Review of Scientific Instruments

Filament identification in wide-angle high speed imaging of the mega amp spherical tokamak

Cite as: Rev. Sci. Instrum. **90**, 093502 (2019); https://doi.org/10.1063/1.5109470
Submitted: 09 May 2019 . Accepted: 13 August 2019 . Published Online: 03 September 2019

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, and J. W. Bradley

Review of Scientific Instruments ARTICLE scitation.org/journal/rsi

Filament identification in wide-angle high speed imaging of the mega amp spherical tokamak

Cite as: Rev. Sci. Instrum. 90, 093502 (2019); doi: 10.1063/1.5109470
Submitted: 9 May 2019 • Accepted: 13 August 2019 •
Published Online: 3 September 2019

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]

AFFILIATIONS

[1]EURATOM/CCFE Fusion Association, Culham Science Center, Abingdon OX14 3DB, United Kingdom
[2]Department of Electrical Engineering and Electronics, The University of Liverpool, Liverpool L69 3GJ, United Kingdom
[3]Department of Electrical and Electronic Engineering, University of Cagliari, Cagliari, Italy

[a)]Tom.Farley@ukaea.uk

ABSTRACT

A new tomographic inversion technique is presented for the identification of plasma filaments in wide-angle visible camera data. The technique works on the assumption that background subtracted images of filaments can be represented as a superposition of uniformly emitting magnetic equilibrium field lines. A large collection of equilibrium magnetic field lines is traced and projected onto the camera field of view and combined to form a geometry matrix describing the coordinate transformation from magnetic field aligned coordinates to image pixel coordinates. Inverting this matrix enables the reprojection of the emission in the camera images onto a field aligned basis, from which filaments are readily identifiable. The inversion is a poorly conditioned problem which is overcome using a least-squares approach with Laplacian regularization. Blobs are identified using the "watershed" algorithm and 2D Gaussians are fitted to get the positions, widths, and amplitudes of the filaments. A synthetic camera diagnostic generating images containing experimentally representative filaments is utilized to rigorously benchmark the accuracy and reliability of the technique. 74% of synthetic filaments above the detection amplitude threshold are successfully detected, with 98.8% of detected filaments being true positives. The accuracy with which filament properties and their probability density functions are recovered is discussed, along with sources of error and methods to minimize 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 density and temperature 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] fueling,[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 be 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, and intermittent plasma objects (IPOs)].[9,10] Recent forward modeling[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 nonlocal 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 modeling and the informing of analytic filament frameworks.

The bulk of past filament measurements have used Langmuir probe based techniques, analyzing the characteristic non-Gaussian statistics of ion saturation current fluctuations that arise from filaments passing the probe tip, with the strongest non-Gaussian behavior observed in the far SOL, where the relative amplitude of the filament fluctuations is the 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 multiprobe 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 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. First, 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 probes, 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 midplane and the divertor. The passive nature of the measurement also reduces its complexity, avoiding the complications of understanding complex nonequilibrium phenomena such as gas plume propagation or neutral beam-plasma interaction, instead relying on relatively uniform and slowly varying excitation and ionization rates. Second, 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 analyzes 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 simultaneous 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 analyzed. Automated or semiautomated 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 sophisticated techniques set a fixed radial position *a priori*, 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 semiautonomous techniques have focused on narrow, windowed camera views of specific areas of the plasma (e.g., the outboard midplane,[36,38] center column[36] or top outside corner of the plasma[34]), in order to maximize 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 quasitoroidal 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 setup of the camera system for which the technique has been developed, before Sec. III details the technique itself. Section IV describes the synthetic camera diagnostic which has been developed to test the technique, while Sec. 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 discusses a number of further factors affecting the technique, before Sec. VIII concludes the paper by summarizing 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

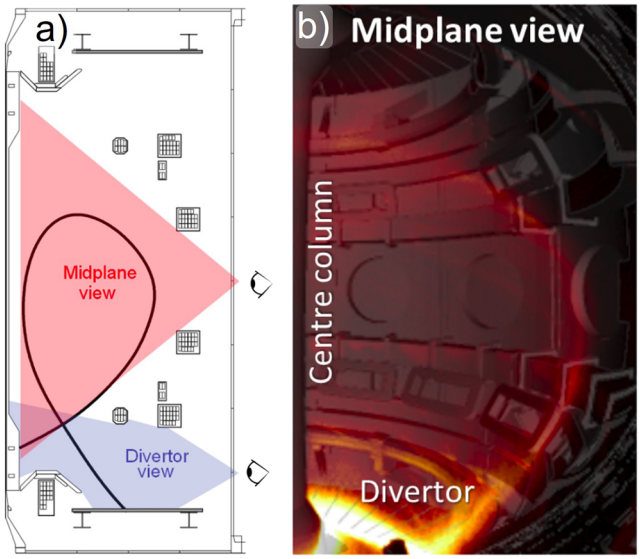

FIG. 1. (a) Schematic of the field of view of the Photron SA-1 fast camera when installed at the midplane or divertor. (b) False color image of a MAST plasma, as viewed by the SA1 camera, with a CAD rendering of the MAST vessel components overlaid.

are available corresponding to the "main chamber" and "divertor" views as shown in Fig. 1(a). This paper concerns the "main chamber" viewing geometry. Also shown in Fig. 1(b) 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.[65] 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.

III. FILAMENT MEASUREMENT TECHNIQUE

A. Overview and assumptions

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

Due to the fact that filaments are 3D objects with no simple exploitable toroidal or poloidal symmetry, conventional tomographic inversion techniques, such as the generalized Abel inversion,[40] cannot be used to study them. Instead, we exploit the helical symmetry of the magnetic field that the filaments conform to, in a similar way to that proposed in Ref. 41. The technique presented here operates by using information about the magnetic geometry to reconstruct the 3D intensity information present in the 2D camera images onto the horizontal R-$R\phi$ plane at the midplane ($Z = 0$), 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 proceed:

(A) Filaments are assumed to be aligned well to the background magnetic field which can be calculated via magnetic reconstruction.
(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 to be expected given the magnitude of parallel transport relative to perpendicular transport, which quickly spreads blobs of plasma along field lines. Furthermore, it is justified by a large base of experimental data from MAST[36] alongside other tokamaks.[19,42]

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 homogeneous neutral distributions away from the divertor.[43] 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.

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. The matrix of camera pixel intensities forming the camera image, I, can therefore be expressed as a weighted sum of basis images of uniformly emitting field lines, I_{basis},

$$I = \sum_{j=0}^{N} \epsilon_j I_{basis,j}, \quad (1)$$

where ϵ_j describes the relative contribution of the jth field line basis image to the camera image. The summation is valid due to the edge plasma being optically thin to visible radiation.[40] Equation (1) represents the mapping of the information in the camera images between a pixel basis set and a field line basis set.

All the information required to construct the field line basis set is contained in the magnetic equilibrium data and the camera spatial calibration that describes how emission from 3D point sources are mapped to pixels on the camera sensor.

We wish to convert the many nonlocal line integrated measurements from each pixel in the camera image into an estimate of the local plasma emissivity. The expression of the camera information in the field aligned basis set facilitates the identification of filaments, since field aligned plasma structures appear as bloblike clusters of high intensity field lines in 2D plots of ϵ_j. By analyzing the regions of enhanced emissivity present in ϵ_j, the filaments' sizes and locations can be deduced in a similar way as for poloidal intensity distributions produced by GPI or BES diagnostics. Therefore, the vector of field line basis weightings, ϵ_j, containing the filament information is the quantity that we seek to calculate.

The rest of this section describes the implementation of the technique, from preprocessing of the camera images and inversion of the data onto the field aligned basis, to the identification of filaments within the inverted data.

B. Image preprocessing

A series of preprocessing steps are applied to the camera images before the data are inverted. The output of each step is shown in Fig. 2. The preprocessing technique is motivated by the fact that filaments in the SOL represent positive density perturbations on top of a background so that perturbations in the light emission are also positive. Therefore, we define the background as the minimum in the signal over a time window about the frame of interest. 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 pixelwise 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 pixelwise FFT) component of the total emission. In the first two cases, this led to nonphysical 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 bilateral median filter[44] is applied, which is a gradient preserving noise removal method. The reduced dynamic range resulting from the background subtraction amplifies the relative contribution of the shot noise to the image, making noise removal particularly useful. The noise removal is followed by a Gaussian de-blur (weighted subtraction of the image convolved with a Gaussian kernel)[45] which helps to resharpen the image after the smoothing of the noise removal.

The application of regularization (discussed shortly) has a similar smoothing effect so that the inversion can proceed without the additional smoothing and sharpening steps. However, the effects of the additional smoothing on the results of the full inversion are found to be negligible, while the image smoothing produces significant improvements for nonregularization based techniques such as that discussed in Appendix A.

