UKAEA-CCFE-PR(20)67

S.A. Sabbagh, J.W. Berkery, Y.S. Park, J.H. Ahn, Y. Jiang, J.D. Riquezes, J.G. Bak, R.E. Bell, M.D. Boyer, L. Delgado-Aparicio, A.H. Glasser, J. Hollocombe, J. Kim, A. Kirk, J. Ko, W. Ko, L. Kogan, B.P. LeBlanc, J.H. Lee, C.E. Myers, L. Terzolo, A. Thornton, S.W. Yoon

Disruption Event Characterization and Forecasting in Tokamaks

Disruption Event Characterization and Forecasting in Tokamaks

S.A. Sabbagh, J.W. Berkery, Y.S. Park, J.H. Ahn, Y. Jiang, J.D. Riquezes, J.G. Bak, R.E. Bell, M.D. Boyer, L. Delgado-Aparicio, A.H. Glasser, J. Hollocombe, J. Kim, A. Kirk, J. Ko, W. Ko, L. Kogan, B.P. LeBlanc, J.H. Lee, C.E. Myers, L. Terzolo, A. Thornton, S.W. Yoon

Disruption Event Characterization and Forecasting in Tokamaks

S.A. Sabbagh[1], J.W. Berkery[1], Y.S. Park[1], J.H. Ahn[1], Y. Jiang[1], J.D. Riquezes[1], J.G. Bak[2], R.E. Bell[3], M.D. Boyer[3], L. Delgado-Aparicio[3], A.H. Glasser[4], J. Hollocombe[5], J. Kim[2], A. Kirk[5], J. Ko[2], W. Ko[2], L. Kogan[5], B.P. LeBlanc[3], J.H. Lee[2], C.E. Myers[6], L. Terzolo[2], A. Thornton[5], S.W. Yoon[2]

[1]Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, USA

[2]National Fusion Research Institute, Daejeon, Republic of Korea

[3]Princeton Plasma Physics Laboratory, Princeton University, Princeton, NJ, USA

[4]Fusion Theory and Computation, Inc., Kingston, WA, USA

[5]Culham Centre for Fusion Energy, UKAEA, Abingdon, UK

[6]Sandia National Laboratories, Albuquerque, NM, USA

Abstract. Disruption prediction and avoidance is a critical need for next-step tokamaks such as ITER. The Disruption Event Characterization and Forecasting Code (DECAF) is used to fully automate analysis of tokamak data to determine chains of events that lead to disruptions and to forecast their evolution allowing sufficient time for mitigation or full avoidance. Disruption event chains related to local rotating or global MHD modes and vertical instability are examined with warnings issued for many off-normal events including density limits, plasma dynamics, confinement transitions, and profile variations. Along with Greenwald density limit evaluation, a local radiative island power balance theory is evaluated and compared to the observation of island growth. Automated decomposition and analysis of rotating tearing modes produce physical event chains leading to disruptions. A total MHD state warning model comprised of 15 separate criteria produces a disruption forecast about 180 ms before a standard locked mode detector warning. Single DECAF event analyses have begun on KSTAR, MAST, and NSTX/-U databases with thousands of shot seconds of device operation using from 0.5 - 1 million tested sample times per device. An initial multi-device database comparison illustrates a highly important result that plasma disruptivity does not need to increase as β_N increases. Global MHD instabilities such as resistive wall modes (RWM) can give the least amount of warning time before disruption. In an NSTX database with unstable modes, the RWM onset, loss of boundary and current control, and disruption event warnings are found in all cases and vertical displacement events are found in 91% of cases. An initial time-dependent reduced physics model of kinetic RWM stabilization created to forecast the disruption chain predicts instability 84% of the time for experimentally unstable cases with relatively low false positive rate. Instances of disruption event chain analysis illustrate dynamics including H-L back transitions for rotating MHD and global RWM triggering events. Disruption warnings are issued with sufficient time before the

disruption (on transport timescales) to potentially allow active profile control for disruption avoidance, active mode control, or mitigation.

1. Introduction

Disruption prediction and avoidance is a critical need for next-step tokamaks such as ITER, since plasma disruptions[1,2] can place significant thermal heat loads and electromagnetic forces on the device and can potentially lead to damage from runaway electrons.[3] Meeting these challenging goals with the high reliability required for ITER and future tokamaks goes beyond active instability control alone and will require multiple approaches, including an understanding of the connection between events leading to disruptions, and the ability to forecast such events well before they occur. Studies often aim to predict the onset of a disruption with sufficient time to successfully trigger disruption mitigation systems. This criterion will mitigate potential damage to the device in question, however, it will also terminate the plasma and is expected to require time to reset the tokamak to a proper operational state to continue operation. With sufficiently early forecasting of a potential disruption, appropriate control systems and actuators could be used to alter the plasma state in a way that would avoid disruption entirely. The Disruption Event Characterization and Forecasting Code (DECAF), under development for this purpose, is used to automate analysis of tokamak data to determine chains of events that lead to disruptions and to forecast their evolution to inform plasma profile and mode control systems aimed to most preferably avoid plasma disruption entirely, or if needed to mitigate the deleterious effects of a disruption. The DECAF approach also provides the understanding required to control the plasma best avoid unfavorable plasma operational states.

The disruption chain "events" as defined in DECAF largely follow the paradigm established by the analysis performed by de Vries, et al.[4] for JET. These studies, in which the results were produced by individual, manual examination of each discharge studied, established a framework for quantifying disruption "events". The temporal combination of these events were considered in that work, which have the analog of disruption event "chains" in DECAF. The DECAF approach aims to automatically determine the relation of the events and quantify their appearance to characterize the most probable and deleterious event chains. It additionally aims to forecast the onset of the events and chains, especially for events that experimentally manifest in close time proximity to the disruption and would elude disruption avoidance control systems, or even disruption mitigation systems. Earlier work by de Vries, et al.[5] using the JET database was also an important element of the evolution of disruption research, especially as it showed fairly low disruptivity in carbon wall operation (dropping to near 5%), but did not address disruption chain events and their interconnection. This, and the studies in Reference 4 contrast the higher disruptivity in JET experienced with the ITER-like

wall operation, illustrating the need for a more physics-based assessment to understand the causes of the disruptions more universally.[6] Other studies have attempted an automatic disruption classification in a statistical manner utilizing generative topographic mapping applied to the JET operation with the ITER-like wall.[7] Initial studies of real-time assessment allowing early reaction to potential disruption causes, including attempts at disruption avoidance have been conducted on machines such as ASDEX-U and TCV.[8]

The DECAF approach differs from these studies. It is primarily physics-based and aims to provide a quantitative and, importantly, a deterministic (rather than a statistical) predictor for disruptions. It also aims to provide an understanding of the dynamics of the events leading to disruptions to best ensure disruption prediction extrapolability to future devices. This is highly important in high fusion power devices such as ITER in which the production of disruptions to teach purely automated model building approaches is highly restricted. Still, the DECAF approach and code are highly flexible and allow a large range of models from simple empirical comparisons, to reduced explicit analytic models based on computationally intensive first-principles physics analysis, machine learning reductions of first-principles physics models, or hybrid machine learning models that use both physics-based and pure machine-based techniques. DECAF events are not simply labels. Instead, they contain both attributes and methods in an object-oriented programming sense. In this way, the collected understanding of these events can be programmed and continually validated against tokamak data to improve their general validity. To best validate the expanding models being added to DECAF, significant effort is being placed on testing the algorithms against *full* tokamak databases on multiple tokamak devices throughout the world. As shown later, this approach is required to avoid errant determination of plasma parameters from databases limited to time periods that are only in close time proximity to the disruption. A larger variety of devices also provides essential depth in testing physics models and determining uniqueness and commonality in the events and their chains leading to disruptions. In the present work, the KSTAR, MAST, and NSTX/NSTX-U databases are examined, with analysis expanding to the DIII-D and TCV databases that are also available. Importantly, DECAF meets all of the disruption predictor requirements outlined by Humphreys, et al.[9] that a predictor must (i) predict specific phenomena, (ii) provide a continuous variable quantifying proximity to disruptive states, and that can trigger actions, (iii) provide sufficient lead time for mitigation or avoidance, (iv) be extrapolable to new devices, and (v) be real-time calculable. To clarify, the goal of the research presented here is the automated characterization of disruptions by event chains and their forecasting using input data and analysis that are not computed in real-time. However, the implementation and development of the present DECAF code has been performed to efficiently allow analogous analysis to be performed in real-time. Future work

is planned to do this, and also to compare the performance of the offline and real-time DECAF analysis.

The following sections examine an important subset of the event analysis in the code and insights gained on the connection of plasma dynamics to the events. Section 2 describes the DECAF approach further, including the concept of disruption warning levels (with connection to past work). Section 3 describes the continuing development of the physical models in DECAF, with examples including density limits and MHD instabilities – including both their automated characterization and present forecasting capabilities. Section 4 describes the initial investigation of large, general databases (from KSTAR, NSTX, and MAST), including full disruption event chains, the importance of analysing long periods of the plasma evolution in the devices, the increasing capability of the code to produce early disruption forecasting, and a brief summary of the disruption prediction model performance evolution. A summary and discussion are included in Section 5.

