

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

Gábor Bíró,¹ Mihály Pocsai,¹ Imre F. Barna,¹ Joshua T. Moody,² and Gábor Demeter¹

¹*Wigner Research Centre for Physics, Budapest, Hungary*

²*Max Planck Institute for Physics, Munich, Germany*

(Dated: May 26, 2022)

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 a tenuous vapor and machine learning techniques are tested for extracting quantitative information from the images. By building a database of simulated signals with a range of plasma parameters for training Deep Neural Networks, we demonstrate that they can extract 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.

PACS numbers:

I. 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 difficulties that wakefield acceleration poses. The AWAKE Collaboration 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 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 ($\mathcal{N} = 10^{14} - 10^{15} \text{ cm}^{-3}$) atomic vapor. The setup is similar to the one 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. 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.

II. SCHLIEREN IMAGING OF A PLASMA CHANNEL CROSS-SECTION

A. Measurement principle

In the novel accelerator device of the AWAKE Collaboration, a 10-meter-long rubidium vapor source is used, with a TW power laser pulse propagating along the axis to ionize the vapor. Plasma for accelerator operation must satisfy very stringent constraints with respect to homogeneity of density. This can be fulfilled by carefully tailoring rubidium vapor density in the chamber and achieving single-electron ionization of the atoms with a probability very close to unity [30]. The propagation of the ultra-short, TW ionizing pulse along the vapor source is itself a complex nonlinear process [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 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 [31] 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 $\lambda = 780.311$ nm, close to the $\lambda = 780.241$ nm D_2 resonance wavelength of rubidium. With this choice, the anomalous dispersion around the resonance line yields a refractive index change of $\delta n_v = 10^{-4} - 10^{-3}$ caused by the vapor for $\mathcal{N} = 10^{14} - 10^{15}$ cm $^{-3}$ densities. 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. The Rayleigh-length of the probe beam with these parameters is $R \approx 27$ m, so the transit of the probe beam across the 4 cm thick layer of rubidium vapor amounts to a phase-shift of the Gaussian beam profile due to vapor dispersion, combined with an attenuation factor due to absorption. 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.

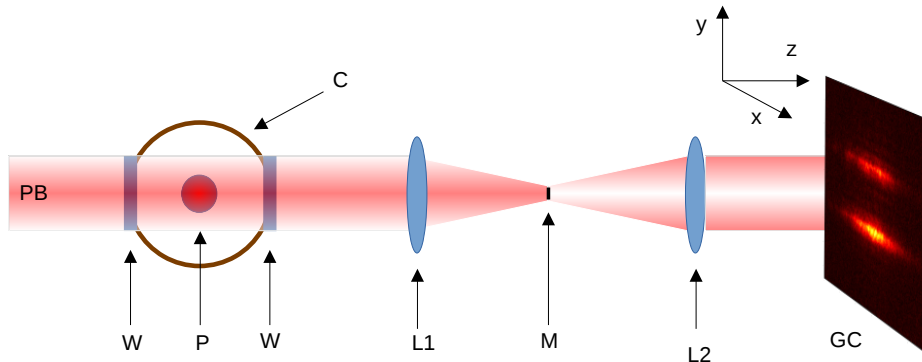


FIG. 1: Sketch of the Schlieren imaging measurement setup (not drawn to scale). PB - probe beam, C - vapor source chamber cross-section, W - viewport, P - plasma channel cross section, L1, L2 - 75 cm focal length lenses in 4f setup, M - mask, GC - gated camera.

An example of a measured image can be seen on Fig. 2 a). Now the probe beam diameter (\sim mm) is negligible with respect to the spatial scale (\sim m) at which the plasma channel cross-section changes along the x axis (the direction of propagation for the TW ionizing pulses). Therefore, the refractive index variation (and hence probe beam phase modulation) changes only along the y direction. Changes along the x direction on the measured image are only

due to the variation of the probe beam amplitude along this coordinate. Thus the measured image contains stripes parallel with the x axis and it is convenient to create a 1D lineout along y by taking a region of interest (ROI) from the region of x where the probe beam is the most intense and averaging this along x (Fig. 2 b)). This procedure helps reduce the noise level of the signal and produces a 1D curve that contains the same information on the plasma as the 2D camera image does. The task is to extract quantitative information on the plasma channel from this curve.

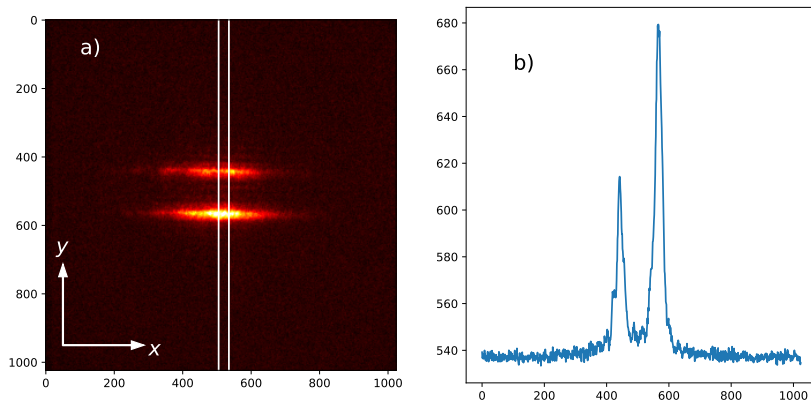


FIG. 2: a) Schlieren image on the gated camera. Lines in the middle mark the region of interest (ROI) from which we calculate the lineout around the probe beam center. b) Lineout taken from the Schlieren image ROI by averaging along x .

B. Obtaining plasma parameters

To help interpret images obtained in the measurement, we first note, that if the plasma distribution in the $y - z$ 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 [32]. (Note that the vapor densities used here require that we augment the description of [32] with a pressure broadening term in the homogeneous lineshape [33].) 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 using the standard

formulas for Fourier optics [31]. 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 ionizing pulse energies, ionization probability saturates to values very close to unity in the central part of the beam [27, 30]. 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. The plasma density is thus assumed to have the following form:

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

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 unionized vapor.

The functional form written here can account for a weakly ionized vapor, when $r_0 = 0$ and the plasma density distribution is an axisymmetric Gaussian. 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, $r_0 > 0$ happens only for $P_{max} \approx 1$.