C. Tomographic inversion

The next stage of analysis is an inversion of the signal (as results from the previous preprocessing step) onto real space coordinates. A set of basis field line images are produced by projecting magnetic field line trajectories onto the camera image plane. The field lines are traced using a 4th order Runge-Kutta integrator with the magnetic field structure provided by an EFIT++[46,47] kinetic equilibrium reconstruction. The field lines are parameterized by their launch positions at the midplane $(R, R\phi)$, about a central toroidal angle ϕ_0. The field lines are followed for one toroidal rotation, so that they only enter the camera field of view once. The spatial calibration of the camera from Calcam is used to project the 3D coordinates describing the points along the field lines onto 2D pixel

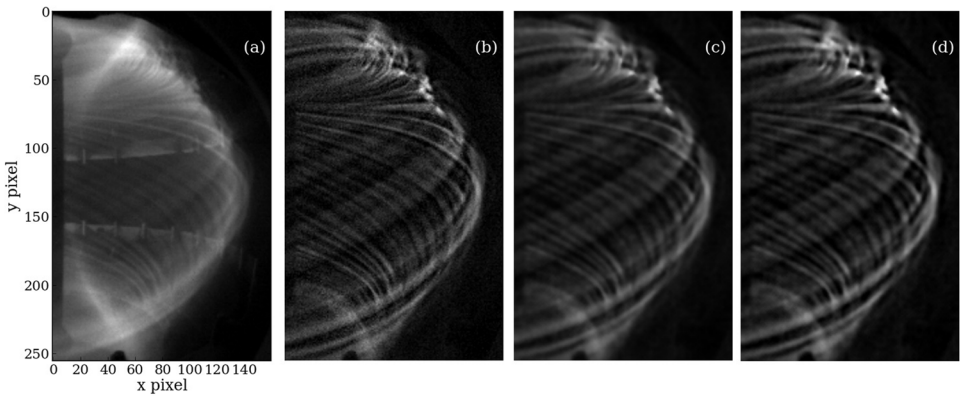

FIG. 2. (a) Raw image from MAST shot 29841 at 0.224 24 s. (b) Background subtracted image using 10 frames prior to the desired frame to construct a background. (c) Application of the bilateral median filter. (d) Application of a Gaussian de-blur.

coordinates in the camera image plane, according to the camera field of view and distortion model (see Ref. 48). With the pixels imaging each relevant field line identified, an image is produced of each field line, accounting for the inverse square reduction in intensity with distance from the camera pupil and volumetric line of sight integration effects. An example of a basis field line image is given in Fig. 3(a). Summing a collection of similar localized field line basis images according to Eq. (1) results in the image of the uniformly emitting filamentary flux tube in Fig. 3(b). Note that the flux tube is fainter at the top of the image due to fanning of the field lines and consequent stretching of the flux tube's cross section as the field lines get closer to the X-point. This results in the observed drop in emission which is also seen in experimental images.

Field lines are traced for a regular grid of launch positions in the R-$R\phi$ plane so that the set of points along all the field lines describe a volume of the SOL that is visible in the camera field of view. It is important to note that many of the field lines overlap in the field of view of the camera so that the field line basis set is *not* orthogonal. This 2D grid determines the resolution of the image of the inverted field aligned emission, ϵ. Figure 4 illustrates the relation between image plane and the inverted emission plane onto which the camera data are mapped. Note that due to the narrow radial width of the SOL, the toroidal extent of the inversion domain of 120 cm is much greater than the radial extent of 18.4 cm. In real space, the inversion domain takes the form of a curved trapezoidal plane, although for simplicity it is represented as rectangular in Fig. 4(a).

The inversion problem can be rigorously described if it is cast into linear algebra, with the camera image thought of as a matrix of pixel intensities with dimensions $m \times n$. The image matrix can be collapsed into a single contiguous image vector, $\mathbf{I}$, with elements I_i, and dimensions $mn \times 1$ ($i = 1, 2, \ldots, mn$). The resolution of the camera used in this work is $m \times n = 256 \times 160$, so that $\mathbf{I}$ is a vector of $mn = 40\,960$ pixel intensities. Likewise, the result of the inversion describing the field aligned emission in the toroidal plane with dimensions $p \times q$ can be flattened into a vector, $\boldsymbol{\epsilon}$, with elements ϵ_j and dimensions $pq \times 1$, equal to the number of field lines in our emission basis ($j = 1, 2, \ldots, pq$). Then, we can compose a *geometry matrix*, $\mathbb{G}_{ij} = [\mathbf{I}^{(0)}_{basis}, \mathbf{I}^{(1)}_{basis}, \ldots, \mathbf{I}^{(pq-1)}_{basis}, \mathbf{I}^{(pq)}_{basis}]$, where the column vectors, $\mathbf{I}^{(j)}_{basis}$, correspond to the flattened basis field line images [see the example image in Fig. 3(a)]. Here, $\mathbb{G}_{ij}$ represents the element from the ith row and jth column of matrix $\mathbb{G}$. The resulting geometry matrix, $\mathbb{G}_{ij}$, has dimensions $mn \times pq$ and contains all the necessary information about the magnetic equilibrium structure and camera viewing geometry.

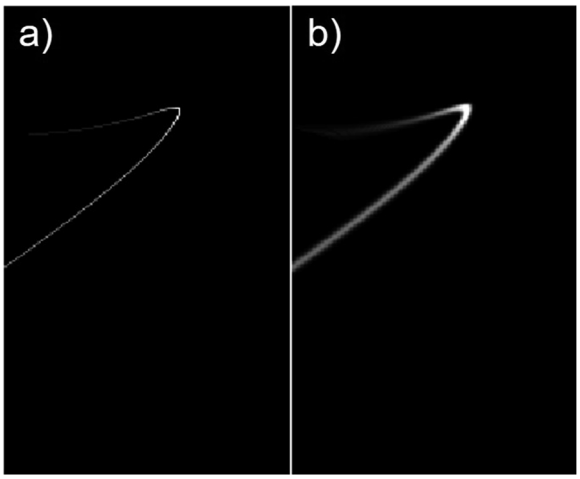

FIG. 3. (a) Example of a basis field line image used to construct the background subtracted camera images. The varying intensity of the field line accounts for the inverse square reduction in intensity with distance from the camera pupil and volumetric line of sight integration effects. Flattened basis images such as this form columns of the geometry matrix $\mathbb{G}_{ij}$. (b) Emission from a bundle of field lines with emission amplitudes, ϵ_j, given by a 2D Gaussian in the toroidal plane.

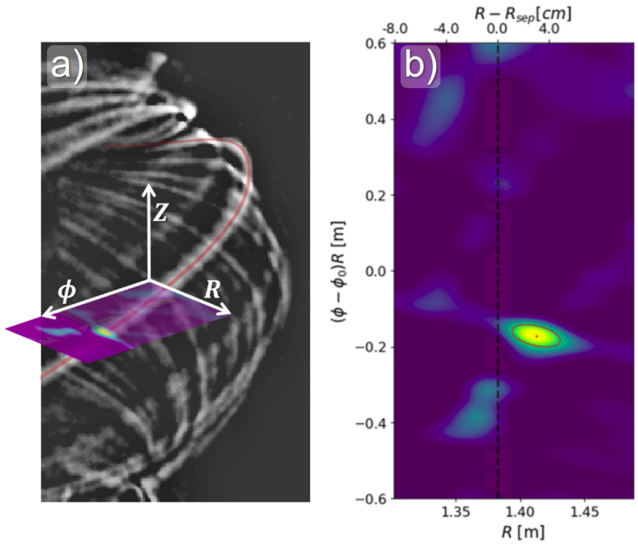

FIG. 4. (a) An experimental camera frame from MAST shot 29852, with the preprocessing step applied. (b) Inversion calculated using the frame in (a). The inversion in (b) is also superimposed on the frame in (a) to illustrate the (R, ϕ) midplane coordinate system.

It has been found empirically that to be able to represent a midplane camera image with a basis of field-lines in this manner requires a well resolved grid, typically with a radial spacing $\lesssim$4 mm and a toroidal spacing $\lesssim 0.2^\circ$. For the work presented here, the resolution of the inversion is typically of the order $p \times q = 47 \times 123$, corresponding to a total of $pq = 5781$ distinct field lines in the basis. Further details on the choice of inversion resolution will be discussed later.

Equation (1) can now be expressed in matrix format

$$\underbrace{I_i}_{mn\times 1} = \underbrace{\mathbb{G}_{ij}}_{mn\times pq}\ \underbrace{\epsilon_j}_{pq\times 1}, \tag{2}$$

where the Einstein summation convention is used to represent sums over common indices. Once the matrix $\mathbb{G}$ has been computed it can be applied to describe any number of frames from a movie for which the magnetic equilibrium is reasonably constant.

Equation (2) represents an overdetermined system of equations, such that $\mathbb{G}$ does not have a true inverse. Therefore, we instead

take an ordinary least squares approach to finding the "best" value of the inverted emission vector $\hat{\boldsymbol{\epsilon}}$ that satisfies

$$\hat{\boldsymbol{\epsilon}} = \underset{\boldsymbol{\epsilon}}{\operatorname{argmin}} \, \|\mathbb{G}\boldsymbol{\epsilon} - \mathbf{I}\|^2, \tag{3}$$

where "argmin" is the operator returning the value of $\boldsymbol{\epsilon}$ that minimizes its argument. The ordinary least squares solution is then given by[49]

$$\begin{aligned} \hat{\boldsymbol{\epsilon}} &= (\mathbb{G}^T\mathbb{G})^{-1}\mathbb{G}^T\mathbf{I} \\ &= \mathbb{H}^{-1}\mathbb{G}^T\mathbf{I} \\ &= \mathbb{H}^{-1}\boldsymbol{\mathcal{E}}, \end{aligned} \tag{4}$$

where $\mathbb{G}^T$ represents the transpose of $\mathbb{G}$, $\mathbb{H} = \mathbb{G}^T\mathbb{G}$ is termed the *Gramian matrix* of $\mathbb{G}$, and $\boldsymbol{\mathcal{E}} = \mathbb{G}^T\mathbf{I}$ is the *moment matrix*. Given that the $\mathbb{H}$ is square, it has a true inverse, $\mathbb{H}^{-1}$, satisfying $\mathbb{H}\mathbb{H}^{-1} = \mathbb{I}$, where $\mathbb{I}$ is the identity matrix (not to be confused with the image vector $\mathbf{I}$). Thus, the problem of finding $\hat{\epsilon}$ has been reformulated to involve the inversion of a square matrix, $\mathbb{H}$.

To calculate the inverse of $\mathbb{H}$, a non-negative SART[50] (simultaneous algebraic reconstruction) algorithm with Laplacian regularization is used, as it has been found to be many times faster than similar methods, such as standard non-negative least squares,[51] singular value decomposition (SVD),[52] and More-Penrose pseudoinverse matrix inversion algorithms,[53] while being more robust to noise than the QR factorization algorithm[54] and producing equally good results. The non-negativity constraint was added to avoid unphysical negative emission in the inversion solution. As the density field in the SOL is a smoothly varying quantity, we apply Laplacian regularization. This ensures that the resulting emission is smoothly varying and provides the constraint needed to identify a unique solution, producing good results. The strength of the regularization parameter,[55] λ, was tuned empirically so as to avoid spurious sharp banding patterns in $\hat{\epsilon}$ resulting from noise, while avoiding over smoothing of the large scale physical structures in the inversion.