2. Disruption chain events and warning levels

Fig. 1 simply illustrates the paradigm that DECAF follows in providing automated understanding of the dynamics leading to a tokamak disruption along with an example from experiment. Continuous tokamak plasma operation at high fusion performance is desired (Fig. 1a). However, at some point this "normal" operational plasma state can be altered by various "events" ranging from purely technical considerations (e.g. magnetic field power supply interruption) to more complex reasons such as the onset of plasma instabilities, loss of heating power balance, or loss of torque balance. This alteration is considered as a chain of individual events, starting with a trigger event and evolving toward the plasma disruption (here labelled by the acronym DIS representing the plasma current quench). DECAF analysis of device databases aims to automatically determine not only the frequency of occurrence of events, but also understanding of this chain of events. Future real-time implementations of DECAF diagnostic interpretation and forecasting models of such events can then be used to trigger disruption avoidance systems. This expanded strategy can be contrasted to present disruption avoidance systems, e.g. MHD instability control systems[10] that essentially wait until the "disruption precipice" to address avoidance of the oncoming disruption. Fig. 1b) shows a DECAF analysis of a plasma with "mid-range" normalized beta $\beta_N \equiv 10^8 <\beta_t> aB_0/I_p$ (where $\beta_t \equiv 2\mu_0 <p>/B_0^2 = 4.5$ toroidal beta, p is the plasma pressure, B_0 is the vacuum toroidal magnetic field at the plasma geometric center, and a is the plasma minor radius at the midplane) in the NSTX[11] device. In this case, a global magnetohydrodynamic instability (resistive wall mode, RWM) is identified by DECAF as the event chain trigger, which in the past was thought to be the direct precursor to the disruption (DIS). However, DECAF identifies several interceding events. Next in the chain, 7 ms after the RWM event, is a

vertical displacement event (VDE). The algorithm for determining whether a VDE event has reached a critical warning level is described in the next section. Five milliseconds later, a wall proximity warning (WPC) is issued indicating that the plasma boundary is about to touch the device first wall (and does soon after the warning). A low plasma density warning (LON) is issued 2 ms later, followed 5 ms later by the IPR event warning that the feedback control target plasma current request is no longer being met. By now, the original separatrix-limited plasma is in contact with the wall and is decreasing in size, with the edge plasma safety factor, q, decreasing as plasma poloidal flux is lost. At 9 ms later, a low q warning (LOQ) is issued. Finally the time of the disruption (DIS) is found, based on the plasma current quench.[12] The DIS event occurs over 30ms after the trigger RWM is issued, which is expected to be just enough time in ITER to trigger the disruption mitigation system effectively. However, this relatively short-duration disruption chain would be better handled if the RWM event itself was forecast at an earlier time. DECAF presently has a model to do this, as discussed in Section 3.3.2. At present, DECAF event warning levels are determined by a flexible diagnostic and physics model "point" system similar to that successfully used for NSTX[13] but significantly expanded including results from general analysis of the evolving diagnostic input and plasma equilibria with a continuous warning level determination. A key expansion of the present DECAF approach is that several event criteria can be used in conglomerate to determine combined "levels" that allow DECAF to issue event warnings. For example, at present, 15 separate criteria are used to determine the total MHD warning level for rotating MHD modes (see Section 3.2.1).

3. Physical Model Development

A profound power of the DECAF approach is the ability to test any physical model developed by the fusion research community for practical use as part of a disruption prediction model ensemble. Models that can quantitatively forecast disruptions more accurately across all devices can then objectively be chosen as being more desirable using quantitative figures of merit. Over 50 disruption chain events are presently identified, with over 20 events that have diagnostic evaluation and physical models providing warning levels. Simpler evaluations examine key diagnostics in combination to compute warning levels, with comparison to critical levels to determine when DECAF issues event warnings. For example, the VDE event combines a comparison of axis position ($|Z|$), axis velocity ($|dZ/dt|$), and Z dZ/dt against threshold levels set in the model. Critical levels of such models will differ for each machine. The validation of the technical and physics-based models for each of the 5 devices in the present DECAF database now comprises the primary near-term research effort. More desirable are models that more transparently reproduce the behavior of all tokamak devices. The simplest models in this class are empirical models such as the Greenwald

density limit. A next level includes models that are more closely based on first-principles physics, examples of which are discussed below.

3.1 Density Limits

The Greenwald density limit[14] (event GWL) is included in DECAF as a universal empirical model for disruption forecasting. Recently several theories have been developed to explain the observed global Greenwald limit in tokamaks, including a ballooning stability limit at the separatrix[15] and a local island power balance theory[16,17]. In the latter theory, power balance in an island between input Ohmic heating and radiated power loss results in a maximum local density that scales with local current density. If the density at the island exceeds the limit, or alternatively if the radiated power at the island exceeds the input power ($P_{loss} > P_{input}$), then the island grows and can lead to plasma disruption. The limit can be written either in a form which mimics the global Greenwald density limit in a local form, or one that mimics a radiated power fraction localized to the magnetic island surface. This model has been added to the DECAF code including the radiated power, resistivity, and current density profiles as inputs.

The radiated power profile (P_{loss}) can either be measured directly or can be estimated from density profiles and calculated cooling rates of deuterium and impurities such as carbon, which depend on electron temperature[18]. Fig. 2 shows both the measured and calculated profiles for an NSTX discharge. The P_{loss} is calculated as $P_{loss} = n_e \sum n_Z L_Z$, where the species Z considered in this case are limited to deuterium and carbon and the cooling rates L in Wm^3 are given for deuterium by $L_D = 5.35*10^{-37} T_e^{1/2}$ with electron temperature in keV (Ref. 15), and for carbon by tabulated formulae in Ref. 18. The input power profile which P_{loss} is compared to in Fig. 2 is calculated from $P_{input} = \eta j^2$. The resistivity profile, η, is calculated based on electron temperature and the effective charge Z_{eff} (formula from Ref. 15) which are measured by Thomson scattering and charge exchange recombination spectroscopy. The current density profile used here is the total surface-averaged current density profiles from various sources (Ohmic, bootstrap, beam-driven), which are computed by the TRANSP code.

The power balance model is a local condition for island growth, therefore mode marginal stability would occur when $P_{loss}/P_{input} > 1$ at the location of the island. This defines the DECAF event "island power balance" (IPB) shown in Fig. 2. The rotating MHD mode growth that arises when $P_{loss}/P_{input} = 1$ (Fig. 3) is measured as having toroidal mode number $n = 1$. The lowest order rational surface in the plasma is $q = 2$, so $m/n = 2/1$ activity is the most likely candidate. Therefore, the local power balance criterion is evaluated at the $q = 2$ surface. The $n = 1$ mode onset in Fig. 3 is highly correlated with the IPB event warning in the plasma shown. Also shown is the computed Greenwald fraction evolution and the DECAF event GWL defined as the line-averaged plasma density equal to the Greenwald density. At

the IPB event, the Greenwald fraction is ~ 0.9. While this correlation is positive, the present state of analysis shows the quantitative evaluation of the IPB event to be sensitive to the accuracy of the local input criterion (e.g. position of the $q = 2$ surface). Present DECAF analysis shows that the local island power balance evolution follows the evolution of the global Greenwald fraction. For 13 discharges tested, the Greenwald fraction ranges from 0.75 to 1.05 at the time of MHD onset and the local island power balance fraction has a range of about 0.60 to 1.50. Continued analysis is focussed on reducing this variation and eliminating the need for full TRANSP analysis for each plasma, for example through neural net evaluation of a representative set of TRANSP runs to determine the required input for the IPB event.

3.2 Rotating MHD instabilities

Automated analysis of rotating MHD modes with tearing characteristics has started by using a DECAF module to produce physical event chains leading to disruptions through slowing of the modes by resonant field drag mechanisms and subsequent locking. An algorithm portable across tokamaks devices has been developed that processes the spectral decomposition and signal phase matching of magnetic probe signals for mode discrimination. Multiple modes occurring simultaneously are tracked and bifurcation of the toroidal rotation frequency and locking for each mode due to the loss of torque balance under resonant braking are detected.

3.2.1 Disruption Event Characterization

The information analysed for these modes along with plasma rotation profile and other plasma measurements produces predictive warnings for the individual modes, along with a total MHD event warning signal showing initial success as a disruption forecaster. These capabilities are illustrated in Fig. 4 for the same plasma shown in Fig. 3. In the plasma illustrated, rotating MHD instabilities thought to be non-linearly saturated and slowly evolving resistive modes are found using a generalized phase matching algorithm in DECAF using an array of toroidal magnetic probes typically available in tokamaks. The code discriminates the toroidal mode number of the instabilities and tracks all modes that have greater than a specified amplitude. Modes that approach the disruption are indicated by the chevrons in the diagram (which show the mode n number). DECAF events based on the mode evolution are also shown, including the bifurcation of the modes (BIF-n1,2) (loss of torque balance leading to rapid loss of mode rotation), and events marking the locking of the modes in the laboratory frame of reference (LTM-n1,2). A single "total" MHD warning signal that varies with time is also shown. This warning is created by a set of criteria and can be used as a disruption predictor, as described in the next section.