Given the above explicit functional form, we can calculate simulated signals on the gated camera for any set of plasma parameters $\{y_0, r_0, t_0, P_{max}\}$, some examples can be seen on Fig. 3 b). The mathematical task is now to determine the plasma parameters that had been used to give rise to the given (possibly noisy) signal. Because of the nonlinear, integral-

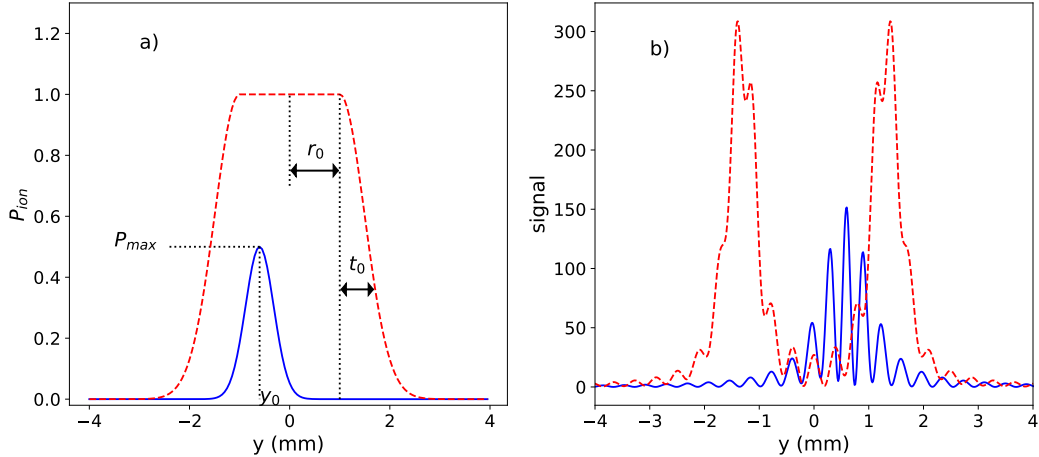


FIG. 3: a) Ionization probability for a weakly ionized ($P_{max} < 1, r_0 = 0$), narrow plasma channel (solid blue line) and a saturated core $P_{max} = 1, r_0 > 1$, wide plasma channel (dashed red line). b) The corresponding simulated signals (signal for a weakly ionized plasma has been scaled up for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

type relationship between the parameters and the Schlieren signal, 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.

III. 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 to the input data with significant non-linearities [34]. A typical data-based application of a feedforward artificial neural network is the non-linear regression, which is aimed to infer some parameters from the input data:

$$y_{j, Pred} = f(x) = A \left(\sum_{i=1}^N w_{ij} x_i + b_i \right), \quad (2)$$

where A is some non-linear activation function, N is the number of the neurons in the layer, b_j is a bias vector and the w_{ij} matrix contains the trainable parameters. A network may consists of multiple such layers—this is the case when it is called a *deep* neural network. During a supervised training cycle (*epoch*) of the network, the training input data is evaluated

and compared to a reference output according to well defined loss function, $\mathcal{L}(y_{Pred}, y_{True})$. 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.

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 [35]) or some weight parameters in a multi-component loss function [36]. However, one of the most crucial and necessary element of the training is undoubtedly a good quality training dataset.

A. Dataset generation

Based on the formulation discussed in II B, we have implemented a Python script that is able to simulate the Schlieren signal from a given set of plasma parameters. We have utilized this script to generate datasets for the training, validation and testing of the DNNs. The datasets contained the simulated Schlieren signals sampled at 1024 points, paired with the set of generating plasma parameters $\{y_0, r_0, t_0, P_{max}\}$ used to obtain them. For each set of parameters, 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 $[P_0, 1]$:

$$P_{max} \in [P_0^n, 1]^{1/n} \quad (3)$$

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 rule:

$$r_0 < \frac{Q}{(1 - P_{max})^2} \quad (4)$$

with $Q = 0.25 \mu\text{m}$. This value 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. 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. The reason for this is that, such samples resemble only noise, and no peaks or interference patterns can be extracted from them. Then, we dropped samples with the absolute value of their mean less than 1.0. Our argument behind it is that there are samples with a maximum amplitude greater than 5.0 resembling noise and having a single large value somewhere, however, further raising the minimum accepted signal amplitude would exclude samples with acceptable signal/noise ratio. 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. This filtering of the data greatly improves the performance of the network, but it does not introduce any artificial, unwanted bias.

In most of the machine learning applications, it is crucial to standardize the data with some pre-processing method. However, in our case, we have full control over the generation of the simulated datasets, therefore only the followings have been considered:

1. The input is a vector of 1024 elements, representing the detector image. 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 a constant factor to scale them into an approximately uniform range:

$$\tilde{P}_{max} = F_P \cdot P_{max}, \tag{5a}$$

$$\tilde{y}_0 = F_Y \cdot y_0, \tag{5b}$$

$$\tilde{t}_0 = F_T \cdot t_0, \tag{5c}$$

$$\tilde{r}_0 = F_R \cdot r_0. \tag{5d}$$

The scaling factors have been chosen as: $F_P = 1$, $F_Y = F_T = 10^2$ and $F_R = 10^3$. The overall distributions of the parameters after these scalings are plotted on Figure 4.