We now present an example of how the technique is applied to the inversion of experimental camera data from MAST. Figure 5 illustrates the two main calculation steps in the inversion process, along with a reprojection of the inversion output back onto the camera field of view. Figure 5(a) shows an experimental background subtracted camera image from MAST discharge 29852 which is to be inverted. Figure 5(b) shows the intensities in the moment matrix, $\boldsymbol{\mathcal{E}}$, resulting from the product of the transpose of the geometry matrix with the image vector produced by flattening the image in Fig. 5(a). The range of toroidal angles spanned by the inversion domain are also shown in white. Figure 5(c) shows the inversion resulting from the application of the SART algorithm. A number of bright blobs of field aligned emission are visible corresponding to filaments in the camera image. These structures are identified and measured using the filament identification algorithm described shortly, with the resulting detections indicated by colored ellipses. Finally, Fig. 5(d) shows the reprojection of the information captured in $\hat{\boldsymbol{\epsilon}}$ back onto the camera field of view. The red line bounds the inversion plane in which the emission is mapped.

The reprojection is produced by taking the product of the geometry matrix with the inverted emission vector (discussed further in Sec. IV). Field lines at the centers of the detected filament ellipses in Fig. 5(c) are overplotted in Figs. 5(a) and 5(d) in the same colors. The identified structures align well with the filaments visible in the camera image, giving a visual indication of the quality of the detections.

The design choice to analyze a subset of the toroidal domain is motivated by the large computational gains that can be made by making the geometry matrix smaller. This reduces the size of the geometry matrix by around an order of magnitude, leading to a significant gain in computation time. Given many of the field lines excluded from the analysis domain are either outside the field of view

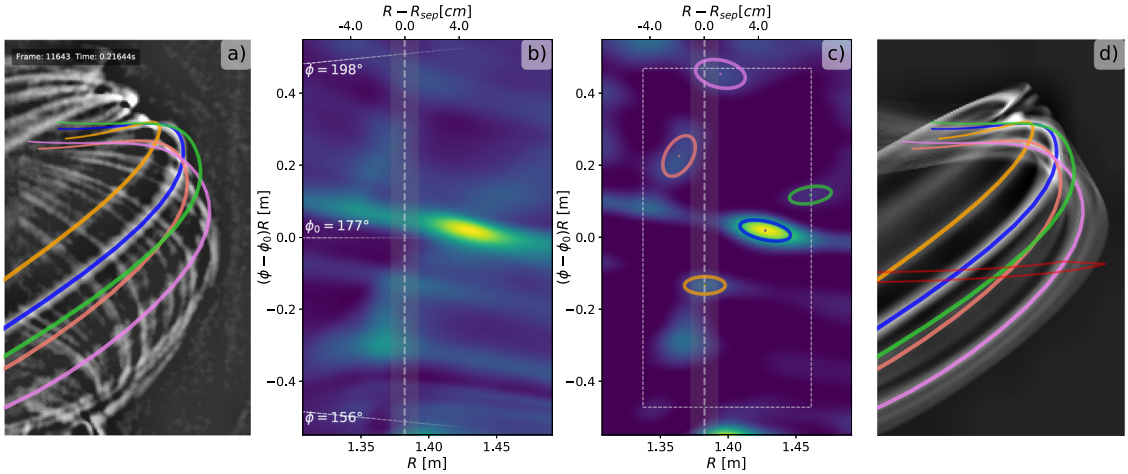

FIG. 5. (a) Preprocessed camera frame field lines superimposed marking the centers of the detected filaments. (b) Emissivity data from the moment matrix $\boldsymbol{\mathcal{E}}$. The range of toroidal angles covered by the inversion domain is indicated. (c) Inverted field aligned emissivity represented by $\hat{\boldsymbol{\epsilon}}$. Colored 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.

of the camera are only partially visible or are in regions of the image where the filaments are poorly resolved, minimal useful information is lost, while information about the most readily observable filaments is retained.

The chosen analysis domain focuses on the foreground filaments. This is because these filaments are closest to the camera so their widths and positions can be best resolved. The field lines at the top and bottom of the images are excluded from the analysis domain since only short portions of them are in the field of view of the camera, so that the properties of the filaments there cannot be well constrained either.

While the background filaments do not fall within the analysis domain, they do overlap with the foreground filaments we wish to measure. However, they are near-perpendicular to the foreground filaments so that their contribution to the intensity is similar for all foreground field lines. Therefore, to first order, the additional intensity from background filaments appears as a background offset in the foreground filament intensities and does not significantly impact the analysis of the foreground filaments. This is discussed further in Sec. V H.

A consequence of the reduced inversion domain is that, at its boundaries where there is a sharp cut off in the basis, the SART algorithm struggles to distribute emission correctly. For this reason, filament detection is only applied to structures with centers within the region defined by the dotted white border box in Fig. 5(c).

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, ϕR), widths (δ_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 is found to originate from noise or artifacts in the inversion [see white contour in Fig. 6(a)]. Alternatively, the threshold can be set to a number of standard deviations above the average inversion intensity across all the frames, akin to many past measurements in the literature.[9,56] However, the first method has the benefit that all filaments above the noise floor of the technique can potentially be analyzed, as opposed to imposing a somewhat less physically justifiable; e.g., $2.5\sigma_I$ cutoff in the amplitude of the coherent filamentary structures that are analyzed. Further details of the selection of this parameter are given in Sec. 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 labeled background (not filaments). Dilated regions around 2D local maxima with intensities above ϵ_{thresh} are labeled foreground (potential filament centers). With these regions as inputs [Fig. 6(b)], the watershed algorithm sorts the remaining unassigned regions into background [purple regions in Fig. 6(c)] or distinct foreground regions [colored regions in Fig. 6(c)]. Note that each pixel in Fig. 6(b) represents a single basis field line, showing the resolution of the analysis grid. The watershed algorithm can be pictured in terms of different colored 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 color. On contact with another region, the area assigned to that color (filament) is frozen. This process is illustrated in Fig. 7.

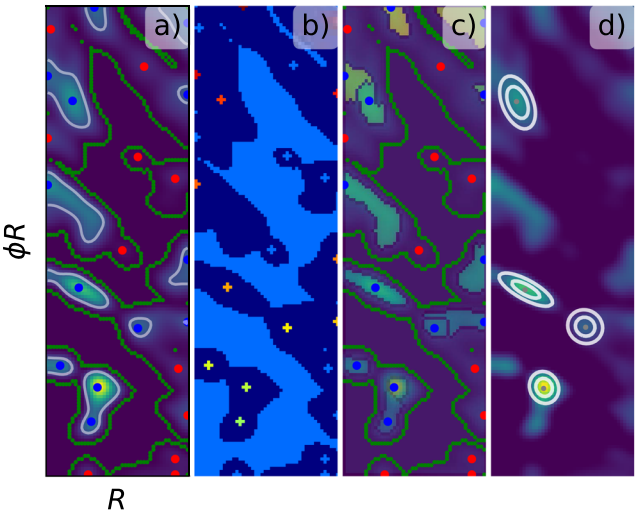

FIG. 6. (a) Emissivity data, with the minimum peak blob intensity threshold indicated as a white contour. Local maxima above and below the threshold are indicated as blue and red dots, respectively, while local minima and boundaries of zero emission are indicated in green. (b) Regions given as input to the watershed algorithm. Definite background is in light blue (not a blob), dark blue areas are yet to be classified by the watershed algorithm, and all other colors indicate foreground regions (sources for potential blobs). Note that each pixel in the image represents a single basis field line. (c) Colored regions show contours returned by the watershed algorithm. Colored points are as in (a). (d) Contours of 2D Gaussians fitted to the data in the colored regions shown in (c). Some regions have been filtered out using a range of criteria and thus do not have fits.

The identified regions are filtered according to a number of conditions. First, the contours are assessed on their solidity (ratio of area of the contour to the 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 (δ_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

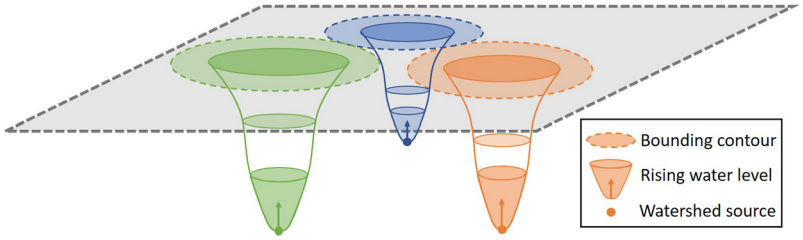

FIG. 7. Illustration of the watershed algorithm, with different colored water sources rising until they touch and define a bounding contour.

in the inversion. The Gaussian fits yield the filaments central positions, 1σ major and minor axes widths, amplitudes, and inclinations [see contours in Fig. 6(d)]. 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 bloblike 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. First, it handles interacting filaments very effectively, preventing them from being merged into a single larger blob and avoiding problems from multimodal emissivity structures producing multiple contours for a given filament. Second, it efficiently maximizes 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 neighboring filament (whichever is closest). When combined with the fitting of 2D Gaussians to the contours, this means that large secluded filaments can utilize 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. 5(c), the detected filaments are indicated by colored ellipses showing the 1σ contours of the fitted Gaussians. The central field lines for each fitted filament are indicated by colored lines of the same color in Figs. 5(a) and 5(d).

Geometric line of sight effects from wide angle imaging lead to some complications correctly identifying filaments and extracting their properties from the inverted emission data. Insight into artifacts in the reconstructed emission can be gained by considering the structures present in the moment matrix, shown in Fig. 5(b). The moment matrix gives the sum total emission in the camera image along each field line in the basis rather than a self-consistent distribution of the emission among the field lines. Therefore, any remaining artifacts from line of sight effects in the inverted emission, originate in stronger forms in the moment matrix. The moment matrix captures the main structural features of the inversion solution, but these structures are distorted over smaller length scales by the interfield line dependencies, resulting in banding and blurring in $\mathcal{E}$.