3.2.2 Forecasting

A significant part of DECAF research is determining the best criteria to create predictive warnings. The warning model shown in Fig. 4 is comprised of 15 separate criteria, also shown in the figure displayed as a heat map. This summary of the criteria used to produce the warning level provides a useful illustration of how the total warning level reaches high values, indicating a disruption onset. A total warning level of 4 indicates close proximity to the disruption for this model. The heat map also gives us an understanding of what is happening in the plasma to create the undesirable plasma states approaching the disruption. Early in the discharge (before t = 0.25s), MHD modes are also found, and core plasma rotation is low as the plasma starts up and typically transitions from counter-NBI rotation to co-NBI rotation. However, the mode frequencies are relatively high at this time, which is generally a safe condition. Later, near t = 0.25s, the MHD warning level increases as modes are again found but now with decreased and decreasing rotation frequency. However, these frequencies are not critically low (no mode bifurcations are found) and plasma rotation is not low, so the warning level remains low. However, after t ~ 0.6s the heat map clearly shows more negative (destabilizing) criteria occurring simultaneously including an increased mode amplitude, decreasing mode frequencies, and decreasing plasma rotation across the profile. Near t ~ 0.7s more negative criteria occur: mode frequencies are now below a recently evaluation of mode bifurcation frequency levels, the modes drop to very low frequency, and core plasma rotation (three channels of the plasma rotation profile are considered – core, middle, and outer) is critically low. Late in the evolution in close time proximity to the disruption (t ~ 0.8s), a critical level of locked mode amplitude occurs. Such a locked mode detector signal is typically used as the primary, and occasionally the only indicator to predict a possible disruption, but this indication occurs very late in the evolution. We see here that the DECAF analysis starts to show a significant change in the total MHD warning level about 180 ms earlier, providing far better advanced notice of the potential disruption allowing the potential for control systems to alter plasma stability to avoid disruption. Additionally, and of critical importance, the DECAF analysis provides physical understanding of the negative evolution of the plasma state as it moves toward the disruption. Further forecasting of resistive MHD stability using the resistive DCON code is being investigated through supporting KSTAR research[19].

3.3 Global MHD Instabilities

Global MHD instabilities such as external kink/ballooning modes or resistive wall modes (RWMs)[20] can cause the most rapid disruptions (e.g. Fig. 1) and give the least amount of pre-disruption warning time. Therefore, attention needs to be put toward forecasting such events to cue profile control systems well before instability develops.

3.3.1 Disruption Event Characterization

To examine disruption event chains with global MHD triggers, DECAF analysis was performed on a database of 45 NSTX discharges that were pre-determined to have unstable RWMs which lead to disruptions. Tearing modes were stable during these discharges to focus on global MHD in this analysis. A typical disruption event chain with an RWM trigger was shown in Fig. 1. In this database, the RWM, loss of boundary control (WPC), LOQ, IPR, and DIS events are found in 100% of the plasmas, and VDE events are found in 91% of the plasmas. A pressure peaking warning (PRP) occurred on a majority of the discharges analyzed (35 of 45), but typically occurred with or after the RWM, not before in this database. The GWL event warning is found in a few cases when the warning level is set at a Greenwald fraction of 0.9. Interestingly, GWL can start the RWM disruption chain and is explained by the correlation of reduced plasma rotation caused by increasing plasma density, leading to RWM instability by a destabilizing change in the plasma rotation profile, discovered in NSTX[21]. Analysis shows that 61% of RWM events in a shot occur within 20 conducting wall current diffusion times, τ_w, of the disruption. The other RWM events found occur earlier but are not false positives as they cause significant thermal collapses or "minor disruptions" of the plasma with subsequent recovery (plasma stored energy can drop by 30% or more over tens of ms, much larger than the largest ELMs in tokamaks which cause far smaller stored energy decreases up to ~ 6%). It is useful to examine which events are commonly associated with this dynamic. For example, the low safety factor condition (LOQ) was clearly seen to happen often in close conjunction (either just before or just after) the designated time of disruption (DIS). One way of seeing this is to examine a histogram of some of the timing of the events before the time of disruption (DIS), shown in Fig. 5. Here only the events within 14 τ_w of the disruption are shown, where τ_w is taken to be 5 ms; there are some RWM events at earlier times which are not shown here. It is clear that LOQ and IPR events occur close to the time of disruption, and these are often preceded by VSC, and RWM events which peak around 30 ms prior to the disruption

Examining the common chains of events more closely can provide insight into how to cue avoidance systems to return to normal plasma operations (see Fig. 1). For example, if the RWM can be detected in real-time by a growing mode amplitude signal from an array of external magnetic sensors, it is useful to know what the typical routes of plasma behavior directly follow the RWM so that plasma control systems may be employed to avoid them. Even with the limited dataset examined here, we can find interesting trends. Of the 26 RWM events occurring within 100ms of the disruption, they were followed immediately by VDE (related to bulk plasma motion) 15 times. Further, looking at the two-event chains that

happened directly after this set of RWMs, we find that even though there are theoretically 56 two-event combinations that could occur from the eight currently tested, just two two-event chains accounted for 50% of the cases (VDE ➔ PRP and VDE ➔ WPC) and five accounted for 77%.

3.3.2 Forecasting

Kinetic RWM analysis has shown high success over years of quantitative comparison to experiment to determine the mode marginal stability allowed through plasma precession drift and bounce orbit resonances, collisionality, ν, and energetic particle effects[21,22]. To allow rapid processing, full kinetic RWM computations (using the MISK code[21,22]) that have been used successfully to predict mode stability on NSTX and DIII-D have been used to create a reduced model of the kinetic RWM stability growth rate in DECAF (Fig. 6a). Gaussian functions with parameters fit from full MISK calculations of NSTX marginally stable equilibria are used to define the kinetic energy functional δW_K as functions of *ExB* frequency and collisionality. The model also incorporates expressions dependent on plasma pressure peaking, internal inductance, and aspect ratio for the ideal MHD no-wall and with-wall beta limits computed from thousands of DCON calculations using experimental equilibria[23].The modelled growth rate can be used to forecast RWM instability based on plasma equilibrium reconstructions and rotation measurements and is time-dependent based on the equilibrium evolution. Fig. 6a) illustrates the evolution of a high normalized beta NSTX experimental plasma as it becomes RWM unstable near a predicted marginal stability contour (while not shown, the growth rate contours on Fig. 6a change as the plasma evolves). This reduced kinetic RWM stability model in DECAF, detailed in Reference 23, performed well in its first incarnation against a larger database of plasmas to determine the proximity of discharges to marginal stability (Fig. 6b). The model predicted instability 84% of the time (stringent marginal stability evaluation) for experimentally unstable cases with a relatively low false positive rate. DECAF also showed 44% of plasmas were predicted unstable within 320 ms (~ 60 τ_w) of the disruption time, and 33% were predicted unstable within 100 ms of a minor disruption. Stability was predicted in 77% of experimentally stable cases. The evolution of discharges that were RWM stable were notably separate on the (ExB frequency, collisionality) stability map, not crossing the computed marginal stability contour.

4. Initial Investigation of General Databases

4.1 Individual disruption chain events

The DECAF code has recently produced an initial analysis of large databases for multiple tokamak devices for a small set of disruption characterization events. The analysis is

conducted over the full duration of the planned plasma current flat-top, rather than a limited period near the disruption time as might be available from a disruption database. Thousands of shot seconds are available in the databases, with upwards of $0.5 - 1$ million tested sample times per database. For example, if the DIS event is used, the analysis produces the equivalent of "disruptivity diagrams" showing the probability of a disruption occurring within a given parameter space of tokamak operation. These diagrams are shown for NSTX, MAST, and KSTAR in Fig. 7 expressed as standard stability operational space (l_i, β_N) figures (l_i is the plasma internal inductance). This multi-device comparison illustrates a highly important and still largely unappreciated result separately published for DIII-D and NSTX[24] for smaller datasets that plasma disruptivity does not need to increase (and can actually decrease) as β_N increases. However, as will be shown in the next section, the high beta regions of low disruptivity are in fact key areas for DECAF algorithms to analyse events that can lead to disruptions.

Unlike standard disruptivity plots, DECAF can provide additional insight by illustrating where in parameter space events other than DIS happen. For example, the VDE event detects the loss of vertical stability. When plotted in the parameter space of elongation, κ, vs. l_i, it becomes clear that vertical stability shows a strong dependence on these parameters (Fig. 8a) and that, additionally, the location in parameter space of an event preceding the disruption (like VDE) can be far from where the actual disruption event occurs (DIS, shown in Fig. 8b).