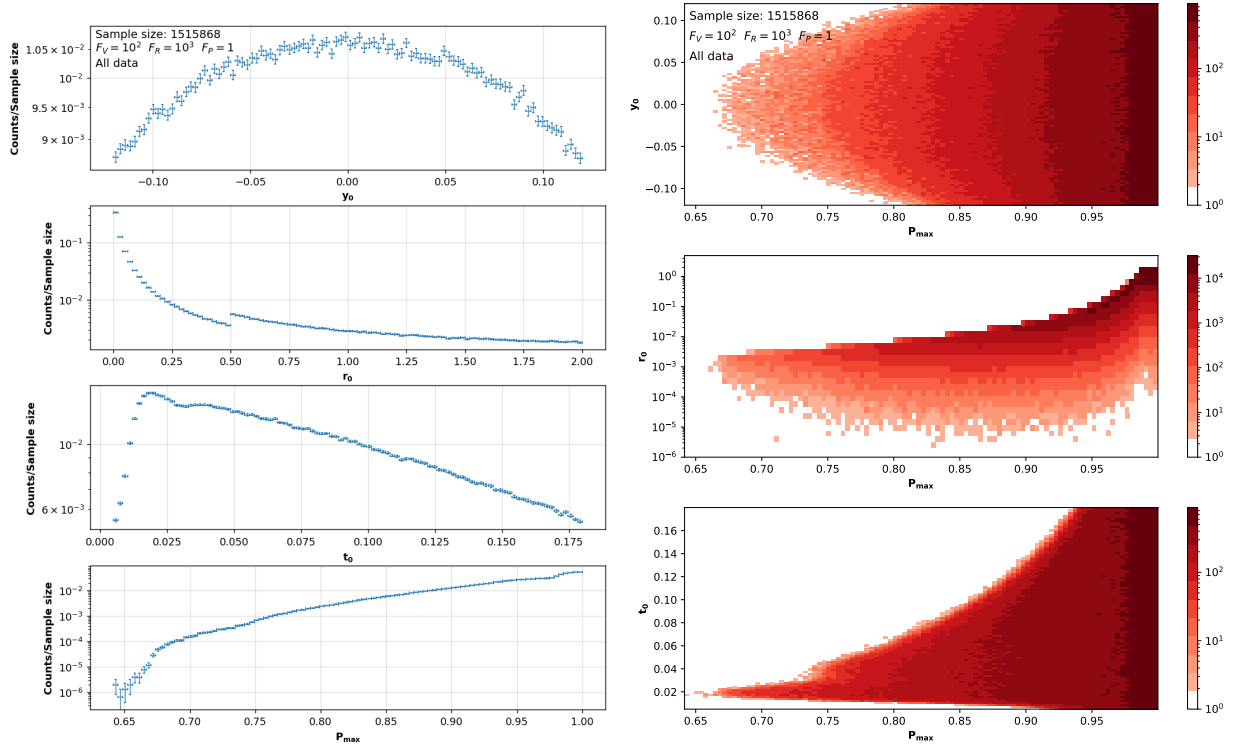


FIG. 4: The distribution of the parameters in the training data.

Both for training and validating purposes, we generated realistic and noiseless datasets with different distributions given by $n = 1, 3, 5, 7$ and 10 , respectively. First, we trained and validated the networks on distinct datasets, that is, using datasets corresponding only to $n = 1$ with noise, then e.g. $n = 5$ with noise, etc. We accepted a configuration only if the validation was convincing both for realistic and noiseless datasets. We also tested cross-validation in terms of noise and distribution in P_{max} . That is, we tried to validate a network trained with realistic data on noiseless data and vice versa, and also tried validating a network such that the training and validating datasets have different distributions in P_{max} . In these cases, we experienced bad correlations between the true and the predicted parameters. According to these experiences, first we decided to use unified datasets both for training and validating purposes, i.e. the datasets contained realistic and noiseless samples from all P_{max} distributions, given above. Later on, we used only realistic training and validating datasets with mixed distributions in P_{max} .

The unified and filtered datasets contained a total amount of $1.40M$ training and $0.23M$ validating samples, respectively.

According to the argument in II B, there is a strong correlation between the parameters

P_{max} and r_0 : the $r_0 > 0$ values are favored only when $P_{max} \sim 1$. In order to improve the statistics of the training dataset, an additional set was also generated, that allowed configurations only with $r_0 > 0.5$ mm—hence there is a small jump in the r_0 distribution of Figure 4. The correlations between the parameters are also visualized in Figure 5, which presents the r_{xy} Pearson correlation coefficients of the training data with a population size of $n = 1,515,868$ samples, defined by the following equation:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (6)$$

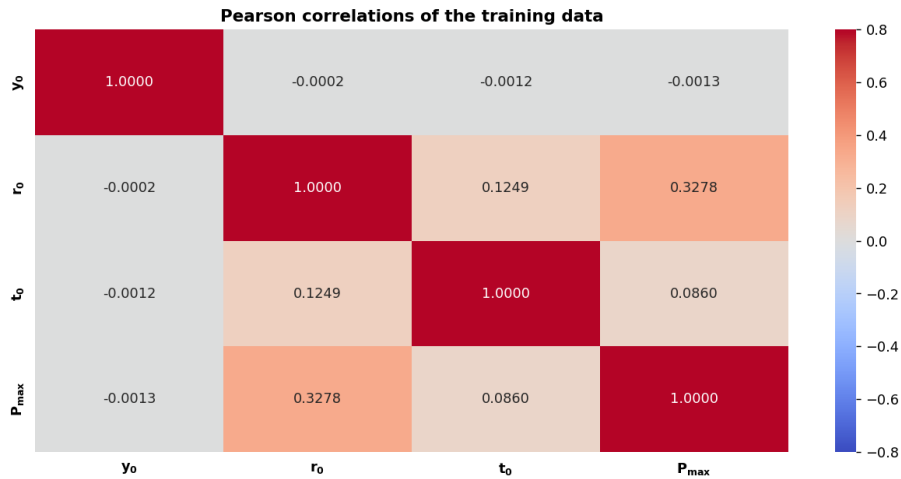


FIG. 5: The Pearson correlations in the training data.

As the figure shows, the r_0 (the radius of the plasma) is strongly correlated with the maximum ionization probability, 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.

B. Methodology

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 [37, 38]. The basic building block of the framework, referred as *Dense block* is sketched on Figure

6, which consists of a fully connected layer, followed by a batch normalization, a scaled exponential linear unit (SELU [39]) and a dropout layer with fixed dropout rate of 0.1.

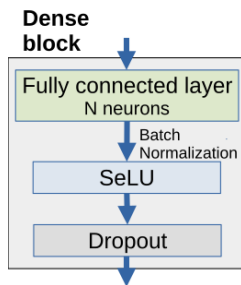


FIG. 6: A basic building block of the applied neural networks.

Our neural networks have been composed of such building blocks as it is depicted on Figure 7, which has three distinct parts and several configurable parameters. In the first part (marked with light pink background) one or more feature extraction block process the input data, with D_i parallel Dense blocks and 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. The L index specifies the $L - 1$ number of the consecutive feature extraction blocks, and afterwards 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 hidden layer follows, with N_{HL} neurons in the given layer. Finally, the last fully connected layer represents the four plasma parameters with linear activation. The mean absolute error has been chosen for the loss function, while the the optimization was performed with the Adam algorithm [35]. The initial learning rate has been slowly decreased with a linear decay.

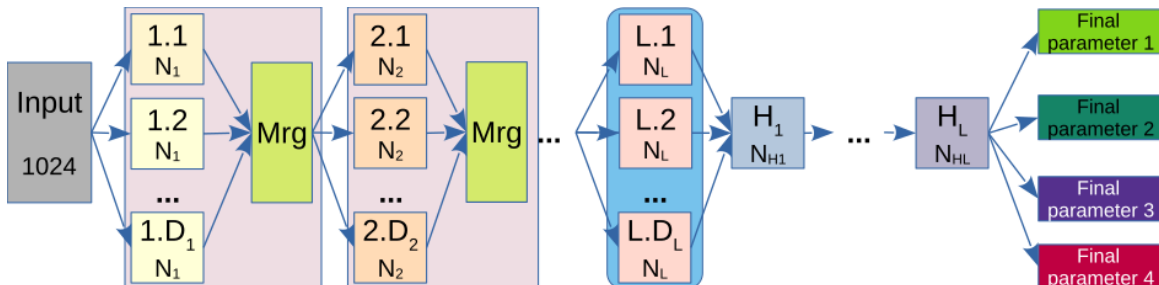


FIG. 7: The structure of the implemented neural networks.

