Silicon Photonic Integrated Circuits for Soliton Based Long Haul Optical Communication

Alvaro Moscoso-Mártir, Jonas Koch, Olaf Schulz, Juliana Müller, Ali Tabatabaei Mashayekh, Arka D. Das, *Graduate Student Member, IEEE*, Florian Merget, Stephan Pachnicke, *Senior Member, IEEE*, and Jeremy Witzens, *Senior Member, IEEE*

(Invited Paper)

***Abstract*—We demonstrate a silicon photonic transmitter for soliton-based communications enabling the interleaving of QPSK and APSK modulated soliton pulses overlapping in both the time and frequency domains. The nonlinear Fourier transform is used to both determine adequate soliton launch conditions and facilitate nonlinear equalization at the receiver. A spectral efficiency of 0.5 b/s/Hz per polarization is reached over a 5400 km transmission distance and constitutes, to the best of our knowledge, a record for soliton transmission over such long distances. Linear and nonlinear minimum mean square error and artificial neural network-based equalizers are also experimentally benchmarked, with artificial neural networks resulting in particularly effective equalization in this system.**

***Index Terms*—Nonlinear Fourier transform, nonlinear optics, optical fiber communication, optical solitons, optical transmitters, photonic integrated circuits, silicon photonics.**

I. Introduction

NONLINEAR channel distortions such as phase modulation induced by the Kerr-effect currently prevent increase in signal power beyond a few mW from improving spectral efficiencies [1]. Hence, efficient solutions need to be found to mitigate nonlinear impairments and to shift upwards the nonlinear capacity limit.

Soliton pulses have been an early approach to accommodate fiber nonlinearities. They balance fiber dispersion with Kerr-nonlinearities and maintain their shape as they propagate down the fiber [2], [3]. Due to problems such as soliton-collision induced timing jitter in wavelength division multiplexed (WDM) links [4] and the low spectral efficiencies of the solitons themselves, the interest in solitons for optical communications waned over the years. However, recent developments in coherent detection and in the signal modulation paradigm based on the Zakharov-Shabat scattering problem – referred to as the nonlinear Fourier transform (NFT) [5]–[7] – have renewed the interest in solitons. The NFT decomposes an optical signal into a discrete and continuous spectrum, wherein the first corresponds to solitonic transmission and the second to dispersive waves to which the toolkit of conventional coherent communications can be applied. The NFT enables joint exploitation of both the continuous and discrete spectra [8], [9], so that soliton based transmission could potentially become a "top-off", increasing the overall channel capacity. While it has not yet been possible to implement this in an effective manner leading to an increase of spectral efficiencies beyond that of state-of-the-art coherent communications, in particular due to interaction between the continuous and discrete spectra in the presence of noise that requires advanced equalization techniques [10] as well as signal power limitations in case of *b*-modulation [11], this remains an active field of investigation.

Here, we tackle the partial problem of increasing the spectral efficiency of soliton transmission with an architecture that is scalable to filling an entire communication band. The proposed solution reaches a spectral efficiency of 0.5 b/s/Hz for a single polarization over a distance of 5400 km, to the best of our knowledge a record for such long distances. To achieve this spectral efficiency, we apply amplitude phase shift keying (APSK) modulation to fundamental soliton pulses that are densely packed in both frequency and time and launched by an integrated silicon photonics transmitter implementing four soliton channels.

In order to place this performance in context, we start by briefly reviewing the state-of-the-art of soliton transmission in Section II. We then proceed by describing the architecture concept (Section III), its implementation (Section IV), and the experimental transmission results (Section V). Two appendices provide the reader with the mathematical framework for the utilized data processing techniques.

Manuscript received August 31, 2021; revised December 1, 2021 and March 15, 2022; accepted April 18, 2022. Date of publication April 26, 2022; date of current version May 25, 2022. This work was supported by the German Research Foundation through Electronic-Photonic Integrated Systems for Ultrafast Signal Processing Program and Nonlinear Fourier Transform based Optical Transmission using Electronic-Photonic Signal Processing Project under Grants PA1705/1-2 and JW4137/X-2. *(Corresponding author: Alvaro Moscoso-Mártir.)*

Alvaro Moscoso-Mártir, Juliana Müller, Ali Tabatabaei Mashayekh, Arka D. Das, Florian Merget, and Jeremy Witzens are with the Institute of Integrated Photonics, RWTH Aachen University, 52074 Aachen, Germany (e-mail: amartir@iph.rwth-aachen.de; jmueller@iph.rwth-aachen.de; amashayekh@iph.rwth-aachen.de; adas@iph.rwth-aachen.de; fmerget@iph.rwth-aachen.de; jwitzens@iph.rwth-aachen.de).

Jonas Koch, Olaf Schulz, and Stephan Pachnicke are with the Chair of Communications, Kiel University, 24143 Kiel, Germany (e-mail: jonas.koch@tf.uni-kiel.de; olaf.schulz@tf.uni-kiel.de; stephan.pachnicke@tf.uni-kiel.de).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JLT.2022.3170250.

Digital Object Identifier 10.1109/JLT.2022.3170250

II. State-of-the-Art Soliton Transmission

A lot of attention has been given to soliton-based transmission in the late 1990s [12]–[16], with a particular focus on dispersion managed solitons (DMS) [17] to better deal with the constraints of lumped amplification. Indeed, in a link with lumped amplifiers, the distance between them is normally required to stay much below the length over which solitons of different channels collide [4]. Otherwise, soliton collisions result in a permanent shift of their center frequencies, which in turn results in the Gordon-Haus timing jitter after propagation through the remaining link [18]. Since the collision length scales with the pulse duration, this critically limits the minimum pulse duration and baud rate for a given distance between amplifiers. In DMS links, on the other hand, balancing of the dispersion in each fiber segment removes this requirement, so that much higher baud rates can be obtained. This way, high single polarization pulse rates of 40 GBd were reached and further doubled by interleaving pulses with alternating polarizations [16]. Using this polarization interleaving technique, that reduces nonlinear interaction between the pulses and thus enables closer pulse packing [14], spectral efficiencies as high as 0.4 b/s/Hz were obtained in dense (D)WDM systems with aggregate bandwidths of 1.02 Tb/s [13] and 640 Gb/s [16] transported over 1000 km. Higher transmission distances of 9000 km were also achieved, albeit at a reduced spectral efficiency of 0.133 b/s/Hz [15]. While in this case polarization diversity was not used in a given WDM channel, adjacent channels were again multiplexed together with orthogonal polarizations. The authors of that publication modeled that an increased spectral efficiency of 0.2 b/s/Hz might be achievable with their architecture by reducing the channel spacing.

Rekindled interest in eigenvalue communications has arisen more recently in the context of coherent communications. The new technological context has however led to a different set of constraints. For one, most works now focus on non-dispersion managed links with lumped amplification, in which stretches of a single type of fiber – typically non-zero dispersion shifted fiber (NZDSF) – are simply combined with an erbium doped fiber amplifier (EDFA) without alternating the fiber type. This limits the baud rate, as explained above, so that the focus has shifted from obtaining high serial data rates to realizing the best possible spectral efficiencies under these conditions. Swapping the polarization of spectrally adjacent solitons, so as to reduce their nonlinear interaction while leaving corresponding spectral gaps of unused spectrum in the complementary polarization, is also no longer an attractive technique in contemporary coherent links in which both polarizations can be exploited. Lastly, instead of generating solitons from the pulses of mode-locked lasers or by using separate analog pulse carvers [16], they are now directly synthesized in the digital domain and generated by an arbitrary waveform generator (AWG) followed by an IQ-modulator. This allows the use of contemporary transceiver architectures and is also a requirement for joint synthesis of the discrete and continuous parts of the nonlinear spectrum, if these are to be jointly exploited [9].

To modulate a single soliton, four degrees of freedom can be taken into account: (i) The shape of the soliton itself, i.e., its peak power and its width can be varied, (ii) its center frequency can be modulated, (iii) the timing of the soliton pulse can be dithered, and (iv) its phase can be varied. In the context of the NFT, these degrees of freedom are respectively described by (i) the imaginary part of the eigenvalue λ_k, (ii) the real part of λ_k, (iii) the amplitude of the nonlinear coefficient $b(\lambda_k)$ and (iv) the phase of $b(\lambda_k)$, see Appendix I for details. Since $b(\lambda_k)$ serves in lieu of a complex-valued signal amplitude, modulation schemes that vary both the timing and the phase of the soliton pulse are referred to as amplitude phase shift keying (APSK) in the framework of the NFT [19], [20].

By using the inverse NFT (INFT), higher-order solitons consisting of multiple complex-modulated discrete eigenvalues can be created to increase the data rate. In 2015, data transport was shown by using solitonic waveforms repeating at 0.5 GBd with three co-propagating eigenvalues each modulated with on-off keying (OOK) [21]. This was followed shortly thereafter by work showing quadrature phase shift keying (QPSK) of two co-propagating eigenvalues, with a pulse repetition rate of 1 GBd [22], [23]. However, the utilization of higher-order solitons does not appear to have an obvious advantage in terms of spectral efficiency, since higher-order solitons also require more spectrum. For example, in [21] eigenvalues corresponding, stand-alone, to pulses with the (amplitude) shapes $\mathrm{sech}(t/T_0)$, $\mathrm{sech}(2t/T_0)$ and $\mathrm{sech}(3t/T_0)$ were multiplexed. Here, the longest pulse $\mathrm{sech}(t/T_0)$ limits the baud rate, while the shortest pulse $\mathrm{sech}(3t/T_0)$ limits the spectrum, that is three times wider than otherwise required. Moreover, this approach is also not scalable to an ever increasing number of eigenvalues per pulse, due to constraints placed on the computational power to perform the INFT as well as the growing hardware requirements in terms of bandwidth and effective number of bits (ENOB). Thus, dense interleaving of fundamental soliton pulses, in both the time and frequency domain, appears a more straightforwardly scalable solution and has been adopted here. However, it should be noted that interleaving higher-order solitons each generated by a corresponding modulator would make better use of the modulation speed afforded by contemporary modulators.

The focus later shifted to increasing the spectral efficiency by using higher-order modulation formats, with 16APSK and two additional bit eigenvalue modulation applied to a single eigenvalue in [19], resulting in a spectral efficiency estimated as 0.417 b/s/Hz at 4 GBd based on the reported system parameters for distances up to a 1000 km. 16APSK was then later applied to interleaved soliton pulses of alternating center frequency [20].

