



# Deep learning reconstruction of ultrashort pulses from 2D spatial intensity patterns recorded by an all-in-line system in a single-shot

RON ZIV,<sup>1</sup> ALEX DIKOPOLTSEV,<sup>2</sup> TOM ZAHAVY,<sup>1</sup> ITTAI RUBINSTEIN,<sup>2</sup>  
PAVEL SIDORENKO,<sup>2</sup>  OREN COHEN,<sup>2</sup> AND MORDECHAI SEGEV<sup>2,\*</sup>

<sup>1</sup>*Department of EE, Technion, Haifa 32000, Israel*

<sup>2</sup>*Department of Physics and Solid State Institute, Technion, Haifa 32000, Israel*

\**msegev@technion.ac.il*

**Abstract:** We propose a simple all-in-line single-shot scheme for diagnostics of ultrashort laser pulses, consisting of a multi-mode fiber, a nonlinear crystal and a camera. The system records a 2D spatial intensity pattern, from which the pulse shape (amplitude and phase) are recovered, through a fast Deep Learning algorithm. We explore this scheme in simulations and demonstrate the recovery of ultrashort pulses, robustness to noise in measurements and to inaccuracies in the parameters of the system components. Our technique mitigates the need for commonly used iterative optimization reconstruction methods, which are usually slow and hampered by the presence of noise. These features make our concept system advantageous for real time probing of ultrafast processes and noisy conditions. Moreover, this work exemplifies that using deep learning we can unlock new types of systems for pulse recovery.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

## 1. Introduction

Ultrashort femtosecond-scale laser pulses are a key ingredient in time-resolved investigations of ultrafast phenomena [1–3], such as chemical reactions, electron dynamics in atoms and molecules etc., hence their complete characterization (amplitude and phase) is of great importance. However, sensor technology does not yet have short enough response time to recover ultrashort pulses directly. Consequently, ultrashort pulses are currently recovered indirectly, often through algorithmic methods. A widespread method is the Frequency-Resolved Optical Gating (FROG [4]) which is based on gating a pulse with a time shifted replica of itself inside a nonlinear medium, and measuring the power spectrum of the nonlinear signal as function of the delay between the pulses (the 2D FROG trace). A more recent method that gains popularity, known as d-scan (dispersion scan [5]), is based on varying the dispersion experienced by the probe pulse before the pulse passes through the nonlinear crystal. D-Scan relies on forming a 2D trace by stacking the nonlinear spectra at the different degrees of dispersion. The reconstruction algorithms used in both FROG and d-scan to recover the probed pulse from the recorded spectrograms are iterative phase-retrieval algorithms [4–6]. Recently, with the progress in machine learning [7,8], deep-learning-based reconstruction of ultrashort pulses was demonstrated in both FROG [9], d-scan [10] and other techniques [11], displaying considerable improvements in terms of speed of reconstruction and noise robustness, due to the intrinsic ability of deep learning to filter out noise [12]. However, in the field of diagnostics of ultrashort laser pulses, deep learning techniques were so far employed only for improving the performance of already existing schemes, but never as the pulse recovery method associated with a completely new scheme.

Conventional FROG and d-scan devices work in the multi-shot regime, which requires trains of (almost) identical pulses. However, in some experiments [13,14], the probed pulse is not part of a train of identical pulses, hence it is often desired to characterize a pulse using a single-shot characterization method. Indeed, single-shot variants of FROG (termed GRENOUILLE [15])

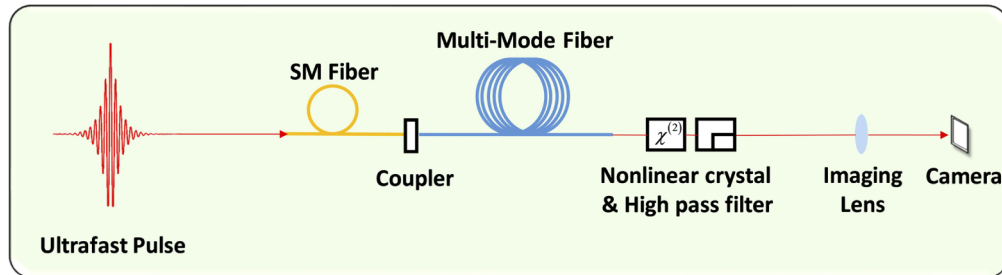
and of d-scan [16] were developed. In GRENOUILLE, a Fresnel bi-prism splits an incoming pulse beam into two non-collinear stripe beams, where the delay between the pulses is mapped to a spatial axis. The beams overlap in a thick nonlinear crystal, acting also as a spectrometer by utilizing the narrow bandwidth frequency-angle phase-matching dependence. In this scheme, the group velocity mismatch and crystal dispersion result in a tradeoff between the spectral bandwidth and temporal duration of the pulse, limiting the time-bandwidth product of the pulses such a device can recover. In single-shot d-scan, different transverse parts of the beam experience different degrees of dispersion by using a prism and imaging system. The number and accuracy of sampled dispersion points is limited by the resolution of the imaging system. Another single-shot method is Spectral Phase Interferometry for Direct Electric—Field Reconstruction (SPIDER [17]), which is based on spectral interferometry between the probed pulse and a frequency shifted replica of itself. The probed pulse is calculated directly from the measured 1D interferogram. However, SPIDER devices are generally more complicated than GRENOUILLE and single-shot d-scan. Also, generally single-shot devices are inherently highly sensitive to noise (as there is no averaging by multiple pulses and the power of the incoming signal is always limited), hence noise robustness of the reconstruction algorithm is especially critical in single-shot pulse recovery systems.