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

C. Network implementations

In this study, several different architecture implementations have been investigated. During training, the mean absolute error was utilized as the loss function, defined as:

$$\mathcal{L}(y_{Pred}, y_{True}) = \frac{1}{N} \sum_{i=1}^N |y_{Pred} - y_{True}| \quad (7)$$

Furthermore, the training process was monitored with the mean squared error and Log-Cosh errors as additional metrics [40], and 15% of the training data was used for validating each epoch. Hereinafter, we refer to the calculation of the loss for every epoch as the measurement of training quality. Obviously, bad convergence in the measurement of training quality shows bad correlation between the true and predicted parameters. However, it is still possible that with quick and clear convergence, the correlations are still poor. We found that it is sufficient to train the networks for 20,000 epochs to get the best achievable quality of predictions.

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 bad. 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 $3 \cdot 10^{-5}$.

We also 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 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.

In Table I the specific values of the configurable parameters and the final loss value for the listed variants are summarized. In the followings, the best three variants are compared.

Models named *Bush* and *Skirun* included only feature extraction parts without additional hidden layers, while the model named *LongBottle* is a special case of the network, where no explicit feature extraction blocks have been applied, therefore it is considered as a traditional deep neural network. For a reference of the models, see Table I.

TABLE I: The configurable hyperparameters.

	Skirun	Bush	LongBottle
D_L	16, 8, 4	16, 4	1, 1, 1, 1, 1
N_L	512, 256, 128	256, 64	1024, 512, 256, 128, 64
N_{HL}	0	0	0
Trainable parameters	9.6M	4.3M	1.75M
Final loss	0.00402	0.00667	0.00954

IV. RESULTS

In order to test the performance of the trained networks, the plasma parameters have been predicted from the validation dataset described in the previous section. Figure 8 presents the Pearson correlation coefficients of the predicted parameters for the different networks.

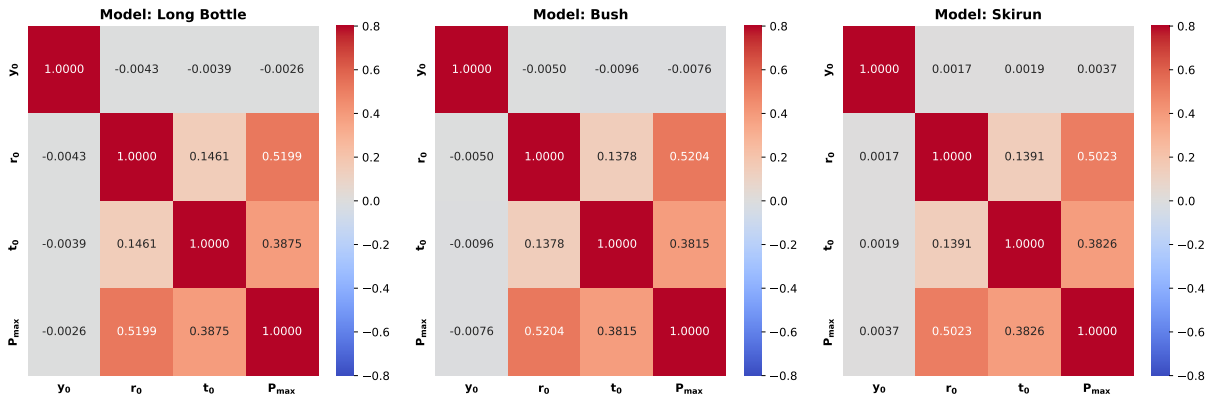


FIG. 8: The learned Pearson correlations.

The different architectures learned the same correlations, there is both qualitative and quantitative agreement. Comparing the results with the reference plotted in Figure 5, the following conclusions can be drawn:

1. the y_0 parameter is basically uncorrelated, which is well reproduced by all variants;
2. the networks predict $\sim 25\%$ higher r_{t_0, r_0} correlation than the reference;
3. the predicted r_{P_{max}, r_0} correlation is $\sim 60\%$ higher, than the reference;
4. in the reference, the r_{P_{max}, t_0} correlation is only 8.6%, while it is much higher, $\sim 38\%$ from all the predictions.

Note that the uncorrelatedness of the y_0 parameter means that the plasma parameters have a translation invariance with respect to y_0 . This exactly agrees 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. It is also important to emphasize that we did not put any constraint on the plasma parameters, like $P_{max} \in [0, 1]$. All the presented network variants recognized the physical range of the parameters along with the correlations between them, e.g. larger values for r_0 are favored at higher values of P_{max} .

The predicted parameters are plotted versus the true values for all the investigated networks in Figure 9. All of the network variants produced a fairly adequate prediction of the plasma parameters—however, there are some significant differences. The *LongBottle* 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.

To quantify the goodness of the predictions, first, a linear fit was performed to all parameters, and the mean error along with the mean absolute error is shown for all case. Second, the amplitude and phase errors were investigated according to the following considerations.

Let y_{Pred} , y_{Noisy} and y_{True} denote the predicted and the simulated Schlieren signals with and without the simulated noise, respectively. That is,

$$y_{Noisy}(x) \approx y_{True}(x) + n(x) \quad (8)$$

with $n(x)$ denoting the noise such that

$$\int_{x_{min}}^{x_{max}} n(x)dx = 0. \quad (9)$$

Introducing the local noise to signal ratio

$$\eta(x) = \frac{n(x)}{y_{True}(x)}, \quad (10)$$

equation (8) can be rewritten as

$$y_{Noisy}(x) \approx y_{True}(x) [1 + \eta(x)]. \quad (11)$$

Recall that for $P_{max} \approx 1$, the local noise to signal ratio can be neglected. However, the smaller values P_{max} takes, the larger will be the fraction of the $[x_{min}, x_{max}]$ interval where