Due to the narrow width of the SOL, the variation in magnetic field and thus pitch angle across the radial extent of the inversion domain is rather small. Therefore, the geometry of the basis field lines varies slowly in the radial direction. The field lines in close radial proximity to a filament can therefore have raised emission along a portion of their length even if the field line as a whole does not properly align with the filament. This explains why many of the banding structures are predominantly radial in nature. We term these bands extending outward from the filaments the "shadows" of the filaments, in the sense that the filaments "cast" emission onto their neighboring field lines.

While these shadows are heavily attenuated in the inverted emission, some remnant of them remains. The most direct consequence of the shadows is that they broaden the emission from the filaments so that their widths will generally be overestimated. However, they also have a second important consequence, arising from the fact that some basis field lines can be weakly approximated by the linear combination of more spatially separated field lines. This is illustrated in Fig. 8 where the flux tubes represented by the red and orange dashed lines overlap with significant

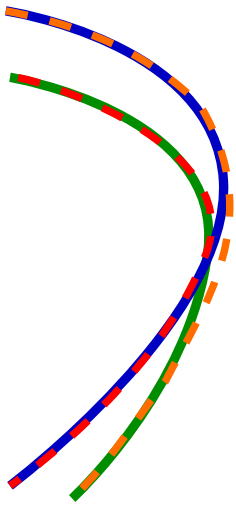

FIG. 8. Illustration of the origin of "ghost" filaments. The blue and green lines represent true filaments that overlap in the line of sight of the camera. Red and orange dashed lines represent possible field lines that overlap with significant portions of the blue and green filaments and thus generate an artificial ghost filament.

portions of the blue and green flux tubes. Therefore, when filaments lie along the blue and green flux tubes, the average emission along the length of the red and orange flux tubes will also be high despite there being little genuine localized emission along those field lines. This leads to secondary regions of enhanced emission in the inversion, that have the same appearance as filaments. These spurious filament structures we term "ghost" filaments. In the emission plots, these ghost filaments occur where the shadows of multiple parent filaments overlap to form new local maxima in intensity.

Bright spots in the camera images, typically occurring above the midplane, 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 ghost filaments.

The green filament in Fig. 5 is an example of a ghost filament arising from the overlapping of sections of the red and blue filaments in the camera line of sight.

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 modeling of simulation outputs for comparison with experimental measurements (as applied in Ref. 57). The principle used is the converse of the inversion process: Eq. (2) is applied to produce a synthetic camera image $\mathbf{I}_{synth}$ for a predefined emission profile $\boldsymbol{\epsilon}_{synth}$ and geometry matrix $\mathbb{G}$. The resulting image contains emission that is equivalent to that in background subtracted experimental camera frames.

In principle, the input emission vector can be any 2D pattern. If an experimentally derived emission vector is used as input, we produce a synthetic reprojection of the inversion data, demonstrating the visual form of the structures that are captured in the emission vector, as was shown in Fig. 5(d). The input emission vector can also be provided by numerical simulations[57] and in this application, the technique provides forward modeling. A third option is when entirely artificial filament images are produced for the purposes of benchmarking and technique optimization and each filament's emission profile at the midplane is set to a 2D Gaussian.

In the latter case, 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 the synthetic diagnostic composes images through a linear combination of field line basis images, it can only produce images of field-aligned structures and no variations in emission along field lines is possible in its current form. This limitation could be relaxed by including further basis functions of increasing complexity in $\mathbb{G}$, although for the reasons outlined in Sec. III A, the emission of interest is strongly field aligned and such an increase in complexity is not warranted at this point.

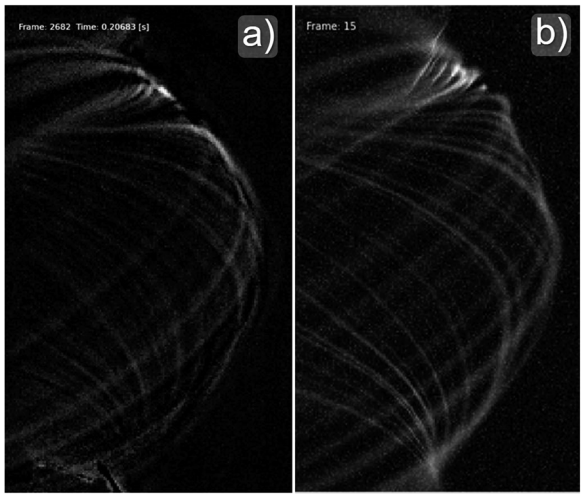

FIG. 9. (a) Experimental camera frame from MAST shot 29852, with the preprocessing described in Sec. III B applied. (b) Frame from the reference synthetic data set produced using the synthetic diagnostic, containing 40 synthetic filaments, with positive definite Gaussian noise applied.

Figure 9 compares output from the synthetic diagnostic to experimental data, showing a good visual qualitative similarity, further justifying this choice. Some differences are seen in the upper and lower divertor regions due to poorer constraint of the camera calibration in those regions. However, they are outside of the analysis region that is inverted and so do not impact the results below.

V. BENCHMARKING

As discussed in Sec. III C, geometrical distortions that occur in the inversion process can lead to errors and false detection of ghost 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 analyzing 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 data set, A, 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. These synthetic filaments were distributed uniformly around the full toroidal extent of the machine. A range of numbers of filaments per frame was explored and 40 filaments per frame were chosen as it corresponds to a commonly observed toroidal quasimode number seen in MAST[12] and produced synthetic frames which most closely visually resembled experimental camera data. The toroidal quasimode number describes the average number of filaments observed in a full toroidal circuit of the machine. The chosen value lies at the upper end of literature values for quasitoroidal 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 quasimode number is likely to appear lower than the input mode number.

Positive Gaussian noise with standard deviation equal 5% of the image dynamic range was applied to all synthetic frames. This amplitude of noise is representative of that found in experimental camera data.

For this application, the geometry matrix is calculated with a higher grid resolution of $\Delta R = \Delta(\phi R) = 2.5$ mm in order to avoid aliasing effects in the resulting images.

The synthetic filaments are produced with Gaussian cross sections and their properties (positions, widths, and amplitudes) are drawn from realistic probability distributions based on experimental findings. The details of the chosen distributions are summarized 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,[58] indicating that filaments are randomly distributed toroidally around the machine with no mode number (hence the use of the term *quasimode 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 59.

The camera calibration and magnetic equilibrium used to produce the synthetic data are for MAST discharge 29 852. The radial extent of the analysis region was $R = 1.290$–1.474 m, spanning the outer midplane SOL and including several centimeters inside the separatrix, given the MAST midplane separatrix was located at $R \approx 1.36$ m for this discharge. The analysis region extended toroidally from $\phi R = -0.55$ m to $\phi R = +0.548$ m, corresponding to an angular range in machine coordinates of $\phi = 153^\circ$–215° centered about $\phi_0 = 177^\circ$ (c.f. camera tangency plane at $\phi = 215^\circ$).

The analysis region focuses on foreground filaments (i.e., those that intersect the midplane in the foreground). Of the 200 000 filaments in the data set, 20 243 (10.1%) fall within the analysis domain indicated by the white dashed box in Fig. 5(c), corresponding to an average of 4.05 filaments in the analysis domain 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 5781 distinct field lines. This is the same resolution used to invert experimental camera data. 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

In order to quantitatively assess the quality of our inversion technique, we need to classify the detected filaments as either "true positives" that can be assigned to corresponding input synthetic filament or spurious "false positives" that cannot be matched to a synthetic filament. Before we proceed, we define the *detected filaments* as the output from the filament identification technique and the *synthetic filaments* as the input filaments used to generate the synthetic camera images.

The method by which the set of detected filaments $\{D_{\text{all}}\}$ and the set of synthetic filaments $\{S_{\text{all}}\}$ are compared and matched is as follows:

1. As discussed earlier, since the synthetic filaments are produced homogeneously around the whole machine, only some of them will fall in the toroidally localized analysis domain. These will form the subset $\{S_{\text{domain}}\}$ of filaments inside the analysis region [white dashed box in Fig. 5(c)]. Of the filaments in $\{S_{\text{domain}}\}$, those that (1) lie within the analysis domain and (2) have amplitudes above the detection threshold, ϵ_{thresh}, [defined in Sec. III D] will from an even smaller subset that we call $\{S_{\text{analyze}}\}$.
2. Each detected filament is compared to each unmatched synthetic filament in $\{S_{\text{domain}}\}$. The detected filaments are examined in order of decreasing amplitude.
   (a) If a single synthetic filament center lies within 1σ of the Gaussian fitted to the detected filament, the input and detected filaments are added to the sets of matched filaments $\{S_{\text{matched}}\}$ and $\{D_{\text{matched}}\}$, 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. Overlapping filaments can be generated due to the statistical nature of the synthetic frame generation. However, we do not expect them to occur very frequently in actual experiments.
3. Synthetic filaments in $\{S_{\text{analyze}}\}$ that are neither matched nor overlapping are labeled "missed" and added to $\{S_{\text{missed}}\}$. Detected filaments that are inside the analysis region and are not matched are labeled "false" detections and added to set $\{D_{\text{false}}\}$.

To summarize, we hence have that

$$\{S_{\text{all}}\} > \{S_{\text{domain}}\} > \{S_{\text{analyze}}\} \approx \{S_{\text{matched}}\} + \{S_{\text{missed}}\} + \{S_{\text{overlap}}\} \tag{5}$$

and

$$\{D_{\text{all}}\} = \{D_{\text{matched}}\} + \{D_{\text{false}}\}. \tag{6}$$

During this matching procedure a number of additional cases are also handled which result in the approximate equality in Eq. (5) and 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_{analyze} and the filaments are considered matched as above. If a synthetic filament in S_{analyze} is matched to a detected filament outside the analysis region (i.e., in the border region subject to basis boundary effects) the synthetic filament is removed from S_{analyze} as it does not manifest in the analysis domain and thus would not be observed experimentally.

Finally, with all detected filaments assessed for synthetic matches, we define the set

$$\begin{aligned}\{S_{\text{isolated}}\} &= \{S_{\text{analyze}}\} - \{S_{\text{overlap}}\} \\ &= \{S_{\text{matched}}\} + \{S_{\text{missed}}\},\end{aligned} \tag{7}$$

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 (i.e., we want to minimize $\{S_{\text{missed}}\}$).

