T. Farley, N. R. Walkden, F. Militello, M. Sanna, J. Young, S. S. Silburn, J. Harrison, L. Kogan, I. Lupelli, S. S. Henderson, A. Kirk, J.W. Bradley

Filament identification in wide-angle high speed imaging of the Mega Amp Spherical Tokamak

Filament identification in wide-angle high speed imaging of the Mega Amp Spherical Tokamak

T. Farley, N. R. Walkden, F. Militello, M. Sanna, J. Young, S. S. Silburn, J. Harrison, L. Kogan, I. Lupelli, S. S. Henderson, A. Kirk, J.W. Bradley

Filament identification in wide-angle high speed imaging of the Mega Amp Spherical Tokamak

T. Farley,[1, 2, a)] N. R. Walkden,[1] F. Militello,[1] M. Sanna,[3] J. Young,[1] S. S. Silburn,[1] J. Harrison,[1] L. Kogan,[1] I. Lupelli,[1] S. S. Henderson,[1] A. Kirk,[1] and J.W. Bradley[2]

[1)]*EURATOM/CCFE Fusion Association, Culham Science Centre, Abingdon, OX14 3DB, UK*

[2)]*Department of Electrical Engineering and Electronics, The University of Liverpool, L69 3GJ, UK*

[3)]*Department of Electrical and Electronic Engineering, University of Cagliari, Cagliari, Italy*

(Dated: 7 May 2019)

A new tomographic inversion technique is presented for the identification of plasma filaments in wide-angle visible camera data. Direct inversion of camera data onto a field aligned basis is a poorly conditioned problem which is overcome by breaking the analysis into a 'psuedo-inversion' step followed by a 'point spread function correction' step. Camera images are first prepossessed, applying background subtraction, noise reduction and sharpening enhancements to bring out the transient filaments structures in the images. A large collections of equilibrium magnetic field lines are traced and projected onto the camera field of view. These field line images are combined to form a geometry matrix which is used to produce a pseudo-inversion which is obtained from a convolution of each individual field line image with the camera image. A tractable inversion is then performed on a point spread function matrix which is derived from the geometry matrix. The resulting 2D intensity distribution parameterised by the field line machine coordinates at the mid-plane of the machine is a good approximation of the direct inversion problem. Blobs of high intensity are identified using the 'watershed' or 'valley-filling' algorithm and 2D Gaussians are fitted to get the positions, widths and amplitudes of the filaments. A synthetic diagnostic producing artificial camera data containing experimentally representative filaments is utilised to rigorously benchmark the accuracy and reliability of the technique. 74% of synthetic filaments above the detection amplitude threshold are successfully detected and only 1.2% of detected filaments are found to be false positives. The accuracy with which filament properties and their probability density functions are recovered is discussed, along with sources of error and methods to minimise them.

I. INTRODUCTION

The competition between transport processes parallel and perpendicular to the magnetic field in the scrape-off layer (SOL) of a tokamak determines the radial profiles of variables such as temperature and density that impinge on divertor and first-wall surfaces[1]. These profiles impact many aspects of the machine operation including damage to plasma-facing components[2], fuelling[3], penetration of resonant heating waves[4], impurity accumulation[5], detachment onset[6] and tritium retention[7,8]. With this in mind, it is important that a sound understanding of perpendicular transport in the SOL is established. In the last decade or so it has become increasingly clear that a large portion of the radial flux of particles and heat in the SOL is carried intermittently by mesoscale coherent structures known as filaments (alternatively blobs, avaloids, IPOs)[9,10]. Recent forward modelling[11] on MAST has shown that particle[12] and heat-flux[13] profiles at divertor targets can be reconciled with a transport based on filament motion. The presence of filaments leads to a non-local relationship between fluxes and gradients in the SOL[14] and a fuller understanding of the physics underlying filaments is required before predictions can be made

for future devices. This requires extensive experimental measurements of filaments for comparison with modelling and the informing of analytic filament frameworks.

The bulk of past filament measurements have used Langmuir probe based techniques, analysing the characteristic non-Gaussian statistics of ion saturation current fluctuations that arise from filaments passing the probe tip, with the strongest non-Gaussian behaviour observed in the far SOL, where the relative amplitude of the filament fluctuations is largest[15]. Positively skewed fluctuation statistics are observed universally in toroidal plasma devices[10], with in depth studies on tokamaks including Alcator C-Mod[16], TCV[15], MAST[17] and JET[18], among others[19], providing information about filament amplitudes, dimensions, velocities and waiting times between filaments. With the exception of a few multi-probe studies[20], Langmuir probes, whether fixed or reciprocating, are limited to studies of filament population statistics and cannot provide information about filaments throughout their lifetimes (i.e. once they have passed the probe tip); information important for a full understanding filament dynamics. Beam Emission Spectroscopy (BES) diagnostics on NSTX[21,22], DIII-D[9,23] and TEXTOR[24], as well as Gas Puff Imaging (GPI) diagnostics on NSTX[22,25–28], Alcator C-Mod[25,29,30] and ASDEX Upgrade[31] provide 2D density profiles in the plane perpendicular to the magnetic field, from which filaments can be identified and their motion studied. While

---

these techniques provide excellent high spatial and temporal resolution measurements in the poloidal plane, they are limited to measuring a small area of the SOL and are unable to provide information about the parallel structure of filaments.

In this paper we look at the application of direct wide-angle fast visible imaging, which has the potential to provide information about the perpendicular and parallel structure of filaments across their full lifetimes.

Fast framing camera technology has improved drastically since the earliest visual observations of filaments in the 1980s[32,33]. Modern fast framing cameras are now capable of recording the whole plasma at resolutions, frame rates and exposure times sufficient for resolving filament motion. Visible imaging can now provide similar resolution data for filament positions, sizes and velocities, as other techniques, while having a number of benefits which enable the collection of larger quantities of filament data and simplify implementation. Firstly, visible imaging is a passive technique and hence can operate over a wide range of conditions, for the whole duration of a pulse, without temporal constraints or the need for perturbative probe, gas or neutral beam injection. This also means that the measurement is not spatially constrained to a small region local to a gas nozzle or beam path and can instead explore large regions of the SOL, both around the mid-plane and the divertor. The passive nature of the measurement also reduces its complexity, avoiding the complications of understanding complex non-equilibrium phenomena such as gas plume propagation or neutral beam-plasma interaction, instead relying on relatively uniform and slowly varying excitation and ionisation rates. Secondly, visible imaging can provide information about the full parallel structure of filaments, informing us about their 3D nature to a degree that other diagnostics cannot. With these strengths, fast visible imaging is well positioned to significantly improve our understanding of filaments.

Past fast camera analyses of filaments have either been performed manually, making them time consuming and subjective, or have achieved automation of the measurement of a limited subset of filament parameters. Past simultanious measurements of both toroidal and radial filament properties (positions, widths and velocities)[34–36] have involved the manual alignment of projected field lines with filamentary structures in camera images by eye; a challenging exercise owing to their translucent, diffuse nature and their interactions with one another, greatly reducing the quantity of data that has been analysed. Automated or semi-automated measurements of radial (toroidal) positions, widths and velocities have previously been made by considering the time varying intensity of specific radial[12,34,36] (toroidal[34,37]) arcs of image pixels to infer the passage of filaments along the length of the arc. These techniques rely on finding areas of an image where the camera line of sight is such that pixel coordinates along a line can approximately be mapped directly to radial (toroidal) coordinates and typically assume that only one filament is passing the arc at any one time. More so-phisticated techniques set a fixed radial position *a prioiri*, typically close to or at the separatrix, and located maxima in the average intensity of superimposed field lines as a function of their toroidal position[12,35,36,38]. This incurs uncertainties in the measured quantities (e.g. typically error in toroidal widths[35]) due to the uncertainty in the precise radial position of the filaments and fails to capture information about the radial properties (positions, widths and velocities) of the filaments. Furthermore, with the exception of Ref. 12, all applications of these autonomous or semi-autonomus techniques have focused on narrow, windowed camera views of specific areas of the plasma (e.g. the outboard mid-plane[36,38], centre column[36] or top outside corner of the plasma[34]), in order to maximise the camera's frame rate, thus limiting the number of observable filaments and restricting the length of filaments along which parallel structure can be explored.

This paper describes a new technique that has been developed to automate the identification and analysis of filaments in fast camera data, with the goal of retrieving as much individual filament data as possible. The technique is designed to simultaneously measure the positions, widths, amplitudes and velocities of individual filaments, across a large extent of the plasma. Further quantities can be derived from these measurements such as the quasi-toroidal mode number and the distributions of filament separations and waiting times. As all quantities can be measured for individual filaments, inter-relationships between different filament properties can be investigated. In this way a large database of filament properties can be generated, enabling big data approaches to the understanding of the statistical properties of filament parameters and their dependence on engineering and physics parameters. A good understanding of these statistics will enable the validation of analytic frameworks such as that developed by Militello and Omotani[11] and provide the inputs for these frameworks, greatly improving our understanding of the filamentary generation of SOL density profiles.

Section II discusses the specification and set-up of the camera system for which the technique has been developed, before Section III details the technique itself. Section IV describes the synthetic camera diagnostic which has been developed to test the technique, while Section V discusses its implementation in benchmarking the technique and quantifying the technique's systematic and random errors. Section VI demonstrates the application of the technique to experimental data and interprets the results applying insight from the benchmarking analysis. Section VII concludes the paper by summarising the key findings and giving the outlook for future refinement and applications of the technique.

II. DIAGNOSTIC SPECIFICATION

The open design of the MAST vessel facilitates deep-field, wide angle views of the MAST plasma. Two tangential viewing geometries are available corresponding
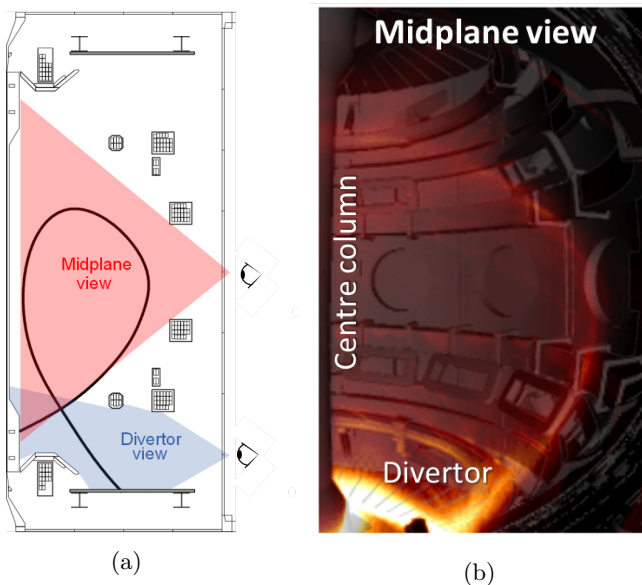
FIG. 1: a) Schematic of the field of view of the Photron SA-1 fast camera when installed at the mid-plane or divertor. b) False color image of a MAST plasma, as viewed by the SA1 camera, with a CAD rendering of the MAST vessel components overlaid.

to the 'main chamber' and 'divertor' views as shown in Fig. 1a. This paper concerns the 'main chamber' viewing geometry. Also shown in Fig. 1b is a false color image of a MAST plasma with a CAD rendering of the MAST vessel components overlaid. The center-column and P4 poloidal magnetic field coil are highlighted to guide the eye.

The camera used was a Photron SA1.1 fast framing camera[1]. The frame rate of the camera was 100 kHz with an integration time of 3 μs and a pixel resolution of 256×160, corresponding to a spatial resolution at the tangency plane of 5 mm. The camera was unfiltered such that the light emission recorded was dominated by D_α emission. The poloidal plane perpendicular to the camera line of sight falls at the toroidal angle $\phi = 215°$, where the toroidal machine coordinate ϕ is 0° at the start of the first octant and continues round clockwise.

The camera position in real space is calibrated using the Calcam[39]code which maps points selected in a 3D rendering of the MAST vacuum vessel onto 2D pixel coordinates of the camera image. This camera registration then allows for paths in real space to be cast onto the image plane of the camera. In this way, 3D magnetic information can be mapped onto the camera field of view.

---

[1] Photron SA1 specification: https://photron.com/fastcam-sa1-1

III. FILAMENT IDENTIFICATION TECHNIQUE

A. Overview and Assumptions

The goal of the technique we have developed is to take as inputs 1) camera images, 2) a camera calibration and 3) a magnetic equilibrium description and return the positions, widths and orientations of filaments within the frames.

Due to the fact that filaments are 3D objects with no exploitable toroidal or poloidal symmetry, conventional tomographic inversion techniques, such as an Abel inversion, cannot be used to study them. The technique presented here operates by using information about the magnetic geometry to restructure the 2D intensity information present in the camera images onto the horizontal R-$R\phi$ plane at the mid-plane ($Z = 0$), where R is major radius, ϕ is the angular toroidal position and Z is the vertical machine coordinate. $R\phi$ is the toroidal arc length at the outer midplane and is used in place of ϕ as the toroidal coordinate, so that both dimensions have units of length, from which meaningful filament dimensions can be found.

Two central assumptions about the nature of filaments are made in order to progress. A) Filaments are assumed to be aligned well to the background magnetic field which can be calculated via magnetic reconstruction. This is to be expected given the magnitude of parallel transport relative to perpendicular transport, which quickly spreads blobs of plasma along field lines. Secondly, B) the light emission from a filament is treated as being constant in the direction parallel to the magnetic field.

