Disruption Prevention Via Interpretable Data-Driven Algorithms On DIII-D And EAST

Cristina Rea[1]
crea@mit.edu

with K.J. Montes[1], W. Hu[2], J. Barr[3], K. Erickson[4], R.S. Granetz[1]
Q.P. Yuan[2], B. Sammuli[3], D. Chen[2], B. Shen[2], B.J. Xiao[2]
the DIII-D, and the EAST team

[1]MIT PSFC, Cambridge, MA, USA
[2]ASIPP, Hefei, China
[3]General Atomics, San Diego, CA, USA
[4]PPPL, Princeton, NJ, USA

28th IAEA Fusion Energy Conference
Virtual Meeting
May 10-15, 2021

Detailed outline

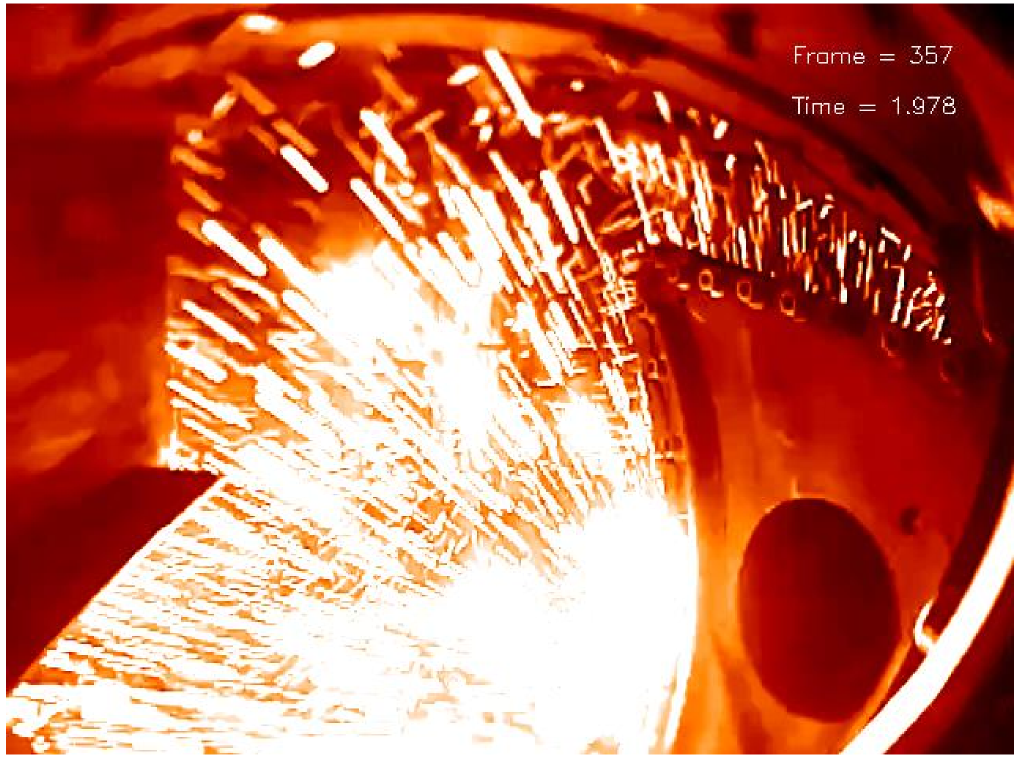
I. **Intro and Motivations [slides 2-5]**
   a) Disruptions as final loss of control (3), interpretable algorithms to aid in active monitoring of soft and hard limits (4-5)

II. **Disruption Prediction via Random Forest (DPRF) [slides 6-18]**
   a) Previous results (6)
      1. More details on RF methodology in backup slides (23-25)
   b) DIII-D upgrades: DPRF2.0 (7-14)
      1. Improved training set (7) and input features (8-9)
      2. Off-normal detection closed-loop experiments (11-12)
      3. Proximity control integration (13-14)
   c) EAST implementation and closed-loop experiments (15-18)

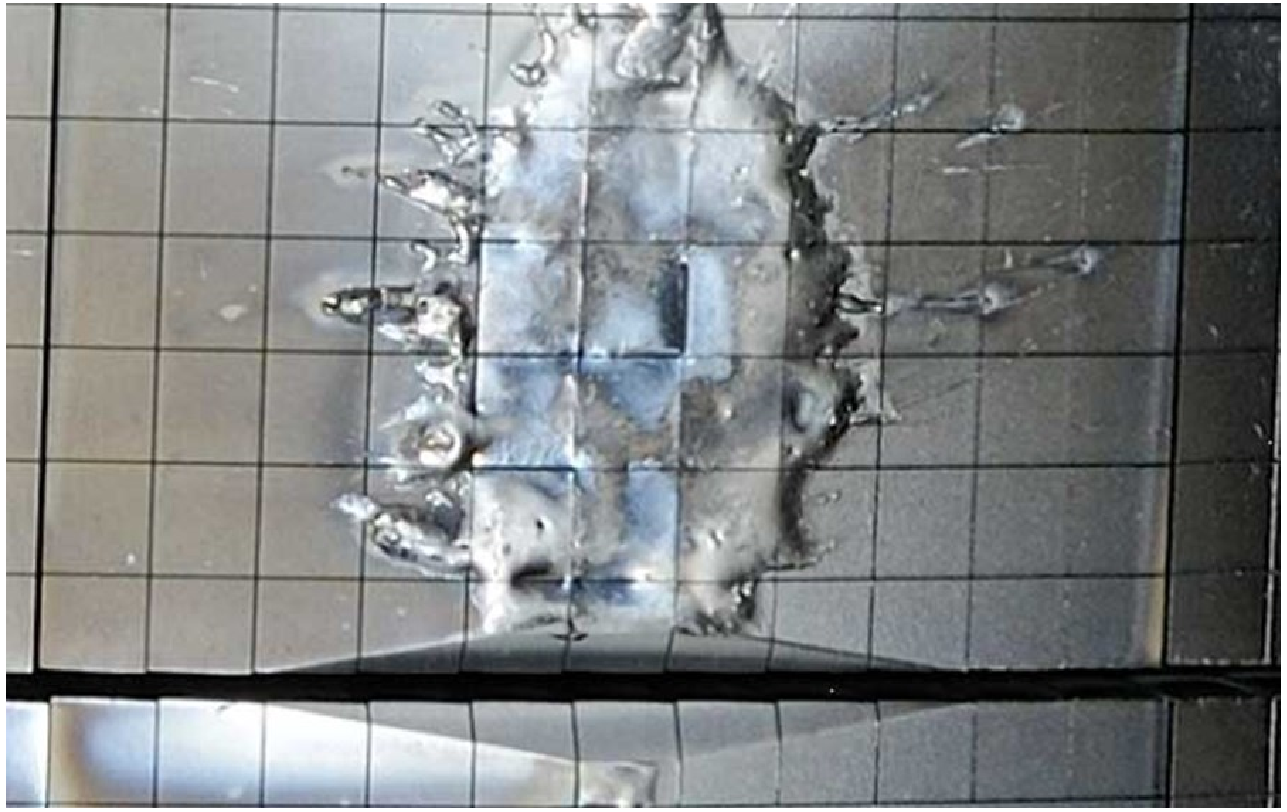
III. **Summary And Conclusions [slides 19-20]**

Plasma pushed close to operational limits often leads to instabilities onset or control faults: unintentional disruptions

- Disruptions related to **peak plasma performances:** higher stored energy, longer confinement times…
  - Consequences: melting/ablation of plasma facing components, thermal loads, mechanical stresses,…
- **Real-time prediction** and **avoidance**, with **mitigation**, mandatory when scaling to reactor sizes and forces.



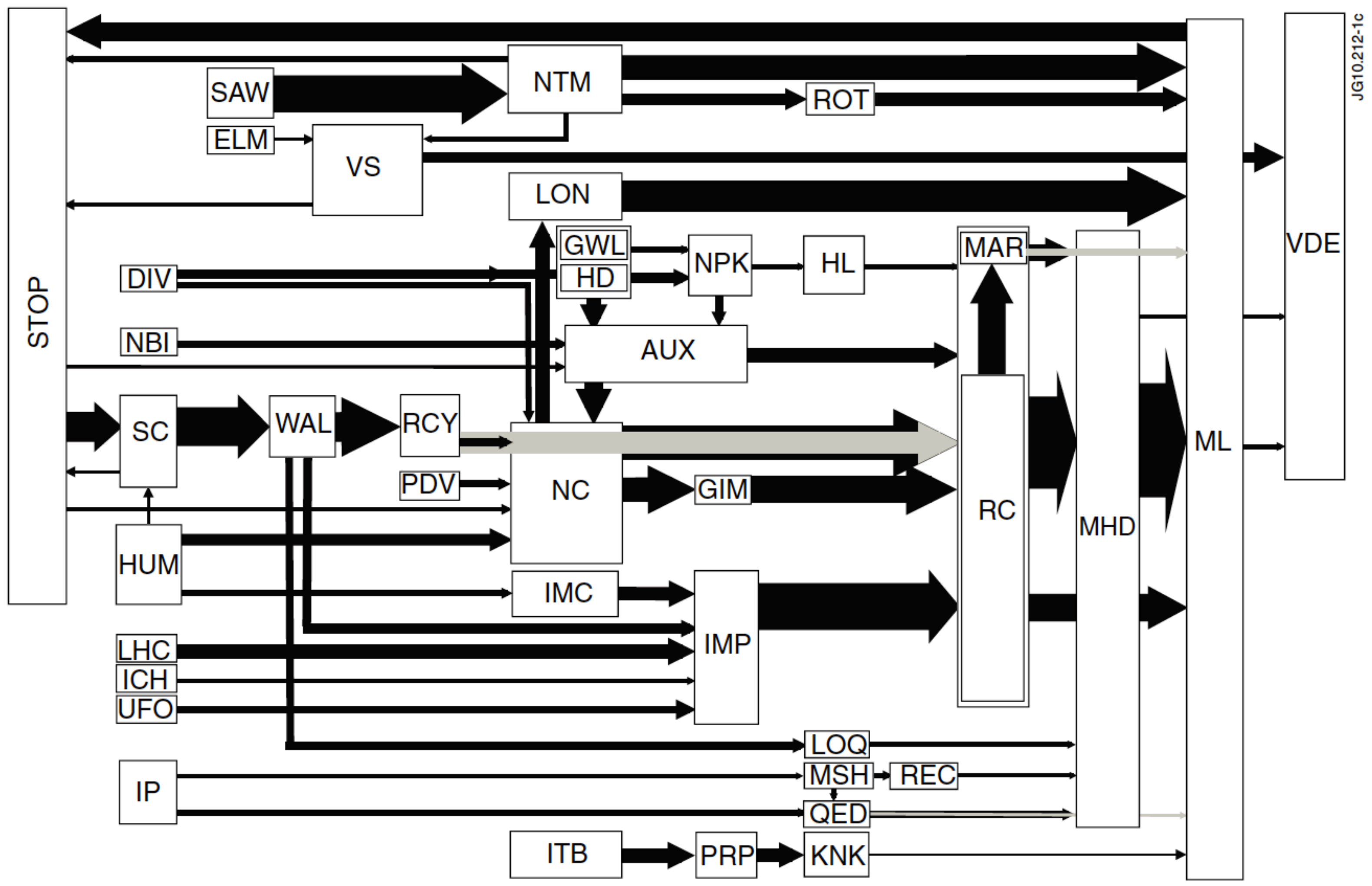
View from visible camera of disruption on Alcator C-Mod. Courtesy R.A. Tinguely.



JET runaway damage.
https://www.iter.org/newsline/-/2234

Statistical studies show complex chains of events: need timely identification of unstable events

possible disruptive chains of events

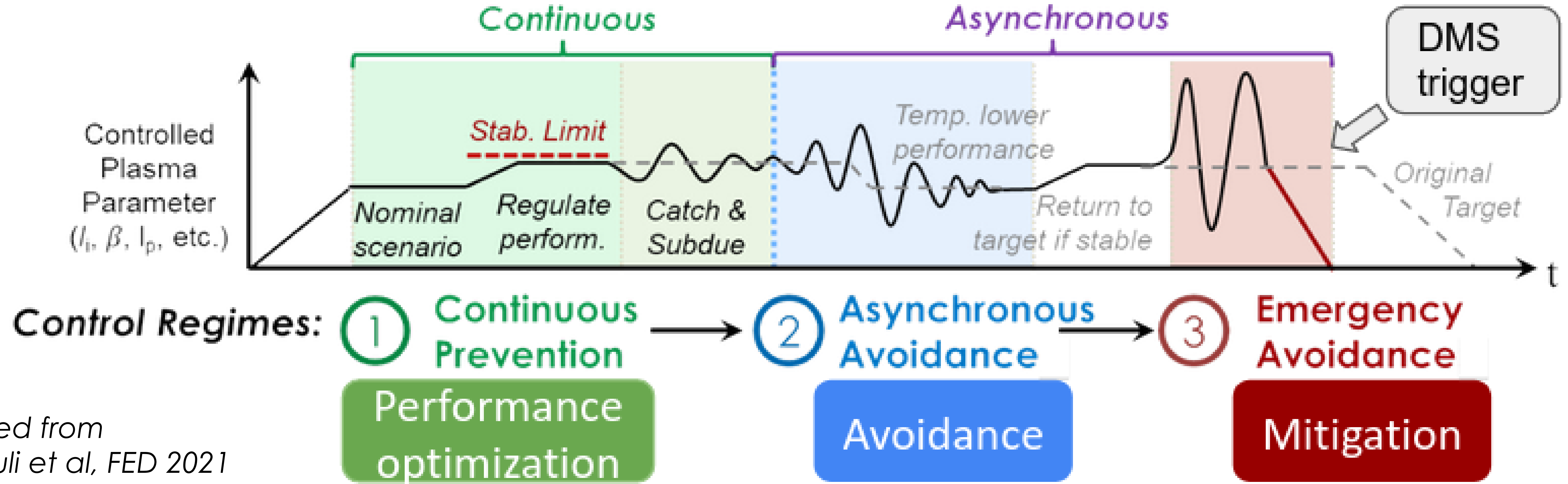


De Vries et al. NF 51 (2011) 053018 "Survey of disruption causes at JET"

- Statistics of the sequence of events for ~10yrs of **unintentional disruptions** at **JET**: width of the connecting arrows indicates the **frequency** of **event occurrence**;
- **Similar studies** are **not always available** across different tokamaks.

Disruptions as final loss of control: successful pre-disruptive event identification can inform plasma controller on proper actuators to use.