Here, we propose and demonstrate in simulations, a simple single-shot pulse characterization system, based on all-in-line propagation of the pulsed beam using off-the-shelf fibers and a  $\chi^{(2)}$  nonlinear crystal. The output of the crystal is imaged onto a camera that records a 2D speckle intensity pattern. We show that by using deep learning techniques, we can successfully reconstruct the pulse from the recorded data, even at low SNR, a previously unattainable feat for these types of measurement systems. We prove that this system can provide a practicable single shot measurement apparatus for diagnosis of ultrashort pulses, and validate the robustness of the reconstruction to physical variations in the system. Last but not least, the multi-mode fiber in our scheme can be replaced by other components that mix the beam's spectral and spatial degrees of freedoms (e.g. a thick diffuser), which would offer greater resolution and dynamic range.

## 2. Proposed system

The system is based on spatiotemporal coupling and a nonlinear measurement (see Fig. 1). The first element of the system is a single-mode fiber. When a pulse enters the single-mode fiber, its spatial profile is coupled (projected) to the spatial mode of the single-mode fiber, while retaining its spectrum and spectral phase. We assume that the field propagates linearly in the single-mode fiber, experiencing only dispersion. Next, for the purpose of creating spatiotemporal coupling, we use a coupler and a multi-mode fiber. The coupler is actually a diffuser that acts to project the field from the single-mode fiber onto the different modes of the multi-mode fiber, exciting each mode with a different amplitude and phase. The diffuser is positioned far enough from the SM fiber, such that the wave front incident upon the diffuser is approximately spherical, hence the coupling from the diffuser into the modes of the multimode fiber is of equal strength. In practice, any physical realization that will couple deterministically between the (one) mode exiting the SM fiber to the many modes of the multi-mode fiber would be suitable. In the multi-mode fiber, each mode has a different spatial profile and a different rate of phase accumulation according to the modal dispersion and the wavelength. Next, we introduce a  $\chi^{(2)}$  nonlinear medium with a large enough numerical aperture such that the highly multimode field emerging from the fiber is coupled to the crystal without being truncated (readily done because these crystals are bulk materials whose numerical aperture can be fairly large). Under the common assumptions of slowly varying envelope and non-depleted pump approximation, the electric field emerging from the crystal is composed of the outcome of all the  $\chi^{(2)}$  processes occurring among the frequency components of the pulse: sum frequency generation, difference frequency and rectification (for simplicity, we neglect cascaded effects where  $\chi^{(2)}$  operates twice

or more). The emerging field is then passed through an optical high pass filter, keeping only the sum frequencies while blocking the original frequencies associated with the power spectrum of the original pulse. This complex high-frequency field is then directly imaged onto a camera that measures the time-averaged interference pattern created by the sum-frequency field of the pulse. The nonlinearity of this process, combined with the spatiotemporal mixing due to the diffusive coupler and the propagation through the multi-mode fiber, creates an interference pattern that depends on the amplitude and spectral phase of the pulse. As described below, this interference pattern allows for deciphering the amplitude and phase of the incoming pulse.



**Fig. 1.** Proposed system for ultrafast pulse measurement. The ultrashort pulse is passed through a single mode (SM) fiber, and subsequently through a coupler that couples to multiple modes of a multimode fiber. The resultant spatio-temporal pulse exiting the multimode fiber is passed through a nonlinear crystal (NLC) and then through a high-pass frequency filter (HPF) that passes only the sum-frequencies (blocking the frequencies associated with the spectrum of the original pulse). The ensuing interference pattern is directly imaged onto a camera.

We choose this system due to its simplicity and small size, and because it offers an easily controllable method of interference between different frequencies, by simply changing the propagation length in the dispersive medium (multi-mode fiber). We note that such a system, but without the nonlinearity, was used for high resolution low loss spectrometry [18]. However, that system measures only the power spectrum of the pulse while the spectral phase is lost, preventing complete pulse reconstruction.

We wish to note that the efficiency of the above  $\chi^{(2)}$  process depends on the selectivity of the phase matching condition, just like in any conventional FROG system. The phase matching selectivity depends on the spectral bandwidth of the pulses and the polarization. The spectral bandwidth of the pulses simulated here are commonly handled by a FROG system, with the same tradeoff and considerations as ours. Desirably, the polarization of the light going through the nonlinear crystal should typically be linear (like in a FROG system). In our system, light existing from a multi-mode fiber can be in mixed polarization state. To rectify that, we suggest using a polarization-maintaining multimode fiber, which can in principle be rather short – as long as it leads to accumulate sufficient phase differences between modes due to modal dispersion. Otherwise, the mixed polarization state of the light exiting the multimode fiber can also be corrected by a polarizer, and then accounted for in the modeling of the network and while training the network. In what follows, we assume for simplicity that the light emerging from the multimode fiber maintains the linear polarization.

