

Full length article

## Machine learning methods for schlieren imaging of a plasma channel in tenuous atomic vapor

Gábor Bíró<sup>a,1</sup>, Mihály Pocsai<sup>a,1</sup>, Imre F. Barna<sup>a</sup>, Gergely G. Barnaföldi<sup>a</sup>, Joshua T. Moody<sup>b</sup>, Gábor Demeter<sup>a,\*</sup>

<sup>a</sup> Wigner Research Centre for Physics, Budapest, Hungary

<sup>b</sup> Max Planck Institute for Physics, Munich, Germany

### ARTICLE INFO

#### Keywords:

Machine learning  
Schlieren imaging  
Plasma diagnostic  
Near-resonant laser probe

### ABSTRACT

We investigate the usage of a schlieren imaging setup to measure the geometrical dimensions of a plasma channel in atomic vapor. Near resonant probe light is used to image the plasma channel in the tenuous vapor and machine learning techniques are tested for extracting quantitative information from the images. We build a database of simulated signals with a range of plasma parameters for training Deep Neural Networks, and demonstrate that they can estimate, from the schlieren images, reliably and with high accuracy the location, the radius and the maximum ionization fraction of the plasma channel as well as the width of the transition region between the core of the plasma channel and the unionized vapor. We test several different neural network architectures with supervised learning and show that the parameter estimations supplied by the networks are resilient with respect to slight changes of the experimental parameters that may occur in the course of a measurement.

### 1. Introduction

Rapid developments in computing and information science in recent years led to increasingly sophisticated Machine Learning (ML) implementations. The list of possible applications is ever growing, including (but not limited to) autonomous driving [1], healthcare [2], speech recognition [3] and various high-energy physics studies [4,5]. Machine learning methods have been used for some time also for evaluating optical diagnostic measurements in plasma physics, for example, tomographic measurements of radiation from fusion plasmas [6–9].

Schlieren imaging is a sensitive method for the detection of refractive index variations in transparent media, used widely in aeronautics and fluid dynamics [10]. The method is also extensively used for the investigation of plasma processes in atmospheric gases [11] and, in particular for a wide range of processes involving laser induced plasma [12–16]. Quite recently, ML techniques have been proposed to extract information from schlieren imaging measurements of flows and shocks [17–19].

Plasma wakefield acceleration is a technology that promises a new generation of compact particle accelerators for scientific and commercial uses [20,21]. Numerous research groups and collaborations are working worldwide to overcome the technological challenges that wakefield acceleration poses. The Advanced Wakefield Experiment

(AWAKE) hosted at CERN is a project where a high-energy proton bunch is used to drive plasma wakefields for electron acceleration [22, 23]. At the heart of the novel accelerator device, a 10-meter-long plasma channel achieves the modulation of the energetic proton bunch and the acceleration of witness electron bunches in the emerging wakefields. Created via photoionization using a terawatt (TW) power laser system in a rubidium vapor source chamber [24,25], plasma channel generation is in itself a complex problem of laser beam propagation/filamentation [26–28]. Optical diagnostic tools monitoring the plasma channel can thus have a significant role in optimizing, improving the accelerator device and understanding wakefield physics.

In this paper, we consider using a schlieren imaging setup as a diagnostic tool to determine vital parameters of a narrow plasma channel in tenuous ( $N = 10^{14} - 10^{15} \text{ cm}^{-3}$ ) atomic vapor. The setup is similar to that tested to observe atomic excitation in rubidium vapor [29] and is geared toward determining the precise location and diameter of the rubidium plasma channel as well as the characteristic length for the spatial decay of plasma density. We test the use of Deep Neural Networks (DNNs) as universal function approximators to extract quantitative information on the plasma from the schlieren images. We build datasets of simulated measurements to train networks with different architectures to estimate the parameters of the plasma and study the

\* Corresponding author.

E-mail address: [demeter.gabor@wigner.hu](mailto:demeter.gabor@wigner.hu) (G. Demeter).

<sup>1</sup> These authors contributed equally.

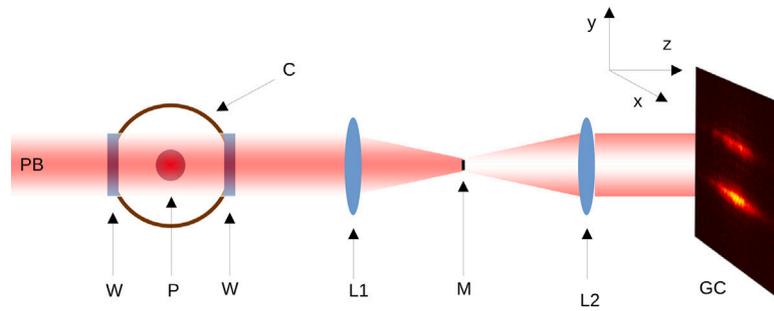


Fig. 1. Sketch of the schlieren imaging measurement setup (not drawn to scale). PB – probe beam, C – vapor source chamber cross-section, W – viewport, P – plasma channel cross section, L1, L2 – 75 cm focal length lenses in 4f setup, M – mask, GC – gated camera.

accuracy of their estimates. To prove their usefulness in practice, we also investigate the accuracy of parameter prediction in cases where the signals processed by the networks are derived under experimental conditions that differ somewhat from that of the training dataset. We demonstrate that schlieren imaging and machine learning techniques can be used effectively together to obtain information crucial for the operation of a proton-driven wakefield accelerator. On a more general note, the method can readily be applied for investigating the properties of plasma channels that arise during laser filamentation in atmospheric gases [30].

## 2. Schlieren imaging of a plasma channel cross-section

### 2.1. Measurement principle

In the novel electron accelerator device of the AWAKE facility, a 10-meter-long plasma channel is used as an energy-exchange medium between the high-energy proton driver and the witness electrons to be accelerated. Plasma for accelerator operation must satisfy very stringent constraints with respect to the spatial dependence of density [24]. The AWAKE plasma concept is to use a rubidium atomic vapor source, where the vapor density in the chamber can be tailored precisely by strict temperature control. A TW class laser system is used to propagate a 120 fs duration laser pulse along the chamber axis to achieve single-electron ionization of the atoms with a probability very close to unity [22,31]. Finely controlled vapor density is thus converted to precisely engineered plasma density.

The propagation of the ultra-short, TW ionizing pulse along the vapor source is itself a complex process of nonlinear optical interaction [26], especially because it is resonant with the rubidium  $D_2$  transition line [27,28]. Validating the quality of the plasma can be done near the downstream end of the vapor source, where a pair of observation ports on opposite sides of the chamber allow the passage of a probe beam transverse to the plasma channel axis. Precise measurements are hampered by the fact that plasma density distribution should be observed on a timescale much shorter than the  $\sim 10 \mu\text{s}$  recombination/diffusion timescales and that the vapor (and hence the plasma) is extremely tenuous, its  $\mathcal{N} = 10^{14} - 10^{15} \text{ cm}^{-3}$  density being 4–5 orders of magnitude smaller than the normal atmospheric density.

The schlieren imaging measurement setup that can be used for the required observation is sketched on Fig. 1. A Gaussian probe beam with beam waist parameter  $w_0 = 2.6 \text{ mm}$  transits the chamber of the vapor source along the  $z$  axis through a pair of viewports. The diameter of the chamber cross-section is 4 cm and the probe beam waist is positioned near the center of the chamber. Two lenses with focal lengths  $f = 75 \text{ cm}$  are placed in a ‘4f’ setup [32] after the chamber, with a  $D = 1.5 \text{ mm}$  diameter circular mask positioned on the optical axis in the back focal plane of the first lens. A gated, image intensified camera (GC) detects the probe beam, triggered about 100 ns after the ionizing laser and gated to collect light for 100 ns exposure time. The probe beam is from a diode laser tuned to  $\lambda = 780.311 \text{ nm}$ ,

close to the  $\lambda = 780.241 \text{ nm}$   $D_2$  resonance wavelength of rubidium. With this choice, the vapor’s anomalous dispersion around the resonance line yields a refractive index change of  $\delta n_v = 10^{-4} - 10^{-3}$ . At the same time, refractive index change due to the plasma dispersion is  $\delta n_p = \sqrt{1 - \omega_p^2/\omega^2} - 1 = (-2.7) \cdot 10^{-8} - (-2.7) \cdot 10^{-7}$ , over 3 orders of magnitude smaller in the region where the vapor is ionized. Thus the probe beam acquires a spatially dependent phase-shift and slight attenuation due to absorption upon transit. The circular mask at the focal plane between the two lenses acts as a high pass filter that blocks all of the probe light, unless a plasma channel (the schlieren object) modulates the probe beam phasefront sufficiently such that some of the probe is deflected around the mask’s edge.