The first of these assumptions is justified by a large base of experimental data from MAST[36] alongside other tokamaks[19,40]. The second assumption contains two factors. First, filaments are assumed to be homogeneous along magnetic field lines in the camera field of view. This is justified a posteriori by observation of the fact that filaments are always observed to span the entire length of field lines within the camera's view. While some apparent parallel variation in light emission is observed, it is many times weaker than the cross-field variation. This is expected as parallel transport is very efficient on the relevant time scales and so will tend to smooth out differences along field lines. In the absence of a precise description of the parallel variation of the filaments, treating them as homogeneous is sufficient for us to proceed. The second factor is the assumption that the neutral density in MAST remains homogeneous along the length of the filament. This is motivated by the open vessel design of MAST which leads to homogenous neutral disributions away from the divertor[41]. While this assumption breaks down close to the divertor surface, we are concerned with the main chamber where this is not an issue. In principle both of these assumptions could be relaxed through suitable modifications to the technique. However the complexity this introduces was not deemed to be worthwhile at this stage. The technique also benefits from strong magnetic
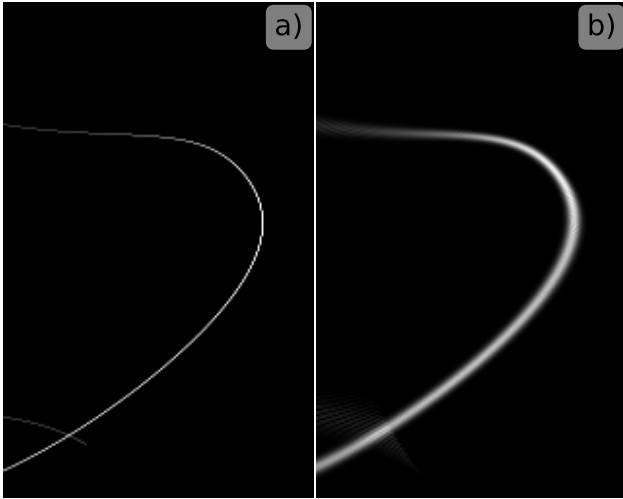
FIG. 2: a) Field line image composing a single column in the geometry matrix G_{ij}. b) Emission from a bundle of field lines with amplitudes given by a 2D Gaussian.

FIG. 3: a) An experimental camera frame from MAST shot 29852, with the pre-processing described in Section III B applied. b) Pseudo-inversion calculated using the frame in a). The Inversion in b) is also superimposed on the frame in a) to illustrate the (R, ϕ) mid-plane coordinate system.

shear in the SOL, as strong variation in the magnetic pitch angle helps to constrain the radial position of the filaments. Therefore, MAST with its tight aspect ratio and thus higher level of shear close to the separatrix than in most tokamaks, is well suited to this analysis.

Combining assumptions A and B, the filament structures observed in fast camera images can be formed by a linear superposition of images of uniformly emitting field-lines from the equilibrium magnetic field, cast onto the camera field of view. This can be expressed as a linear operation as follows using the Einstein summation convention:

$$I_i = \epsilon_j G_{ij} \qquad (1)$$

Here the vector I is the 'image vector' which is a contiguous vector containing each pixel in the camera image such that I_i is the ith pixel value in the camera image. This is a measured quantity. G is the 'field-line geometry matrix' and is constructed from the individual 'field-line images' from the chosen basis of equilibrium field lines such that G_{ij} is the ith pixel value of the image vector of the jth magnetic field line. An example of a field line image that forms a column of G is shown in Fig. 2a). G can be constructed given a description of the magnetic equilibrium and a calibration of the camera field of view. ϵ is the 'emissivity vector' and gives the contribution of each field line in the field line geometry matrix to the camera image, such that ϵ_j is the emissvity of the jth magnetic field line. This is the quantity we seek.

Fig. 2b) is an example of a reshaped image vector I resulting when Eq. (1) is applied to an emissivity vector ϵ representing a Gaussian bundle of field line emitters.

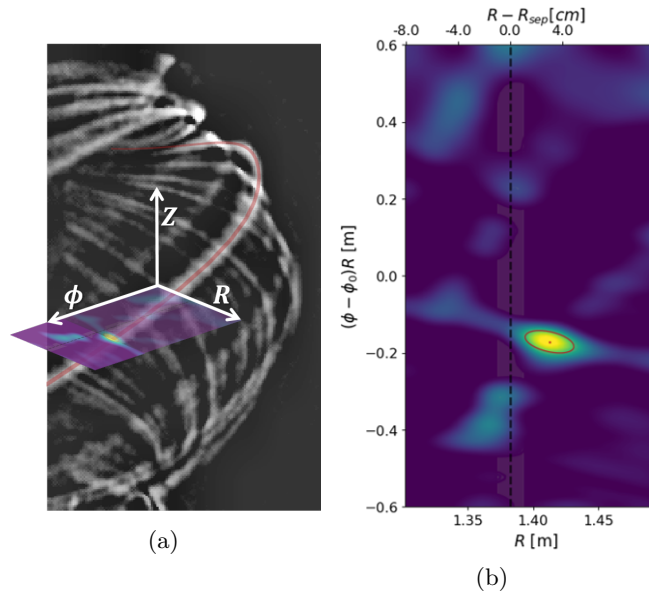Eq. (1) represents casting the information in the camera images from a pixel basis set onto a field line basis set, which facilitates the identification of field aligned plasma structures. Thus, filaments appear as blob like regions of high intensity field lines in ϵ_j, from which the filaments' sizes and locations can be deduced in a similar way as for poloidal intensity distributions produced by GPI or BES diagnostics.

Eq. (1) is an inversion problem where the goal is to find ϵ_j by inverting the matrix G_{ij}. This is usually solved as a linear least squares problem. Unfortunately, given the camera resolution, viewing geometry and density of filaments in the field of view, the problem is strongly under resolved, so that G_{ij} is so close to being singular, making the procedure intractable.

This problem is overcome by calculating a pseudo-inversion, ξ, that benefits from being robust to noise and being applicable to a subset of the toroidal and radial SOL domain. We then recover ϵ though the inversion of a point spread function (PSF) matrix which removes distortions inherent to ξ. Though this technique, inversion of the camera images becomes a tractable problem, with reliable results. Fig. 3 illustrates the toroidal emission plane onto which the camera data is inverted.

The rest of this section describes the implementation of the technique, from pre-processing of the camera images, the psuedo-inversion and PSF correction steps, to the identification of filaments within the inverted data.
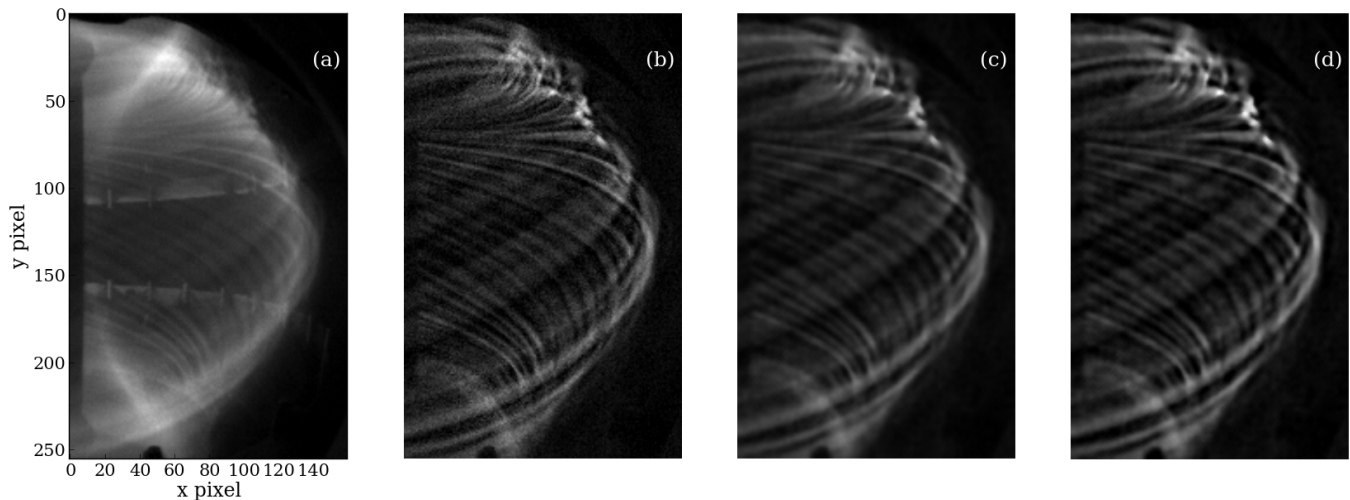
FIG. 4: a) Raw image from MAST shot 29841 at 0.22424s. b) Background subtracted image using 10 frames prior to the desired frame to construct a background. c) Application of the bi-lateral median filter. d) Application of a Gaussian de-blurr.

B. Pre-processing

A series of pre-processing steps are applied to the camera images before the extraction of filamentary structures is performed. The technique is motivated by the fact that filaments in the SOL represent positive perturbations on top of a background, therefore perturbations in the light emission are also positive, making the background the minimum in the signal over time. Thus a background subtraction technique is applied to the movie to remove the low-frequency background component of the light emission and extract just the fluctuating part containing the filaments. This step is necessary to avoid the filaments being washed out by background light emission in the SOL and prevent first wall structures confounding the analysis. The background calculation, which follows in a similar manner to that described by Dudson[35], is achieved by taking the pixel-wise minimum in the light intensity over a set of 10 frames (0.1 ms) that preceded the frame of interest in time. This technique has previously been applied in Refs. 12, 35, 36, and 38 and discussion of the validity of the technique can be found therein. Other background calculations were considered by instead taking the background as represented by either the mean, median or the low frequency (extracted from a pixel-wise FFT) component of the total emission. In the first two cases this led to non-physical negative perturbations in the SOL which could not be reconciled with Langmuir probe measurements[17]. The latter performed similarly to the minimum technique, however the calculation was considerably slower, so the minimum based background subtraction was adopted. Following background subtraction a bi-lateral median filter[42] is applied which is a gradient preserving noise removal method. The reduced dynamic range of the background subtracted images reduces the signal to noise ratio due to shot noise making noise removal particularly useful. Noise removal is followed by a Gaussian de-blur (weighted subtraction of the image convolved with a Gaussian kernel) which helps to re-sharpen the image after the smoothing of the noise removal. These two filters, while not vital elements of the technique like the background subtraction, have been found to effectively reduce the propagation of noise through the analysis. Fig. 4 shows an example of an image with each of the techniques described in this section applied in turn.

C. Inversion

The next stage of analysis is an inversion of the signal (as results from the previous three stages of pre-processing) onto real space coordinates, $\underline{\epsilon}_{R\text{-}R\phi}$.

With the previously stated assumptions in place the analysis proceeds by projecting magnetic field line trajectories onto the camera image. Magnetic field lines are traced using a 4th order Runge-Kutta integrator with the magnetic field structure provided by an EFIT++[43,44] equilibrium reconstruction. The magnetic field lines are parametrised by their launch positions at the mid-plane $(R, R\phi)$, about a central toroidal angle ϕ_0. A grid of field lines is traced describing a volume SOL in the camera field of view. The calcam code is used to project the field lines onto the camera field of view and images are produced accounting for volumetric integration effects [Fig. 2a)]. Each image is then flattened into a contiguous vector and used to form a column of G. G need only be recomputed when the magnetic equilibrium or camera calibration changes and thus once calculated can be utilised to invert a large collection of frames.

Multiplying both sides of Eq. (1) by G yields the

'pseudo-inversion emissivity vector',

$$\xi_j = G_{ij} I_i = G_{ij} \epsilon_k G_{ik} \qquad (2)$$

such that ξ_j is the pseudo-inversion intensity of the j$^{\text{th}}$ magnetic field line for the image vector I.

