SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Dipartimento di / Department of

Fisica

Dottorato di Ricerca in / PhD program Fisica Ciclo / Cycle 37

Curriculum in (se presente / if it is)

TITOLO TESI / THESIS TITLE

Cognome / Surname Bonalumi Nome / Name Luca

Matricola / Registration number 777689

Tutore / Tutor: Paola Mantica

Cotutore / Co-tutor:
(se presente / if there is one)

Supervisor: Carlo Sozzi
(se presente / if there is one)

Coordinatore / Coordinator: Stefano Ragazzi

ANNO ACCADEMICO / ACADEMIC YEAR 2023/2024

University of the Study Milano-Bicocca

Physics department

Investigation of the (neoclassical) tearing instabilities from the trigger to the disruption in JET and DTT

Supervisor:
Carlo Sozzi

Candidate:
Luca Bonalumi

Academic year 2023/2024

Contents

Summary

Nuclear fusion represents one of the most promising energy sources due to its high energy density, abundant fuels, and intrinsic safety. The process involves two light nuclei that fuse, producing energy that can be gathered and converted into electricity. The temperature necessary to achieve nuclear fusion forces the matter to the state of plasma, a system of charged particles, globally neutral, with collective behaviour. The tokamaks are machines used to confine a plasma of hydrogen isotopes, employing a combination of magnetic fields, allowing the plasma to be heated until the energy of the particles overcomes the coulomb barrier and the nuclei fuse.

The plasma exhibits instabilities that develop over various time and spatial scales. This thesis focuses on a resistive magnetohydrodynamical instability called the neoclassical tearing mode ((N)TM). The onset of the (N)TM manifests with a change in the topology of the confining magnetic field, forming a magnetic island that can grow, reducing performance and possibly leading to a complete loss of confinement of the plasma (disruption). The investigation of the (N)TM involves different aspects: the trigger of the mode, its evolution and control, and the role (N)TM can play in disruptive phenomena. In this work, different aspects addressed above have been investigated by exploiting both experimental data from the Joint European Torus (JET) and predictive simulations of (N)TM evolution and control based on available scenarios from the Divertor Tokamak Test (DTT).

After a general introduction (Chapter 1) and a description of DTT and JET (Chapter 2), a detailed theoretical analysis related to the physics of (N)TM is set in Chapter 3.

Chapter 4 covers the topic of NTM control. The (N)TM evolution can be controlled by injecting powerful Electron Cyclotron Waves that drive a current within the island. Predictive simulations in the DTT full power scenario were carried out to address some requirements for designing the electron cyclotron systems and investigate possible suppression strategies.

The onset of the pre-disruptive (N)TM has been modeled in JET disruptive plasma. A physical picture where the ion polarization current covers a role in the trigger of tearing instabilities has been proposed in a published work (Appendix A) and introduced in Chapter 5.

The topic of the early identification of disruptive conditions with deep learning is addressed in Chapter 6 using a dataset composed of JET disruptive plasma preceded by the onset of an (N)TM. The classification of a deep neural network has been analyzed via explainable artificial intelligence (XAI) methods to verify if the network can distinguish between different disruptive patterns that involve the trigger of a 2/1 mode. Section 6.4 presents the published work (reported in Appendix B)

Chapter 1

Introduction

The controlled fusion of hydrogen isotopes in plasma is considered one of the most promising future energy sources, offering high efficiency and improved environmental safety compared to nuclear fission. In laboratory settings, the most energetically favorable reaction for thermonuclear fusion involves the fusion of deuterium (D) and tritium (T), which produces a α particle and a neutron, releasing 17.6 MeV of energy per reaction:

$$\mathrm{D}_2^1 + \mathrm{T}_3^1 \rightarrow \mathrm{He}_4^2 + \mathrm{n}_1^0 \tag{1.1}$$

This reaction is particularly advantageous because it has the highest fusion cross-section at relatively low energies, making it more accessible than other potential fusion reactions. The Coulomb barrier, which causes repulsion between positively charged nuclei, typically limits the fusion probability, especially at low energies. However, the deuterium-tritium reaction can proceed at temperatures achievable in current experimental setups thanks to quantum mechanical tunneling, which allows nuclei to overcome the Coulomb barrier more easily. As shown in Fig.1.1, other fusion reactions, such as deuterium-deuterium (D-D) and deuterium-helium-3 (D-He^3), are possible but at significantly lower cross-sections at equivalent energy levels. For instance, the D-D reaction, which can yield either a proton and tritium or a neutron and helium-3, requires higher temperatures and delivers less energy per reaction (around 4 MeV). Similarly, the D-He^3 reaction produces a proton and helium-4 but necessitates even higher plasma temperatures than D-T fusion, offering a peak energy output of 18.3 MeV. These alternatives are less feasible in current technology due to their higher energy thresholds and lower reaction rates. In a future D-T fusion reactor, the neutrons produced in the fusion process would be used to breed tritium from lithium, which is abundant on Earth, through interactions with the reactor's walls. The α particles produced in the fusion reaction contribute to heating the plasma. When confinement conditions are sufficiently optimized, the temperature can be maintained by this α particle heating alone, overcoming energy losses. When this occurs, the plasma reaches the "ignited" state. The progress toward ignition is commonly described by the triple product of density (n), temperature (T), and energy confinement time (τ_E) as:

$$nT\tau_E > 3 \cdot 10^{21}\mathrm{m}^{-3}\mathrm{KeV\ s} \tag{1.1}$$

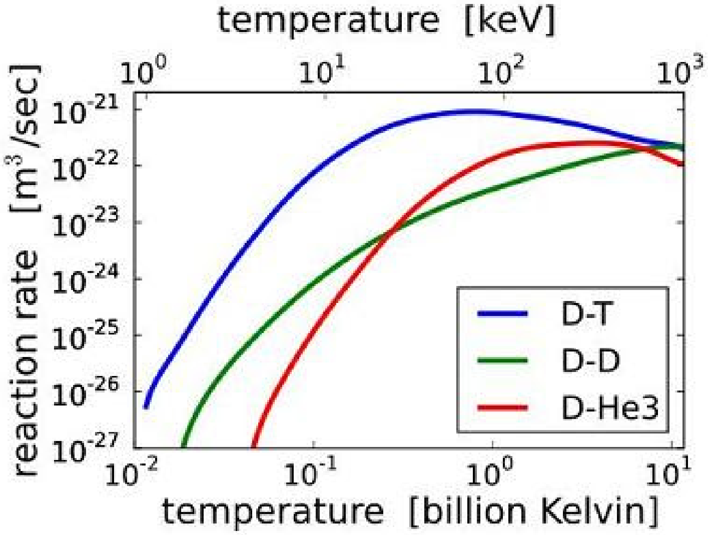


Figure 1.1: Reaction rate of the main nuclear fusion reactions as a function of the temperature [116].

At the extremely high temperatures necessary for fusion, on the order of hundreds of millions of degrees Celsius, the hydrogen isotopes transition into a plasma state. A plasma is a state of matter in which the particles are fully ionized, consisting of free electrons and nuclei. In this state, the high kinetic energy of the particles can overcome the Coulomb barrier, which repels positively charged nuclei. The *tokamak* is currently the most advanced prototype of a nuclear fusion reactor. It uses powerful magnetic fields to confine plasma in a toroidal shape, creating conditions for fusion reactions. The tokamak's magnetic confinement helps sustain the plasma at the extremely high temperatures—around 150 million degrees Celsius—required for efficient fusion.

1.1 Tokamak

As pointed out before, the energy required to overcome the coulomb barrier forces the particles to change state in a plasma. The energy must be confined for a time τ_E sufficient to satisfy Lawson's criteria and reach the condition for the ignition. Two approaches are currently followed: inertial confinement fusion, where the plasma is confined for a time that depends on its diffusion (and hence connected to its inertia), and magnetic confinement fusion, which tries to maximize the confinement time by means of a magnetic field. Indeed, the plasma, composed of charged particles, interacts with the magnetic field so that, in principle, it is possible to think of a magnetic configuration where the plasma is detached from the cold wall, maintaining the temperature necessary for achieving a sufficient amount of fusion reactions. As the motion of the charged particles is unbounded along the magnetic field lines, the plasma needs to be confined over surfaces where, at ev-

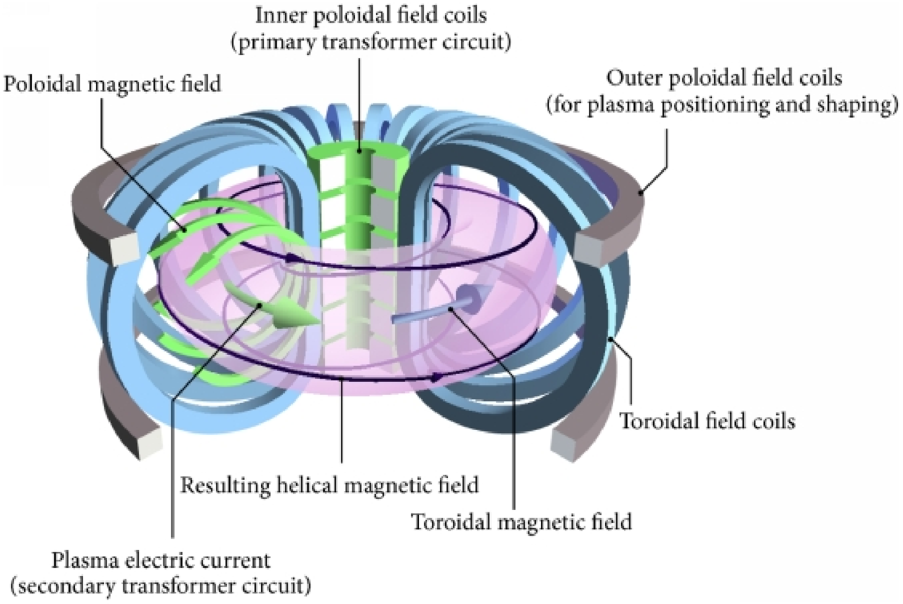


Figure 1.2: Illustration of a tokamak, showing the toroidal magnetic field line, produced by the toroidal field coils. The plasma current, induced by the central solenoid, that produces the poloidal field, twisting the magnetic field lines [67]

ery point, the confining magnetic field is non-vanishing and tangent to the surface. However, the well-known Poincare's theorem constrains the possible geometries that can be used, stating that the only topology where it is possible to define a non-vanishing tangent field in every point is a torus. One of the most promising configurations for obtaining a future fusion reactor prototype is the tokamak, a toroidal chamber where a magnetic field confines the plasma. The primary component of the magnetic field, denoted as B_ϕ, is oriented in the toroidal direction and is generated by conducting coils encircling the chamber. The plasma particles move along the field lines following a helical trajectory, with their radius determined by the Larmor radius $r_L = \frac{mv_\perp}{eB_\phi}$. According to Ampere's law, B_ϕ follows the curvature of the machine and scales as $1/R$. These inhomogeneities in curvature and intensity produce drifts that depend on the charge of the particles, leading to charge separation and generating an electric field. This electric field causes the plasma column to move vertically, quickly disrupting the confinement. This behavior, known as vertical displacement instability, can be handled by introducing a secondary magnetic field pointing in the poloidal direction. The poloidal magnetic field B_θ is generated by a current induced within the plasma, functioning analogously to the secondary circuit of a transformer. In this context, the primary circuit is constituted by a central solenoid through which the primary current flows. This configuration is stable (actually metastable, as explained at the end of the chapter), and the equilibrium can be described through the magnetohydrodynamics (MHD), a single-fluid model. In the next section, a brief description of the MHD equilibrium will be given.