4.2 Disruption event chain analysis for arbitrary discharges

DECAF event characterization and event chain analysis shows that disruption forecasting analysis often start during plasma states that can appear safe. This is illustrated using the disruptivity database plot shown in Fig. 7a) and the figures in this section. The regions of high disruptivity in Fig. 7a) may be thought to be the most important based on human inspection. However, an apparent problem is that the region of high disruptivity at low β_N and mid-range l_i is not physically understood to be a dangerous operational region. The enigma is resolved by understanding that the plasma state can evolve significantly from more usual high performance parameters to the point at which the disruption actually occurs. This fact is completely missed, for example, by disruption database studies that only process data near the disruption time. Even worse, such studies may parameterize disruptive limits based on these misguided terminal states. In contrast, DECAF disruption event chain analysis of two discharges in Fig. 7a) that disrupt (DIS event in DECAF, marked by red and green X's in the figure) show that the *start* of the event chains appears in the region indicated by the red and green circles – which are *far* from what might be expected. This also illustrates why the use of a numerical tool that would focus only on the regions of high disruptivity (such as a "black box" machine learning approach) would produce an inaccurate assessment of the

plasma states that produce early disruption forecasts. The disruption event chain for these plasmas in Fig. 7a) are shown in Fig. 9 along with the DECAF MHD mode decomposition and total MHD warning level. As before, we see this warning level rising toward and past the critical value of 4.0 as the disruption is approached. The DECAF mode decomposition adds information showing that the mode evolution toward lower rotation frequencies is relatively slow. This is one reason why the plasma disrupts far from the plasma state at the trigger event.

The DECAF event chains in Fig. 9 provide a wealth of information. In Fig. 9a), we see a critical warning for the individual $n = 1$ rotating MHD mode (MHD-n1) as a starting point for the chain. Note from the top frame that the low frequency $n = 1$ mode itself was detected far earlier – near t ~ 0.22s. However, the warning level for the activity was not determined to be sufficiently high then. The mode bifurcation (event BIF-n1) occurs 5 ms later. The mode locks (event LTM-n1) 45 ms after the bifurcation. Then, a different dynamic occurs, as DECAF finds a pressure peaking event warning (PRP) happening 23 ms later. While the warning literally flags that the pressure peaking factor is exceedingly high, it also importantly indicates that an H-L energy confinement back-transition has occurred, the H-mode pedestal is lost, and the neutral beams have better penetration increasing the plasma pressure peakedness. The IPR warning occurs 5 ms after PRP and simultaneously the plasma makes a close approach to the vessel wall (WPC). Finally, the plasma disrupts 4 ms after the WPC event. It is also interesting that the VDE event warning occurs 3 ms after DIS. Usually the events are reversed in time. This indicates that the plasma remains mainly on the midplane during the evolution, uncharacteristic of NSTX disruptions. Fig. 9b) shows a relatively slow RWM-triggered disruption (i.e. compared to Fig. 1b)). In this disruption event chain, the PRP warning again indicates an H-L back transition and a VDE is produced approximately 10 τ_W after the RWM trigger occurs. As shown, the disruptions in these two plasmas occur 77 ms and 101 ms after the initial DECAF warnings. These intervals represent transport timescales (a few energy confinement times) and so would allow sufficient time for active profile control for disruption avoidance, or easily allow time for active mode control or disruption mitigation.

4.3 Disruption Prediction Performance

Disruption prediction research using the DECAF approach also importantly allows *quantifiable* figures of merit (most importantly the plasma disruptivity) to assess any prediction models produced. This figure of merit allows an objective assessment of the relative performance of different models, and allows an assessment of how close the predictor would come compared to ITER needs. Fig. 10 shows a progression of DECAF disruption forecasting models. The earliest models included about 10 events and were run in

databases for which the events that led to the disruption were known. For such databases, DECAF produced very high performance (e.g. 100% true positives). A next evaluation of models focused on earlier forecasting once the first forecasting model was implemented in the code. True positives were found to be ~84%, which was a measure mainly of the single forecasting model. The addition of more forecasting models (such as the MHD analysis shown) could improve that performance with further development. The most recent testing of the code has been on large databases ~ 10,000 shot*seconds of plasma run time tested. This was done with a smaller number of events due to computer RAM limitations. With 5 events, applied to all plasma shots from an NSTX database, DECAF has produced performance levels of over 91% true positive disruption predictions. False positives in this analysis reached 8.7% which is fairly high. However, further code development that allows the events to poll each other will reduce this level considerably.

5. Summary and Discussion

The Disruption Event Characterization and Forecasting Code is a physics-based, fully automated analysis paradigm that continues to compile the knowledge of many years of tokamak research by implementing a collection of models to solve the critical issue of plasma disruptions in tokamaks. The approach considers the evolution from normal tokamak operation toward a plasma disruption as occurring through a linked series, or "chain" of events, most or all being "off-normal". The approach has several advantages including an analysis that produces greater understanding of the events as a collective, and allowing control systems to be guided by the events and events chains in taking early actions (e.g. on transport timescales) to avoid disruptions before they happen via the forecasting element of the analysis. The approach is inherently deterministic rather than statistical. The DECAF approach meets all of the disruption predictor requirements outlined by Humphreys, et al.[9] Four of the five requirements have been demonstrated in this paper. The DECAF events comprise both attributes and methods which include the ability to predict the specific plasma physics phenomena covered by the scope of each event. The events in combination provide a continuous variable quantifying the proximity to disruptive states. The events themselves, and more powerfully the event chains can be used to more intelligently trigger actions by control systems (e.g. by using gradients of the underlying physical processes to determine how to change the plasma state) than a system that only provides a binary classification of the plasma state. Present model and forecasting development has produced analysis that forecasts the onset of a disruption on plasma transport timescales, thereby allowing sufficient lead time to avoid plasma disruption through plasma control techniques (e.g. profile control, shape control, plasma stored energy control, plasma mode control). The DECAF paradigm puts high priority on the ability to characterize and forecast plasma disruptions from the full

databases of multiple tokamaks to allow more stringent validation of the underlying physics models across several devices, thereby producing superior extrapolability to future devices.

While DECAF connects to the databases of several tokamaks, once read into the code the data used is abstracted from the specific local data names to general names (e.g. CES toroidal plasma rotation profile on KSTAR[25] and CHERS plasma rotation profile on NSTX[26] becomes "plasma rotation" in DECAF, with a generalized input describing the measurement geometries. This allows straightforward comparisons between quantities and consistency with future common data management systems, such as IMAS for ITER[27], the exact common variable names for which are presently being implemented.

Modern machine learning (ML) techniques were envisioned to be used in specific ways in DECAF analysis since its origin once they were found to be needed. The first instance of this was thought to be the reproduction of the mode growth rate results for the kinetic MHD global mode model detailed in Reference 23 by a deep learning neural net approach[28]. However, as shown in the reference, basis functions that produced good fits to the analysis results were found based on the underlying physics solutions, and so that approach was utilized. Having analytic forms for the models are superior to purely numerical approaches in that gradients of the key quantities used for forecasting and control algorithms can be computed analytically, producing smooth results to reduce analysis output noise. However, finding such basis functions is not straightforward, and so is not possible in general. Therefore, we have now started to implement machine learning for DECAF, but in significantly different ways than have been used for disruption prediction research to date, which typically approach the problem treating a disruption database as a binary classification problem, and applying deep learning or other techniques to the database as a "black box". Machine learning in DECAF follows a philosophy that is more amenable to produce human understanding of the results and allow greater flexibility for use in control systems. Specifically, we are presently adopting three approaches of using machine learning to support DECAF. First is the reduction of results from certain complex physical models by deep learning neural nets to allow rapid (including real time) determination of quantities used in DECAF models. Two such machine learning techniques following this approach above have already been applied as analysis supporting the DECAF code, specifically, deep learning neural nets, and non-linear random forest regression analysis. These are presently being used to train on DCON ideal MHD stability code calculations and will be the subject of a future paper. Second, a "hybrid model" approach is envisioned to comprise the bulk of ML use in DECAF. In this approach, both a physics model and a database that the physics model is meant to reproduce are given. The difference between the computed physical quantities produced by the model and the data, δ_e, is typically interpreted as the error between these

two, for which a physical model is not known. The hybrid model approach focusses on reproducing δ_e, rather than the entire physical model. This approach has two advantages: (i) the key independent variables needed to produce the ML model of δ_e will be a smaller set than required to reproduce the entire database using ML alone, and (ii) ML techniques have a greater probability of exposing key physics variables of δ_e, which can produce human understanding of a physical model that explains it. In that case, the new physics discovered that reproduces δ_e can be added to the original physical model, with the opportunity to iterate this process to discover more of the underlying physical model. The approach of addressing δ_e is analogous to the use of an optimum observer in control system design.[29] Finally, ML techniques will be used to evaluate more general linkages between DECAF events.

