

# Plasma Tomography with Machine Learning

Diogo Duarte Parente Godinho Soares de Carvalho

diogo.d.carvalho@tecnico.ulisboa.pt

Physics Department, Instituto Superior Técnico, Universidade de Lisboa, Portugal

## Abstract

Due to the reduced number of lines of sight available in fusion reactors, computing tomograms with the desired resolution is a very time consuming process. In this work, we approached this problem from a machine learning perspective. Two different methods to perform plasma tomography, based on previously existing reconstructions and machine learning frameworks, were used. Tests were conducted using JET bolometric system and COMPASS soft X-ray setup, both methods having produced tomograms with a quality equivalent to that of the state of the art algorithms in substantially less time. In particular, Artificial Neural Networks produced better results for events where the magnetic flux surfaces shift considerably. The possibility of developing real-time control systems based on these algorithms is also addressed.

**Keywords:** Plasma Physics, Computed Tomography, Gradient Descent, Artificial Neural Networks

## 1. Introduction

Despite the efforts, since it was proposed in the 1940s, no fusion reactor [1] capable of confining the plasma long enough to produce energy on a desired industrial level was yet built. The difficulty of obtaining the necessary density, temperature and confinement times for self-sustained fusion is due to instabilities and other transport phenomena. It is then mandatory to develop diagnostic methods that provide us with a better understanding of what is happening inside the device. One such diagnostics is tomography, with a lot of research being done both in the electronic and algorithm parts [2, 3].

### 1.1. Plasma Tomography

In plasma tomography, contrary to medical applications, the plasma itself is the source of radiation which is generated by several mechanisms such as line radiation, bremsstrahlung and recombination [1]. For the spectral range in which this emission happens, it is typically assumed that the plasma is optically thin, i.e. that there exists no re-absorption of the emitted radiation. One then tries to obtain the emissivity profile  $g$  [ $\text{W m}^{-3}$ ] which is closely related to the temperature, density and impurity distribution. A detector will measure the chord brightness  $f$  [ $\text{W m}^{-2}$ ] that can be approximated as the integral of  $g$  along the line of sight, for the wavelength range to which the detector is sensible. If one has access to  $N_d$  different detectors, to obtain  $g$  we then have to solve a system of  $N_d$  integral equations of the form

$$f_i = \int_{L_i} g \cdot dl, \quad i = 1, \dots, N_d \quad (1)$$

where  $f_i$  is the measurement along the line of sight  $L_i$ . The system is under-determined, as an infinite number of detectors  $f_i$  would be needed to find a unique

solution.

While for medical tomography around  $10^5$  lines of sight are available, in plasma tomography normally only around  $10^2$  exist [4] (example in Figure 1). It is then necessary to use inversion algorithms that surpass this limitation and allow a feasible reconstruction of a poloidal plasma cross-section (Figure 2). They will normally require *a priori* knowledge or enforce some conditions on the solution, normally related to its smoothness.

Several important features can then be determined depending on the type of tomographic detectors available. For example, bolometers [2], sensible to radiation between the infrared and a few tens of keV (X-ray), can be used to determine the total power radiated over the entire plasma volume. One of such setups is currently in operation at JET [5] (Figure 1).

Another possibility is to use soft X-ray (SXR) to-

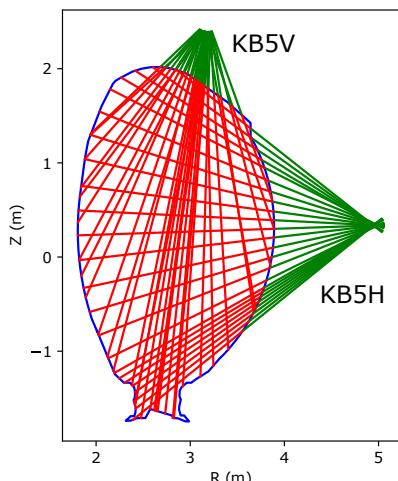


Figure 1: KB5 Bolometer system at JET. It has two cameras, KB5H and KB5V, having in total 56 lines of sight

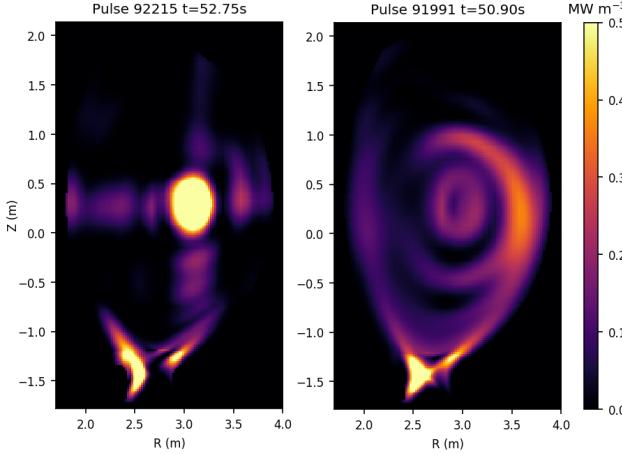


Figure 2: Example of JET tomographic reconstruction using a pixel grid of size  $196 \times 115$

mography [2]. Due to the high  $Z_{\text{eff}}$  dependence of bremsstrahlung, it is possible to characterize the heavy ion impurity localization and associated transport phenomena by analyzing the SXR emissivity profile. One example of such setup is available at COMPASS [6]. It should also be possible to perform control of the plasma position since the centroid of the SXR tomogram provides an estimation of the plasma center [7].

## 1.2. State of the art algorithms

Even though there exists a broad set of algorithms for plasma tomography, such as Abel Inversion [8], Cormack-based methods [9] or Maximum Entropy [4], currently one of the most widely used are Tikhonov Regularization algorithms [3].

When using this family of methods one tries to find the emissivity values pixel-wise. Using a square pixel grid with  $N_x$  horizontal and  $N_y$  vertical pixels we define the system:

$$\mathbf{f} = \mathbf{T} \cdot \mathbf{g} \quad (2)$$

where  $\mathbf{f}$  is a column vector with size  $N_d$  representing the output of the detectors,  $\mathbf{g}$  is the representation of the emissivity profile, originally a matrix reshaped into a column vector of size  $N_{\text{pixel}} = N_x \times N_y$ , and  $\mathbf{T}$  the projection matrix of size  $N_d \times N_{\text{pixel}}$  where each element  $T_{ij}$  corresponds to the contribution of pixel  $j$  to the value obtained by detector  $i$ . If one neglects the angular width of each chord the value of  $T_{ij}$  can be defined as the length of chord  $i$  in pixel  $j$  (see Figure 3).

