1].)

finite-genus solutions to construct a reliable and effective fibre-optic communication system. That proof-of-concept approach requires further development, which is described below as a report of another research. Constructing a more realistic communication system with a non-zero gain/loss profile and optimized data carriers was implemented. Moreover, the full power of neural networks was harnessed to solve the direct problem and compensate for the deterministic distortions related to real systems.

3.3.2 Fibre-optic communication system with the CNN-based receiver

This work continued the application of finite-genus solutions to the NLSE in fibre optic communication systems. Similar to the previous section, information was encoded to the phases of FGS while the main spectrum served to adjust signals' parameters. The RHP solver generated signals $q(t, z_0)$ for the given scattering data at the transmitter. However, there were no extension prefixes: signals' distortions due to their overlapping were compensated with the same CNN at the receiver that performed the phase detection. Another distinction in this research (compared to the previous section) was implementing a more practical amplification scheme with a nonzero gain/loss profile (EDFA). Finally, the key element of the method was the convolutional neural network-based receiver. It detected the phases of finite-genus solutions and performed equalization, including the phases' rotation. Moreover, as mentioned before, it compensated for such deterministic effects as periodic attenuation and amplification, symbols overlapping due to the action of chromatic dispersion, and the impact of noise. The results of this research were published in work [J2].

Finite-genus solutions of a generic type as data carriers

In this numerical experiment, the genus-4 solutions were chosen to encode information into their phases. The main spectrum with the configuration of eq. (3.2) was exploited where parameter a controlled signals' power. However, the typical values of a considered in the research were high enough to affect the signal's duration and bandwidth. In other words, for these values of the imaginary part of the main spectrum, finite-genus solutions have non-commensurable frequencies of nonlinear modes. Also, when a is high enough, it broadens the linear spectrum (see Fig. 3.2). And finally, the phases influence the bandwidth: while a signal propagates over an optical fibre, its bandwidth changes, which relates to the evolution of the phases because the main spectrum is invariant. But this impact is negligible, as was reported in [34].

The frequencies of nonlinear modes (C_f^1, C_f^2...) were calculated for every configuration of the mains spectrum. Then, information symbols were calculated with the RHP solver at the interval T. This value was defined as $T = 2\pi/C_f^{min}$, where $C_f^{min} = \min_j C_f^j$ is the minimum frequency across all nonlinear harmonics. Therefore, T contains at least one period of each nonlinear mode because it is maximum among them. In Fig. 3.9, three consecutive symbols at the transmitter are depicted. For the central one, the periods of nonlinear harmonics, T_1, T_2, T_3, T_4, are provided[3]. It is seen that the maximum period T_1 (defined by the minimum frequency C_f^{min}) contains all other T_2, T_3, T_4. Thus, it is enough to choose the symbol's duration as $T = T_1$ because the full information about each nonlinear mode is encapsulated in its one period. These symbols were then concatenated into a sequence without extension prefixes, one by one, as depicted in Fig. 3.9.

Compensation for dispersion-induced overlapping was performed with the CNN receiver that simultaneously processed three consecutive symbols. There exist remarkable gaps between the neighboring symbols both in amplitude and phase. The amplitudes' gaps can be avoided by shifting each symbol in time. In other words, the symbol is defined not on the interval $[0, T]$, but on $[\Delta t, T + \Delta t]$, where Δt is different for each symbol and chosen to have no gaps. The total phase of each symbol can be adjusted to remove the gaps in phase between the neighbors. However, each operation requires sacrificing one of the phases used for modulation, obviously reducing the system's capacity. Therefore, signals were propagated without any processing to eliminate these

[3]The genus-4 solution has 5 nonlinear modes, but one has zero frequency, and its period is undefined.

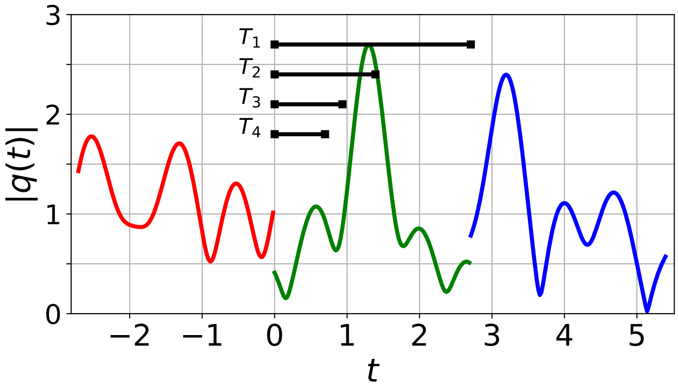

Figure 3.9: The genus-4 symbols at the transmitter generated with parameter $a = 0.6$ and random phases. The central symbol with duration $T = T_1$ encapsulates nonlinear harmonics with smaller periods T_2, T_3, T_4. The symbols are concatenated one by one without extension prefixes. The amplitude and time are provided in dimensionless units. (Cited from [J2].)

gaps. As demonstrated later, the CNN-based receiver could fully compensate for disruptions related to this effect.

Communication system and a convolutional neural network-based receiver

The communication system under the study contained the following key elements: (i) a genus-4 symbol's phase modulator with random data; (ii) the RHP-based generator of signals $q(t, z_0)$ in the time domain with the main spectrum specified by desired power; (iii) N spans of an optical fibre with EDFA scheme; (iv) the convolutional neural network-based receiver that retrieves the phases and compensate for their evolution. In some experiments, optical bandpass filters were applied both at the transmitter and receiver, as well as downsampling at the CNN input. The scheme of the system is in Fig. 3.10.

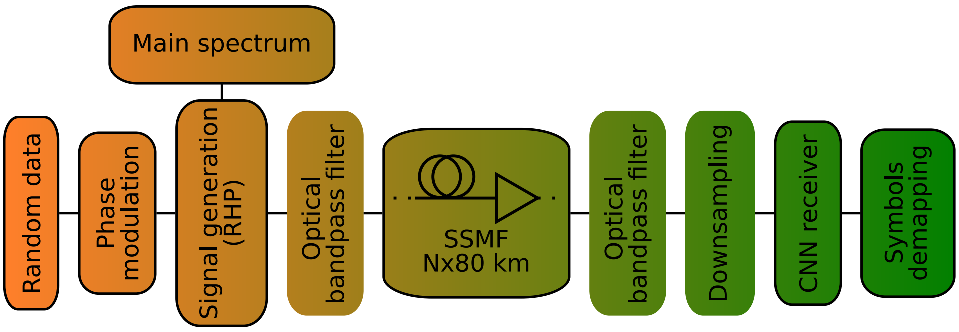

Figure 3.10: The scheme of the fibre-optic communication system. The RHP solver generates information-carrying symbols for the given main spectrum and phases at the transmitter. Then, signals propagate through N optical fibre spans with erbium-doped amplifiers. The described CNN processes the signals at the receiver. In some scenarios, optical bandpass filters and downsampling are applied. (Cited from [J2].)

The equation governing a signal dynamic in the optical fibre with attenuation is:

$$iq_z - \frac{\beta_2}{2} q_{tt} + \gamma |q|^2 q = -i\frac{\alpha}{2} q + n(t, z), \tag{3.3}$$

where in contrast to eq. (3.1) the loss term characterized parameter α is introduced. However, the finite-genus solutions correspond to the NLSE with only dispersion and nonlinear terms, eq. (3.1). Any deviation from this model breaks integrability, and finite-genus signals do not hold their properties, including the linear evolution of the phases, eq. (2.27). However, when signals propagate in the medium with attenuation and amplification, its dynamic averaged over the path still obeys the lossless model, the so-called lossless path-averaged model [58, 136].

As specified above, the key component of this method is the convolutional neural network-based receiver. The CNN solves the direct problem of the periodic NFT; details are given in J1 and CP1. In the current approach, it also compensates for the phases' evolution. Their rotation (governed according to eq. (2.27)) is incorporated into the neural network at the receiver. This is achieved by training the NN with the propagated signals as input and the phases from the transmitter as labels. The CNN has three consecutive convolutional layers, one fully connected layer, and an output layer. The same CNN and hyperparameters were utilized in the previous experiment; see Fig. 3.6 and Table 3.3.

In this study, the neural network at the receiver took three consecutive symbols. This was implemented to consider dispersion-induced memory. There was no chromatic dispersion compensation or extension prefixes; symbols were stacked one by one. To guarantee that it is enough to account for only three symbols, the following estimation of signal linear broadening can be done [136]:

$$\Delta T = 2\pi |\beta_2| B L. \tag{3.4}$$

Here $\beta_2 = -21.67\,\mathrm{ps}^2/\mathrm{km}$ is the group velocity dispersion, $B = 6\,\mathrm{GHz}$ is a typical bandwidth, $L = 1040\,\mathrm{km}$ is a typical propagation distance. The expression above provides the signal broadening of $\Delta T = 0.85\,\mathrm{ns}$. At the same time, the symbol duration in the experiment was fixed and chosen of $1\,\mathrm{ns}$. Obviously, it is enough to consider only three symbols at the CNN input to account for memory effects. Moreover, it was reported in other research that finite-genus solutions experience even smaller broadening than those provided with eq. (3.4) due to their nonlinear nature [73].