There is one final disruption predictor requirement remaining to be demonstrated by DECAF – that it be real-time calculable. This characteristic can also be met by all models discussed in this paper, and that are presently available in DECAF. The present offline modeling and analysis is starting to be implemented for real-time use in the KSTAR device. In such an implementation, only causal analysis techniques based on real-time measurements and real-time equilibrium reconstruction quantities can be used. Comparison of disruption prediction performance results between the two approaches will be made once the results from real-time analysis are completed.

Figure Captions

Fig. 1: (a) Schematic diagram illustrating a paradigm of plasma state evolution away from normal operation toward a plasma disruption as a series of events that form a disruption event chain, (b) automated evaluation of the disruption event chain for a plasma discharge with $\beta_N = 4.5$, (c) and (d) higher time resolution illustration of n = 1 RWM amplitude and plasma toroidal rotation as the disruption is approached.

Fig. 2: Profiles of calculated (deuterium and carbon) and measured total radiated power density and calculated input power density for NSTX discharge 134020 at 0.60s.

Fig. 3: (a) spectrogram of rotating MHD activity from a toroidal array of magnetic probes for NSTX discharge 134020, illustrating n = 1 mode growth near the time of the loss of power island power balance (DECAF event IPB), (b) Greenwald fraction and local power balance criterion.