1.2 Ideal magnetohydrodynamics

As pointed out before, toroidal geometry is the only topology in which it is possible to define a field tangent to the surface everywhere. A surface densely covered by a magnetic field in a tokamak is called *irrational flux surface*. In a toroidal geometry, the plasma is confined on a flux surface, and a particle can pass from one surface to another solely due to collisions. The equilibrium of the plasma on the flux surfaces can be described using the MHD, derived in terms of single-fluid quantities:

$$\rho = m_i n \tag{1.2}$$

$$\vec{v} = \vec{u}_i \tag{1.3}$$

$$\vec{J} = en(\vec{u}_i - \vec{u}_e) \tag{1.4}$$

$$p = p_i + p_e \tag{1.5}$$

$$T = \frac{(T_i + T_e)}{2} \tag{1.6}$$

where the subscripts e, i indicate the electrons and ions,n is the density, p is the pressure, T is the temperature, u is the species velocity. These definitions are derived under the assumption of quasi-neutral plasma ($n_i = n_e$). Two fluid equations, derived from kinetic models, are added to create a set of single fluid equations containing the single fluid quantities:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = 0 \tag{1.7}$$

$$\rho \frac{d\vec{v}}{dt} - \vec{J} \times \vec{B} + \nabla p = 0 \tag{1.8}$$

$$\vec{E} + \vec{v} \times \vec{B} = 0 \tag{1.9}$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \tag{1.10}$$

$$\nabla \times \vec{B} = \mu_0 \vec{J} \tag{1.11}$$

$$\nabla \cdot \vec{B} = 0 \tag{1.12}$$

Eq. 1.7 represents the mass continuity equation without considering any possible source or sink of particles (nuclear fusion, recombination, charge exchange). Eq. 1.8 is the ideal limit of the balance momentum equation. Eq. 1.9 is the Ohm's law, neglecting the resistivity. The system is closed by Maxwell's equation, neglecting the $\nabla \cdot E$ because, due to quasi-neutrality, the equation is trivial. The MHD equilibrium inside a tokamak is represented by *nested closed toroidal magnetic flux surfaces*. The magnetic field lines covering the flux surfaces define two different kinds of magnetic surfaces depending on the winding step: if the magnetic field line closes on itself after an integer number of

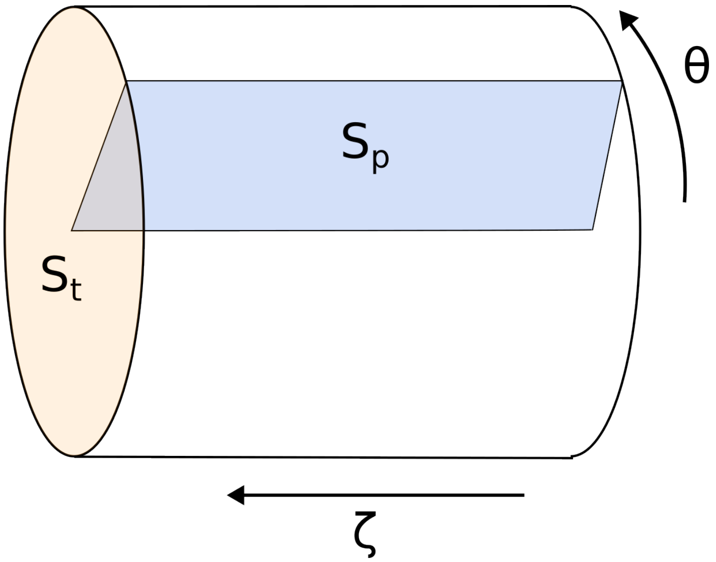


Figure 1.3: Illustration of the surfaces S_p and S_t that define respectively the poloidal and toroidal magnetic field flux.

toroidal turns, the surface is said to be rational. Otherwise, the surface is called irrational. A function that is constant along a flux surface is called flux function, and its value can be used to “label” the flux function. A flux function must satisfy the relation:

$$\vec{B} \cdot \nabla f = 0 \tag{1.13}$$

For a particular magnetic field line that defines the strength of the magnetic field $\vec{B}$ in a specific part of the magnetic surface, the relation 1.13 expresses the propriety of the quantity f to vary only perpendicularly to the field line. The first relevant flux functions are the poloidal and toroidal magnetic fluxes defined as:

$$\Psi_t = \int_{S_t} dS\, \vec{n} \cdot \vec{B} \tag{1.14}$$

$$\Psi_p = \int_{S_p} dS\, \vec{n} \cdot \vec{B} \tag{1.15}$$

Of course, the definition of flux function is automatically satisfied for a sufficiently regular magnetic field so that the toroidal and poloidal magnetic fluxes are flux functions and can be used to label the magnetic surfaces. The fundamental outcome of ideal MHD is that the magnetic flux enclosed by any arbitrary open surface moving with the plasma remains constant, implying that the flux is “frozen” into the plasma. This can be shown by evaluating the time derivative of the poloidal flux across a general surface S_p:

$$\frac{d\psi_p}{dt} = \int_{S_p} \frac{\partial \vec{B}}{\partial t} \cdot \vec{n} dS - \oint \vec{v} \times \vec{B} \cdot dl \tag{1.16}$$

Using Eq.1.10:

$$\frac{d\psi_p}{dt} = -\oint (\vec{E} + \vec{v} \times \vec{B} \cdot dl \tag{1.17}$$

