Towards Open Machine Learning Benchmarks for Tokamak Event Prediction from MAST

P. Sharma[1]*, S. Jackson[1], N. Cummings[1], C. J. Ham[1], J. Hodson[1], A. Kirk[1], K. Lawal[1], D. Ryan[1], S. Pamela[1], and The MAST Team

[1] United Kingdom Atomic Energy Authority (UKAEA), Culham Campus, United Kingdom

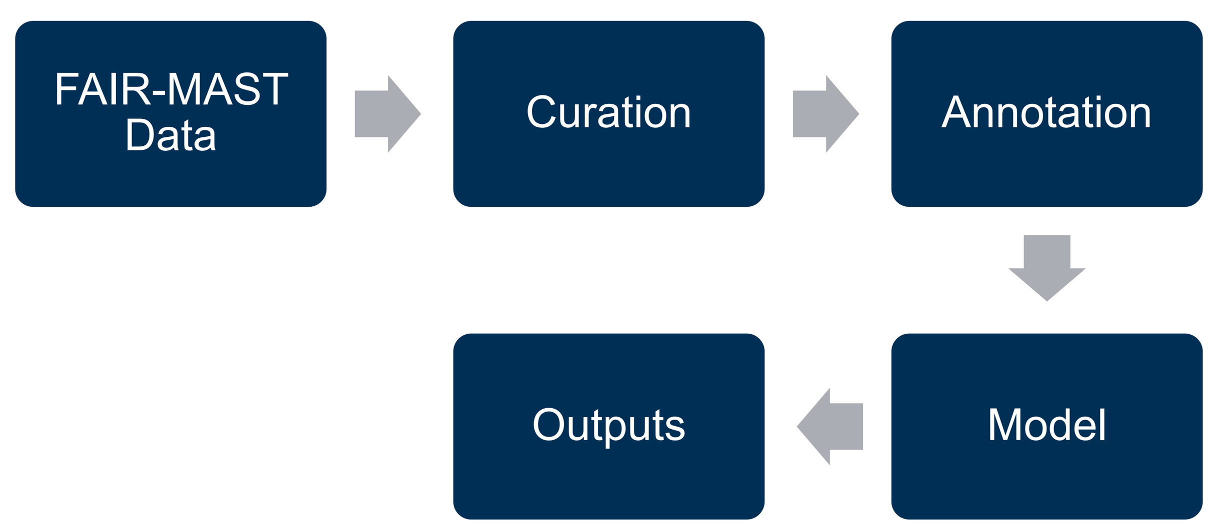
UK Atomic Energy Authority

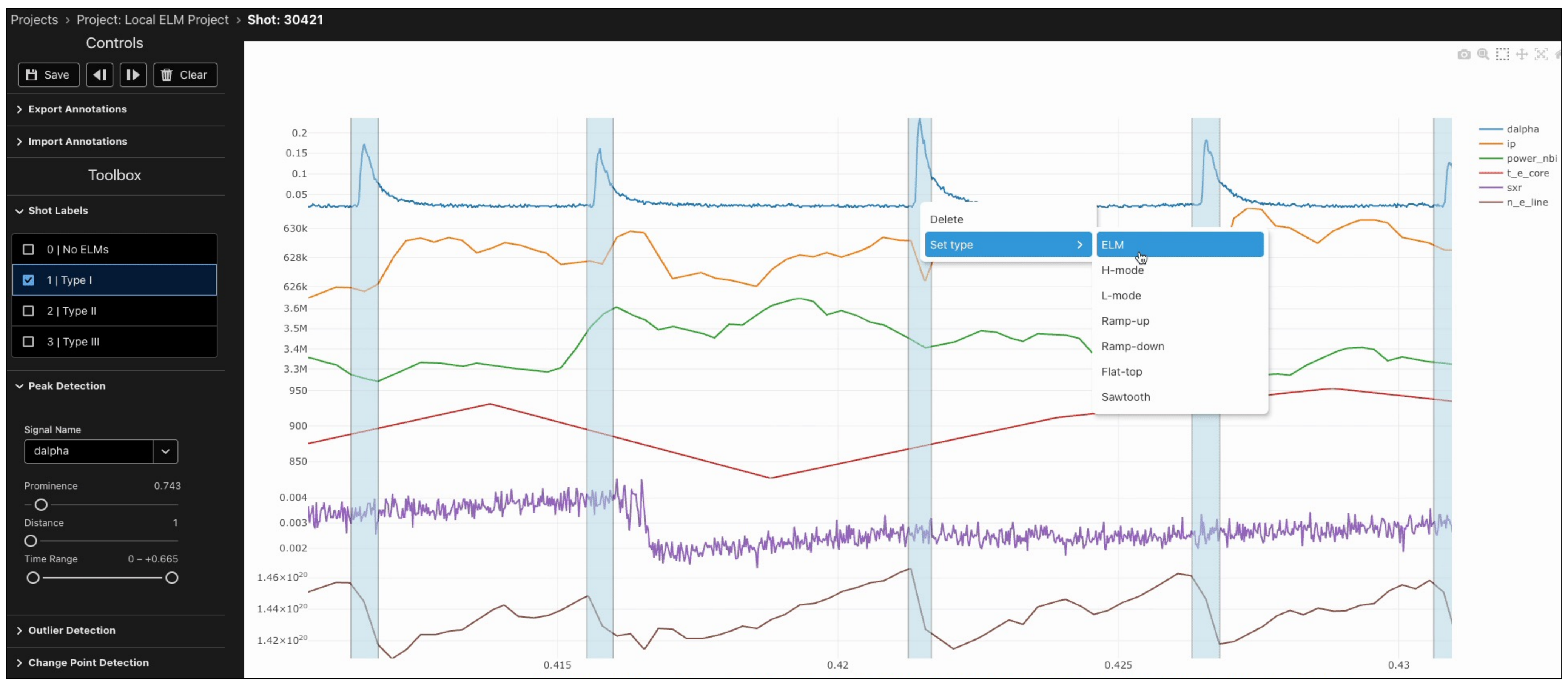
Abstract

- Reliable prediction of disruptions, MHD modes, confinement transitions, and ELMs is essential for safe tokamak operation.
- FAIR-MAST [1,2] provides open access to MAST diagnostic data, but further processing is required to make it "AI-ready".
- We curate annotations and develop baseline ML models for four tasks: disruption, MHD segmentation, confinement mode, and ELMs.
- Baselines provide starting points for reproducible studies and future benchmarks.

Introduction

- FAIR-MAST signals curated and prepared for ML tasks.
- Annotation process (see right).
- Label review + model feedback loop (under construction).
- Outputs: baseline models and metrics.





Disruption prediction

- Theory: Sudden loss of plasma confinement; must predict with warning time [3].
- Ground truth: Auto-detected from plasma current (417 shots).
- Model: Stacked BiLSTM with weighted sampling and sliding time window.
- Limitation: Ambiguous cases without sharp current drop remain unresolved.
- Early predictions when flat-top phase is unclear.
- Performance declines as lead time increases (tested 10/30/60 ms).

MHD mode segmentation

- Theory: Plasma instabilities (LLM, fishbones, NTMs, sawteeth) reduce performance and may trigger disruptions [4].
- Ground truth: Semi-automated spectrogram annotation from Mirnov coils (85 shots; 51 containing LLM).
- Model: Mask R-CNN with ResNet-101 backbone.
- Limitation: High-frequency structures hard to label; non-expert annotations.
- IoU is extremely low because LLMs are thin structures; small misalignments penalise overlap heavily.