For these reasons, we have opted in this work for using QPSK and 8APSK modulation of densely packed fundamental soliton pulses, increasing the number of launched subchannels to four, the spectral efficiency per polarization to 0.5 b/s/Hz and the transmission distance to 5400 km. Since the soliton pulses are independently synthesized, the system architecture is straightforwardly parallelizable by duplication of the hardware, as opposed to joint processing with the INFT.

III. System Concept

First-order (fundamental) solitons that are partially overlapping in the time and in the frequency domain, but share a

common polarization, are QPSK or APSK modulated, optically merged and transmitted without dispersion management [24]–[26].

Mathematically, the launched wavefunction is described as having the amplitude

$$Q(t) = A \sum_{n=0}^{N-1} e^{i2\pi\Delta f \cdot t + i\varphi_n} \cdot \text{sech}\left(\frac{t}{T_0} - n\Delta t - t_n\right) \quad (1)$$

where A is the required amplitude to reach the soliton launch condition, N is the number of interleaved pulses, Δf the frequency spacing between channels, φ_n the applied data dependent phase, t the time, T_0 the characteristic time constant of the sech pulses, Δt the time increment between interleaved pulses, and, in case of APSK modulation, t_n the data dependent time increment applied to the pulse launch time. In the four channel configuration implemented here we follow the pulse sequencing described by (1). With increasing number of soliton channels, soliton pulses that are far apart in frequency have decreasing interaction and could be launched completely overlapping in time [27]. This would result in a more general description

$$Q(t) = A \sum_{n=0}^{N-1} e^{i2\pi\Delta f \cdot t + i\varphi_n} \cdot \text{sech}\left(\frac{t}{T_0} - mod\,(n, M)\,\Delta t - t_n\right) \quad (2)$$

with M corresponding to the number of channels that need to be sequenced in time before starting over.

This sequencing requires tight control on the differential frequency and timing of the solitons. Integrated platforms enable robust systems with precise pulse interleaving for optical time division multiplexing [28], [29] and have enabled control on pulse interleaving in the picosecond range [30]. Here, an integrated silicon photonics (SiP) chip has been developed that can modulate and multiplex together up to four first-order soliton channels with tight control over their time and frequency spacing. In addition to the transmitter (Tx) centric work, the effectiveness of different equalization methods applied to the received NFT spectrum at the receiver (Rx) is experimentally verified for the explored transmission scenarios. Artificial neural networks (ANNs) in particular show a clear increase in the maximum achievable distance.

After modulating the different eigenvalues, the conceptually most straightforward approach to create the combined soliton pulses is to directly synthesize the waveform with the INFT in the digital domain, as this ensures that nonlinear interaction between the pulses does not lead to a loss of the corresponding eigenvalues, i.e., that the soliton pulses are actually conserved. This is, however, limited in scalability due to the finite bandwidth of electro-optic transceivers. Therefore, a solution has to be found to stitch together lower bandwidth signals such as simpler first-order soliton pulses at the Tx, while at the same time guaranteeing that linear superposition in the optical domain does not lead to excessive distortion of the soliton pulses via nonlinear interaction as they propagate down the link. In extreme cases, soliton pulses even disappear via nonlinear pulse fusion or pulse destruction. Here, the NFT can be employed as a useful tool not for in-line processing of the data, but for establishing allowable launch conditions to guarantee conservation of each of the individual pulses [31].

This is exemplified in Fig. 1, that shows the pulse properties, in terms of extracted eigenvalues, obtained after linearly superposing two fundamental soliton pulses of different center frequencies, e.g., with a directional combiner-splitter (DCS) or a 2-by-1 multimode interferometer (MMI). The initial sech pulses are assumed to have a (power) full width at half maximum (FWHM) of 105 ps (corresponding to $\text{Im}\,(\lambda_k) = 0.5i$) as used in the following experiments, and to have zero temporal offset. The four panels correspond to different frequency offsets, while in each the phase difference between the pulses was swept. At a frequency offset of 16 GHz, which is large compared to the 10.1 GHz optical pulse bandwidth at -20 dB power, the eigenvalues can be seen to remain unperturbed after the linear superposition, which means that the soliton pulses do not suffer from significant distortion. This situation changes, however, as the frequency difference is reduced, as a consequence of the nonlinearity of the NFT. At 10.5 GHz, the real parts of the eigenvalues are being slightly perturbed by a phase dependent amount that leads to time-domain jitter after propagation, but are otherwise conserved. A more drastic modification of the eigenvalues is seen at 5 GHz frequency spacing, as here for some phases one eigenvalue is lost and the remaining one gains in power (imaginary part), with some of the power also converted into dispersive waves in the continuous spectrum. This shows that under these launch conditions one of the fundamental soliton pulses and the corresponding data are completely lost. Finally, at very small frequency differences one can also observe the generation of higher-order solitons with two co-propagating eigenvalues that share a common real part. However, this process is phase sensitive and difficult to exploit in case of phase modulation, as used in this work. Consequently, these graphs show that packing of these soliton pulses at 5 GHz or below, as implemented by linear superposition in the optical domain, will not work without irreversible loss of information. The exact choice of parameters, determining how densely pulses can be packed in the frequency and time domain, also has to take the time-sequencing of the pulses into account, as further analyzed below.

Fig. 2 shows the time and frequency domain overlaps between pulses that are to be interleaved in the optical domain for two scenarios corresponding to (a), (b) a 105 ps FWHM and (c), (d) a 210 ps FWHM. In both cases, the pulse spacing and frequency offset are 150 ps and 10 GHz. While in the first scenario both time- and frequency-domain overlaps are present, in the second the frequency-domain overlap is negligible at the cost of a much increased time-domain/spatial overlap. The baud rates and, for a given modulation scheme, the spectral efficiencies are the same in both scenarios, so that these correspond merely in a trade-off between the magnitudes of the time and frequency domain overlaps between soliton pulses.

For both scenarios, we have numerically analyzed how much further the frequency spacing can be reduced between the channels while maintaining the soliton pulses. In order to do this, the nonlinear spectrum resulting from linearly superposing two sech pulses with the targeted FWHM and time spacing was computed

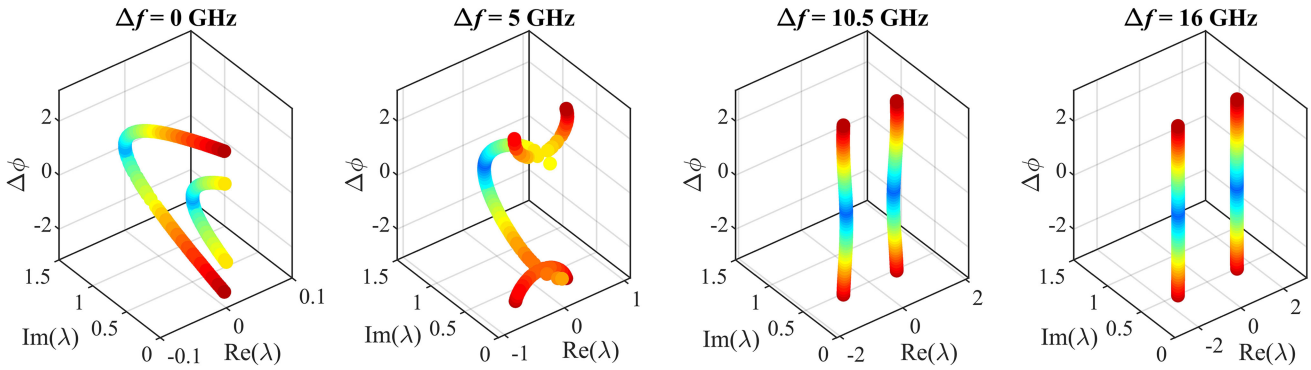

Fig. 1. Position of eigenvalues after linear multiplexing in the optical domain. The time offset between solitons was set to 0, the frequency difference is set according to the figure titles. The phase offset ($\Delta\phi$) between the initial soliton pulses was swept and is also indicated by the color coding.

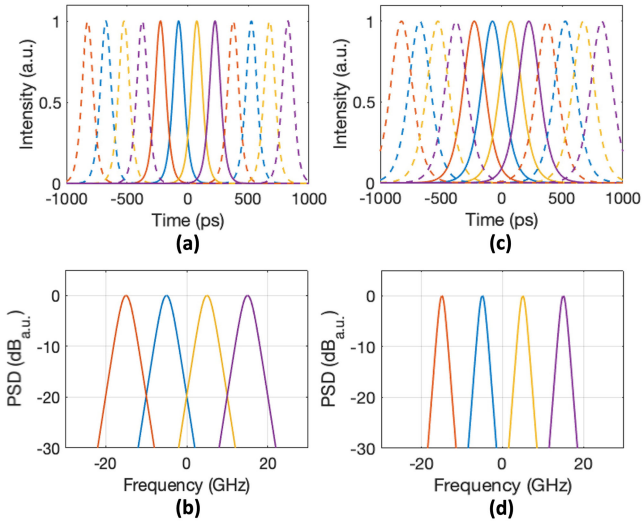

Fig. 2. Time domain intensity profiles and spectra of fundamental solitons prior to interleaving for two different scenarios. In all cases, pulses are spaced by 150 ps and carrier frequencies by 10 GHz. Pulses have a FWHM of (a), (b) 105 ps and (c), (d) 210 ps. In (a) and (b) only the four central pulses have been drawn with a solid line to facilitate the visualization, there is however no other significance to the adjacent pulses being represented as dashed curves.