Active monitoring and prediction of soft/hard limits necessary to inform transition across ops boundaries



Adapted from Sammuli et al, FED 2021

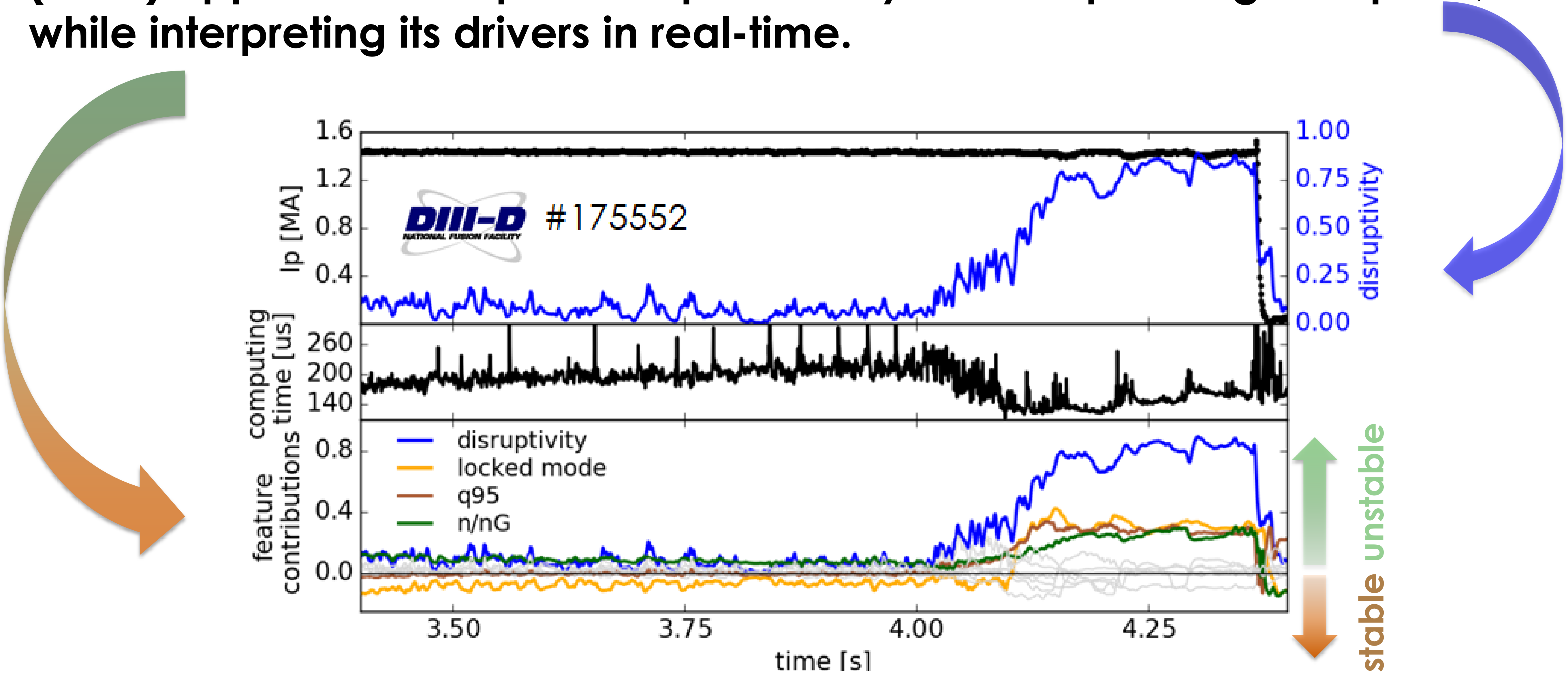
- **Proximity to stability boundaries needs to be actively controlled by the PCS, managing different actuators for different tasks.**
- **Disruption Free Protocol* @DIII-D qualify solutions for different control regimes.**

*J. Barr et al 2021, 28th IAEA FEC, EX/5-TH/6

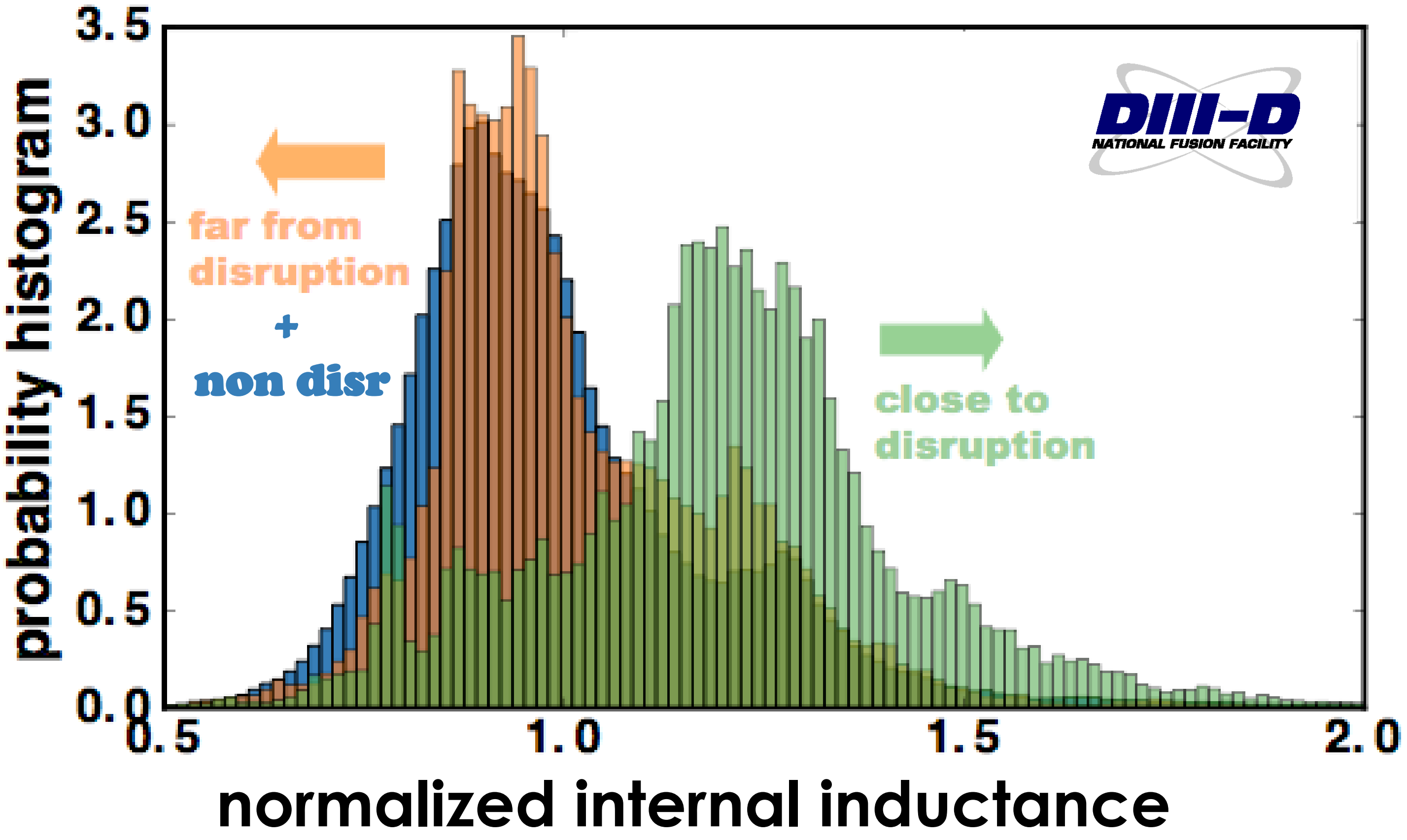
Interpretable data-driven models provide general proximity to unstable ops space.

Interpretable ML models for disruption prediction useful resources to identify stability boundaries in real-time

- **DIII-D and EAST: the Disruption Prediction via Random Forest algorithm (DPRF) applied to compute the probability of an impending disruption, while interpreting its drivers in real-time.**



DPRF supervised binary classification algorithm: identify transitions from non disruptive to disruptive phases



DPRF is based on the **Random Forest*** ensemble algorithm → collection of decision trees: 

Provides **metrics of interpretability**.

**L. Breimann, Machine Learning 45, 5–32 (2001)*

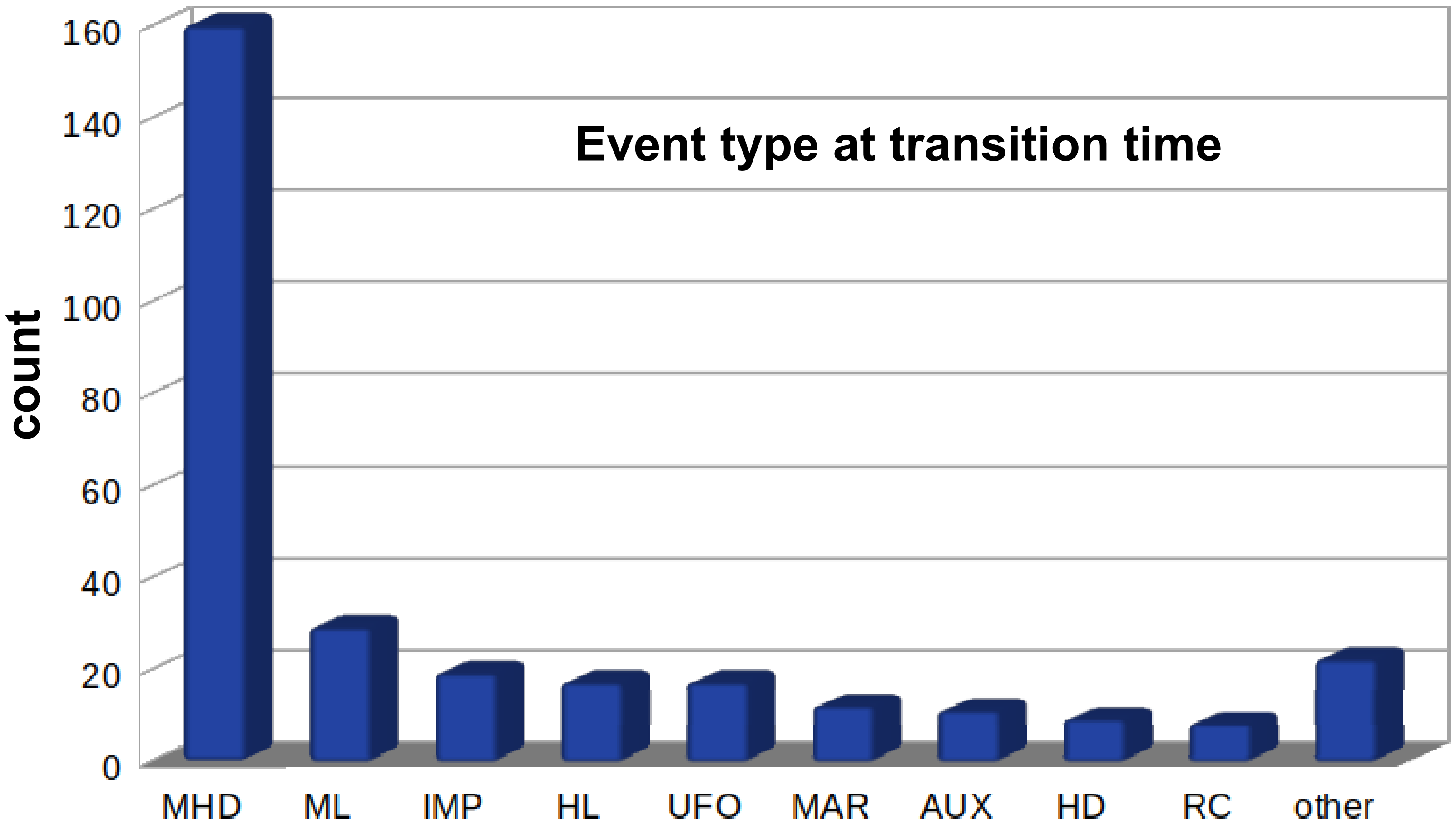
- **Fixed time for transition from safe to disruptive operational space.**
- **Training set: thousands of discharges, agnostic to disruption type.**
- **Offline cross-machine investigation 0-D features (flattop data).**
- **DIII-D Real-time implementation in FY18-19.**

→ DPRF 2.0

C. Rea and R.S. Granetz, Fus. Science Tech. 74 (2018)
C. Rea et al., Plasma Phys. Control. Fusion 60 (2018)
C. Rea et al., Nucl. Fusion 59 (2019)
K. Montes, C. Rea et al., Nucl. Fusion 59 (2019)

Upgrades to DIII-D DPRF through improved training set and input features: DPRF 2.0

- **Improved label classification by detecting transitions between specific operational boundaries on a shot-by-shot basis.** A. Pau et al., *Nucl. Fusion* 59 (2019)



Tags from De Vries et al. NF 51 (2011) 053018
"Survey of disruption causes at JET"

- **Unstable events manually identified** > 300 DIII-D discharges (Montes).
- ML algorithms: training composition can affect the model sensitivity towards certain scenarios.
- **Need (automated) identification of disruption causes.**