Confinement mode classification

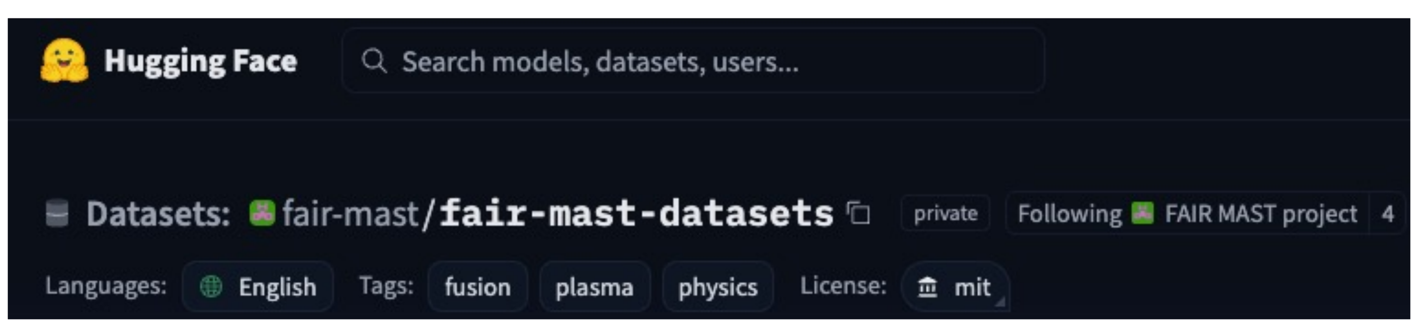
- Theory: Transition between L-mode (low confinement) and H-mode (high confinement) [5].
- Ground truth: H-mode intervals hand-labelled by expert (85 shots).
- Model: 1D U-Net with sliding time window.
- Limitation: Label boundaries may misalign by tens of ms.

ELM spike identification

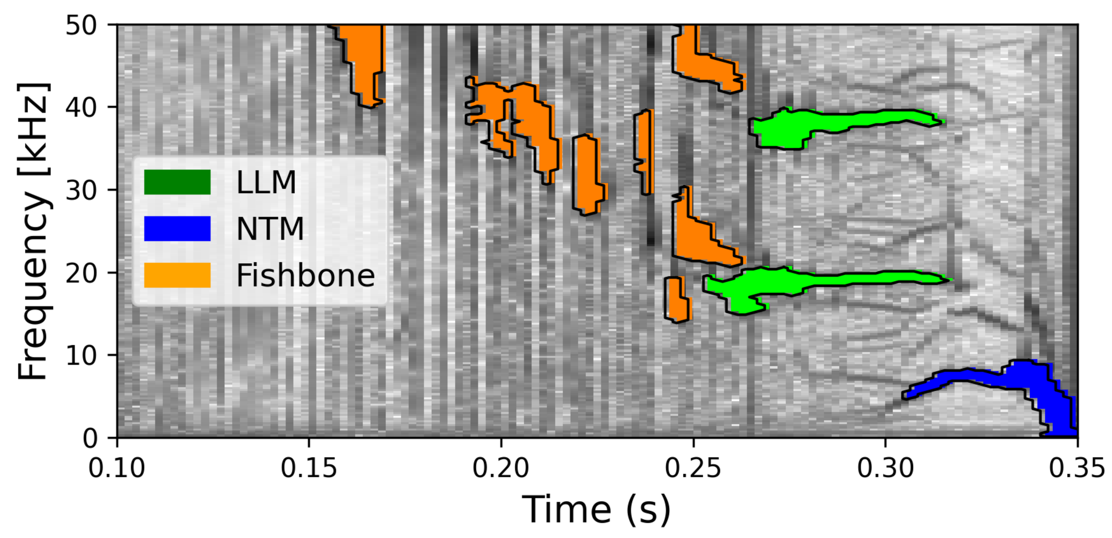
- Theory: Short bursts during H-mode, crucial for plasma–wall interaction [6].
- Ground truth: Thresholding on Dα + manual verification (101 shots).
- Model: 1D U-Net with sliding time window.
- Limitation: Narrow spikes → metrics sensitive to small misalignments.