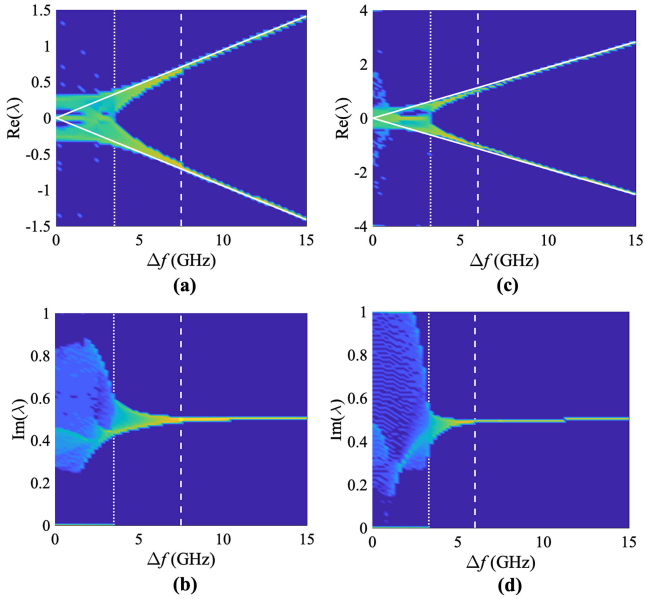

Fig. 3. Real and imaginary parts of the eigenvalues after linear superposition of two soliton pulses spaced by 150 ps as a function of the frequency offset. Traces corresponding to different relative phases have been overlaid, resulting in the smearing out of the overall diagram. In (a), (b), the FWHM of the pulses is assumed to be 105 ps, in (c), (d) 210 ps. Solid white lines in (a), (c) indicate the position of the eigenvalues before superposition. Dashed white lines indicate the minimum frequency offset below which significant smearing out of the traces is observed and are at (a), (b) 7.5 GHz and (c), (d) 6 GHz. The diagrams go through a bifurcation below 3.5 GHz and 3.3 GHz frequency spacing for the shorter and for the longer pulses, respectively, as shown by the dotted boundaries.

for both scenarios as a function of frequency difference and for all possible phase offsets. Results are shown in Fig. 3(a) and 3(b) for the first scenario (105 ps FWHM and 150 ps time separation), that depict the real and imaginary parts of the resulting eigenvalues as a function of the frequency offset. The data for the different phase offsets is overlaid, resulting in the observed smearing out of the traces below 7.5 GHz frequency offset and resulting in timing jitter at the Rx (boundary shown by dashed line). This becomes worse as the frequency difference is reduced until 3.5 GHz, at which point a bifurcation occurs in the diagram and only a single eigenvalue survives (boundary shown by dotted line). Therefore, this is the lower limit on channel spacing irrespective of how powerful equalization techniques can become. Fig 3(c) and 3(d) correspond to the second scenario, in which the time spacing was maintained at 150 ps, but the pulse width increased to 210 ps. It is apparent that even though the pulse overlap is now much bigger in the time domain, this results in a slightly better tolerance to dense pulse packing in the frequency domain as a result of the reduced spectral widths. The real part of the eigenvalues now stays insensitive to phase offsets down to about 6 GHz (dashed boundary) and the bifurcation occurs at a 3.3 GHz frequency offset (dotted boundary).

Due to the developed photonic integrated circuit (PIC) currently limiting frequency offsets between channels (see Section IV), experimental demonstration will be done using 150 ps pulse-to-pulse spacing with four channels spaced by 10 GHz from each other, as shown in Fig. 2 and corresponding to a spectral efficiency of 0.333 b/s/Hz for QPSK and 0.5 b/s/Hz for 8APSK. However, data from Fig. 3 suggests that a tighter

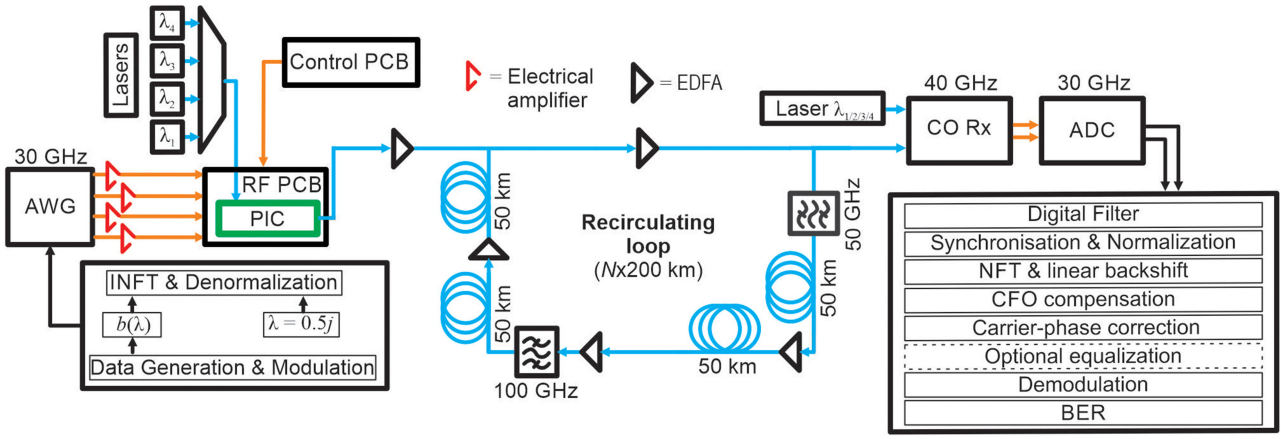

Fig. 4. Block diagram of the investigated link (electrical lines are color coded in orange and optical lines in blue).

frequency spacing should be possible, in particular for the longer pulses that are somewhat easier to merge (see above). These longer pulses also suffer less from lumped amplification inside the fiber channel due to the increased collision length [4] and are easier to demultiplex at the Rx as a consequence of the disjoint spectra. Consequently, even higher spectral efficiencies might be achievable in the future.

IV. Experimental Setup

We experimentally demonstrate merging of first-order solitons in the optical domain by using a PIC at the Tx. All the experiments have been performed using the long-haul transmission system with block diagram shown in Fig. 4. In the digital domain, 60 blocks of 2000 first-order solitons ($T_{\mathrm{FWHM}} = 105$ ps) are generated and QPSK or 8APSK modulated on the Tx side. On the Rx side, the detected signal is filtered after the analog-to-digital conversion using a 5th-order Butterworth filter with a 3-dB cutoff frequency of 6.6 GHz, isolating the signal corresponding to a single channel. Afterwards, the signal is synchronized, normalized into NFT units, its NFT is calculated and the fiber-induced linear phase rotation is de-rotated. In the next step, carrier-frequency-offset (CFO) compensation and phase recovery are applied. CFO in particular becomes trivial after applying the NFT, as it then consists in simply translating the real part of the eigenvalues. Before demodulation, it is possible to employ linear or nonlinear minimum mean square error (N)LMMSE [32], [33] or ANN [33] equalization to increase the maximum reach of the system. Finally, the received signal is demodulated and the bit error ratio (BER) calculated.

The I- and Q-parts of the first-order soliton pulses are converted to the analog domain by an arbitrary waveform generator (Keysight M8196A, 84 GSa/s for QPSK and 85 GSa/s for 8APSK, 30 GHz analog bandwidth) with four electrical channels, thus supporting up to two independent complex valued optical channels (larger channel counts are later emulated by time shifting waveforms, as described below). It creates pulses with an optical bandwidth (BW) of 10.1 GHz (99% power) on the Tx side. The electrical signal is then amplified by two pairs

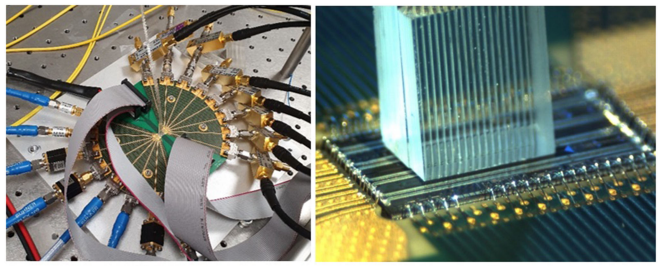

Fig. 5. (a) Photograph of the RF PCB. (b) Photograph of the system PIC attached to the fiber array.

of phase-matched RF amplifiers up to a drive voltage of 2 V_{pp}. The electrical signal is injected into the PIC via a custom-made multi-layer RF PCB, see Fig. 5(a). On the Rx side, the electrical signal is directly converted into the digital domain after coherent reception using a Keysight DSOZ334A oscilloscope (80 GSa/s), so that the signal can be post-processed as explained above.

In the optical domain, four discrete external cavity lasers (100 kHz linewidth) are multiplexed on the Tx side and injected into the PIC from a fiber array via a single grating coupler (GC), see Fig. 5(b), to provide the carriers needed to optically modulate the first-order solitons. Since the PIC receives the unmodulated light through a single input port, it is also compatible with operation with a comb source [34]. Its block diagram and characteristics are described in the following subsection. At the output of the PIC, the modulated and multiplexed first-order solitons are re-amplified with an EDFA to reach an average launch power that fulfills the soliton condition, after which they are launched into a fiber-optical recirculating loop consisting of four spans of 50 km True-Wave NZDSF (with dispersion $D = 4.5$ ps/(nm $\cdot$ km), losses $\alpha = 0.2$ dB/km and nonlinear parameter $\gamma = 1.6$ $\mathrm{W^{-1}km^{-1}}$) and EDFAs (Fig. 4). On the Rx side, the signal is amplified once more by another EDFA, bandpass filtered using a tunable optical filter with a 20 GHz 3-dB passband centered on the channels under test and its polarization adjusted. Finally, the signal is detected by a coherent

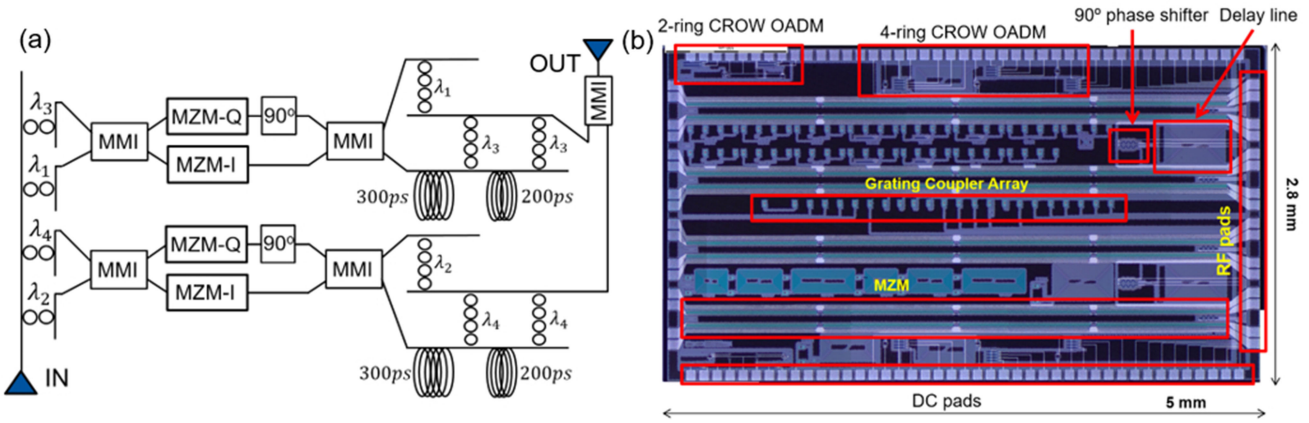

Fig. 6. (a) PIC block diagram. (b) Micrograph of the system PIC.