For each power, an independent neural network was trained. While the CNNs had the same architecture and hyperparameters for all power levels, the layers' weights differed. The training data consisted of 4×10^5 symbols, while the test dataset contained 2×10^4 symbols (this value was expanded in some experiments to provide at least 100 error symbols). Each symbol was discretized into 128 or 64 samples in different simulations; thereby, the input of NN was 3×128 or 3×64 values. As labels, the complex-valued points on the unit circle were chosen as specified above. The NN was implemented with TensorFlow 2.0 framework and trained over $5\,000$ epochs with the Adam optimizer and learning rate of 10^{-4}.

Results and discussion

The results of the numerical implementation of the data transmission system (as specified in Fig. 3.10) are provided below. At the transmitter, the sequence of 1020 symbols was generated, concatenated into one signal, and transmitted through an optical fibre over one polarization and one frequency channel. The propagation path consisted of N spans of SSMF, each $80\,\mathrm{km}$ in length. The EDFA scheme was implemented, and noise was introduced at the end of each span. The first and last 10 symbols were removed after the signal propagation to avoid boundary effects. Then, the CNN-based receiver calculated the phases and compensated for any deterministic distortions. Finally, the system performance was evaluated by directly counting error bits.

The information symbols were generated with the main spectrum specified in eq. (3.2). The signal's power was modulated with the parameter a. The duration of each symbol was normalized to be $1\,\mathrm{ns}$ for any value of a[4]. 32-PSK modulation format was exploited to encode information into each of the five phases of the genus-4 solution. The parameters of SSMF were the following: the group velocity dispersion $\beta_2 = -21.67\,\mathrm{ps}^2/\mathrm{km}$, the nonlinear factor $\gamma = 1.27 \times 10^{-3}\,\mathrm{mW}^{-1}\mathrm{km}^{-1}$, and the fibre loss coefficient $\alpha = 0.2\,\mathrm{dB/km}$. The simulation of the signal propagation was performed utilizing the splitstet Fourier method (SSFM) with 128 samples per symbol and 128 step per fibre span[5]. After each span, the Gaussian white noise was introduced. Its spectral power density is $N_{ASE} = (e^{\alpha L_A} - 1)h\nu_s NF/2$, with $L_A = 80\,\mathrm{km}$ being the span length, $h\nu_s$ the photon energy (h is Plank's constant and $\nu_s = 193.4\,\mathrm{THz}$ is the carrier's frequency), and $NF = 10^{4.5/10}$ a noise figure ($NF/2$ in the formula corresponds to the single polarization) [138].

First, data transmission performance with the chosen genus-4 solutions was estimated in terms of BER as a function of the signal's launch power after propagation on $1040\,\mathrm{km}$ (13 spans). The results are depicted in Fig. 3.11. The launch power was varied from $-4\,\mathrm{dBm}$ to $3\,\mathrm{dBm}$ that corresponded to the parameter a changed in the interval from 0.4 to 2. Two different sampling rates at the receiver (input of the CNN) were considered: $128\,\mathrm{GSam/s}$ and $64\,\mathrm{GSam/s}$. The first provided 128 samples per symbol (that had a duration of $1\,\mathrm{ns}$), while the second provided 64 samples. The value $64\,\mathrm{GSam/s}$ is achievable with modern hardware; thereby, this scenario can be implemented in a real system. Whereas $128\,\mathrm{GSam/s}$ is out of common availability, this

[4]The description of the normalization procedure is provided in [26]
[5]More details about SSFM can be found in [1]

regime was studied to observe the system's behavior in an ideal scenario. However, in both settings, the system demonstrated the data transmission with BER below a hard decision FEC (HD-FEC) threshold of 3.8×10^{-3} (a 7% overhead) [137]. The optimal power for the $128\,\mathrm{sam/sym}$ scenario (where BER achieved its minimum value) was $-1.72\,\mathrm{dBm}$ ($a = 0.6$). At the same time, $-0.57\,\mathrm{dBm}$ ($a = 0.75$) corresponded to the best performance for $64\,\mathrm{sam/sym}$ sampling rate. One can observe the remarkable behavior of the $128\,\mathrm{sam/sym}$ curve at low powers. It demonstrates smaller values of BER than expected if compared with the $64\,\mathrm{sam/sym}$ dependence. This can be explained by the ability of a CNN to filter noise [86]: more samples at the CNN input provide better filtering. In Fig. 3.12, the constellation diagram for the optimal power of $64\,\mathrm{sam/sym}$ scenario is presented.

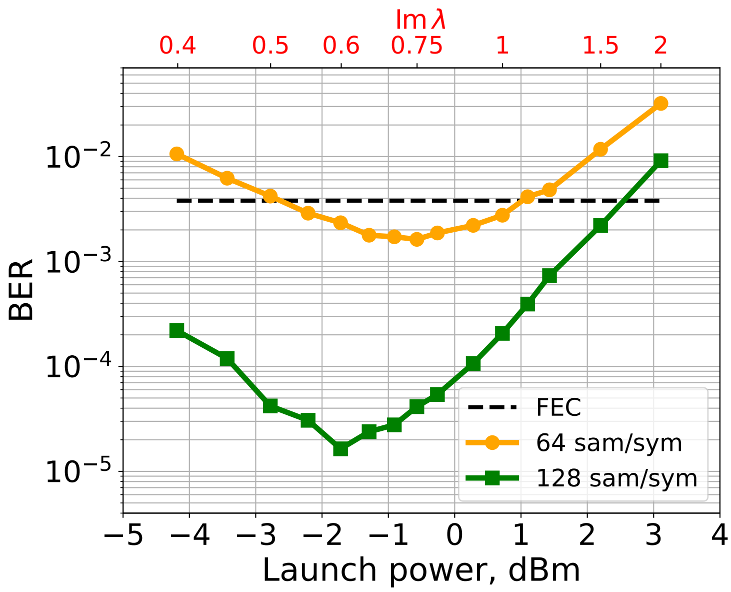

Figure 3.11: BER estimated in dependence on the signal's launch power after propagation on $1040\,\mathrm{km}$ (13 spans). Two different sampling rates at the receiver $128\,\mathrm{sam/sym}$ (green) and $64\,\mathrm{sam/sym}$ (yellow) are considered. The HD-FEC threshold is 3.8×10^{-3}. The upper x-axis depicts the values of parameter a corresponding to the power on the low x-axis. (Cited from [J2].)

Second, the performance of both scenarios – $128\,\mathrm{sam/sym}$ and $64\,\mathrm{sam/sym}$ sampling rates at their optimal powers – was evaluated as a function of distance (the number of fibre spans); see Fig. 3.13. When 128 samples per symbol were applied, BER remained below the HD-FEC threshold up to the distance $1840\,\mathrm{km}$ provided with 23 spans. The maximum number of spans with data signaling below the threshold for the $64\,\mathrm{sam/sym}$ sampling rate was 15, corresponding to $1200\,\mathrm{km}$.

In the following experiment, the effectiveness in terms of spectral efficiency was estimated. The scenario with the parameter $a = 0.75$ corresponding to the optimal power for $64\,\mathrm{sam/sym}$ sampling rate was considered. The linear spectrum of the signal at

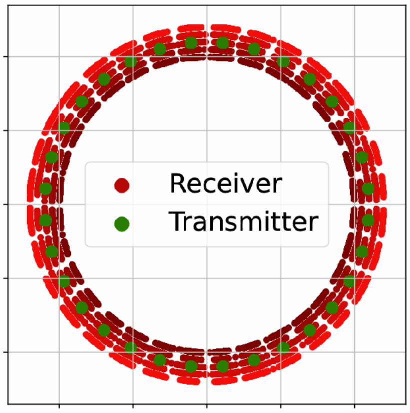

Figure 3.12: The distribution of the phases at the receiver at the optimal power for $64\,\mathrm{sam/sym}$ scenario, $a = 0.75$ (see Fig. 3.11). The shades of red depict the phases from different nonlinear harmonics, while the green points are referred values at the transmitter. (Cited from [J2].)