Next, we analyze our proposed system theoretically and simulate examples. To that end, we calculate the spatiotemporal coupling of the pulse emerging from the single-mode fiber to the modes of the multi-mode fiber by writing an analytic description of the propagation of the pulse in the system. For simplicity, we use the modes of a step-index multi-mode fiber as our mathematical basis. We will limit our basis to the family of linearly polarized modes, under the weakly guided approximation (a full description of the modes can be found in [19]). Thus, we

can construct a transfer operator which will evolve an incoming single mode field, in the form of a pulse emerging from the single-mode fiber, through the multi-mode fiber. We write it in the following form:

$$T(r, \theta, z, \omega) = \sum_{l,m} \alpha_{l,m}(\omega) E_{l,m}(r, \theta, z, \omega) \quad (1)$$

where  $E_{l,m}$  are the spatial profiles of the  $(l, m)$  mode including also the relevant phase accumulation along the propagation axis  $z$ , and  $\alpha_{l,m}$  are complex valued coefficients representing the coupling created by the diffuser coupler and the single mode field input. These coefficients represent both magnitude and phase coupling. Therefore, from this point onwards we can disregard the spatial profile of the incoming field and let  $\alpha_{l,m}$  encode this information directly. The coefficients  $\alpha_{l,m}$  may (or may not) depend on the frequency, depending on the coupling method between the single mode and the multimode fibers. In our simulations described below, we take these coefficients to be random, hence their explicit dependence on the wavelength is immaterial.

Using this transfer operator, we can write the complete field evolving in the multi-mode fiber, given an arbitrary input field  $E_{input}(\omega) = A(\omega)e^{i\phi(\omega)n}$  (of a given power spectrum and spectral phase):

$$\begin{aligned} E'(r, \theta, z, \omega) &= E_{input}(\omega) \cdot T(r, \theta, z, \omega) \\ &= A(\omega)e^{i\phi(\omega)} \cdot \sum_{l,m} \alpha_{l,m}(\omega) E_{l,m}(r, \theta, z, \omega) \end{aligned} \quad (2)$$

For simplicity, in our analysis and results, we will focus on obtaining the amplitude and phase of the pulse entering the multi-mode-fiber, noted as  $E_{input}(\omega)$ . As the propagation in the single-mode fiber is a linear operation, it can be readily back propagated to relate back to the original pulse and thus we can disregard this effect for our needs. Using (2) we write an analytic description of the sum frequency field generated in the nonlinear crystal, and the resultant interference pattern measured by camera:

$$M_{nonlinear}(r, \theta) = \int_{-\infty}^{\infty} I^2(r, \theta, z = L, t) dt = \int_{-\infty}^{\infty} |E(r, \theta, z = L, t)|^4 dt \quad (3)$$

where  $I(r, \theta, z = L, t)$  is the intensity of the field  $E(r, \theta, z = L, t)$  at the output of the fiber, as given by Eq. (2) and  $L$  is the length of the fiber. We want to describe the nonlinear interference process described by Eq. (3) through the spectral fields. For a linear process this is straightforward by using Parseval's theorem, while for this nonlinear interference we take a direct approach and write the inverse Fourier transform:

$$\begin{aligned} M_{nonlinear}(r, \theta) &= \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} E(r, \theta, z = L, \omega) e^{i\omega t} d\omega \right|^4 dt \\ &= \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} A(\omega) e^{i\phi(\omega)} \cdot T(r, \theta, z = L, \omega) e^{i\omega t} d\omega \right|^4 dt \end{aligned} \quad (4)$$

This is the first indication that the nonlinearity of the sum frequency generation process creates a non-trivial functional dependence on the spectral phase. It is important to note that the functional that maps an input pulse to the nonlinear interference is not injective, which means that the inversion of this system contains ambiguities. This can be seen easily by exploring two common trivial functional ambiguities: a global phase and a time shift (or a linear phase in the Fourier domain). But these ambiguities are trivial and generally of no practical importance, hence their effect can be mitigated by working in a predefined subspace of pulses with equal global phase and a fixed time shift.

### 3. Reconstruction method

To the best of our knowledge, there is no analytic solution for the extraction of spectrum and spectral phase from the spatiotemporal nonlinear interference pattern. This poses the question on

how can we reconstruct the incoming pulse out of the recorded intensity pattern. In principle, this is a regression problem and it can be stated as: given an interference pattern, we want to regress and find the pulse which created it. Mathematically, we can write the relation between a pulse and the interference pattern arising from that pulse, as the following functional:

$$\Gamma : E(\omega) \rightarrow M(r, \theta) \quad \begin{array}{l} E(\omega) \in \mathbb{C}^n \\ M(r, \theta) \in \mathbb{R}^{k \times k} \end{array} \quad (5)$$

where  $M(r, \theta)$  is a matrix representing the discrete sampled interference pattern,  $E(\omega)$  is a vector representing the discrete sampled complex valued pulse, and  $\Gamma$  is the discrete case of the integral derived in Eq. (4). In the regression problem we are trying to find the inverse of this relation.