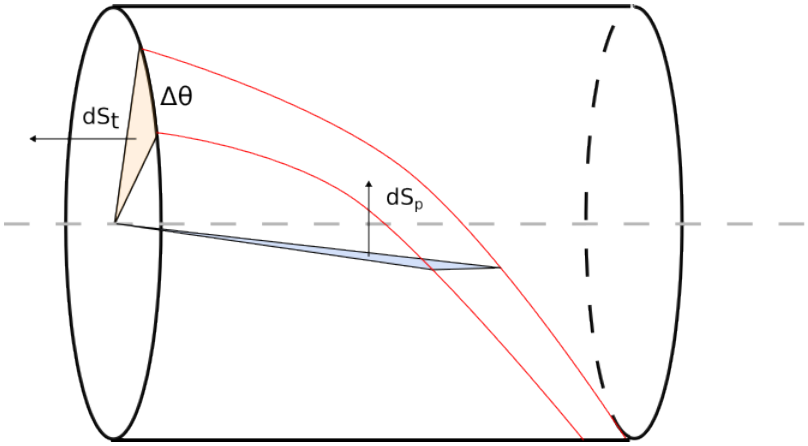


Figure 1.4: Illustration of the volume considered in [63] to provide a physical interpretation of the safety factor q. The red line represents a single field line at $\zeta = 0$ and the same field line at $\zeta = 2\pi$. After a toroidal turn, there is a difference $\Delta\theta$ in the field line in the poloidal plane. The flux across the surface S_p is equal to the flux across S_t, leading to the interpretation that the rotational transform indicates the winding step of the magnetic surface.

Eq. 1.17 is identically equal to zero, according to the ideal approximation of Ohm's law (Eq.1.9). Considering the proprieties of the flux label, for a fixed flux surface, it is possible to express the poloidal flux as a function of the toroidal flux:

$$\Psi_p = \Psi_p(\Psi_t) \tag{1.18}$$

and define the rotational transform as:

$$\iota = 2\pi \frac{d\Psi_p}{d\Psi_t} \tag{1.19}$$

The rotational transform defines the average $\Delta\theta$ in the poloidal plane between a magnetic field line at $\zeta = 0$ and $\zeta = 2\pi$, namely after a toroidal turn. In general, the $\Delta\theta$ that characterized the field line is not uniform, and its average gives a measure of the twist of the magnetic field line over the rational surface. A physical picture is provided in [63]. Choosing a magnetic surface ergodically covered by a magnetic field line. Following a field-line segment, it is possible to build a volume composed by the surface dS_t that covers the angle $\Delta\theta$ at a fixed ζ, and the surface dS_p at a fixed θ that intersects the field line as shown in Fig. 1.4. As no flux can flow across the surface by definition, the divergence-free condition for the magnetic field (that applies also to the flux for a sufficiently regular magnetic field) states that:

$$d\Psi_t = d\Psi_p \tag{1.20}$$

The volume, in general, can change as $\Delta\theta$ in principle is not uniform; however, every time the field line passes, the condition 1.20 holds. The field line will complete a certain number of toroidal and poloidal turns after one can say that the line closes on itself with a certain accuracy. Defining as m and n respectively the number of toroidal and poloidal circuits before the line closes on itself and iteratively increasing $\Delta\theta$ following the trajectory of the field line with the same procedure defined before, when the line closes on itself:

$$n\, d\Psi_t = m\, d\Psi_p \tag{1.21}$$

So that:

$$\frac{\iota}{2\pi} = \frac{d\Psi_p}{d\Psi_t} = \frac{n}{m} \tag{1.22}$$

gives the average pitch of the magnetic field lines of a magnetic surface. Of course, the rotational transform is a flux function. For historical reasons, the considered quantity is the inverse of the rotational transform:

$$q = \frac{2\pi}{\iota} = \frac{d\Psi_t}{d\Psi_p} = \frac{m}{n} \tag{1.23}$$

this quantity is called *safety factor*. In general, q can be a rational or irrational number. If q is irrational, the magnetic field line covers ergodically the magnetic surface (this is why, in the previously adopted procedure, the line was considered to close on itself, “within a certain accuracy”). On the other hand, if q is a rational number, the line closes on itself. Physically, the equilibrium of the magnetic surfaces is described by the ideal MHD:

$$\vec{J} \times \vec{B} = \nabla p \tag{1.24}$$

It is straightforward to show that the flux surfaces are also isobaric:

$$\vec{B} \cdot \nabla p = \vec{J} \times \vec{B} \cdot \vec{B} = 0 \tag{1.25}$$

that states that the pressure is a surface function.

1.2.1 MHD equilibrium: the Grad-Shafranov equation

The bidimensional shape of the flux surfaces assuming axial symmetry can be obtained by solving an elliptic partial differential equation, the Grad-Shafranov, whose solution provides an expression for the function $\psi(R,Z)$, where ψ represents a label function. Considering an axisymmetric system, described by cylindrical coordinates (R,Z,ϕ) $(\partial/\partial\phi = 0)$, the magnetic field can be written as a function ψ, defined as:

$$\vec{B} = B_\phi \hat{e}_\phi + \vec{B}_{pol} \tag{1.26}$$

where B_{pol}:

$$B_{pol} = \frac{1}{R} \nabla\psi \times \hat{e}_\phi \tag{1.27}$$

It can be shown easily that Eq. 1.26 satisfies the divergence-free condition $\nabla \cdot \vec{B} = 0$. Current density $\vec{J}$ is defined by the Ampere's law:

$$\mu_0 \vec{J} = \nabla \times \vec{B} \tag{1.28}$$

using Eq. 1.26, the current becomes:

$$\mu_0 J = -\frac{1}{R}\Delta^* \psi \hat{e}_\phi + \frac{1}{R}\nabla(RB_\phi) \times \hat{e}_\phi \tag{1.29}$$

$$\Delta^* \psi \equiv R^2 \nabla \cdot \left(\frac{\nabla \psi}{R^2}\right) = R\frac{\partial}{\partial R}\left(\frac{1}{R}\frac{\partial \psi}{\partial R}\right) + \frac{\partial^2 \psi}{\partial Z^2} \tag{1.30}$$