An example of a measured image can be seen on Fig. 2(a). The probe beam diameter ( $\sim \text{mm}$ ) is negligible compared to the spatial scale ( $\sim \text{m}$ ) at which the plasma channel cross-section changes along the direction of propagation for the TW ionizing pulses (the  $x$  coordinate on Figs. 1 and 2). Therefore the measured image contains stripes parallel with the  $x$  axis and only variations along  $y$  are meaningful. It is convenient to create a 1D lineout along  $y$  by taking a region of interest (ROI) of  $x$  around the probe center and averaging the signal along  $x$  (Fig. 2(b)). This procedure helps to improve the signal-to-noise ratio. The task is to extract quantitative information on the plasma channel from this  $S(y)$  signal curve.

### 2.2. Parameters describing the plasma

To help interpret experimental images, we first note, that if the plasma distribution in the  $y - z$  cross-section plane is given, we can easily calculate the measured signal. The precise refractive index and absorption parameter for the probe beam wavelength can be obtained from the composite lineshape function using the material parameters of the rubidium  $D_2$  line [33]. (Note that the vapor densities used here require that we augment the description of [33] with a pressure broadening term in the homogeneous lineshape [34].) The integrated phaseshift of the probe beam and the overall attenuation can then be computed, and the transit across the 4f system with the mask can be calculated [32]. Therefore we start the analysis by assuming some functional form for the ionization probability in the vapor, and observing the *simulated* signal that this plasma distribution would produce.

To find a physically meaningful set of functions for the plasma density, we note that for relatively small ionizing pulse energies, we expect the ionization probability to be some power of the pulse intensity in general for multiphoton ionization. For large pulse energies, ionization probability saturates to values very close to unity in the central part of the beam [27,31]. The plasma channel is assumed to be axisymmetric in the  $y - z$  plane, with center relatively close to the axis of the vapor source (and optical axis). Since a shift of the plasma in the  $z$  direction (parallel to the probe beam propagation) cannot be detected by the setup (the quantities we measure arise as integrals along  $z$ ), we characterize the plasma center location with a single coordinate  $y_0$ , measured from the optical axis (i.e.  $y_0$  is the vertical coordinate

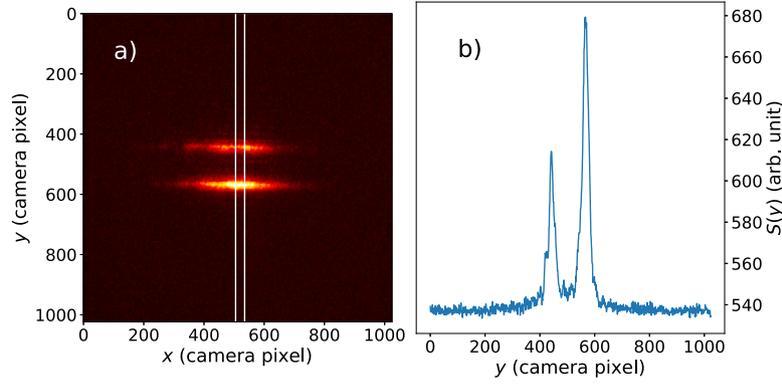


Fig. 2. (a) Schlieren image on the gated camera. Vertical lines near the middle mark the region of interest (ROI) from which we calculate the lineout  $S(y)$  around the probe beam center. (b) Lineout taken from the schlieren image ROI by averaging along  $x$ .

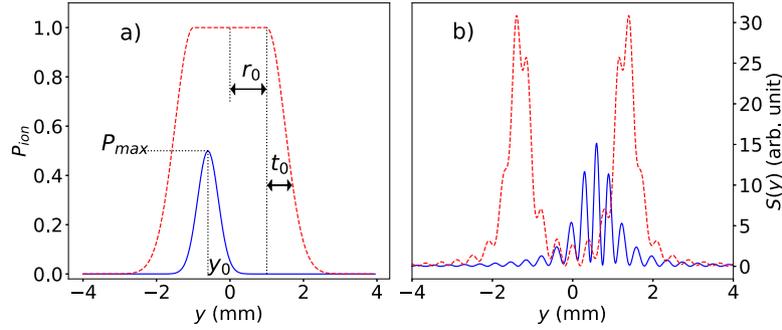


Fig. 3. (a) Ionization probability for a weakly ionized ( $P_{max} < 1, r_0 = 0$ ), narrow plasma channel (solid blue line) and a saturated core  $P_{max} = 1, r_0 > 1$ , wide plasma channel (dashed red line). (b) The corresponding simulated signals (signal for a weakly ionized plasma has been scaled up for better visibility).

of the plasma center). The plasma density is thus assumed to have the following form:

$$\mathcal{N}_p = \begin{cases} \mathcal{N}_0 P_{max}, & \text{if } r \leq r_0, \\ \mathcal{N}_0 P_{max} \exp\left(-\frac{(r-r_0)^2}{t_0^2}\right), & \text{if } r > r_0. \end{cases} \quad (1)$$

where the formula contains the set of plasma parameters:

$$C = \{y_0, r_0, t_0, P_{max}\}. \quad (2)$$

Here  $\mathcal{N}_0$  is the vapor density,  $P_{max}$  is the maximum ionization probability of the vapor at the plasma channel center,  $P_{max} \in [0, 1]$ .  $r = \sqrt{(y-y_0)^2 + z^2}$  is the geometric distance from the plasma channel center, located at  $(y, z) = (y_0, 0)$ ,  $r_0$  is the radius of the plasma channel core where the ionization fraction (and hence the plasma density) is constant, and finally  $t_0$  is the parameter that characterizes the width of the transition region between the plasma channel center and the completely ionized vapor.

The functional form written here can account for a weakly (not fully) ionized vapor, when  $r_0 = 0$  and the plasma density distribution is an axisymmetric Gaussian - a dependence we may expect for multiphoton ionization by a Gaussian laser beam. It can also account for the opposite case, when there is a sizeable domain of fully ionized vapor  $P_{max} = 1, r_0 > 0$  and a Gaussian shaped transition region surrounding it (see Fig. 3(a)). Clearly, not all parameter sets are physically realistic. Since the central, constant density region is associated with a saturation of the ionization fraction,  $r_0 > 0$  is compatible only with  $P_{max} \approx 1$ .

Given the above explicit functional form, we can calculate the simulated signal  $S$  on the gated camera for any set of plasma parameters  $C$ . This is a 1024 element vector of pixel values, some examples can be seen on Fig. 3(b). The mathematical task is now to determine the  $C$  that had been used to give rise to the given schlieren signal  $S$ . Because of the nonlinear, integral-type relationship between  $C$  and  $S$ , and the fact that

$S$  will also contain measurement noise, this is a difficult task. Therefore, in the following section we propose a novel method for processing the schlieren signals with the application of DNNs.

### 3. Inferring the plasma parameters with neural networks

Machine learning techniques have been successfully utilized in many fields, where it is an essential necessity to provide a precise and quick evaluation of the input data with significant non-linearities [35]. A typical data-based application of a feedforward artificial neural network is the non-linear regression, which aims to predict some parameters from the input data:

$$C_{j,pred} = f(S) = A\left(\sum_{i=1}^N w_{ij} s_i + b_j\right), \quad (3)$$

where  $A$  is some non-linear activation function,  $N$  is the number of the neurons in the layer,  $s_i$  are the discrete values of the signal vector  $S$ ,  $b_j$  is a bias vector and the  $w_{ij}$  matrix contains the trainable parameters. A network may consist of multiple such connected layers — in case it has at least one hidden layer between the input and the output, it is called a *Deep Neural Network*. If a suitable database of input data  $S_k$  is available with the corresponding reference parameters  $C_k$ , supervised training of the network is possible. During a supervised training cycle (*epoch*) of the network, the training input data  $S_k$  is evaluated and compared to the desired reference output  $C_k$  according to a well defined loss function,  $\mathcal{L}(C_{pred}, C_{true})$ . The objective then is to minimize this loss function by optimizing the weights in the  $w_{ij}$  matrices, which is performed by the backpropagation: the weights receive updates that are proportional to the partial derivatives of the loss function with respect to the weights. This process is then repeated iteratively in several epochs until some stopping condition. The final validation of training quality is done by evaluating the network predictions on a subset of the reference data — a part that was not included in the training process.

During a supervised training of a DNN model, there are a variety of tunable parameters that are specific to the given architecture and training method (so called *hyperparameters*), like the learning rate (which controls the extent of update that the weights receive during backpropagation), the moments of the stochastic gradient descent optimizer (like in the popular Adam algorithm [36]) or some weight parameters in a multi-component loss function [37]. However, one of the most crucial and necessary element of the training is undoubtedly a good quality training dataset.