We will take an approach similar to the one taken in [9] and construct a deep neural network to solve the regression problem to reconstruct the pulse from the recorded data. The deep neural network is a parametric function that can represent high-dimensional nonlinear functions. By learning the parameters of the network we are able to train the network to solve specific tasks. This concept has been around for some time now [20], but only in recent years, with the exponential growth in computational power and improved network architectures [21], deep networks are proving to be a formidable technique that can solve difficult and complex problems. By constructing an optimization problem on the parameters of a given network and using ground truth inputs and outputs (samples and labels), we can learn the correct parameters by solving the underlying optimization problem. Usually, direct solution is not possible due to the complexity of the problem, hence the problem is solved by iterations, using a variant of SGD (Stochastic Gradient Descent). In our problem, we pass an input image, i.e. the nonlinear interference pattern, through the computational layers of the network and receive an output vector which represents the temporal electric field.

This process of optimizing the network parameters is referred to as the training stage of the network. For this stage, we use two disjoint datasets of known pulses, “training set” and “test set”, each with its simulated intensity pattern recorded by the camera in our proposed setup. The first is used for training the network. Namely, in every iteration, each data sample changes the network parameters, with the purpose of decreasing the overall reconstruction error. The second dataset is used for testing the performance: a sample data is passed through the network to produce a reconstruction and find its error, but the network parameters are not varied by the test dataset. In training the network, it is important to avoid overfitting the parameters of the trained network to the specific data used for training, because tight overfitting might hamper the ability of high quality recovery of pulses that were not part of the training set. Hence, to prevent overfitting, we train the network on data from the training set, and then test its performance on the (disjoint) “test set”. This is done iteratively, until we find the best network parameters that yield the best fit on the testing data. In this way, we avoid overfitting and find the most suitable network parameters to recover the pulse.

In our problem, we train the network to approximate the inversion of the functional described in Eq. (5) by feeding into the network nonlinear interference patterns as input and the pulses as labels, and we aim to minimize the L1 loss between the output of the network and the given label, i.e.:

$$w = \arg \min_{w'} \|DNN(I; w') - E(t)\|_1 \quad (6)$$

where  $w$  are the parameters we wish to optimize,  $DNN(I; w')$  is the output of the network,  $I$  is the sum frequency generation interference pattern and  $E(t)$  is the time domain complex envelope of the field, where  $E(t) = F[A(\omega)e^{i\phi(\omega)}]$ . The L1 criteria is chosen as it promotes sparser solutions which are more robust to noise (we observe this empirically, as training with a L2 norm results in slightly noisier solutions).

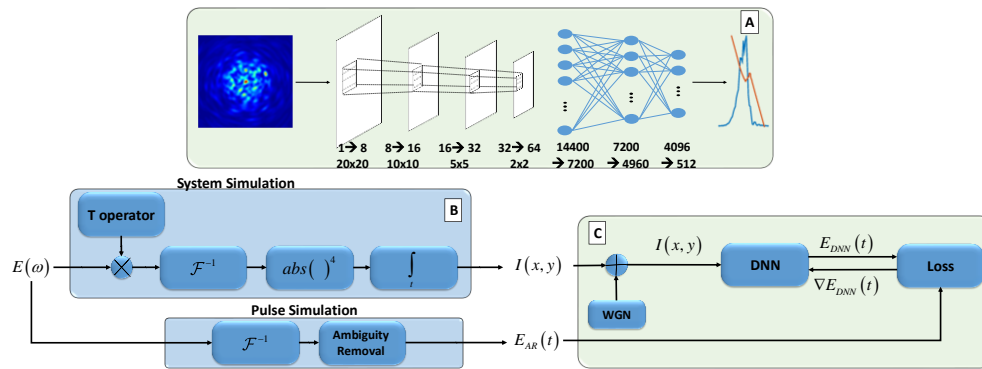


In practice, we will define our label as the real and imaginary part of the pulse, and concatenate them into a single vector. This means that we will try and find a  $\mathbb{R}^{2n}$  vector rather than a  $\mathbb{C}^n$  vector. This representation is less ambiguous than separating into amplitude and phase, and has shown better results in previous works [9].

Figure 2 shows the optimized architecture of our regression network used for reconstruction of pulses from nonlinear interference patterns. The network is composed of two major components, convolutional neural network (CNN) and a rectifier linear unit (ReLU). A CNN is a type of neural network which preforms cross-correlation between an input object and a learned kernel, meaning that each output value is composed of a local weighted average defined by the kernel. ReLU is a type of nonlinear activation function with the following form:

$$f(x) = \max(0, x) \quad (7)$$

Our network is constructed by a 4-layer CNN followed by three fully connected layers with ReLU nonlinear activation between them. Figure 2 also shows the channels and filter sizes in each layers.



**Fig. 2.** (A) Regression network architecture: four CNN layers followed by three fully connected layers. The input to this network is a sum frequency interference pattern, which is passed through the computational layers, until a final output is produced in the form of a vector of the real and imaginary parts of the temporal electric field. (B) Block diagram of sum frequency interference measurement and of the label generation from an input spectral pulse, as described in section 2 and section 3. The input of the simulation is passed on to the regression network. (C) Supervised training of the regression network. Each interference pattern is passed through the network to create a reconstruction. The error between a reconstructed pulse and its ground truth pulse shape is used in back propagation and gradient descent to train the network and improve the network parameters.