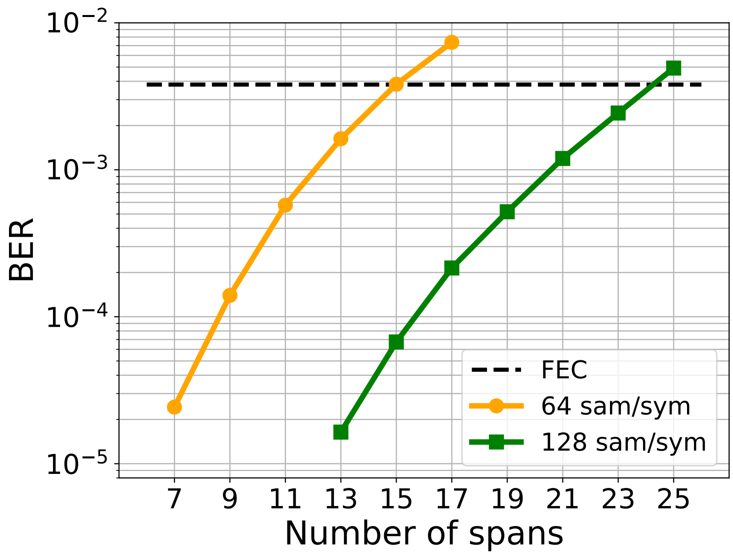

Figure 3.13: BER for different numbers of optical fibre spans (each $80\,\mathrm{km}$). The $128\,\mathrm{sam/sym}$ scenario is depicted in green, while $64\,\mathrm{sam/sym}$ is in yellow. The HD-FEC threshold equal to 3.8×10^{-3} (7$\%$ overhead) also shown. (Cited from [J2].)

the transmitter consisting of 1020 symbols is plotted; see Fig. 3.14. The overtone-like patterns on the sides effectively broaden the spectrum (depicted in grey in Fig. 3.14). 99% of energy is contained in the bandwidth of $21\,\mathrm{GHz}$. Such value makes the time-bandwidth product unacceptably large to use in practice. However, the CNN-based receiver can be trained to process the signals with a truncated spectrum. The narrow-band filter was applied at the transmitter to cut undesirable side structures in the spectrum; see the scheme in Fig. 3.10. The same filter at the receiver cleaned the signal from noise. Even in the scenario when all overtone-like structures are removed, the central part of the spectrum (green in Fig. 3.14) still contains the majority of the energy

(95%). Therefore, the properties of finite-genus solutions are kept.

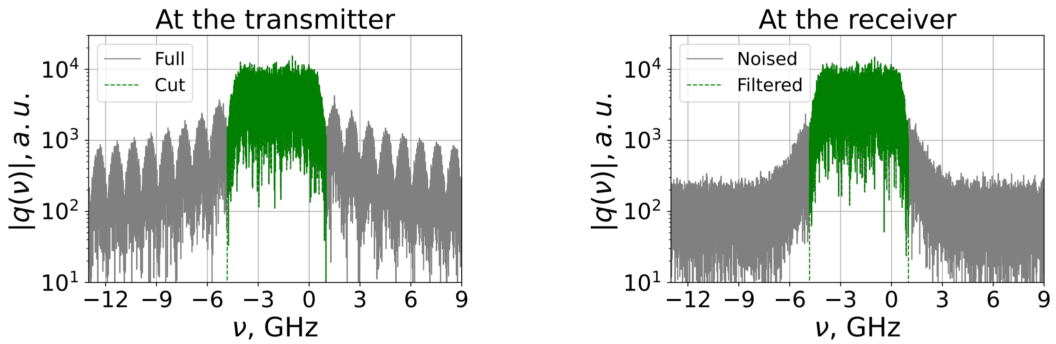

Figure 3.14: The linear spectra for the signal at the optimal power $a = 0.75$ at the transmitter (left) and receiver (right). The grey depicts the spectrum before applying a narrow-band filter, and the green is after filtering. (Cited from [J2].)

To see how truncation of the spectrum affects the performance, the scenarios with different numbers of overtones were considered, starting from zero and finishing with five overtones on each side. A wider spectrum provides less deviation from the original finite-genus signals. In other words, such systems keep the properties of finite-genus solutions but suffer from a high time-bandwidth product and are more affected by noise. For the signals transmitted with different numbers of overtones (different bandwidth), SE and BER were calculated for the practical case with $64\,\mathrm{sam/sym}$ sampling rate and $1040\,\mathrm{km}$ $(13$ spans) propagation distance. The results are presented in Table 3.4. To estimate SE, the following expression was used:

$$SE = \frac{5 \times AIR_{av}}{T_s \times B}, \tag{3.5}$$

where T_s and B are the corresponding symbol duration and bandwidth. The achievable information rate (AIR) was calculated assuming the received symbols have Gaussian statistics and using the method described in [139]. AIR_{av} was averaged over all symbols in the received signal and all nonlinear modes in one symbol (5 modes per symbol). Therefore, $5 \times AIR_{av}$ is the average value of the achievable information rate per symbol.

Number of overtones	0	1	2	3	4	5
Bandwidth, GHz	5.83	7.95	9.96	11.96	13.95	15.97
SE, bits/s/Hz	4.28	3.14	2.51	2.09	1.79	1.56
BER, $\times 10^{-5}$	8.64	2.95	2.10	2.39	2.66	2.99

Table 3.4: SE, and BER for signals with different bandwidths (cited from [J2]).

The scenarios with the truncated spectra provided better system performance (low BER) in comparison with the full bandwidth case with $\mathrm{BER} = 1.63 \times 10^{-3}$ (see Fig. 3.11).

This is explained by effective noise filtering. When all overtones were removed, and only the central part of the spectrum was used, the performance degradation related to the deviation from exact finite-genus solutions was observed. However, this regime still resulted in data transmission with $\mathrm{BER} = 8.64 \times 10^{-5}$ significantly smaller than the HD-FEC threshold. The achieved spectral efficiency in the system was $4.28\,\mathrm{bits/s/Hz}$. With the occupied bandwidth of $5.83\,\mathrm{GHz}$ the capacity of the system was $25\,\mathrm{Gbits/s}$. To increase the system's throughput, a wider bandwidth needs to be exploited. This can be implemented by increasing the genus of solutions that carry information or by occupying bandwidth with many channels of the type considered in this study. It is worth noting that reducing the symbol's duration to have a larger bandwidth requires a higher sampling rate to provide the same oversampling factor and reliable work of the CNN-based receiver. The optimization and dimensionality reduction of the CNN input are out of the scope of the current research.

Investigating the communication system described in this section and estimating its effectiveness are key results of this chapter. To my knowledge, SE reported here is the maximum achieved in the systems with the periodic NFT. Therefore, it is useful to compare this approach with others implemented based on the conventional NFT[6]. The first example is the data transmission with the b-modulation method and noise-signal correlation analysis [83]. The transmission parameters are close to the ones characterizing the system from this chapter: $5\,\mathrm{GHz}$ bandwidth and $960\,\mathrm{km}$ propagation distance, providing spectral efficiency of $5.51\,\mathrm{bits/s/Hz}$. Second, NFT scheme with nonlinear frequency division multiplexing over both polarizations and the Hermite-Gaussian-based carriers [69]. SE of $12\,\mathrm{bits/s/Hz}$ (over both polarizations) was reported while the bandwidth was $4.75\,\mathrm{GHz}$ and the transmission distance was $800\,\mathrm{km}$. The system presented in this chapter showed a lower but comparable SE than these referred values. However, the communication systems based on the periodic NFT and finite-genus solutions still require further investigation.

3.3.3 Complexity reduction of the CNN receiver

Compression techniques such as weight clustering, pruning, and quantization can substantially reduce neural network complexity. A detailed description of these methods can be found in [124, 140, 141, 142]. Application of these techniques specifically to the convolutional neural networks is given in [123]. The computational complexity of the CNN-based receiver exploited in the previous sections was studied independently. The weight clustering compression technique was applied to simplify the neural network. It was published in the conference paper [CP2].

First, the computational complexity analysis of the CNN at the receiver was per-

[6]These particular works are mentioned in the introduction.

formed. The complexity of prediction in terms of the number of real-valued multiplications was calculated. Training of the NN can be conducted offline, and therefore, the complexity of this stage is out of interest. In the prediction phase, an initial signal passes through all layers of the CNN consecutively. For the 1-D convolutional layer (CL), the complexity is expressed as:

$$C_{CNN} = \left(\frac{n_{in} + 2p - d(n_k - 1) - 1}{s} + 1 \right) \times n_k \times n_{feat} \times n_{filt}, \tag{3.6}$$

with n_{in} being the input vector's size, n_k size of the kernel, n_{filt} the number of filters, and n_{feat} the number of features. Parameters p, d, and s are padding, dilation, and stride. A fully connected layer (FCL) has the following number of multiplications:

$$C_{FCL} = n_{in} \times n_n, \tag{3.7}$$

with n_{in} being layer's input length, and n_n the number of neurons in the layer. The total complexity of CNN is just the sum of independent layers. There are three 1-D convolutional layers, one fully connected and one output layer in the considered architecture. Therefore, the total computational complexity is:

$$C_{CNN} = C_{CL\,1} + C_{CL\,2} + C_{CL\,3} + C_{FCL} + C_{out}. \tag{3.8}$$

The CNN considered in the previous section[7], CL1 (the first convolutional layer) has the length of the input $n_{in} = 3 \times 128$ or $n_{in} = 3 \times 64$ because three consecutive symbols are processed. The signal's real and imaginary parts (in-phase and quadrature components) provide two features, $n_{feat} = 2$. The input for CL2 is determined by the number of steps a kernel takes over the input on the previous layer (expression in the brackets in (3.6)), while the number of features equals the number of filters from the previous layer. The input of CL3 is defined similarly. The stride $s = 2$ for 3×128 input and $s = 1$ for 3×64 input for all convolutional layers as well as $p = 0$ and $d = 1$ in all scenarios. The output of CL3 is flattened before being fed into the FCL, which processes a one-dimensional array. The output layer consists of ten neurons and returns the real and imaginary parts of $e^{i\phi_j}$ for the predicted phase $\{\phi_j, j = 0, .., 4\}$ of a genus-4 solution. A more detailed analysis of the complexity can be found in [123, 124].