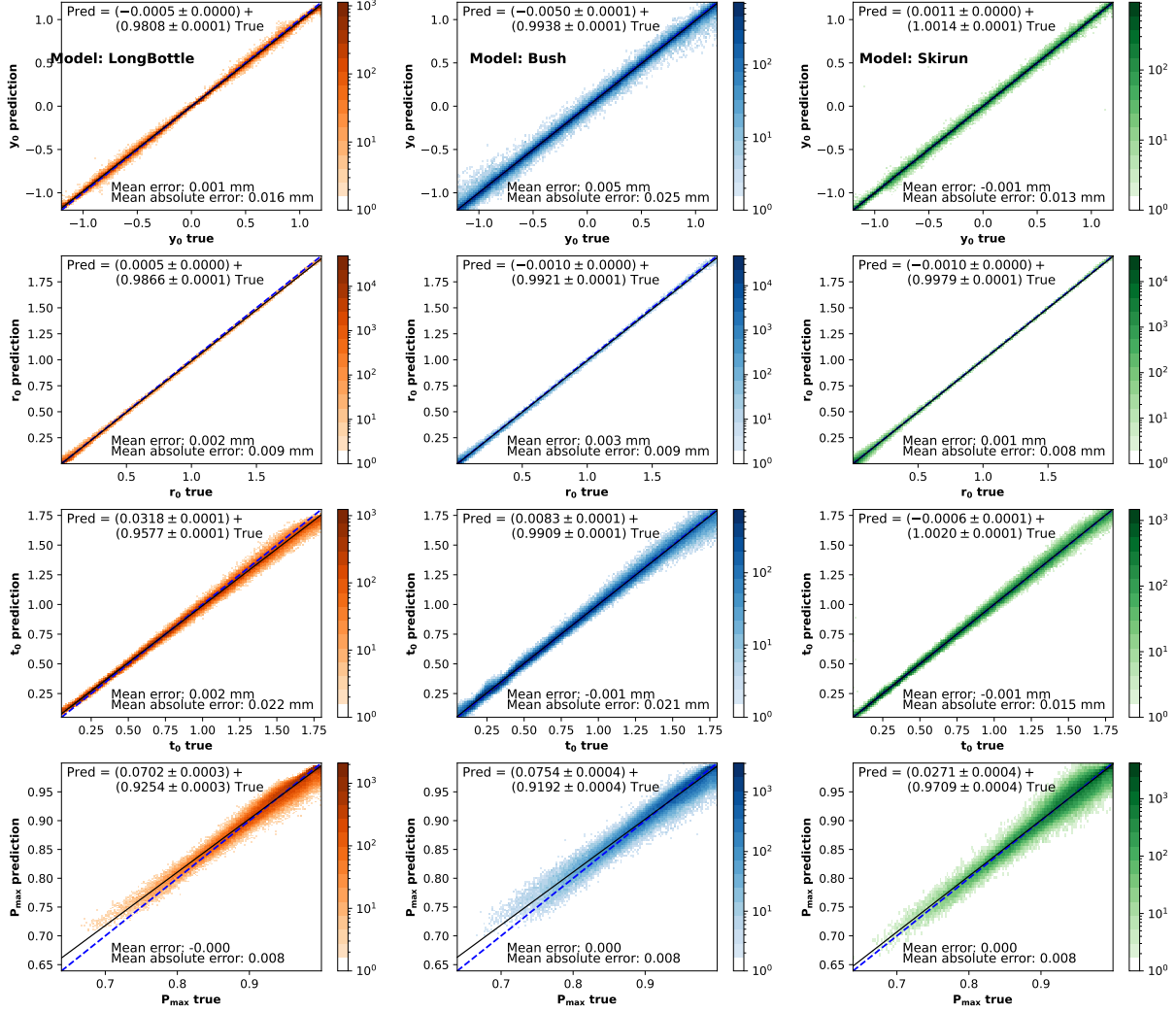


FIG. 9: The learned parameter correlations.

$\eta(x)$ plays a significant role. This can be seen on the top and bottom panels of Fig. 11, respectively.

The high accuracy of our models suggests that after some slight phase and amplitude corrections, the true signals can be restored from the predicted ones. That is,

$$y_{Pred}(x) \approx a \cdot y(x - x_{ph,corr}) \quad (12)$$

with $x_{ph,corr}$ denoting the phase correction and a being close to unity. $y(x)$ can either be $y_{Noisy}(x)$ or $y_{True}(x)$. As discussed above, for $P_{max} \approx 1$, the relation $y_{True}(x) \approx y_{Noisy}(x)$ holds. In order to see how the constant a measures the quality of the fitting, first, consider

the normalized signals in the followings:

$$\tilde{y}(x) = \frac{y(x)}{C} \quad (13)$$

with

$$C = \int_{x_{min}}^{x_{max}} |y(x)| dx. \quad (14)$$

Let the phase dependent amplitude error defined as

$$A_{err}(x_{ph}) = \int_{x_{min}}^{x_{max}} |\tilde{y}(x) - \tilde{y}_{Pred}(x - x_{ph})| dx \quad (15)$$

with $\tilde{y}(x)$ being either $\tilde{y}_{True}(x)$ or $\tilde{y}_{Noisy}(x)$, and x_{ph} being a small phase shift in the predicted signal. Due to the high accuracy of the predictions, we can assume that for some small value of $x_{ph,min}$, $A_{err}(x_{ph,min})$ will be minimal. Therefore, we define $A_{err} = A_{err}(x_{ph,min})$ as the amplitude error and $x_{ph,err} = x_{ph,min}$ as the phase error. Using the approximation showed in Eq. (12), it is easy to see that comparing the true, i.e. noiseless signal to the predicted one, the expression of the amplitude error reduces to

$$A_{err} = |1 - a|, \quad (16)$$

meaning that A_{err} is a good estimate for the relative error of the amplitude of the signal. When comparing the predicted signal to the noisy one, and considering the cases where the local noise to signal ratio is negligible, we get the same approximating expression for A_{err} and therefore its meaning remains unchanged. However, when $\eta(x)$ is not negligible, A_{err} neither does have such a nice graphical meaning as before, nor can be approximated with a simple formula. In contrast, when interpreting the phase error, we recall that both for the noiseless and noisy cases, the relation $x_{ph,min} = x_{ph,corr}$ holds. Therefore, $x_{ph,err} = x_{ph,corr}$, that is, the phase error is the slight correction with which the original signal can be restored from the predicted one with high accuracy.

We used the quantities defined above to measure the quality of the predictions as follows. For every test sample, we calculated A_{err} and $x_{ph,err}$ and made a histogram from them. In the histogram of the amplitude error we divided the $[0, A_{err,max}]$ interval to 1 024 bins, while in the histogram of the phase error we took 51 bins since the phase error ranged from -25 pixels to 25 pixels. We calculated the midpoints of the intervals representing the bins and assigned the counts of the bins to the corresponding midpoints. Then, we divided the number of counts of the bins with the total number of samples to obtain the probability

of the signal having A_{err} amplitude error and x_{err} phase error, respectively. From these probabilities we calculated the mean values of A_{err} and $|x_{err}|$ with

$$\langle A_{err} \rangle = \sum_i p(A_{err,i}) A_{err,i} \quad (17)$$

and

$$\langle |x_{err}| \rangle = \sum_i p(x_{err,i}) |x_{err,i}|, \quad (18)$$

respectively.