## 4. Results

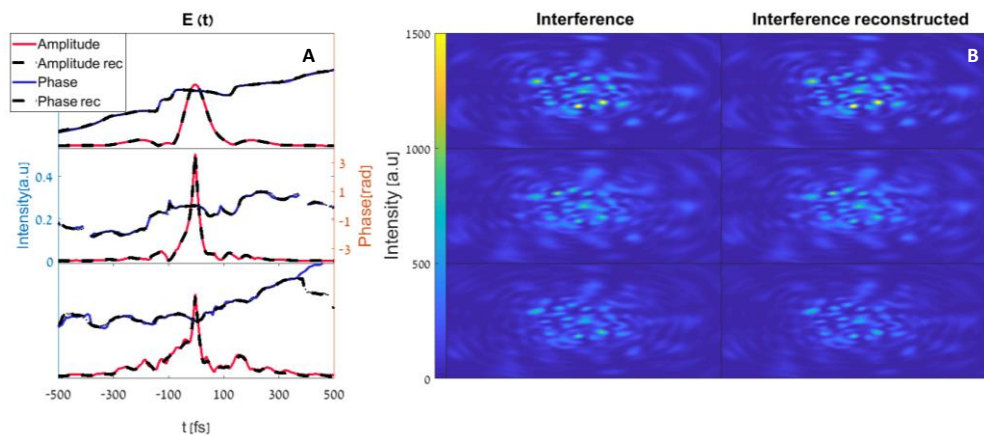
In order to learn the functional mapping described in Eq. (6) we need to create a dataset representing the physical relation between the pulse and the nonlinear interference pattern measured by the camera. Our simulated system consists of a multi-mode fiber of length 1cm, core radius of  $50\mu\text{m}$  and refractive indices of  $n_{\text{core}} = 1.4699$ ,  $n_{\text{clad}} = 1.4533$  (parameters of a commercially available multimode fiber). This results in a numerical aperture of  $\text{NA} \approx 0.22$  for the fiber (a reasonable NA for phase matching in crystals that are not too long and coherence length that is not too short). Using these parameters we construct a transfer operator in the manner described in (1), with  $l \in [0, 19]$ ,  $m \in [0, 10]$  and we choose the coupling coefficients,  $\alpha_{l,m}$ , as  $\alpha_{l,m} = e^{ix}$  where  $x \sim \text{Uni}[0, 2\pi]$ . In a real physical system each coefficient will have a

slightly different amplitude with some dependence on the wavelength; however, for a large enough number of modes in the multi-mode fiber, we assume that this choice of coupling coefficient is reasonable for a proof-of-concept demonstration.

In our simulations, we use a mixture of pulses with either a Gaussian or a Lorentzian power spectrum with random varying width,  $\text{width}_{\text{Gaussian}} = 68 \pm 39\text{Thz}$ ,  $\text{width}_{\text{Lorentzian}} = 45 \pm 2.6\text{Thz}$ , with a randomly generated spectral phase. The spectral phase is smoothed by removing high frequencies in order to ensure the pulse to be time limited yet complex, selected to have pulses with time bandwidth product in the range of  $[0.5, 5]$ . As noted, to be able to invert the functional relation, we remove the trivial phase and time shift ambiguities by setting the phase at the center of the pulse to be zero and centering the max amplitude of the pulse to the center of the vector.

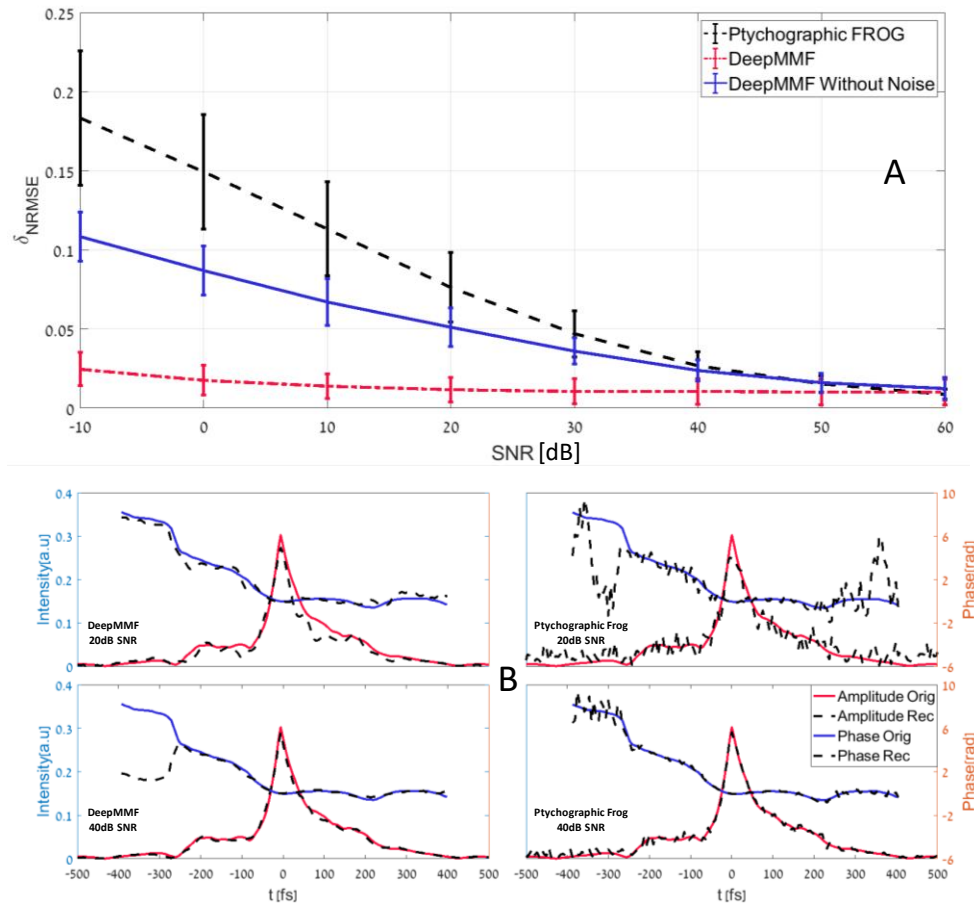
Working in the digital domain, we discretize the above functions using a linearly spaced spectral grid points in the range of  $\omega \in [1.93\text{Phz}, 3.01\text{Phz}]$  with spacing of  $\Delta\omega = 4.24\text{Thz}$ , equivalent to temporal resolution of 5.8fs, corresponding to 256 spectral points and wavelength span in the range  $\lambda \in [625\text{nm}, 975\text{nm}]$ . The 2D linear spatial grid is chosen such that each pixel is of the size of  $0.21\mu\text{m} \times 0.21\mu\text{m}$  with  $256 \times 256$  spatial points corresponding to  $(x, y) \in [-27.5\mu\text{m}, 27.5\mu\text{m}] \times [-27.5\mu\text{m}, 27.5\mu\text{m}]$ , preserving resolution in regions of interest yet keeping computational effort low. The produced image represents the center section of the fiber, where the rest of the image from the fiber is cropped out and is not fed into the network (so as to reduce the input dimension of the data for computational efficiency). We do not see any performance degradation due to this cropping.