Figs. 5a) and 5b) show an experimental camera frame and its corresponding pseudo-inversion respectively. The pseudo-inversion, which is the convolution of each individual field line image with the camera image, has some favourable properties. As it can be calculated using a simple matrix product, it is computationally far lighter than solving Eq. (1) directly. It negates the need for a full coverage of the toroidal cross-section because the dominant contribution to the pseudo-inversion is from magnetic field lines with trajectories close to the selected one, with other field lines far from the selected one having minimal impact. This allows the use of much smaller, lower resolution matrices than before. Additionally, it has a very suppressive effect on noise in the camera image because of the inherent averaging along field lines. However, it is affected by a set of distortions that make it differ from the desired emissivity vector, ϵ, resulting from a direct inversion. These distortions occur due to line of sight effects which cause different field lines to overlap along portions of their length in the camera field of view. Overlapping of neighbouring field lines leads to smearing, while overlapping of more separated field lines can produce spurious areas of high intensity.Finally, overlapping of background filament structures outside of the inversion domain contributes low level noise in the pseudo-inversions. The frequency and extent of overlapping field lines varies throughout the camera field of view in a complex manner. Fortunately, information about the overlapping of field lines and the resulting distortions is contained in G. Taking the dot product of G with the transpose of itself gives the 'point spread function' (PSF) matrix,

$$P_{jk} = G_{ij} G_{ik}^T \qquad (3)$$

where P_{jk} gives the relative contribution of the k$^{\text{th}}$ field line to the observed emission along the j$^{\text{th}}$ field line. Given a 3D field line is represented by a point on the horizontal $(R, \phi R)$ mid-plane, the PSF represents the resulting spread in ξ, synonymous with the conventional case of a PSF relating the response of a camera to a δ-function source of light. Mathematically the PSF is the convolution of one field line with all others, expressing how inverted emission spreads onto overlapping field lines. With this quantity defined, Eq. (2) can be rearranged to give

$$\xi_j = \epsilon_k G_{ij} G_{ik}^T = \epsilon_k P_{ik} \ , \qquad (4)$$

where we exploit the fact that we can change the order of multiplication by taking the matrix transpose of the field line geometry matrix. Eq. (4) is again a standard linear inversion problem where ξ_j is a measurable, the matrix P_{ik} is pre-computable and known, and we want the unknown emission vector, ϵ_k, which is exactly the same problem as the one in Eq. (1). The advantage now is that the matrix to invert is square and much smaller and so is much more amenable to inversion and less computationally demanding. Eq. (4) is solved using the non-negative constrained SART (Simultaneous Algebraic Reconstruction Technique) inversion algorithm described in Ref. 45 with Laplacian regularisation. This produces good results after several hundred iterations which corresponds to around a couple of minutes of computation time for typically sized inversion domains. The use of Laplacian regularisation, which has the effect of smoothing the inverted emissivity, follows from the fact that in reality the basis of field-lines we choose to compose G will never exactly describe all the fine structure observed in the camera images. The effect of the PSF correction step is to reduce blurring and banding that arises from geometric line of sight effects that affect the convolution and thus it better constrains filament widths and reduces spurious regions of emission, although some artefacts do remain. The effect of the PSF correction is illustrated in Fig. 5c). Applying Eq. (1) to the emission vector represented in Fig. 5c) we can re-project the emission back into the camera plane [discussed further in Section IV] as shown in Fig. 5d) . Many of the most prominent filamentary structures in 5a) are reproduced. Due to the reduced domain of the inversion the re-projection only describes emission in that volume containing field lines that cut through the horizontal plane indicated in red, analogous to a flux-tube often used in simulations. For this reason, the re-projection does not contain structures at the upper or lower divertor regions or on the far side of the plasma. An artefact is often observed where excess emission is placed at the borders of the inversion domain, particularly visible in the far SOL. This arises due to the SART algorithm struggling to distribute emission correctly at the boundaries of the inversion domain. For this reason, filament detections are limited to the region within the dotted white border box in Fig. 5.

D. Filament Identification

To extract filament properties from the inverted emissivity vector a blob detection algorithm based on the watershed algorithm is used. 2D Gaussians are fitted to the identified blobs in order to extract the position $(R, \phi R)$, widths $(\delta_R, \delta_{\phi R})$, amplitudes (A) and orientation (θ) of the filaments.

The watershed algorithm requires the identification of marker points which are inside definite filament regions (foreground) and inside regions that are definitely not filaments (background). To produce the markers, an empirical field line emissivity amplitude threshold is set, ϵ_{thresh}, below which a significant proportion of local peaks in field line emissivity are found to originate from noise or remnant pseudo-inversion distortion effects [see white contour in Fig. 6a)]. Alternatively, the threshold can be set to a number of standard deviations above the average
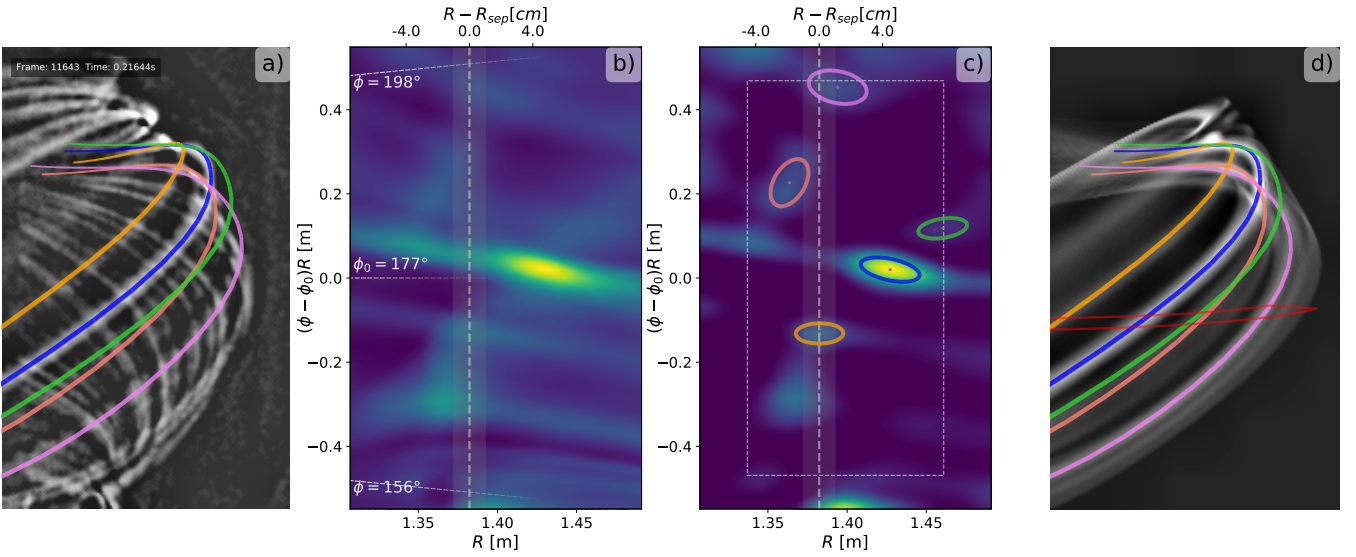
FIG. 5: a) Pre-processed camera frame with superimposed filament centre field lines. b) Emissivity data resulting from the pseudo-inversion. The range of toroidal angles covered by the inversion domain are indicated. c) Emissivity data after the PSF correction. Coloured ellipses indicate the filaments projected on the frame images. The white dashed rectangle indicates the analysis region which avoids boundary effects and the white vertical dashed line shows the location of the separatrix. d) Reprojection of the emissivity data in c). The red line indicates the inversion plane from which the emission is projected along field lines.

pseudo inversion intensity across all the frames, akin to many past measurements in the literature[9,46]. However, the first method has the benefit that all filaments above the noise floor of the technique can potentially be analysed, as opposed to imposing a somewhat arbitrary e.g. $2.5\sigma_I$ cut-off in the amplitude of the coherent filamentary structures that are analysed. Further details of the selection of this parameter are given in Section V D. Clusters of high intensity field lines with a peak amplitude above this intensity threshold, are considered to be filament candidates. Dilated regions around all 2D local minima and local maxima with intensities below ϵ_{thresh} are labelled as background (not filaments). Dilated regions around 2D local maxima with intensities above ϵ_{thresh} are labelled as foreground (potential filament centres). With these regions as inputs [Fig. 6b)], the watershed algorithm sorts the remaining unassigned regions into background [purple regions in Fig. 6c)] or distinct foreground regions [coloured regions in Fig. 6c)]. The watershed algorithm can be pictured in terms of different coloured pools of water originating at the distinct foreground markers. The water in each pool is allowed to rise and spread according the terrain of the data until it comes into contact with a background marker or water of a different colour. On contact with another region, the area assigned to that colour (filament) is frozen.

These regions are filtered according to a number of conditions. First, the contours are assessed on their solidity (area of the contour/area of its convex hull), which is a measure of how uniformly convex the contour is and thus how elliptical it is in shape. Contours with a low

solidity (typically $< 60\%$), are rejected as their irregular shape indicates they cannot be described well by a fitted 2D Gaussian and they may be the results of interacting filaments which the watershed algorithm was unable to separate. Next contours are rejected if they extend to the edge of the analysis grid and have significant intensities there (typically $> 50\%$ of the peak contour amplitude) as these contours extend outside of the analysis grid and thus are missing information required for an accurate Gaussian fit. Finally, any excessively large regions that are too large to be a filament ($\delta_R > 15$ cm, $\delta_{R\phi} > 20$ cm) are rejected.

The remaining contours are fitted with tilted 2D Gaussians with a background level set to the average intensity of the local minima in the inversion. The Gaussian fits yield the filaments central positions, 1σ major and minor axes widths, amplitudes and inclinations [see contours in Fig. 6d)]. A Gaussian description of the filaments is chosen because it is the simplest fitting function that produces a representative width and amplitude for the observed blob-like data. The combination of line of sight integration effects and limited spatial resolution mean it is not possible to reliably discern any more complex internal filament structure.

The watershed algorithm has a number of important advantages over previously tested algorithms that relied upon contours generated from thresholding the intensity data at a suitable level identified for each peak. Firstly, it handles interacting filaments very effectively, preventing them from being merged into a single larger blob and avoiding problems from multi-modal emissivity struc-
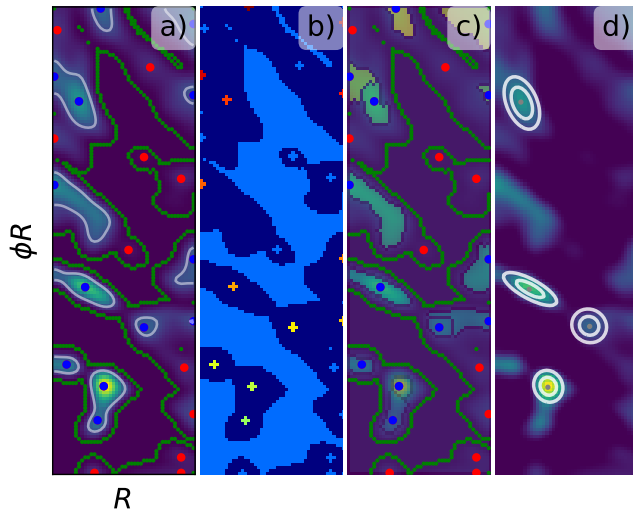
FIG. 6: a) Emissivity data, with the minimum peak blob intensity threshold indicated as a white contour. Local maxima above and below the threshold are indicated as blue dots and red dots respectively, while local minima are indicated in green. b) Markers given as input to the watershed algorithm. Light blue is definite background (not a blob), dark blue areas are yet to be classified by the watershed algorithm and all other colours indicate foreground regions (sources for potential blobs). c) Coloured regions show contours returned by the watershed algorithm. Coloured points are as in a). d) Contours of 2D Gaussians fitted to the data in the coloured regions shown in c). Some regions have been filtered out using a range on criteria and thus do not have fits.

tures producing multiple contours for a given filament. Secondly, it efficiently maximises the amount of relevant intensity data assigned to each filament, with each filament assigned to the largest possible surrounding region out to the closest local minima in intensity or the closest neighbouring filament (whichever is closest). When combined with the fitting of 2D Gaussians to the contours, this means that large secluded filaments can utilise intensity information spanning the whole structure leading to particularity high accuracy fits. In the case of small, interacting filaments this makes the detection of the filaments possible by relying on the upper sections of the amplitude distributions, which would not have been correctly identifiable by the previously tested algorithms, due to the absence of the required contour level (e.g. 50%) within the filament's enclosing contour.

In Fig. 5c) the detected filaments are indicated by coloured ellipses showing the 1σ contours of the fitted Gaussians. The central field lines for each fitted filament are indicated by coloured lines of the same colour in Fig. 5a) and 5d).

Geometric line of sight effects from wide angle imaging
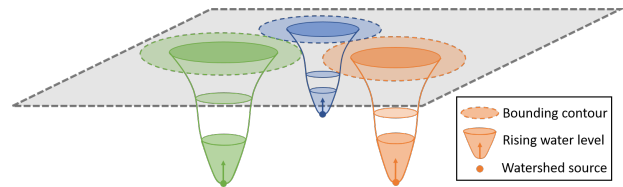

FIG. 7: Illustration of the watershed algorithm, with different coloured water sources rising until they touch and define a bounding contour.

lead to some complications correctly identifying filaments and extracting their properties from the inverted emission data. Being a spherical tokamak, the high level of magnetic shear in MAST produces stronger variation in field line pitch angle with R than in most machines, aiding constraint of the filaments' radial positions. However, the pitch angle variation is still sufficiently small across the SOL, that constraining the radial width of filaments is particularly challenging. When considering the camera images, this translates to variation in the radial position of a field line from the centre of a filament, still resulting in much of the field line overlaying high intensity regions of the filament in the camera field of view. This means that field lines positioned radially either side of a filament have raised intensities despite being outside the filament's flux tube, resulting in radial stripes of high intensity in the inversion. Examples of these can be seen as near horizontal bands in Fig. 5 b) and c). We term these radial bands the 'shadows' of the filaments. These shadows give the filaments enlarged footprints, particularly exaggerating the radial widths of the filaments. The PSF correction step significantly reduces the shadows, but does not eliminate them. In addition to increasing the widths of the detected filaments, in some cases, the shadows of multiple filaments can interact to form new local maxima in intensity which are labelled as filaments by the technique. We term these spurious detections 'shadow filaments'. Shadow filaments arise from the overlapping of filaments in the camera's field of view, such that a third unrelated magnetic flux tube aligns well with a portion of the resulting structure, producing high average emission along those field lines despite there being little genuine localised emission along those field lines. Therefore, shadow filaments typically occur at lower or higher major radii than either of the true filaments from the shadows of which they form.

Bright spots in the camera images, typically occurring above the mid-plane where sections of field lines are nearly tangential to the camera field of view, can lead to further enlarging of a filament's shadow in both the radial and toroidal directions by raising the average intensities of all field lines that pass through the bright spot. Such filaments are particularly susceptible to producing shadow filaments.