Rx (Neophotonics μICR-Class 40, BW = 40 GHz) using a local oscillator with a 1 kHz linewidth for the first series of experiments (250 ps pulse-spacing) and 10 kHz in the higher power mode for the rest of the experiments (150 ps pulse spacing). This configuration allows single channel reception with the maximum bit and sampling resolution afforded by the oscilloscope. While all the experiments reported in this paper have been done with single-channel reception, we have also made first experiments with joint multi-channel reception via a wideband receiver, with results described in [35].

A. System PIC and Transmitter Link Budget

The system PIC has been designed and fabricated in the standard silicon photonics process of Advanced Micro Foundry (AMF), Singapore. It implements electro-optic modulation and reconfigurable interleaving of soliton pulses with different time and frequency spacings. Fig. 6 shows a block diagram of the integrated system together with a micrograph of the fabricated chip. Inside the chip, four 2nd-order coupled (ring-) resonator optical waveguide (CROW) optical add-drop multiplexers (OADMs) send the different carriers to one of the two input ports of two IQ Mach-Zehnder modulators (MZMs).

This configuration allows to assess the performance of a four-channel system using only two different IQ signals and MZMs. Channels 3 and 4 carry, respectively, the same information as channels 1 and 2 [see wavelength labelling in Fig. 6(a)], but are delayed after modulation using 300 ps or 500 ps delay lines that can be selected by using one of two redundant CROW OADMs. These delays are respectively used for configurations in which pulses of adjacent channels are spaced by 150 ps and 250 ps and consequently result in the correct sequencing for channels 1 and 3/channels 2 and 4 (the sequencing between channels 1 and 2/channels 3 and 4 is programmed by the AWG). While the data between these channel pairs is not decorrelated in the launch windows, pulses propagate at different speeds, collide several times with pulses from other launch windows, and reshuffle before arriving at the Rx [36]. Moreover, since the four channels are supplied with carriers by four different solitary lasers, the relative phase between channels is also random and drifts over time.

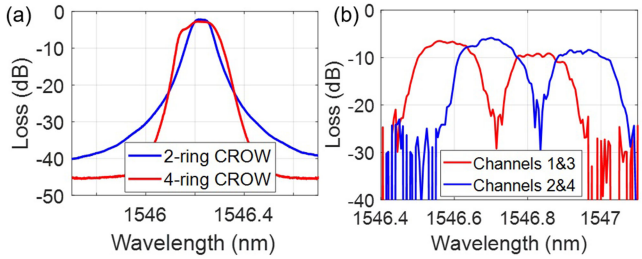

Fig. 7. (a) 2-ring and 4-ring CROW OADM wavelength response after alignment. (b) Reconfigurable output network response after alignment by control PCB.

Since spectra of adjacent channels are overlapping in the investigated launch conditions, regular WDM cannot be used to superimpose the fields. To address this, only even or odd channels are multiplexed with 4th-order CROW OADMs to common bus waveguides. Since channel indices are then two apart, spectra are non-overlapping and the OADM passbands can be chosen to couple the pulses over without crosstalk or spectral distortion. These two output bus waveguides, that respectively carry the even and odd channels, are finally combined into a single output waveguide using an MMI. This final superposition occurs without spectral filtering, as required to combine overlapping spectra without distortion, but comes at the cost of 3 dB additional insertion losses due to reciprocity. The penalty is however incurred only once, irrespective of the number of channels.

The critical elements of the system PIC are the 2nd- and 4th-order CROW OADMs and the IQ-MZMs, that have been independently characterized. The 2nd-order CROW OADMs can be shifted by more than one full free spectral range (FSR). They have an insertion loss (IL) of 2 dB and a 3-dB BW of 7.5 GHz after alignment of the rings, as shown in Fig. 7(a). Thermal phase tuners have a resistance of 1.6 kΩ and require 6 mA (9.6 V) to tune individual rings up to one FSR. The 4th-order CROW OADMs can also be tuned by more than one full FSR. They have 3 dB IL and a 3-dB BW of 15 GHz after alignment, Fig. 7(a). For these, thermal phase shifters have a resistance of 0.97 kΩ and require 5.5 mA (5.34 V) to tune individual rings by one FSR. As a result of the wide 3-dB bandwidth of the 4th order

CROW filters used to multiplex the solitons, 15 GHz, that is significantly larger than the −20 dB bandwidth (99% power) of the soliton pulses, and of the distortionless combination of the two bus waveguides by the 2-by-1 MMI, the output of the transmitter chip is very close to a simple linear superposition of the four fundamental soliton pulses.

The IQ-MZMs have been designed with slotted transmission lines, with phase recovery loops to phase match the RF and the optical signals, and with a GSSG driving scheme that suppresses RF cross-talk between complementary phase shifters operated in push-pull configuration [37], [38]. They have 4.4 mm long phase shifters, an electro-optic cutoff frequency of 13.8 GHz targeted for the present application, a $V_\pi L$ of 2.86 V·cm and below 4 dB IL. When operated with 2 V_{pp} signals and biased at minimum power for QPSK/8APSK modulation, they introduce an IL and modulation penalty of 12.3 dB, defined here as the attenuation of the peak power after modulation since nested modulators are biased at the zero-transmission point. It should be noted that it is not the cutoff frequency of the electro-optic modulators that limits the pulse width here, but rather the lumped amplification used in the link and the requirement to have sufficiently long soliton collision lengths. The V_π of the modulator and the overall ILs of the PIC on the other hand have a direct impact on the noise performance of the link [26] and have been prioritized during device optimization within the constraints of the available process in a multi-project wafer run.

We have derived the link budget of the Tx system in order to ensure the pulse powers required for the soliton condition can be met in the fiber [39]. The output power of the four discrete cavity lasers is set to 12 dBm for channels 3 and 4, and to 8 dBm for channels 1 and 2. This difference in power equalizes the extra IL losses introduced by the delay lines for channels 3 and 4 [Fig. 6(a)]. i) The external multiplexer introduces an IL of 7 dB; ii) the input and output GCs a cumulative IL of 7.5 dB; iii) 2nd and 4th-order CROW OADMs a cumulative IL of 5 dB; iv) the IQ-MZMs an insertion and modulation penalty of 12.3 dB for 2 V_{pp} drive voltage; v) delay lines an IL of 4 dB; vi) the output MMI an IL of 3 dB due to reciprocity (with negligible excess losses); and vii) interconnect waveguides and monitor taps a cumulative 3 dB IL. We obtain a soliton-pulse peak-power of −29.8 dBm at the PIC output. These values have been used to adjust the required amplification at the PIC output to reach the soliton peak power launch condition of 3.9 dBm calculated based on the path-average conditions.

B. Control System and High-Speed PCB

The overall temperature of the system PIC is controlled by a thermoelectric cooler (TEC) placed on the bottom of the RF PCB to avoid large temperature fluctuations. Once the PIC temperature is stable, the two IQ-MZMs, four 2nd-order CROW OADMs and six 4th-order CROW OADMs, corresponding to a total of 44 thermally tuned phase shifters, are controlled in real time by a custom-made control PCB. Each CROW OADM has a monitor PD connected to its drop port generating a feedback signal for closed-loop control, based on maximization of the dropped power [40]. An example of stabilized transfer functions,

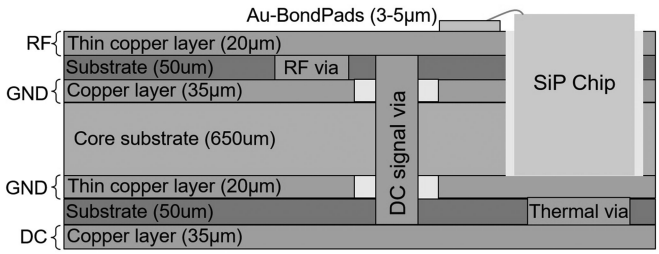

Fig. 8. Multilayer stack of the RF PCB.

as measured from the system PIC from taps located between the output of the IQ MZMs and the outputs of the reconfigurable networks, is shown in Fig. 7(b). As expected, channels 3 and 4 show larger IL than channels 1 and 2 due to the extra delay lines in the corresponding optical paths.

The RF PCB has to support the high density of control signals, the grounded co-planar waveguide (GCPW) RF transmission lines, and has to enable short wire bonds to the PIC to minimize RF insertion losses and back-reflections at the PCB to PIC interface. These requirements are supported by the multilayer stack shown in Fig. 8. A thin substrate is used on the top side between the RF lines and the underlying ground planes to shrink the lateral dimensions of the transmission lines and enable a compact RF pad frame. Symmetric bottom layers are used for the control signal fan-out. A thick substrate core is placed in the middle to enable a sufficiently deep cavity in which the chip can be inserted with a top surface nearly flush to the high-speed lanes, minimizing the required wire-bond lengths. A high-density array of metal vias below the PIC provides a thermal path to the underlying TEC, attached to the bottom side of the PCB. Multiple test structures have been fabricated to characterize the PCB performance, showing a return loss below −11 dB for all channels (including wire bonds, estimated by wire bonding two traces with each other in a test structure), IL lower than 3.5 dB at 20 GHz and an inter-channel crosstalk below −30 dB at 15 GHz when driving with fully differential signaling.

V. System Measurements

The implemented long-haul transmission link has been used to perform single polarization QPSK and 8APSK modulated 4-channel experiments with 105 ps FWHM pulses. In a first series of experiments, we defined the QPSK modulated pulses of adjacent channels to be separated by 250 ps and their center frequencies to be offset by 15 GHz. This corresponds to a very conservatively chosen set of experimental conditions in which soliton pulses are fully disjoint in time and frequency, and serves to validate the soliton pulse propagation as well as the equalization methods. We have started this initial experiment by sweeping the launch power of a single 105 FWHM pulse repeating at a 1 GBd rate, that was QPSK modulated with a 2 V_{pp} signal and transmitted over 3000 km, with results shown in Fig. 9. Based on the theoretical optimum soliton peak power of 3.9 dBm, taking the path-average conditions into account, we expect an optimum average power of −5.3 dBm

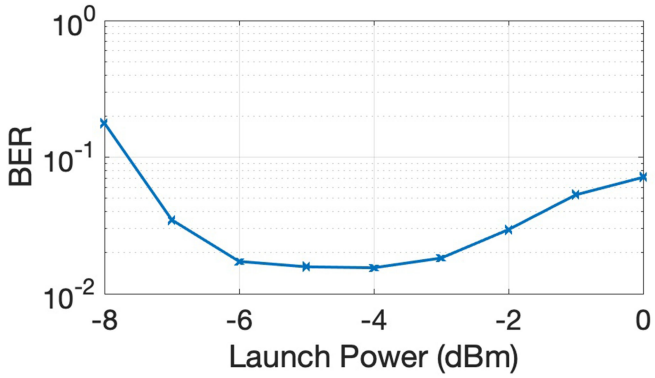

Fig. 9. BER as a function of average launch power for a single 105 ps FWHM pulse train that is QPSK modulated with a 1 GBd repetition rate, after a transmission distance of 3000 km.