Putting Eqs. 1.26 and 1.29 in Eq. 1.28 and combining the solutions projected along the $\nabla\psi$, $\vec{B}$ and $\vec{J}$ directions:

$$\Delta^* \psi = -\mu_0 R^2 \frac{dp}{d\psi} - \frac{1}{2}\frac{dF^2}{d\psi} \tag{1.31}$$

where $F \equiv B_\phi R$. Eq. 1.31 is the Grad-Shafranov equation obtained in static axisymmetric equilibrium. The ideal MHD model cannot express the pressure profile $p(\psi)$ and the function $F(\psi)$ related to the toroidal magnetic field. For this reason, the Grad-Shafranov equation is considered an underdetermined equation that needs another model involving plasma transport to define these free functions. The Grad-Shafranov equation is obtained assuming a static equilibrium condition, namely, the time derivatives equal zero. However, the equilibrium in a plasma is unstable, and the plasma is subjected to instabilities that can break the confinement, leading to disruptions. In the next section, a brief introduction to plasma instabilities is presented.

1.3 Plasma stability

The MHD equilibrium of the plasma is obtained by balancing the pressure gradient and the $\vec{J} \times \vec{B}$ force. Despite the ideal MHD admits an equilibrium, namely $d/dt = 0$, it does not guarantee a stable equilibrium. A system is said to be in stable equilibrium if, for every perturbation, the system remains in a neighborhood of the equilibrium point. This is exemplified by a pendulum: when displaced from its equilibrium position, the pendulum begins to oscillate around that equilibrium point. In unstable systems, small deviations from equilibrium favour the onset of instabilities that can grow, diverging the system from its initial conditions. The plasma can become unstable because of the large gradients (current and pressure) typical of the tokamak equilibria, and the onset of instabilities is facilitated by the free energy available in the system. There are various instabilities, and this work will focus on those that can be described using the MHD model. A key distinction is based on the physical models required to explain these instabilities. Some are ideal magnetohydrodynamic instabilities, meaning they do not require resistivity for their modeling. Examples include the internal or external kink, which are particularly dangerous as they grow on an Alfvén timescale and can quickly destroy confinement. The inclusion of the resistivity breaks the frozen field line theorem and allows to account for phenomena that involve a diffusion of the magnetic field lines. This occurs in various resistive instabilities, which, despite their differing characteristics, share a common underlying mechanism: magnetic reconnection. Examples are the Sawtooth instabilities and the Edge Localized Mode (ELM) instabilities. This thesis will be focused on the (neoclassical) tearing mode (NTM) and in the next section, a brief physical, phenomenological description will be given.

1.3.1 Neoclassical tearing mode

The NTM instabilities arise at rational surfaces, where magnetic field lines, over resistive timescales, undergo magnetic reconnection through a diffusion process. This reconnection changes the topology of the magnetic surfaces, resulting in the formation of what is known as a magnetic island. The very first phase of the evolution of the magnetic island is linear and the stability of the rational surface is quantified by an index, the Δ_0', related to the logarithmic jump of the perturbed magnetic flux across the rational surface. In this context, the literature refers to the instability as Tearing Instability (TM). The resistive diffusion admits forming a new topology, called magnetic island. Beyond a certain width, the dynamic of the TM becomes non-linear, influenced by neoclassical effects such as the bootstrap current and the curvature of the magnetic field lines. These effects will be extensively analysed in Chapter 3. The nonlinearities give rise to characteristic behaviors in the tearing mode dynamics, such as the emergence of a threshold width required to trigger the mode and the establishment of a saturation width for the magnetic island. Nonlinearities also allow the trigger of the mode as a consequence of the interaction between different plasma instabilities, such as the sawtooth crash. The plasma within the magnetic island rapidly equilibrates, causing the temperature and the pressure profile to flatten. This enhances radial energy transport, reduces the plasma β, and weakens the confinement. If the magnetic island exceeds a certain threshold size, the plasma geometry becomes so distorted that confinement is significantly compromised. This could potentially cause the plasma to disrupt, unloading the energy on the first wall of the tokamak. This rapid event occurs on extremely short timescales and poses a serious threat, especially for the future generation of nuclear fusion reactors. One approach to NTM control is the injection of radio-frequency waves, particularly using Electron Cyclotron Current Drive (ECCD), which targets the resonant surface to compensate for the loss of bootstrap current within the island. This thesis will focus on the study of the $m = 2$, $n = 1$ NTM, the most alarming instability that can grow and lead to disruption. This thesis aims to investigate neoclassical tearing modes (NTMs) across different phases. First, the triggering of disruptive NTMs in JET will be analyzed, offering an interpretation based on the role of the ion polarization current in triggering the mode. Next, the evolution of the instability will be simulated in DTT scenarios using the integrated transport code JETTO-Jintrac, providing useful indication for the design of the stabilizing ECCD. Finally, innovative techniques will be applied to a neural network trained for disruption prediction, providing insights into the most significant physical quantities involved in the disruption prediction.

Chapter 2

DTT and JET

This thesis primarily relies on a simulated scenario performed on the Divertor Tokamak Test (DTT) facility, currently under construction in Italy, and experimental data from the Joint European Torus (JET) in England. This chapter will provide a brief description of these devices. The DTT has been designed as a machine dedicated to studying the divertor, a component for managing heat and impurities in the plasma. Its purpose is to tackle the challenges associated with heat dissipation in future reactors like ITER and DEMO. Specifically, the DTT aims to test advanced solutions for controlling plasma-wall interactions, ensuring efficient power flux management, and extending the operational life of wall materials. On the other hand, JET, which was the largest operational tokamak for many years, has been an important contributor to nuclear fusion research since the 1980s. Thanks to its design and capabilities, JET has served as a testing ground for many of the technologies and discoveries that are now driving the development of ITER. Among its contributions is its ability to operate with deuterium-tritium fuel mixtures, critical for replicating the conditions required for large-scale fusion reactions.