The green filament in Fig. 5 is an example of a shadow filament arising from the overlapping of sections of the red and blue filaments in the camera line of sight. In
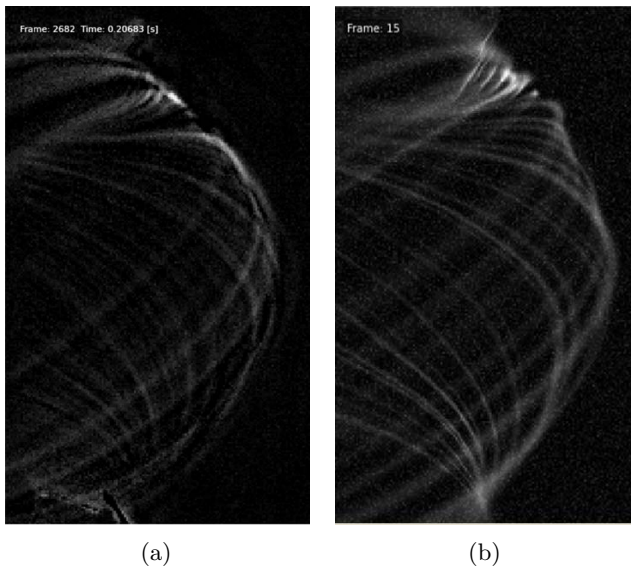
FIG. 8: a) Experimental camera frame from MAST shot 29852, with the preprocessing described in Section III B applied. b) Frame from the reference synthetic data set produced using the synthetic diagnostic, containing 40 synthetic filaments, with positive definite Gaussian noise applied.

this example the position of the violet filament has also been slightly perturbed by strong emission where several filaments overlap.

Given these complicating factors, it is important to benchmark the technique thoroughly and assess its accuracy, limitations and errors. The remainder of the paper discusses the methodology and results of how this has been achieved.

IV.   SYNTHETIC DIAGNOSTIC

A synthetic camera diagnostic has been created to aid in the development and benchmarking of the identification technique, as well as facilitating the forward modelling of simulation outputs for comparison with experimental measurements (as applied in 47). The principle is the converse of the inversion, instead calculating I from Eq. (1) for a predefined ϵ and G.

The geometry matrix is calculated with a higher resolution than for the inversion procedure in order to avoid aliasing effects in the resulting images. The product of G with ϵ, produces a synthetic image, as viewed by the camera, of the local field aligned emission at the mid plane, as described by the emission vector. This emission is equivalent to that in background subtracted experimental camera data. In principle the emission vector can be any flattened 2D pattern parameterised by the major radius, R, and the toroidal position, $R\phi$, at the mid plane. If an experimentally derived emission vector is used as input we produce a synthetic re-projection of the inversion data,

demonstrating the visual form of the structures that are captured in the emission vector as shown in Fig. 5d). When it is used to forward model simulation results[47], the ADAS atomic physics library[48] is used to calculate the light emission for the output distributions of temperature and density, producing the required vector. In cases where entirely artificial filament images are produced for the purposes of benchmarking and technique optimisation, each filament's emission profile at the mid-plane is set to a 2D Gaussian. Images of synthetic filaments can be combined additively to produce full synthetic frames. In order to closely match experimental data, positive definite Gaussian noise can be added to the frames to simulate noise from the camera sensor. As for the inversion process, the synthetic diagnostic only produces images of field-aligned structures and cannot produce variations in emission along field lines in its current form. Output from the synthetic diagnostic is compared to experimental data in Fig. 8, showing a strong qualitative and quantitative resemblance. The upper and lower divertor regions where some differences are seen are outside of the analysis region that is inverted.

V.   BENCHMARKING

As discussed in Section III C, geometrical distortions that occur in the inversion process can lead to errors and false detection of shadow filaments, complicating the interpretation of the outputs of the technique. Benchmarking is required to quantify and understand these effects in order to draw strong conclusions from future applications of the technique. By analysing large quantities of synthetic data and matching detected filaments returned from the technique to input synthetic filaments and comparing their properties, the accuracy and reliability of the technique can be assessed.

A.   Synthetic Data Set

A reference synthetic data set, $A_{exp,40}$, of 5000 synthetic fast camera frames was generated for this study with 40 synthetic filaments per frame, corresponding to a total of 200,000 filaments with random positions and characteristics. A range of numbers of filaments per frame were explored and 40 filaments per frame was chosen as it corresponds to a commonly observed toroidal quasi-mode number (average instantaneous number of filaments in the SOL) seen in MAST[12] and produced synthetic frames which visually resembled experimental camera data. The value chosen lies at the upper end of literature values for quasi-toroidal mode number[12,35]. This is appropriate given that literature values are likely to have missed the very faintest filaments which are challenging to measure, but which are important to include in the synthetic data. Furthermore, the potential overlap of filaments means the apparent quasi-mode number is likely to appear lower

than the input mode number.

The synthetic filaments have Gaussian cross-sections with their properties (positions, widths and amplitudes) drawn from realistic distributions based on experimental findings. The details of the chosen distributions are summarised in Appendix B. The radial distribution of the filaments follows a log-normal distribution[27,28] starting at the separatrix ($R = 1.36$ m) and peaking 3 cm into the SOL. Analysis of experimental camera data with the inversion technique has shown an exponential distribution of toroidal filament separations, indicating that filaments are randomly distributed toroidally around the machine with no mode number (hence the use of the term quasi-mode number to refer to the average number of filaments). Therefore, we have adopted a uniform distribution for the toroidal locations. The radial and toroidal widths are drawn from log-normal distributions informed by the width distributions observed in Refs. 12 and 36. Filament amplitudes are distributed exponentially as observed in Refs. 18, 30, and 49.

The radial extent of the analysis region was $R = 1.290$-1.474 m, spanning the outer mid-plane SOL and including several centimetres inside the separatrix, given the MAST mid-plane separatrix was located at $R \approx 1.36$ m for this discharge. The analysis region extended toroidally from $\phi R = -0.8$ m to $\phi R = +0.8$ m, loosely corresponding to an angular range in machine coordinates of $\phi = 153°$-$215°$ centred about $\phi_0 = 177°$ (c.f. camera tangency plane at $\phi = 215°$). Thus the analysis region focuses on foreground filaments (i.e. those that intersect the mid-plane in the foreground). Foreground filaments are chosen due to their proximity to the camera increasing their size in the camera field of view and thus maximising the accuracy of their width measurements. Of the 200,000 filaments in the data set, 20,243 (10.1%) fall within the analysis domain, corresponding to an average of 4.05 visible filaments per frame.

Toroidal positions and widths are generated as angles (ϕ and δ_ϕ) and converted to lengths ($R\phi$ and $\delta_{R\phi}$). The analysis grid had a resolution of 4 mm in R and 9 mm in $R\phi$, corresponding to 47 radial points and 123 toroidal points, giving a total of 5,781 distinct field lines. This resolution was chosen so as to best constrain the radial and toroidal widths of the filaments, while keeping the computational burden manageable.

Given the decaying exponential amplitude distribution peaks at zero, many of the synthetic filaments will be very faint, as may also be true in the experimental case, and therefore cannot reasonably be expected to be detected against background noise. However, inclusion of these very faint background filaments, while not currently a focus of measurement, is important for reproducing the full complexity of the likely distributions of the experimental data.

In order for the amplitudes of the synthetic filaments to be consistent with those in experimental data, the synthetic images are scaled so that histograms of the pixel intensities in the synthetic images are consistent with those from experimental frames. This calibration is important for results from synthetic data at a given filament amplitude threshold to translate well to experimental measurements.

B. Matching synthetic and detected filaments

The goal is to classify the detected filaments from the technique as either 'true positives' that can be assigned to a true input synthetic filament or 'false positives' that cannot be matched to a true filament.

The method by which the set of detected detected filaments, D_{all} and the set of synthetic synthetic filaments S_{all} are compared and matched is as follows.

1. The synthetic filaments are vetted into a subset S_{domain} of filaments inside the analysis region [white dashed box in Fig. 5c)] and a smaller set of filaments, S_{analyse}, that both 1) lie within the analysis domain and 2) have amplitudes above the detection threshold, ϵ_{thresh}.

2. The detected filaments are examined in order of decreasing amplitude. Each detected filament is compared to each unmatched synthetic filament.

   (a) If a synthetic filament centre lies within 1σ of the Gaussian fitted to the detected filament, the input and detected filaments are added to the sets of matched filaments S_{matched} and D_{matched} respectively.

   (b) If multiple synthetic filaments are within 1σ, the highest amplitude is taken as the 'match' and the rest are considered to 'overlap' with the 'matched' filament as they are too close to it to be distinguished.

3. Synthetic analysis filaments that are neither matched nor overlapping are labelled as 'missed' and added to S_{missed}. Detected filaments that are inside the analysis region and are not matched are labelled as 'false' detections and added to set D_{false}.

During this matching procedure a number of edge cases are also handled which influence the figures of merit which are discussed shortly. If a detected filament within the analysis region is matched to a synthetic filament outside the analysis set (i.e. a synthetic filament with an amplitude slightly below ϵ_{thresh}, or located just outside the edge of the analysis region), then the synthetic filament is added to S_{analyse} and the filaments are considered matched as above.

If a synthetic filament in S_{analyse} is matched to a detected filament outside the analysis region the synthetic filament is removed from S_{analyse} as it does not manifest in the analysis domain and thus would not be observed experimentally.

Finally, with all filaments assessed for matches, we define the set $S_{\text{isolated}} = S_{\text{analyse}} - S_{\text{overlap}}$, which is the set of filaments that 1) are in the analysis domain, 2)

have amplitudes above ϵ_{thresh} and 3) can be resolved individually. This is the set of filaments we wish the technique to identify as accurately and reliably as possible.

Fig. 9 shows a frame illustrating each of the categories of synthetic and detected filament (typical frames do not display this complexity). Synthetic filaments are indicated by dashed ellipses and detected filaments by solid ellipses. Matched filaments are shown in green ($S_{\mathrm{matched}}/D_{\mathrm{matched}}$) and false ($D_{\mathrm{false}}$) or missed ($S_{\mathrm{missed}}$) detections are shown in red. Overlapping synthetic filaments (S_{overlap}) are shown in blue, while synthetic filaments with low amplitudes or that are outside the analysis domain (i.e. filaments not in S_{analysis}) are shown in light pink.

C. Benchmarking Figures of Merit

In order to quantify the performance of the technique, a number of numerical figures of merit (FoM) have been developed. These can be used to objectively assess the effects of changes to the technique, informing further development of the technique as well as giving metrics for benchmarking overall success. Each type of FoM is described below.

a. Sensitivity   We define the sensitivity FoMs (also known as recall) as

$$f_{\mathrm{sens,domain/isolated}} = \frac{|S_{\mathrm{matched}}|}{|S_{\mathrm{domain/isolated}}|} \times 100\%. \quad (5)$$

$f_{\mathrm{sens,domain}}$ gives the proportion of all true filaments in the analysis domain that are detected, while $f_{\mathrm{sens,isolated}}$ gives the proportion of isolated analysis filaments that are detected.

b. Precision   We define the precision FoM as

$$f_{\mathrm{prec}} = \frac{|D_{\mathrm{matched}}|}{|D_{\mathrm{matched}}| + |D_{\mathrm{false}}|} \times 100\%. \quad (6)$$

This is the proportion of all detections that are true positives.

The above FoMs are termed *identification* FoMs as they describe how effectively filaments are identified with the technique.

c. Parameter fractional errors   By calculating the average absolute fractional differences between the measured and synthetic filament parameters the technique can be tuned to minimise errors individually in the position, width and amplitude of the filaments. The percentage error in a generic parameter p can be expressed as

$$f_{\mathrm{p}} = \left\langle \frac{|p_{\mathrm{measured}} - p_{\mathrm{input}}|}{p_{\mathrm{input}}} \right\rangle_{\mathrm{matched}} \times 100\%. \quad (7)$$

Parameter fractional errors are calculated for filament position, widths and amplitudes.