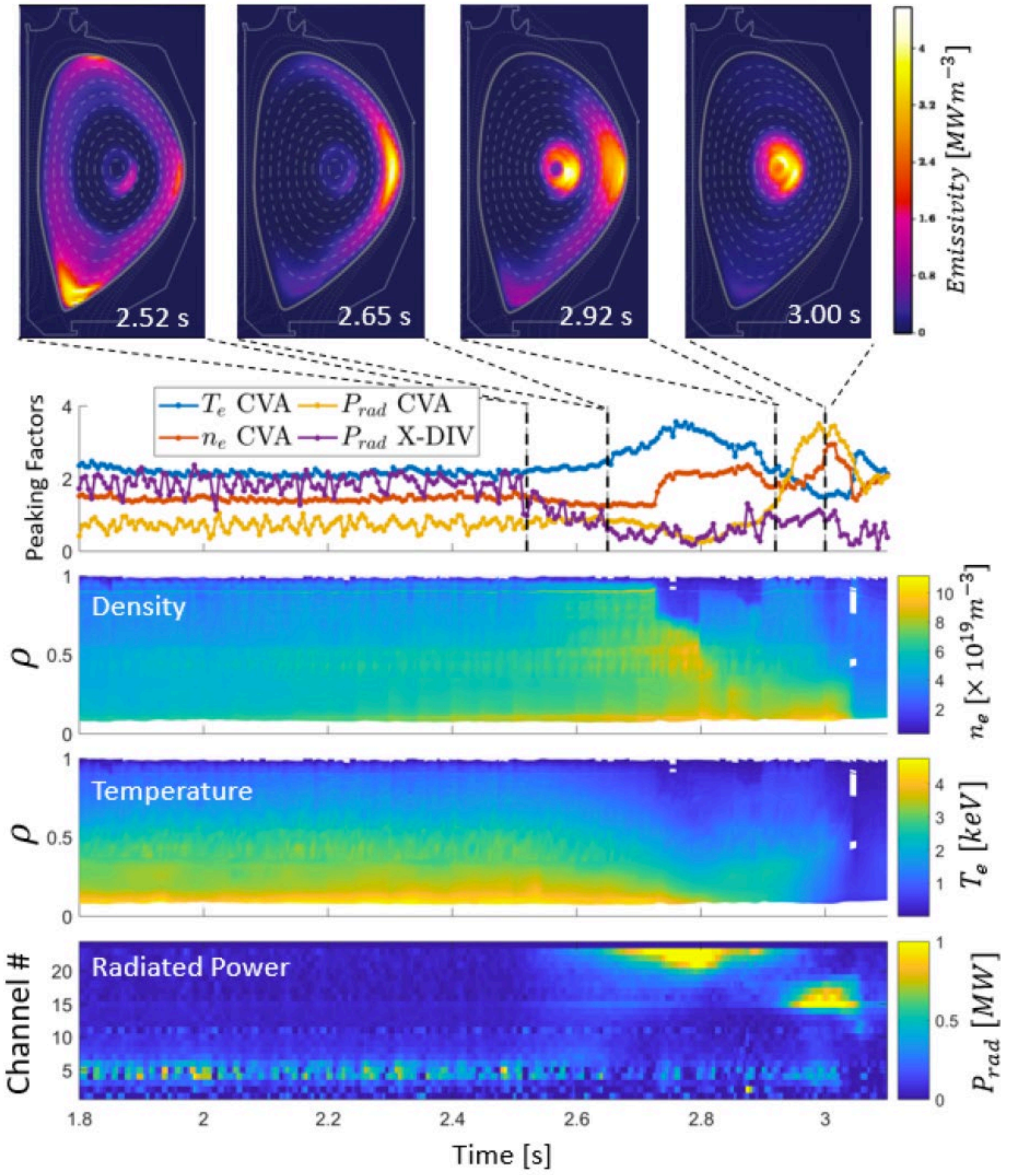
DPRF 2.0: radial peaking factors added to other 0-D inputs to detect earlier instability onset

- **1D/2D profile information** compressed into **peaking factors**.

- Profile diagnostics mapped onto flux surfaces or core / divertor regions.

Peaking factors are interpretable, easy to calculate in real-time

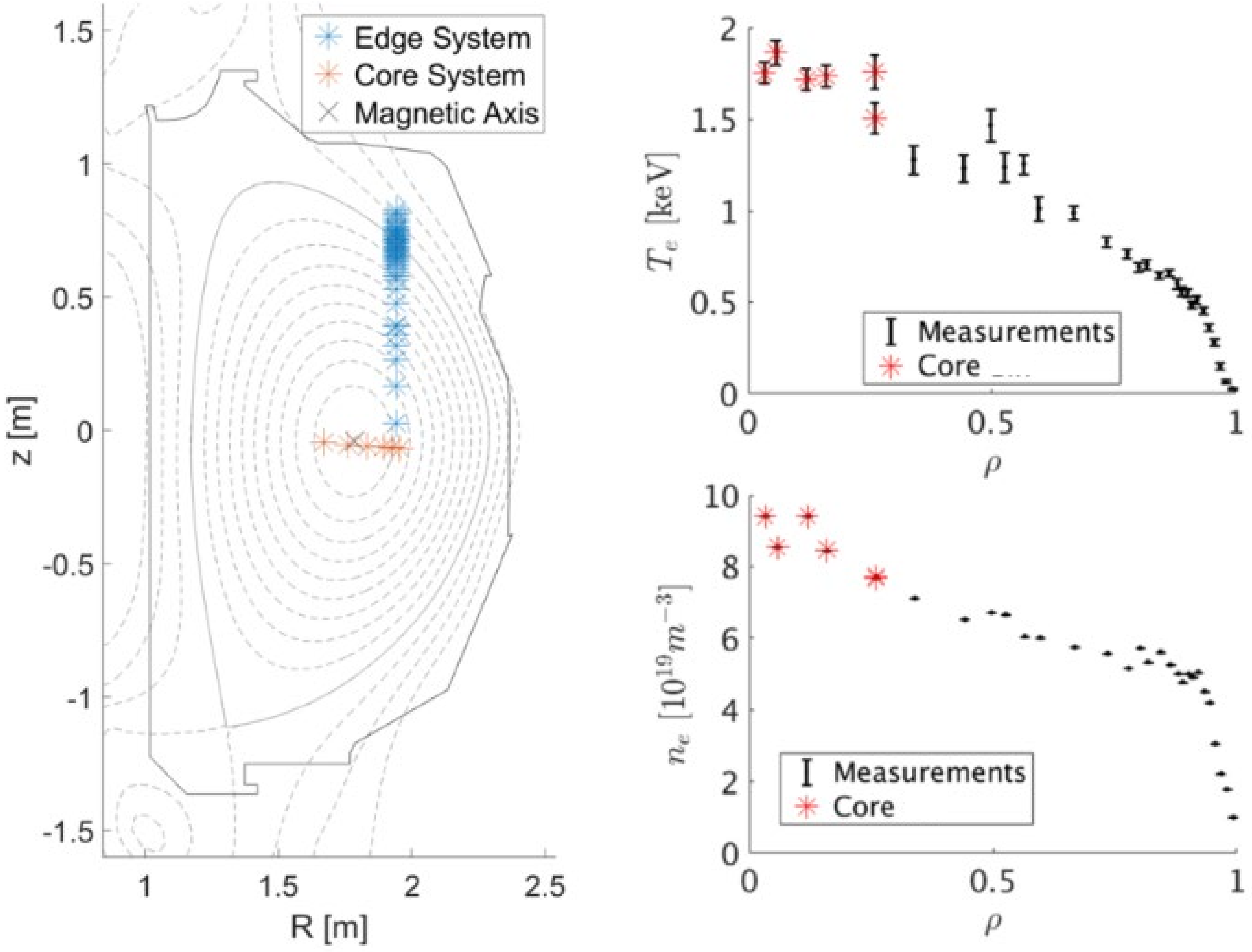
A. Pau et al., *IEEE TPS*, 46 (2018)
C. Rea, K.J. Montes, A. Pau, R.S. Granetz, O. Sauter, "Progress Towards Interpretable Machine Learning-based Disruption Predictors Across Tokamaks", *Fus. Science Tech.* (2020)



DIII-D NATIONAL FUSION FACILITY

T_e and n_e remapped onto ρ to extract relative importance of the core vs full profile + P_{rad} peakings

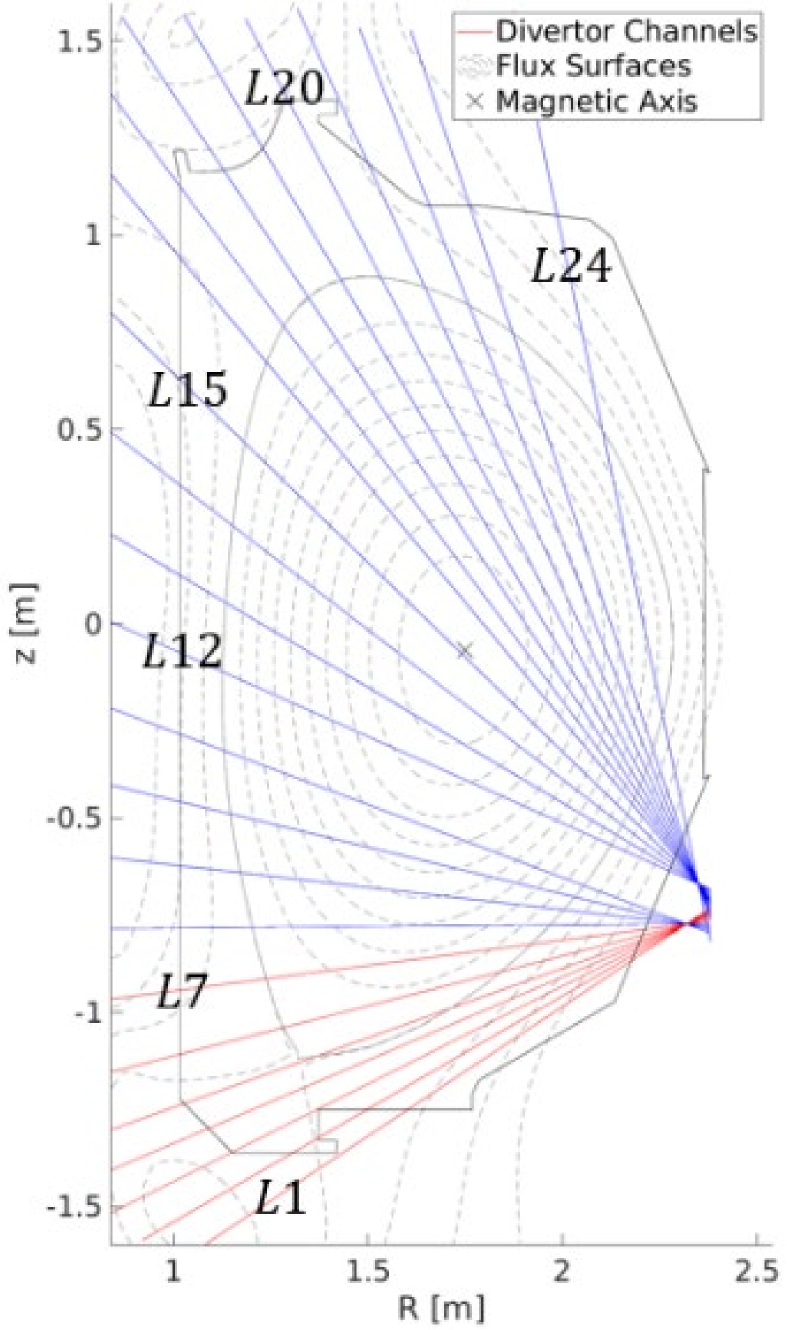
DIII-D Thomson Scattering System



T_e peaking
n_e peaking
P_{rad} Core Peaking
P_{rad} Divertor Peaking

Lower bolometer fan used to isolate P_{rad} contribution in the core or the lower X-point region.

DIII-D Lower Bolometer Fan

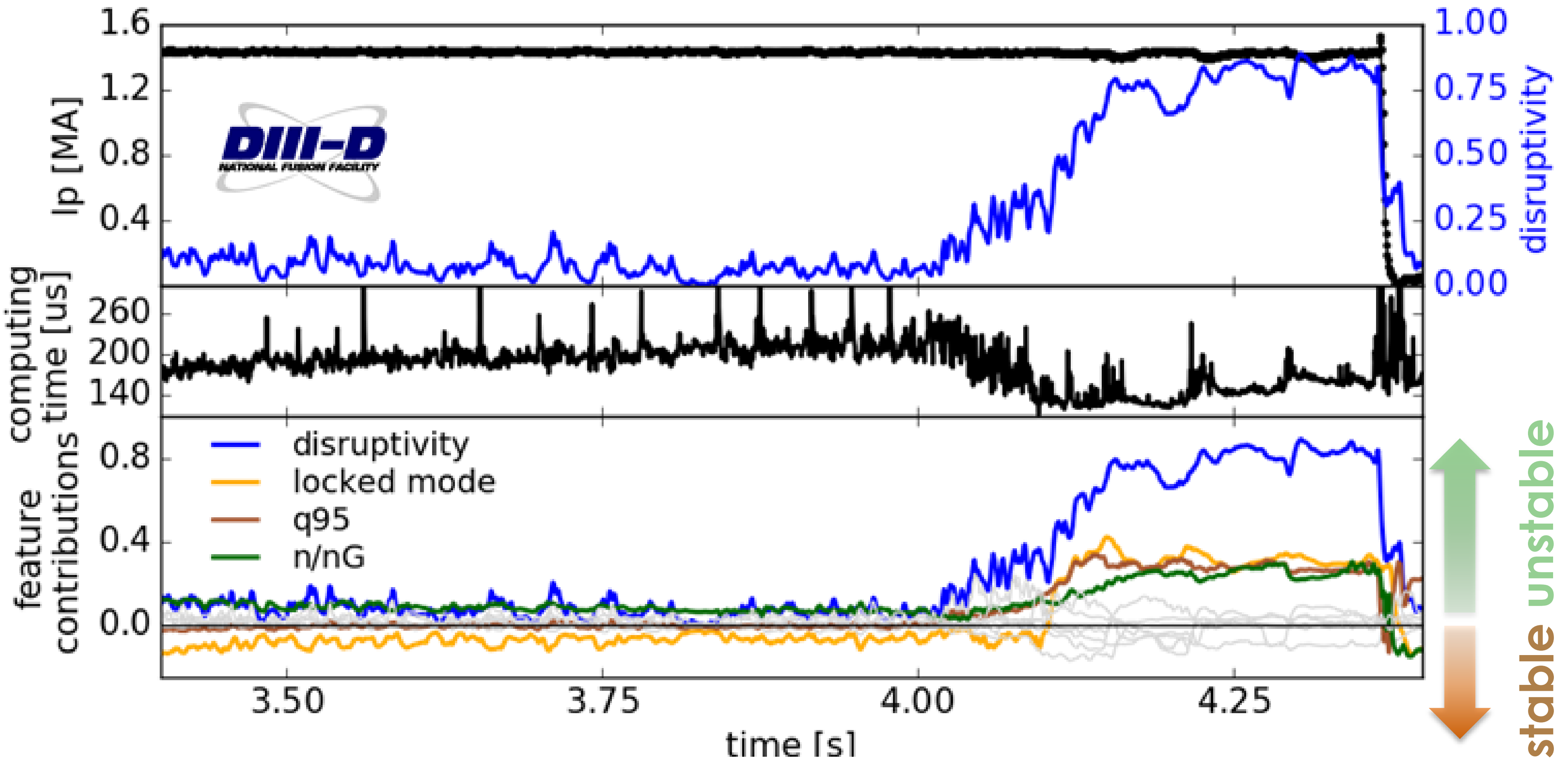


C. Rea, K.J. Montes, A. Pau, R.S. Granetz, O. Sauter, "Progress Towards Interpretable Machine Learning-based Disruption Predictors Across Tokamaks", *Fus. Science Tech.* (2020)

DIII-D NATIONAL FUSION FACILITY

DPRF 2.0: use feature contributions to identify disruptivity drivers in real-time and inform PCS