In the case where  $N_d \geq N_{\text{pixel}}$  one could try to use least-squares method to minimize the  $\chi^2$  value

$$\chi^2 = (\tilde{\mathbf{T}} \cdot \mathbf{g} - \tilde{\mathbf{f}})^{\top} \cdot (\tilde{\mathbf{T}} \cdot \mathbf{g} - \tilde{\mathbf{f}}) \quad (3)$$

where  $\tilde{T}_{ij} = T_{ij}/\sigma_i$ ,  $\tilde{f}_i = f_i/\sigma_i$  and  $\sigma_i$  is the expected error from sensor  $i$  [4]. Unfortunately, this does not happen in plasma tomography. Since  $N_d < N_{\text{pixel}}$  there is an infinite number of solutions and it would always over-fit, obtaining  $\chi^2 = 0$ .

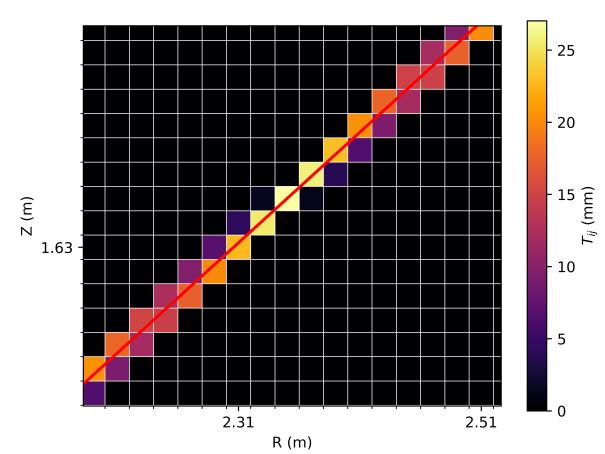


Figure 3: Example of contribution from line of sight  $i = 31$  of JET bolometer system to matrix  $\mathbf{T}$ . Each value  $T_{ij}$  represents the length the line of sight travels inside the pixel, having each pixel a size  $2 \times 2$  cm.

Since minimizing solely  $\chi^2$  won't provide a unique solution, a new functional  $\phi$  can be defined:

$$\phi = \frac{1}{2} \chi^2 + \alpha \mathcal{R} \quad (4)$$

where  $\mathcal{R}$  is a regularizing functional and  $\alpha$  is a positive constant usually called the regularization parameter. If now one tries to minimize  $\phi$  then a unique solution is provided that balances how well the reconstruction fits the original detectors measurements while imposing some characteristics on it [4].

Normally, anisotropic versions of the gradient operators are used to enforce smoothing along magnetic flux surfaces while allowing steeper gradients across them. This is currently what is being done at JET [10]. Other possibility is to use Minimum Fisher Regularization (MFR) [4]. One example of such implementation with the anisotropic version of the functional is currently used at COMPASS [11]. Recently, a modification of the algorithm that shortcuts its iterative nature was also proposed and it is expected to work in real-time [12].

Independently of the regularization functional chosen, the main issue with these types of algorithms are the size of the matrices involved. Due to this, the required operations, such as matrix inversions, are computationally expensive. Additionally, one still needs to tune the regularization parameter [13] until a certain convergence condition is reached. Normally one uses [4]:

$$\chi^2/N_d \simeq 1. \quad (5)$$

Another possible solution to compute tomograms, although they require the existence of previous computed reconstructions, are Artificial Neural Networks (NNs). Given the fast development of deep learning capabilities NNs became a successful tool in a broad variety of fields, being plasma tomography one of them [14, 15]. Since once trained a NN can rapidly output an image

without any fitting, matrix inversion or other computationally costly operation, they become a formidable candidate for real-time applications.

### 1.3. Original Contributions

Firstly, a method to obtain the general matrix corresponding to the tomographic inversion, based on existing reconstructions, is proposed. With this approach, the tomographic inversion becomes a single matrix product, significantly speeding up the process. In order to benchmark the method, tests using both JET bolometric system and COMPASS SXR setup were made.

This method is then compared with a more complex and non-linear model, namely NNs. While for JET bolometric system a NN to perform the full reconstruction was already developed [15], for COMPASS, and more specifically to SXR tomographic setups, such has not yet been done with real data. For this reason, a NN based on the one developed for JET was designed and trained for COMPASS, thus demonstrating the applicability of the architecture to different tomographic diagnostics and devices.

Lastly, several methods to improve the robustness of the NNs to the occurrence of malfunctioning detectors and changes in the camera system (adding, removing, changing lines of sight) are proposed and benchmarked.

## 2. Fitting the Inverse Matrix

Instead of using Tikhonov Regularization methods, one can try to define a single (inverse) transformation matrix  $\mathbf{M}$  that best fits an existing data-set of detector measurements  $\mathbf{f}$  and the corresponding reconstructions  $\mathbf{g}$ .

To determine  $\mathbf{M}$ , first, one must define an error metric, usually referred to as loss function [16], which an optimization algorithm will try to minimize. The chosen loss function  $\mathcal{L}$  was the mean absolute error defined as

$$\mathcal{L} = \frac{1}{N_{rec}} \sum_{k=1}^{N_{rec}} \sum_{i=1}^{N_{pixel}} \frac{|\tilde{g}_i - g_i|}{N_{pixel}}, \quad \tilde{\mathbf{g}} = \mathbf{M} \cdot \mathbf{f} \quad (6)$$

where  $N_{rec}$  is the number of reconstructions available.

To find the minimum of (6) one can use the gradient descent algorithm [16]. At each iteration  $n+1$ , the loss  $\mathcal{L}$  is computed and the values of  $\mathbf{M}$  are updated according to

$$\mathbf{M}^{(n+1)} = \mathbf{M}^{(n)} - \eta \frac{\partial \mathcal{L}}{\partial \mathbf{M}} \Big|_{\mathbf{M}=\mathbf{M}^{(n)}}, \quad (7)$$

where  $\eta > 0$  is called the learning rate that sets how much  $\mathbf{M}$  changes along the direction of the gradient. If  $\eta$  is too small, convergence takes too much time, if it is too big the algorithm will diverge. All entries of matrix  $\mathbf{M}$  are initialized as zero.

There exist several improved versions of gradient descent [17], that enable a faster convergence rate,

namely through the use of momentum. In each iteration an extra step is added according to:

$$\Delta \mathbf{M}^{(n+1)} = \gamma \Delta \mathbf{M}^{(n)} - \eta \frac{\partial \mathcal{L}}{\partial \mathbf{M}} \Big|_{\mathbf{M}=\mathbf{M}^{(n)}} \quad (8)$$

$$\mathbf{M}^{(n+1)} = \mathbf{M}^{(n)} + \Delta \mathbf{M}^{(n+1)} \quad (9)$$

where the new term  $\Delta \mathbf{M}$  stores the information about the last update and  $0 < \gamma < 1$  is the momentum parameter.