The FoMs calculated using Eq. (7) are termed *measurement* FoMs as they concern the accuracy with which each parameter is measured.
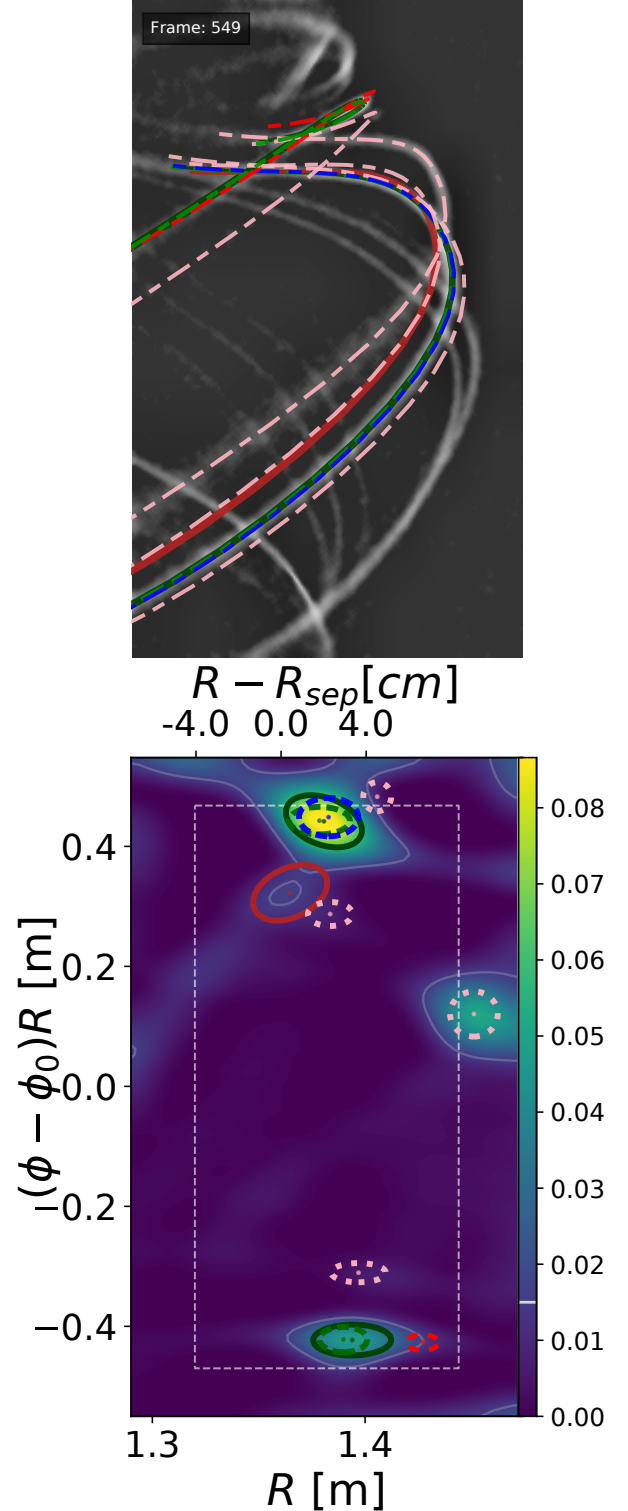

FIG. 9: a) Synthetic camera frame and b) corresponding inversions with detected (synthetic) filaments represented by ellipses with solid (dashed) lines. Matched filaments are shown in green, false (missed) filaments in red, overlapping filaments in blue and filaments outside the analysis region (white dashed rectangle) or with low amplitudes in pale pink. The white contour and colour bar indicate $\epsilon_{\mathrm{thresh}}$.

D. Filament amplitude detection threshold

The parameter ϵ_{thresh}, determining the minimum intensity of filaments detected using the technique is chosen so as to maximise the precision and sensitivity figures of merit, while sampling a sufficiently large portion of the total filament population. The first function of ϵ_{thresh} is to ensure the exclusion of small local maxima in the inversion below the noise floor in the technique. However, further increases in ϵ_{thresh} yield improvements in the FoMs due to the exclusion of spurious shadow intensity structures in the inversion, arising from the interactions of successively larger synthetic filaments. At large ϵ_{thresh} only the largest filaments in the population remain in the analysis set, leading to increased sensitivity since the detectable objects are brighter and hence more distinct. The precision also increases since only spurious structures from interactions between the very strongest filaments remain, leading to fewer false positives. Fig. 10 shows how much of the total filament population ($f_{\text{sens,domain}}$) is excluded for successive gains in precision (f_{prec}) and sensitivity to the targeted filament population ($(f_{\text{sens,isolated}})$) as ϵ_{thresh} is increased.

The optimal value, $\epsilon_{thresh} = 1.5 \times 10^{-2}$, is chosen, as above this value there are relatively modest gains in precision and target group sensitivity. At this level, 36% of the total filament population lies above the amplitude threshold, of which 72% are successfully detected, while only 1.2% of detections are false positives. These FoM values are summarised in Table III.

The dependence of $f_{\text{sens, domain}}$ on ϵ_{thresh} is broadly exponential, rolling over slightly at low amplitudes as the noise floor of the technique is approached. The distribution peaks at 58%, indicating that, for a decaying exponential amplitude distribution, over half of the total filament population is potentially detectable (i.e. above the noise floor of this technique), with the drawback that at this amplitude threshold there are almost equally as many false positives ($f_{\text{prec}} \approx 55\%$). The precision saturates around $\epsilon_{thresh} = 2.0 \times 10^{-2}$, above which which nearly all detections are true positives.

E. Breakdown of filament detections

Table I gives a breakdown of the number of filaments that are assigned to each category by the matching algorithm for the reference synthetic data set, analysed with $\epsilon_{thresh} = 1.5 \times 10^{-2}$. Around 10% of all the synthetic filaments spawned around the full toroidal extent of the machine fall within 46° under analysis and within the borders of analysis domain. 50% of those filaments have amplitudes above ϵ_{thresh}, while 49% both have amplitudes above ϵ_{thresh} and do not overlap with higher amplitude filaments. Around 1% of filaments within the analysis domain are obscured by higher amplitude filaments, which present themselves together as a single local maxima in intensity. While these underlying filaments, obscured by
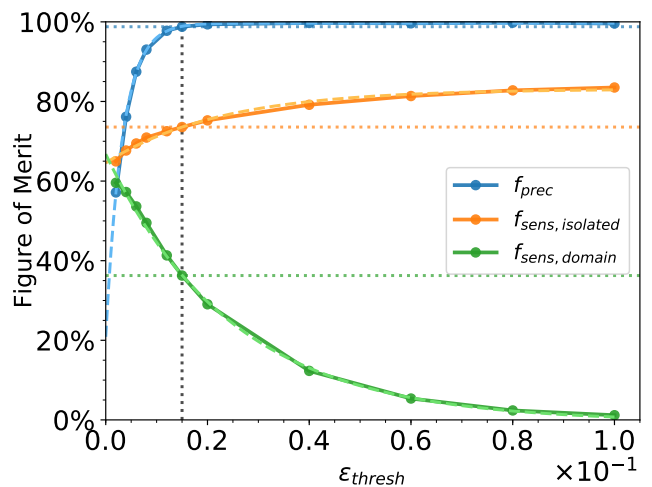

FIG. 10: Dependence of the precision (blue), analysis sensitivity (orange), and total sensitivity (green) FoMs on filament amplitude detection threshold, ϵ_{thresh}. Dashed lines show the FoM values achieved at the value of $\epsilon_{thresh} = 1.5 \times 10^{-2}$ used in the analysis.

TABLE I: Breakdown of the numbers of synthetic filaments in each analysis category for the reference synthetic data set, analysed with $\epsilon_{thresh} = 1.5 \times 10^{-2}$.

Set S_{x}	$|S_{\text{x}}|$	$\frac{|S_{\text{x}}|}{|S_{\text{domain}}|}$	$\frac{|S_{\text{x}}|}{|S_{\text{analysis}}|}$	$\frac{|S_{\text{x}}|}{|S_{\text{isolated}}|}$
S_{domain}	20243	100%	-	-
S_{analysis}	10134	50%	100%	-
S_{isolated}	9978	49%	98%	100%
S_{matched}	7341	36%	72%	74%
S_{missed}	2637	13%	26%	26%
S_{overlap}	156	0.8%	1.5%	-

larger filaments cannot be detected, they raise the amplitude of the larger detected filaments. However, we do not expect these filament interactions to occur experimentally.

Table II gives a breakdown of the detected filaments. Of the 7441 detections only 91 were false positives that did not match to a synthetic filament.

TABLE II: Breakdown of the numbers of detected filaments in each analysis category for the reference synthetic data set, analysed with $\epsilon_{thresh} = 1.5 \times 10^{-2}$.

Set D_{x}	$|D_{\text{x}}|$	$\frac{|D_{\text{x}}|}{|D_{\text{domain}}|}$
D_{domain}	7432	100.0%
D_{matched}	7341	98.8%
D_{false}	91	1.2%

TABLE III: Summary of the figures of merit from analysing the reference synthetic data set with $\epsilon_{thresh} = 1.5 \times 10^{-2}$. The first three FoMs should be maximised and the remainder minimised. Values prior to the systematic offset correction are given in brackets. $f_{R-R_{sep}}$ has been normalised relative to the 5 cm SOL width between the LCFS and the first limiting structure.

Quantity, p	Error		FoM
	μ_p	σ_p	f_p
f_{prec}	-	-	98.8%
$f_{\mathrm{sens,isolated}}$	-	-	74%
$f_{\mathrm{sens,domain}}$	-	-	36%
$R - R_{sep}$ [cm]	0.32	0.14	4% (7%)
$R\phi$ [cm]	-0.25	0.18	15% (22%)
δ_R [cm]	0.56	0.27	33% (70%)
$\delta_{R\phi}$ [cm]	1.1	0.73	26% (67%)

F. Error Distributions and Corrections

For each detected filament, it is possible to calculate the absolute error on a parameter p, as defined by $\xi_p = p_{detected} - p_{synthetic}$. Since different filaments can have different errors, the latter are statistically distributed. Assessing the average errors, $\mu_{\xi p}$ identifies systematic errors arising from distortions in the inversion which can be corrected for through offsets in future measurements. Assessing the standard deviation in the errors, $\sigma_{\xi p}$, quantifies the random errors that remain after the systematic errors are corrected for.

Fig. 11 shows the resultant distributions of absolute errors, ξ_p, on the position and width measurements for all matched output (S_{matched}) filaments from the reference synthetic data set. Each parameter shows a broadly Gaussian distribution of errors over at least an order of magnitude in frequency, although long low frequency tails are present. The mean and standard deviation of the fitted Gaussians are listed in Table III. The table also gives the f_{p} values given by Eq. (7) for each parameter (note $f_{R-R_{sep}}$ has been normalised relative to the 5 cm SOL width between the LCFS and the first limiting structure).

The average errors in the R and ϕR positions are small, at 0.3 cm and -0.2 cm respectively. The widths of the Gaussian fitted to the positional error distributions are also small, each around 0.2 cm, although the raised wings and long, largely symmetric tails with Fisher kurtosis values of 8.9 and 12.3 respectively, indicate that rare large errors of greater than 1 cm can occur with greater frequency than suggested by the Gaussian widths (excess kurtosis is the fourth standardised moment which describes tail extremity and is equal to zero for a Gaussian distribution). Almost 3% of R positions having errors greater than 1 cm (5σ) and around 13% of ϕR positions having errors greater than 1 cm (5σ). The high accuracy of the position measurements is important for accurate
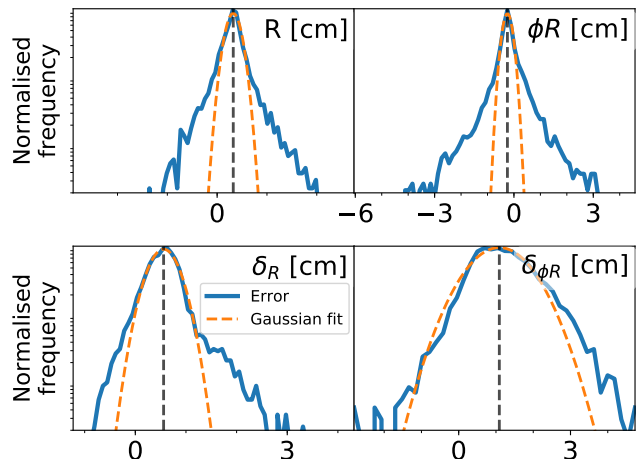

FIG. 11: Distribution of errors in detected filament parameters for radial position, toroidal position, radial width and toroidal width (left to right, top to bottom respectively). The solid blue lines shows the distributions of errors, while the dashed orange lines shows Gaussian fits to the distributions.

filament velocity measurements, corresponding to velocity errors of the order 0.3 km/s in both radial and toroidal directions. For a typical filament travelling radially at 0.7 km/s and toroidally at 4 km/s, this corresponds to percentage errors of 40% and 7% respectively. The systematic errors in the width δ_R and $\delta_{\phi R}$ are more significant at 0.6 cm and 1.8 cm respectively. These over estimations of widths arise from a number of factors, most importantly, the finite analysis grid resolution, overlap of filaments and distortions during the inversion process. The error distributions are described well by the fitted Gaussian, with widths of 0.3 cm and 0.7 cm respectively, although again the tails of the distributions are slightly more pronounced, with kurtosis values of 7.9 and 1.1 respectively.

With the systematic errors in these parameters, $\mu_{\xi p}$, identified, they are applied as systematic error offset corrections in the analysis that follow, so that measurements should only be subject to random measurement errors.

G. Position and Width Distribution Measurements

The ability to accurately measure the statistical properties of filaments is essential to build reliable first principle models and to assess their intermittent loads the plasma facing components. We now wish to understand how well our technique is able to reproduce the distributions of the synthetic filaments.