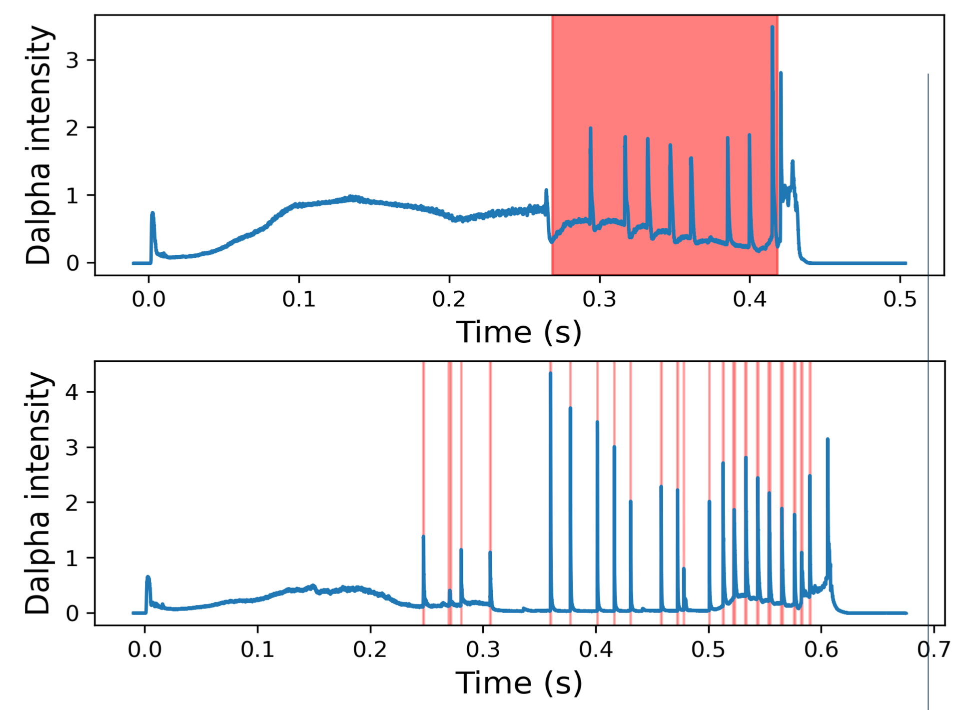
Conclusion & Future Work

- Baselines provide starting points but are not yet full benchmarks.
- Limitations include annotation noise and label misalignments.
- Current metrics do not fully capture thin/filamentary structures.
- Future work:
  - Improve label quality via review + model feedback.
  - Extend baselines towards an open benchmark suite with annotation and data tools.
  - Release datasets and models openly for community use.