This optimization algorithm was implemented in Python using the Theano library [18] that enables the parallelization of operations in GPU. All computations were performed using a Nvidia Titan X GPU and an Intel(R) Core(TM) i7-3820 CPU @ 3.60GHz processor.

### 2.1. Application to JET

As a proof of concept, 600 reconstructions from JET bolometer database were used. These tomograms have been carefully curated so not to contain any artifacts, such as highlighted lines of sight or non-physical solutions with radiation appearing in places where it makes no sense for it to exist. All tomograms from this dataset possess a resolution of  $196 \times 115$  pixels and were computed using the measurements from JET bolometer system (see Figure 1) and the algorithm developed by Ingesson et al. [10].

As a starting point, the data was divided into two sets, 540 reconstructions for fitting, and 60 to which later some performance metrics will be calculated.

It was found that setting to zero the detectors which are known to be faulty increases the convergence rate (by a factor of 10) and does not produce significant differences in the outcomes. For this reason the choice was made to henceforth apply this pre-processing step.

Using  $\eta = 0.01$  and  $\gamma = 0.9$  we stopped gradient descent in iteration  $10^5$  (approximately 4 minutes) since  $\mathcal{L}$  decreased less than  $10^{-6} \text{ kW m}^{-2}$  between consecutive iterations. The minimum error obtained (pixel-wise) was about  $6 \text{ kW m}^{-3}$ , which is low compared with the usual dynamic range ( $1 \text{ MW m}^{-3}$ ).

After the fitting process it is important to evaluate how well can this matrix perform new reconstructions. For this purpose the 60 tomograms left out of the training process were used. Comparisons between the original reconstructions and the ones performed with the inverse matrix are presented in Figure 4. In these examples we highlight a case where the methods produced good results but also one where it severely failed (even though events like this one are less frequent).

To quantify how well these new reconstructions can mimic their originals several metrics are defined as follows:

- Structural similarity (SSIM) [19], measures the similarity along several windows of two images. Ranging from 0 to 1, where 1 is a perfect match.

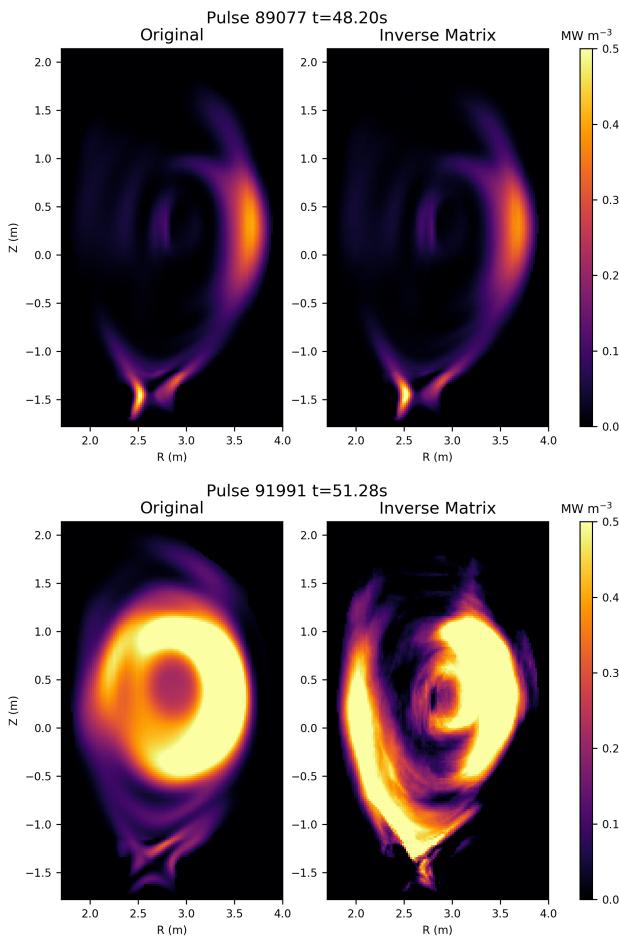


Figure 4: Comparison between original reconstructions from JET bolometer database and the ones done using the inverse tomographic matrix. Difference between images is calculated as original minus new reconstruction.

- Peak signal to noise ratio (PSNR) [20], used to measure the quality of reconstructions from compressed images. It is a logarithmic measure in dB, typical values are between 30 – 50 dB for JPEG images when compared to their originals uncompressed versions [21].
- Normalized root mean squared error (NRMSE) [22].
- Absolute relative error of the total power emitted ( $e_{power}$ ), where the total power is considered to be the sum of the emissivity over all pixels [ $\text{W m}^{-3}$ ].

Both the metrics obtained for the reconstructions from Figure 4 as well as the mean values for the full test set are given in Table 1. It is important to mention that only a few reconstructions with very high errors exist and they are the reason for the high standard deviation values obtained. These results are worse than those obtained in previous related works with a NN [15], although the NN was trained on a different and much bigger set. Still, the mean metrics values achieved indicate that in most cases the emissivity profile shape

is similar (high SSIM and PSNR) while the radiation peaks are slightly out of position (high NRMSE). Another conclusion is that the total power emitted by the plasma is correctly estimated (low  $e_{power}$ ).

Pulse	SSIM	PSNR (dB)	NRMSE	$e_{power}$
89077	0.977	36.37	0.10	0.01
91991	0.475	14.13	0.74	0.08
Full Set	$0.875 \pm 0.093$	$30.07 \pm 5.39$	$0.38 \pm 0.21$	$0.03 \pm 0.08$

Table 1: Evaluation metrics obtained using the inverse matrix method for the reconstructions in Figure 4 and on the full test set of JET tomograms

### 2.1.1 Regularization Patterns

Each value  $M_{ij}$  indicates how much detector  $j$  contributes to the value of pixel  $i$ . For this reason each column  $j$  of  $\mathbf{M}$  indicates the contribution of detector  $j$  to the whole reconstruction. It is then possible to observe the regularization pattern that  $\mathbf{M}$  is enforcing by plotting each column reshaped into the original image size ( $196 \times 115$ ). Some examples are presented in Figure 5.

Measures along most lines of sight contribute to curvatures that resemble magnetic flux surfaces. For lines of sight directly pointed at the divertor or the plasma core the detector contribution is mainly focused in these regions and the magnetic flux surface structures do not appear as highlighted. While the overall shape of the regularization patterns seems plausible, there are some regions with associated regularization that appear to have no relation to their corresponding line of sight.