2.1 The Divertor Tokamak Test (DTT)

The Divertor Tokamak Test (DTT) project addresses one of the key challenges in fusion energy research: managing heat and particle fluxes in future reactors like ITER and DEMO [1]. As these reactors will operate at much higher power levels, ensuring efficient heat dissipation and maintaining the integrity of plasma-facing materials are critical for their success. In particular, DTT aims to develop advanced solutions for heat exhaust and plasma-wall interactions (PWI), essential for the long-term operation of fusion devices. In the framework of the European Fusion Roadmap [27] and the eight missions proposed to realize fusion energy, DTT specifically contributes to Mission 2, which focuses on developing a power exhaust system capable of sustaining the large heat loads expected in fusion reactors. The roadmap highlights the need to optimize a conventional divertor for detached conditions (such as in ITER) and to explore alternative, more advanced config-

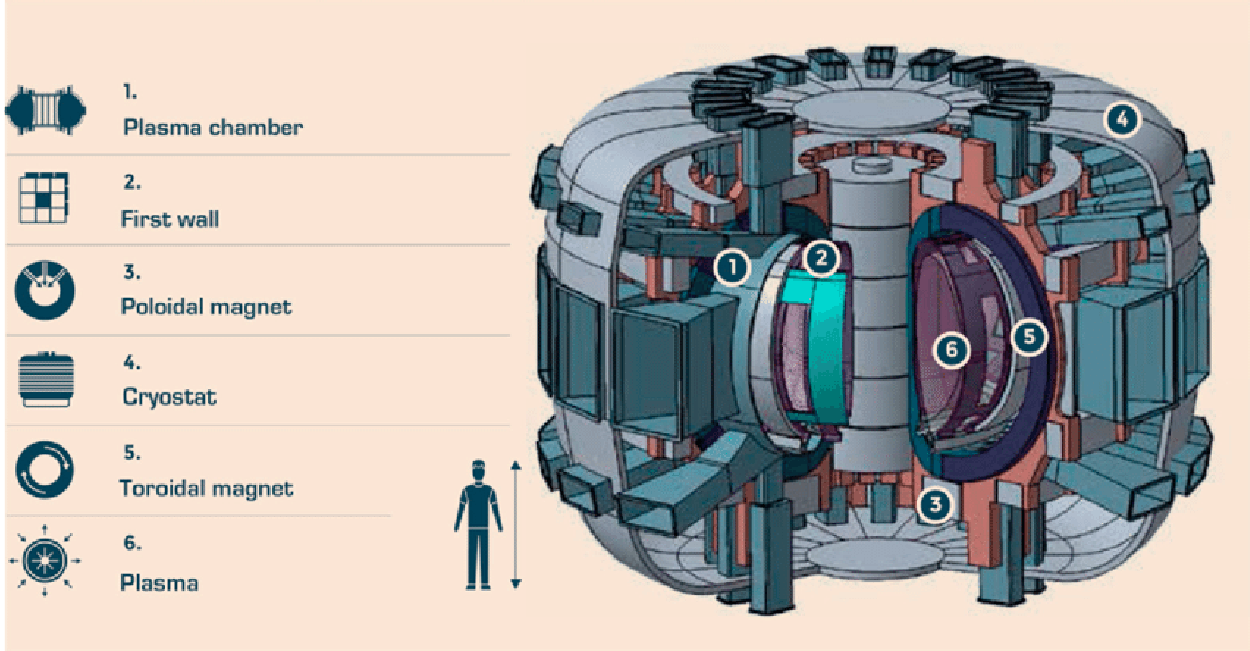


Figure 2.1: Illustration of DTT facility, showing its main components [87].

urations supporting DEMO design. To fill the gap between ITER and DEMO in terms of power handling capabilities, DTT is designed to test solutions that include advanced magnetic configurations, making it a highly flexible and integrated testbed for plasma exhaust research. The DTT's flexibility will allow it to test various divertor solutions while maintaining the ability to adapt to unforeseen technical challenges that might arise during the development of DEMO. In addition to testing conventional and alternative divertor designs, the DTT project aims to assess the performance of various plasma-facing materials, including liquid metals, to ensure the durability and efficiency of the divertor systems under extreme conditions. DTT is characterized by its flexible divertor design, which allows for testing both conventional and advanced divertor configurations. The primary goal

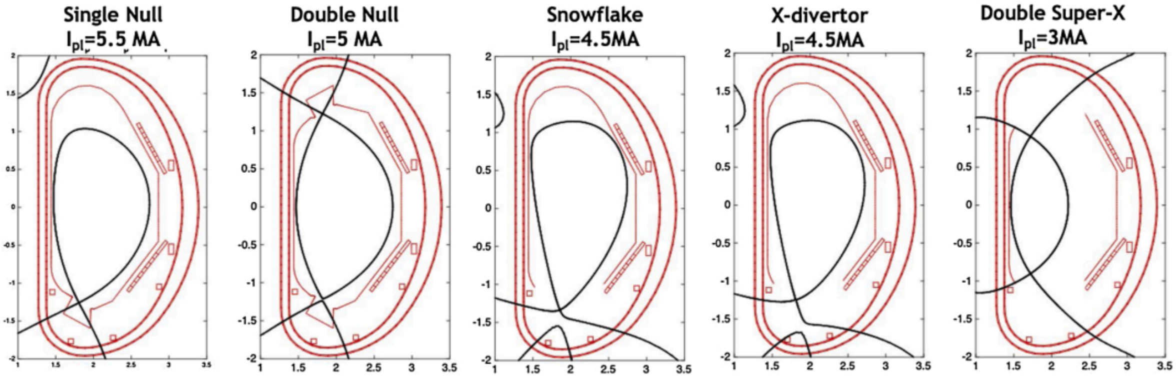


Figure 2.2: Divertor configurations available in DTT [1]