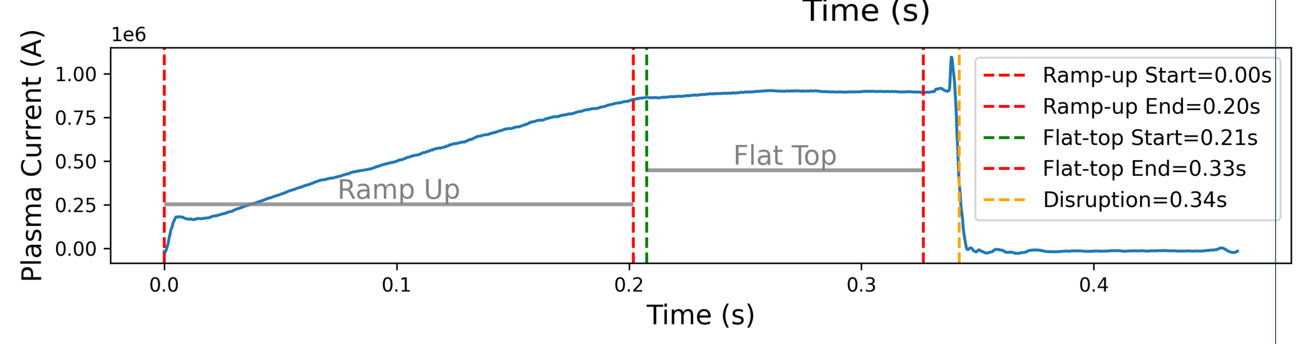


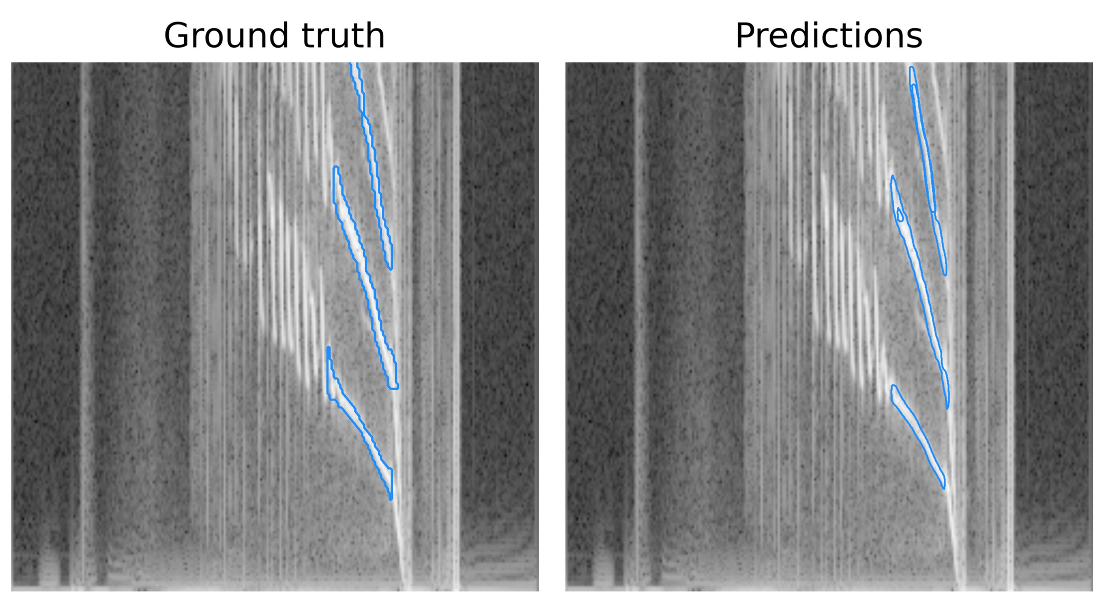
AI-ready FAIR-MAST dataset (under construction)