We also attempted to fit the inverse matrix to all 600 reconstructions. As a result, some of the regularization patterns changed significantly, especially in regions where it is uncommon to have emissivity peaks. Hence, if in the future more reconstructions with similar quality become available, a more diverse dataset can be used to obtain the inverse matrix, thus mitigating some of the observed errors.

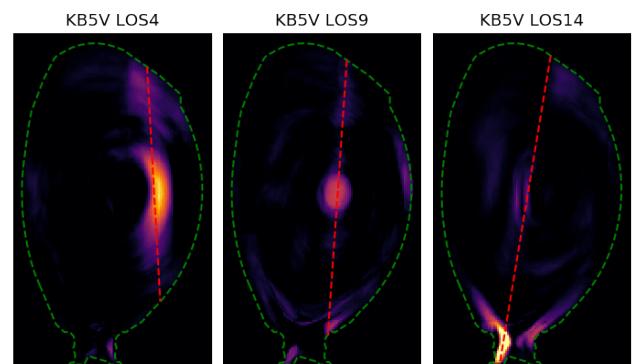


Figure 5: Example of regularization patterns obtained for different lines of sight (LOS) of JET bolometer setup (dashed red) with vessel contour (dashed green)

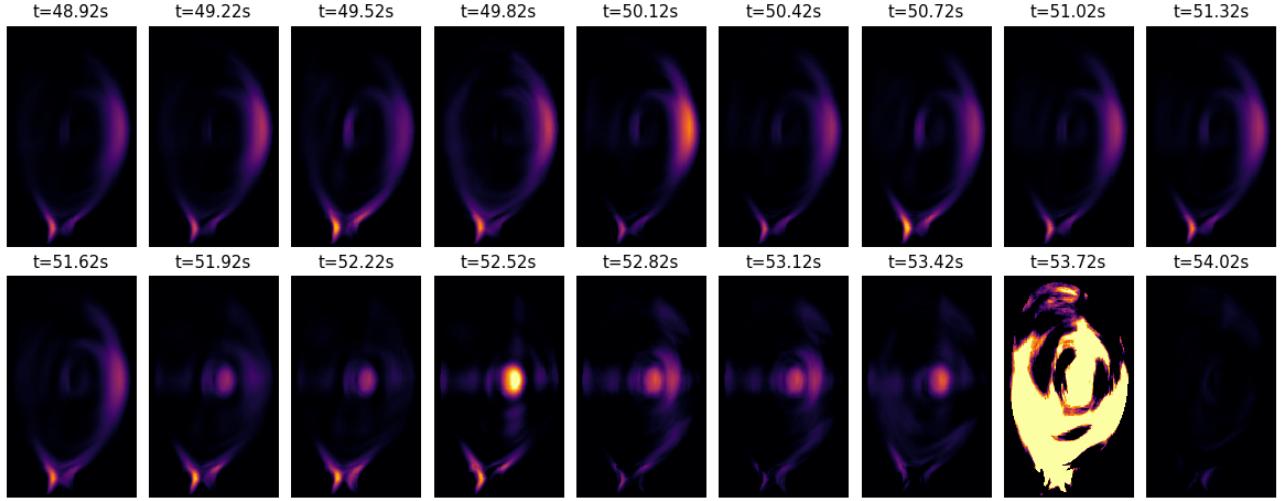


Figure 6: Reconstructions obtained, using the fitted inverse matrix, for pulse 92213 from  $t = 48.72$  s to  $t = 53.92$  s with  $\Delta t = 0.3$  s. The same dynamic range from 0 to  $1.5 \text{ MW m}^{-3}$  is used in all tomograms.

### 2.1.2 Full-Pulse Reconstruction

Using the NumPy implementation for matrix product in average one needs less than 0.4 ms to compute one reconstruction. It is then possible to compute reconstructions in real-time at half the sampling rate of the JET bolometer system (5 kHz). This time-stamp can be reduced given a faster implementation for matrix product is used, possibly taking into account the sparse nature of the matrices involved.

The reconstruction speed-up enables full pulse reconstructions as the one presented in Figure 6 allowing the study of several physical phenomena such as impurity sputtering, transport events or even disruption prediction. In this case the heavy ion motion, due to the tungsten removed from the vessel walls, causes the emission of a considerable amount of radiation, whose contribution is captured by the bolometers and observable later in the corresponding reconstructions.

## 3. Application to COMPASS

The COMPASS experiment possesses a SXR camera system for tomographic reconstructions [6] with a sampling frequency of 2 MHz. Along the years this setup has suffered several changes, both in the number of cameras used and the lines of sight viewing angles.

In this work we will start by using tomograms corresponding to GEOM\_201410 (Figure 7). This setup is equipped with 3 cameras A,B (35 lines of sight each) and F (20 lines of sight), making a total of 90 lines of sight. Later with GEOM\_201701, camera F was removed and, because of this, only 70 LOS remained. Also, the lines of sight of camera B slightly changed, covering now a bigger solid angle.

The tomograms performed at COMPASS use the Minimum Fisher Regularization algorithm described by Mlynar et al. [11], with anisotropic regularization along magnetic flux surfaces. Given the iterative na-

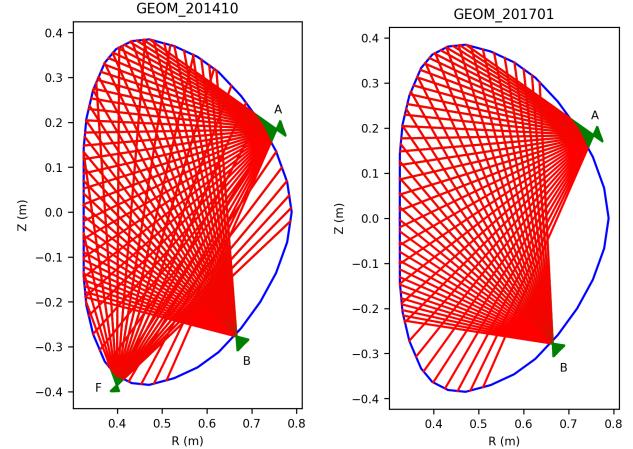


Figure 7: COMPASS SXR setup for GEOM\_201410 and GEOM\_201701

ture of Minimum Fisher Regularization, there is a need to define the convergence condition for which a reconstruction is considered to be correct. The condition used is

$$|\chi^2/N_d - 1| \leq 0.05, \quad (10)$$

where  $\chi^2$  is defined as in (3) and  $N_d$  is the number of detectors. To obtain tomograms with a resolution of  $135 \times 105$  some seconds are needed, considerably less time than at JET but still not applicable to real-time scenarios.