### 3.1. Dataset generation

As discussed in Section 2.2, in our case it is possible to create a training and validation database by randomly generating plasma parameter sets on physically relevant parameter intervals and calculating the corresponding simulated schlieren signals. The datasets contain the computed signals  $S$  sampled at discrete points, paired with the set of generating plasma parameters  $C$ . Signal sampling is defined by the 1024 pixel wide, 13  $\mu\text{m}$  pixelsize detector of the camera used in the measurements. For each  $C$ ,  $y_0$  and  $t_0$  were randomly chosen, distributed uniformly on a given interval.  $P_{max}$  was chosen such that its  $1/n$ th power was a uniform random number on the interval  $[P_0, 1]$ :

$$P_{max} \in [P_0^n, 1]^{1/n} \quad (4)$$

This method skews the probability distribution of  $P_{max}$  to favor values close to 1 somewhat, the exact amount depending on the positive integer generating parameter  $n$ . Note that  $n = 1$  corresponds to uniformly distributed  $P_{max}$  values on the  $[P_0, 1]$  interval. For  $r_0$ , we enforced the following relationship:

$$r_0 < \frac{Q}{(1 - P_{max})^2} \quad (5)$$

with  $Q = 0.25 \mu\text{m}$ . This constraint guarantees  $r_0$  to have substantial values only when  $P_{max}$  is close to 1.

In order to perform the training process with physically realistic data, the signal was slightly smeared with additive and multiplicative Gaussian noises. The noise levels relevant for the experimental setup were deduced from test measurements. We then filtered the signals to reject samples whose signal-to-noise ratio was judged too small to evaluate reliably. First, we dropped samples whose maximum amplitude was less than 5.0 units (a value deduced from the level of background noise). Such samples resemble only noise, and no peaks or interference patterns can be reliably extracted from them. Then, we dropped samples with the absolute value of their mean less than 1.0. This filtering of the data greatly improves the performance of the network, while it does not introduce any artificial, unwanted bias. Note that the filtering also affects the statistical distribution of generating parameters in the final dataset created. Parameter sets that yield a plasma that is “undetectable” at the given noise level – e.g. because the ionization probability is too small, or the plasma is shifted too far out of the probe beam to be detected – are excluded.

In machine learning applications, it is typical to standardize the data with some pre-processing method, which in general leads to improved numerical stability and faster learning. As we have full control over the generation of the simulated datasets, the followings have been considered:

1. The input  $S$  is a vector of 1024 elements, representing the detector image lineout. Since the signal amplitude is sensitive to the degree of ionization, we did not apply any scaling to the input.
2. In order to improve the learning process and reduce numerical instabilities, the plasma parameters have been multiplied with constant factors to scale them to a comparable numerical range:

$$\tilde{P}_{max} = F_P \cdot P_{max}, \quad (6a)$$

$$\tilde{y}_0 = F_Y \cdot y_0, \quad (6b)$$

$$\tilde{t}_0 = F_T \cdot t_0, \quad (6c)$$

$$\tilde{r}_0 = F_R \cdot r_0. \quad (6d)$$

We chose the scaling factors as:  $F_P = 1$ ,  $F_Y = F_T = 10^2$ , and  $F_R = 10^3$ . Apart from improving the convergence during training, scaling also allows us to rank parameters in terms of importance. Parameters scaled to greater numerical range will carry greater weight in the loss function, so the relative error of their estimation can be expected to be smaller after training. Our choice above was motivated by the fact that the plasma core radius  $r_0$ , originally approximately the same size as the other two geometric parameters, is the most important of the three for practical reasons. In what follows, we denote the 4-element vector of scaled plasma parameters by  $\tilde{p} = (\tilde{P}_{max}, \tilde{y}_0, \tilde{r}_0, \tilde{t}_0)$  and specific components by  $p_i$ .

For training, validation and testing purposes, we generated datasets with different distributions given by  $n = 1, 3, 5, 7$  and 10 and mixed them. Because  $r_0 > 0$  values are favored only when  $P_{max} \sim 1$ , it also became necessary to generate an additional subset, that allowed configurations only with  $r_0 > 0.5$  mm. This improves the statistics of the training dataset and thus also the final precision of the predicted  $r_0$  parameters. Beside the realistic, noisy sample signals, the “pure” (i.e. noiseless) signals were also retained during the generation process for testing and uncertainty estimation purposes. The overall distributions of the parameters in the training set are plotted on Fig. 4. The  $r_0 > 0.5$  mm subset causes a visible discontinuity in the distribution for  $r_0$ . The overall unified dataset contained a total amount of  $N = 1.1 \cdot 10^6$  samples, and has been split into training, validation and test datasets with ratios of 0.65 : 0.15 : 0.20 respectively.

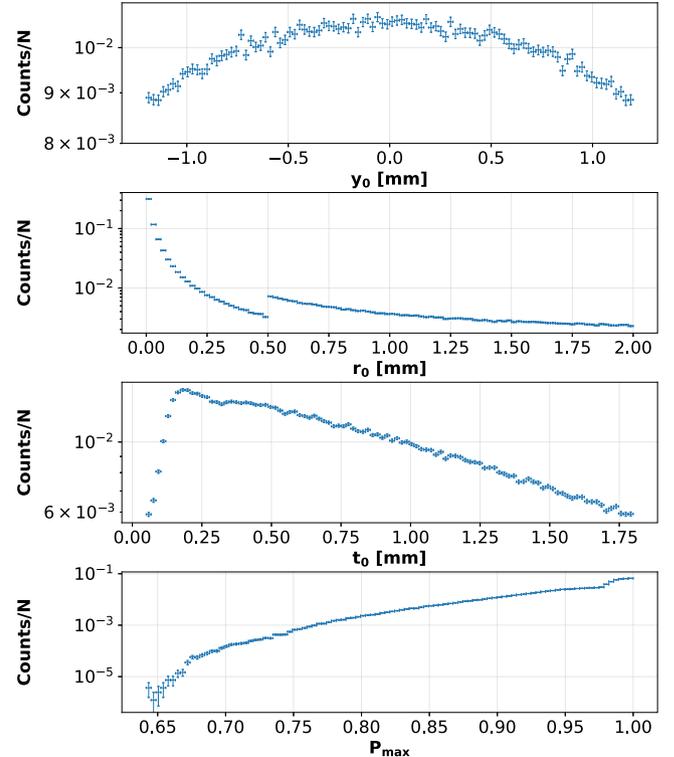


Fig. 4. Plasma parameter distributions in the training dataset in counts/N where  $N$  is the sample size. The discontinuity in the distribution of  $r_0$  is due to the separate  $r_0 > 0.5$  mm sample subset (see main text).

It is instructive to visualize the correlations between the parameters  $p_i$  in the dataset, defined as the Pearson correlation coefficients of the training data:

$$r_{ij} = \frac{\sum_{k=1}^N (p_{k,i} - \bar{p}_i)(p_{k,j} - \bar{p}_j)}{\sqrt{\sum_{k=1}^N (p_{k,i} - \bar{p}_i)^2} \sqrt{\sum_{k=1}^N (p_{k,j} - \bar{p}_j)^2}} \quad (7)$$

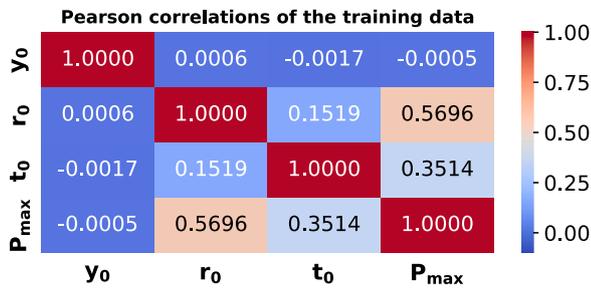


Fig. 5. The Pearson correlations in the training data.

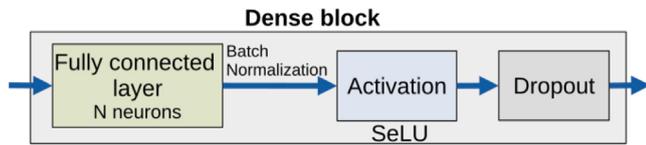


Fig. 6. A basic building block of the applied networks.

The correlations are shown in Fig. 5, which shows that  $r_0$  is strongly correlated with  $P_{max}$ , while the transition region width,  $t_0$  is only slightly correlated with  $P_{max}$  and  $r_0$ . The  $y_0$  location is basically uncorrelated with the other parameters.