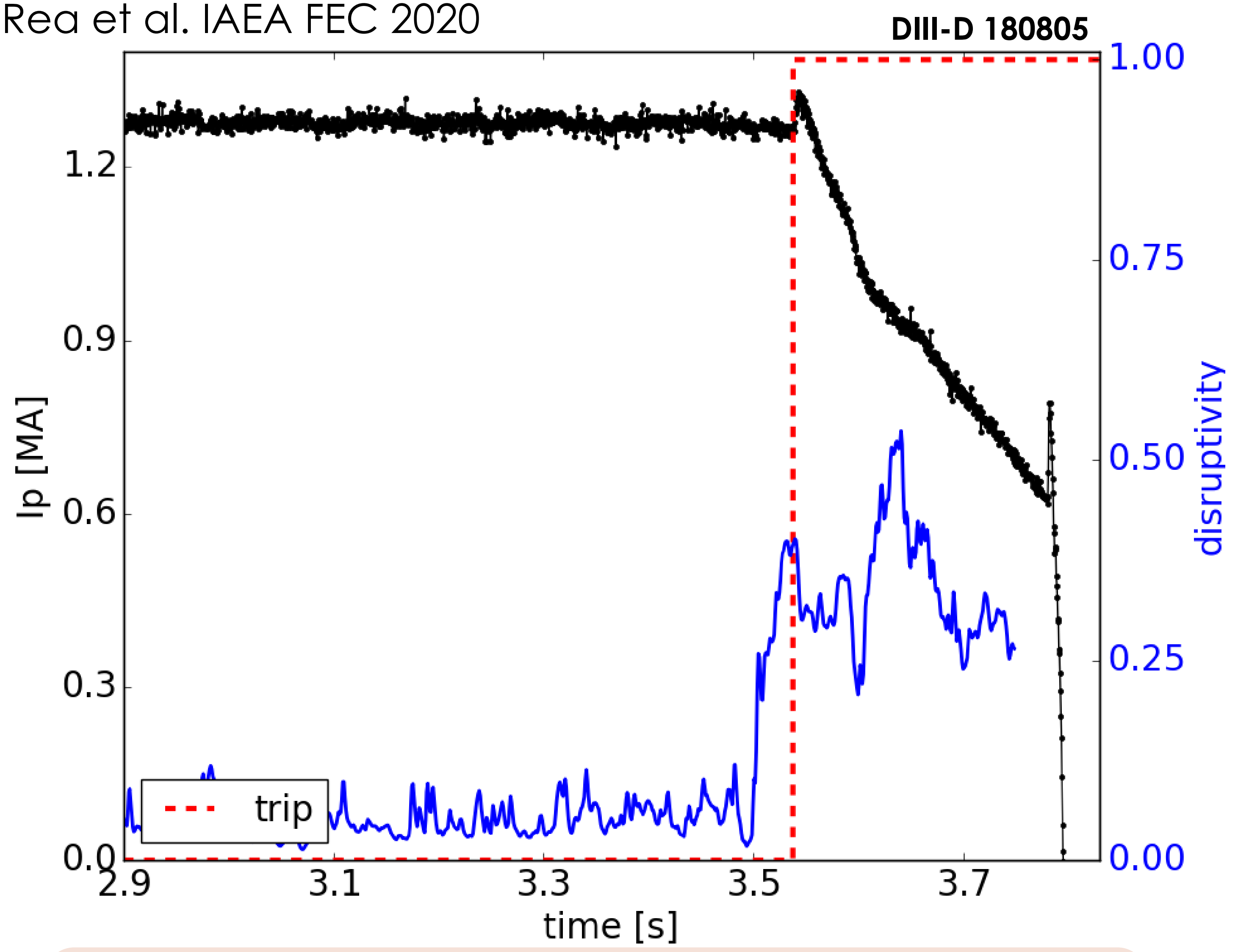
DPRF 2.0
n/n_G
W_{mhd}
β_n
$(I_p - I_{prog})/I_{prog}$
ℓ_i
$B_r^{n=1}/B_\phi$ Locked mode
κ Elongation
δ Triangularity
ζ Squareness
V_{loop}
T_e peaking
n_e peaking
P_{rad} Core Peaking
P_{rad} Divertor Peaking



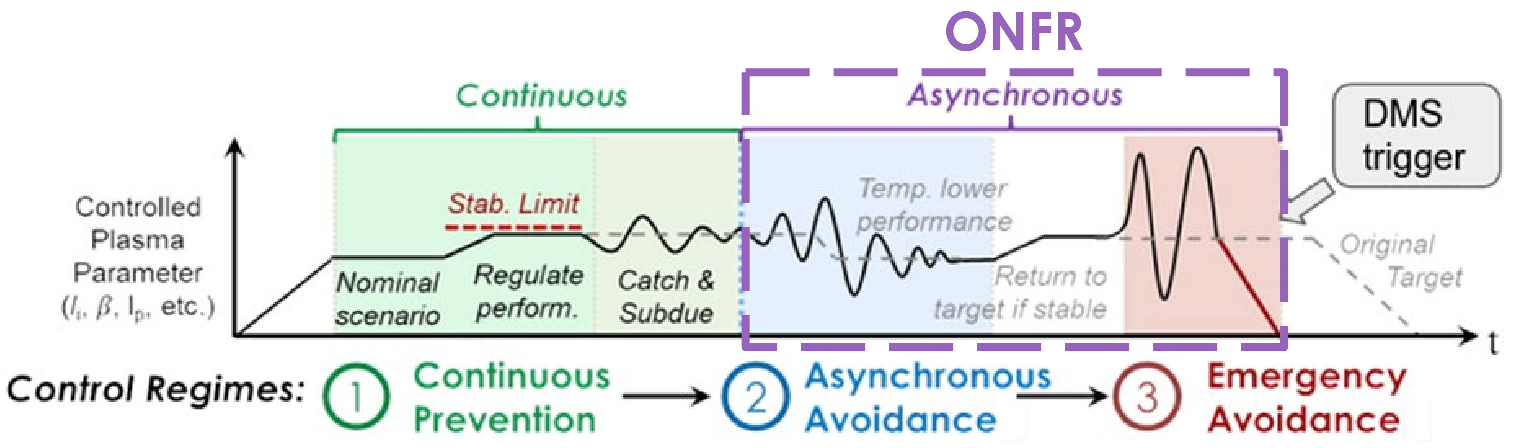
Access to disruptivity drivers in real-time enables monitoring of unstable plasma features

DPRF 2.0 shows real-time feature contribution computation (~ 200 μs) and successful ONFR* integration

C. Rea et al. IAEA FEC 2020



- **Fast shutdown** triggered by preset disruptivity threshold.
- Alarm communicated to **ONFR**, in line with **disruption-free protocol** for **asynchronous control** and **emergency response**.



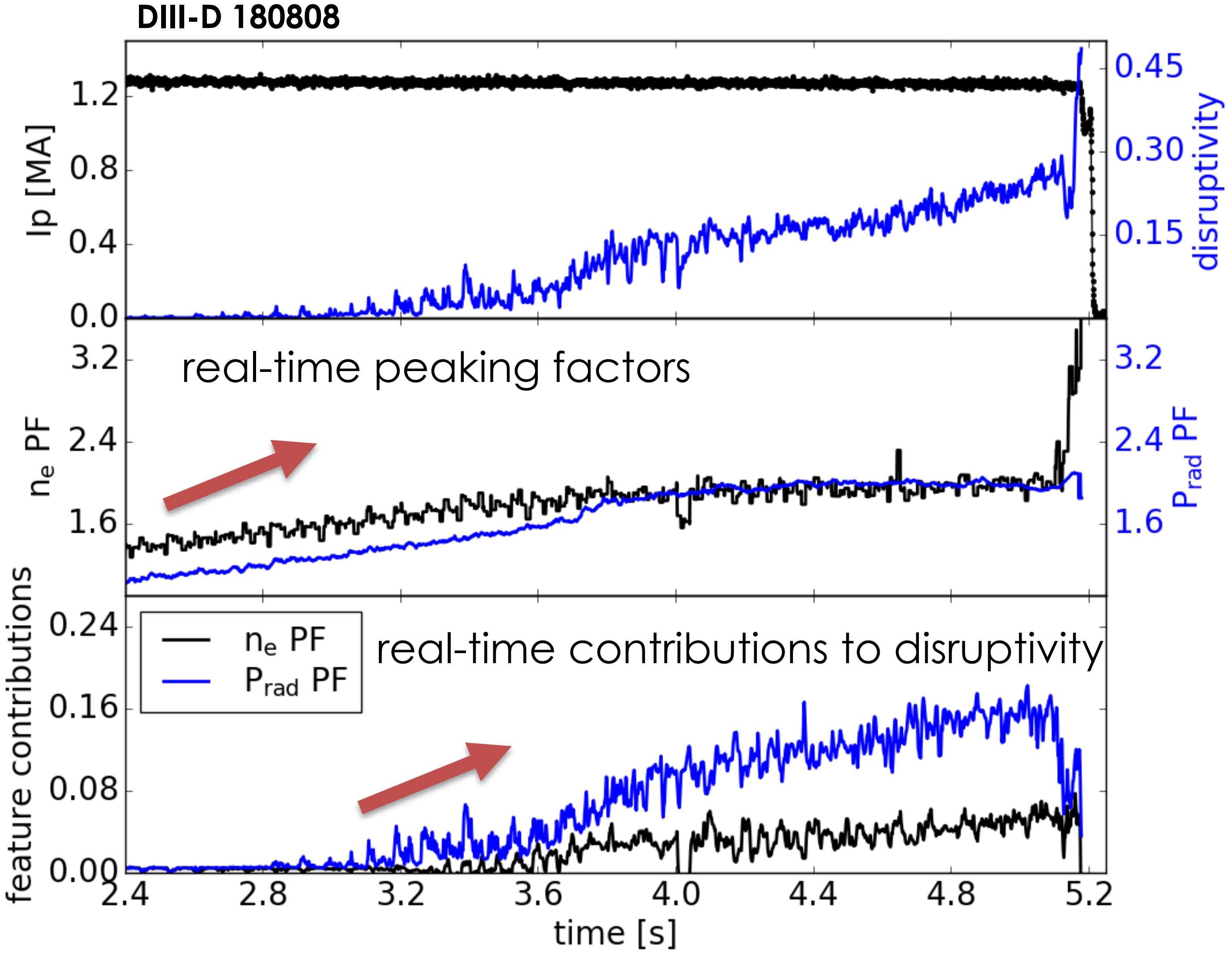
Closed the loop in the PCS by triggering early rapid shutdown, MGI, and ECH

*Off-Normal Fault Response → **Asynchronous and Emergency response**.
N. Eidietis et al., 2018 *Nucl. Fusion* 58 056023

DPRF 2.0 shows real-time feature contribution computation (~ 200 μs) and successful ONFR* integration

C. Rea et al. IAEA FEC 2020

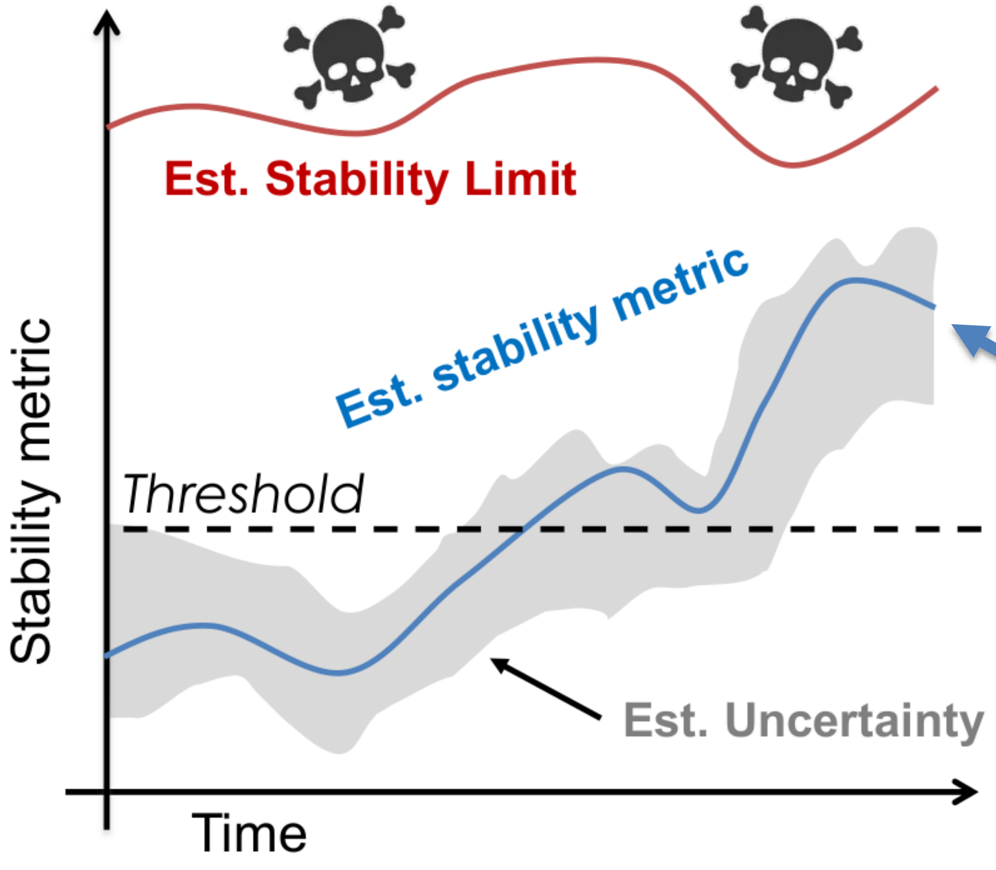
- Flattop disruption with an **impurity accumulation** event: puffing Ar starting at ~ 2s.
- **Peaking factors** reflect changes in profiles due to impurity accumulation, leading to an increase in calculated disruptivity.
- Real-time **feature contributions show stronger signature** of such event.