Fig. 4: Rotating MHD mode discrimination capabilities in DECAF. The upper left frame shows the mode discrimination and decomposition into DECAF events. The lower left panel shows a total MHD warning level that increases as the disruption is approached. The right panel shows a heat map illustration of 15 event criteria that comprise the total MHD warning level.

Fig. 5: Histogram of the timing of RWM event triggered disruption chain events before the time of disruption.

Fig. 6: (a) Stability map vs. ExB frequency and collisionality from DECAF reduced kinetic RWM stability model; (b) statistics illustrating results of the model in forecasting instability for RWM unstable NSTX plasmas.

Fig. 7: Event probability diagram of DECAF event DIS during I_p flat-top showing that disruption probability does not have to increase as plasma normalized beta is increased based on large databases from the NSTX, MAST, and KSTAR tokamaks.

Fig. 8: Event probability diagrams of DECAF events VDE (a) and DIS (b) for a large database from the NSTX tokamak.

Fig. 9: (top to bottom) DECAF decomposition of rotating MHD in relatively slow evolutions toward disruption; total MHD warning signal; DECAF event chains leading to disruption. Plasma in frame (a) is triggered by rotating MHD, (b) by RWM.

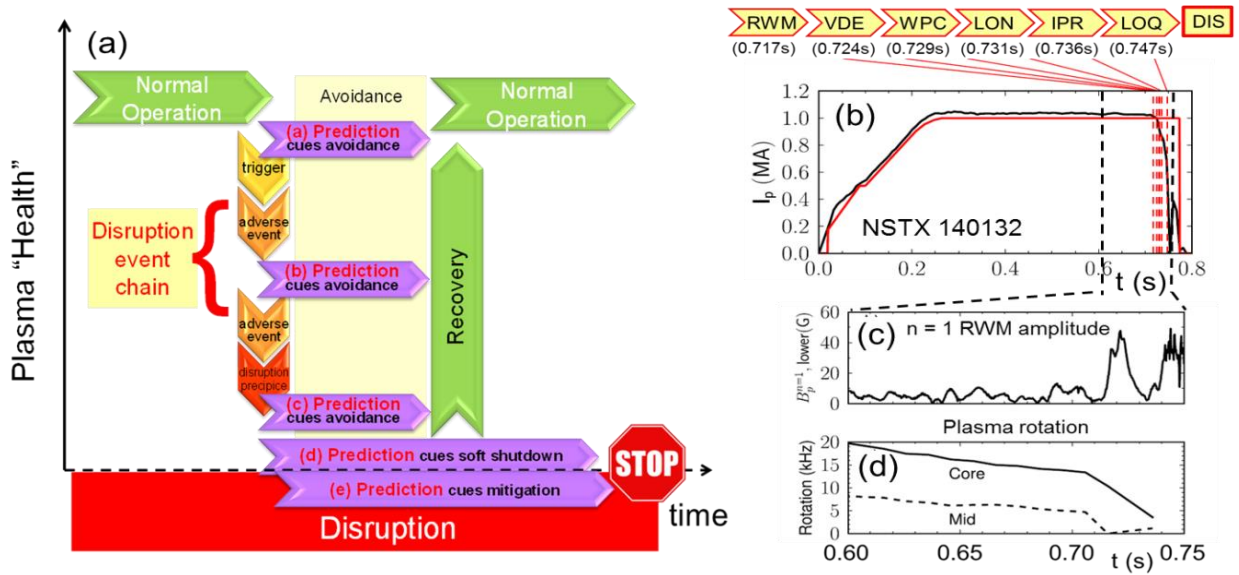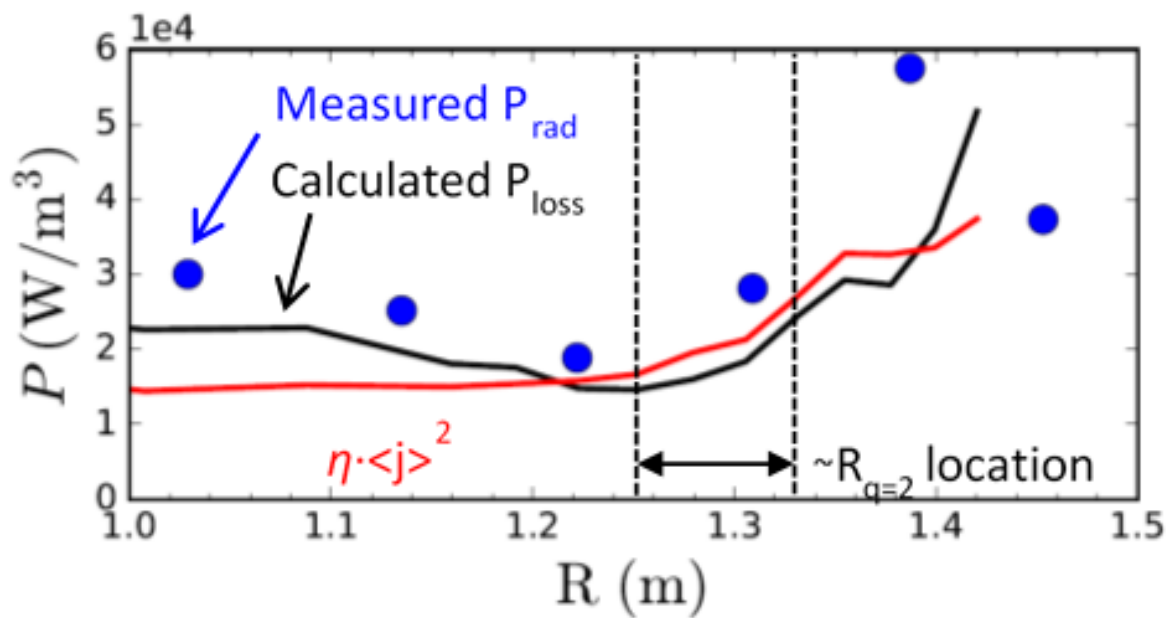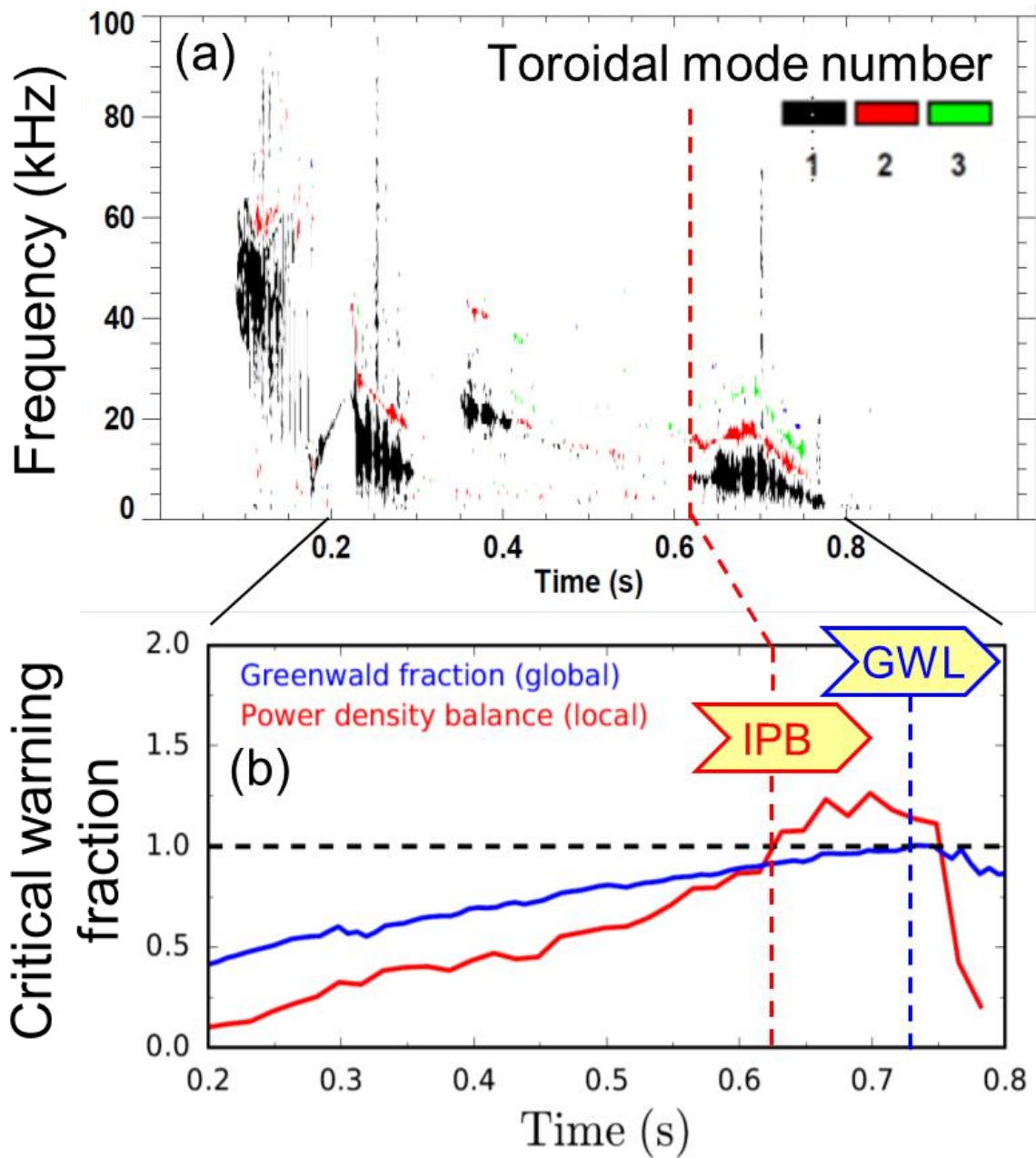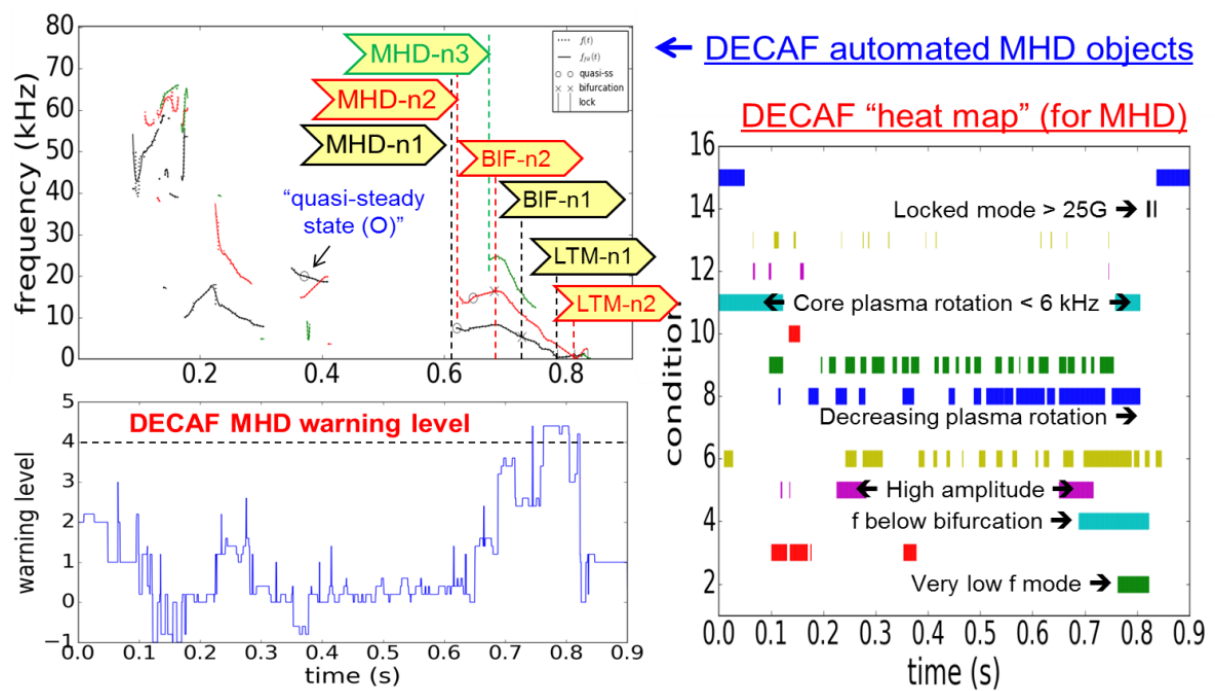*Fig. 10: DECAF model performance evolution (true positive disruption forecast).*