In Fig. 10, we show a particular frame that displays each of the categories of synthetic and detected filament (typical synthetic frames do not generally display this complexity). Synthetic filaments are indicated by dashed ellipses and detected filaments by solid ellipses. Matched filaments are shown in green (belonging to $\{S_{\text{matched}}\}$ or $\{D_{\text{matched}}\}$) and false ($\{D_{\text{false}}\}$) or missed ($\{S_{\text{missed}}\}$) detections are shown in red. Overlapping synthetic filaments ($\{S_{\text{overlap}}\}$) are shown in blue, while synthetic filaments outside the analysis domain or with amplitudes below the threshold (filaments not in $\{S_{\text{analysis}}\}$) are shown in pink.

C. Benchmarking figures of merit

In order to quantify the performance of the inversion technique, a number of numerical figures of merit (FoM) have been developed. These are instrumental to an objective assessment of the technique, informing further development, as well as giving metrics for benchmarking overall success. Each type of FoM is described below. Here, the notation $|S_{\text{matched}}|$, represents the number of elements in the set $\{S_{\text{matched}}\}$ and the same convention will be used for all the other sets.

1. Sensitivity

We define *sensitivity* (also known as *recall*) as

$$f_{\text{sens,domain}} = \frac{|S_{\text{matched}}|}{|S_{\text{domain}}|} \times 100. \tag{8}$$

This gives the proportion of all filaments in the analysis domain that are detected and is expressed as a percentage. We also define

$$f_{\text{sens,isolated}} = \frac{|S_{\text{matched}}|}{|S_{\text{matched}}| + |S_{\text{missed}}|} \times 100, \tag{9}$$

which gives the proportion of isolated analysis filaments that are detected.

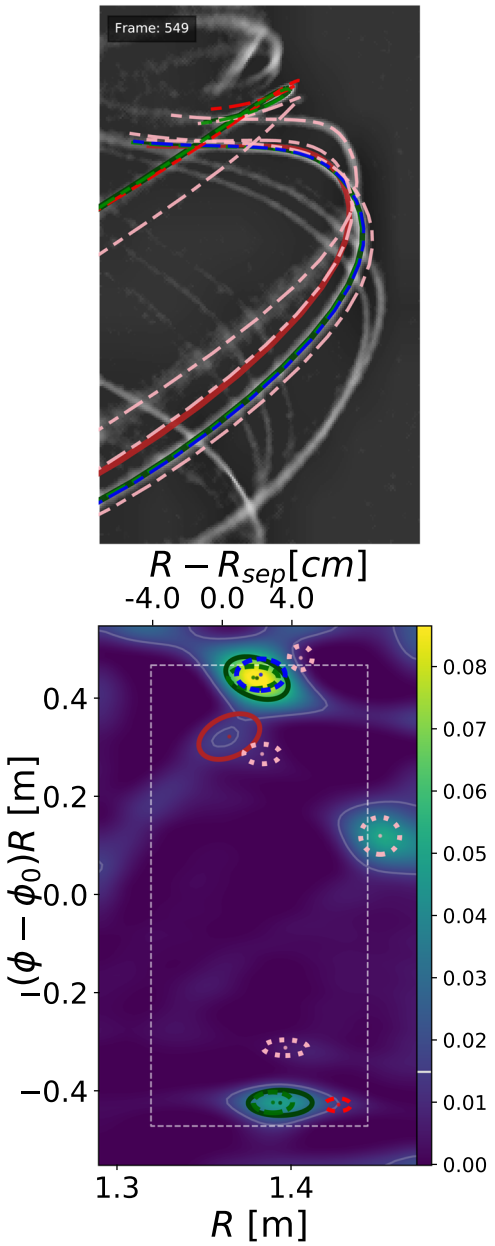

FIG. 10. (a) Synthetic camera frame and (b) corresponding inversions with detected (synthetic) filaments represented by ellipses with solid (dashed) lines. Matched filaments are shown in green, false (missed) filaments in red, overlapping filaments in blue and filaments outside the analysis region (white dashed rectangle) or with low amplitudes in pale pink. The white contour and color bar indicate ϵ_{thresh}. The color scale depends on the arbitrary synthetic amplitudes in the synthetic images.

2. Precision

We define *precision* as

$$f_{\text{prec}} = \frac{|D_{\text{matched}}|}{|D_{\text{matched}}| + |D_{\text{false}}|} \times 100. \tag{10}$$

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

These are termed *identification* FoMs as they describe how effectively filaments are identified by the technique.

3. Parameter fractional errors

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

$$f_{\mathrm{p}} = \left\langle \frac{|p_{\mathrm{measured}} - p_{\mathrm{input}}|}{p_{\mathrm{input}}} \right\rangle_{\mathrm{matched}} \times 100, \tag{11}$$

where $\{p_{\mathrm{measured}}\}$ and $\{p_{\mathrm{input}}\}$ are the sets of detected and synthetic filament parameter values, respectively, and the angular bracket represents the ensemble average over all matched filaments. Parameter fractional errors are calculated for filament position, widths, and amplitudes.

The FoMs calculated using Eq. (11) are termed *measurement* FoMs as they concern the accuracy with which each parameter is measured.

D. Filament amplitude detection threshold

The analysis of synthetic images provides an objective means of setting the threshold amplitude for filament identification described in Sec. III D. The parameter $\epsilon_{\mathrm{thresh}}$ is chosen so as to maximize the precision and sensitivity figures of merit, while sampling a sufficiently large portion of the total filament population.

The first function of $\epsilon_{\mathrm{thresh}}$ is to ensure the exclusion of small local maxima below the noise floor. However, further increases in $\epsilon_{\mathrm{thresh}}$ yield improvements in the FoMs due to the exclusion of spurious ghost filaments.

At large $\epsilon_{\mathrm{thresh}}$, only the largest filaments in the population remain in the analysis set, leading to increased sensitivity to filaments in $\{S_{\mathrm{analyze}}\}$, 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.

Figure 11 shows how much of the total filament population (f_{sens}) is excluded for successive gains in precision (f_{prec}) and sensitivity to the targeted filament population ($f_{\mathrm{sens,isolated}}$) as $\epsilon_{\mathrm{thresh}}$ 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 isolated sensitivity. At this level, 50% of the total filament population lies above the amplitude threshold. The technique detects 36% of the total filament population and 74% of isolated analysis filaments. This is achieved with a 98.8% true positive detection rate. These FoM values are summarized in Table III.

The dependence of $f_{\mathrm{sens,domain}}$ 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 59%, 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 ($f_{\mathrm{prec}} \approx 57\%$). The precision saturates around $\epsilon_{thresh} = 2.0 \times 10^{-2}$, above which nearly all detections are true positives.

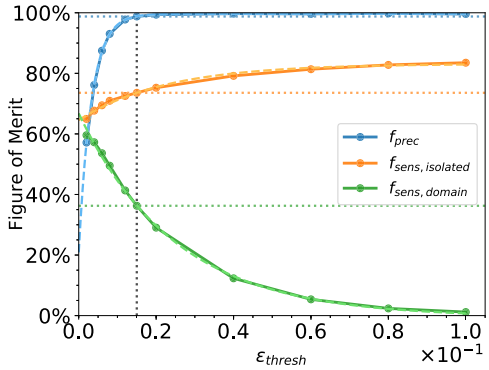

FIG. 11. Dependence of the precision (blue), analysis sensitivity (orange), and total sensitivity (green) FoMs on filament amplitude detection threshold, $\epsilon_{\mathrm{thresh}}$. Dashed lines show the FoM values achieved at the value of $\epsilon_{thresh} = 1.5 \times 10^{-2}$ used in the analysis.

E. Breakdown of filament detections

With the figures of merit in place and the threshold set, we can now quantify the performance of the inversion technique. 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, analyzed 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 the window in toroidal angle of 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 larger filaments cannot be detected, they raise the amplitude of the larger detected filaments. However, we do not expect these filament interactions to occur routinely in experiment.

Table II gives a breakdown of the detected filaments. Of the 7441 detections, only 91 (1.2%) were false positives that did not match to a synthetic filament.

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

Set S_{x}	$\lvert S_{\mathrm{x}}\rvert$	$\frac{\lvert S_{\mathrm{x}}\rvert}{\lvert S_{\mathrm{domain}}\rvert}$ (%)	$\frac{\lvert S_{\mathrm{x}}\rvert}{\lvert S_{\mathrm{analysis}}\rvert}$ (%)	$\frac{\lvert S_{\mathrm{x}}\rvert}{\lvert S_{\mathrm{isolated}}\rvert}$ (%)
S_{domain}	20 243	100	...	...
S_{analysis}	10 134	50	100	...
S_{isolated}	9 978	49	98	100
S_{matched}	7 341	36	72	74
S_{missed}	2 637	13	26	26
S_{overlap}	156	0.8	1.5	...

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

Set D_x	$\lvert D_x \rvert$	$\frac{\lvert D_x \rvert}{\lvert D_{domain} \rvert}$ (%)
D_{domain}	7432	100.0
$D_{matched}$	7341	98.8
D_{false}	91	1.2

F. Error distributions and corrections

After quantifying the detection capabilities, we now address the problem of assessing the quality of the measurements. 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.

Figure 12 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. (11) for each parameter (note $f_{R-R_{sep}}$ has been normalized relative to the 5 cm SOL width between the last closed flux surface (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 fitted Gaussian are also small, each around 0.2 cm, although the raised wings

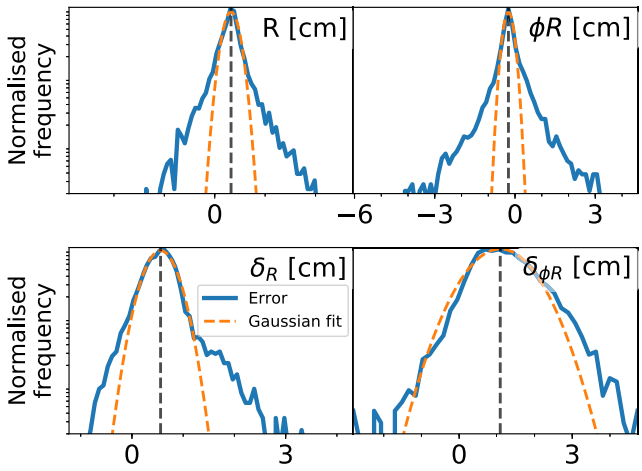

FIG. 12. Distribution of errors in detected filament parameters for radial position, toroidal position, radial width and toroidal width (left to right, top to bottom, respectively). The solid blue lines shows the distributions of errors, while the dashed orange lines shows Gaussian fits to the distributions.