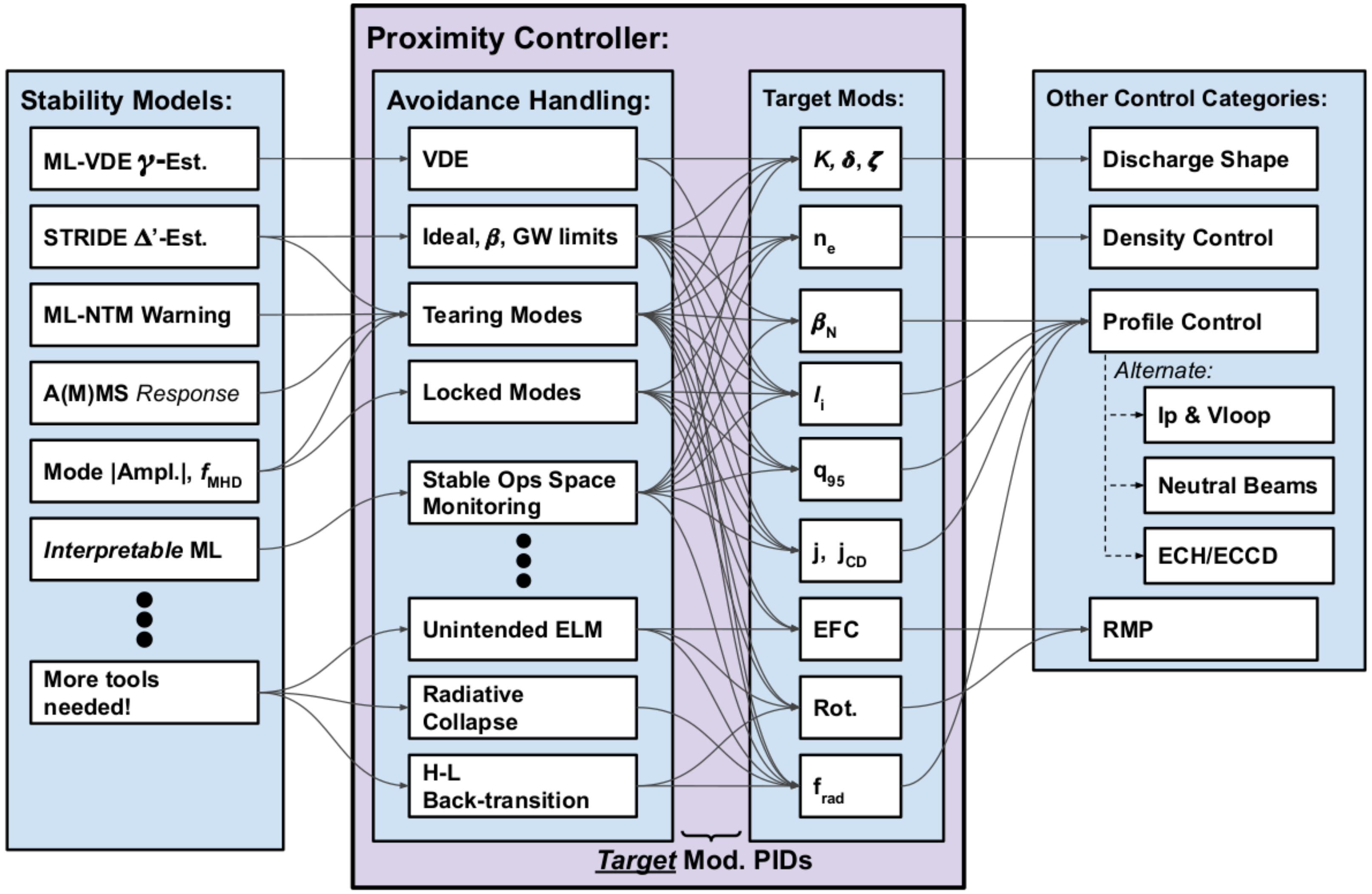
Assessed peaking factors as relevant metrics in ITER baseline scenario on DIII-D

*Off-Normal Fault Response → **Asynchronous and Emergency response**.
N. Eidietis et al., 2018 *Nucl. Fusion* 58 056023

New (FY20) in DIII-D PCS: Proximity Controller, glue code between stability models & actuators regulation



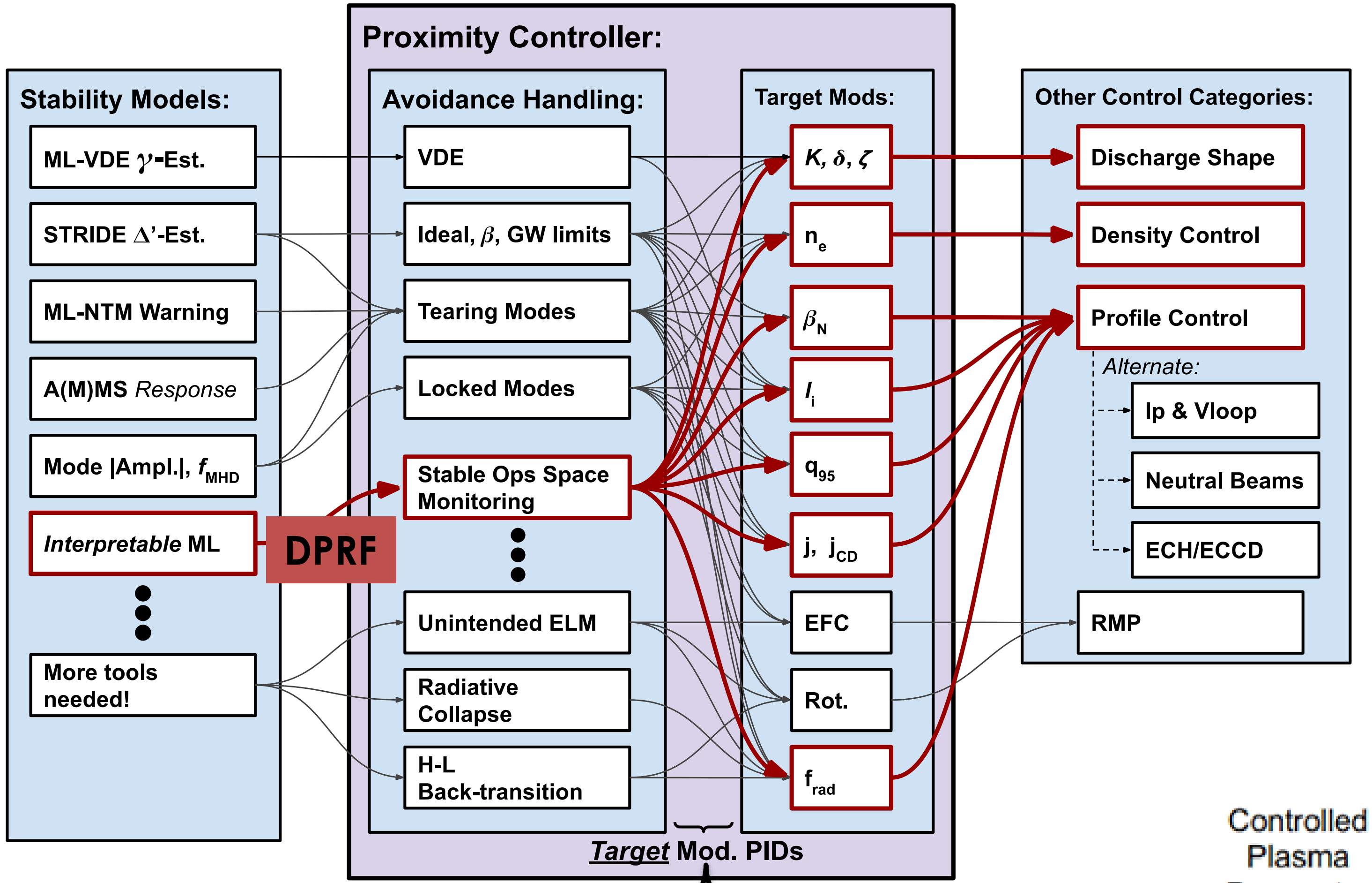
- **Generalized architecture block** connecting multiple input **stability models** to actuators categories for **active regulation**:



- tunable PIDs/matrices mapping stability "errors" to target mods, relative to nearness to stability limit. Ex: (metric-ref)/(lim-ref)

Adapted from J. Barr, 3DSP Group Mtg, 02/22/2021

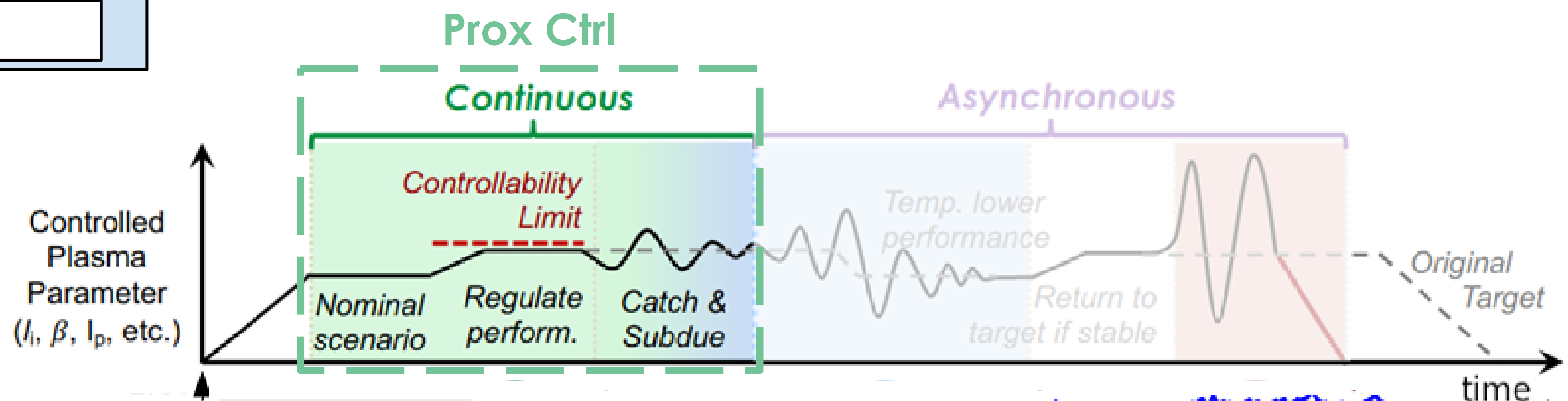
DPRF included in DIII-D proximity controller, being tested right now to regulate plasma stability and performance

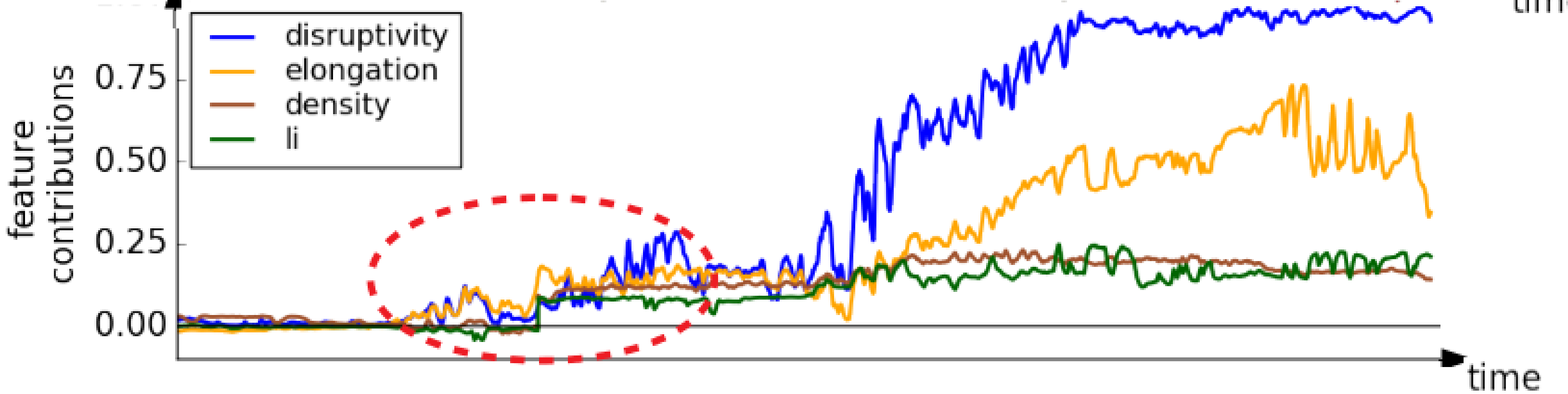


$$\Delta\kappa = PID\left[f_{danger} * f_{\kappa,contrib} * sign\left(\frac{d\kappa}{dt}\right)\frac{\Delta\kappa_{target}}{\Delta f_{\kappa,contrib}}\right]$$

Disruptivity is a measure of proximity to an unstable operating space

Feature contributions can be mapped onto controllable plasma parameters to regulate stability





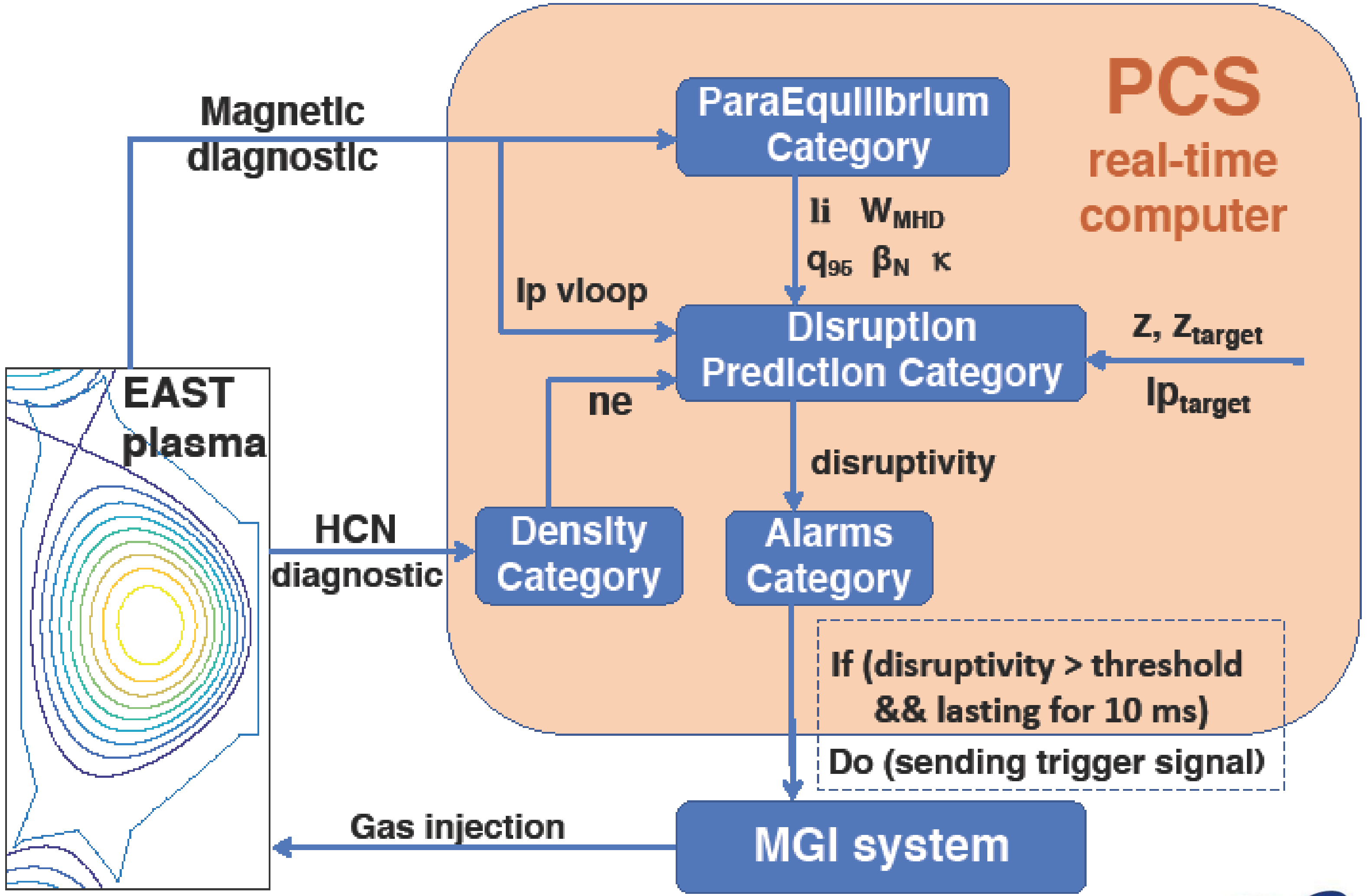
J. Barr, "Control Solutions Supporting Disruption Free Operation on DIII-D and EAST", IAEA TM PDM July 2020
C. Rea, APS-DPP 2020