The dataset used in the following sections possesses 5784 reconstructions, belonging to pulses in the interval 9860–11150. The data is divided into 3 groups (and not 2 as previously done) for training, validation and testing according to the ratio 80%/10%/10%. This is the common procedure used for training a NN [23] and since we will be comparing the performance of a NN against the inverse matrix method we divide the dataset this way. For the matrix fitting process only the training and test set will be used.

### 3.1. Fitting the Inverse Matrix

By applying the procedures described in Section 2, one obtains the inverse tomographic matrix now for COMPASS SXR. Using  $\eta = 1$  and  $\gamma = 0.9$  after  $10^5$  iterations an error of around  $17 \text{ W m}^{-3}$  was reached. Since the error decreased less than  $10^{-6} \text{ kW m}^{-3}$  in consecutive epochs no more iterations were performed.

Comparisons between the original tomograms and the ones performed with the inverse matrix are presented in Figure 8.

The evaluation metrics (already explained in Section 2.1) for both Figure 8 and the full test set are presented in Table 2. It is also of interest to measure the absolute error of the SXR emissivity centroid, since it is closely related with the plasma center position. This measurement is done both in  $R$  and  $Z$  axis,  $e_R$  and  $e_Z$ . The  $\chi^2$  value of the new reconstructions is calculated as well, considering this is the criterion used to validate the originals.

The average values of SSIM, PSNR and NRMSE are significantly better than those obtained for JET when

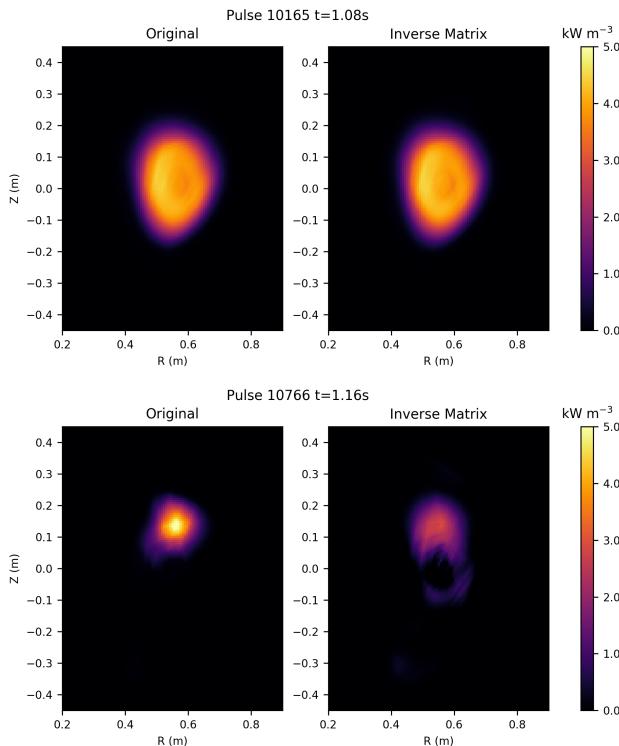


Figure 8: Comparison between original reconstructions from COMPASS SXR database and the ones done using the inverse tomographic matrix

applying the same method (see Table 1) and similar to those obtained using a NN by Ferreira et al. [15], although the standard deviation values increased. This performance improvement is mainly due the SXR profiles resemble each other more, making it easier to define one general regularization. Still, the inverse matrix struggles to capture major changes in the plasma shape (as can be seen in Figure 8). For this reason, while the average errors in the centroid position are smaller than 4 mm, in some reconstructions it can go up to several centimeters. Also, while all the original 578 reconstructions from the test set fulfilled (10), only in 63 tomograms obtained with the inverse matrix is the same condition satisfied.

#### 3.1.1 Regularization Patterns

Several of the obtained lines of sight regularization patterns are given in Figure 9. As in Section 2.1 for JET, it is possible to observe, for most of the lines of sight, what seem to be magnetic flux surfaces. For detectors pointing straight at the plasma center the pattern is not as smooth, being most of the contribution concentrated in the center. It is still observable that some regularization patterns include regions that seem to have no correlation with the corresponding line of sight.

#### 3.1.2 Full Pulse Reconstruction

Using the NumPy implementation for matrix product the time necessary to perform one reconstruction is 0.4 ms, the same as the one measured for JET in Section 2.1 due to the matrix dimensions not changing significantly.

New reconstructions for full pulses can also be computed but one must be careful to verify which ones are physically meaningful or not. For example by determining the value of  $\chi^2$  over the entire pulse 10188 (Figure 10) it can be seen that in the beginning and very end of the shot the values are specially high. A similar behavior for  $\chi^2$  is observed for other pulses.

These errors can partially be justified by the lack of training data for both the beginning and end of the shot. When generating the dataset of reconstructions for training, a threshold to disregard time-stamps with very low power measurements was used, meaning there is a lack of reconstructions for the beginning of the shot. On the other end, when very near to the disruption, the original Minimum Fisher algorithm stops converging. While there is small interest in producing

Pulse	SSIM	PSNR (dB)	NRMSE	$e_{\text{power}}$	$e_R$ (mm)	$e_Z$ (mm)	$\chi^2$
10165	0.998	47.22	0.02	0.01	0.19	0.77	0.96
10766	0.879	28.01	0.43	0.15	8.99	21.71	8.58
Full Set	$0.981 \pm 0.027$	$37.06 \pm 6.83$	$0.13 \pm 0.23$	$0.07 \pm 0.15$	$1.86 \pm 2.37$	$3.25 \pm 4.01$	$1.18 \pm 0.83$

Table 2: Evaluation metrics obtained using the inverse matrix method for the reconstructions in Figure 8 and on the full test set of COMPASS SXR tomograms

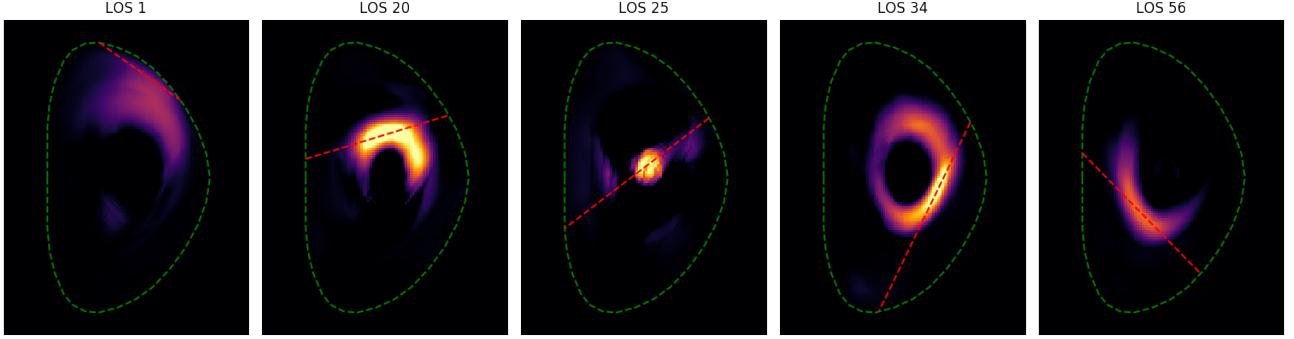


Figure 9: Example of regularization patterns obtained for different lines of sight (LOS) of COMPASS SXR setup (dashed red) with vessel contour (dashed green)