### 3.2. Network implementations

Our aim is to develop a robust framework that is able to infer the plasma parameters from the schlieren signals with high accuracy. To achieve this, a customizable DNN framework has been implemented in Python, using Keras v2.7.0 with Tensorflow v.2.7.0 backend [38,39]. The basic building block of the framework, called a *Dense block* is sketched on Fig. 6, which consists of a fully connected layer, followed by a batch normalization, a scaled exponential linear unit activation (SELU [40]) and a dropout layer with fixed dropout rate of 0.1 [41–43].

Our neural networks, which have three distinct parts and several configurable parameters are composed of such building blocks as it is depicted on Fig. 7. In the first part, (marked with light pink background) one or more feature extraction blocks process the input data, consisting of  $D_i$  parallel Dense blocks, with  $N_i$  neurons in each block. Subsequently, the output of the  $D_i$  Dense blocks are merged, which is then followed by a SoftMax activation. Next, after  $L - 1$  consecutive feature extraction blocks, a concatenation layer merges the parallel blocks in the  $L$ th layer (as an analogy for flattening, marked with light blue background). In the third part,  $H_L$  optional hidden layers follow, with  $N_{HL}$  neurons in the given layer. We note, that though this specific parametrization provides good flexibility, we have found that  $N_{HL} = 0$  eventually leads to satisfactory results, therefore in the this study we did not consider the  $N_{HL} > 0$  variations. Finally, the last fully connected layer represents the four plasma parameters. We

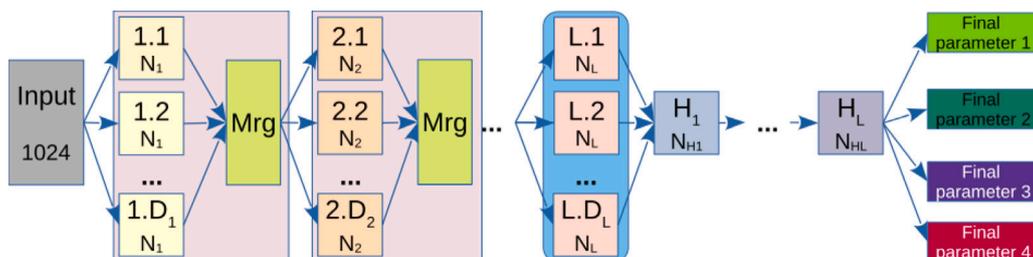


Fig. 7. The general structure of the implemented neural networks.

tried different kinds of activation functions for the last layer in our networks, and found that the best choice is the linear activation, which is a common choice for regression tasks. Since in this way the possible values of the predicted parameters were not limited neither from below nor from above, the network was forced to learn the physically relevant value ranges of the plasma parameters.

### 3.3. Network training

We investigated several different architecture implementations, in what follows, we compare the best three variants. Models codenamed *FE1* and *FE2* included only feature extraction parts without additional hidden layers, while the model named *FEO* is a special case of the network, where no explicit feature extraction blocks have been applied, therefore it is considered as a traditional multilayer perceptron. See Table 1 for the specific values of the configurable parameters in each model architecture. During training, the mean absolute error of the vector of scaled parameters was utilized as the loss function:

$$\mathcal{L}(\vec{p}_{pred}, \vec{p}_{true}) = \frac{1}{M} \sum_{k=1}^M |\vec{p}_{k,pred} - \vec{p}_{k,true}|, \quad (8)$$

where  $M$  is the size of the training dataset, and the optimization was performed with the Adam algorithm with default settings [36]. The initial learning rate has been slowly decreased with a linear decay.

The training process was monitored with the mean squared error and Log-Cosh errors as additional metrics [44], and 15% of the full dataset was used for validating each epoch (validating set). By investigating the convergence of the validation loss and the other metrics, we found that it is sufficient to train the networks for 20,000 epochs to get the best achievable quality of predictions without overfitting the training data. The final loss value for the listed variants are shown in Table 1.

Considering the initial learning rate, we tried different values in the  $[10^{-4}, 10^{-2}]$  ranges. Taking lower initial values results in clearer and faster convergence in the measurement of training quality, however, with too small values, the correlations become unrealistic. Taking too high initial values results in high fluctuations and very slow convergence in the measurement of training quality with bad correlations. We found that the best value for the initial learning rate is  $5 \cdot 10^{-3}$ .

The training, evaluating and testing were performed on the GPU clusters of the Wigner Scientific Computing Laboratory (WSCLAB).

Table 1

The configurations for the applied architectures and the final value of the loss function at the end of training.

	FE2	FE1	FEO
Feature extraction blocks	2	1	0
$D_L$	16, 8, 4	16, 4	1, 1, 1, 1
$N_L$	512, 256, 128	256, 64	1024, 512, 256, 128, 64
Trainable parameters	9.6M	4.3M	1.75M
Final loss	$4.02 \cdot 10^{-3}$	$6.67 \cdot 10^{-3}$	$9.54 \cdot 10^{-3}$

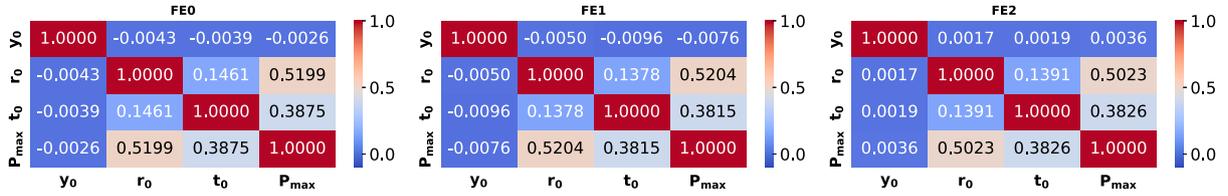


Fig. 8. The Pearson correlations for the parameter prediction on the test data by the various models.

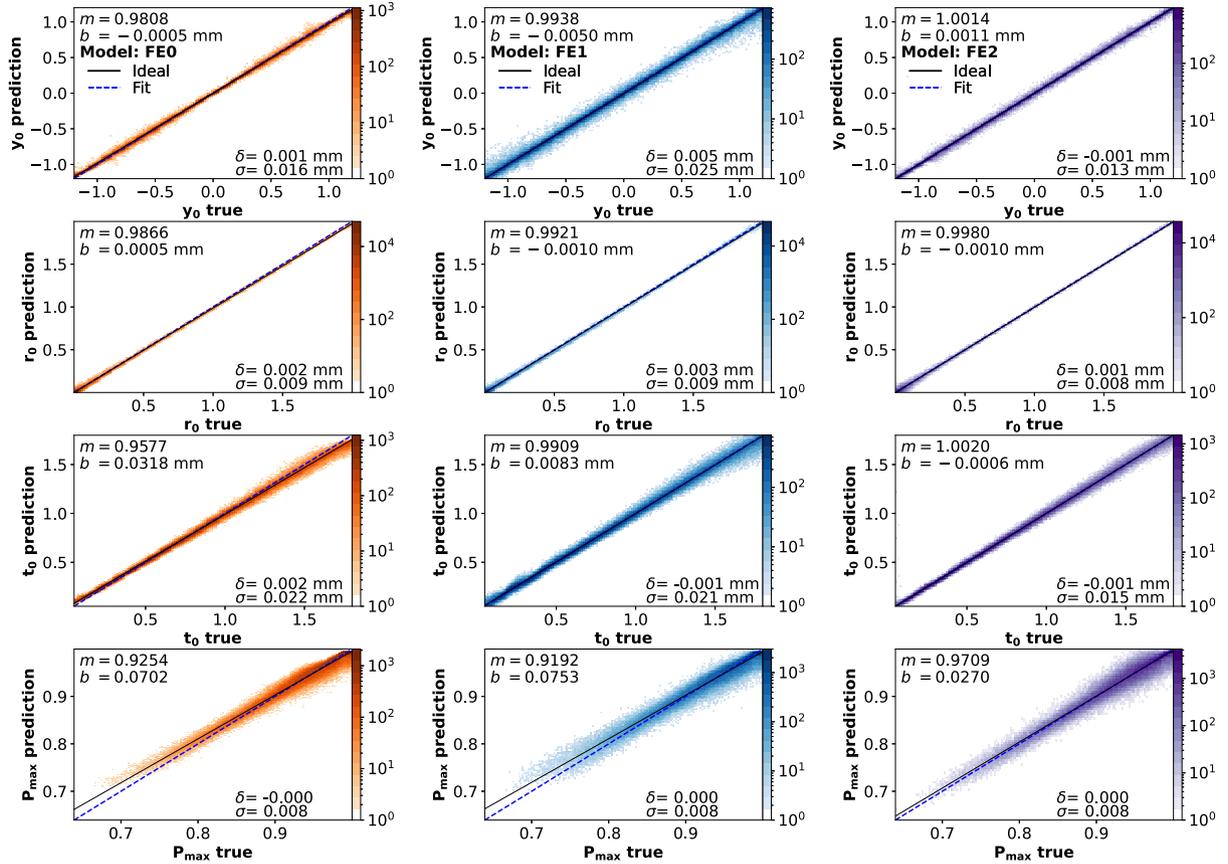


Fig. 9. 2D histogram plots of predicted plasma parameters vs. true values for the models. Each column of subfigures corresponds to parameter estimates of a single network architecture. Note that color is coded with a logarithmic scale for outlier estimates to be visible. The ideal line of prediction equals the true value is indicated with a solid line, while the best fit line to the predicted data is indicated with a dashed line. The slopes  $m$  and intercepts  $b$  of the best fit lines (see Eq. (10)), as well as the values of mean absolute error  $\sigma$  and mean error  $\delta$  (Eqs. (9)) are indicated on each subfigure. Note, that in most of the cases the fit (dashed blue line) is overlapping with the ideal case (solid black line). The difference is discernible only for the  $P_{max}$  parameter (lowest row of figures).