Fig. 12 compares the frequency density distribution functions of filament positions and widths for the original synthetic filaments (S_{domain}, orange dashed lines) and the detected output filaments (D_{domain}, black solid lines). The detected values have been corrected by the
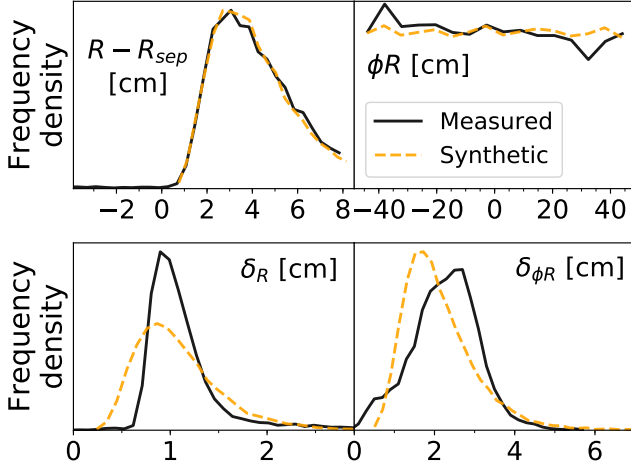
FIG. 12: Comparison of input and output frequency density distributions of filament positions and widths for the reference synthetic data set. Distributions of synthetic input filaments are given by orange dashed lines and detected distributions by black solid lines.
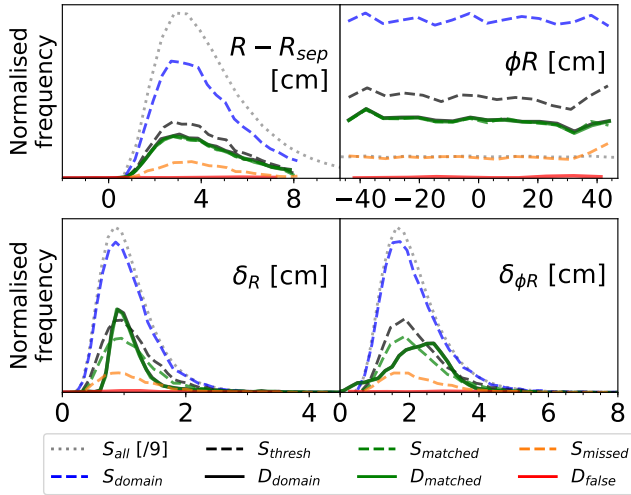

FIG. 13: Comparison of input and output normalised frequency distributions of filament positions and widths for the reference synthetic data set. Dashed and dotted lines indicate sets of synthetic filaments, while solid lines indicate sets of detected filaments recovered from the analysis procedure. The frequencies for the full set of synthetic filaments (dotted grey line) have been divided by 9 to fit on the same axes.

systematic offsets identified in the previous section. The distributions functions of radial and toroidal positions are very well recovered, closely following the synthetic distributions. The measured width distributions preserve the overall shape of the original distributions, although the distribution of δ_R is more peaked than the original distribution and the output distribution of $\delta_{\phi R}$ has its peak shifted to the larger widths by around 0.6 cm. Thus

we can have confidence that measured distributions of filament positions and widths accurately reflect the forms of the true underlying physical distributions.

The distributions of matched (S_{matched}) and missed (S_{missed}) filaments collapse onto the overall distribution of synthetic filaments (i.e. differ only in amplitude and thus can be scaled to match one another), showing that filaments are not selectively detected and thus the full extent of the distributions are evenly sampled in the detections. This is strong evidence that filaments have the observed properties and are not for instance all a single size. Similarly the distribution of matched detections (D_{matched}) collapses onto the overall distributions of detections. The distributions of false positive detections (D_{false}) show separate structure. False detections are over twice as likely to occur at low R values, which is anticipated since line of sight effects mean shadow filaments typically occur at lower R values than the two true filaments which overlap to produce them. The distribution of false positive detections has two peaks in ϕR located around $\phi R =$ -20 cm and +30 cm, indicating locations where line of sight effects are particularly prone to producing shadow filaments. Both width distributions have higher relative false positive rates for larger width filaments, especially in the case of $\sigma_{\phi R}$. Fig. 13 shows the normalised frequency distributions of the original and measured filaments for each of the matching categories. Dashed and dotted lines show the original distributions, while solid lines show the measured distributions. The dotted grey lines show the distributions for all 200,000 synthetic filaments in the data set distributed around the full toroidal extent of the machine (divided by 9 to fit on the same axes), while all other lines refer to sets of filaments within the analysis domain. Here the proportion of filaments in each matching category is clear, demonstrating how few filaments are missed or false detections.

H. Amplitude Distribution Measurements

In order to compare the synthetic and detected filament amplitude distributions the synthetic amplitudes must be mapped to corresponding amplitudes in the inversion data. Fig. 14 shows the inversion amplitude measured at the true centre of each synthetic filament in the reference synthetic data set as a function of the uncalibrated synthetic input amplitude. While there is some scatter resulting from the overlap of filaments and distortions in the inversion, there is a strong linear dependence with a Pearson correlation coefficient of 0.79. Excluding high amplitude outliers, the gradient of a linear fit to the data yields the calibration factor required to compare input amplitudes and inversion amplitudes directly.

Fig. 15 compares the distributions of detected filament amplitudes and calibrated synthetic amplitudes, using line styles consistent with Fig. 13. The vertical grey line indicates the amplitude threshold $\epsilon_{thresh} = 1.5 \times 10^{-2}$, below which detections are not attempted. The output distribu-
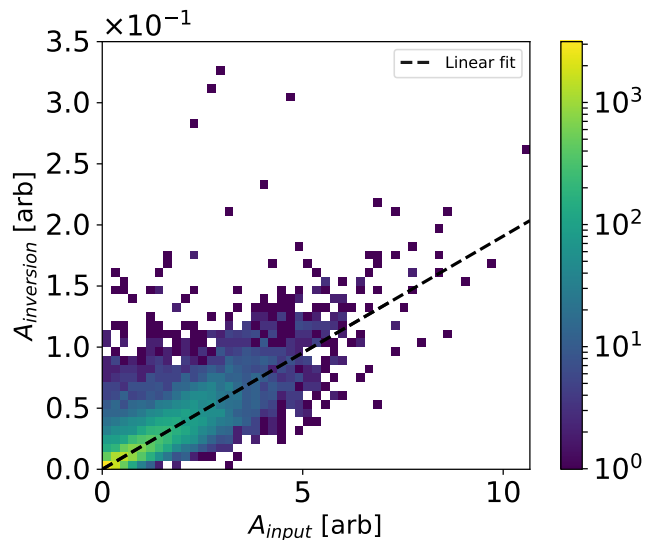
FIG. 14: Relationship between synthetic input field line emission amplitude and resulting inversion field line emission amplitude.

tion successfully captures the exponential nature of the input distribution, but for a roll over at low amplitudes, arising from from the increased proportion of missed filaments at low amplitude. The majority of false positive detections also occur at low amplitudes. The matched synthetic filaments curve continues below ϵ_{thresh} due to the subset of synthetic filaments with amplitudes below ϵ_{thresh} that manifest in the inversion with amplitudes above ϵ_{thresh} and are detected.

In order to assess the fidelity of the amplitude measurements and ensure that measured exponential amplitude distributions are not an artefact of the measurement technique, other data sets with different amplitude distributions have been analysed, each containing 5000 frames.

Data set $A_{exp,1}$ shares the same experimentally representative parameter distributions as the reference synthetic data set, $A_{exp,40}$, but contains only one filament per frame. This data set is used to assess the effect of filament overlap on results from the reference dataset.

Data set $A_{\delta,1}$ contains a single randomly positioned filament per frame, each with the same fixed amplitude and filament size. This heavily simplified data set collapses the input amplitude and width distributions onto delta functions, so that the width of the recovered amplitude distribution is solely due to the intrinsic inaccuracies in the analysis technique. The output amplitude distribution them provides an estimate of the minimum intrinsic random error on an amplitude measurement.

Data set $A_{\delta,40}$ contains the same distributions $A_{\delta,1}$, but with 40 filaments per frame, so that the effects of filament overlap can be readily identified. Further details of the filament parameter distributions used in these data sets are given in Appendix B.

Fig. 16 shows the calibrated synthetic amplitude dis-

tributions for each of these data sets (dashed lines) and the inversion amplitudes measured at the true centre of each filament in the analysis domain (solid lines). In the case of the simple $A_{\delta,1}$ data set, the input delta function in amplitude is broadened, resulting in a Gaussian width of around 15% of the mean.

When multiple filaments are introduced, as in the $A_{\delta,40}$ data set, interactions between filaments further broaden the measured distribution, giving a Gaussian width around 25% of the mean. There is also a slight shift to higher amplitudes and an enlargement of the high amplitude tail. Filament overlap effects can be divided into two categories. Firstly, filaments with large physical separations can overlap in the camera's line of sight. This will occur experimentally. Secondly, the synthetic datasets do not preclude the physical overlap of filaments. This is not anticipated experimentally. Therefore, a component of the high amplitude tail and broadening will be exaggerated in these results.

In the case of the single filament, physically representative data set, $A_{exp,1}$, the shape of the measured amplitude distribution closely follows the input distribution. This indicates that in the absence of filament overlap, amplitude distributions are recovered reliably. The introduction of filament overlap in $A_{exp,40}$ introduces a constant positive amplitude offset and a roll over at low amplitudes. The constant offset occurs due to overlap with foreground and background filaments, raising the average intensity of each projected field line. As the filaments are uniformly distributed around the machine, on average they constitute a largely uniform background of emission in the images. However, variation in field line intersection in different parts of the image due to line of sight effects will lead to some dispersion in the background offset. It is this dispersion that produces the roll over in measured amplitudes. Given the roll over occurs around the blob detection amplitude threshold and the constant offset does not affect the distribution shape, these effects should have little impact on experimental conclusions about measured amplitude distributions.

VI. EXPERIMENTAL PARAMETER DISTRIBUTIONS

To give an example of the potential of our new technique, we now apply it to a real discharge carried out on MAST. The analysis was performed on 4000 frames from discharge 29852, over the time window 0.205-0.245 s, with $\epsilon_{thresh} = 0.015$ and the systematic offset corrections applied as determined by the synthetic filament study. This was a double null discharge, with a plasma current of $I_p = 700$ kA and an injected neutral beam power of $P_{NBI} = 2$ MW.

Fig. 17 shows the measured distributions of filament positions and widths for the experimental data set in black. Orange dashed lines have been added to sketch, without any pretence of rigour, possible distributions for each parameter that can be inferred given the findings
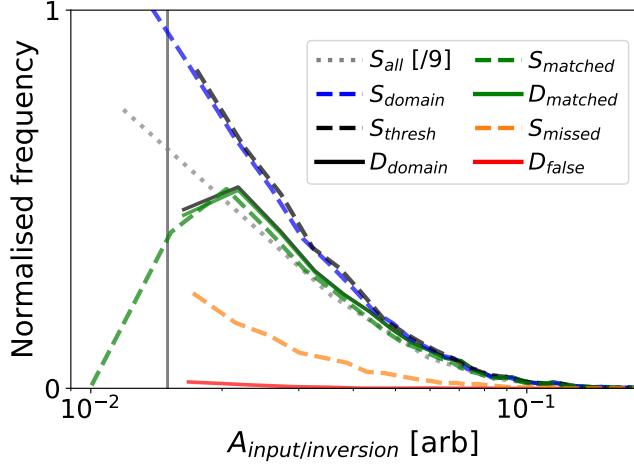
FIG. 15: Input and output amplitude distributions. Line styles are consistent with Fig. 13. The vertical grey line indicates the amplitude threshold $\epsilon_{thresh} = 1.5 \times 10^{-2}$.
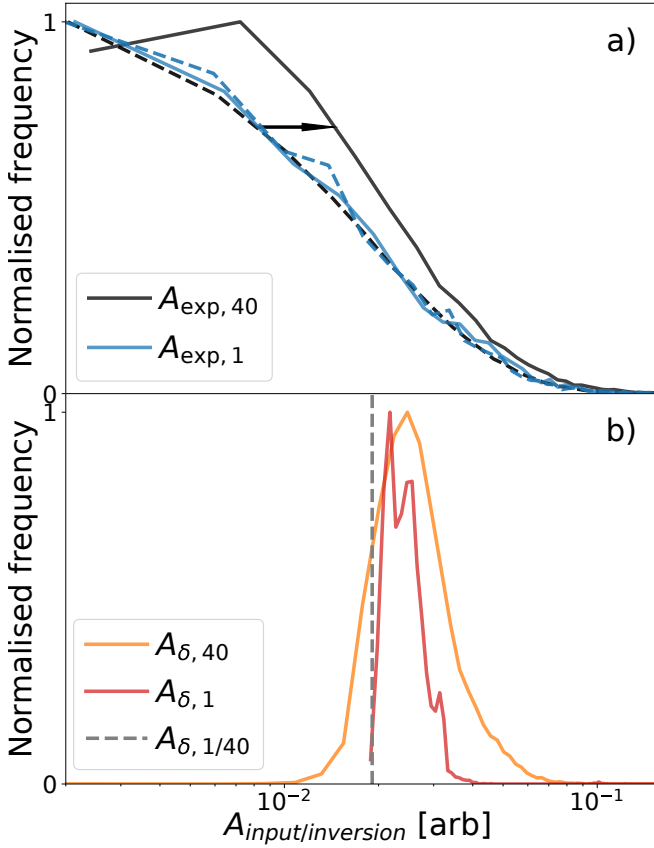

FIG. 16: Calibrated original and measured inversion amplitude distributions for a) datasets $A_{\mathrm{exp},40}$ and $A_{\mathrm{exp},1}$ with exponential amplitude distributions and b) datasets $A_{\delta,40}$ and $A_{\delta,1}$ with delta function amplitude distributions. Dashed lines are the calibrated input amplitude distributions, while solid lines show the distributions of inversion amplitudes. The vertical dashed grey line indicates the delta function in input amplitudes for the the $A_{\delta,40}$ and $A_{\delta,1}$ datasets.
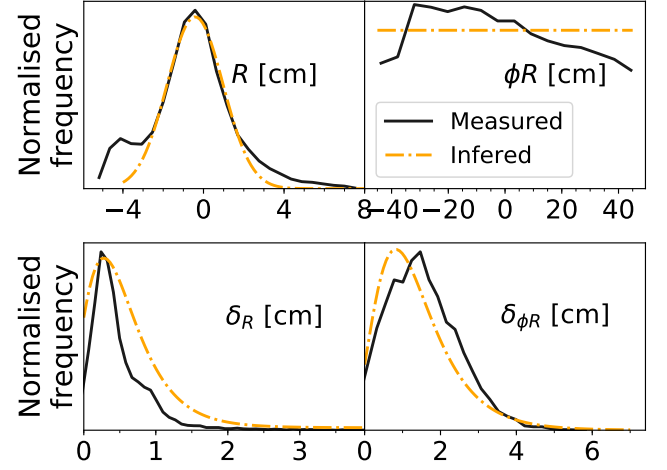