DIII-D and EAST PCS similarities enable portability of DPRF as general disruption alarm

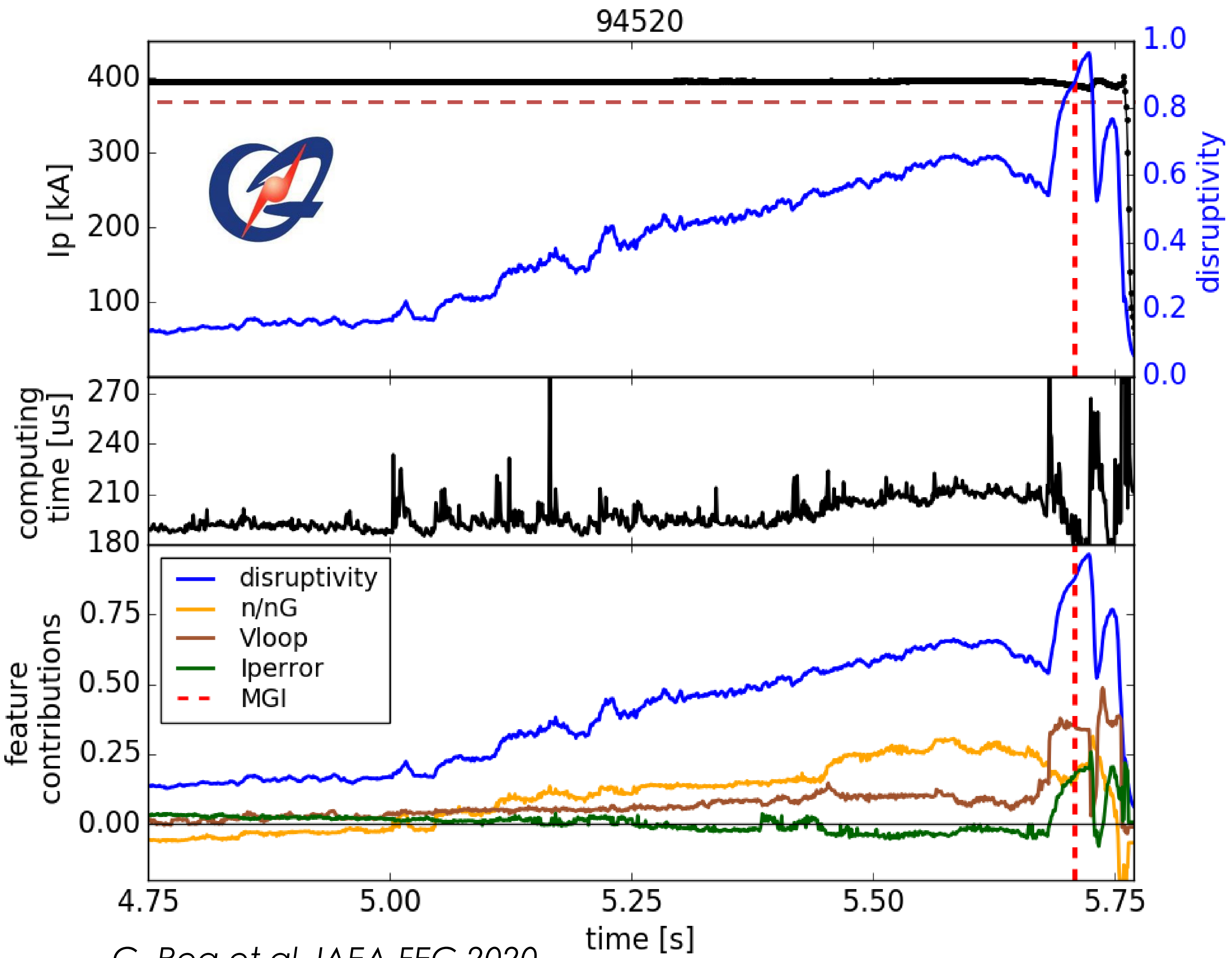
- DPRF version ported in EAST PCS during 2019-2020, gathered stats on performances during 2020 campaign.
- Few dedicated discharges to **test DPRF as MGI trigger**.

0D input features

EAST
n/n_G
V_{loop}
$(I_p - I_{prog})/I_{prog}$
ℓ_i
q_{95}
W_{mhd}
$(z_{cur} - z_{prog})/a$
β_n
κ



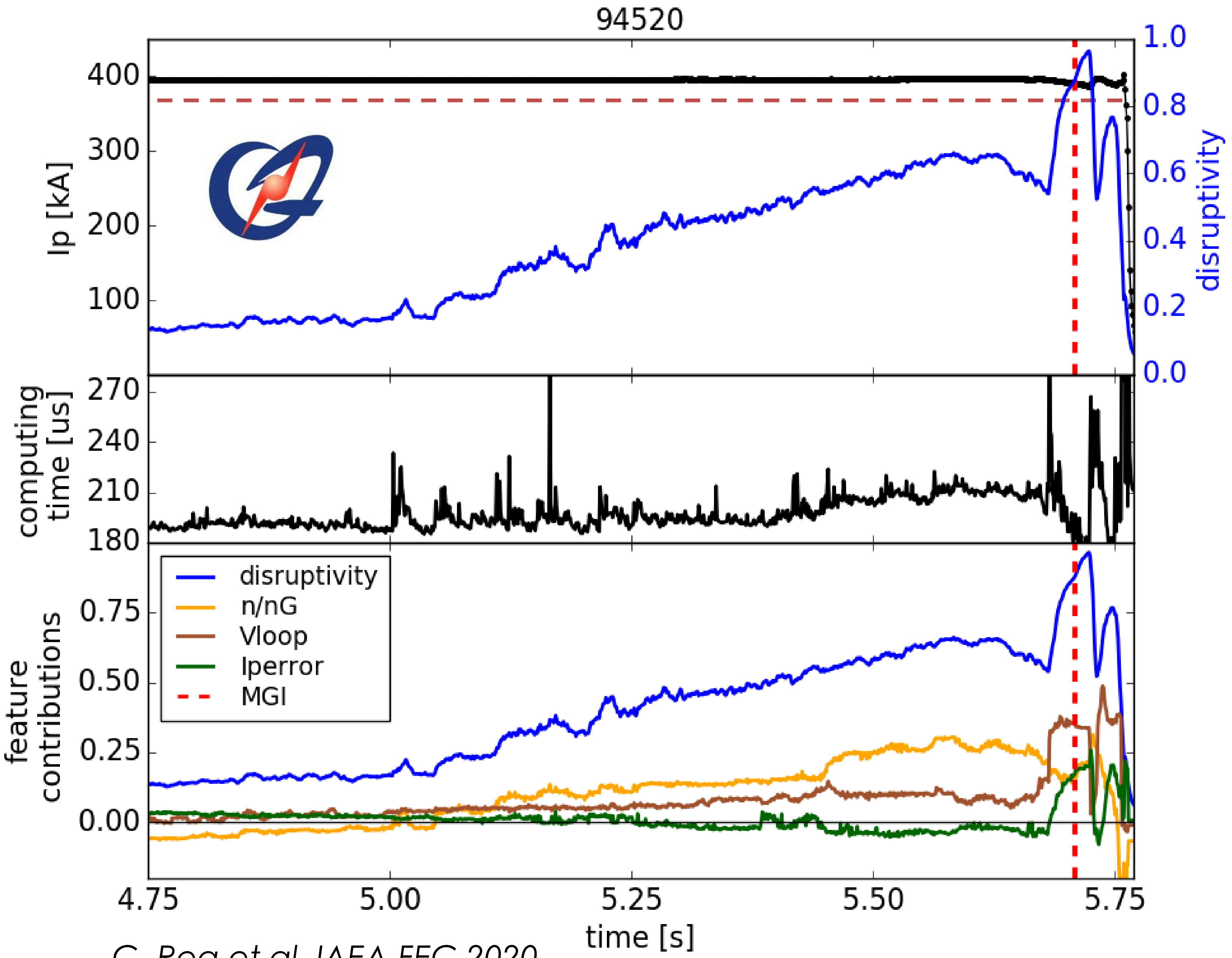
DPRF installed in EAST PCS: feature contributions and disruptivity calculated in real-time in ~ 200 µs



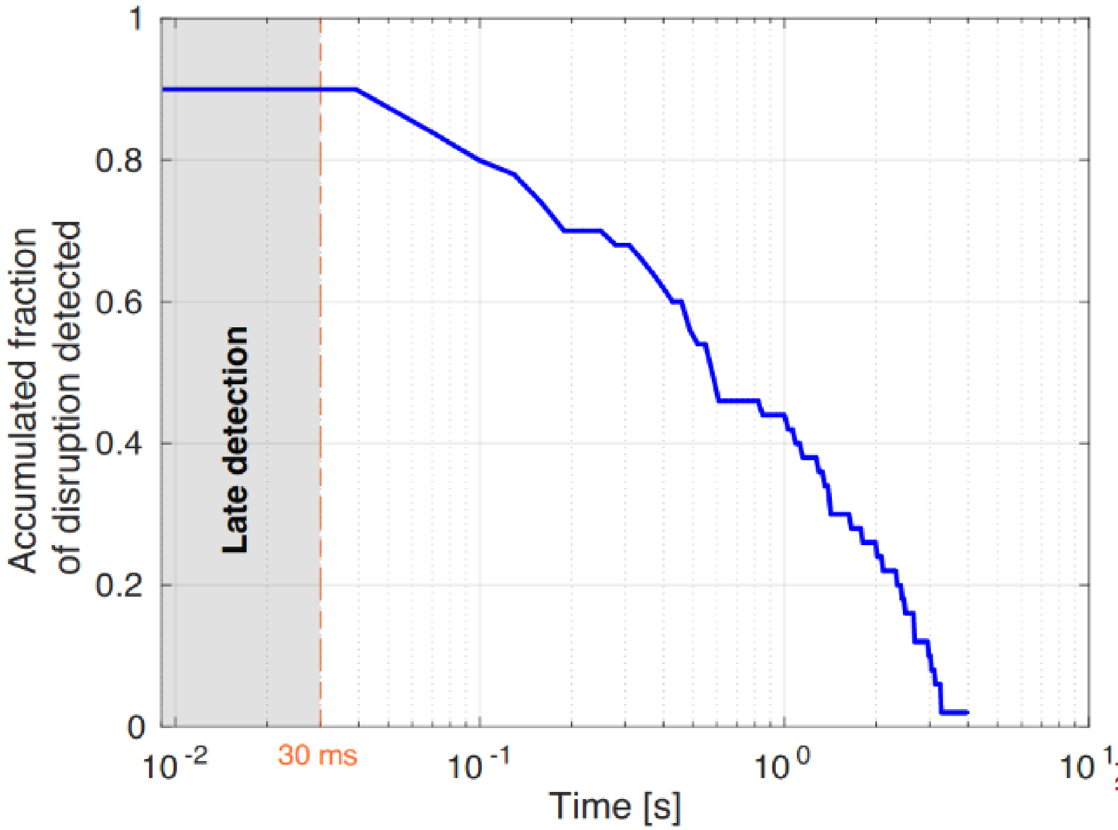
C. Rea et al. IAEA FEC 2020
W. Hu, C. Rea et al., Nucl Fusion 2021 in review

- DPRF trained using ~400 high density (n_e/n_G > 0.8) disruptions and ~400 non-disruptive data.
- Tested in real-time on shots with similar conditions.
- Tested in **closed-loop to fire mitigation system.**

EAST DPRF: disruptivity threshold of 0.8 guarantees TP ~ 92% and FP ~ 10% and avg warning time >1 s



- Performance plateau 40-50 ms before the disruption, guaranteeing > 90% of correct classifications, while keeping the false alarm rate at values less than 10%.