## 4. Results

### 4.1. Plasma parameter estimation

In order to test the performance of the trained networks, plasma parameters have been predicted for signals in the test dataset (20% of the full reference dataset described in the previous section). Fig. 8 presents the Pearson correlation coefficients of the predicted parameters for the different networks. While the goal of the trained neural networks is to be capable of predicting individual parameter sets, it is also useful to investigate the network's perceptions on the statistical correlations. In terms of usability, the models should be able to reproduce the test data both at these microscopic and macroscopic levels.

The structure of the extracted correlations by the different architectures are very similar to those we presented for the training data. Comparing the results with the reference plotted in Fig. 5, we can see that the  $y_0$  parameter is basically uncorrelated for the parameter estimations of all network variants. The correlations  $r_{P_{max}, y_0}$  for the

network estimates are about 5% lower than the reference, while the correlations  $r_{P_{max}, r_0}$  are about 3% higher.

Note that the uncorrelatedness of the  $y_0$  parameter means that the other plasma parameters have a translation invariance with respect to  $y_0$ . This agrees with the fact that the properties of the plasma channel are independent of the exact position of the ionizing laser pulse if the vapor has a uniform density distribution. The signal itself is not translationally invariant however, due to the varying probe beam intensity, but the networks learned to ignore this variation in the signal. It is also important to emphasize that we did not enforce any artificial constraint on the plasma parameter predictions (e.g.  $P_{max} \in [0, 1]$ ). All the presented network variants recognized the physical range of the parameters along with the correct correlations between  $r_0$  and  $P_{max}$ .

As the most important measure of network prediction accuracy, we compare the network parameter predictions  $C_{pred}$  to their true values  $C_{true}$  over the test dataset. The predicted parameters are plotted versus the true values for all the investigated networks in Fig. 9. To quantify the accuracy of the predictions, we calculated the mean absolute error  $\sigma$  for the parameter predictions, as well as the mean error  $\delta$  for all

parameters  $X \in C$ :

$$\begin{aligned}\sigma_X &= \frac{1}{M} \sum_k |X_{k,pred} - X_{k,true}|, \\ \delta_X &= \frac{1}{M} \sum_k (X_{k,pred} - X_{k,true}).\end{aligned}\quad (9)$$

Furthermore, a linear fit was also performed to the ensemble of parameter predictions (marked as a dashed line on the plots):

$$X_{pred} = m_X \cdot X_{true} + b_X \quad (10)$$

Clearly, the deviation of the slope  $m_X$  from 1, as well as the non-zero value of intercept  $b_X$  and the mean error  $\delta_X$  all characterize systematic errors in the parameter prediction.

As is visible on Fig. 9, all of the network variants produced a very good prediction of the plasma parameters. The mean absolute errors for  $y_0$ ,  $r_0$  and  $t_0$  are around (and sometimes even below) the camera pixel size of 13  $\mu\text{m}$ , and never above twice that. Mean errors are a mere fraction of the pixel size. The predictions for  $r_0$  are significantly better than for  $y_0$  and  $t_0$  for all models, a trait controlled by the scaling Eqs. (6). There are some significant differences between models, however. The *FEO* variant produced the largest deviation for all parameters, but it is worthwhile to note that among the investigated architectures this contained the least trainable parameters — only 1.75M, which is less than the half of the second network. On the other hand, we have found that just increasing the number of trainable parameters is not enough for significant improvement, the feature extraction blocks are also necessary.

#### 4.2. Predicted signal accuracy

In order to further estimate the reliability of the models, and to quantify the error of the predictions, we re-calculated a *predicted signal* from the predicted parameters and compared it to the original *pure* (i.e. noiseless) signal. We found that in practically every case, the predicted signal followed very closely the original pure signal.

However, for small amplitude, (i.e. small signal-to-noise ratio) cases, it can happen that the predicted signal is offset slightly from the original one, while still being very similar in shape (see Fig. 10 as an example). Therefore to quantify signal accuracy, a simple integral over the absolute difference between predicted and original signals proved unsuitable and the following procedure was followed.

Let  $S_{pred}$  and  $S_{true}$  denote the predicted and the originally simulated schlieren signals without any additional noise, respectively, and we define the normalized signals as:

$$\tilde{S}(y) = \frac{S(y)}{\int S(y)dy}. \quad (11)$$

We can write the offset-dependent error between the normalized signals as:

$$\varepsilon(y_{off}) = \int |\tilde{S}_{true}(y) - \tilde{S}_{pred}(y - y_{off})| dy \quad (12)$$

and we expect that this has a clear minimum for some small offset value  $y_{err}$ , that we can call the *phase error*, while the *amplitude error* is defined as the minimum of  $\varepsilon$ :

$$A_{err} = \min \varepsilon(y_{off}) = \varepsilon(y_{err}). \quad (13)$$

Clearly, for a discretized signal  $y_{err}$  is some small integer value.

To characterize the quality of network predictions, we calculated  $A_{err}$ ,  $y_{err}$  for all predicted signals of the test dataset and plotted their probability distributions on Fig. 11. For every model, the error of the normalized amplitude has a mean value of around 5%. Considering the phase errors, the plots suggest that the predictions carry a very small amount of phase error, well centered around 0. More precisely said, for all three models around half of the predictions yield  $y_{err} = 0$ , while around 90% of the predictions have  $|y_{err}| \leq 2$ . Only a negligible fraction of the predictions carry a phase error greater than 10 pixels.

#### 5. Robustness of the parameter predictions

Trained DNNs can truly be useful for the evaluation of real measurements only if the parameter predictions are not too sensitive to

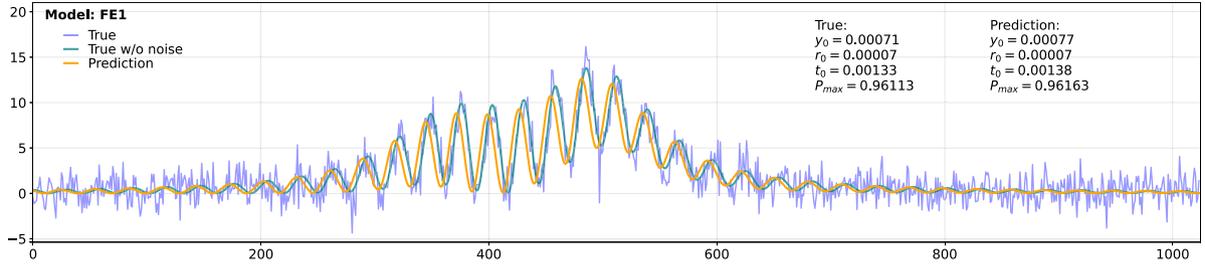


Fig. 10. Example for the true, noisy and reconstructed schlieren signals, and the predicted plasma parameters.