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

	Error		
Quantity, p	μ_p	σ_p	FoM f_p (%)
f_{prec}	...	...	98.8
$f_{sens,isolated}$	...	...	74
$f_{sens,domain}$	...	...	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)

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. 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 filament velocity measurements, corresponding to velocity errors of the order 0.3 km/s in both radial and toroidal directions. For a typical filament traveling 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 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 principles models and to assess their intermittent loads on the plasma facing components. We now wish to understand how well our technique is able to reproduce the distributions of the synthetic filaments.

Figure 13 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 systematic offsets identified in Sec. V F.

The distribution functions of radial and toroidal positions are very well recovered, closely following the synthetic distributions. The detected width distributions preserve the overall shape of the

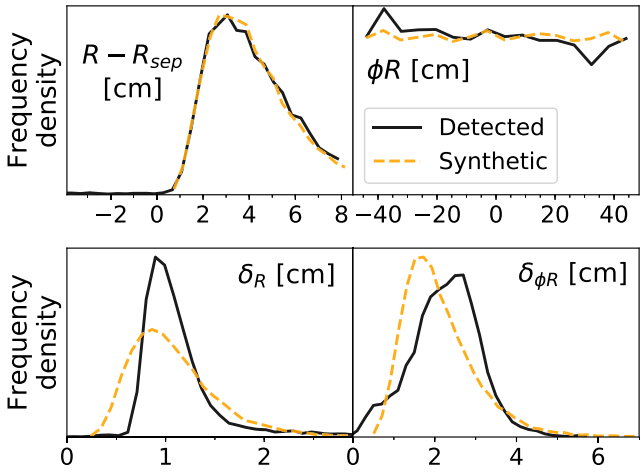

FIG. 13. Comparison of input and output frequency density distributions of filament positions and widths for the reference synthetic data set. Distributions of synthetic input filaments are given by orange dashed lines and detected distributions by black solid lines.

originals, although the distribution of δ_R is more peaked and the distribution of $\delta_{\phi R}$ has its peak shifted to the larger widths by around 0.6 cm. This gives us confidence that detected distributions of filament positions and widths accurately reflect the forms of the true underlying physical distributions.

It is useful to analyze these results in more detail, in particular, the partial distributions of specific subsets in order to determine trends. These partial distributions are displayed and compared in Fig. 14, which we will now discuss. The distributions of matched ($\{S_{matched}\}$) and missed ($\{S_{missed}\}$) filaments collapse onto the overall distribution of synthetic filaments ($\{S_{domain}\}$) showing that filaments are not selectively detected and thus the full extent of the distributions are evenly sampled. 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 distribution of false positive detections ($\{D_{false}\}$), however, shows separate structure that cannot be resolved in Fig. 14 due to their low frequency. Indeed, false detections are over twice as likely to occur at low R values, which is anticipated since line of sight effects mean ghost 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 ghost filaments. Both width distributions have higher relative false positive rates for larger width filaments, especially in the case of $\delta_{\phi R}$. Dashed and dotted lines show the original distributions, while solid lines show the measured distributions. The dotted gray 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.

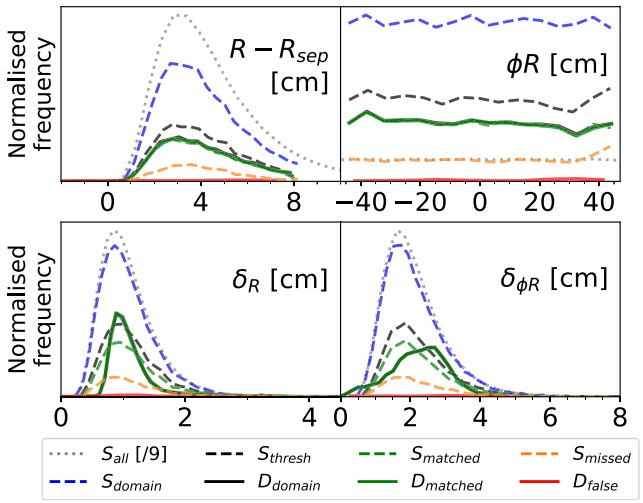

FIG. 14. Comparison of input and output normalized frequency distributions of filament positions and widths for the reference synthetic data set. Dashed and dotted lines indicate sets of synthetic filaments, while solid lines indicate sets of detected filaments recovered from the analysis procedure. The frequencies for the full set of synthetic filaments (dotted gray line) have been divided by 9 to fit on the same axes.

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.

Figure 15 shows the inversion amplitude measured at the true center 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

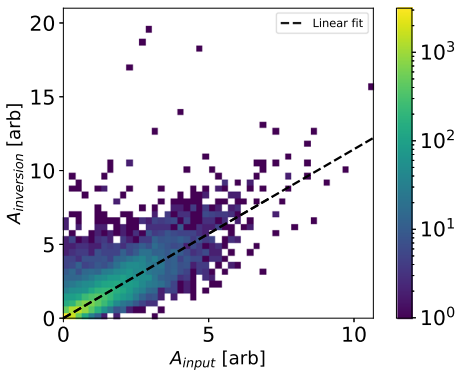

FIG. 15. 2D histogram showing the relationship between synthetic input field line emission amplitude and resulting inversion field line emission amplitude. The color indicates the frequency of each bin (pixel).

outliers, the gradient of a linear fit to the data yields the calibration factor required to compare input amplitudes and inversion amplitudes directly.

Figure 16 compares the distributions of detected filament amplitudes and calibrated synthetic amplitudes, using line styles consistent with Fig. 14. The vertical gray line indicates the amplitude threshold $\epsilon_{thresh} = 1.5 \times 10^{-2}$, below which detections are not attempted. The output distribution successfully captures the exponential nature of the input distribution, but for a roll over at low amplitudes, arising 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 artifact of the measurement technique, other data sets with different amplitude distributions have been analyzed, each containing 5000 frames.

Data set A_1 shares the same experimentally representative parameter distributions as the reference synthetic data set, A, but contains only one filament per frame. This data set is used to assess the effect of filament overlap on results from the reference data set.

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 then provides an estimate of the intrinsic random error on an amplitude measurement.

Data set A_δ contains the same distributions as $A_{\delta,1}$, but with 40 filaments per frame, so that the effects of filament overlap can be readily identified.

Figure 17 shows the calibrated synthetic amplitude distributions for each of these data sets (dashed lines) and the inversion amplitudes measured at the true center 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 dispersed but maintains

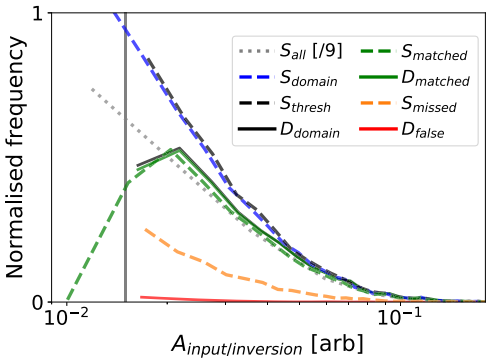

FIG. 16. Synthetic and detected amplitude distributions. Line styles are consistent with Fig. 14. The vertical gray line indicates the amplitude threshold $\epsilon_{thresh} = 1.5 \times 10^{-2}$.

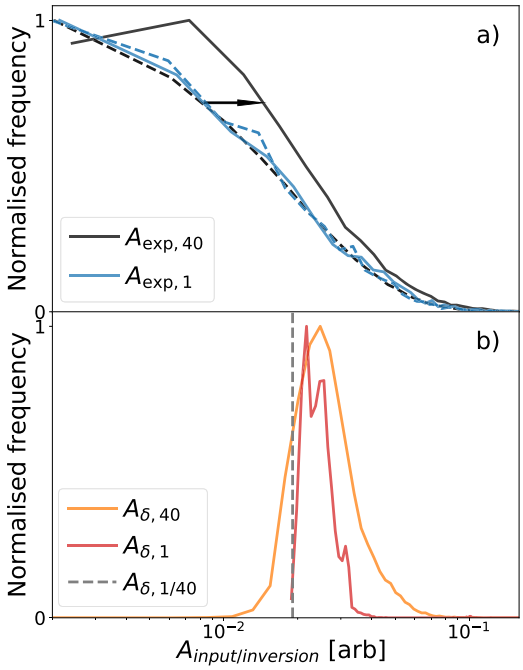

FIG. 17. Calibrated original and measured inversion amplitude distributions for (a) data sets A and A_1 with exponential amplitude distributions and (b) data sets A_δ and $A_{\delta,1}$ with delta function amplitude distributions. Dashed lines are the calibrated input amplitude distributions, while solid lines show the distributions of inversion amplitudes. The vertical dashed gray line indicates the delta function in input amplitudes for the A_δ and $A_{\delta,1}$ data sets.

the peaked, broadly symmetrical shape. Given there is only one filament per frame, this data set demonstrates the intrinsic dispersion in amplitudes resulting from the inversion technique without any filament overlap. The Gaussian width of the measured distribution is around 15% of the mean, indicating that this is the typical size of error that can be expected on amplitude measurements, without filament overlap effects.

When multiple filaments are introduced as in the A_δ data set, the highly peaked distribution shape is retained, but the interactions between filaments lead to further broadening of the distribution to a Gaussian width around 25% of the mean, a shift to higher amplitudes and an enlarged high amplitude tail. It should be noted that while the bulk of these effects will primarily be the result of the overlap of foreground and background filaments and are thus relevant to experimental measurements, the high amplitude tail and a degree of the broadening will be due to overlap of nearby filaments which are not anticipated experimentally.