FIG. 17: Distributions of filament parameters for experimental data from MAST shot 29852, t=[0.205,0.245] s, with $\epsilon_{thresh} = 1.5 \times 10^{-2}$. Solid black lines show the measured distributions with the systematic error offset correction applied. Dashed orange lines are sketches of the inferred most probable true distributions shapes given the insight gained from the synthetic analysis.

from the synthetic data analysis. A bi-modal distribution is recovered for the R positions, with the peak at low major radius likely due to shadow filaments which have their highest relative frequently at low R. The primary peak is described well by a log-normal distribution as observed previously and assumed for the synthetic data sets. The peak in detections occurs at the separatrix where both the density of filaments and neutral particals is high. Inside the separatrix the detection rate falls of sharply due to the rapid fall off in neutral density (neutral ionisation mean free path at the separatrix is ~ 1 cm). A reduced detection rate is observed at large ϕR which is significantly stronger than is seen in the synthetic data set, but is still believed to be a diagnostic artefact from line of sight effects. The region of increased detection density at low ϕR does not coincide with the highest region of false detections in the synthetic data, so it is likely the increased detections at low ϕR are a result of greater sensitivity rather than increased false positives. Both radial and toroidal width distributions have similar shapes to the those measured from the reference synthetic data set, indicating the widths are well described by log-normal distributions. Fig. 18 shows the measured distributions of filament amplitudes for the experimental data set in black, again with orange dashed lines added to sketch out possible distribution that can be inferred from the synthetic data analysis. An exponential amplitude distribution is observed with a roll over around ϵ_{thresh}, supporting previous findings.
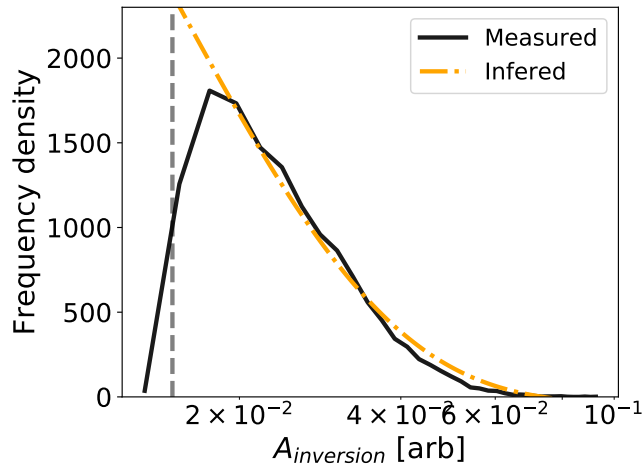
FIG. 18: Distributions of filament amplitudes for experimental data from MAST shot 29852, t=[0.205,0.245] s, with $\epsilon_{thresh} = 1.5 \times 10^{-2}$. The solid black line shows the measured distribution. The dashed orange line is a sketch of the inferred most probable true distributions shapes given the insight gained from the synthetic analysis. The vertical dashed line indicates ϵ_{thresh}.

VII. DISCUSSION

The fast camera data processing technique discussed here has been thoroughly investigated to understand its accuracy and reliability. A number of further factors should be considered that affect the precision and reliability of the technique.

The grid resolution of field lines used in the geometry matrix plays a critical role in the precision of technique's output. This must be sufficiently high so as to resolve the smallest filaments of interest, which are of the order of several millimetres across and in general a few ion Larmor radii. However, the nature of the PSF is such that it enlarges structures in R-$R\phi$ space, such that, provided these effects are accounted for as in the systematic error correction, the grid resolution can be slightly courser than the structures being measured. The resolution that can practically be used is limited by the computational time per analysis filament and the extent of the analysis region. It is desirable that the analysis region be large, so as to both make best use of the available data by analysing as many filaments as possible so as to minimise statistical error from small sample sizes and maximising the proportion of data that is not affected by boundary effects (which scales as the product of the grid dimensions). The grid resolution used in Section V was chosen so as to achieve an equitable balance, such that the necessary features could be resolved, without increasing the computation cost to an extent that it would prohibitively limit the amount of data that could be analysed.

The camera spatial resolution, limited by its pixel reso-

lution and viewing perspective, also limits the accuracy of the inversion. Filament widths in the camera field of view are often only of the order of a few pixels across over much of their length, leading to pronounced discretisation of the intensity distribution across their width. This leads to sharp discontinuities in pixel intensity along field lines, particularly in areas where the field lines pass at 45° to the image axes. This can result in artificial inhomogeneities along field lines and striations in inversion intensity, which complicate identification. While the camera sensor has a 12 bit bit-depth (0-4095), the OpenCV image processing library used in the prepossessing stages of the analysis require the images be converted to 8 bit bit-depth (0-255). This reduction in dynamic range propagates through to the inversion and reduces the fidelity of filament contour selection. If the dynamic range reduction can be avoided in future, this should lead to a slight increase the accuracy of width measurements, particularly for small filaments.

At 100 kHz the camera's integration time is 3.3 μs, during which exceptionally fast filaments[9] with total velocities of $\sim$ 4 km/s can be expected to move up to 1.3 cm. Kirk et al.[12] has previously investigated the effects of exposure time on filament measurements under similar conditions and has found it not to be significant.

A limitation of the synthetic error analysis is that the images are analysed with the same magnetic equilibrium and camera calibration descriptions, which ensures a strong mapping between the inversions and synthetic images. The magnetic equilibrium and camera calibrations will, by contrast, not be exact mappings in the case of experimental data. Thus the synthetic analysis indicates how well the technique can perform with optimal equilibria and camera calibrations. Any inaccuracies in these inputs will lead to further inaccuracies in the inversions and inferred filament properties. The effects of these sources of error will be assessed in a follow up publication.

The technique has thus far been developed looking at L-mode filaments as a first case. This is because they are brighter and more frequent than inter-ELM filaments[36], yet avoid the higher energies and currents involved in ELMs, in which significant ballooning is expected to occur, distorting the filament's local magnetic field from that of the magnetic equilibrium and weakening the assertion that filaments are well aligned to the known magnetic equilibrium (in the absence of reliable non-equilibrium magnetics data). However, this should not prohibit applications to inter-ELM and ELM filaments and these will be explored in future work.

The technique will first be applied across the large archive of existing fast camera footage from MAST, before being applied to MAST-U, which will be equipped with two Photron SA-X2 cameras. MAST-U studies will provide an excellent opportunity to understand the influence of alternative divertor configurations on filamentary transport. The technique should also be imported to other machines, provided that good equilibrium reconstruction and sufficiently bright filaments are present.

VIII. CONCLUSIONS AND OUTLOOK

A novel technique has been detailed for the processing of visual imaging data and the identification of plasma filaments moving in the region of the edge and scrape-off layer above the X-point (i.e. upstream), under the assumptions that filaments are strongly aligned to the equilibrium magnetic field and are relatively homogeneous along their length. The technique has been benchmarked using synthetic camera data, containing filaments with known properties. This has enabled the precision, reliability and limitations of the technique to be assessed, in order to properly inform experimental applications and further development of the technique. Of those filaments in the analysis region of interest with sufficient amplitudes, the technique has been shown to successfully identify 74%, with only 1.2% of detections being false positives. Standard errors on filament's radial and toroidal positions are around ± 2 mm, while standard errors on radial and toroidal widths are around ± 3 mm and ± 7 mm respectively.

Given a sufficiently large sample, the measured filament data has been shown to qualitatively reproduce features of the input distributions. An active area of interest for future development of the technique is in the use of convolution neural networks for identification of filaments in inverted camera data, which has the potential to overcome geometric line of sight effects and significantly improve detection accuracy[50]. Application of the technique to stereoscopic fast camera data shall also be performed to assess the effects of viewing angle and explore inversion techniques independent of the assumption that filaments are strongly field aligned.

The next step will be to apply the technique to large quantities of past camera data, in order to compose a large database of filaments and their properties. The technique can be readily extended to include tracking of filaments in order to provide information about filament radial and toroidal velocities. This should provide the breadth and quantity of data necessary to perform detailed statistical analyses of filament properties to understand their roles in shaping SOL profiles and their dependence on engineering and physics quantities. While the identification technique assumes filaments are homogeneous along field lines, once the filaments have been found their parallel structure in the images can also be investigated, opening up many new possibilities.

ACKNOWLEDGMENTS

Appendix A: Alternative Intensity Metrics

Section III C describes the standard inversion technique used in this paper. However, if in Eq. (4), instead of taking the dot product of G with both sides of Eq. (1), we take the Hadamard (element-wise) product we instead produce a 'weighted field line matrix',

$$F_{ij} = G_{ij} \circ I_i = G_{ij} \circ \epsilon_k G_{ik}, \qquad (A1)$$

such that F_{ij} is the intensity of the ith pixel in the camera image multiplied by the corresponding intensity of the ith pixel in the field line image of the jth field line. The result is a convolution of the field line image with the camera image and effectively picks out the intensity in the camera image that lies only along the jth field line.

The familiar pseudo-inversion vector used above can be recovered by taking the sum over each column of F_{ij},

$$\xi_j = \sum_i F_{ij} . \qquad (A2)$$

The primary advantage of taking the sum of the pixel intensities along each field line is that, as a linear operation, it preserves the ability to perform the subsequent PSF inversion correction step to reduce the distortion of the pseudo-inversion. The use of a linear field line intensity metric also has the benefit of preserving relative filament amplitudes.

However, representative quantities other than the sum can be used to provide useful information for correctly identifying and measuring field aligned plasma structures. Many quantitative measures of field line intensity can be conceived of and have been tested in order to most reliably identify filaments. Each have different strengths for constraining different filament properties, but are often accompanied with a range of weaknesses and limitations.

The geometric mean (G.M.), $(\prod_{i=1}^n F_{ij})^{\frac{1}{n}}$, which penalizes small values in the data series more effectively than the arithmetic mean (A.M.), $\frac{1}{n} \sum_{i=1}^n F_{ij}$, is effective at differentiating field lines that consistently lie on top of filaments from those that do not. While it cannot benefit from a subsequent PSF correction due to the non-linearity of the operation, the G.M. produces good identification results with far less computational expenditure, at the expense of applying non-linear scalings to measured amplitudes and widths. Of all the functional forms investigated, the geometric mean squared (G.M.^{2}) of the intensity along the field lines has been found to have the best all round properties for constraining the identification of filaments. However, metrics such as percentiles in the distribution of intensities along a field line have potential for identifying shadow filaments. This is due to their greater sensitivity to drops in intensity at the cross over of a shadow field line between different filaments overlapping in the cameras line of sight, which have

TABLE IV: Distribution information for filament parameters in synthetic data sets $A_{exp,40}$ and $A_{exp,1}$.

Property		Distribution	Parameters	Refs.
Position	R	$\lambda\mathcal{L} + \nu$	$\mu = 0$, $\sigma = 0.5$ $\lambda = 0.04$, $\nu = 1.36$	[27, 28]
	ϕ	$\lambda\mathcal{U} + \nu$	$\lambda = 355$, $\nu = 2$	N/A
Width	δ_R	$\lambda\mathcal{L}$	$\mu = 0$, $\sigma = 0.4$ $\lambda = 0.01$	[12, 36]
	$\delta_{\phi R}$	$\lambda\mathcal{L}$	$\mu = 0$, $\sigma = 0.4$ $\lambda = 0.8$	[12, 36]
Amplitude	A	$\mathcal{E}$	$\mu = 0$, $\sigma = 0$	[18, 30, 49]

TABLE V: Distribution information for filament parameters in synthetic data sets $A_{\delta,1}$ and $A_{\delta,40}$.

Property		Distribution	Parameters
Position	R	$\lambda\mathcal{L} + \nu$	$\mu = 0$, $\sigma = 0.5$ $\lambda = 0.04$, $\nu = 1.36$
	ϕ	$\lambda\mathcal{U} + \nu$	$\lambda = 355$, $\nu = 2$
Width	δ_R	δ	$\mu = 1.5$
	$\delta_{\phi R}$	δ	$\mu = 1.5$
Amplitude	A	δ	$\mu = 0.02$

little impact on the average intensity along the shadow field line. Therefore, it may be beneficial to combine information from multiple field line metrics, with some utilised to identify filaments and reject false positives as reliably as possible (e.g. non-linear metrics), while others are utilised to measure the parameters of the identified filaments (e.g. linear metrics).

These metrics are of particular interest as multiple channel inputs for convolutional neural network filament identification algorithms. These algorithms' performance can improve dramatically with a broader range of input information, so these applications will be explored further and will be the subject of a future publication.