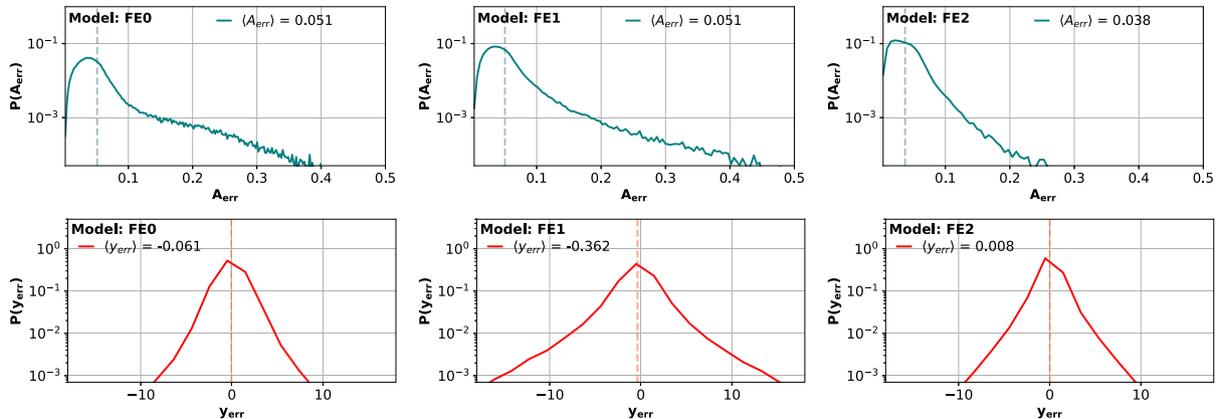


Fig. 11. The distributions of the normalized amplitude and phase errors (see text for details).

slight changes of the experimental circumstances. In the followings, the robustness of the trained networks are investigated.

### 5.1. Variation of experimental parameters

Parameters of the experimental setup are always known with finite accuracy. Some may actually fluctuate or drift somewhat during the course of a measurement. It is essential to verify that DNNs that have been trained on samples calculated with a given set of experimental parameters, can estimate plasma parameters sufficiently accurately even for signals coming from slightly different experimental circumstances.

In the setup considered here, the vapor density and the probe laser intensity are the most important parameters that may vary slightly. Vapor density in the AWAKE vapor source can be held constant to better than 0.2% accuracy [24], while the probe laser power may change possibly by a few percents. To explore prediction robustness with respect to changes that surpass these uncertainties, we generated sets of test samples with vapor density decreased/increased by 2%, 5% and 10%, respectively, as well as sets where the probe laser power was varied by the same amount. Plasma parameter selection, signal generation and filtering followed the same procedure as for the original dataset. Parameters were then predicted by the networks and  $\sigma$  and  $\delta$  were calculated in each case as in Section 4.

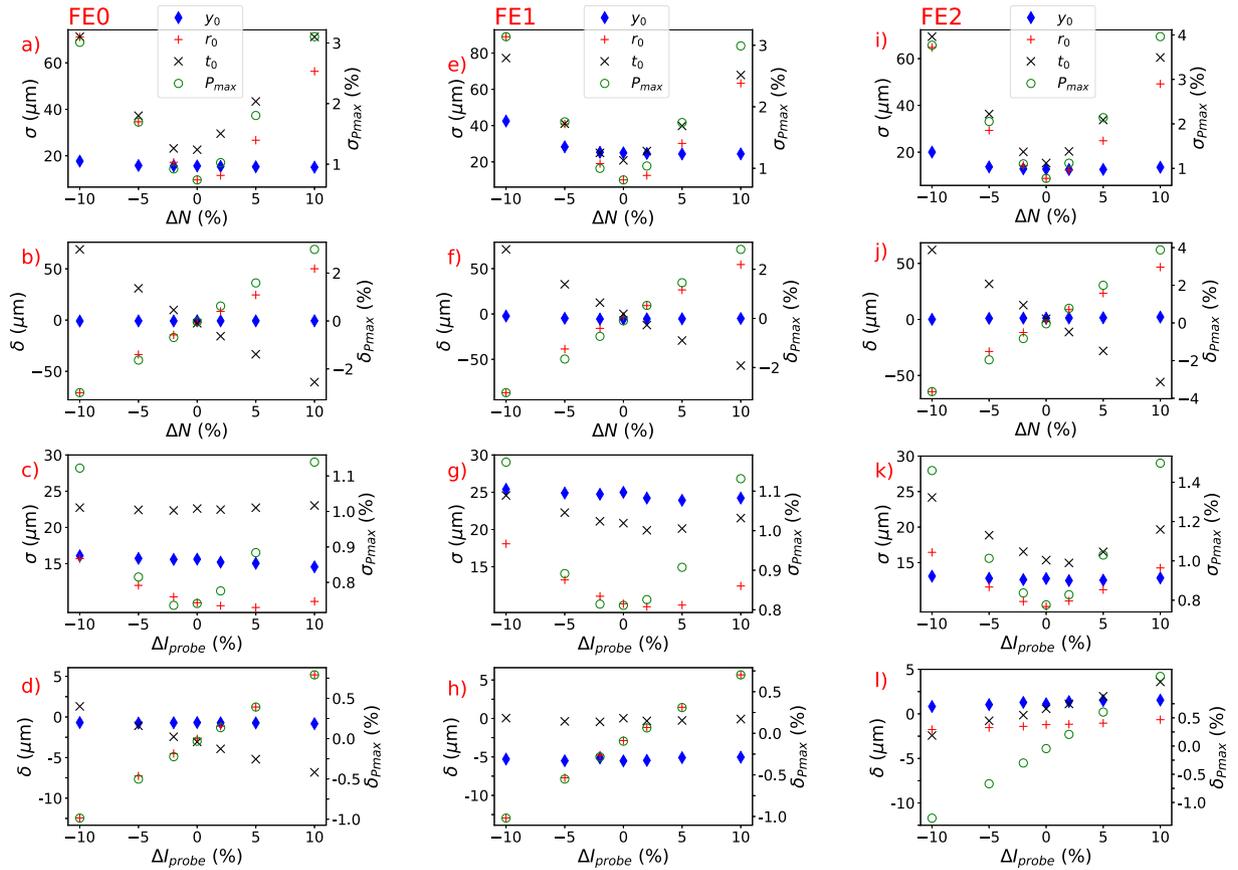
Fig. 12 show the increase of parameter prediction error  $\sigma$  and  $\delta$  as the vapor density and probe laser intensity deviates from the values used during training. Several interesting conclusions can be drawn from the figures. First, changing vapor density causes a much greater error in parameter prediction than changing the probe amplitude. This is especially evident for the systematic error  $\delta$ . The reason is that changing the probe amplitude changes only the signal amplitude, whereas

changing vapor density not only causes the signal amplitude to change (because of absorption), but much more importantly, the signal peak locations and relative heights (because the phaseshift the probe beam acquires upon transit depends on the density). Second, an increase (decrease) in vapor density causes an overestimation (underestimation) of the maximum ionization probability. This is to be expected, as both changes have similar effect on the transmitted probe beam phase and amplitude. Third, an increase (decrease) in vapor density leads to a systematic overestimation (underestimation) of  $r_0$  and a corresponding underestimation (overestimation) of  $t_0$ . Probe laser power changes have a similar effect in general, but on smaller scale.

It is also evident that for a parameter change of 2%, plasma parameter prediction is not affected greatly for any of the networks, i.e. they seem sufficiently robust to be used. Even a 5% change in the vapor density causes an error of about 40  $\mu\text{m}$  in the plasma parameter estimations, a value that is still well acceptable. Systematic errors for  $r_0$ ,  $t_0$  and  $P_{max}$  do become the same order of magnitude as the random errors, however. It is notable, that the network containing the least trainable parameters (FE0) does not fare in any way worse in these tests of robustness than any of its more elaborate counterparts. While some of its predictions on the ‘pure’ problem are somewhat less accurate, the predictions on the test samples with distorted experimental parameters are comparably precise.

### 5.2. Variation of the plasma profile

One further question is whether plasma parameter predictions remain sensible for plasma distribution profiles that are slightly different than the functional dependence of Eq. (1). Indeed, the general nature of a plasma channel that has a core with near constant (saturated)



**Fig. 12.** Robustness of network parameter predictions. Each column of plots corresponds to four parameters of a specific model: FE0: (a)–(d), FE1: (e)–(h), FE2: (i)–(l). Row 1 [(a), (e), (i)]: mean absolute error  $\sigma$  as a function of  $\Delta\mathcal{N}$ . Row 2 [(b), (f), (j)]: mean error  $\delta$  as a function of  $\Delta\mathcal{N}$ . Row 3 [(c), (g), (k)]:  $\sigma$  as a function of  $\Delta I_{probe}$ . Row 4 [(d), (h), (l)]:  $\delta$  as a function of  $\Delta I_{probe}$ . Errors for  $y_0$ ,  $r_0$  and  $t_0$  are shown on the left axis, while the error for  $P_{max}$  is shown on the right axis in each subfigure.

**Table 2**

Increase of prediction error for a sigmoid-type plasma profile.  $\sigma, \delta$  are parameter prediction errors on a Gaussian plasma profile,  $\sigma', \delta'$  are values on a sigmoid profile.