To train the system, we create a dataset of 100,000 pulses and their corresponding nonlinear interference pattern, as described in section 2. Through random sampling of pulses from the train set over many iterations, we pass the interference patterns through the network. This gives us the reconstructed pulses. By computing the error between reconstruction and ground truth pulses, we back-propagate the error and optimize the network parameters, by the process of gradient descent. In practice, we use a popular variant of gradient descent named ADAM [22]. This process is illustrated in Fig. 2. By this process of training, we create two models (the set of parameters that describe the network): one trained without noise present in measurements and one trained with an additive white Gaussian noise of  $\text{SNR} \sim \text{Uni}[0\text{dB}, 30\text{dB}]$ .



**Fig. 3.** Three examples of pulse reconstruction with varying complexity: Top simple, middle medium, bottom most complex. (A) Temporal amplitude and phase of an original and reconstructed pulse. (B) The nonlinear interference of the original pulse and its reconstructed counterpart.

Having trained the network, we now test its performance by creating a test set of 7000 pulses, which are not part of the training set, that is, the test pulses have not been previously seen by the network. By passing the interferences images of the test set pulses through the net, we obtain a set of reconstructed pulses. Examples of such test pulses and their reconstructions can be seen in Fig. 3 for the noiseless model and with no noise present in the test set measurements. Using the L1 error criterion between ground truth pulses and reconstructed pulses on noiseless measurements, on average we obtain a final reconstruction error of  $0.91 \times 10^{-3} \pm 0.71 \times 10^{-3}$  on the test set and  $0.83 \times 10^{-3} \pm 0.57 \times 10^{-3}$  on the train set. This error is very low, compared to the pulses norm which is normalized to unity. We also compare the reconstruction quality in terms of the NRMSE (normalized root mean squared error), defined as  $\delta_{\text{NRMSE}} = \sqrt{\frac{\|E_G - E_R\|_2^2}{N \cdot \text{Max}(E_G)}}$  between the ground truth and the reconstructed pulses and  $N$  being the vector length. We

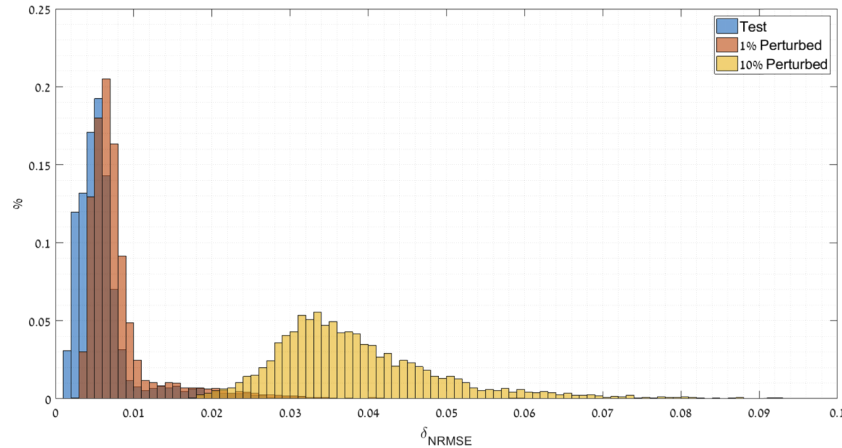


**Fig. 4.** (A) Error as a function of SNR for reconstruction of 30 pulses using ptychography reconstruction on FROG measurements (black line), and using the two DeepMMF models: one trained with AWG noise (red line) and the other trained without noise (blue line), both trained on measurements of pulses in our system. The error bars are  $\sigma_{STD}$  of the error. (B) Reconstruction examples for one pulse, using the two reconstruction algorithms and at two SNR points. Left Column displays DeepMMF reconstruction, right is ptychography. Top Row is at 20dB SNR, Bottom 40dB SNR.



obtain  $\delta_{\text{NRMSE}} = 0.64 \pm 0.59\%$  which indicates good reconstruction quality as on average each reconstructed pulse is highly correlated with its corresponding ground truth.