Below are the results of applying the weight clustering technique to the CNN model. The particular regime was investigated: the propagation at signal's launch power of $-0.6\,\mathrm{dBm}$ and a distance of $1040\,\mathrm{km}$ (13 spans). These settings were chosen to have the BER of the baseline model (that was below 10^{-4}) significantly smaller than the

[7] As an example, the CNN from section 3.3.2 is considered.

HD-FEC threshold (3.8×10^{-3} corresponding to a 7% overhead). Therefore, while the degradation in performance is expected for the clustered models, their BER can still be below the FEC threshold. As before, three consecutive symbols were on the NN input (each with 128 samples). BER was calculated after processing the signals with the CNN of different degrees of clustering to estimate the performance of compressed models and compare them with the baseline.

The performance of the clustered models in terms of BER as a function of the number of clusters (degree of compression) is in Fig. 3.15. The models demonstrate the performance below the HD-FEC starting from six clusters. Then, their efficiency increases while the number of clusters grows to 32. At this value, the clustered model behaves the same as the original one. This trend is obvious: the more clusters in the model, the better it approximates the weights distribution of the baseline model. A detailed description of the compression of this model with the weight clustering technique is given in the previous chapter; see section 2.4.

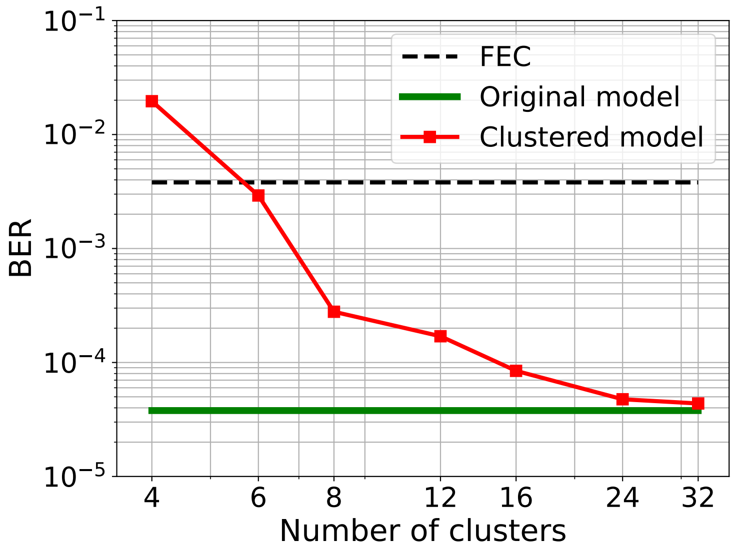

Figure 3.15: BER provided with the clustered model in dependence on the degree of compression (the number of clusters). For comparison, the BER of the baseline model and the HD-FEC threshold of 3.8×10^{-3} are provided. (Cited from [CP2].)

3.4 Chapter conclusion

A numerical investigation of fibre-optic communication systems based on finite-genus solutions of a generic type was performed in this chapter. This research was motivated by the lack of methods to use the full potential of finite-genus signals. In the frame of the periodic NFT, the particular implementations of such transmission systems were provided in works [33, 34]. However, as mentioned before, these approaches suffer from restrictions imposed by monodromy matrix formalism and force the systems to function in a quasi-linear regime. Moreover, the technique based on the RHP approach is limited to operating with the shortened interval of the phases from 0 to π. As a result, the effectiveness of the proposed data transmission systems with FGS as data carriers was underestimated. The method described in the previous chapter was adopted to provide a fair analysis of FGS-based communications. It is worth noting that an analytical approach to solve the direct problem for finite-genus solutions in a general case (not in a quasi-linear regime) still does not exist. Therefore, a neural network-based solver was developed to calculate the scattering data of finite-genus solutions.

In the first numerical experiment, the simulation of the fibre-optic communication system with the phase detection based on the proposed method was performed, 3.3.1 (also [CP1]). The phases of genus-4 solutions were used to encode information, while the main spectrum provided desirable signal parameters: power, duration, and bandwidth. The RHP approach generated information symbols at the transmitter. The data transmission was performed in the ideal conditions: information symbols were attributed with excessively long extension prefixes to exclude from consideration the distortions induced by dispersion, and signals propagated through a lossless fibre to avoid impairments caused by attenuation and amplification. In such settings, the conditions were close to correspond to the integrability of the NLSE and to keep the properties of finite-genus solutions. The method of phase recovery with the CNN was exploited at the receiver, which proved its feasibility and efficiency.

In the following work, the data transmission system was implemented in more realistic conditions, 3.3.2 or [J2]. To have higher spectral efficiency, the CNN at the receiver was harnessed to process symbols without extension prefixes. Moreover, analysis of the signals with truncated spectra was performed. All these techniques allowed a maximum spectral efficiency of $4.28\,\mathrm{bits/s/Hz}$ comparable with values provided by the conventional NFT. Also, a non-constant gain/loss profile was introduced into the system, making the analysis more practical. However, operation in such conditions significantly deviates the model from integrability, breaking the properties of finite-genus solutions. For these reasons, the CNN at the receiver was exploited to retrieve the phases of FGS and compensate for any deterministic effects corresponding to non-ideal conditions.

The core of the proposed receiver is a convolutional NN to solve the direct problem for FGS-GT. The neural network takes a sampled signal $q(t_k, z_0)$ for a fixed value of z_0 and returns the phases $\{\phi_j\}_{j=0}^{N}$ (where N is a genus of the solution). Although the method provided a valuable result and solved the direct problem of finite-genus solutions, being only an option due to the lack of analytical methods, implementing the NN-based calculation in hardware is still complicated. This is mainly because of the high computational complexity of NN models. The complexity reduction with weights clustering was performed to see how a compressed model can provide the phase retrieving and other receiver operations; see 3.3.3 or [CP2]. Whether CNN just solves the direct problem or functions as a receiver, it allows significant compression up to 99% (only 1% of multiplications of the original model can be used). Although a high degree of compression is available for the model, its implementation in hardware requires further exploration.

Despite the progress in developing data transmission systems with finite-genus solutions, there are still open problems that must be solved to construct effective communications. Here, I mention some of them to sketch the possible directions for further exploration. First, the approach developed here is based on the FGS to the NLSE. This means that only a single polarization is used for signal propagation, halving capacity. Therefore, constructing the system based on the finite-genus solution to Manakov's system, which provides propagation over both polarizations, is required. However, to the best of my knowledge, the theory behind these FGS has not been developed. Another direction of investigation is building high-genus solutions. They are important for effectively occupying bandwidth and, finally, for enhancing spectral efficiency. Some advanced approaches for such calculation were provided in [102]. However, more efforts are needed to make the method practical. Also, the following challenges can be addressed: implementation of data encoding with the amplitudes of nonlinear modes together with the phases. This potentially can increase the throughput of the information channel. The method proposed in the previous chapter can be easily expanded to predict the main spectrum (or the imaginary parts of the main spectrum defining the amplitudes of nonlinear harmonics). Finally, comprehensive optimization of finite-genus signals can be performed to fully utilize their potential.

Chapter 4

Conclusion

When I started this research, I aimed to provide a fair capacity estimation of fibre-optic communications based on the periodic NFT for the NLSE. A few works completed by that time reported underestimated performance. This is mainly because they were based on the periodic finite-genus solutions and, for this reason, were restricted to operate in a quasi-linear regime where the performance of the communications degrades due to the noise of optical amplifiers. The lack of theoretical methods to solve the direct problem for finite-genus solutions in the general case made it impossible to adopt the full potential of such signals. The direct problem solver was the only missing component in building the complete NFT framework for finite-genus solutions to the NLSE, while the inverse problem and evolution of scattering data were known. However, neural networks could approximate this transformation and provide a reliable technique to process finite-genus solutions of a generic type.

Chapter 2 is devoted to that method. It starts from the description of two ways to parameterize finite-genus solutions: the algebro-geometric approach and the Riemann-Hilbert problem-based method. The first approach allows for a signal parametrization with the main and auxiliary spectra, while the RHP method describes it in terms of the main spectrum and phases. Then, I introduce the finite-genus solutions of a generic type. To have a signal periodic, the frequencies of its nonlinear harmonics must be commensurable, or, in other words, some interval T_{com} that contains an integer number of periods for each nonlinear mode must exist. To retrieve the scattering data, the signal $q(t, z_0)$ needs to be defined on whole T_{com}. In this case, the direct problem of NFT can be solved with the monodromy matrix formalism. However, in practice, the periods of nonlinear modes are arbitrary, and T_{com} either does not exist or is too big to be processed with existing algorithms effectively. These signals are coined as finite-genus solutions of a generic type. Still, there is no theoretical approach to solve the direct problem in such a scenario. For this reason, a convolutional neural network was adopted to retrieve the scattering data of such finite-genus solutions.