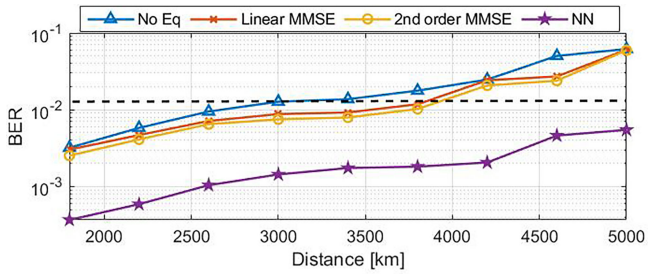

Fig. 10. BER for QPSK 4-channel transmission as a function of distance with and without equalization. The channels are separated by 15 GHz and the pulses sequenced with a 250 ps time interval, resulting in a 1 GBd overall symbol repetition rate.

consistent with the measurements. We then continued with four channel measurements, with a total 4-pulse launch power of 0.7 dBm in line with the optimum determined from the previous measurement.

Fig. 10 shows the BER as a function of distance before and after equalization. Several equalization algorithms have been implemented to post-process the received signal, including LMMSE, 2nd order NLMMSE [32], [33] and ANN that are applied to data preprocessed by the NFT [33], see Appendix II.

While equalization techniques could also have been applied to sampled pulses without the intermediate NFT step, we chose to follow this path in anticipation of future work on more complex modulation formats also involving the continuous spectrum. The ANN equalization in particular significantly improves the reach, that is increased from 3000 km to well above 5000 km. It uses two hidden layers with 132 and 569 nodes, respectively, and results in a BER of 6×10^{-3} at 5000 km. It should be highlighted that this performance is maintained even though solitons have undergone multiple collisions at this point [20], [36]. However, the corresponding spectral efficiency of 0.133 b/Hz/s per polarization needs to be further improved.

Fig. 11 shows the recorded eye diagrams and the recovered constellation diagrams at 3000 km, 4600 km and 5400 km. Eye diagrams are overlaid with sech shaped pulses with a FWHM of 105 ps ($T_0 = 60$ ps). It can be seen that the pulse shape is conserved, confirming solitonic pulse propagation.

In a second set of experiments, we reduced the pulse-to-pulse spacing to 150 ps and the channel-to-channel frequency offset to 10 GHz, increasing the baud rate to 1.66 GBd and the spectral efficiency to 0.333 b/Hz/s per polarization. This corresponds to the scenario illustrated in Fig. 2(a) and 2(b) and to a situation in which pulses overlap substantially both in time and frequency. The optical spectrum recorded right after the Tx is shown in Fig. 12 and an initial power sweep in Fig. 13. Here, the power sweep was directly done with pulses of all four channels launched, in order to take interaction between the densely packed channels into account. Because of the higher repetition rate, we expected to find an optimum at 2.9 dBm. The experimentally determined optimum average launch power is however about 6 dB lower.

This can be understood by inspecting the eye diagrams at this reduced launch power, as shown in Fig. 14. Past ∼2000 km, the pulses take a FWHM of about 230 ps, corresponding to the expected soliton width for the utilized optimum launch power, that is then conserved thereafter. We thus conclude that solitonic transmission is maintained, that a better BER is, however, obtained with wider pulses that reduce the frequency overlap between the channels, at the cost of a higher temporal and spatial overlap, as shown in Fig. 2(c) and 2(d). This is not entirely surprising given the calculations shown in Fig. 3, that suggest that it is somewhat easier to merge the longer pulses, while they also suffer less from lumped amplification and are easier to demultiplex at the Rx. The increased collision length inside the fiber channel might in particular be an important factor in the observed performance improvement [4].

Fig. 15 shows the BER as a function of transmission distance, with and without equalization, using the optimum launch condition. It can be seen that even without equalization the BER stays below the targeted soft decision (SD) forward error correction (FEC) floor up to transmission distances of 4500 km. Surprisingly, this is even better than what was obtained in the more conservative experiments with 250 ps and 15 GHz pulse and channel spacing. An improvement to the phase tracking algorithm applied between the two experiments allowed increasing the local oscillator power by 3 dB, with additional phase noise corrected by the algorithm, improving the link's power budget. Here too, ANN based equalization improves the BER very substantially and allows for transmission distances above 5500 km.

In this second set of experiments, the 10 GHz channel spacing was limited by the bandwidth of the 2nd order CROW OADMs used to route the individual carriers to the modulator inputs in the PIC [Fig. 6(a) and 7(a)]. At 10 GHz, the power crosstalk coupled over from adjacent carriers is already in the order of −12 dB, leading to very large signal crosstalk given the coherent nature of the communication scheme (the coherent signal, proportional to the E-field, is rejected by less than 6 dB). Consequently, we could not perform experiments at smaller channel spacings that would require a redesign of the 2nd order CROWs. Based on the analysis performed in Section III, we expect, however, smaller channel spacings to be compatible with this transmission scenario as shown in Fig. 3.

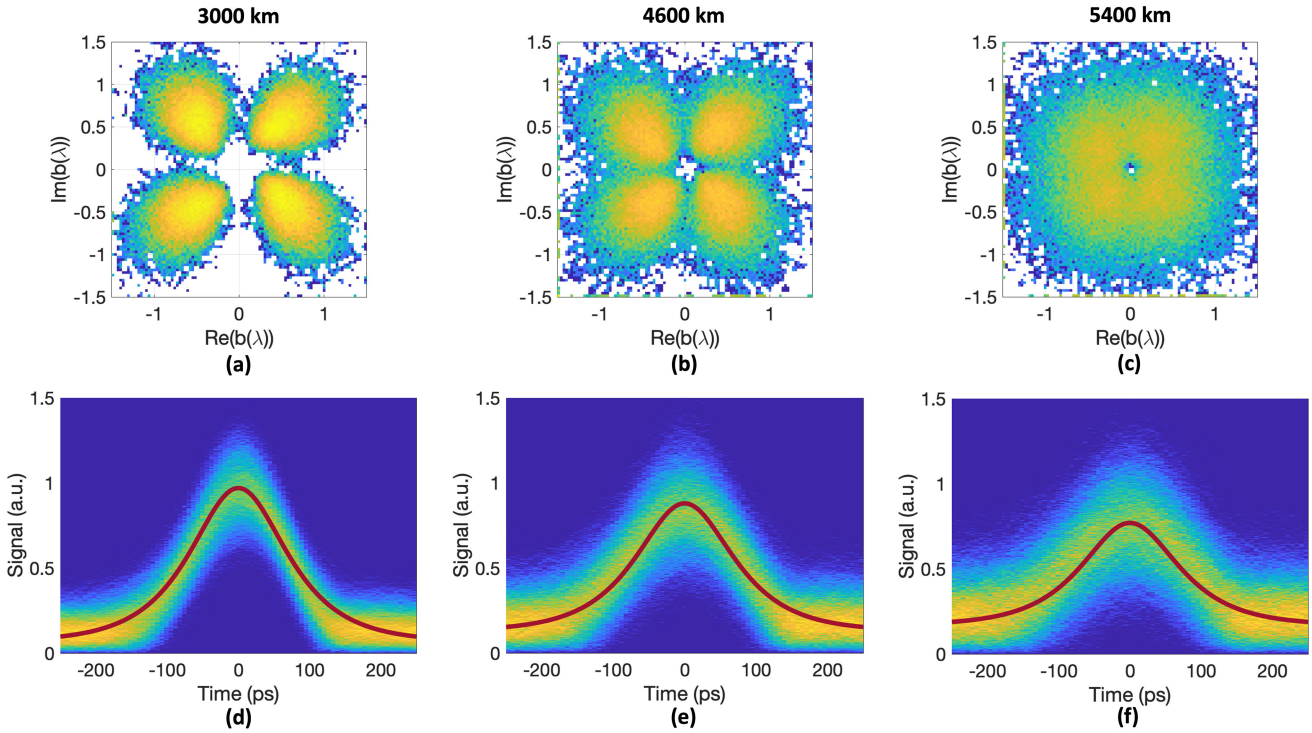

Fig. 11. (a)–(c) Constellation diagrams and (d)–(f) eye diagrams at different propagation distances of 3000 km, 4600 km and 5400 km, for the four-channel transmission experiments with 15 GHz channel and 250 ps pulse spacing (1 GBd). The eye diagrams are overlaid with a sech shaped pulse with a power FWHM of 105 ps and a T_0 of 60 ps, showing that the initial pulse shape has been conserved.

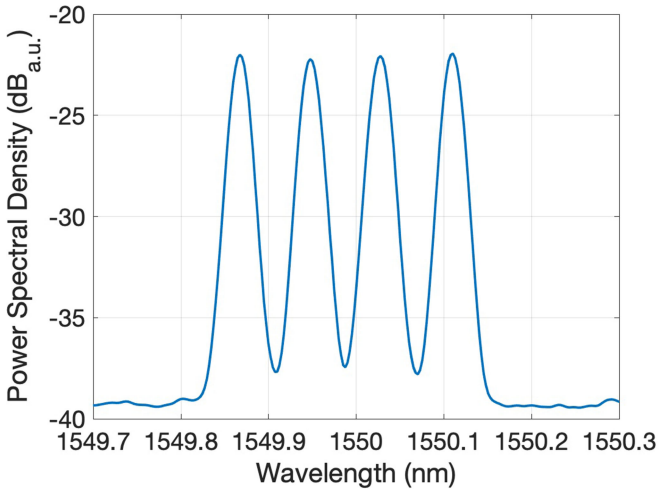

Fig. 12. Optical spectrum recorded right after the Tx when launching four channels with a 10 GHz frequency spacing and a 150 ps pulse spacing. The FWHM of the pulses is kept at 105 ps.