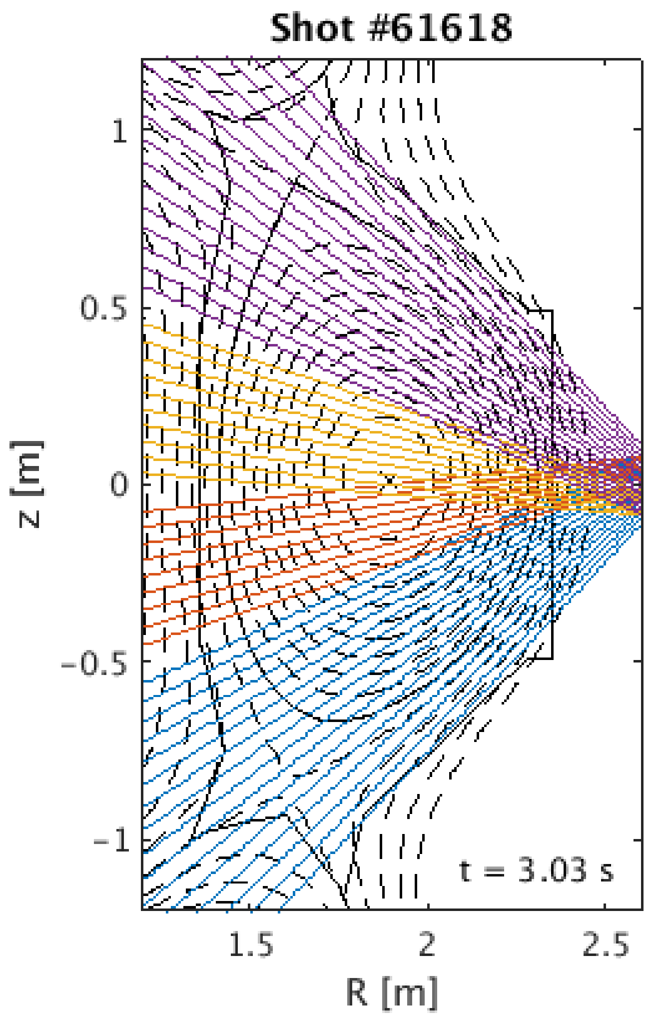


C. Rea et al. IAEA FEC 2020

W. Hu, C. Rea et al., Nucl Fusion 2021 in review

Future work, EAST DPRF upgrades:

- **Shot-by-shot transition time** to unstable operational space;
- Implementation of **radiation profiles peaking factors**, also in **real-time**;
- DPRF tailored on **impurity-driven** (W) **disruptions** in 2021 experiments;
- GA **proximity controller** ported to EAST will enable DPRF as stability model for **continuous prevention.**



AXUV array fans on EAST

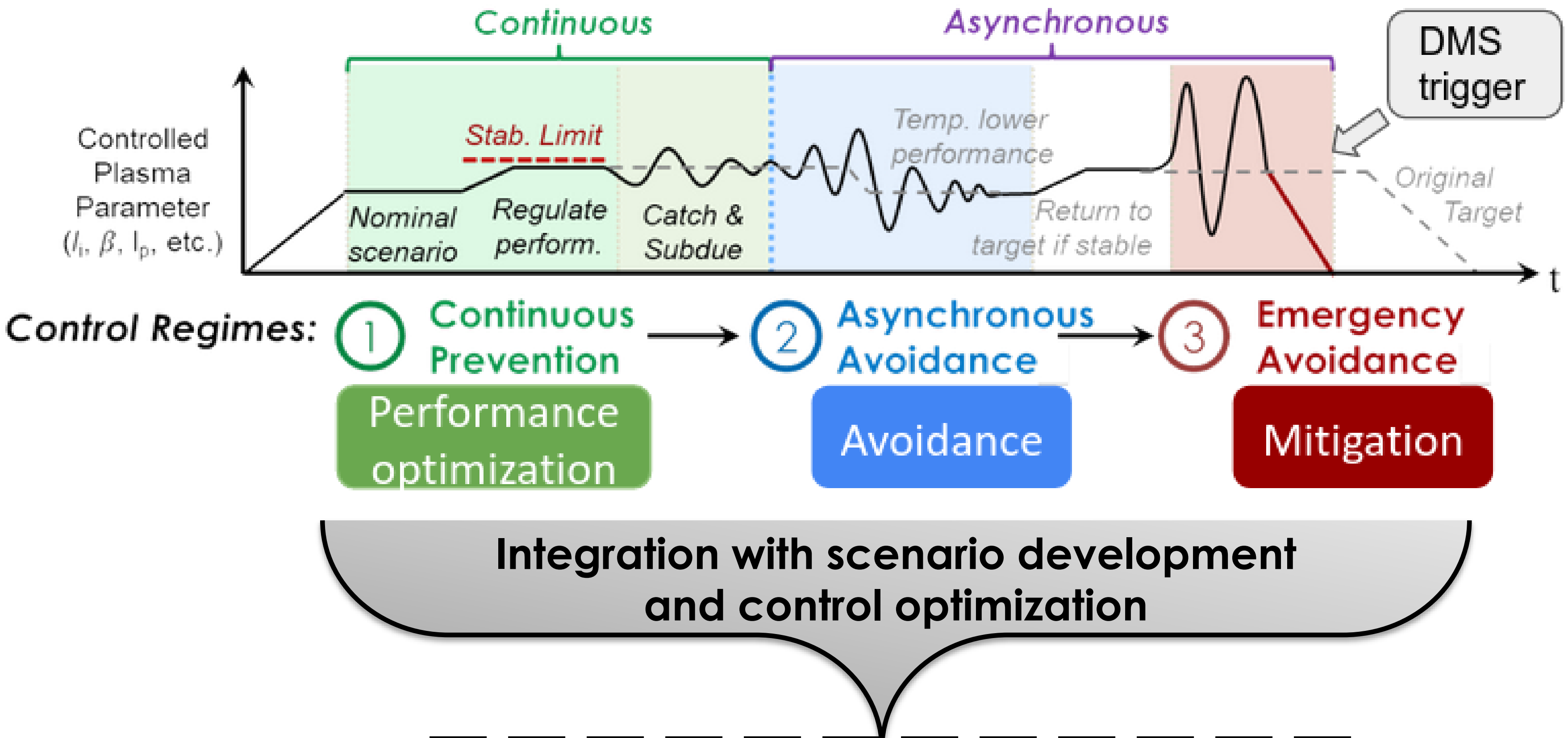
Interpretable ML + control algorithms can be used to regulate the plasma away from stability limits

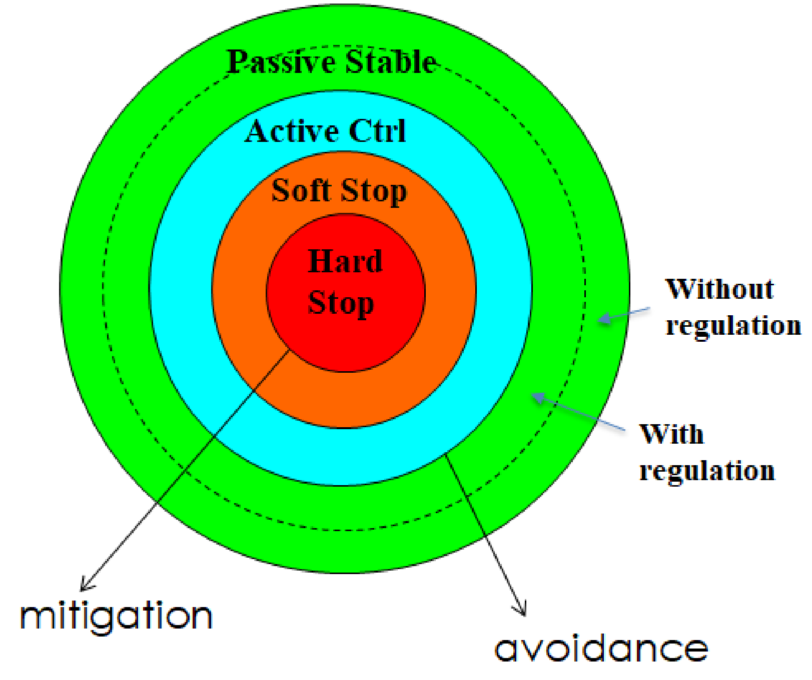
- **DPRF** provides **explainable predictions** – tested on **C-Mod, EAST, DIII-D**:
  - Works as **real-time scenario detector** (DIII-D, EAST).
    - Tested for **asynchronous avoidance** and **emergency response**.
  - Now integrated with **Proximity Controller** for **continuous monitoring** and **stability regulation** (DIII-D).
    - Ongoing real-time control tests.

➢ ***IAEA TM on Plasma Disruptions and their Mitigation*** *material:* *https://conferences.iaea.org/event/217/overview*

➢ *Part of the data analysis was performed using the OMFIT integrated modelling framework.*

Future reactors must operate between passively stable and actively controlled prevention regimes





Adapted from D. Humphreys, USBPO seminar, October 2018

Event and plasma state predictors needed:

- Continuous monitoring;
- Exception handling;

- ❑ Event detection [Montes 2021]
- ❑ Continuous stability monitoring [Rea 2020, 2021]
- ❑ Survival analysis [Tinguely 2019]
- ❑ Binary classification [Zhu 2021] for DMS trigger
- ❑ Time-to-disruption predictions

Additional/Backup slides

Useful references:

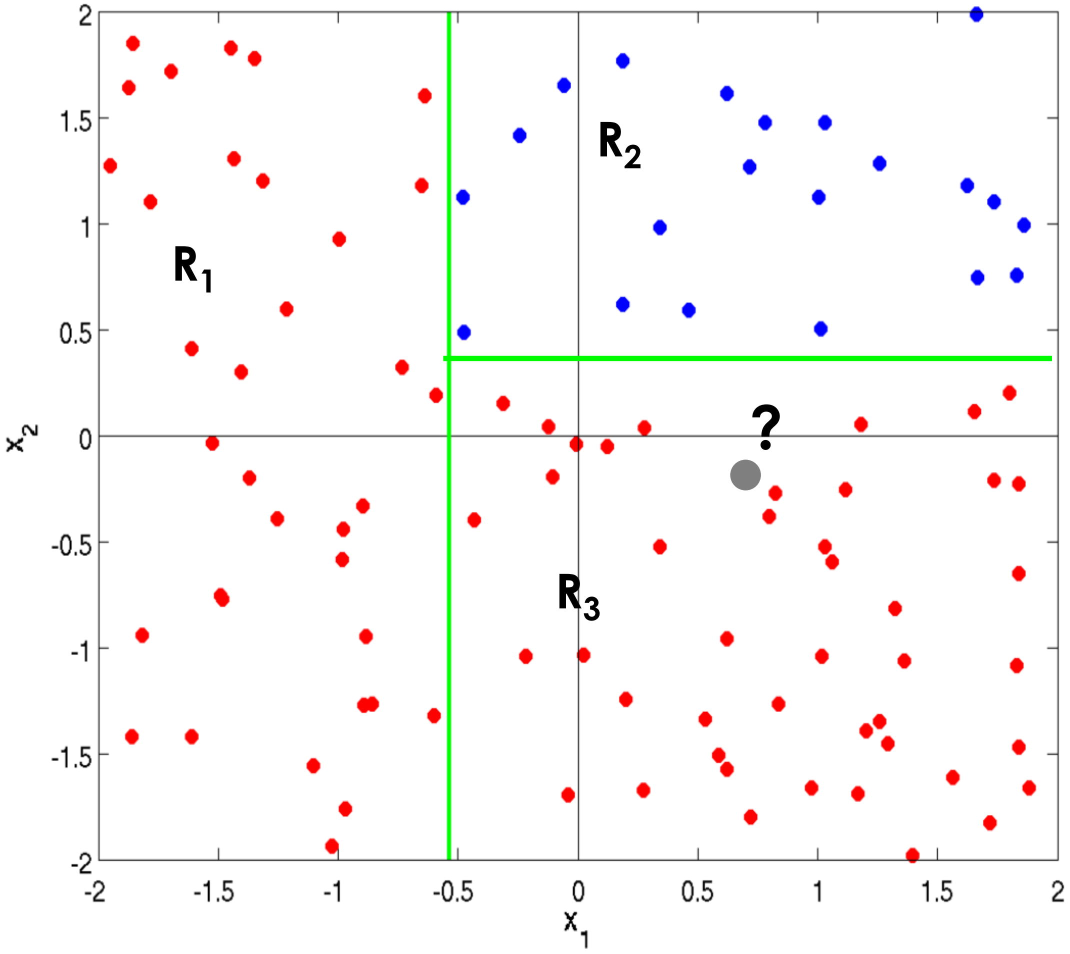
[Barr 2021] J. Barr et al 2021, 28th IAEA FEC, EX/5-TH/6
[Montes 2021] K.J. Montes et al 2021 Nucl. Fusion 61 026022
[Rea 2020] C. Rea et al 2020 Fusion Sci. Technol. 76 912–24
[Rea 2021] C. Rea et al 2021, 28th IAEA FEC, EX/P1-25
[Tinguely 2019] R A Tinguely et al 2019 Plasma Phys. Control. Fusion 61 095009
[Zhu 2021] J.X. Zhu et al 2021 Nucl. Fusion 61 026007

Acknowledgments

Acknowledgement: This material is based upon work supported by the U.S. Department of Energy, Office of Science, and Office of Fusion Energy Sciences, using the DIII-D National Fusion Facility, a DOE Office of Science user facility, under Award(s) DE-FC02-99ER54512, DE-SC0014264, DE-SC0010720, DE-SC0010492, and DE-FC02-04ER54698. Additionally, this work is supported by the National MCF Energy R&D Program of China, Grant No. 2018YFE0302100.

Random Forests* are large collections of randomized and de-correlated decision trees, i.e. CART models

- **CART** (Classification and Regression Trees) algorithms repeatedly partition the input space, to build trees whose end nodes are as pure as possible.
- 2D classification example: 2 features (x_1, x_2) and 2 classes (red, blue).