An idea behind the approach is the following: among periods of all nonlinear modes for the given signal, the maximum one can be chosen as an interval on which $q(t, z_0)$ is defined. It guarantees that each nonlinear harmonic contains at least one period in this interval. Therefore, all information about the spectral data is contained in the signal. The finite-genus solutions of a generic type used for NN training were parametrized with the RHP approach, which means that the scattering data were the main spectrum and phases. The neural network was trained to predict the phases of a finite-genus signal having its samples in the time domain $q(t_n, z_0)$ at the NN input. A convolutional encoder-decoder model proved itself in other tasks in the frame of NFT and was chosen in this research. This architecture was then adjusted with the Bayesian optimization technique to provide better performance in my particular task. The feasibility and flexibility of the method were demonstrated with different finite-genus solutions: genus-4 $\mathrm{Im}[\lambda_j] = 1$, genus-$4$ $\mathrm{Im}[\lambda_j] = 5$, and genus-$8$ $\mathrm{Im}[\lambda_j] = 1$. The first example provided a moderate degree of complexity, the interaction of five nonlinear modes, but weak nonlinearity that does not contribute to the signal dynamic significantly. In the second scenario, genus-4 $\mathrm{Im}[\lambda_j] = 5$, the NN was tested in the regime with high nonlinearity changing the nature of signal evolution. Finally, the NN demonstrated its ability to process high-genus signals with genus-8 $\mathrm{Im}[\lambda_j] = 1$. In all scenarios, the error of the phase prediction was $10^{-3} - 10^{-2}\,\mathrm{rad}$, which, if it is required, can be decreased with more precise input signals, a bigger dataset, and a more complex NN model. Also, the weight clustering technique was applied to compress the NN, reducing the volume of memory that is needed to keep the model and its complexity in terms of the number of multiplications. The high-level compression of 99% was achieved while keeping the prediction performance compared to those of the uncompressed model.

Therefore, with this NN-based direct problem solver, the complete NFT framework for finite-genus solutions was implemented. The inverse problem, that is retrieving a signal from its spectral data, as well as the evolution of the scattering data, can be implemented with existing theoretical approaches, but the direct problem has been unsolved for the finite-genus solution of the generic type. Now, this gap has been filled with the proposed NN-based approach, making it possible to perform the full set of operations: retrieving the scattering data, their evolution, and calculating the signal $q(t, z)$ from them. Although the method was developed to recover the phases of signals only, following the initial goal of exploiting the finite-genus solutions in communications, it can also be expanded to calculate the main spectrum. Moreover, I am convinced that the approach can be applied in other areas where nonlinear waves dynamics are studied. Just because the framework did not exist until recently, the analysis of finite-genus solutions was constrained to specific low-genus scenarios, which allowed for analytical descriptions. The provided approach releases this limitation.

Chapter 3 describes the application of the NFT for the finite-genus solution of a generic type and the developed NN-based approach in fibre-optic communications. The data transmission systems based on finite-genus solutions were proposed before. However, these techniques relied on theoretical methods to solve the direct problem of the NFT and, therefore, operated only with a restricted set of periodic signals. This forced the systems to function in a quasi-linear regime. Moreover, the developed RHP-based communication method exploited a limited set of phases from 0 to π for modulation, halving the throughput. As a result, the capacity of the proposed data transmission systems with the periodic finite-genus solutions as data carriers was underestimated. The complete NFT framework, with the direct problem implemented through a convolutional NN, was adopted. This allowed the design of a communication system that is able to operate out of a quasi-linear regime, use the full set of phases for modulation, and, finally, provide a fair analysis of finite-genus solutions-based communications.

Initially, an idealized system was implemented just to demonstrate the applicability of the developed framework. The channel model was kept close to the integrable NLSE. The genus-4 solutions were used as data carriers. Their main spectrum provided the desirable signal parameters: power, duration, and bandwidth, while the phases were modulated with the transmitted information. At the transmitter the RHP approach generated information symbols. Each symbol had excessively long extension prefixes, diminishing the distortion associated with symbols overlapping due to dispersion. Moreover, signals propagated over the fibre with zero gain/loss profile, excluding impairments related to attenuation and amplification. At the receiver, the neural network retrieved the phases of propagated symbols with the following evolution compensation by the known analytic relation. This communication scenario proved the applicability of the proposed NN-based method for phase retrieving.

In the following analysis, a more practical transmission system was investigated. The same genus-4 solutions were exploited to encode information into their phases while the main spectrum controlled the signal parameters. To achieve maximum spectral efficiency, the symbols were transmitted without any extension prefixes. For the same reason, the signals at the transmitter also undergo truncation of their spectrum. A non-zero gain and loss were introduced, providing a realistic amplification/attenuation scenario based on the erbium-doped fibre amplification scheme. Operation in such settings deviates the channel model from the original integrable NLSE. This, in turn, compromises the properties of finite-genus solutions, breaking the trivial evolution of phases. To fix that problem, the neural network-based receiver was trained not only to retrieve the phases but also to compensate for their evolution and other deterministic phenomena related to non-ideal conditions. The considered data transmission approach demonstrated a high spectral efficiency of $4.28\,\mathrm{bits/s/Hz}$ over a transmission distance of $1040\,\mathrm{km}$ that is comparable with the results provided by the conventional NFT.

Despite an effective application of the proposed NFT framework for finite-genus solutions of a generic type in fibre-optic communications, there are a few open questions that must be solved to improve the performance of the systems. Double capacity can be achieved when using both polarizations for data transmission. In this scenario, Manakov's system describes signal propagation through an optical fibre. However, to the best of my knowledge, still, there are no developed approaches to calculate the finite-genus solutions for Manakov's system. One more way to design an effective communication system is exploiting high-genus solutions that requires the development of corresponding numerical routines. The high-genus solutions are important for effective bandwidth utilization to achieve maximum spectral efficiency.

Finally, the proposed neural network-based NFT for finite-genus solutions, as well as the communication systems analysis, can be improved through further investigations. A few possible research directions are stated above. I hope this modest work will serve as a starting point for new discoveries.

Bibliography

[1] G. P. Agrawal. *Nonlinear fiber optics*. Academic press, 5th edition, 2012.

[2] Y. S. Kivshar and G. P. Agrawal. *Optical solitons: from fibers to photonic crystals*. Academic press, 2003.

[3] L. Debnath. The Nonlinear Schrödinger Equation and Solitary Waves. In *Nonlinear Partial Differential Equations for Scientists and Engineers*, pages 515–556. Birkhäuser Boston, 2012.

[4] A. R. Osborne. Nonlinear Ocean Waves and the Inverse Scattering Transform. In *International Geophysics Series, vol. 97*. Elsevier, 2010.

[5] E. P. Gross. Structure of a quantized vortex in boson systems. *Il Nuovo Cimento (1955-1965)*, 20(3):454–477, 1961.

[6] L. P. Pitaevskii. Vortex lines in an imperfect Bose gas. *Sov. Phys. JETP*, 13(2): 451–454, 1961.

[7] A. C. Newell. *Solitons in mathematics and physics*. SIAM, 1985.

[8] A. S. Sharma and B. Buti. Envelope solitons and holes for sine-Gordon and nonlinear Klein-Gordon equations. *Journal of Physics A: Mathematical and General*, 9(11):1823, 1976.

[9] N. Karjanto. The nonlinear Schröodinger equation: A mathematical model with its wide-ranging applications. *arXiv preprint arXiv:1912.10683*, 2019.

[10] B. D. Fried and Y. H. Ichikawa. On the nonlinear Schrödinger equation for Langmuir waves. *Journal of the Physical Society of Japan*, 34(4):1073–1082, 1973.

[11] A. Hasegawa and F. Tappert. Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers. I. Anomalous dispersion. *Applied Physics Letters*, 23(3):142–144, 1973.

[12] F. Copie, S. Randoux, and P. Suret. The physics of the one-dimensional nonlinear Schrödinger equation in fiber optics: Rogue waves, modulation instability and self-focusing phenomena. *Reviews in Physics*, 5:100037, 2020.

[13] A. Hasegawa and F. Tappert. Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers. II. Normal dispersion. *Applied Physics Letters*, 23 (4):171–172, 1973.

[14] S. K. Turitsyn, B. G. Bale, and M. P. Fedoruk. Dispersion-managed solitons in fibre systems and lasers. *Physics reports*, 521(4):135–203, 2012.

[15] A. R. Its and V. P. Kotljarov. Explicit formulas for the solutions of a nonlinear Schrödinger equation. *Akademiia Nauk Ukrainskoi RSR Dopovidi Seriia Fiziko Matematichni ta Tekhnichni Nauki*, (11):965–968, 1976.

[16] M. J. Ablowitz and H. Segur. *Solitons and the inverse scattering transform*. SIAM, 1981.

[17] G. L. Lamb Jr. *Elements of soliton theory*. New York, 1980.