FE0	$\sigma \rightarrow \sigma'$	$\delta \rightarrow \delta'$
$y_0$ ( $\mu\text{m}$ )	15.6 $\rightarrow$ 15.2	-0.7 $\rightarrow$ -0.5
$r_0$ ( $\mu\text{m}$ )	9.6 $\rightarrow$ 31.7	-2.7 $\rightarrow$ 27.5
$t_0$ ( $\mu\text{m}$ )	22.6 $\rightarrow$ 34.1	-3.1 $\rightarrow$ 30.6
$P_{max}$ (%)	0.74 $\rightarrow$ 0.99	-0.03 $\rightarrow$ 0.21
FE1	$\sigma \rightarrow \sigma'$	$\delta \rightarrow \delta'$
$y_0$ ( $\mu\text{m}$ )	25.0 $\rightarrow$ 31.5	-5.5 $\rightarrow$ -5.9
$r_0$ ( $\mu\text{m}$ )	10.0 $\rightarrow$ 27.7	-2.9 $\rightarrow$ 18.0
$t_0$ ( $\mu\text{m}$ )	20.9 $\rightarrow$ 37.8	0. $\rightarrow$ 25.5
$P_{max}$ (%)	0.81 $\rightarrow$ 1.02	-0.1 $\rightarrow$ 0.24
FE2	$\sigma \rightarrow \sigma'$	$\delta \rightarrow \delta'$
$y_0$ ( $\mu\text{m}$ )	12.8 $\rightarrow$ 15.6	1.1 $\rightarrow$ 0.7
$r_0$ ( $\mu\text{m}$ )	8.8 $\rightarrow$ 39.5	-1.2 $\rightarrow$ 33.2
$t_0$ ( $\mu\text{m}$ )	15.3 $\rightarrow$ 46.7	0.6 $\rightarrow$ 42.1
$P_{max}$ (%)	0.78 $\rightarrow$ 1.26	-0.05 $\rightarrow$ 0.12

ionization and a transition region to unionized vapor surrounding it can be described with many qualitatively similar functions. To investigate this question, we generated a test dataset where the radial dependence of the plasma was described by a sigmoid-type function:

$$\mathcal{N}_{plasma} = \mathcal{N}_0 P_{max} \frac{\exp(-[r - r'_0]/t')}{1 + \exp(-[r - r'_0]/t')} \quad (14)$$

The test dataset was generated by choosing a set of plasma parameters according to the procedure of Section 3.1, computing the radial plasma distribution and then calculating the best fit sigmoid distribution Eq. (14). A plasma channel with this sigmoid distribution was used to calculate the signals that were processed by the networks. We then compared the network predictions to the parameters of the original Gaussian type plasma profile. Table 2 shows the increase of the prediction error when the networks are used to predict parameters from these signals. Systematic errors are now greater, they again become the same order of magnitude as the random errors. Nevertheless, DNNs still provide a fairly accurate estimation for all plasma parameters in question, the error introduced is no greater than that introduced by changing the vapor density by about 5%.

### 5.3. Further comments

The purpose of this section is to demonstrate that network predictions are sufficiently accurate even if they are used to process signals that come from circumstances somewhat different from that of the training signals. This robustness is very important for the networks to be useful in practice. Clearly, the prediction errors shown in this section provide a much more realistic estimate of practical network accuracy, than those calculated in Section 4. We note, that if some experimental parameter can vary a lot, it is possible to increase prediction robustness with respect to this variation by incorporating this parameter as a variable one during training data generation. This of course means that probably a larger training dataset is required, with an increased computational load during training.

The experimentally observable variation of the plasma channel properties (and hence the schlieren images) has several other reasons. The dominant source is the fact that the powerful ionizing laser pulses have a significant shot-to-shot energy and pulse shape fluctuations. The propagation through the vapor is, in itself, also a strongly nonlinear process leading to stochastic fluctuations of the result by the downstream end of the vapor source. These variations are not necessarily a problem however, in fact it is these properties we want to explore using the machine learning technique. On the other hand, one can observe [28] from the transmitted part of the ionizing laser pulse that the pulse may possess some ellipticity, so the plasma channel itself

cannot be expected to be axially perfectly symmetric. In this case the single  $r_0/t_0$  parameter values supplied by the networks can be thought of as some representative values. Furthermore, under some conditions the ionizing laser pulse may break up into several filaments, leading to a plasma with multiple cores. Clearly in this case the network estimates cannot even be expected to be representative. It is therefore important to evaluate the energy distribution of the transmitted ionizing pulse parallel to evaluating the schlieren images to be certain that the plasma parameters supplied by the networks are credible.

## 6. Summary

In this paper we have demonstrated that machine learning techniques can be useful in determining the geometrical properties of a plasma channel in atomic vapor. The plasma is monitored via schlieren imaging, which is a sensitive method that allows us to obtain a signal even in the very low-density vapor used in proton-driven plasma wakefield accelerators. DNNs of several different architectures have been trained on a large set of noisy sample signals to predict the plasma channel location, the radius of the plasma channel core, the width of the transition region between the core and the unionized vapor and the ionization fraction of the atoms at the plasma channel center. The trained DNNs provide a robust and efficient framework to gain high-accuracy, quantitative predictions for the plasma parameters from noisy schlieren signals. Without putting any constraint on the network output, they recognized the physical ranges of the parameters describing the plasma channel. We have also shown, that the network predictions remain accurate enough even if the networks are used to process signals obtained with slightly different experimental parameters or slightly different radial profile than those in the training dataset. This property is crucial for practical application. In particular, we varied the density of the atomic vapor and the intensity of the probe laser, because they can be held constant only to a certain accuracy during the course of a several hour experimental run. We conclude that this method of quantitative parameter estimation can be useful in a number of settings, and can certainly prove efficient for AWAKE, the proton-driven plasma wakefield experiment at CERN.

### CRedit authorship contribution statement

**Gábor Bíró:** Conceptualization, Methodology, Software, Investigation, Visualization, Writing – original draft. **Mihály Pocsai:** Software, Investigation, Validation, Visualization, Writing – original draft. **Imre F. Barna:** Writing – review & editing. **Gergely G. Barnaföldi:** Funding acquisition, Writing – review & editing. **Joshua T. Moody:** Conceptualization. **Gábor Demeter:** Conceptualization, Formal analysis, Methodology, Software, Investigation, Writing – original draft.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Data availability

Data will be made available on request.

### Acknowledgment

The authors are grateful to Gábor Papp for the useful discussions.

## Funding

The research was supported by the Hungarian National Research, Development and Innovation Office (NKFIH) under the contract numbers OTKA K135515, K123815 and NKFIH 2019-2.1.6-NEMZ\_KI-2019-00011, 2020-2.1.1-ED-2021-00179, 2019-2.1.6-NEMZ\_KI-2019-00004, MEC\_R-140947, and the Wigner Scientific Computing Laboratory (WS-CLAB, the former Wigner GPU Laboratory), Hungary. On behalf of Project Awakelaser we are grateful for the usage of ELKH Cloud [45] which helped us achieve the results published in this paper. Author G. B. was supported by the European Union project RRF-2.3.1-21-2022-00004 within the framework of the Artificial Intelligence National Laboratory.