of the divertor system is to manage the heat and particle fluxes generated by the plasma, which in future fusion reactors are expected to exceed 10 MW/m^2. To address this, the DTT can accommodate a variety of divertor geometries, including [1] Single Null configuration, used in most tokamaks, X-Divertor, Snowflake Divertor and Negative triangularity. Various divertor configurations are shown in Fig. 2.2. In addition to these configurations, the DTT is designed to test liquid metal divertors [83], which represent a promising approach for handling heat loads. The liquid metal can absorb and dissipate heat more effectively than solid materials, offering a solution to one of the most important challenges in fusion reactor design. The DTT divertor system emphasizes flexibility and ease of replacement. The divertor modules can be changed rapidly through remote handling systems, ensuring that different configurations can be tested quickly, minimizing downtime between experiments. The DTT employs a combination of Electron Cyclotron Resonance Heating (ECRH), Ion Cyclotron Resonance Heating (ICRH), and Neutral Beam Injection (NBI) to achieve the necessary plasma heating [47]. Each system contributes to the total installed heating power, which in the final operational stage will reach 45 MW. The ECRH system comprises four clusters of eight gyrotrons, each operating at 170 GHz with a power output of 1 MW per gyrotron, providing 32 MW of heating power. The ICRH system delivers 8 MW of heating via two antenna modules, which operate in the frequency range of 60 to 90 MHz. Finally, the NBI system injects 10 MW of power through negative ion injection, operating at an energy level of 500 keV. The magnetic confinement system of the DTT is based on a combination of superconducting and normal conducting magnets, designed to provide control over the plasma and enable long-pulse operations. The magnetic confinement system in DTT includes several key components. The Toroidal Field (TF) coils consist of 18 superconducting Nb_3Sn coils, capable of generating a toroidal field of 6 T at the plasma major radius, which contributes to maintaining stable plasma confinement. The Central Solenoid (CS) is divided into six independently fed modules, each made from Nb_3Sn. The Poloidal Field (PF) coils system consists of six independent coils made from NbTi and Nb_3Sn conductors. These coils are responsible for shaping the plasma and controlling the plasma current during operation, which is critical in maintaining plasma stability. Additionally, two copper coils inside the vacuum vessel provide fast control over vertical instabilities, permitting real-time adjustments of the plasma position and improving the plasma's confinement and overall stability. DTT's operational

Scenario	**I_p (MA)**	**B_T (T)**	**Divertor**	**ECRH**	**ICRH**	**NBI**
A	2	3	SN/XD	8	0	0
B	2	3	SN/XD	16	4.75	0
C	4	6	SN/XD	16	4.75	0
D	5.5	6	SN	16	4.75	5-10
E	5.5	6	SN	32	9.5	10

Table 2.1: DTT Operational Scenarios

scenarios will evolve as the heating capabilities are incrementally enhanced, allowing for higher-power experiments over time. In the first phase (scenario A) of the machine's life, 8MW of ECRH will be used, focusing on reaching an H-mode baseline plasma. During this phase, DTT will explore the single null and X-Divertor configurations using a 3T toroidal magnetic field and 2MA plasma current. From phase B to phase D, the additional power will progressively increase, reaching the full power in scenario E. Tab. 2.1 shows the plasma current, toroidal magnetic field, and the strength of every additional heating system.

2.2 The Joint European Torus (JET)

The Joint European Torus (JET) [81], located in Culham, UK, started in 1970s as an European tokamak, with construction beginning in 1978. The device became operational in 1983 to investigate the behavior of plasma using deuterium-tritium fuel. JET demonstrated the feasibility of D-T fusion, achieving the first controlled D-T fusion reactions in 1991. This led to its 1997 record of 16 MW of fusion power output during a 1-second pulse. In 2022, JET set a new world record by producing 59 MJ of energy over five seconds, showing improvements in plasma confinement and heating technologies. JET's technical contributions include advances in plasma heating (via neutral beam injection and ion cyclotron resonance heating), plasma-wall interactions, and remote handling technologies. The main aims of the JET [24] are centered on advancing the study of plasma behavior in conditions that approximate those necessary for a functioning fusion reactor. This includes scaling and extending plasma parameters such as temperature, density, and confinement time to the levels required for sustained fusion reactions. Due to its iter-like wall, one main objective was to investigate impurities' effects on plasma stability and performance. Since impurities can lead to significant energy losses through radiation, understanding and mitigating these effects optimizes plasma confinement and overall efficiency. Additionally, JET aimed to develop supplementary heating techniques, including NBI and radiofrequency (RF) heating. These methods help achieve the high plasma temperatures needed for fusion, as ohmic heating alone is insufficient to reach interesting temperatures. Another important goal is to study the behavior of alpha particles generated during D-T fusion reactions. Understanding how these particles contribute to plasma heating and how they influence plasma stability is vital for future reactor designs. JET is characterized by a major torus radius of 2.96 meters and a minor radius of 1.25 meters, yielding a plasma volume of approximately 80 cubic meters. The device can produce a toroidal magnetic field of up to 4 T and a plasma current of up to 4.8 MA. JET is equipped with auxiliary heating systems, including 35 MW of NBI power and 6 MW of RF power for ICRH, which provides up to 41 MW of additional power. This allows to raise the plasma temperature to over 30 keV, placing the scenario in a thermonuclear fusion regime. JET can achieve a plasma density of around $1 \times 10^{20}\,\mathrm{m}^{-3}$, with the energy confinement time typically on the order of 1 second. JET is equipped with a comprehensive array of diagnostic tools, including hundreds of magnetic probes, interferometers, and spectrometers, which measure plasma parameters such as density, temperature, and magnetic field profiles. Advanced control systems monitor and stabilize the plasma, providing feedback control for plasma shape, position, and stability. These capabilities allow JET to investigate various plasma instabilities, such as magnetohydrodynamic (MHD) instabilities, sawtooth oscillations, and neoclassical tearing modes. JET allows for two different operational scenarios: the baseline and the hybrid. In the baseline scenario, the normalized β, denoted as β_N, typically ranges between 1.5 and 2. This scenario is designed around the standard inductive operation envisioned for ITER. However, baseline plasmas are limited by the onset of magnetohydrodynamic (MHD) instabilities, particularly neoclassical tearing modes (NTMs). In contrast, hybrid plasmas operate at higher β_N values, usually between 2.5 and 3, allow-