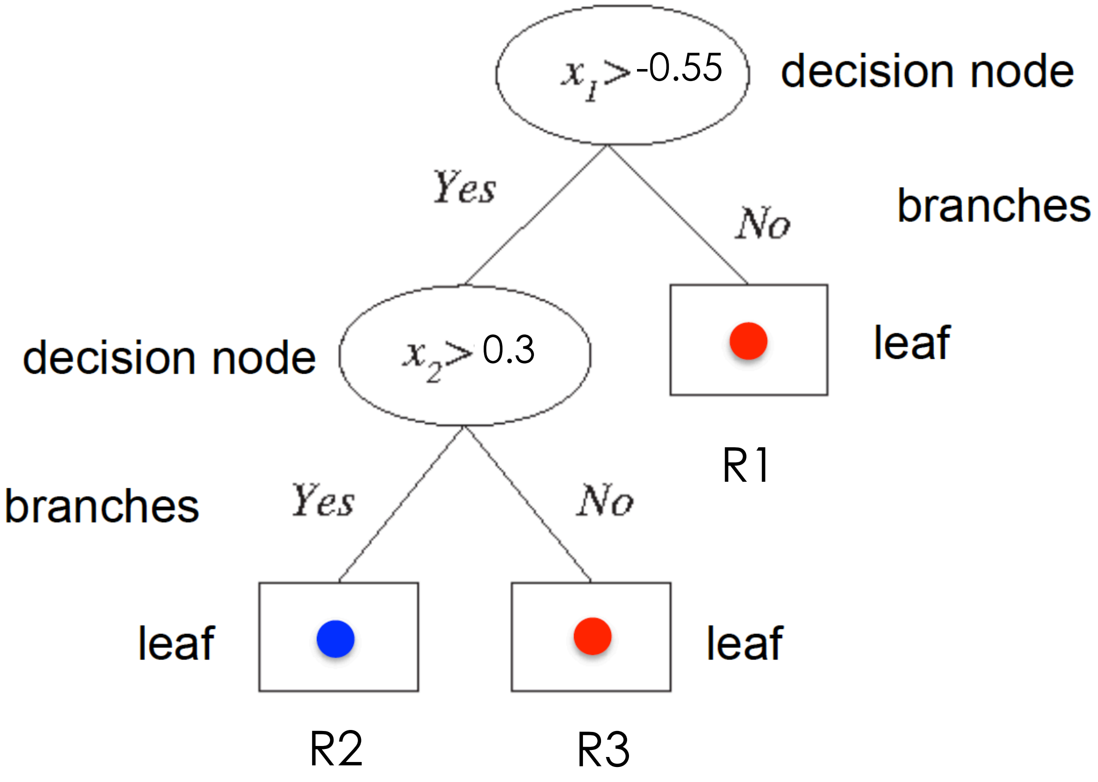
- The algorithm selects the best splitting value to partition the dataset, by **minimizing** an **impurity** measure:

$$\mathcal{I}'_m = -\sum_{j=1}^{n} \frac{N_{mj}}{N_m} \sum_{i=1}^{K} p^i_{mj} \log_2 p^i_{mj}$$

E. Alpaydin, "Introduction to Machine Learning", 2nd edition, MIT Press

- Tree learning via **information gain maximization.**

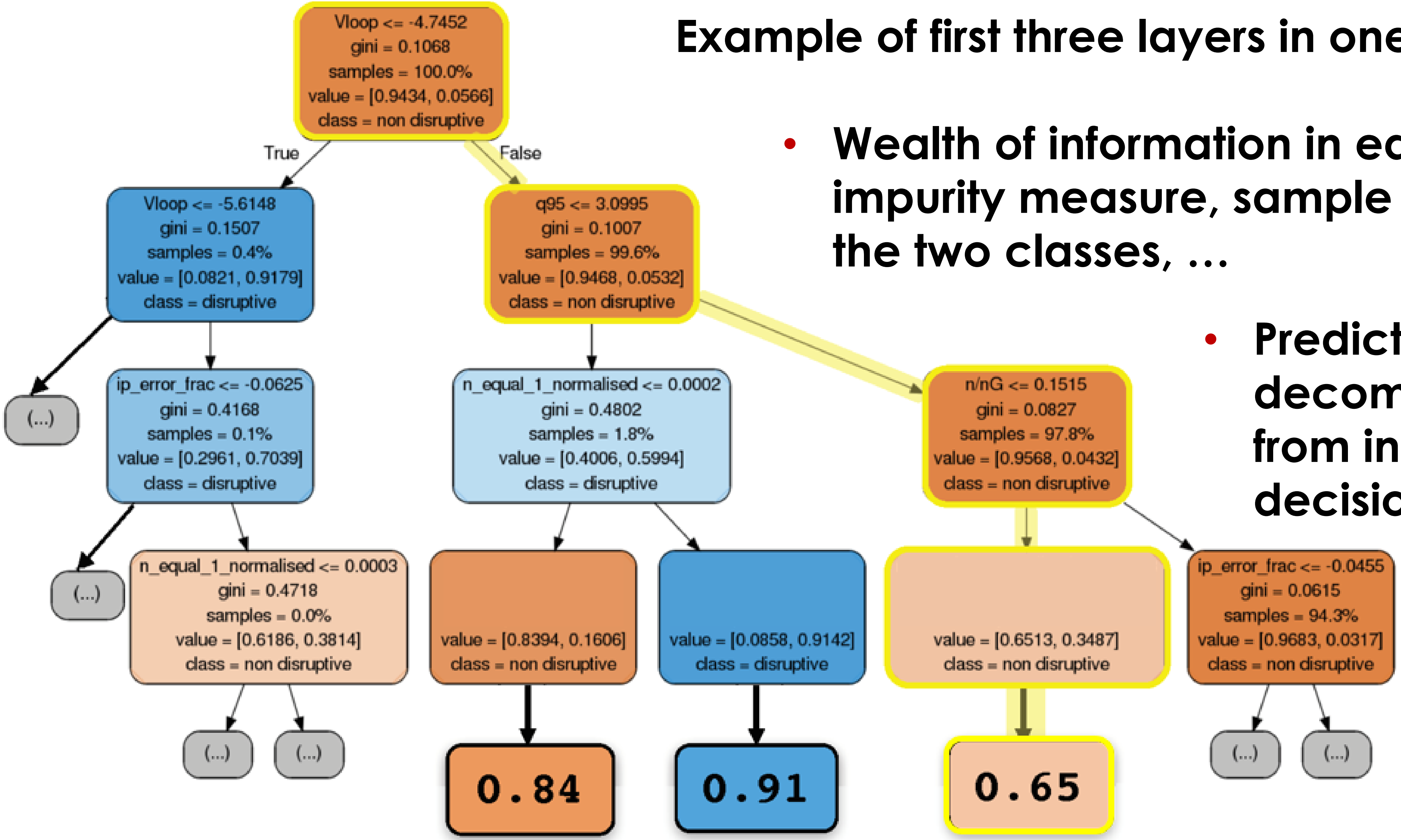
This set of rules, i.e. **collection of decision paths,** is used to classify a new, unseen (test) sample



*L. Breimann, Machine Learning 45, 5–32 (2001)

Decision paths in (DP)RF trees provide wealth of accessible information

https://github.com/andosa/treeinterpreter, A. Saabas,
A. Palczewska et al., Integration of Reusable Systems (2014).

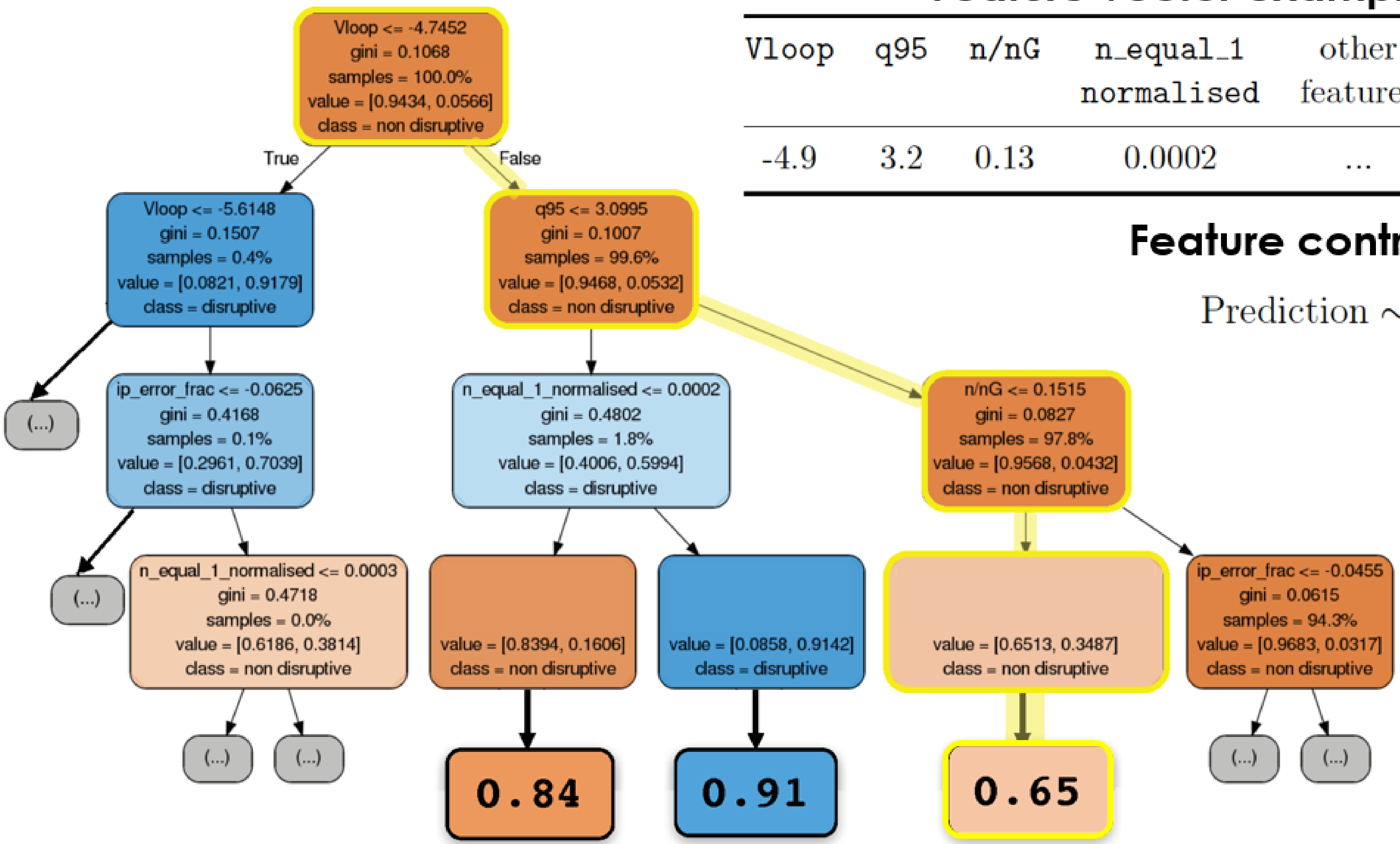


Example of first three layers in one trained RF tree:

- **Wealth of information in each node, e.g. impurity measure, sample distributions in the two classes, …**
- **Predictions on new samples decomposed in contributions from individual features in decision path.**

Decision paths in (DP)RF trees provide local measures of explainability through information gain and loss

https://github.com/andosa/treeinterpreter, A. Saabas,
A. Palczewska et al., *Integration of Reusable Systems (2014)*.



Feature vector example

Vloop	q95	n/nG	n_equal_1 normalised	other features	prediction
-4.9	3.2	0.13	0.0002	...	~ 0.65

Feature contribution breakdown

$$
\begin{aligned}
\text{Prediction} \sim 0.651 \sim\; & 0.943 \quad \text{(bias term)} \\
& + 0.003 \quad (\text{gain, } \texttt{Vloop}) \\
& + 0.010 \quad (\text{gain, } \texttt{q95}) \\
& - 0.305 \quad (\text{loss, } \texttt{n/nG})
\end{aligned}
$$

contributions, or information gain (+) and loss (-)

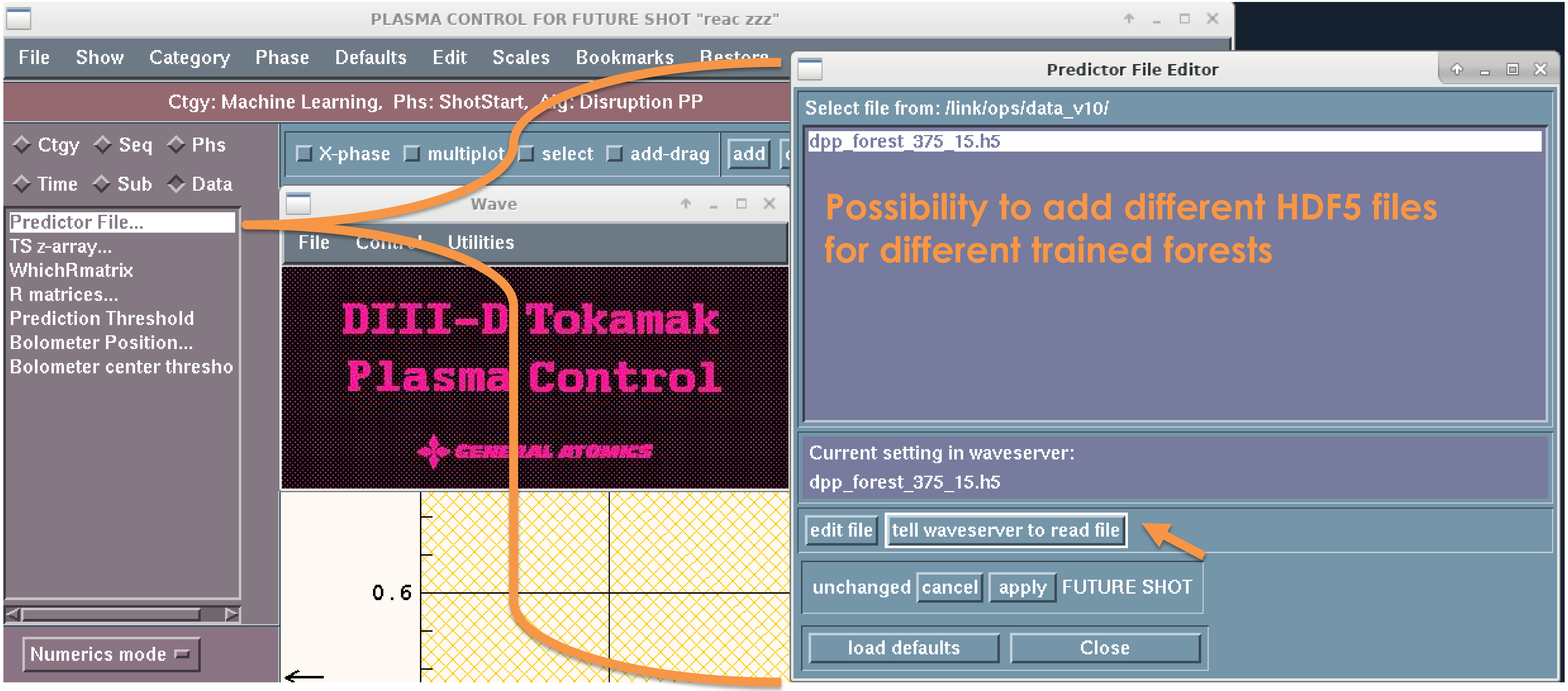
Predictions for forest of *M* trees can be **decomposed** in the *K* **contributions** from each evaluated input feature:

$$F(x) = \frac{1}{M}\sum_{m=1}^{M} \text{bias}_m + \sum_{k=1}^{K}\left(\frac{1}{M}\sum_{m=1}^{M} \text{contrib}_m(x,k)\right)$$

DPRF 2.0 – additional technical changes to the real-time implementation

- **Decision tree collection translates into huge "if-then" PCS external function: slows down PCS compiling;**
- **Remapped DPRF trained structure to hdf5 file:**
  - Model data can be loaded in the PCS (even different data for different phases) at runtime;
  - New general hdf5 interface developed – can be adapted for other data-driven algorithms.
- **Speeds up rebuilding/compiling time of the PCS and allows flexibility on retraining the algorithms between rundays/experiments.**

DPRF 2.0 – additional technical changes to the real-time implementation



C. Rea | 28th IAEA FEC | May 2021