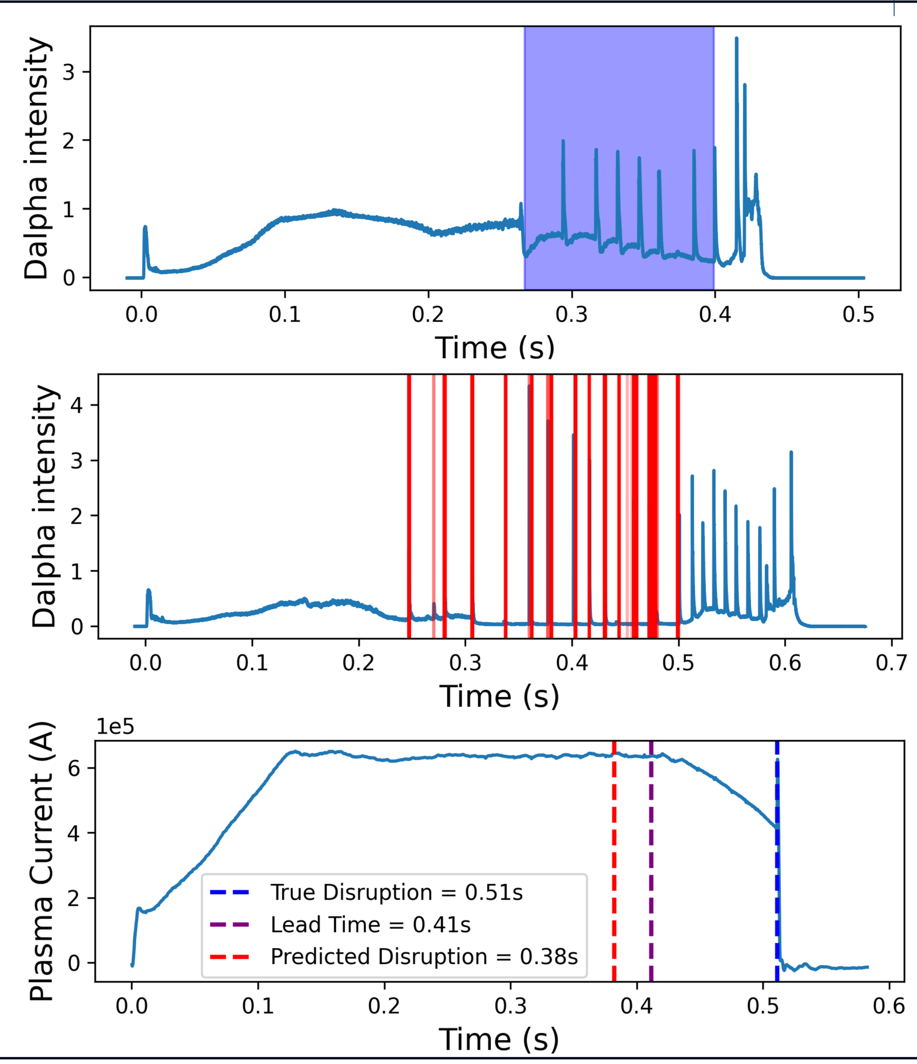




Ground truth examples (different shots shown for illustration):
Top-left: MHD mode segmentation
Top-right: Confinement mode classification
Middle-right: ELM spike identification
Bottom-right: Disruption prediction







Baseline model predictions:
Top-left: MHD segmentation (predicted vs labelled contours).
Top-right: Confinement classification (predicted H-mode interval).
Middle-right: ELM detection (predicted spikes in Dα).
Bottom-right: Disruption prediction (lead time vs true disruption).

Metrics

- **Accuracy omitted:** dominated by background; misleading for event detection.
- **Classification** - *Confinement & ELMs:* Precision/Recall/F1 + ROC AUC. *Disruption:* Precision/Recall/F1 only; **ROC AUC omitted** (unreliable under extreme temporal imbalance + windowing).
- **Disruption early-warning:** hit-rate 0.63; median warning 26.5 ms. Alarms earlier than 50 ms are counted as premature (not true positives).
- **Segmentation (MHD):** report IoU; note it heavily penalises thin modes (LLMs).

Task	Confinement	ELMs	MHD modes	Disruption
Precision	0.82 ± 0.22	0.79 ± 0.20	0.75 ± 0.13	0.84 ± 0.07
Recall	0.83 ± 0.21	0.80 ± 0.20	0.73 ± 0.15	0.94 ± 0.06
F1-score	0.79 ± 0.25	0.78 ± 0.20	0.72 ± 0.10	0.87 ± 0.09
IoU	-	-	0.39 ± 0.01	-
ROC AUC	0.90	0.85	-	-

[1] JACKSON et al. 2024, SoftwareX.
[2] JACKSON et al. 2025, IEEE Trans. Plasma Sci.
[3] SABBAGH et al. 2023, Phys. Plasmas.
[4] THORNTON A., PhD Thesis, University of York (2011).
[5] CONNOR et al. 2000, Plasma Phys. Control. Fusion.
[6] ZOHM et al. 1996, Plasma Phys. Control. Fusion.

The UK Atomic Energy Authority's mission is to lead the delivery of sustainable fusion energy and maximise scientific and economic benefit

Find out more
www.gov.uk/ukaea

Acknowledgements
We thank the MAST team for providing experimental data. Computations used the Cambridge CSD3 cluster.
The MHD modes task was carried out in collaboration with Enrique Zapata Cornejo (PSFC, MIT).