## References

- [1] S. Narayanan, E. Chaniotakis, C. Antoniou, *Transp. Res. C* (ISSN: 0968-090X) 111 (2020) 255–293, URL <https://www.sciencedirect.com/science/article/pii/S0968090X19303493>.
- [2] V.K. Verma, S. Verma, *Materials Today: Proc.* (ISSN: 2214-7853) 57 (2022) 2144–2147, URL <https://www.sciencedirect.com/science/article/pii/S2214785321077749>. International Conference on Innovation and Application in Science and Technology.
- [3] L. Deng, X. Li, *IEEE Trans. Audio Speech Lang. Proces.* 21 (5) (2013) 1060–1089.
- [4] M. Feickert, B. Nachman, 2021, [arXiv:2102.02770](https://arxiv.org/abs/2102.02770).
- [5] G. Bíró, B. Tankó-Bartalis, G.G. Barnaföldi, 2021, [arXiv:2111.15655](https://arxiv.org/abs/2111.15655).
- [6] F.A. Matos, D.R. Ferreira, P.J. Carvalho, *Fusion Eng. Des.* (ISSN: 0920-3796) 114 (2017) 18–25, URL <https://www.sciencedirect.com/science/article/pii/S0920379616306883>.
- [7] J. Mlynar, T. Craciunescu, D.R. Ferreira, P. Carvalho, O. Ficker, O. Grover, M. Imrisek, J. Svoboda, J. contributors, *J. Fusion Energy* 38 (2019) 458–466.
- [8] D.J. Clayton, K. Tritz, D. Stutman, R.E. Bell, A. Diallo, B.P. LeBlanc, M. Podesta, *Plasma Phys. Control. Fusion* 55 (2013) 095015.
- [9] G. Demeter, *Rev. Sci. Instrum.* 68 (3) (1997) 1438–1443, <http://dx.doi.org/10.1063/1.1147630>.
- [10] G.S. Settles, *Schlieren and Shadowgraph Techniques - Visualizing Phenomena in Transparent Media*, Springer-Verlag, Berlin, Heidelberg, ISBN: 978-3-642-63034-7, 2001.
- [11] E. Traldi, M. Boselli, E. Simoncelli, A. Stancampiano, M. Gherardi, V. Colombo, G. Settles, *EPJ Tech. Instrum.* 5 (2018) 4.
- [12] C.E. Clayton, K.-C. Tzeng, D. Gordon, P. Muggli, W.B. Mori, C. Joshi, V. Malka, Z. Najmudin, A. Modena, D. Neely, A.E. Dangor, *Phys. Rev. Lett.* 81 (1998) 100–103, URL <https://link.aps.org/doi/10.1103/PhysRevLett.81.100>.
- [13] O. Iwase, W. Süß, D.H.H. Hoffmann, M. Roth, C. Stöckl, M. Geissel, W. Seelig, R. Bock, *Phys. Scr.* 58 (6) (1998) 634–635, URL <https://doi.org/10.1088/0953-4075/41/15/155004>.
- [14] H. Honda, K. Katsura, E. Takahashi, K. Kondo, *Appl. Phys. B* 70 (2000) 395–398.
- [15] F. Veloso, H. Chuaqui, R. Aliaga-Rossel, M. Favre, I.H. Mitchell, E. Wyndham, *Rev. Sci. Instrum.* 77 (6) (2006) 063506, URL <https://doi.org/10.1088/0953-4075/41/15/155004>.
- [16] D. Batani, J. Santos, P. Forestier-Colleoni, D. Mancelli, M. Ehret, J. Trela, A. Morace, K. Jakubowska, L. Antonelli, D. del Sorbo, M. Manclossi, M. Veltcheva, *J. Fusion Energy* 38 (2019) 299–314.
- [17] I.A. Znamenskaya, I.A. Doroshchenko, *J. Flow Vis. Image Process.* (ISSN: 1065-3090) 28 (4) (2021) 1–26.
- [18] S. Cai, Z. Wang, F. Fuest, Y.J. Jeon, C. Gray, G.E. Karniadakis, *J. Fluid Mech.* 915 (2021) A102.
- [19] B.N. Ubald, P. Seshadri, A. Duncan, Density estimation from schlieren images through machine learning, 2022, URL <https://arxiv.org/abs/2201.05233>.
- [20] C. Joshi, T. Katsouleas, *Phys. Today* 56 (6) (2003) 47.
- [21] W. Leemans, E. Esarey, *Phys. Today* 62 (3) (2009) 44.
- [22] E. Gschwendtner, et al., *Nucl. Instrum. Methods Phys. Res. A* (ISSN: 0168-9002) 829 (2016) 76–82, URL <http://www.sciencedirect.com/science/article/pii/S0168900216001881>. 2nd European Advanced Accelerator Concepts Workshop - EAAC 2015.
- [23] E. Adli, A. Ahuja, O. Apsimon, R. Apsimon, A.-M. Bachmann, D. Barrientos, F. Batsch, J. Bauche, V.B. Olsen, M. Bernardini, et al., *Nature* 561 (7723) (2018) 363.
- [24] E. Öz, P. Muggli, *Nucl. Instrum. Methods Phys. Res. A* (ISSN: 0168-9002) 740 (2014) 197–202, <http://dx.doi.org/10.1016/j.nima.2013.10.093>, URL <http://www.sciencedirect.com/science/article/pii/S0168900213015003>. Proceedings of the first European Advanced Accelerator Concepts Workshop 2013.
- [25] G. Plyushchev, R. Kersevan, A. Petrenko, P. Muggli, *J. Phys. D: Appl. Phys.* 51 (2) (2017) 025203, URL <https://doi.org/10.1088/1361-6463/aa9dd7>.
- [26] A. Couairon, A. Mysyrowicz, *Phys. Rep.* (ISSN: 0370-1573) 441 (2) (2007) 47–189, URL <http://www.sciencedirect.com/science/article/pii/S037015730700021X>.
- [27] G. Demeter, *Phys. Rev. A* 99 (2019) 063423, URL <https://link.aps.org/doi/10.1103/PhysRevA.99.063423>.
- [28] G. Demeter, J.T. Moody, M.Á. Kedves, B. Ráczkevi, M. Aladi, A.-M. Bachmann, F. Batsch, F. Braunmüller, G.P. Djotyan, V. Fedosseev, F. Friebe, S. Gessner, E. Granados, E. Guran, M. Hüther, V. Lee, M. Martyanov, P. Muggli, E. Öz, H. Panuganti, L. Verra, G.Z. Della Porta, *Phys. Rev. A* 104 (2021) 033506, URL <https://link.aps.org/doi/10.1103/PhysRevA.104.033506>.
- [29] A.-M. Bachmann, M. Martyanov, J. Moody, A. Pukhov, P. Muggli, *Nucl. Instrum. Methods Phys. Res. A* (ISSN: 0168-9002) 909 (2018) 387–390, URL <https://www.sciencedirect.com/science/article/pii/S0168900217314687>. 3rd European Advanced Accelerator Concepts workshop (EAAC2017).
- [30] P. Walch, B. Mahieu, L. Arantchouk, Y.-B. André, A. Mysyrowicz, A. Houard, *Appl. Phys. Lett.* 119 (26) (2021) 264101, URL <https://doi.org/10.1063/5.0077635>.
- [31] E. Adli, A. Ahuja, O. Apsimon, R. Apsimon, A.-M. Bachmann, D. Barrientos, et al., AWAKE Collaboration Collaboration Collaboration, *Phys. Rev. Lett.* 122 (2019) 054802, URL <https://link.aps.org/doi/10.1103/PhysRevLett.122.054802>.
- [32] B.E.A. Saleh, M.C. Teich, *Fundamentals of Photonics*, John Wiley and Sons, Inc., ISBN: 9781119506874, 2019.
- [33] P. Siddons, C.S. Adams, C. Ge, I.G. Hughes, *J. Phys. B: At. Mol. Opt. Phys.* 41 (15) (2008) 155004, URL <https://doi.org/10.1088/0953-4075/41/15/155004>.
- [34] A.J. van Lange, P. van der Straten, D. van Oosten, *J. Phys. B: At. Mol. Opt. Phys.* 53 (12) (2020) 125402, URL <https://doi.org/10.1088/1361-6455/ab7fc2>.
- [35] S. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, third ed., Prentice Hall, 2010.
- [36] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2017, [arXiv: 1412.6980](https://arxiv.org/abs/1412.6980).
- [37] A. Dosovitskiy, J. Djolonga, International Conference on Learning Representations, 2020, URL <https://openreview.net/forum?id=HyxY6JHKwr>.
- [38] F. Chollet, et al., Keras, 2015, <https://keras.io>.
- [39] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous distributed systems, 2016, [arXiv: 1603.04467](https://arxiv.org/abs/1603.04467).
- [40] G. Klambauer, T. Unterthiner, A. Mayr, S. Hochreiter, 2017, URL <https://arxiv.org/abs/1706.02515>.
- [41] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, in: Springer Series in Statistics, Springer New York Inc., New York, NY, USA, 2001.
- [42] A. Meyer-Baese, V. Schmid, in: A. Meyer-Baese, V. Schmid (Eds.), *Pattern Recognition and Signal Analysis in Medical Imaging* (Second Edition), second ed., Academic Press, Oxford, ISBN: 978-0-12-409545-8, 2014, pp. 197–243, URL <https://www.sciencedirect.com/science/article/pii/B9780124095458000078>.
- [43] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015, URL <https://arxiv.org/abs/1502.03167>.
- [44] T. Moshagen, N.A. Adde, A.N. Rajgopal, Finding hidden-feature depending laws inside a data set and classifying it using neural network, 2021, URL <https://arxiv.org/abs/2101.10427>.
- [45] M. Héder, E. Rigó, D. Medgyesi, R. Lovas, S. Tenczer, F. Török, A. Farkas, M. Emödi, J. Kadlecik, G. Mezö, Á. Pintér, P. Kacsuk, *Inform. Társad.* 22 (2) (2022) 128, URL <https://doi.org/10.22503/infars.xxii.2022.2.8>.