[18] L. F. Mollenauer and J. P. Gordon. *Solitons in optical fibers: fundamentals and applications*. Elsevier, 2006.

[19] J. Shao, X. Liang, and S. Kumar. Comparison of split-step Fourier schemes for simulating fiber optic communication systems. *IEEE Photonics Journal*, 6(4):1–15, 2014.

[20] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur. The inverse scattering transform-Fourier analysis for nonlinear problems. *Studies in applied mathematics*, 53(4):249–315, 1974.

[21] V. E. Zakharov and A. B. Shabat. Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media. *Sov. Phys. JETP*, 34(1):62, 1972.

[22] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura. Method for solving the Korteweg-deVries equation. *Physical Review Letters*, 19(19):1095, 1967.

[23] P. D. Lax. Integrals of nonlinear equations of evolution and solitary waves. In *Selected Papers Volume I*, pages 366–389. Springer, 2005.

[24] R. J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel. Capacity limits of optical fiber networks. *Journal of Lightwave Technology*, 28(4):662–701, 2010.

[25] A. Hasegawa and T. Nyu. Eigenvalue communication. *Journal of Lightwave Technology*, 11(3):395–399, 1993.

[26] S. K. Turitsyn, J. E. Prilepsky, S. T. Le, S. Wahls, L. L. Frumin, M. Kamalian, and S. A. Derevyanko. Nonlinear Fourier transform for optical data processing and transmission: advances and perspectives. *Optica*, 4(3):307–322, 2017.

[27] M. I. Yousefi and F. R. Kschischang. Information transmission using the nonlinear Fourier transform, Part I: Mathematical tools. *IEEE Transactions on Information Theory*, 60(7):4312–4328, 2014.

[28] V. B. Matveev. 30 years of finite-gap integration theory. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 366(1867):837–875, 2008.

[29] V. Kotlyarov and A. Its. Periodic problem for the nonlinear Schrödinger equation. *arXiv preprint arXiv:1401.4445*, 2014.

[30] V. Kotlyarov and D. Shepelsky. Planar unimodular Baker-Akhiezer function for the nonlinear Schrödinger equation. *Annals of Mathematical Sciences and Applications*, 2(2):343–384, 2017.

[31] E. D. Belokolos, A. I. Bobenko, V. Z. Enolskii, A. R. Its, and V. B. Matveev. *Algebro-geometric approach to nonlinear integrable equations*. Springer, 1994.

[32] P. G. Grinevich. Approximation theorem for the self-focusing nonlinear Schrödinger equation and for the periodic curves in R3. *Physica D: Nonlinear Phenomena*, 152:20–27, 2001.

[33] J.-W. Goossens, H. Hafermann, and Y. Jaouën. Data transmission based on exact inverse periodic nonlinear Fourier transform, Part I: Theory. *Journal of Lightwave Technology*, 38(23):6499–6519, 2020.

[34] M. Kamalian-Kopae, A. Vasylchenkova, D. Shepelsky, J. E. Prilepsky, and S. K. Turitsyn. Full-spectrum periodic nonlinear Fourier transform optical communication through solving the Riemann-Hilbert problem. *Journal of Lightwave Technology*, 38(14):3602–3615, 2020.

[35] G. P. Agrawal. *Fiber-optic communication systems*. John Wiley & Sons, Fifth edition, 2021.

[36] P. J. Winzer, D. T. Neilson, and A. R. Chraplyvy. Fiber-optic transmission and networking: the previous 20 and the next 20 years. *Optics express*, 26(18):24190–24239, 2018.

[37] R. J. Mears, L. Reekie, I. M. Jauncey, and D. N. Payne. Low-noise erbium-doped fibre amplifier operating at 1.54 μm. *Electronics letters*, 19(23):1026–1028, 1987.

[38] E. Agrell, M. Karlsson, A. R. Chraplyvy, D. J. Richardson, P. M. Krummrich, P. Winzer, K. Roberts, J. K. Fischer, S. J. Savory, B. J. Eggleton, et al. Roadmap of optical communications. *Journal of optics*, 18(6):063002, 2016.

[39] K. Kikuchi. Fundamentals of coherent optical fiber communications. *Journal of Lightwave Technology*, 34(1):157–179, 2015.

[40] T. Kobayashi, J. Cho, M. Lamponi, G. De Valicourt, and C. R. Doerr. Coherent optical transceivers scaling and integration challenges. *Proceedings of the IEEE*, 110(11):1679–1698, 2022.

[41] H. Sun, M. Torbatian, M. Karimi, R. Maher, S. Thomson, M. Tehrani, Y. Gao, A. Kumpera, G. Soliman, A. Kakkar, et al. 800G DSP ASIC design using probabilistic shaping and digital sub-carrier multiplexing. *Journal of Lightwave Technology*, 38(17):4744–4756, 2020.

[42] M. Kuschnerov, M. Chouayakh, K. Piyawanno, B. Spinnler, E. De Man, P. Kainzmaier, M. S. Alfiad, A. Napoli, and B. Lankl. Data-aided versus blind single-carrier coherent receivers. *IEEE Photonics Journal*, 2(3):387–403, 2010.

[43] C. Minkenberg, R. Krishnaswamy, A. Zilkie, and D. Nelson. Co-packaged datacenter optics: Opportunities and challenges. *IET optoelectronics*, 15(2):77–91, 2021.

[44] D. M. Marom, Y. Miyamoto, D. T. Neilson, and I. Tomkos. Optical switching in future fiber-optic networks utilizing spectral and spatial degrees of freedom. *Proceedings of the IEEE*, 110(11):1835–1852, 2022.

[45] D. J. Richardson, J. M. Fini, and L. E. Nelson. Space-division multiplexing in optical fibres. *Nature photonics*, 7(5):354–362, 2013.

[46] B. J. Puttnam, G. Rademacher, and R. S. Luís. Space-division multiplexing for optical fiber communications. *Optica*, 8(9):1186–1203, 2021.

[47] D. Soma, Y. Wakayama, S. Beppu, S. Sumita, T. Tsuritani, T. Hayashi, T. Nagashima, M. Suzuki, M. Yoshida, K. Kasai, et al. 10.16-Peta-b/s dense SDM/WDM transmission over 6-mode 19-core fiber across the C+ L band. *Journal of Lightwave Technology*, 36(6):1362–1368, 2018.

[48] T. Hoshida, V. Curri, L. Galdino, D. T. Neilson, W. Forysiak, J. K. Fischer, T. Kato, and P. Poggiolini. Ultrawideband systems and networks: beyond C+ L-band. *Proceedings of the IEEE*, 110(11):1725–1741, 2022.

[49] B. J. Puttnam, R. S. Luís, G. Rademacher, M. Mendez-Astudilio, Y. Awaji, and H. Furukawa. S, C and extended L-band transmission with doped fiber and distributed Raman amplification. In *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3. IEEE, 2021.

[50] B. J. Puttnam, R. S. Luis, I. Phillips, M. Tan, A. Donodin, D. Pratiwi, L. Dallachiesa, Y. Huang, M. Mazur, N. K. Fontaine, et al. 402 Tb/s GMI data-rate OESCLU-band transmission. In *Optical Fiber Communication Conference*, pages Th4A–3. Optica Publishing Group, 2024.

[51] F. Musumeci, C. Rottondi, A. Nag, I. Macaluso, D. Zibar, M. Ruffini, and M. Tornatore. An overview on application of machine learning techniques in optical networks. *IEEE Communications Surveys & Tutorials*, 21(2):1383–1408, 2018.

[52] P. Freire, E. Manuylovich, J. E. Prilepsky, and S. K. Turitsyn. Artificial neural networks for photonic applications—from algorithms to implementation: tutorial. *Advances in Optics and Photonics*, 15(3):739–834, 2023.

[53] Y. V. Kiran, T. Venkatesh, and C. S. R. Murthy. A reinforcement learning framework for path selection and wavelength selection in optical burst switched networks. *IEEE Journal on Selected Areas in Communications*, 25(9):18–26, 2007.

[54] C. Häger and H. D. Pfister. Nonlinear interference mitigation via deep neural networks. In *Optical Fiber Communication Conference*, pages W3A–4. Optica Publishing Group, 2018.

[55] M. A. Jarajreh, E. Giacoumidis, I. Aldaya, S. T. Le, A. Tsokanos, Z. Ghassemlooy, and N. J. Doran. Artificial neural network nonlinear equalizer for coherent optical OFDM. *IEEE Photonics Technology Letters*, 27(4):387–390, 2014.

[56] P. P. Mitra and J. B. Stark. Nonlinear limits to the information capacity of optical fibre communications. *Nature*, 411(6841):1027–1030, 2001.

[57] J. E. Prilepsky, S. A. Derevyanko, K. J. Blow, I. Gabitov, and S. K. Turitsyn. Nonlinear inverse synthesis and eigenvalue division multiplexing in optical fiber channels. *Physical review letters*, 113(1):013901, 2014.

[58] S. T. Le, J. E. Prilepsky, and S. K. Turitsyn. Nonlinear inverse synthesis technique for optical links with lumped amplification. *Optics express*, 23(7):8317–8328, 2015.