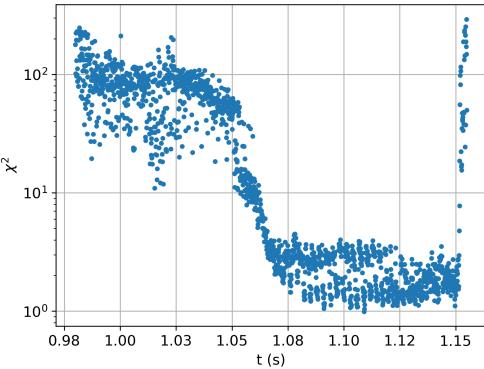


Figure 10: Variation of  $\chi^2$  obtained for reconstructions performed using the inverse matrix for COMPASS pulse 10188

tomographic reconstructions for low emissivity values, the same cannot be said for the end part of the shot, where the description of events that precede disruptions is an important field of study.

Another reason for the higher errors closer to the disruption is the assumption of a single regularization which is profoundly wrong for these time-stamps due to the significant shift of flux surfaces.

### 3.2. Fitting a Deep Neural Network

The NN developed for COMPASS SXR system (Figure 11) uses the same architecture developed for JET bolometer setup [15] but several changes in its dimensions were made. Its implementation was done using the Keras library [24] running on top of Theano.

To enable the usage of the NN for all SXR setup geometries, the number of inputs was increased to 90, since this was the maximum number of SXR detectors ever available. The size of the dense layers was reduced from 7500 to 3315 given no significant gain was obtained and more computational power was needed when using a bigger layer, while the performance decreased significantly for smaller sizes. For the same reasons the number of filters was reduced from 20 to 15. The output of the NN has a resolution of  $136 \times 104$  which differs from the tomograms original resolution

$135 \times 105$ . To solve this issue, a row of zeros is added to the original image and a column is removed. Deleting the column is not a problem since the border of the reconstruction corresponds to emissivity values outside the vessel, which are set to zero. All layers use ReLu activation functions.

The NN was fitted to the same training set used for the calculation of the inverse matrix. This training data was divided into 10 batches of size approximately 435. Instead of updating the parameters of the NN solely once per epoch (using the full training set at once), 10 updates of the optimization algorithm are now performed (one for each batch). Regarding the optimizer, the Adam algorithm [25] with a learning rate  $\eta = 10^{-4}$  was chosen.

The final parameters saved for the NN correspond to those obtained in epoch 25,682 where the minimum error of  $5.4 \text{ W m}^{-3}$  on the validation set was reached after approximately 86 hours of training.

Several comparisons between the original tomograms and the ones performed with the NN are presented in Figure 12 (same as used in Figure 8). Their corresponding evaluation metrics as well as those for the full test set are presented in Table 3. The reconstruction quality and the obtained evaluation metrics are significantly better than those achieved with the inverse matrix.

The SSIM values are practically one, the NRMSE and  $e_{\text{power}}$  became residual and the average error in the emissivity peak position in both axis is considerably smaller than the image resolution (the pixel size is  $6.7 \times 6.7 \text{ mm}$ ). Also, from the full test set of 579 events, for 445 reconstructions performed with the NN the condition (10) is satisfied (compared with 63 obtained with the inverse matrix).

Regarding the computation time needed to perform a reconstruction using the COMPASS NN a value of 1 ms was determined. This is approximately double the time necessary with the inverse matrix method (0.4 ms) and half the time needed for the JET NN (2 ms) since dimensions changed significantly.

Regarding the time-stamps in which the NN can perform physically meaningful reconstructions it was ob-