FIG. 1

FIG. 2

FIG. 3

FIG. 4

FIG. 5
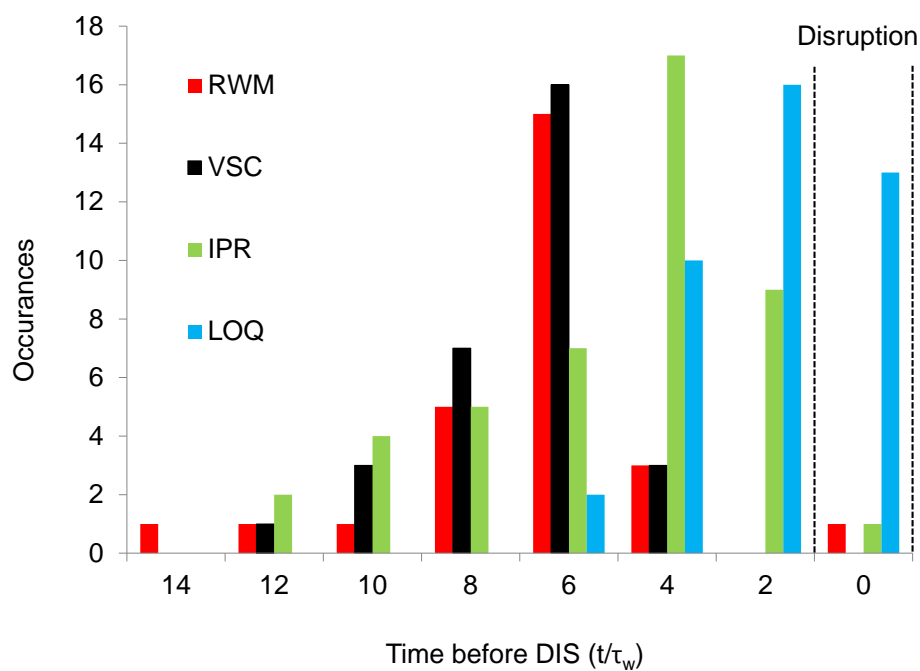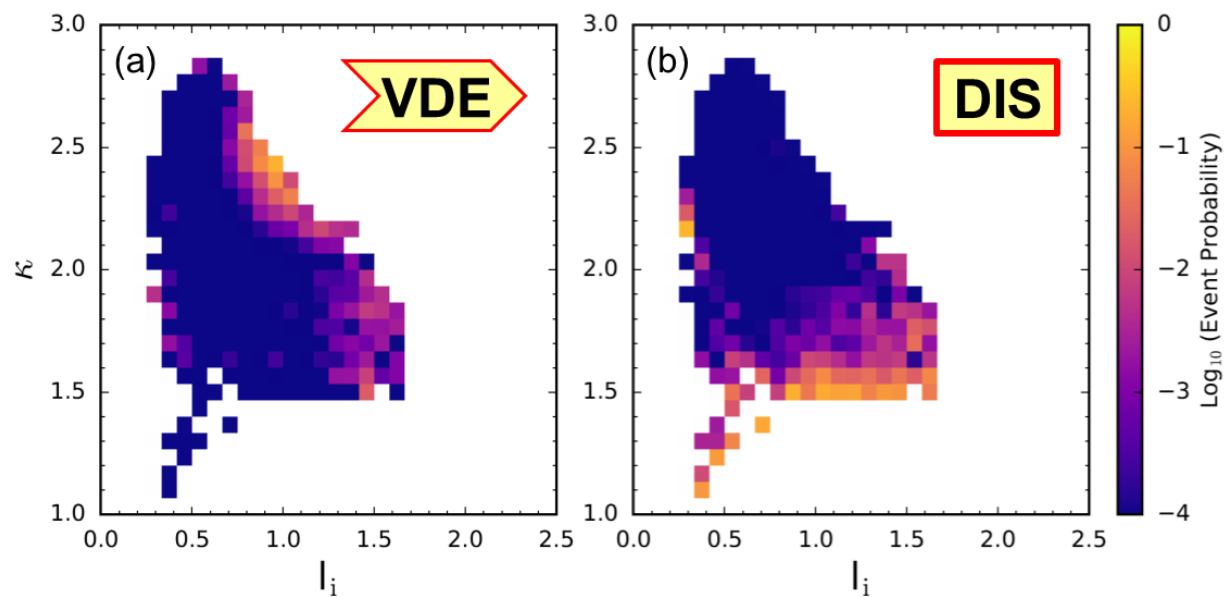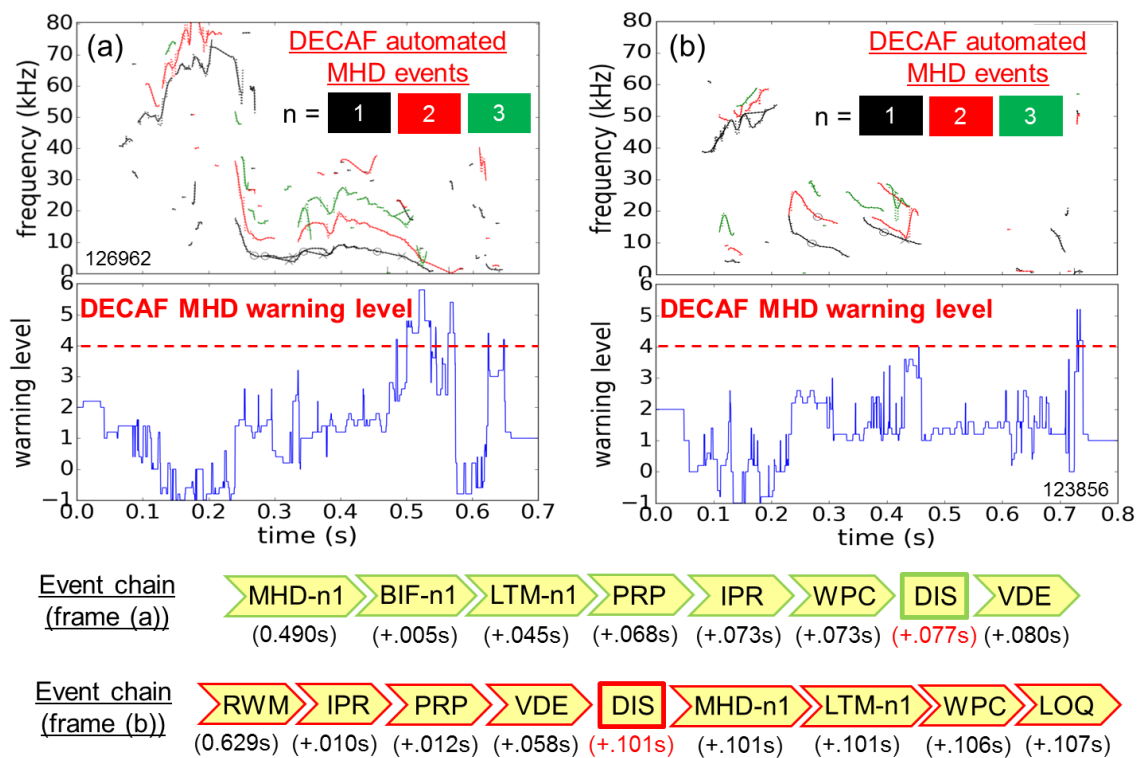
FIG. 6
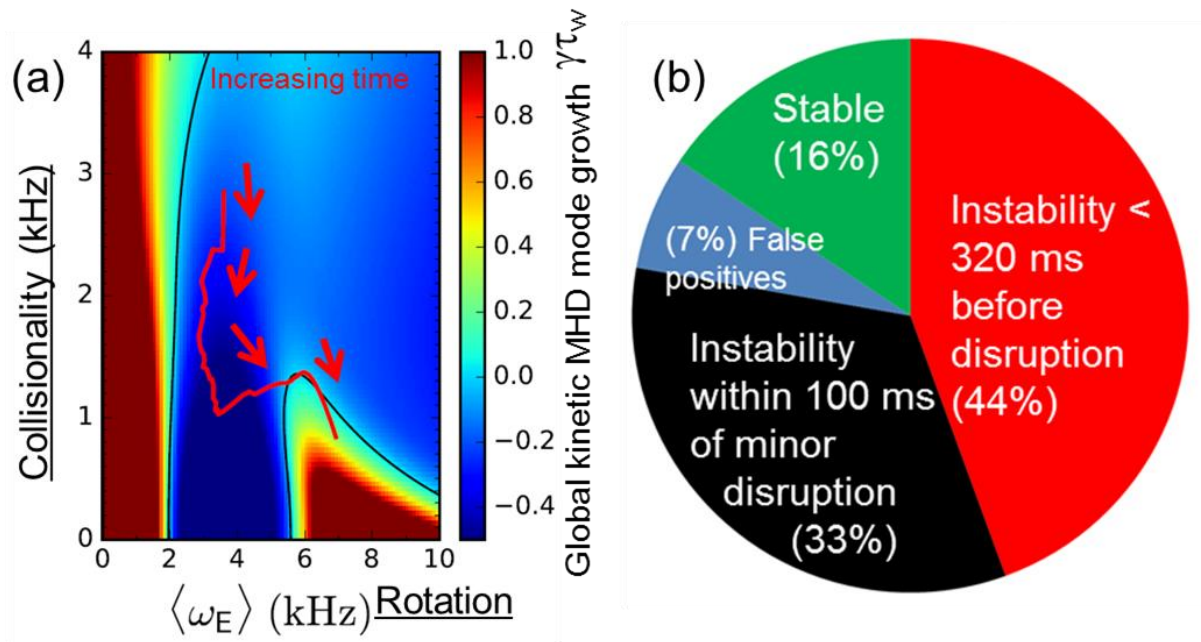
FIG. 7
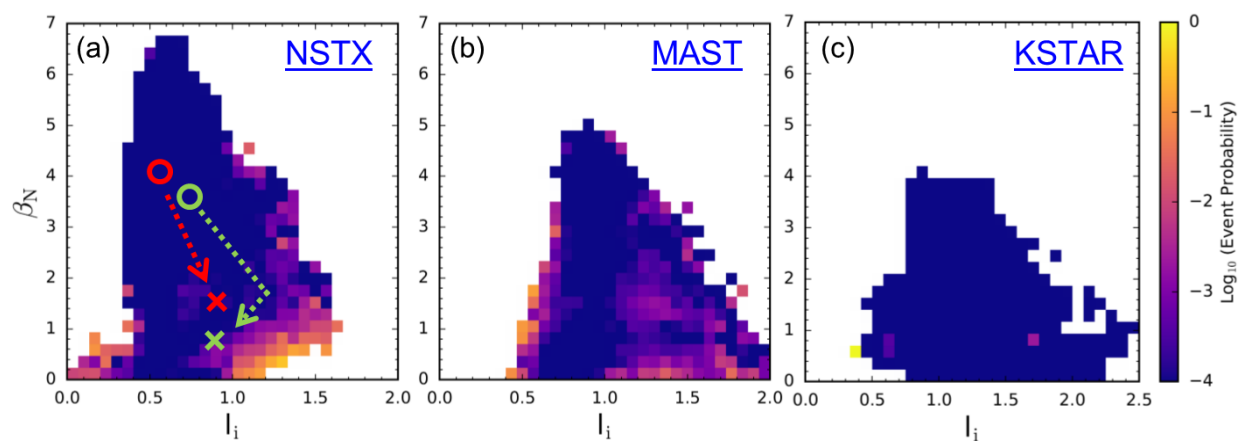
FIG. 8