[59] S. A. Derevyanko, J. E. Prilepsky, and S. K. Turitsyn. Capacity estimates for optical transmission based on the nonlinear Fourier transform. *Nature Communications*, 7(1):12710, 2016.

[60] A. Maruta. Eigenvalue modulated optical transmission system. *2015 Opto-Electronics and Communications Conference (OECC)*, pages 1–3, 2015.

[61] S. Gaiarin, A. M. Perego, E. P. da Silva, F. Da Ros, and D. Zibar. Dual-polarization nonlinear Fourier transform-based optical communication system. *Optica*, 5(3): 263–270, 2018.

[62] Y. Bi, H. Xu, H. Gao, C. Bai, Y. Zhang, L. Xi, N. Cui, L. Yang, W. Sun, and X. Tang. Frequency offset modeling in presence of ASE noise and corresponding low-complexity solution for discrete spectrum modulated nonlinear frequency division multiplexing system. *Optics Communications*, 537:129396, 2023.

[63] T. Kodama, K. Mishina, Y. Yoshida, D. Hisano, and A. Maruta. Transmission of hyper-multi-level eigenvalue-modulated signal using arbitrary optical multi-eigenvalue. *Optics Communications*, 546:129748, 2023.

[64] R. Zhang, L. Xi, J. Wei, J. Deng, S. Du, W. Zhang, X. Zhang, and X. Xiao. Optimal design of eigenvalues for the full-spectrum modulated nonlinear frequency division multiplexing transmission system. *IEEE Photonics Journal*, 15(3):1–8, 2023.

[65] S. T. Le, V. Aref, and H. Buelow. Nonlinear signal multiplexing for communication beyond the Kerr nonlinearity limit. *Nature Photonics*, 11(9):570–576, 2017.

[66] F. Da Ros, S. Civelli, S. Gaiarin, E. P. Da Silva, N. De Renzis, M. Secondini, and D. Zibar. Dual-polarization NFDM transmission with continuous and discrete spectral modulation. *Journal of Lightwave Technology*, 37(10):2335–2343, 2019.

[67] T. Gui, G. Zhou, C. Lu, A. P. T. Lau, and S. Wahls. Nonlinear frequency division multiplexing with b-modulation: shifting the energy barrier. *Optics express*, 26 (21):27978–27990, 2018.

[68] S. A. Derevyanko and Jaroslaw E. Prilepsky. Channel model for the dual-polarization b-modulated nonlinear frequency-division multiplexing optical transmission systems. *Optics express*, 31(12):19686–19702, 2023.

[69] M. Balogun, S. Derevyanko, and L. Barry. Dual-polarization Hermite-Gaussian-based NFDM transmission system. *IEEE access*, 11:125692–125697, 2023.

[70] X. Yangzhang, S. T. Le, V. Aref, H. Buelow, D. Lavery, and P. Bayvel. Experimental demonstration of dual-polarisation NFDM transmission with b-modulation. In *2019 IEEE Photonics Conference (IPC)*, pages 1–4. IEEE, 2019.

[71] M. Kamalian, J. E. Prilepsky, S. T. Le, and S. K. Turitsyn. Periodic nonlinear Fourier transform for fiber-optic communications, Part I: theory and numerical methods. *Optics express*, 24(16):18353–18369, 2016.

[72] M. Kamalian, J. E. Prilepsky, S. T. Le, and S. K. Turitsyn. Periodic nonlinear Fourier transform for fiber-optic communications, Part II: eigenvalue communication. *Optics express*, 24(16):18370–18381, 2016.

[73] J.-W. Goossens, Y. Jaouën, and H. Hafermann. Experimental demonstration of data transmission based on the exact inverse periodic nonlinear Fourier transform. In *Optical Fiber Communication Conference*, pages M1I–6. Optica Publishing Group, 2019.

[74] E. G. Turitsyna and S. K. Turitsyn. Digital signal processing based on inverse scattering transform. *Optics letters*, 38(20):4186–4188, 2013.

[75] S. Wahls, S. T. Le, J. E. Prilepsky, H. V. Poor, and S. K. Turitsyn. Digital backpropagation in the nonlinear Fourier domain. In *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 445–449. IEEE, 2015.

[76] I. T. Lima, V. S. Grigoryan, M. O'Sullivan, and C. R. Menyuk. Computational complexity of nonlinear transforms applied to optical communications systems with normal dispersion fibers. In *2015 IEEE Photonics Conference (IPC)*, pages 277–278. IEEE, 2015.

[77] S. Wahls and H. V. Poor. Introducing the fast nonlinear Fourier transform. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5780–5784. IEEE, 2013.

[78] S. Wahls, S. Chimmalgi, and P. J. Prins. FNFT: A software library for computing nonlinear Fourier transforms. *Journal of Open Source Software*, 3(23):597, 2018.

[79] S. Wahls and V. Vaibhav. Fast inverse nonlinear Fourier transforms for continuous spectra of Zakharov-Shabat type. *arXiv preprint arXiv:1607.01305*, 2016.

[80] V. Vaibhav and S. Wahls. Introducing the fast inverse NFT. In *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3. IEEE, 2017.

[81] S. T. Le, J. E. Prilepsky, and S. K. Turitsyn. Nonlinear inverse synthesis for high spectral efficiency transmission in optical fibers. *Optics express*, 22(22):26720–26741, 2014.

[82] F. Da Ros, S. Gaiarin, R. T. Jones, O. Jovanovic, E. P. da Silva, and D. Zibar. Nonlinear Fourier transform: perpetual research topic or future game-changer?

In *Signal Processing in Photonic Communications*, pages SpTu2I–1. Optica Publishing Group, 2020.

[83] Q. Zhang and F. R. Kschischang. Correlation-aided nonlinear spectrum detection. *Journal of Lightwave Technology*, 39(15):4923–4931, 2021.

[84] R. T. Jones, S. Gaiarin, M. P. Yankov, and D. Zibar. Time-domain neural network receiver for nonlinear frequency division multiplexed systems. *IEEE Photonics Technology Letters*, 30(12):1079–1082, 2018.

[85] W. Q. Zhang, T. H. Chan, and S. A. Vahid. Serial and parallel convolutional neural network schemes for NFDM signals. *Scientific Reports*, 12(1):7962, 2022.

[86] E. V. Sedov, P. J. Freire, V. V. Seredin, V. A. Kolbasin, M. Kamalian-Kopae, I. S. Chekhovskoy, S. K. Turitsyn, and J. E. Prilepsky. Neural networks for computing and denoising the continuous nonlinear Fourier spectrum in focusing nonlinear Schrödinger equation. *Scientific Reports*, 11(1):22857, 2021.

[87] M. Kamalian-Kopae, A. Vasylchenkova, O. Kotlyar, M. Pankratova, J. Prilepsky, and S. Turitsyn. Artificial neural network-based equaliser in the nonlinear Fourier domain for fibre-optic communication applications. In *2019 Conference on Lasers and Electro-Optics Europe & European Quantum Electronics Conference (CLEO/Europe-EQEC)*, pages 1–1. IEEE, 2019.

[88] S. Yamamoto, K. Mishina, and A. Maruta. Demodulation of optical eigenvalue modulated signals using neural networks. *IEICE Communications Express*, 8 (12):507–512, 2019.

[89] O. Kotlyar, M. Pankratova, M. Kamalian, A. Vasylchenkova, J. E. Prilepsky, and S. K. Turitsyn. Unsupervised and supervised machine learning for performance improvement of NFT optical transmission. In *2018 IEEE British and Irish conference on optics and photonics (BICOP)*, pages 1–4. IEEE, 2018.

[90] W. Q. Zhang, T. H. Chan, and S. Afshar. A neural network for forward and inverse nonlinear Fourier transforms for fibre optic communication. *Optics & Laser Technology*, 176:110971, 2024.

[91] A. Its. The Riemann-Hilbert problem and integrable systems. *Notices of the AMS*, 50(11):1389–1400, 2003.

[92] T. Trogdon and S. Olver. *Riemann–Hilbert problems, their numerical solution, and the computation of nonlinear special functions*. SIAM, 2015.

[93] T. Trogdon and B. Deconinck. Numerical computation of the finite-genus solutions of the Korteweg–de Vries equation via Riemann–Hilbert problems. *Applied Mathematics Letters*, 26(1):5–9, 2013.

[94] T. Trogdon and S. Olver. Numerical inverse scattering for the focusing and defocusing nonlinear Schrödinger equations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 469(2149):20120330, 2013.

[95] S. Olver. A general framework for solving Riemann–Hilbert problems numerically. *Numerische Mathematik*, 122:305–340, 2012.

[96] A. Vasylchenkova. *Nonlinear Fourier Transform in application to long-haul optical communications*. PhD thesis, Aston University, 2020.

[97] D. Bilman, P. Nabelek, and T. Trogdon. Computation of large-genus solutions of the Korteweg–de Vries equation. *Physica D: Nonlinear Phenomena*, 449:133715, 2023.