We also compare the trained models performance with and without AWG noise present in test set measurements, which can be seen in Fig. 4. As expected, when SNR is high the two models achieve similar accuracy, yet once noise is introduced and the SNR is lowered - the model trained with noise shows greater noise immunity and reconstruction accuracy, to the extent 22% improvement for high SNR and 300% for very low SNR. For comparison sake, we create FROG traces from the same pulses and compare their reconstruction using the ptychography algorithm [23] as shown in Fig. 4.



**Fig. 5.** Probability distribution of normalized root mean squared error (NRMSE) reconstruction errors of the three test sets, done on 7000 pulses.

Next, we test the robustness to transfer operator estimation mismatch, i.e. the sensitivity of the network to errors in the operator. This test is motivated by the algorithmic challenge this system introduces: the transfer coefficient of a real lab coupler is unknown, hence, in order to simulate a real system, one would need to measure or estimate these coefficients. This will be further elaborated in the discussion. For the sake of comparison, we create 3 instances of operators: a ground truth operator, an operator whose  $\alpha_{l,m}$  coefficients differ by 1% and an operator with 10% difference from the ground truth. The difference is generated by adding a  $x\%$  uniform random phase to the coefficients. We train a network using the train set from the ground truth operator and test it using test sets from the three operators. The results of this test can be seen in Fig. 5. We can see that the 1% difference in coefficients created a small shift in the histogram towards higher errors which means there was a decrease in performance. As this degradation is small, compared to variance of the errors within the pulse distribution, it shows that the system is robust to small deviations in the physical elements.

## 5. Conclusions

We proposed a new simple all in-line system for ultrashort pulse reconstruction from sum frequency field interference measurements, employing deep learning to numerically invert the nonlinear interference pattern to pulse mapping, as a method for analytical inversion is unknown. Using simulated data, we have shown that this method achieves good reconstruction results, with an average NRMSE error of  $0.64 \pm 0.59\%$ .

The system shows good performance even in scenarios of low SNR, which is a key property for single-shot ultrashort pulse measurement system. We expect our method to be especially

useful in cases where single-shot measurements are required and there is a power constraint on the pulses employed (e.g. when studying biological specimen).

As the transfer operator of the presented device is system-dependent, the training of the network must be done on a specific real system. This can be achieved by either taking interference measurements of known pulses (labels) from the system itself or by simulation of measurements of simulated pulses using the parameters of the system. Naturally, using simulated data (instead of data from experiments) for training the network is highly desirable, but this may pose challenges due to discrepancies between real and simulated data (problem known as “sim to real” in deep networks). Therefore, showing that the system is robust to small errors in the transfer operator implies that we can rely on estimating the parameters of the fiber and the coupler rather than measuring them precisely, and thereby train the network on simulated pulses. Although beyond the scope of this work, it is also possible to use the data itself to extract these coupling coefficients and thus improve on the systems performance in general.

We emphasize that our training stage and therefore reconstruction algorithm is system agnostic. That is, we do not assume any prior knowledge (or any information) on the system itself or on the transfer function at the training stage. The network extracts this information by itself during the training. Thus, as long as the data presented to the network in the training stage (interference measurements and pulse shapes) represents the functional physical mapping in the setup - the system will perform well – even if the actual transfer operator is different than the one we simulated here (Fig. 1). For example, non-ideal effects such as mode coupling due to bent fibers, and other additional effects that may arise while constructing such a system in experiments, will be learned directly from the data.

Before closing, we note that the system proposed here introduces a new feature with respect to making use of the redundancy of the measurements. Namely, the recorded data in FROG and d-scan is generally redundant. That is, there are many more measurements than required for reconstructing the pulse uniquely. This redundancy makes these techniques powerful, giving rise to robustness to noise and misalignment as well as allowing reconstructing a pulse at higher temporal resolution from the temporal scanning steps in FROG [23] and reconstructing multiple pulses from a single FROG trace [24,25]. However, the measured data for each scanning step in FROG or for specific dispersion in d-scan is not redundant; in fact, it does not even contain enough information to reconstruct the pulse. Here comes into play another added value of our technique: in our system, due to the strong spatiotemporal mixing in the MMF, the temporal information of the pulse is mixed completely in the 2D image. There is no separation into rows or columns. Hence, in principle, every subset of the image is highly redundant. This endows our technique with possibly greater robustness that can make it advantageous for real time probing of ultrafast processes and under noisy conditions. Also, it is important to note that our technique does not suffer from the stringent physical restrictions characteristic of other single-shot methods, for example, on the time-bandwidth product [15] or on the spatial beam profile [16], avoiding these limitations facilitates greater dynamic range and robustness.

Finally, it is worth noting that the architecture of the network can be further improved by adding a generative model, such as Generative Adversarial Network (GAN) [26], which can be used for enriching the distribution of pulses learned and thus improving robustness of reconstruction. Thus, we believe that this method can achieve state of the art results, comparable to FROG reconstruction accuracy, while improving the noise immunity and thus improving SNR.