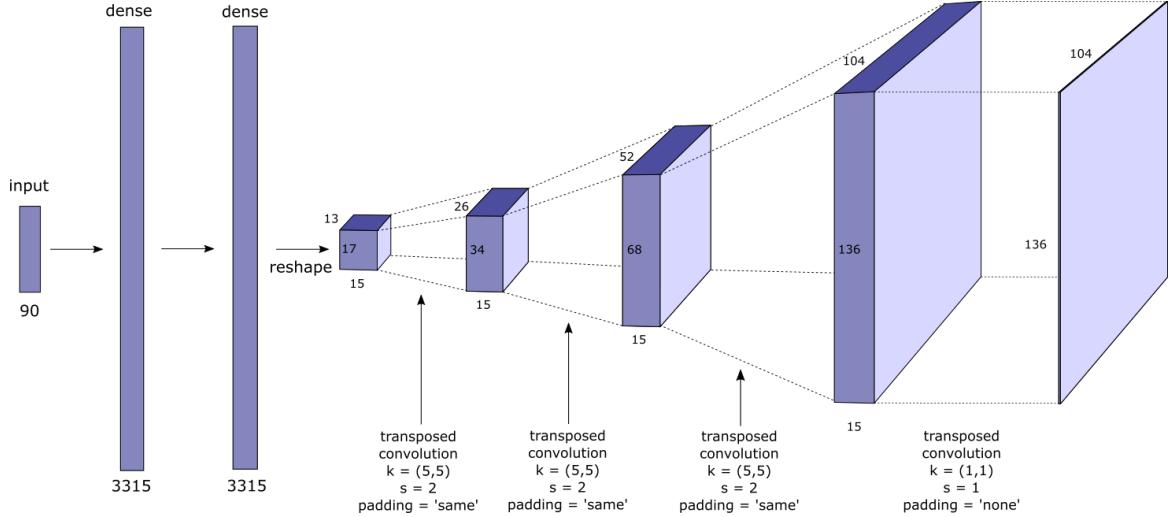


Figure 11: NN architecture used for COMPASS SXR system

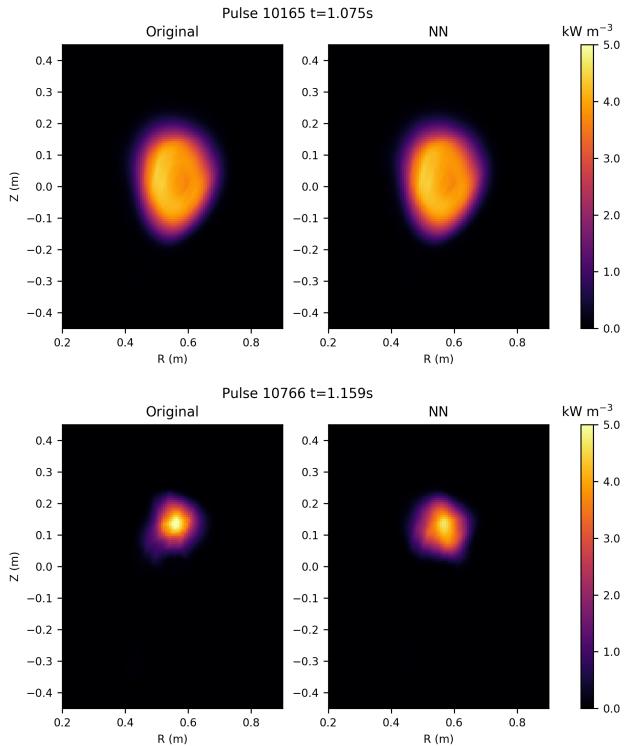


Figure 12: Comparison between original reconstructions from COMPASS SXR database and the ones done using the NN. Difference between images is calculated as original minus new reconstruction.

served that the results are better than those obtained with the inverse matrix method but the problems already explained in the beginning and end part of the shot still remain.

### 3.2.1 Handling Missing Detectors

It might happen during a shot that a certain detector is not working properly or not measuring at all. It is then of interest for the NN to cope with missing detectors without the need to retrain it every time a new detector malfunctions.

To try to solve this problem, in each training iteration, a certain amount of detectors randomly chosen can be shut-down. This process is called dropout [26] and is normally used in neurons located at deeper layers of the NN to prevent overfitting. In this scenario, one is not trying to prevent overfitting but to make the NN able to extrapolate information from missing detectors based only on those available.

To verify the performance of the resulting training, at test time all detectors are set to zero one at a time, and the metrics are calculated for all the test set reconstructions. The same procedure is also applied to all the possible combinations of 2 detectors being shutdown.

Among all tested values (0.01, 0.05 and 0.010), the dropout rate of 0.01 provided the best results. The comparison between its performance and a NN trained without dropout is presented in Table 4. The NN

Pulse	SSIM	PSNR (dB)	NRMSE	$e_{\text{power}}$	$e_R$ (mm)	$e_Z$ (mm)	$\chi^2$
10165	1.000	58.07	0.01	0.01	0.22	0.21	1.02
10766	0.973	30.77	0.31	0.22	4.11	6.89	1.84
Full Set	$0.998 \pm 0.004$	$49.96 \pm 4.63$	$0.02 \pm 0.02$	$0.01 \pm 0.01$	$0.60 \pm 0.67$	$1.00 \pm 1.08$	$1.05 \pm 0.18$

Table 3: Evaluation metrics obtained using the NN for the reconstructions in Figure 12 and on the full test set of COMPASS SXR tomograms

#Shutdown	Dropout	SSIM	PSNR (dB)	NRMSE	$e_{\text{power}}$	$e_R$ (mm)	$e_Z$ (mm)	$\chi^2$
0	0.01	$0.999 \pm 0.004$	$50.23 \pm 4.23$	$0.02 \pm 0.02$	$0.01 \pm 0.01$	$0.52 \pm 0.60$	$0.86 \pm 1.05$	$1.09 \pm 0.17$
	None	$0.987 \pm 0.029$	$41.77 \pm 8.79$	$0.07 \pm 0.06$	$0.02 \pm 0.03$	$1.60 \pm 1.80$	$2.14 \pm 2.44$	$1.64 \pm 2.60$
1	0.01	$0.996 \pm 0.025$	$48.79 \pm 6.06$	$0.02 \pm 0.02$	$0.01 \pm 0.01$	$0.55 \pm 0.63$	$0.90 \pm 1.07$	$1.09 \pm 0.32$
	None	$0.983 \pm 0.020$	$37.99 \pm 7.84$	$0.10 \pm 0.07$	$0.03 \pm 0.04$	$2.32 \pm 2.32$	$3.02 \pm 3.17$	$2.23 \pm 4.18$
2	0.01	$0.998 \pm 0.004$	$48.97 \pm 4.22$	$0.03 \pm 0.02$	$0.01 \pm 0.01$	$0.60 \pm 0.67$	$0.94 \pm 1.11$	$1.10 \pm 0.22$

Table 4: Evaluation metrics obtained for COMPASS NN after shutting down a given number of detectors, using different dropout rates at training time. The values presented are the average over all possible combinations of shutdown detectors.

trained with a dropout of 0.01 is able to maintain a similar performance for all the tested scenarios (1 detector or 2 detectors at a time being shutdown). It still generates better results, when compared to the NN trained with no dropout (Table 3), for all metrics other than  $\chi^2$ , in the event of no detector being shutdown. Higher tested values of dropout failed to obtain better results and perform even worse than the inverse matrix method.

In the case of JET NN it was found that a dropout rate of 0.05 generated the best results. This implies we should tune the dropout rate as any other parameter of the model (layer size, number of filters).

### 3.2.2 Changing Geometry

While setups like JET bolometer system have not changed their geometry for a long time, this is not the case for COMPASS SXR system. Modifications in the detectors drive the previously trained NN obsolete and there is a need to adapt it for this new geometry. One can then attempt to retrain the NN by applying several approaches:

1. Use reconstructions from the new geometry as training samples. This requires new sample tomograms to be collected, which means that it is not possible to utilize the NN from the starting point of the experiments.
2. Use past reconstructions from other geometries together with the new projection matrix  $\mathbf{T}_{\text{new}}$  to determine through (2) what would the new detectors in these past events have measured. Afterwards, one trains the NN with this new “virtual” detector values as input and the corresponding old reconstructions as output. For this to work, it is necessary for the plasma emissivity profile to have similar conditions in both the old and the new geometry.

In this work, the “old” geometry corresponds to GEOM\_201410, and the “new” one corresponds to GEOM\_201701 (see Figure 7), with a new dataset of 6728 reconstructions from pulses in the interval 13213–14884.

When using Method 1, the NN training follows the same procedure previously described in this chapter for GEOM\_201410. As expected this method provides similar results (see Table 5) to those obtained for GEOM\_201410 (see Table 3) thus demonstrating that the NN architecture is applicable to different SXR camera geometries when trained on reconstructions from the same setup.