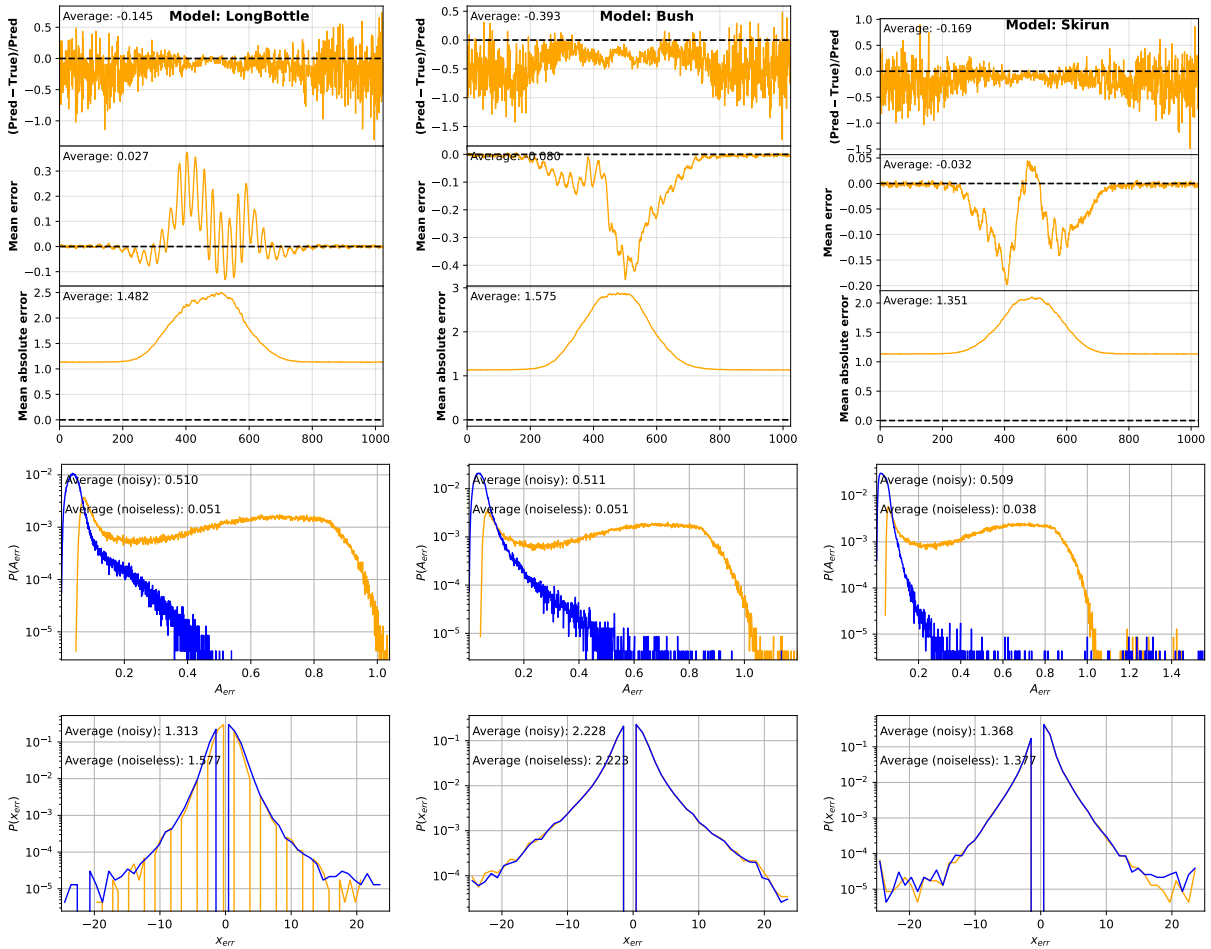


FIG. 10: The average error values (see text for details).

We show the probabilities of errors versus the errors on the bottom two panels of Fig. 10. The orange and blue curves correspond to comparing the predicted signal to the noisy and noiseless signals, respectively. As discussed above, both the probability histogram and the mean value of the amplitude error for the noiseless case shows the validity of the approximation given in Eq. (12). For every model, the relative error of the amplitude remains

below 10% with a mean value of around 5%. The severe distortion of the noise can also be seen on the figures. The long tail of the noisy histograms suggests that the vast majority of the samples are weighted with a significant amount of noise. However, considering the phase errors, we cannot see any difference between the noisy and the noiseless probability histograms as they overlap. This suggests that the predictions carry a very small amount of phase error. More precisely said, most prediction does not suffer a phase error at all. For all the models, only an unnoticeable fraction of the predictions carries a phase error greater than 10 pixels. The mean value of the absolute phase errors reveals that the vast majority of the predictions has a phase error less than 2 pixels, corresponding to a relative phase error of 0.1 – 0.2%.

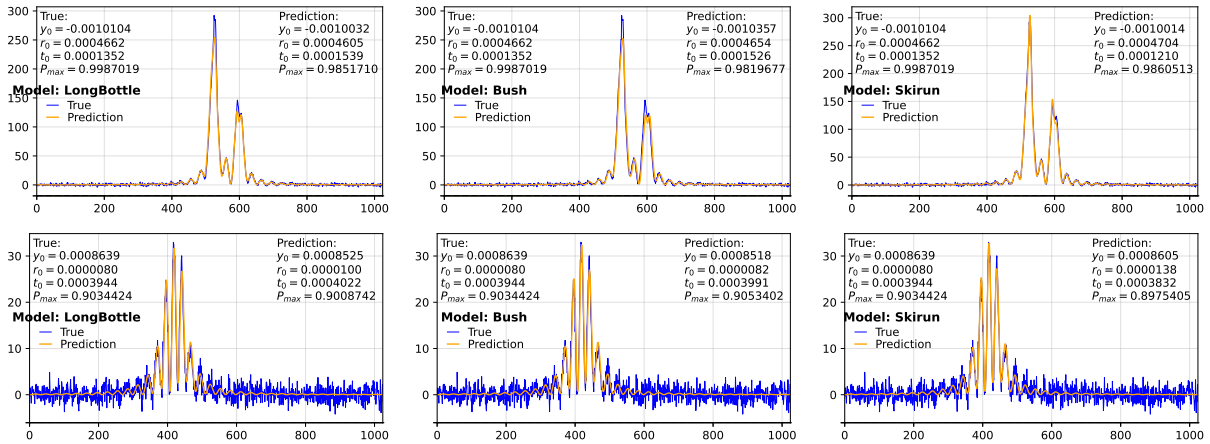


FIG. 11: Example predictions by the three models.

Finally, Figure 11 shows some examples for the predicted parameters and the Schlieren signals for each network architectures.