In the case of the single filament, physically representative data set, A_1, the inversion amplitudes closely follow the input, indicating that in the absence of filament overlap, amplitude distributions are recovered reliably. The addition of multiple filaments per frame in A introduces a constant positive offset due to the overlap of foreground and background filaments raising the average intensity of each projected field line.

As the filaments are uniformly distributed around the machine and there are many possible points of intersection between

different field lines in the field of view of the camera, the average effect of the filaments outside the analysis domain is to provide an effectively uniform background of emission. 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. Hence, a roll over is observed in the measured amplitudes. Therefore, the measured amplitude distribution is representative of the true distribution at all but the lowest amplitudes, except for an approximately constant shift to higher amplitudes due to filament overlap. Given the roll over occurs below the optimal blob detection amplitude threshold, $\epsilon_{thresh} = 1.5 \times 10^{-2}$, and the constant offset does not affect the distribution shape, this should not affect 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 29 852, 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.

Figure 18 shows the measured experimental distributions of filament positions and widths, obtained with the new inversion technique, in black. Given the findings from the synthetic data analysis, orange dashed lines have been added to sketch, without any pretense of rigor, possible inferred distributions for each parameter.

A bimodal distribution is recovered for the R positions, with the peak at low major radius likely due to ghost 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 around the separatrix where both the number of filaments

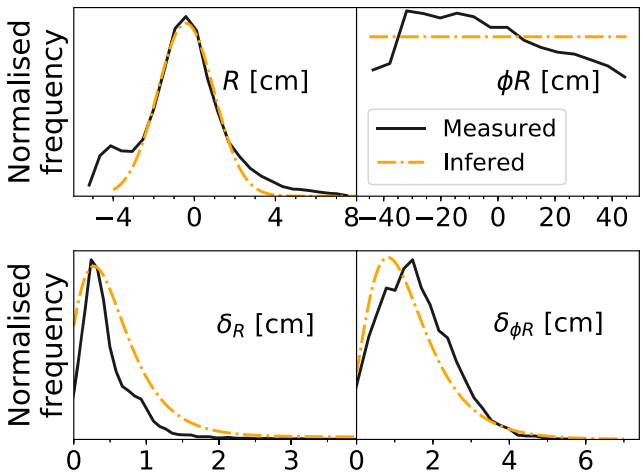

FIG. 18. Distributions of filament parameters for experimental data from MAST shot 29 852, t = (0.205, 0.245) s, with $\epsilon_{thresh} = 1.5 \times 10^{-2}$. Solid black lines show the measured distributions with the systematic error offset correction applied. Dashed orange lines are sketches of the inferred most probable true distributions shapes given the insight gained from the synthetic analysis.

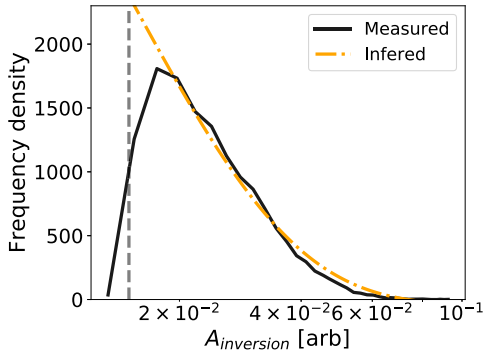

FIG. 19. Distributions of filament amplitudes for experimental data from MAST shot 29 852, t = (0.205, 0.245) s, with $\epsilon_{thresh} = 1.5 \times 10^{-2}$. The solid black line shows the measured distribution. The dashed orange line is a sketch of the inferred most probable true distributions shapes given the insight gained from the synthetic analysis. The vertical dashed line indicates ϵ_{thresh}.

and the neutral particle density is high. Inside the separatrix the detection rate falls off sharply due to the rapid fall off in neutral density, given the neutral ionization mean free path at the separatrix is ~1 cm.

As far as the toroidal distribution is concerned, 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 artifact 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.

Figure 19 shows the measured distributions of filament amplitudes for the experimental data set in black, again with an orange dashed line added to sketch out the possible distribution that can be inferred from the synthetic data analysis. An exponential amplitude distribution is observed with a roll over around ϵ_{thresh}, similar to that observed for the reference synthetic data set. This observation of an exponential amplitude distribution is in agreement with previous measurements with Langmuir probes[15,30] and GPI.[60]

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 an important role in the precision of technique's output. This must be sufficiently high to resolve the smallest filaments of interest, which are of the order of several millimeters across and in general a few ion Larmor radii. However, the distortion and smoothing in the inversion and regularization process 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 coarser 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 for two reasons. First, analyzing a large region makes best use of the available data and produces more filament detections, so that stronger statistical conclusions can be drawn from the larger sample sizes. Second, a large analysis domain maximizes 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 Sec. V was chosen 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 analyzed.

The camera spatial resolution, limited by its pixel resolution and viewing perspective, also limits the accuracy of the filament measurements. 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 discretization 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[61] image processing library used in the prepossessing stages of the analysis requires the images to 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. The smallest, faintest filaments are only a few pixels across and contain pixels with intensities at the lower end of the 8-bit dynamic range (0–255). Therefore, the range of the amplitudes in the contour can be small leading to significant "terracing" of the intensities across the contour. When fitting the 2D Gaussians across only a few points, this can have noticeable effects for fitting to the smallest, faintest filaments. Therefore, avoiding the conversion to 8-bit data for the OpenCV operations would preferable in future.

At 100 kHz the camera's integration time is 3.3 μs, during which fast filaments[9] with total velocities of ~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 source of error that has not been quantified in this benchmarking exercise is that arising from inaccuracies in the magnetic equilibrium and camera calibration, which will never be perfect descriptions of the physical systems. A nonquantitative assessment of these errors has shown they can be significant for poor quality camera calibrations and equilibrium reconstructions. However, if care is taken in producing these inputs their effects on the benchmarking conclusions should not be large. While quantification of these sources of error is an important area of future study, it is beyond the scope of this work.

Future analysis will also explore the effect of impurity emission on the unfiltered camera measurements, particularly measured filament widths. While these effects are expected to be negligible, they shall be assessed using the CHERAB[62] spectrally resolved ray tracing code.

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 nonequilibrium 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.[65] Application of the technique to stereoscopic fast camera data will also be performed to assess the effects of the 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 analyzes 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. While

in the process of finalizing the paper, we learned of a different approach with similar aims presented in Ref. 64, testifying the attention of the community to the subject of experimental filament analysis.

ACKNOWLEDGMENTS

This work has received funding from the Euratom research and training programme 2014–2018 and 2019–2020 under Grant Agreement No. 633053 and from the RCUKEnergy Programme (Grant No. EP/P012450/1) as well as the Fusion CDT Programme through the EPSRC [Grant No. EP/L01663X/1]. To obtain further information on the data and models underlying this paper, please contact PublicationsManager@ccfe.ac.uk.

APPENDIX A: ALTERNATIVE APPROACH TO THE CAMERA ANALYSIS

Section III C describes the standard inversion technique used in this paper. In the moment matrix discussed there, the emission of each field line is given by the sum of the intensities along the field line and it therefore represents the average properties of the emission along the field line. However, a completely different approach to the analysis of the camera data is possible. Indeed, knowing how a given field line projects on the camera image, we can identify how the emission changes along it by tracking down the intensity on each pixel belonging to the said field line. This produces the possibility to investigate different statistical moments of the intensity distribution along the field lines and therefore acquire additional information. By doing this, we have dropped the assumption that the emission is homogeneous along the field lines, which was essential in the main inversion technique. If instead of taking the matrix product of $\mathbb{G}^T$ with $\mathbf{I}$, we take the elementwise product we produce a "weighted geometry matrix,"

$$\mathbb{F}_{ij} = \mathbb{G}_{ij} I_i, \tag{A1}$$

where $\mathbb{F}_{ij}$ is the intensity of the i-th pixel in the camera image multiplied by the corresponding intensity of the i-th pixel in the basis image of the j-th field line. In Eq. (A1), we are suspending the use of the Einstein summation convention. The familiar moment matrix used above can be recovered by taking the sum over each column of F_{ij},

$$\mathcal{E}_j = \sum_i \mathbb{F}_{ij}. \tag{A2}$$

The primary advantage of taking the sum of the pixel intensities along each field line is that the linear operation preserves the ability to perform the subsequent least squares optimization to reduce the distortion of the structures in the moment matrix. The use of a linear field line intensity metric also has the benefit of preserving relative filament amplitudes and internal filament structure.

However, a range of "parallel" operators acting on the columns of weighted geometry matrix, $\mathcal{F}_i(\mathbb{F})$, other than the sum $\mathcal{F}_i = \sum_i$ can quantify the emission along the field lines in different ways that also provide useful information for the analysis of field aligned plasma structures, as expressed by

$$\xi_j = \mathcal{F}_i(\mathbb{F}_{ij}). \tag{A3}$$

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.), $\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$, which penalizes small values in the data series more effectively than the arithmetic mean (A. M.), $\frac{1}{n}\sum_{i=1}^{n} x_i$, is effective at differentiating field lines that consistently lie on top of filaments from those that do not. While it is not compatible with a subsequent least-squares optimization due to the nonlinearity of the operation, the G.M. is effective at identifying and measuring the position of filaments with far less computational expenditure. However, it is less appropriate for measuring filament widths and amplitudes. 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, other metrics such as the 2nd and 98th percentiles of the distribution of intensities along a field line, or parameterizations of smoothness, have potential for identifying ghost filaments. This is because they are sensitive to the short portions of field lines with particularly low intensities that typically occur where parent filaments intersect and the alignment of the shadow field lines to the parent structures is poorest.

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.

APPENDIX B: SYNTHETIC CAMERA DATASETS

Four synthetic datasets each consisting of 5000 frames were used in this work. In 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,

TABLE IV. Distribution information for filament parameters in synthetic data sets A and A_1. Due to the parameterization of the field lines in ϕR space, the inversion domain is slightly wedge shaped. To account for this filaments are not generated in a 5° toroidal wedge. The center of the distribution is set such that this wedge is on the far side of the machine where the field lines are not in the camera field of view.