We recently became aware of a closely related paper that was posted on the arxiv [27], demonstrating a system of a similar concept of using a single-shot scheme for diagnostics of ultrashort laser pulses, consisting of a multi-mode fiber, nonlinear spatio-temporal mixing, and reconstruction via Deep Learning.

## Acknowledgment

We gratefully acknowledge the support of the Israel Science Foundation (ISF).

## Disclosures

The authors declare no conflicts of interest.

## References

1. S. X. Hu and L. A. Collins, "Attosecond Pump Probe: Exploring Ultrafast Electron Motion inside an Atom," *Phys. Rev. Lett.* **96**(7), 073004 (2006).
2. W. Demtroder, "Laser Spectroscopy: Basic Concepts and Instrumentation, Second Enlarged Edition," *Opt. Eng.* **35**(11), 3361 (1996).
3. K. E. Sheetz and J. Squier, "Ultrafast optics: Imaging and manipulating biological systems," *J. Appl. Phys.* **105**(5), 051101 (2009).
4. R. Trebino, *Frequency-Resolved Optical Gating: The Measurement of Ultrashort Laser Pulses* (Springer US, 2000), Vol. 62.
5. M. Miranda, C. L. Arnold, T. Fordell, F. Silva, B. Alonso, R. Weigand, A. L'Huillier, and H. Crespo, "Characterization of broadband few-cycle laser pulses with the d-scan technique," *Opt. Express* **20**(17), 18732 (2012).
6. N. C. Geib, M. Zilk, T. Pertsch, and F. Eilenberger, "Common pulse retrieval algorithm: a fast and universal method to retrieve ultrashort pulses," *Optica* **6**(4), 495 (2019).
7. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst.* **60**(6), 87–90 (2017).
8. A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature* **542**(7639), 115–118 (2017).
9. T. Zahavy, A. Dikopoltsev, D. Moss, G. I. Haham, O. Cohen, S. Mannor, and M. Segev, "Deep learning reconstruction of ultrashort pulses," *Optica* **5**(5), 666–673 (2018).
10. S. Kleinert, A. Tajalli, T. Nagy, and U. Morgner, "Rapid phase retrieval of ultrashort pulses from dispersion scan traces using deep neural networks," *Opt. Lett.* **44**(4), 979 (2019).
11. J. White and Z. Chang, "Attosecond streaking phase retrieval with neural network," *Opt. Express* **27**(4), 4799 (2019).
12. P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. A. Manzagol, "Stacked denoising autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion" *J. Mach. Learn. Res.* (2010).
13. C. Horn, M. Wollenhaupt, M. Krug, T. Baumert, R. de Nalda, and L. Bañares, "Adaptive control of molecular alignment," *Phys. Rev. A* **73**(3), 031401 (2006).
14. Y.-H. Chen, S. Varma, I. Alexeev, and H. Milchberg, "Measurement of transient nonlinear refractive index in gases using xenon supercontinuum single-shot spectral interferometry," *Opt. Express* **15**(12), 7458 (2007).
15. P. O'Shea, M. Kimmel, X. Gu, and R. Trebino, "Highly simplified device for ultrashort-pulse measurement," *Opt. Lett.* **26**(12), 932 (2001).
16. D. Fabris, W. Holgado, F. Silva, T. Witting, J. W. G. Tisch, and H. Crespo, "Single-shot implementation of dispersion-scan for the characterization of ultrashort laser pulses," *Opt. Express* **23**(25), 32803 (2015).
17. C. Iaconis and I. A. Walmsley, "Spectral phase interferometry for direct electric-field reconstruction of ultrashort optical pulses," *Opt. Lett.* **23**(10), 792 (1998).
18. B. Redding and H. Cao, "Using a multimode fiber as a high-resolution, low-loss spectrometer," *Opt. Lett.* **37**(16), 3384–3386 (2012).
19. K. Okamoto, *Fundamentals of Optical Waveguides* (2006), Chap. 3.
20. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature* **323**(6088), 533–536 (1986).
21. Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *ISCAS 2010 - 2010 IEEE International Symposium on Circuits and Systems: Nano-Bio Circuit Fabrics and Systems (2010)*, pp. 253–256.
22. D. P. Kingma and J. L. Ba, "Adam: A method for stochastic gradient descent," *ICLR Int. Conf. Learn. Represent.*, 1–15 (2015).
23. P. Sidorenko, O. Lahav, Z. Avnat, and O. Cohen, "Ptychographic reconstruction algorithm for frequency-resolved optical gating: super-resolution and supreme robustness," *Optica* **3**(12), 1320 (2016).
24. C. Bourassin-Bouchet and M.-E. Couprie, "Partially coherent ultrafast spectrography," *Nat. Commun.* **6**(1), 6465 (2015).
25. G. I. Haham, P. Sidorenko, O. Lahav, and O. Cohen, "Multiplexed FROG," *Opt. Express* **25**(26), 33007 (2017).
26. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," *Adv. Neural Inf. Process. Syst.* **27**, 2672–2680 (2014).
27. W. Xiong, B. Redding, S. Gertler, Y. Bromberg, H. Tagare, and H. Cao, "Deep learning of ultrafast pulses with a multimode fiber," arXiv:1911.00649 (2019).