A. Robustness of the models

Whether trained neural networks may be useful for the evaluation of real measurements also depends on the sensitivity of their results on parameter changes of the experiment. The networks have been trained on samples calculated with a set of fixed parameters, but some parameters can actually fluctuate or drift somewhat during the course of a measurement. This means that the neural networks will be used for evaluating samples that have been obtained with slightly different experimental parameters, therefore it is important to verify

that network predictions are not affected substantially by such slight changes. In the setup considered here, the vapor density and the probe laser intensity are the most important parameters that may change slightly. The vapor density can be held constant to better than 1% accuracy, while the probe laser power may change possibly by a few per cents. In order to evaluate the effect of parameter changes, we have 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 changed by the same amount. The mean absolute error σ of parameter prediction was calculated in each case, as well as the mean error δ , whose increase signifies systematic errors of the prediction. Please note that the purpose of this section is to test the robustness of the trained models—however, these changes in the vapor density and probe laser intensity could be easily incorporated in additional training datasets, that could further improve the neural networks’ ability to give accurate and reliable predictions to the plasma parameters.

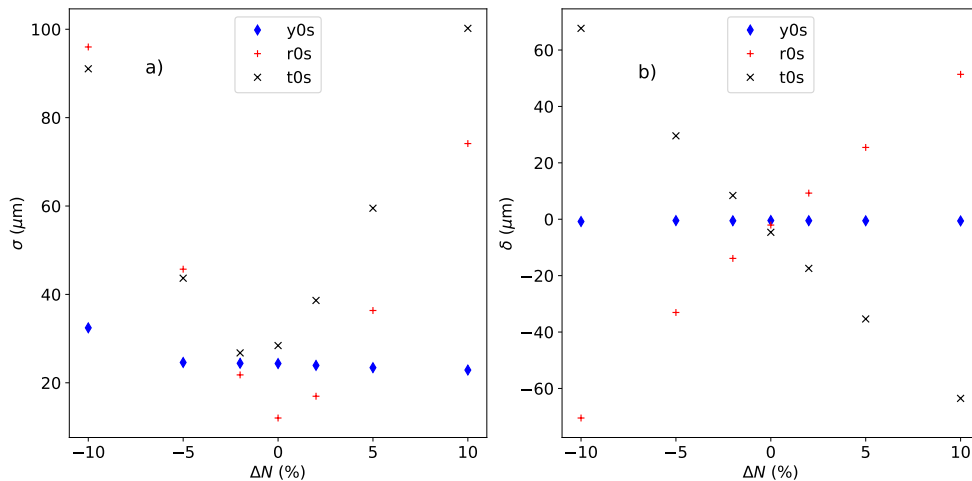


FIG. 12: a) Mean squared error σ and b) mean error δ of parameter prediction as a function of vapor density change from the value used for training.

Figure 12 a) shows the increment of parameter prediction error as the vapor density deviates from the value used during training. The mean error δ for the same parameter predictions can be seen on Fig. 13. (Note that the values are plotted in μm .) It is interesting, that while σ does not increase substantially for a deviation of 2% for any of the three parameters plotted, δ does deviate from the ideal value for r_0 and t_0 in opposite sense. Nevertheless, a systematic error $< 20\mu\text{m}$ for $\Delta N = 2\%$ is completely negligible. Probe laser

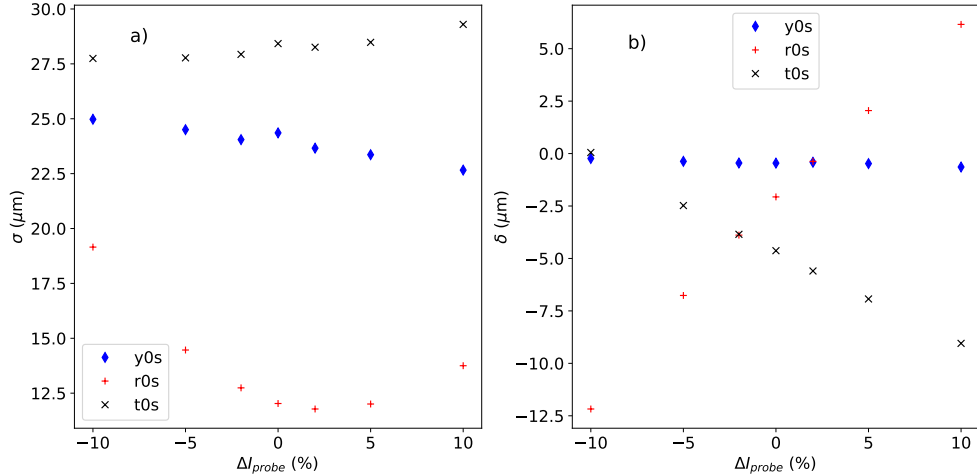


FIG. 13: a) Mean squared error σ and b) mean error δ of parameter prediction as a function of probe laser power change from the value used for training.

power changes have a much smaller effect (see Fig. 13), even for a power variance of $\pm 10\%$.

V. SUMMARY

In this paper a novel method to predict the geometrical dimensions of a plasma channel in atomic vapor has been presented, applying machine learning techniques. Schlieren imaging is an important method used to monitor the plasma properties, which is a crucial task in plasma wakefield acceleration technology. The implemented DNNs provide a robust and efficient framework to predict the plasma parameters from the noisy Schlieren signals with high accuracy. Without putting any constraint on the networks, they recognized the physical ranges of the parameters describing the plasma channel. The robustness of the networks has also been presented with respect to slight changes of the power of the probe laser and the density of the vapor. Improving this robustness may be achieved by further training the presented models on datasets corresponding to different powers of the probe laser and different vapor densities. These results and considerations suggests that after some improvement, the presented networks will be eligible to reliably evaluate the corresponding experimental data.

VI. ACKNOWLEDGMENT

The research was supported by the Hungarian National Research, Development and Innovation Office (NKFIH) under the contract numbers OTKA K135515, K123815 and NKFIH 2019-2.1.11-TÉT-2019-00078, 2019-2.1.11-TÉT-2019-00050, 2019-2.1.6-NEMZ_KI-2019-00011, 2020-2.1.1-ED-2021-00179, and the Wigner Scientific Computational Laboratory (WSCLAB, the former Wigner GPU Laboratory). Author G. B. was supported by the Ministry of Innovation and Technology NRD Office within the framework of the MILAB Artificial Intelligence National Laboratory Program. The authors are grateful to Gábor Papp and Gergely Gábor Barnaföldi for the useful discussions.