[98] Y.-C. Ma and M. J. Ablowitz. The periodic cubic Schrödinger equation. *Studies in applied Mathematics*, 65(2):113–158, 1981.

[99] M. Kamalian-Kopae. *On nonlinear Fourier transform-based fibre-optic communication systems for periodic signals*. PhD thesis, Aston University, 2018.

[100] S. Wahls and H. V. Poor. Fast numerical nonlinear Fourier transforms. *IEEE Transactions on Information Theory*, 61(12):6957–6974, 2015.

[101] E. R. Tracy. *Topics in nonlinear wave theory with applications*. PhD thesis, University of Maryland, 1984.

[102] S. Chimmalgi and S. Wahls. On computing high-dimensional Riemann theta functions. *Communications in Nonlinear Science and Numerical Simulation*, 123: 107266, 2023.

[103] B. Deconinck, M. Heil, A. Bobenko, M. Van Hoeij, and M. Schmies. Computing Riemann theta functions. *Mathematics of Computation*, 73(247):1417–1442, 2004.

[104] T. D. Trogdon. *Riemann-Hilbert problems, their numerical solution and the computation of nonlinear special functions*. PhD thesis, University of Washington, 2013.

[105] M. Kamalian, A. Vasylchenkova, D. Shepelsky, J. E. Prilepsky, and S. K. Turitsyn. Signal modulation and processing in nonlinear fibre channels by employing the Riemann–Hilbert problem. *Journal of Lightwave Technology*, 36(24):5714–5727, 2018.

[106] A. I. Bobenko. *Computational approach to Riemann surfaces*, volume 2013. Springer Science & Business Media, 2011.

[107] A. O. Smirnov. Periodic two-phase “rogue waves”. *Mathematical Notes*, 94:897–907, 2013.

[108] A. Kamchatnov. New approach to periodic solutions of integrable equations and nonlinear theory of modulational instability. *Physics Reports*, 286(4):199–270, 1997.

[109] S. Olver. RHPackage. `https://github.com/dlfivefifty/RHPackage`, 2010.

[110] S. Olver. RiemannHilbert.jl. `https://github.com/JuliaHolomorphic/RiemannHilbert.jl`, 2015.

[111] B. Deconinck, A. S. Fokas, and J. Lenells. The implementation of the unified transform to the nonlinear Schrödinger equation with periodic initial conditions. *Letters in Mathematical Physics*, 111(1):17, 2021.

[112] A. S. Fokas and J. Lenells. A new approach to integrable evolution equations on the circle. *Proceedings of the Royal Society A*, 477(2245):20200605, 2021.

[113] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

[114] A. Van Den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 12, 2016.

[115] M. Pelikan, D. E. Goldberg, E. Cantú-Paz, et al. BOA: The Bayesian optimization algorithm. In *Proceedings of the genetic and evolutionary computation conference GECCO-99*, volume 1, pages 525–532. Citeseer, 1999.

[116] S. Han, H. Mao, and W. J. Dally. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. *arXiv preprint arXiv:1510.00149*, 2015.

[117] S. Son, S. Nah, and K. M. Lee. Clustering convolutional kernels to compress deep neural networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 216–232, 2018.

[118] T. Liang, J. Glossner, L. Wang, S. Shi, and X. Zhang. Pruning and quantization for deep neural network acceleration: a survey. *Neurocomputing*, 461:370–403, 2021.

[119] D. Blalock, J. J. Gonzalez Ortiz, J. Frankle, and J. Guttag. What is the state of neural network pruning? *Proceedings of Machine Learning and Systems*, 2: 129–146, 2020.

[120] M. Augasta and T. Kathirvalavakumar. Pruning algorithms of neural networks—a comparative study. *Open Computer Science*, 3(3):105–115, 2013.

[121] A. Gholami, S. Kim, Z. Dong, Z. Yao, M. W. Mahoney, and K. Keutzer. A survey of quantization methods for efficient neural network inference. In *Low-Power Computer Vision*, pages 291–326. Chapman and Hall/CRC, 2022.

[122] O. Weng. Neural network quantization for efficient inference: A survey. *arXiv preprint arXiv:2112.06126*, 2021.

[123] P. J. Freire, A. Napoli, B. Spinnler, M. Anderson, D. A. Ron, W. Schairer, T. Bex, N. Costa, S. K. Turitsyn, and J. E. Prilepsky. Reducing computational complexity of neural networks in optical channel equalization: From concepts to implementation. *Journal of Lightwave Technology*, 41(14):4557–4581, 2023.

[124] P. J. Freire, Y. Osadchuk, B. Spinnler, A. Napoli, W. Schairer, N. Costa, J. E. Prilepsky, and S. K. Turitsyn. Performance versus complexity study of neural network equalizers in coherent optical systems. *Journal of Lightwave Technology*, 39(19):6085–6096, 2021.

[125] S. V. Manakov. On the theory of two-dimensional stationary self-focusing of electromagnetic waves. *Soviet Physics-JETP*, 38(2):248–253, 1974.

[126] D. Marcuse, C. R. Manyuk, and P. K. A. Wai. Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence. *Journal of Lightwave Technology*, 15(9):1735–1746, 1997.

[127] P. L. Christiansen, J. C. Eilbeck, V. Z. Enolskii, and N. A. Kostov. Quasi–periodic and periodic solutions for coupled nonlinear Schrödinger equations of Manakov type. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 456(2001):2263–2281, 2000.

[128] L. Wu, X. Geng, and G. He. Algebro-geometric solutions to the Manakov hierarchy. *Applicable Analysis*, 95(4):769–800, 2016.

[129] J.-W. Goossens. *Advances in nonlinear Fourier transform based optical fiber communication*. PhD thesis, Institut polytechnique de Paris, 2020.

[130] M. Kamalian, A. Vasylchenkova, J. Prilepsky, D. Shepelsky, and S. Turitsyn. Communication system based on periodic nonlinear Fourier transform with exact in-

verse transformation. In *2018 European Conference on Optical Communication (ECOC)*, pages 1–3. IEEE, 2018.

[131] M. Kamalian, J. Prilepsky, A. Vasylchenkova, D. Shepelsky, and S. Turitsyn. Methods of nonlinear Fourier-based optical transmission with periodically-extended signals. In *2018 IEEE International Conference on the Science of Electrical Engineering in Israel (ICSEE)*, pages 1–5. IEEE, 2018.

[132] M. Kamalian, J. E. Prilepsky, S. T. Le, and S. K. Turitsyn. Periodic nonlinear Fourier transform based transmissions with high order QAM formats. In *ECOC 2016; 42nd European Conference on Optical Communication*, pages 1–3. VDE, 2016.

[133] M. Kamalian, J. E. Prilepsky, S. T. Le, and S. K. Turitsyn. Spectral efficiency estimation in periodic nonlinear Fourier transform based communication systems. In *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3. IEEE, 2017.

[134] M. Kamalian-Kopae, A. Vasylchenkova, D. Shepelsky, J. Prilepsky, and S. Turitsyn. Multisymbol periodic nonlinear Fourier transform communication. In *45th European Conference on Optical Communication (ECOC 2019)*, pages 1–4. IET, 2019.

[135] O. Kotlyar, M. Kamalian-Kopae, J. E. Prilepsky, M. Pankratova, and S. K. Turitsyn. Machine learning for performance improvement of periodic NFT-based communication system. In *45th European Conference on Optical Communication (ECOC 2019)*, pages 1–4. IET, 2019.

[136] M. Kamalian, J. E. Prilepsky, S. T. Le, and S. K. Turitsyn. On the design of NFT-based communication systems with lumped amplification. *Journal of Lightwave Technology*, 35(24):5464–5472, 2017.

[137] E. Agrell and M. Secondini. Information-theoretic tools for optical communications engineers. In *2018 IEEE Photonics Conference (IPC)*, pages 1–5. IEEE, 2018.

[138] G. P. Agrawal. *Fiber-optic communication systems*. John Wiley & Sons, 4th edition, 2012.

[139] P. J. Freire, J. E. Prilepsky, Y. Osadchuk, S. K. Turitsyn, and V. Aref. Deep neural network-aided soft-demapping in coherent optical systems: regression versus classification. *IEEE Transactions on Communications*, 70(12):7973–7988, 2022.

[140] D. A. Ron, P. J. Freire, J. E. Prilepsky, M. Kamalian-Kopae, A. Napoli, and S. K. Turitsyn. Experimental implementation of a neural network optical channel equal-

izer in restricted hardware using pruning and quantization. *Scientific Reports*, 12 (1):8713, 2022.

[141] M. M. Melek and D. Yevick. Nonlinearity mitigation with a perturbation based neural network receiver. *Optical and Quantum Electronics*, 52:1–10, 2020.

[142] Z. Wan, J. Li, L. Shu, M. Luo, X. Li, S. Fu, and K. Xu. Nonlinear equalization based on pruned artificial neural networks for 112-Gb/s SSB-PAM4 transmission over 80-km SSMF. *Optics express*, 26(8):10631–10642, 2018.