FIG. 9

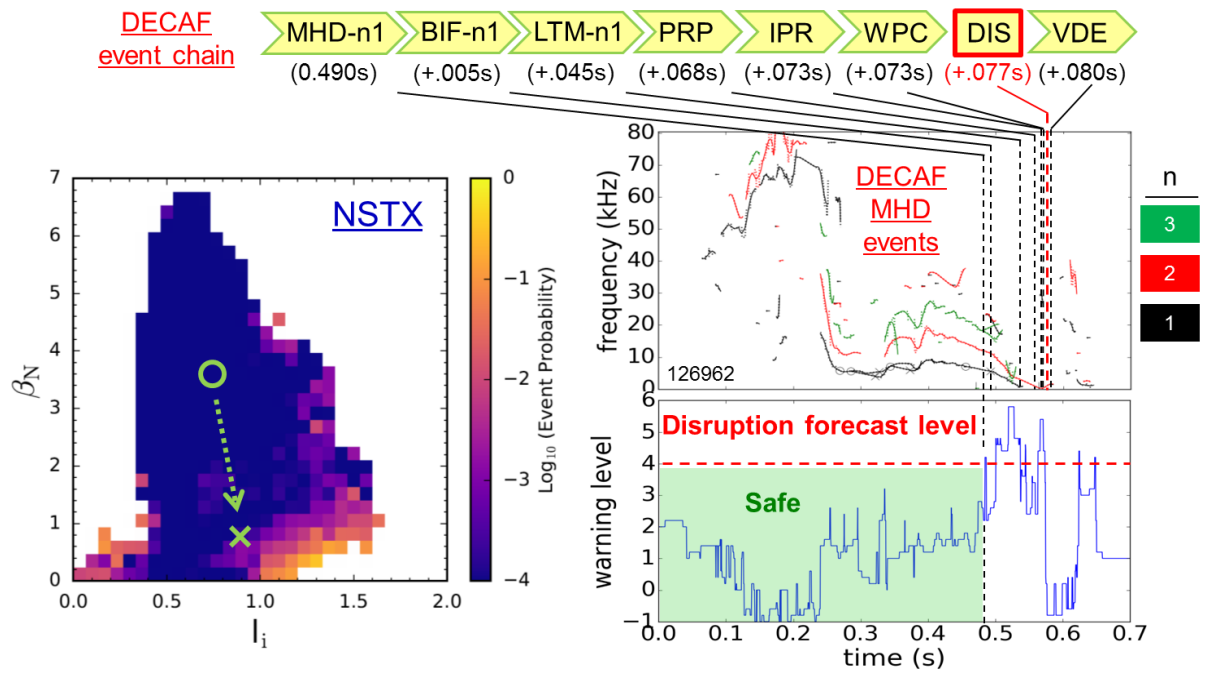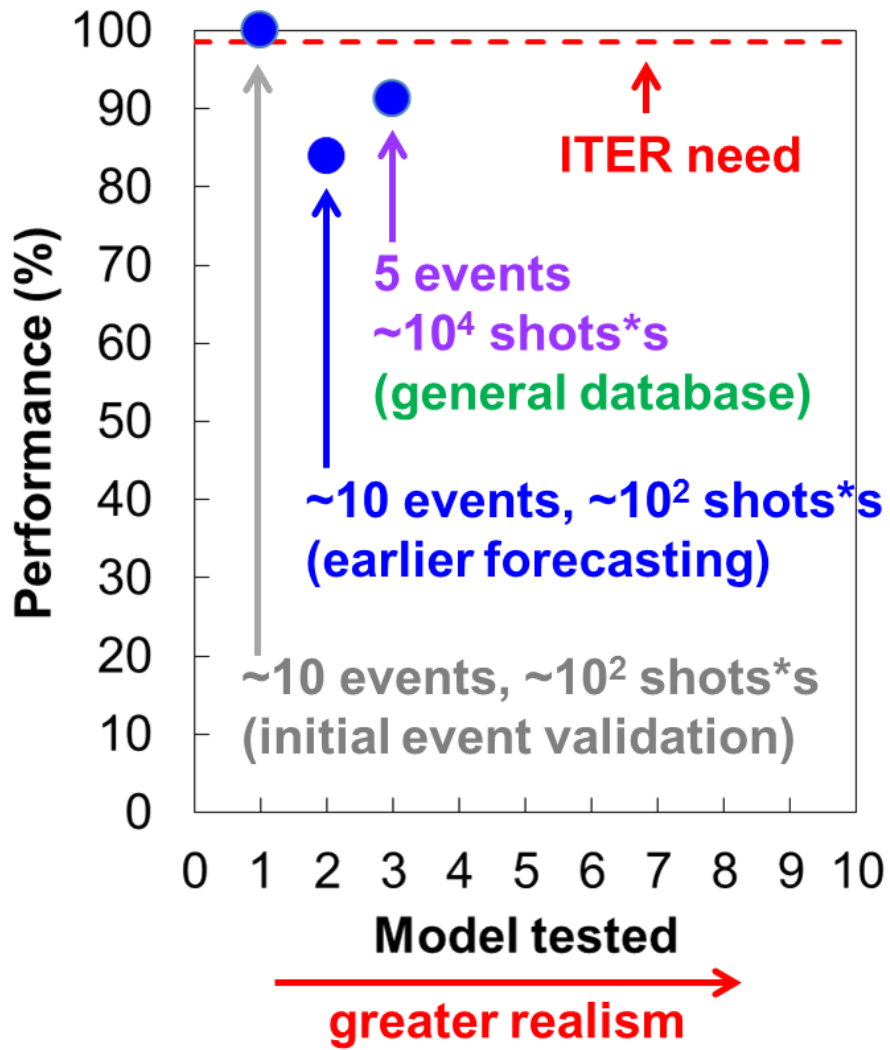============================== (alternate version of Figure 9(a))

FIG. 10

References

[1] T.C. Hender, J.C. Wesley, J.M. Bialek, et al., Nuclear Fusion **47**, S128 (2007).

[2] N.W. Eidietis, S.P. Gerhardt, R.S. Granetz, et al., Nuclear Fusion **55**, 063030 (2015).

[3] M. Sugihara., et al. 2012 Proc. 24th Int. Conf. on Fusion Energy (San Diego, CA, 2012) ITR/P1-14, www-naweb.iaea.org/napc/physics/FEC/FEC2012/index.htm.

[4] P.C. DeVries, M.F. Johnson., Alper, et al., Nuclear Fusion **51**, 053018 (2011).

[5] P.C. de Vries, M.F. Johnson, I. Segui and JET EFDA Contributors, Nuclear Fusion **49**, 055011 (2009).

[6] P. C. de Vries, M. Baruzzo, G. M. D. Hogeweij, S. Jachmich, E. Joffrin, P. J. Lomas, G. F. Matthews, A. Murari, I. Nunes, T. Pütterich, C. Reux, J. Vega, and JET-EFDA Contributors, Physics of Plasmas **21**, 056101 (2014).

[7] B. Cannas, P. C. de Vries, A. Fanni, A. Murari, A. Pau, G. Sias and JET Contributors, Plasma Phys. Control. Fusion **57**, 125003 (2015).

[8] M. Maraschek, A. Gude1, V. Igochine1, et al., Plasma Phys. Control. Fusion **60**, 014047 (2018).

[9] D. Humphreys, et al., Physics of Plasmas **22**, 021806 (2015).

[10] S.A. Sabbagh, R.E. Bell, J.E. Menard, et al., Physical Review Letters **97**, 045004 (2006).

[11] M. Ono et al., Nuclear Fusion **40**, 557 (2000).

[12] C.E. Myers, N.W. Eidietis, S.N. Gerasimov, S.P. Gerhardt, R.S. Granetz, T.C. Hender, G. Pautasso and JET Contributors, Nuclear Fusion **58**, 016050 (2018).

[13] S.P. Gerhardt, D.S. Darrow, R.E. Bell, et al., Nuclear Fusion **53**, 063021 (2013).

[14] M. Greenwald, J.L. Terry, S..M. Wolfe, et al., Nuclear Fusion **28**, 2199 (1988).

[15] T. Eich, R.J. Goldston, A. Kallenbach, et al., Nuclear Fusion **58**, 034001 (2018).

[16] D.A. Gates, and L. Delgado-Aparicio, Physical Review Letters **108**, 165004 (2012).

[17] Q. Teng, D.P. Brennan, L. Delgado-Aparicio, et al., Nuclear Fusion **56**, 106001 (2016).

[18] D.E. Post, et al., Atomic Data and Nuclear Data Tables **20**, 397 (1977).

[19] Y.S. Park, S.A. Sabbagh, J.H. Ahn, et al., 2018 Proc. 27th Int. Conf. on Fusion Energy (Gandhinagar, India) EX/P7-16.

[20] A. Bondeson and D. Ward, Physical Review Letters 72, 2709 (1994).

[21] J.W. Berkery, S.A. Sabbagh, R., Betti, et al., Physical Review Letters **104**, 035003 (2010).

[22] S.A. Sabbagh, J.W. Berkery, R. Betti, et al., Nuclear Fusion **50**, 025020 (2010).

[23] J.W. Berkery, S.A. Sabbagh, R.E. Bell, Physics of Plasmas 24, 056103 (2017).

[24] J.W. Berkery, S.A. Sabbagh, A. Balbaky, et al., Physics of Plasmas **21**, 156112 (2014).

[25] W. Ko, H. Lee, D. Seo. M. Kwon, Rev. Sci. Instrum, 81, 10D740 (2010).

[26] 2B. Stratton, R. J. Fonk, K. P. Jaehnig, N. Schechtman, and E. J. Synakowski, in *Proceedings of the IAEA Technical Committee Meeting on Time Resolved Two- and Three-Dimensional Plasma Diagnostics, Najoya, Japan, 1990* (IAEA, Vienna, 1991) p. 78

[27] F. Imbeaux et al., Nuclear Fusion **55**, 123006 (2015).

[28] F. Chollet, *"Deep Learning With Python",* Manning Publications Co. (2018) ISBN: 9781617294433.

[29] B. Friedland, *"Control System Design: An introduction to State-Space Methods"*, Dover Publications, Inc. (1986) ISBN: 0-486-44278-0.