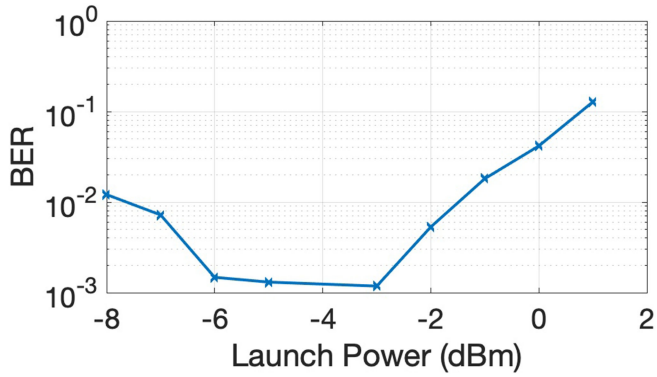

Fig. 13. BER as a function of average launch power for the four channel transmission experiment with 105 ps FWHM QPSK modulated pulses spaced by 150 ps and 10 GHz and repeating with a 1.66 GBd rate, after a transmission distance of 3000 km.

In order to further improve the spectral efficiency, a final set of experiments has been performed keeping the pulse-to-pulse spacing of 150 ps and the channel-to-channel frequency offset of 10 GHz from the second set of experiments (1.66 GBd), but increasing the modulation order from QPSK (2 bits) to 8APSK (3 bits), thus achieving a spectral efficiency of 0.5 b/Hz/s per polarization. In doing so, the phase of b(λ) is still QPSK modulated and a 2-ASK modulation is applied to the amplitude of b(λ). Since the latter determines the timing of the soliton pulses, the 2-ASK modulation can be simply implemented by time shifting the pulses within the time window. As we can see in Fig. 16, the selected time shift has an important impact on the results and the best performance is achieved for a time shift of ±90 ps. This optimal value has been found experimentally and comes from the trade-off between maximizing ASK symbol distance and minimizing the interaction of the solitons in the time domain. Fig. 17 shows the BER as a function of transmission distance, with and without equalization. The BER of the non-equalized

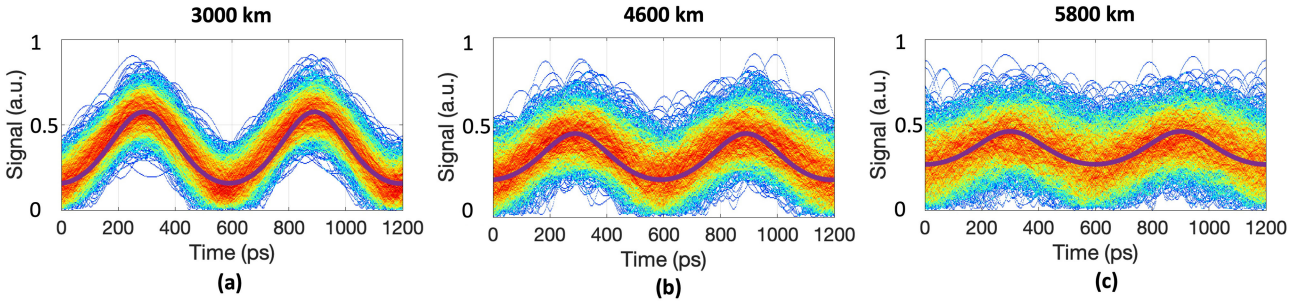

Fig. 14. Eye diagrams recorded after different transmission distances for the denser four-channel QPSK transmission experiments (10 GHz and 150 ps spacing). They are overlaid with repeating sech pulses spaced by 150 ps and with a FWHM of 227 ps ($T_0 = 130$ ps).

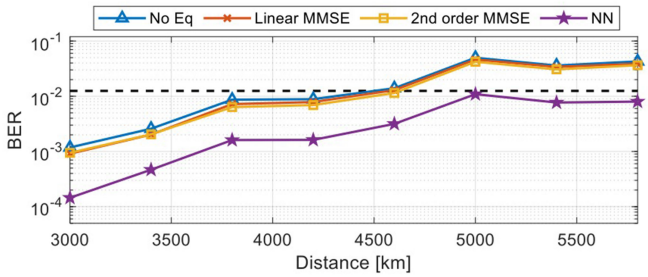

Fig. 15. BER for QPSK 4-channel transmission as a function of distance, with and without equalization. The channels are separated by 10 GHz and the pulses sequenced with a 150 ps time interval (1.66 GBd overall repetition rate).

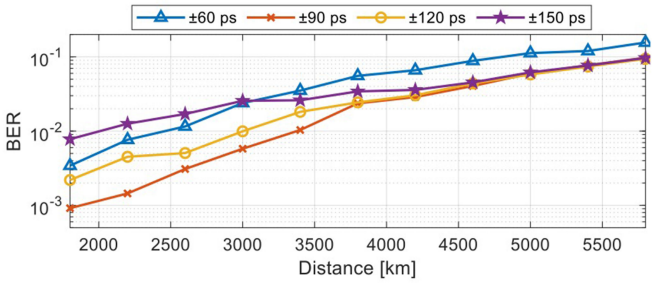

Fig. 16. BER for 8APSK 4-channel transmission as a function of distance for different time shifts without equalization, with channels separated by 10 GHz and the pulses sequenced with a 150 ps time interval.

link stays below the targeted SD-FEC limit up to transmission distances of 3400 km, lower than the result achieved in the previous experiment. However, after (N)LMMSE and NN equalization the transmission distance is increased up to 4600 km and 5400 km, respectively. Therefore, this last experiment has increased the maximum data rate up to 20 Gb/s and the spectral efficiency up to 0.5 b/Hz/s per polarization without incurring significant distance penalties after equalization. To the best of our knowledge, this spectral efficiency is a record for single polarization soliton transmission over such high distances. Moreover, the spectral efficiency could be further improved after redesigning the PIC with a lower BW in the 2nd order CROW OADMs, so channels can be more densely packed as numerically shown in Fig. 3. However, the spectral efficiency still remains modest compared to the capabilities of state-of-the-art

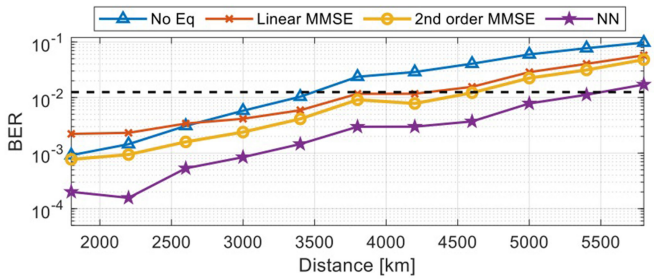

Fig. 17. BER for 8APSK 4-channel transmission as a function of distance, with and without equalization. Data rate and spectral efficiency of this experiment are 20 Gb/s and 0.5 b/Hz/s, respectively.

coherent communications, about 4 b/s/Hz per polarization at the ∼5000 km distances relevant for the work reported in this paper [41]. Consequently, effective and scalable joint exploitation of both the continuous and discrete spectra is a critical prerequisite to make soliton transmission competitive and remains a key issue to further investigate.

VI. Conclusion

In this paper, we have proposed an architecture for a scalable soliton transmitter with a record high spectral efficiency of 0.5 b/s/Hz per polarization at distances above 5000 km that is expected to be straightforwardly improved by further reducing the channel spacing after redesigning the PIC. Artificial-neural-network-based equalization has also been confirmed to be particularly effective in improving the performance of this link. Scalable joint exploitation of the continuous and discrete spectra is the next challenge that will have to be tackled with this architecture.

Appendix I
Nonlinear Fourier Transform

The NFT is a means to solve integrable, nonlinear partial differential equations [5], [42]. The relevant equation that fits this description for fiber transmission is the nonlinear Schrödinger equation (NLSE), which describes the evolution of an optical pulse amplitude $Q(\tau, l)$ through an ideal single-mode fiber

without loss nor noise

$$\frac{\mathrm{d}Q(\tau,l)}{\mathrm{d}l} = \frac{j\beta_2}{2}\frac{\mathrm{d}^2Q(\tau,l)}{\mathrm{d}\tau^2} - j\gamma Q(\tau,l)\left|Q(\tau,l)\right|^2 \tag{3}$$

with l being the travelling distance and τ the time in the reference-frame travelling forward at the group velocity v_g at the center frequency (obtained from time T as $T - l/v_g$). Here, β_2 is the chromatic dispersion coefficient and γ the nonlinearity parameter. In a real system using lumped amplification, the power of the pulses periodically varies as they transit through the fiber link, so that corrections to (3) such as path-average conditions have to be considered [4].

Since the NFT is a nonlinear transformation, it is convenient to work in a normalized system of units with quantities given by

$$q = \frac{Q}{A},\ z = \frac{l}{L},\ t = \frac{\tau}{T_0},$$

with $L = 2T_0^2/|\beta_2|, A = \sqrt{|\beta_2|/(\gamma T_0^2)}$ and T_0 being a freely chosen parameter. In this work, T_0 is set to 60 ps and corresponds to the characteristic time constant of the sech^2 pulses synthesized as fundamental solitons, with a (power) FWHM of $1.763 \cdot T_0 = 105$ ps. In combination with anomalous dispersion ($\beta_2 < 0$), this leads to the normalized NLSE

$$jq_z(t,z) = q_{tt} + 2|q(t,z)|^2 q(t,z). \tag{4}$$

The NFT is based on the Zakharov-Shabat scattering problem derived from the normalized amplitude $q(t,z)$ given by [5], [43]

$$\frac{\mathrm{d}}{\mathrm{d}t}v = Pv, \tag{5}$$

where

$$v = \begin{pmatrix} v_1(\lambda;t,z) \\ v_2(\lambda;t,z) \end{pmatrix},\ P = \begin{pmatrix} -\mathrm{j}\lambda & q(t,z) \\ -q^*(t,z) & \mathrm{j}\lambda \end{pmatrix}.$$

Here, $*$ indicates the complex conjugation and v is an eigensolution of the scattering problem. The nonlinear spectral parameter λ is analogous to frequency in the linear Fourier transform (FT), since, in the limit of low optical power levels, it reduces to $\lambda \simeq -\omega/2$, with ω the angular frequency from the linear Fourier transform. If q fulfills the vanishing boundary condition $q(t,z) \to 0, t \to \pm\infty$, then considering (5) at $t \to -\infty$ leads to a solution

$$v \to \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-\mathrm{j}\lambda t}, t \to -\infty. \tag{6}$$

By propagating v using (5) and starting at (6), we obtain the so-called Jost- or nonlinear Fourier coefficients (NFC)

$$a(\lambda) = \lim_{t\to\infty} v_1 e^{\mathrm{j}\lambda t} \tag{7a}$$

$$b(\lambda) = \lim_{t\to\infty} v_2 e^{-\mathrm{j}\lambda t}. \tag{7b}$$

During propagation through an ideal fiber, the NFCs are simply changing according to

$$a(\lambda; z_0 + z) = a(\lambda, z_0) = a(\lambda) \tag{8a}$$

$$b(\lambda; z_0 + z) = b(\lambda; z_0)\, e^{-4\mathrm{j}\lambda^2 z}, \tag{8b}$$

where z_0 is the position of the Tx [5].