Appendix B: Synthetic camera datasets

Four sysnthetic datasets each consisting of 5000 frames were used in this work. In the Tables IV and V, $\mathcal{U}$ represents the uniform distribution, $\mathcal{L}(x; \sigma, \mu) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right)$ the log-normal distribution, $\mathcal{E}(x; \frac{1}{\beta}) = \frac{1}{\beta} \exp\left(-\frac{x}{\beta}\right)$ the decaying exponential distribution and $\delta(x; \mu) = \delta(x - \mu)$ represents the Dirac delta distribution. λ and ν are used as scaling and offset parameters respectively.

a. *Reference Synthetic Data Set* The reference synthetic data set is the primary synthetic dataset used in this paper. The distributions for each filament parameter

are chosen so as to be best representative of experimental data [see Table IV]. Each frame contained 40 filament, giving good resemblance to experimental data.

b. *Single filament physical data set* The $A_{exp,1}$ data set was produced with the same filament parameter distributions as the reference synthetic data set, but with only one filament in each frame, each of which were ensured to be within the inversion domain. Having only a single filament in each inversion removes the effect of filament overlap and interation.

c. *40 filament simplified data set* The $A_{\delta,40}$ data set was produced with the same spacial distribution of filaments as in the reference synthetic data set, but the amplitude and width distributions were replaced by delta functions [see Table V].

d. *Single filament simplified data set* The $A_{\delta,1}$ data set was produced with the same filament parameter distributions as the $A_{\delta,40}$ data set, but with only one filament in each frame.

[1] P. Stangeby, *The Plasma Boundary of Magnetic Fusion Devices*, edited by P. Stott and H. Wilhelmsson (IOP Publishing, London, 2000).

[2] W. Fundamenski, Journal of Nuclear Materials **390-391**, 10 (2009).

[3] A. Loarte, B. Lipschultz, A. Kukushkin, G. Matthews, P. Stangeby, N. Asakura, G. Counsell, G. Federici, A. Kallenbach, K. Krieger, A. Mahdavi, V. Philipps, D. Reiter, J. Roth, J. Strachan, D. Whyte, R. Doerner, T. Eich, W. Fundamenski, A. Herrmann, M. Fenstermacher, P. Ghendrih, M. Groth, A. Kirschner, S. Konoshima, B. LaBombard, P. Lang, A. Leonard, P. Monier-Garbet, R. Neu, H. Pacher, B. Pegourie, R. Pitts, S. Takamura, J. Terry, E. Tsitrone, t. I. S.-o. L. Group, and Diver, Nuclear Fusion **39**, S203 (1999).

[4] J. R. Myra and D. A. D'Ippolito, AIP Conference Proceedings **1187**, 153 (2009).

[5] B. LaBombard, A. Kuang, D. Brunner, I. Faust, R. Mumgaard, M. Reinke, J. Terry, N. Howard, J. Hughes, M. Chilenski, Y. Lin, E. Marmar, J. Rice, P. Rodriguez-Fernandez, G. Wallace, D. Whyte, S. Wolfe, and S. Wukitch, Nuclear Fusion **57**, 076021 (2017).

[6] S. Krasheninnikov, A. Pigarov, and D. Sigmar, Physics Letters A **214**, 285 (1996).

[7] B. Lipschultz, X. Bonnin, G. Counsell, A. Kallenbach, A. Kukushkin, K. Krieger, A. Leonard, A. Loarte, R. Neu, R. Pitts, T. Rognlien, J. Roth, C. Skinner, J. Terry, E. Tsitrone, D. Whyte, S. Zweben, N. Asakura, D. Coster, R. Doerner, R. Dux, G. Federici, M. Fenstermacher, W. Fundamenski, P. Ghendrih, A. Herrmann, J. Hu, S. Krasheninnikov, G. Kirnev, A. Kreter, V. Kurnaev, B. LaBombard, S. Lisgo, T. Nakano, N. Ohno, H. Pacher, J. Paley, Y. Pan, G. Pautasso, V. Philipps, V. Rohde, D. Rudakov, P. Stangeby, S. Takamura, T. Tanabe, Y. Yang, and S. Zhu, Nuclear Fusion **47**, 1189 (2007).

[8] J. Roth, E. Tsitrone, A. Loarte, T. Loarer, G. Counsell, R. Neu, V. Philipps, S. Brezinsek, M. Lehnen, P. Coad, C. Grisolia, K. Schmid, K. Krieger, A. Kallenbach, B. Lipschultz, R. Doerner, R. Causey, V. Alimov, W. Shu, O. Ogorodnikova, A. Kirschner, G. Federici, and A. Kukushkin, Journal of Nuclear Materials **390-391**, 1 (2009).

[9] J. A. Boedo, D. L. Rudakov, R. A. Moyer, G. R. McKee, R. J. Colchin, M. J. Schaffer, P. G. Stangeby, W. P. West, S. L. Allen, T. E. Evans, R. J. Fonck, E. M. Hollmann, S. Krasheninnikov, A. W. Leonard, W. Nevins, M. A. Mahdavi, G. D. Porter, G. R. Tynan, D. G. Whyte, and X. Xu, Physics of Plasmas **10**, 1670 (2003).

[10] D. A. D'Ippolito, J. R. Myra, S. J. Zweben, D. a. D'Ippolito, J. R. Myra, S. J. Zweben, D. A. D'Ippolito, J. R. Myra, and S. J. Zweben, Physics of Plasmas **18**, 060501 (2011).

[11] F. Militello and J. T. Omotani, Nuclear Fusion **56**, 104004 (2016), arXiv:1604.08790v1.

[12] A. Kirk, A. J. Thornton, J. R. Harrison, F. Militello, N. R. Walkden, the MAST Team, and the EUROfusion MST1 Team, Plasma Physics and Controlled Fusion **58**, 085008 (2016), arXiv:1602.03021.

[13] A. J. Thornton, G. Fishpool, A. Kirk, the MAST Team, the EUROfusion MST1 Team, t. E. M. M. Team, and t. E. M. M. Team, Plasma Phys. Control. Fusion **57**, 115010 (2015).

[14] O. Garcia, R. Pitts, J. Horacek, A. Nielsen, W. Fundamenski, J. Graves, V. Naulin, and J. J. Rasmussen, Journal of Nuclear Materials **363**, 575 (2007).

[15] J. P. Graves, J. Horacek, R. a. Pitts, and K. I. Hopcraft, Plasma Physics and Controlled Fusion **47**, L1 (2005).

[16] B. LaBombard, R. L. Boivin, M. Greenwald, J. Hughes, B. Lipschultz, D. Mossessian, C. S. Pitcher, J. L. Terry, and S. J. Zweben, Physics of Plasmas **8**, 2107 (2001).

[17] F. Militello, P. Tamain, W. Fundamenski, A. Kirk, and V. Naulin, Plasma Phys. Control. Fusion **55**, 025005 (2013), arXiv:arXiv:1305.5064v1.

[18] N. R. Walkden, F. Militello, J. Harrison, T. Farley, S. Silburn, and J. Young, Nuclear Materials and Energy **12**, 175 (2017).

[19] G. Y. Antar, G. Counsell, Y. Yu, B. Lombard, P. Devynck, B. Labombard, and P. Devynck, Physics of Plasmas **10**, 419 (2003).

[20] S. H. Müller, A. Diallo, A. Fasoli, I. Furno, B. Labit, and M. Podest, Physics of Plasmas **14** (2007), 10.1063/1.2813193.

[21] S. Zweben, R. Maqueda, D. Stotler, A. Keesee, J. Boedo, C. Bush, S. Kaye, B. LeBlanc, J. Lowrance, V. Mastrocola, R. Maingi, N. Nishino, G. Renda, D. Swain, J. Wilgen, and t. N. Team, Nuclear Fusion **44**, 134 (2004).

[22] Y. Sechrest, D. Smith, D. P. Stotler, T. Munsat, and S. J. Zweben, Physics of Plasmas **22** (2015), 10.1063/1.4921215.

[23] J. A. Boedo, D. Rudakov, R. Moyer, S. Krasheninnikov, D. Whyte, G. McKee, G. Tynan, M. Schaffer, P. Stangeby, P. West, S. Allen, T. Evans, R. Fonck, E. Hollmann, A. Leonard, A. Mahdavi, G. Porter, M. Tillack, and G. Antar, Physics of Plasmas **8**, 4826 (2001).

[24] A. Huber, U. Samm, B. Schweer, and P. Mertens, Plasma Phys. Control. Fusion **47**, 409 (2005).

[25] R. J. Maqueda, G. A. Wurden, D. P. Stotler, S. J. Zweben, B. LaBombard, J. L. Terry, J. L. Lowrance, V. J. Mastrocola, G. F. Renda, D. A. D'lppolito, J. R. Myra, and N. Nishino, Review of Scientific Instruments **74**, 2020 (2003).

[26] J. R. Myra, D. A. D'Ippolito, D. P. Stotler, S. J. Zweben, B. P. LeBlanc, J. E. Menard, R. J. Maqueda, and J. Boedo, Physics of Plasmas **13**, 092509 (2006).

[27] S. Zweben, W. Davis, S. Kaye, J. Myra, R. Bell, B. LeBlanc, R. Maqueda, T. Munsat, S. Sabbagh, Y. Sechrest, and D. Stotler, Nuclear Fusion **55**, 093035 (2015).

[28] S. J. Zweben, J. R. Myra, W. M. Davis, D. a. D'Ippolito, T. K. Gray, S. M. Kaye, B. P. LeBlanc, R. J. Maqueda, D. a. Russell, and D. P. Stotler, Plasma Physics and Controlled Fusion **58**, 044007 (2016).

[29] R. Kube, O. E. Garcia, B. LaBombard, J. L. Terry, and S. J. Zweben, Journal of Nuclear Materials **438**, S505 (2013).

[30] O. E. Garcia, S. M. Fritzner, R. Kube, I. Cziegler, B. Labombard, and J. L. Terry, Physics of Plasmas **20**, 055901 (2013).

[31] G. Fuchert, G. Birkenmeier, D. Carralero, T. Lunt, P. Manz, H. W. Müller, B. Nold, M. Ramisch, V. Rohde, and U. Stroth, Plasma Physics and Controlled Fusion **56**, 125001 (2014).

[32] D. Goodall, Journal of Nuclear Materials **111-112**, 11 (1982).

[33] S. J. Zweben and S. S. Medley, Phys. Fluids B **1**, 2058 (1989).

[34] A. Kirk, N. Ben Ayed, G. Counsell, B. Dudson, T. Eich, A. Herrmann, B. Koch, R. Martin, A. Meakins, S. Saarelma, R. Scannell, S. Tallents, M. Walsh, H. R. Wilson, and M. Team, Plasma Physics and Controlled Fusion **48**, B433 (2006).

[35] B. D. Dudson, *Edge turbulence in the Mega-Amp Spherical Tokamak*, Ph.D. thesis, University of Oxford (2007).

[36] N. Ben Ayed, A. Kirk, B. Dudson, S. Tallents, R. G. L. Vann, and H. R. Wilson, Plasma Physics and Controlled Fusion **51**, 035016 (2009).

[37] a. Kirk, T. Eich, A. Herrmann, H. W. Muller, L. D. Horton, G. F. Counsell, M. Price, V. Rohde, V. Bobkov, B. Kurzan, J. Neuhauser, H. Wilson, t. A. U. Teams, and Mast, Plasma Physics and Controlled Fusion **47**, 995 (2005).

[38] B. D. Dudson, N. B. Ayed, A. Kirk, H. R. Wilson, G. Counsell, X. Xu, M. Umansky, P. B. Snyder, B. LLoyd, t. M. Team, N. Ben Ayed, A. Kirk, H. R. Wilson, G. Counsell, X. Xu, M. Umansky, P. B. Snyder, and B. LLoyd, Plasma Physics and Controlled Fusion **50**, 124012 (2008).

[39] S. Silburn, J. Harrison, M. Smithies, A. Wynn, T. Farley, and J. Cavalier, "Calcam (Version 1.9.4)," (2018).

[40] O. Grulke, J. L. Terry, B. LaBombard, and S. J. Zweben, Physics of Plasmas **13**, 012306 (2006).

[41] E. Havlíčková, J. Harrison, B. Lipschultz, G. Fishpool, A. Kirk, A. Thornton, M. Wischmeier, S. Elmore, and S. Allan, Plasma Physics and Controlled Fusion **57**, 115001 (2015).

[42] C. Tomasi and R. Manduchi, in *Sixth International Conference on Computer Vision* (Bombay, 1998) pp. 839–846, arXiv:arXiv:1011.1669v3.

[43] L. C. Appel and I. Lupelli, Computer Physics Communications **223**, 1 (2018).

[44] L. C. Appel, G. T. A. Huysmans, L. L. Lao, P. J. McCarthy, D. G. Muir, E. R. Solano, J. Storrs, D. Taylor, and W. Zwingmann, in *33rd EPS Conference on Plasma Physics*, Vol. 2 (2006) pp. 1235–1238.

[45] T. Hobiger, T. Kondo, and Y. Koyama, Earth Planets and Space (EPS) **60**, 727 (2008).

[46] D. Carralero, G. Birkenmeier, H. Müller, P. Manz, P. DeMarne, S. Müller, F. Reimold, U. Stroth, M. Wischmeier, and E. Wolfrum, Nuclear Fusion **54**, 123005 (2014), arXiv:arXiv:1407.3618v1.

[47] F. Militello, N. R. Walkden, T. Farley, W. A. Gracias, J. Olsen, F. Riva, L. Easy, N. Fedorczak, I. Lupelli, J. Madsen, A. H. Nielsen, P. Ricci, P. Tamain, and J. Young, Plasma Physics and Controlled Fusion **58**, 105002 (2016).

[48] H. P. Summers, "ADAS User Manual, v2.6," (2004).

[49] O. Garcia, J. Horacek, and R. Pitts, Nuclear Fusion **55**, 062002 (2015).

[50] B. Cannas, S. Carcangiu, A. Fanni, T. Farley, F. Militello, A. Montisci, F. Pisano, G. Sias, and N. Walkden, Fusion Engineering and Design , 0 (2019).