References

- [1] S. Narayanan, E. Chaniotakis, and C. Antoniou, *Transportation Research Part C: Emerging Technologies* **111**, 255 (2020), ISSN 0968-090X, URL <https://www.sciencedirect.com/science/article/pii/S0968090X19303493>.
- [2] V. K. Verma and S. Verma, *Materials Today: Proceedings* **57**, 2144 (2022), ISSN 2214-7853, international Conference on Innovation and Application in Science and Technology, URL <https://www.sciencedirect.com/science/article/pii/S2214785321077749>.
- [3] L. Deng and X. Li, *IEEE Transactions on Audio, Speech, and Language Processing* **21**, 1060 (2013).
- [4] M. Feickert and B. Nachman (2021), 2102.02770.
- [5] G. Bíró, B. Tankó-Bartalis, and G. G. Barnaföldi (2021), 2111.15655.
- [6] F. A. Matos, D. R. Ferreira, and P. J. Carvalho, *Fusion Engineering and Design* **114**, 18 (2017), ISSN 0920-3796, 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, and J. contributors, *Journal of Fusion Energy* **38**, 458 (2019).
- [8] D. J. Clayton, K. Tritz, D. Stutman, R. E. Bell, A. Diallo, B. P. LeBlanc, and M. Podesta, *Plasma Physics and Controlled Fusion* **55**, 095015 (2013).
- [9] G. Demeter, *Review of Scientific Instruments* **68**, 1438 (1997), <https://doi.org/10.1063/1.1147630>, URL <https://doi.org/10.1063/1.1147630>.
- [10] G. S. Settles, *Schlieren and Shadowgraph Techniques - Visualizing Phenomena In Transparent Media* (Springer-Verlag, Berlin, Heidelberg, 2001), ISBN 978-3-642-63034-7.
- [11] E. Traldi, M. Boselli, E. Simoncelli, A. Stancampiano, M. Gherardi, V. Colombo, and G. Settles, *EPJ Techniques and Instrumentation* **5**, 4 (2018).
- [12] C. E. Clayton, K.-C. Tzeng, D. Gordon, P. Muggli, W. B. Mori, C. Joshi, V. Malka, Z. Najmudin, A. Modena, D. Neely, et al., *Phys. Rev. Lett.* **81**, 100 (1998), 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, and R. Bock,

- Physica Scripta **58**, 634 (1998), URL <https://doi.org/10.1088/0031-8949/58/6/017>.
- [14] H. Honda, K. Katsura, E. Takahashi, and K. Kondo, Applied Physics B **70**, 395 (2000).
- [15] F. Veloso, H. Chuaqui, R. Aliaga-Rossel, M. Favre, I. H. Mitchell, and E. Wyndham, Review of Scientific Instruments **77**, 063506 (2006), <https://doi.org/10.1063/1.2206773>, URL <https://doi.org/10.1063/1.2206773>.
- [16] D. Batani, J. Santos, P. Forestier-Colleoni, D. Mancelli, M. Ehret, J. Trela, A. Morace, K. Jakubowska, L. Antonelli, D. del Sorbo, et al., Journal of Fusion Energy **38**, 299 (2019).
- [17] I. A. Znamenskaya and I. A. Doroshchenko, Journal of Flow Visualization and Image Processing **28**, 1 (2021), ISSN 1065-3090.
- [18] S. Cai, Z. Wang, F. Fuest, Y. J. Jeon, C. Gray, and G. E. Karniadakis, Journal of Fluid Mechanics **915**, A102 (2021).
- [19] B. N. Ubald, P. Seshadri, and A. Duncan, *Density estimation from schlieren images through machine learning* (2022), URL <https://arxiv.org/abs/2201.05233>.
- [20] C. Joshi and T. Katsouleas, Physics Today **56**, 47 (2003).
- [21] W. Leemans and E. Esarey, Physics Today **62**, 44 (2009).
- [22] E. Gschwendtner et al., Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **829**, 76 (2016), ISSN 0168-9002, 2nd European Advanced Accelerator Concepts Workshop - EAAC 2015, URL <http://www.sciencedirect.com/science/article/pii/S0168900216001881>.
- [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**, 363 (2018).
- [24] E. Öz and P. Muggli, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **740**, 197 (2014), ISSN 0168-9002, proceedings of the first European Advanced Accelerator Concepts Workshop 2013, URL <http://www.sciencedirect.com/science/article/pii/S0168900213015003>.
- [25] G. Plyushchev, R. Kersevan, A. Petrenko, and P. Muggli, Journal of Physics D: Applied Physics **51**, 025203 (2017), URL <https://doi.org/10.1088/1361-6463/aa9dd7>.
- [26] A. Couairon and A. Mysyrowicz, Physics Reports **441**, 47 (2007), ISSN 0370-1573, URL <http://www.sciencedirect.com/science/article/pii/S037015730700021X>.
- [27] G. Demeter, Phys. Rev. A **99**, 063423 (2019), URL <https://link.aps.org/doi/10.1103/PhysRevA.99.063423>.

- [28] G. Demeter, J. T. Moody, M. A. Kedves, B. Ráczkevi, M. Aladi, A.-M. Bachmann, F. Batsch, F. Braunmüller, G. P. Djotyan, V. Fedosseev, et al., *Phys. Rev. A* **104**, 033506 (2021), URL <https://link.aps.org/doi/10.1103/PhysRevA.104.033506>.
- [29] A.-M. Bachmann, M. Martyanov, J. Moody, A. Pukhov, and P. Muggli, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **909**, 387 (2018), ISSN 0168-9002, 3rd European Advanced Accelerator Concepts workshop (EAAC2017), URL <https://www.sciencedirect.com/science/article/pii/S0168900217314687>.
- [30] E. Adli, A. Ahuja, O. Apsimon, R. Apsimon, A.-M. Bachmann, D. Barrientos, et al. (AWAKE Collaboration), *Phys. Rev. Lett.* **122**, 054802 (2019), URL <https://link.aps.org/doi/10.1103/PhysRevLett.122.054802>.
- [31] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics* (John Wiley and Sons, Inc., 2019), ISBN 9781119506874.
- [32] P. Siddons, C. S. Adams, C. Ge, and I. G. Hughes, *Journal of Physics B: Atomic, Molecular and Optical Physics* **41**, 155004 (2008), URL <https://doi.org/10.1088/0953-4075/41/15/155004>.
- [33] A. J. van Lange, P. van der Straten, and D. van Oosten, *Journal of Physics B: Atomic, Molecular and Optical Physics* **53**, 125402 (2020), URL <https://doi.org/10.1088/1361-6455/ab7fc2>.
- [34] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach* (Prentice Hall, 2010), 3rd ed.
- [35] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization* (2017), 1412.6980.
- [36] A. Dosovitskiy and J. Djolonga, in *International Conference on Learning Representations* (2020), URL <https://openreview.net/forum?id=HyxY6JHKwr>.
- [37] F. Chollet et al., *Keras*, <https://keras.io> (2015).
- [38] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, et al., *Tensorflow: Large-scale machine learning on heterogeneous distributed systems* (2016), 1603.04467.
- [39] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter (2017), URL <https://arxiv.org/abs/1706.02515>.
- [40] T. Moshagen, N. A. Adde, and 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>.