Besides the linear transfer function of the NFT spectrum, one other major difference to the FT is the existence of two spectra, namely the continuous spectrum

$$\hat{q}(\lambda) = \frac{b(\lambda)}{a(\lambda)},\ \lambda \in \mathbb{R}, \tag{9}$$

and the discrete spectrum emerging at high optical power levels

$$\tilde{q}(\lambda) = \frac{b(\lambda_k)}{da(\lambda_k)/d\lambda|_{\lambda=\lambda_k}},\ a(\lambda_k) = 0,\ \lambda_k \in \mathbb{C}^+, k \in \mathbb{N}. \tag{10}$$

for which λ_k is also an eigenvalue of the associated Lax operator [5].

The continuous spectrum is comparable (especially for low signal powers) to the spectrum of the FT and dispersive in the time domain. The discrete spectrum consists of a finite number of discretely scattered points in the upper half of the complex plane (with positive imaginary values). Each of these points would correspond, on their own, to a fundamental soliton in the time domain, which does not broaden during transmission. Together, they form the nonlinear spectrum of higher-order solitonic pulses.

Appendix II
Equalization Techniques

The NFT is defined for the idealized NLSE without noise and loss. This leads to non-trivial distortions of the spectra, if noise [44] or lumped amplification [45] are introduced. Additionally, if $\tilde{q}(\lambda_k)$ or $b(\lambda_k)$ are used for information transmission, error propagation of eigenvalue deviations in (7), (8) and (10) impacts the signal quality. This can be physically explained by changes of the center frequency of the soliton (related to the real part of λ_k) leading to faster or slower transmission speeds due to dispersion, which lead to a temporal shift inside the traveling reference-frame and in turn to a deviation of the amplitude of b. Likewise, changes of the imaginary value of λ_k lead to higher or lower peak amplitudes and hence changes in the Kerr-effect induced phase shift. Due to cross-phase modulation, changes of eigenvalues in a higher-order soliton or merged solitons can have an effect on NFCs associated to other eigenvalues. If the expected transmitted λ_k are known at the Rx, e.g., if they are not themselves used for modulation, deviations can be used to equalize the received NFCs [44], [46]. Exemplarily, the correlation between deviations of the modulated $b(\lambda_k)$ and the eigenvalue λ_k are depicted in Fig. 18. For this plot, the transmission of 10000 first-order solitons over 60 spans of 50 km NZDSF using lumped amplification with an EDFA noise-figure of 5 dB was simulated. Additionally, if only $b(\lambda_k)$ is modulated, $a'(\lambda_k)$ can also be fixed and used for equalization purposes. This has been shown to be advantageous for eigenvalue communications [32] and is the modulation scheme that will be pursued in the following.

To equalize the amplitude (or the phase) of $b(\lambda_k)$, an error vector $\mathbf{n}_\mathrm{L} = [\Delta a'_R, \Delta a'_I, \Delta\lambda_R, \Delta\lambda_I]$ consisting of real and imaginary parts of the known deviations can be used to describe the LMMSE criterion according to [32]

$$\mathrm{argmin}_\mathbf{c} \mathrm{E}\left[\left(\Delta|b| - \mathbf{c}^T\mathbf{n}_\mathrm{L}\right)^2\right]. \tag{11}$$

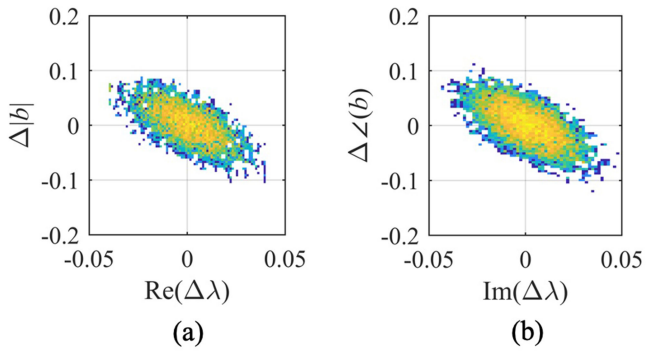

Fig. 18. Scatter plots of (a) deviation of the amplitude of b versus the deviation of the real part of λ and (b) deviation of the phase of b versus the deviation of the imaginary part of λ.

Here, E denotes the expectation value and $\mathbf{c}$ can be estimated using training symbols as

$$\mathbf{c}^T = \mathrm{E}\left[\Delta\,|b|\,\mathbf{n}_{\mathrm{L}}^T\right] \cdot \mathrm{cov}(\mathbf{n}_{\mathrm{L}})^{-1}. \tag{12}$$

To expand this LMMSE nonlinearly, the error vector $\mathbf{n} = [|b(\lambda_k)|,\ \Delta a'_R,\ \Delta a'_I,\ \Delta\lambda_R,\ \Delta\lambda_I]$ can be expanded [32] as

$$\begin{aligned}\mathbf{n}_{\mathrm{NL}_r} = [&\mathbf{n},\\ &\underbrace{\mathbf{n}\cdot\mathbf{n}^{(1)},\ \mathbf{n}\cdot\mathbf{n}^{(2)},\ \mathbf{n}\cdot\mathbf{n}^{(3)},\ \mathbf{n}\cdot\mathbf{n}^{(4)},\ \mathbf{n}\cdot\mathbf{n}^{(5)}}_{\mathbf{n}_{\mathrm{NL}_2}},\\ &\underbrace{\mathbf{n}_{\mathrm{NL}_2}\cdot\mathbf{n}^{(1)},\ \ldots,\ \mathbf{n}_{\mathrm{NL}_2}\cdot\mathbf{n}^{(5)}}_{\mathbf{n}_{\mathrm{NL}_3}},\ldots,\\ &\mathbf{n}_{\mathrm{NL}_{r-1}}\cdot\mathbf{n}^{(1)},\ \ldots,\ \mathbf{n}_{\mathrm{NL}_{r-1}}\cdot\mathbf{n}^{(5)}],\end{aligned}$$

where r stands for the order of nonlinearity. Again, (12) can be used to calculate the coefficient vector $\mathbf{c}$. While the number of coefficients for the linear equalizer is four times the number of considered eigenvalues, it is given by the following expression for the nonlinear equalizer after removing redundant factors

$$N_{\mathrm{coeffs}}\ (r) = \sum_{m=1}^{r} \frac{1}{m!} \prod_{n\,=\,0}^{m-1} (1 + 4\cdot K + n)\,, \tag{13}$$

with K being the number of eigenvalues considered by the equalizer. Since K is generally taken to be small, the total amount of coefficients can be kept small when compared to conventional nonlinear (e.g., Volterra) filters employed in dispersive communication systems.

A further step to refine the equalization is the application of ANNs [33]. ANN receivers in the framework of the NFT can be used in the time domain to fully replace the NFT at the Rx [47], or as equalizers that work with the eigenvalues and NFCs computed by the NFT. Here, the latter is used analogous to the aforementioned (N)LMMSE equalizers. An ANN consists of multiple layers. The input layer consists of a fixed set of nodes, defined by the ANN input $\mathbf{n}_{\mathrm{ANN}} = [\log(|b(\lambda_k)|),\ \angle(b(\lambda_k)),\ \Delta a'_R,\ \Delta a'_I,\ \Delta\lambda_R,\ \Delta\lambda_I]$. The ANN is able to equalize the amplitude and the phase jointly and can also include NFCs and deviations of multiple eigenvalues. The amount and size of the hidden layers is not predefined and has to be selected according to the specifics of the problem. In contrast to conventional equalizers, this is not always straightforward and has to be adapted to the use-case. Each node of the hidden layers consists of an activation function whose argument is the weighted sum of the outputs of the preceding layer. Again, the activation function can be freely chosen and has to be adapted. The weights of the input sum are the result of the training process. The activation function used in this work for the input and inner layers is the rectified linear unit (ReLU), which can be mathematically described by $f\,(x) = \max(0, x)$. The output layer size is again fixed to the use-case and type of output. In this work, instead of streaming out the symbols like the equalizers, the ANN is trained on the index of the transmitted symbol, which is the decimal representation of the bits. The activation function of the output layer is the softmax function given by

$$g(x_j) = \frac{\exp(x_j)}{\sum_{m=1}^{M}\exp(x_m)}\,. \tag{14}$$

Here, M is the amount of output nodes and j is the index of the output. The resulting values are between 0 and 1, where the highest output identifies the predicted symbol. The value of the highest output also represents the certainty of the ANN. This way, an SD-FEC can follow.

To train the ANNs, a gradient descent algorithm optimized by an Adam algorithm has been used. The training goal was to minimize the cross-entropy loss between the desired and predicted outputs.

To estimate the complexity of ANNs for online equalization, the complexity of the ReLU can be expressed by one multiplier and one comparator. Hence, the total amount of multiplications required for online equalization is

$$N_{\mathrm{mult}} = \sum_{l\,=\,1}^{L-1} k_l k_{l+1} + \sum_{l\,=\,2}^{L-1} k_l, \tag{15}$$

where L is the number of layers and k_l is the number of nodes in layer l.

REFERENCES

[1] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, Feb. 2010.

[2] A. Hasegawa and F. Tappert, "Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers," *Appl. Phys. Lett.*, vol. 23, pp. 142–144, Aug. 1973.

[3] L. F. Mollenauer, J. P. Gordon, and M. N. Islam, "Soliton propagation in long fibers with periodically compensated loss," *IEEE J. Quantam Electron.*, vol. 22, no. 1, pp. 157–173, Jan. 1986.