Property		Distribution	Parameters	References
Position	R	$\lambda\mathcal{L}+\nu$	$\mu=0,\sigma=0.5$ $\lambda=0.04,\nu=1.36$	27 and 28
	ϕ	$\lambda\mathcal{U}+\nu$	$\lambda=355,\nu=2$	58
Width	δ_R	$\lambda\mathcal{L}$	$\mu=0,\sigma=0.4$ $\lambda=0.01$	12 and 36
	$\delta_{\phi R}$	$\lambda\mathcal{L}$	$\mu=0,\sigma=0.4$ $\lambda=0.8$	12 and 36
Amplitude	A	$\mathcal{E}$	$\mu=0,\sigma=0$	18, 30, and 59

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

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	$\mu=0.02$

and $\delta(x;\mu)=\delta(x-\mu)$ represents the Dirac delta distribution. λ and ν are used as scaling and offset parameters, respectively.

1. Reference synthetic data set

The reference synthetic data set, A, is the primary synthetic data set 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 filaments, giving good resemblance to experimental data.

2. Single filament physical data set

The A_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 integration.

3. 40 filament simplified data set

The A_δ data set was produced with the same spatial distribution of filaments as in the reference synthetic data set, but the amplitude and width distributions were replaced by delta functions (see Table V).

4. Single filament simplified data set

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

REFERENCES

[1] P. Stangeby, *The Plasma Boundary of Magnetic Fusion Devices*, edited by P. Stott and H. Wilhelmsson (IOP Publishing, London, 2000).
[2] W. Fundamenski, J. Nucl. Mater. **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, and ITPA Scrape-off Layer and Divertor Physics Topical Group, Nucl. Fusion **39**, S203 (1999).
[4] J. R. Myra and D. A. D'Ippolito, AIP Conf. Proc. **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, Nucl. Fusion **57**, 076021 (2017).
[6] S. Krasheninnikov, A. Pigarov, and D. Sigmar, Phys. Lett. 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, Nucl. 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, J. Nucl. Mater. **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, Phys. Plasmas **10**, 1670 (2003).
[10] D. A. D'Ippolito, J. R. Myra, and S. J. Zweben, Phys. Plasmas **18**, 060501 (2011).
[11] F. Militello and J. T. Omotani, Nucl. Fusion **56**, 104004 (2016); e-print arXiv:1604.08790v1.
[12] A. Kirk, A. J. Thornton, J. R. Harrison, F. Militello, N. R. Walkden, MAST Team, and EUROfusion MST1 Team, Plasma Phys. Controlled Fusion **58**, 085008 (2016); e-print arXiv:1602.03021.
[13] A. J. Thornton, G. Fishpool, A. Kirk, MAST Team,and EUROfusion MST1 Team, Plasma Phys. Controlled. Fusion **57**, 115010 (2015).
[14] O. Garcia, R. Pitts, J. Horacek, A. Nielsen, W. Fundamenski, J. Graves, V. Naulin, and J. J. Rasmussen, J. Nucl. Mater. **363**, 575 (2016); e-print arXiv:1604.08790.
[15] J. P. Graves, J. Horacek, R. A. Pitts, and K. I. Hopcraft, Plasma Phys. 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, Phys. Plasmas **8**, 2107 (2001).
[17] F. Militello, P. Tamain, W. Fundamenski, A. Kirk, V. Naulin, and A. H. Nielsen, Plasma Phys. Controlled Fusion **55**, 025005 (2013); e-print arXiv:1305.5064v1.
[18] N. R. Walkden, F. Militello, J. Harrison, T. Farley, S. Silburn, and J. Young, Nucl. Mater. Energy **12**, 175 (2017).
[19] G. Y. Antar, G. Counsell, Y. Yu, B. Labombard, and P. Devynck, Phys. Plasmas **10**, 419 (2003).
[20] S. H. Müller, A. Diallo, A. Fasoli, I. Furno, B. Labit, and M. Podestà, Phys. Plasmas **14**, 110704 (2007).
[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 NSTX Team, Nucl. Fusion **44**, 134 (2004).
[22] Y. Sechrest, D. Smith, D. P. Stotler, T. Munsat, and S. J. Zweben, Phys. Plasmas **22** (2015).
[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, Phys. Plasmas **8**, 4826 (2001).
[24] A. Huber, U. Samm, B. Schweer, and P. Mertens, Plasma Phys. Controlled 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, Rev. Sci. Instrum. **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, J. Boedo, D. A. D'Ippolito, D. P. Stotler, S. J. Zweben, B. P. LeBlanc, J. E. Menard, R. J. Maqueda, and J. Boedo, Phys. 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, Nucl. 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 Phys. Controlled Fusion **58**, 044007 (2016).
[29]R. Kube, O. E. Garcia, B. LaBombard, J. L. Terry, and S. J. Zweben, J. Nucl. Mater. **438**, S505 (2013).
[30]O. E. Garcia, S. M. Fritzner, R. Kube, I. Cziegler, B. Labombard, and J. L. Terry, Phys. Plasmas **20**, 055901 (2013); e-print arXiv:1801.00612.
[31]G. Fuchert, G. Birkenmeier, D. Carralero, T. Lunt, P. Manz, H. W. Müller, B. Nold, M. Ramisch, V. Rohde, and U. Stroth, Plasma Phys. Controlled Fusion **56**, 125001 (2014).
[32]D. Goodall, J. Nucl. Mater. **111-112**, 11 (1982).
[33]S. J. Zweben and S. S. Medley, Phys. Fluids B **1**, 2058 (1989).
[34]A. Kirk, N. B. 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 Phys. 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. B. Ayed, A. Kirk, B. Dudson, S. Tallents, R. G. L. Vann, and H. R. Wilson, Plasma Phys. 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, ASDEX Upgrade, and MAST Teams, Plasma Phys. 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. B. Ayed, A. Kirk, H. R. Wilson, G. Counsell, X. Xu, M. Umansky, P. B. Snyder, and B. LLoyd, Plasma Phys. 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]L. C. Ingesson, B. Alper, B. J. Peterson, and J. C. Vallet, Fusion Sci. Technol. **53**, 528 (2008).
[41]R. N. Van Yen, N. Fedorczak, F. Brochard, G. Bonhomme, K. Schneider, M. Farge, and P. Monier-Garbet, Nucl. Fusion **52**, 013005 (2012).
[42]O. Grulke, J. L. Terry, B. LaBombard, and S. J. Zweben, Phys. Plasmas **13**, 012306 (2006).
[43]E. Havlíčková, J. Harrison, B. Lipschultz, G. Fishpool, A. Kirk, A. Thornton, M. Wischmeier, S. Elmore, and S. Allan, Plasma Phys. Controlled Fusion **57**, 115001 (2015).
[44]C. Tomasi and R. Manduchi, in *Sixth International Conference on Computer Vision* (Bombay, 1998), pp. 839–846, e-print arXiv:1011.1669v3.
[45]B. M. ter Haar Romeny, L. M. J. Florack, M. de Swart, J. Wilting, and M. A. Viergever, Proc. SPIE **2299**, 139–148 (1994).
[46]L. C. Appel and I. Lupelli, Comput. Phys. Commun. **223**, 1 (2018).
[47]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* (EPS, 2006), Vol. 2, pp. 1235–1238.
[48]J. Heikkila and O. Silven, in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE, San Juan, 1997).
[49]R. C. Thurber, B. Aster, and C. H. Borchers, *Parameter Estimation and Inverse Problems*, 3rd ed. (Elsevier, 2013).
[50]T. Hobiger, T. Kondo, and Y. Koyama, Earth, Planets Space **60**, 727 (2008).
[51]M. G. H. Bell, Transp. Res., Part B: Methodol. **25**, 13 (1991).
[52]G. Golub and W. Kahan, J. Soc. Ind. Appl. Math., Ser. B: Numer. Anal. **2**, 205 (2005).
[53]G. Strang, in *Linear Algebra and its Applications*, 4th ed. (Thomson Brooks, Belmont, 2006), Vol. 1, e-print arXiv:0712.0689.
[54]J. Demmel and W. Kahan, SIAM J. Sci. Stat. Comput. **11**, 873 (1990).
[55]P. C. Hansen, Inverse Probl. **8**, 849 (1992).
[56]D. Carralero, G. Birkenmeier, H. Müller, P. Manz, P. DeMarne, S. Müller, F. Reimold, U. Stroth, M. Wischmeier, and E. Wolfrum, Nucl. Fusion **54**, 123005 (2014); e-print arXiv:1407.3618v1.
[57]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 Phys. Controlled Fusion **58**, 105002 (2016).
[58]F. Militello, T. Farley, K. Mukhi, N. Walkden, and J. T. Omotani, Phys. Plasmas **25**, 056112 (2018).
[59]O. Garcia, J. Horacek, and R. Pitts, Nucl. Fusion **55**, 062002 (2015).
[60]O. E. Garcia, I. Cziegler, R. Kube, B. LaBombard, and J. L. Terry, J. Nucl. Mater. **438**, S180 (2013); e-print arXiv:1205.6996v1.
[61]G. Bradski, "The OpenCv Library," Dr. Dobb's Journal of Software Tools **2000**, 2236121.
[62]M. Carr, A. Meakins, M. Bernert, P. David, C. Giroud, J. Harrison, S. Henderson, B. Lipschultz, and F. Reimold, Rev. Sci. Instrum. **89**, 083506 (2018).
[63]B. Cannas, S. Carcangiu, A. Fanni, T. Farley, F. Militello, A. Montisci, F. Pisano, G. Sias, and N. Walkden, Fusion Eng. Des. **146**, 374 (2019).
[64]J. Cavalier, N. Lemoine, F. Brochard, V. Weinzettl, J. Seidl, S. Silburn, P. Tamain, R. Dejarnac, J. Adamek, and R. Panek, Nucl. Fusion **59**, 056025 (2019).
[65]See https://photron.com/fastcam-sa1-1 for Photron SA1 specification.