Method 2 was firstly tested using the dataset previously used for GEOM\_201410 (which considers measurements from cameras A,B,F). The dataset was divided into 2 groups for training/validation according to a ratio of 90%/10%. On the other hand, the test set reconstructions belong to pulses where the new camera geometry is adopted, and are the same as used for Method 1. This attempt failed to obtain good results because it produces high  $\chi^2$  values (see Table 5).

Afterwards, all reconstructions for GEOM\_201410 were again computed but now without using the information from camera F. The values obtained, when training the NN with these new tomograms, are slightly better than the ones calculated using the three cameras, but still far away from the desired level attained with Method 1.

Given that the two geometries are very similar, the last result is unexpected. It was observed that the high  $\chi^2$  values are associated to the detectors from camera B with lines of sight closer to the lower border of the plasma. For the new geometry, these detectors values are always practically zero and, for this reason, the error associated is also very low. On the other hand, when looking at measures from lines of sight of the old geometry that cross approximately the same region, they are not zero. This is an issue, since the NN is

Method	SSIM	NRMSE (dB)	NRMSE	$e_{\text{power}}$	$e_R$ (mm)	$e_Z$ (mm)	$\chi^2$
1	$0.999 \pm 0.002$	$51.25 \pm 5.01$	$0.02 \pm 0.02$	$0.01 \pm 0.01$	$0.39 \pm 0.39$	$0.60 \pm 0.61$	$1.33 \pm 0.79$
2 - A,B,F	$0.967 \pm 0.014$	$31.57 \pm 2.16$	$0.17 \pm 0.03$	$0.09 \pm 0.04$	$1.79 \pm 1.51$	$5.28 \pm 3.77$	$5.49 \pm 9.51$
2 - A,B	$0.974 \pm 0.011$	$33.02 \pm 1.92$	$0.14 \pm 0.03$	$0.07 \pm 0.04$	$2.14 \pm 1.71$	$5.20 \pm 4.24$	$4.16 \pm 11.72$

Table 5: Evaluation metrics obtained for reconstructions relating to COMPASS SXR GEOM\_201701 setup. Determined values correspond to tomograms produced with COMPASS NN using different training methods.

being trained with a set of inputs that do not represent the conditions in which later the NN will be used.

General conclusions regarding the applicability of Method 2 are then difficult to produce since the starting assumption made when proposing it was that the plasma conditions are similar. In this case this was not verified (being it for actual physical reasons or a mis-calibration of the detectors).

## 4. Conclusions

The results of this work show that it is possible to reduce significantly the time necessary to perform tomographic reconstructions, while maintaining the tomograms quality, by utilizing machine learning techniques such as regularization extraction via gradient descent or Artificial Neural Networks. It is also shown that the NN architecture used can be trained and applied to different fusion devices and tomographic diagnostics by only adjusting some of its parameters (layer size, number of filters). Additionally, its robustness to malfunctioning detectors can be improved by using a dropout rate on the inputs at training time, thus adding another tunable parameter to the model.

The usage of NNs provides considerably better results than considering only one regularization, especially for events where the magnetic flux surfaces shift considerably. The main downside is that NNs take more time to train when compared to the fitting of the inverse matrix (hours/days vs. a few minutes). It also takes approximately double the time to produce a reconstruction (1 ms vs. 0.4 ms for COMPASS SXR) and NNs are more prone to overfitting (which one tries to minimize during training). Nevertheless, due to their applicability to a broader range of profiles, the expected improvements in computation time with developments in GPU technology (in both hardware and software) plus the creation and maintenance of reliable tomographic datasets, NNs should become a very effective and widely used tool for tomographic research in fusion devices. In an effort to support this statement all code developed for this work is made available at [27].

## Acknowledgements

This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014–2018 under grant agreement No 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission. IPFN activities received financial support from Fundação para a Ciência e a Tecnologia (FCT) through project UID/FIS/50010/2013. The work developed at COMPASS was funded by FuseNet Master student internship support program. The Titan X GPU used in this work was donated by NVIDIA Corporation.

## References

- [1] J. P. Freidberg, *Plasma physics and fusion energy*, Cambridge University Press (2008)
- [2] L. C. Ingesson et al., *Fusion Science and Technology* **53**, 2, 528 (2008)
- [3] J. Mlynar et al., *Journal of Fusion Energy* 1–9 (2018)
- [4] M. Anton et al., *Plasma Physics and Controlled Fusion* **38**, 11, 1849 (1996)
- [5] A. Huber et al., *Fusion Engineering and Design* **82**, 5–14, 1327 (2007)
- [6] V. Weinzettl et al., *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **623**, 2, 806 (2010)
- [7] J. Mlynar et al., *Plasma physics and controlled fusion* **45**, 2, 169 (2003)
- [8] I. H. Hutchinson, *Principles of Plasma Diagnostics*, Cambridge University Press, 2 ed. (2002)
- [9] P. Carvalho, Ph.D. thesis, Universidade Técnica de Lisboa, Instituto Superior Técnico (2010)
- [10] L. Ingesson et al., *Nuclear fusion* **38**, 11, 1675 (1998)
- [11] J. Mlynar et al., *Review of Scientific Instruments* **83**, 10, 10E531 (2012)
- [12] V. Loffelmann et al., *Fusion Science and Technology* **69**, 2, 505 (2016)
- [13] T. Odstrčil et al., *Review of Scientific Instruments* **87**, 12, 123505 (2016)
- [14] F. A. Matos et al., *Fusion Engineering and Design* **114**, 18 (2017)
- [15] D. R. Ferreira et al., *Fusion Science and Technology* 1–10 (2018)
- [16] I. Goodfellow et al., *Deep learning*, vol. 1, MIT press Cambridge (2016)
- [17] S. Ruder, arXiv preprint arXiv:1609.04747 (2016)
- [18] Theano Development Team, arXiv e-prints **abs/1605.02688** (2016)
- [19] Z. Wang et al., *IEEE transactions on image processing* **13**, 4, 600 (2004)
- [20] Q. Huynh-Thu et al., *Electronics letters* **44**, 13, 800 (2008)
- [21] F. Ebrahimi et al., in *Applications of Digital Image Processing XXVII*, vol. 5558, 300–309, International Society for Optics and Photonics (2004)
- [22] J. R. Fienup, *Applied optics* **36**, 32, 8352 (1997)
- [23] Y. LeCun et al., *nature* **521**, 7553, 436 (2015)
- [24] F. Chollet et al., Keras, <https://keras.io> (2015)
- [25] D. P. Kingma et al., arXiv preprint arXiv:1412.6980 (2014)
- [26] N. Srivastava et al., *The Journal of Machine Learning Research* **15**, 1, 1929 (2014)
- [27] D. D. Carvalho, *PlasmaTomoML Repository*, <https://github.com/diogodcarvalho/PlasmaTomoML> (2018)