[4] L. F. Mollenauer, S. G. Evangelides, and J. P. Gordon, "Wavelength division multiplexing with solitons in ultra-long distance transmission using lumped amplifiers," *J. Lightw. Technol.*, vol. 9, no. 3, pp. 362–367, Mar. 1991.

[5] M. Yousefi and F. R. Kschischang, "Information transmission using the nonlinear fourier transform, Part I–III," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4312–4369, Jul. 2014.

[6] J. E. Prilepsky, S. A. Derevyanko, and S. K. Turitsyn, "Nonlinear spectral management: Liearization of the lossless fiber channel," *Opt. Exp.*, vol. 21, no. 20, pp. 24344–24367, Oct. 2013.

[7] S. T. Le, J. E. Prilepsky, and S. K. Turitsyn, "Nonlinear inverse synthesis for high spectral efficiency transmission in optical fibers," *Opt. Exp.*, vol. 22, no. 22, pp. 26720–26741, Oct. 2014.

[8] S. T. Le, V. Aref, and H. Buelow, "Nonlinear signal multiplexing for communication beyond the Kerr nonlinearity limit," *Nature Photon.*, vol. 11, no. 9, pp. 570–576, 2017.

[9] V. Aref, S. T. Le, and H. Buelow, "Demonstration of fully nonlinear spectrum modulated system in the highly nonlinear optical transmission regime," in *Proc. Eur. Conf. Opt. Commun.*, 2016, pp. 19–21.

[10] V. Aref, S. T. Le, and H. Buelow, "Modulation over nonlinear fourier spectrum: Continuous and discrete spectrum," *J. Lightw. Technol.*, vol. 36, no. 6, pp. 1289–1295, Mar. 2018.

[11] T. Gui *et al.*, "Nonlinear frequency division multiplexing with b-modulation: Shifting the energy barrier," *Opt. Exp.*, vol. 26, no. 21, pp. 27978–27990, 2018.

[12] L. F. Mollenauer, P. V. Mamyshev, and M. J. Neubelt, "Demonstration of soliton WDM transmission at 6 and 7×10Gbit/s, error free over transoceanic distances," *Electron. Lett.*, vol. 32, no. 5, pp. 471–473, 1996.

[13] D. L. Guen *et al.*, "Narrow band 1.02 Tbit/s (51×20 Gbit/s) soliton DWDM transmission over 1000 km of standard fiber with 100 km amplifier span," in *Proc. Opt. Fiber Commun. Conf.*, 1999, Paper PD4.

[14] M. Zitelli, F. Favre, D. L. Guen, and S. D. Burgo, "Numerical and experimental investigation of power and wavelength margins for a 20-Gb/s dispersion-managed soliton transmission system on standard fiber," *Photon. Technol. Lett.*, vol. 11, no. 7, pp. 904–906, Jul. 1999.

[15] L. F. Mollenauer, P. V. Mamyshev, J. Gripp, M. J. Neubelt, and N. Mamyscheva, "Demonstration of massive wavelength-division multiplexing over transoceanic distances by use of dispersion-managed solitons," *Opt. Lett.*, vol. 25, no. 10, pp. 704–706, 2000.

[16] M. Nakazawa, H. Kubota, K. Suzuki, E. Yamada, and A. Sahara, "Recent progress in soliton transmission technology," *Chaos*, vol. 10, no. 3, pp. 486–514, 2000.

[17] S. K. Turitsyn, B. G. Bale, and M. P. Fedoruk, "Dispersion-managed solitons in fibre systems and lasers," *Phys. Rep.*, vol. 521, pp. 135–203, 2012.

[18] J. P. Gordon and H. A. Haus, "Random walk of coherently amplified solitons in optical fiber communications," *Opt. Lett.*, vol. 11, pp. 665–667, 1986.

[19] T. Gui, C. Lu, A. P. T. Lau, and P. K. A. Wai, "High-order modulation on a single discrete eigenvalue for optical communications based on nonlinear Fourier transform," *Opt. Exp.*, vol. 25, no. 17, pp. 20286–20297, Aug. 2017.

[20] G. Zhou, T. Gui, C. Lu, A. P. T. Lau, and P. K. A. Wai, "Improving soliton transmission systems through soliton interactions," *J. Lightw. Technol.*, vol. 38, no. 14, pp. 3563–3572, Jul. 2019.

[21] Z. Dong *et al.*, "Nonlinear frequency division multiplexed transmissions based on NFT," *Photon. Technol. Lett.*, vol. 27, no. 15, pp. 1621–1623, Aug. 2015.

[22] V. Aref, H. Bülow, K. Schuh, and W. Idler, "Experimental demonstration of nonlinear frequency division multiplexed transmission," in *Proc. Eur. Conf. Opt. Commun.*, 2015, pp. 1–3.

[23] V. Aref and H. Bülow, "Design of 2-soliton spectral phase modulated pulses over lumped amplified link," in *Proc. Eur. Conf. Opt. Commun.*, 2016, pp. 409–411.

[24] J. Koch, S. Li, and S. Pachnicke, "Transmission of higher order solitons created by optical multiplexing," *J. Lightw. Technol.,* vol. 37, no. 3, pp. 933–941, Feb. 2019.

[25] J. Koch *et al.*, "Experimental demonstration of a silicon-photonics WDM NFT soliton transmitter," in *Proc. Opt. Netw. Commun. Conf.*, Jun. 2021, Paper W6A.34.

[26] A. Moscoso-Mártir *et al.*, "Silicon photonics integrated circuits for nonlinear Fourier transform based transmission," in *Proc. 47th Eur. Conf. Opt. Commun.*, Sep. 2021, pp. 1–4.

[27] A. F. Benner, J. R. Sauer, and M. J. Ablowitz, "Interaction effects on wavelength-multiplexed soliton data packets," *J. Opt. Soc. Amer. B*, vol. 10, pp. 2331–2340, Dec. 1993.

[28] A. A. Aboketaf, A. W. Elshaari, and S. F. Preble, "Optical time division multiplexer on silicon chip," *Opt. Exp.*, vol. 18, no. 13, pp. 13529–13535, Jun. 2010.

[29] J. Verbist *et al.*, "4:1 Silicon photonic serializer for data center interconnects demonstrating 104 Gbaud OOK and PAM4 transmission," *J. Lightw. Technol.*, vol. 37, no. 5, pp. 1498–1503, Mar. 2019.

[30] Y. Xie, L. Zhuang, and A. J. Lowery, "Picosecond optical pulse processing using a terahertz-bandwidth reconfigurable photonic integrated circuit," *Nanophotonics*, vol. 7, no. 5, pp. 837–852, 2018.

[31] G. Weerasekara and A. Maruta, "Eigenvalue based analysis of soliton fusion phenomenon in the frame work of nonlinear Schrödinger equation," *IEEE Photon. J.*, vol. 9, no. 3, Jun. 2017, Art. no. 7903612.

[32] T. Gui, T. H. Chan, C. Lu, A. P. T. Lau, and P.-K. A. Wai, "Alternative decoding methods for optical communications based on nonlinear Fourier transform," *J. Lightw. Technol.*, vol. 35, no. 9, pp. 1542–1550, May 2017.

[33] J. Koch, K. Chan, C. G. Schaeffer, and S. Pachnicke, "Signal processing techniques for optical transmission based on eigenvalue communication," *J. Sel. Topics Quantam Electron.*, vol. 27, no. 3, May/Jun. 2021, Art. no. 5100214.

[34] J. Müller *et al.*, "Silicon photonics WDM transmitter with single section semiconductor mode-locked laser," *Adv. Opt. Technol.*, vol. 4, no. 2, pp. 119–145, 2015.

[35] A. Zazzi *et al.*, "Optically enabled ADCs and application to optical communications," *IEEE Open J. Solid-State Circuits Soc.*, vol. 1, pp. 209–221, 2021.

[36] J. Koch *et al.*, "Silicon photonics DWDM NLFT soliton transmitter," in *Proc. 21st ITG-Symp. Photon. Netw.*, 2020, pp. 93–100.

[37] S. S. Azadeh *et al.*, "Advances in silicon photonics segmented electrode Mach-Zehnder modulators and peaking enhanced resonant devices," in *Proc. SPIE*, vol. 9288, 2014, Art. no. 928817.

[38] F. Merget *et al.*, "Silicon photonics plasma-modulators with advanced transmission line design," *Opt. Exp.*, vol. 21, no. 17, pp. 19593–19607, 2013.

[39] A. Moscoso-Mártir *et al.*, "Silicon photonics DWDM NLFT soliton transmitter implementation and link budget assessment," in *Proc. Eur. Conf. Integr. Opt.*, Jun. 2020, pp. 1–3. [Online]. Available: https://arxiv.org/abs/2003.08358

[40] J. C. C. Mak *et al.*, "Automatic resonance alignment of high-order microring filters," *IEEE J. Quantam Electron.*, vol. 51, no. 11, Nov. 2015, Art. no. 0600411.

[41] P. J. Winzer, D. T. Neilson, and A. R. Chraplyvy, "Fiber-optic transmission and networking: The previous 20 and the next 20 years," *Opt. Exp.*, vol. 26, no. 18, pp. 24190–24239, Sep. 2018.

[42] S. K. Turitsyn *et al.*, "Nonlinear Fourier transform for optical data processing and transmission: Advances and perspectives," *Optica*, vol. 4, no. 3, pp. 307–322, 2017.

[43] V. E. Zakharov and A. Shabat, "Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media," *J. Exp. Theor. Phys.*, vol. 34, no. 1, pp. 62–69, 1972.

[44] F. J. García-Gómez and V. Aref, "Statistics of the nonlinear discrete spectrum of a noisy pulse," *J. Lightw. Technol.*, vol. 37, no. 14, pp. 3563–3570, Jul. 2019.

[45] M. Kamalian, J. E. Prilepsky, S. T. Le, and S. K. Turitsyn, "On the design of NFT-based communication systems with lumped amplification," *J. Lightw. Technol.*, vol. 35, no. 24, pp. 5464–5472, Dec. 2017.

[46] V. Aref, "Control and detection of discrete spectral amplitudes in nonlinear Fourier spectrum," 2016. [Online]. Available: http://arxiv.org/abs/1605.06328

[47] W. Q. Zhang, T. H. Chan, and S. Afshar, "Direct decoding of nonlinear OFDM-QAM signals using convolutional neural network," *Opt. Exp.*, vol. 29, no. 8, pp. 11591–11604, Apr. 2021.