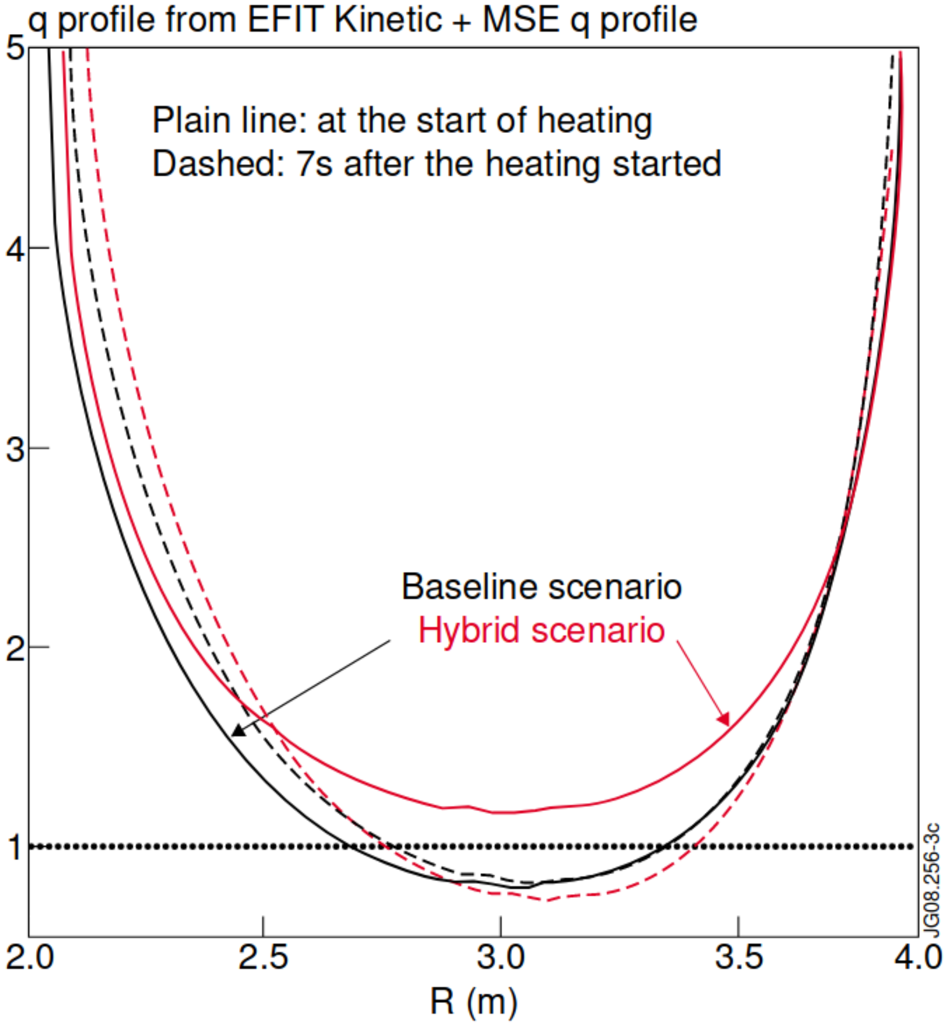


Figure 2.3: Example of q-profiles from jet pulse number 67934 (hybrid) and 67946 (baseline), at the start of the heating (solid line) and after 7s of heating, taken from Ref. [59]

ing for better confinement and improved stability. This higher pressure is possible due to a more stable current profile with lower magnetic shear, reducing MHD instabilities. The confinement factor, represented by H_{98}, is also significantly different between the two scenarios. In baseline plasmas, H_{98} is typically close to 1, providing adequate confinement, but it is limited in terms of overall plasma performance. On the other hand, hybrid plasmas achieve higher confinement with H_{98} values reaching up to 1.5, indicating improved energy retention and enhanced global confinement. This improvement is attributed to better core and pedestal confinement, particularly in high-pressure operations, where the plasma is maintained at lower density and higher temperatures. A significant difference between the two scenarios lies in their strategies for controlling plasma current and magnetic shear. In baseline plasmas, the current is higher, which results in a more peaked current profile. Hybrid plasmas operate at lower plasma currents, which broaden the current profile and lower the magnetic shear. Fig. 2.3 compares the hybrid and baseline scenarios. The hybrid and baseline scenarios are distinguished by controlling current penetration. Regarding MHD stability, the baseline scenario is more prone to developing NTMs and other core instabilities, which can limit the achievable β_N and degrade confinement. These instabilities are primarily caused by the peaked current profile in baseline plasmas, which leads to higher magnetic shear. In the hybrid scenario, these instabilities are minimized by the broader current profile and reduced magnetic shear. In the last experimental campaign, JET [70] achieved a new world record related to the produced energy due to D-T reaction. The time trace of the power gathered from the nuclear fusion reactions is shown in Fig.2.4. The pulse 99971 obtained 59 MJ of energy from D-T reactions with an average power of 10.1 MW. The ratio between energy produced and energy injected by additional heating is 0.33. The pulse used a tritium-rich plasma mixture where the deuterium abundance ranges between 5% to 17% with respect to the total D and T densities. The additional heating were pure deuterium neutral beam injection (D-NBI) and ICRH.

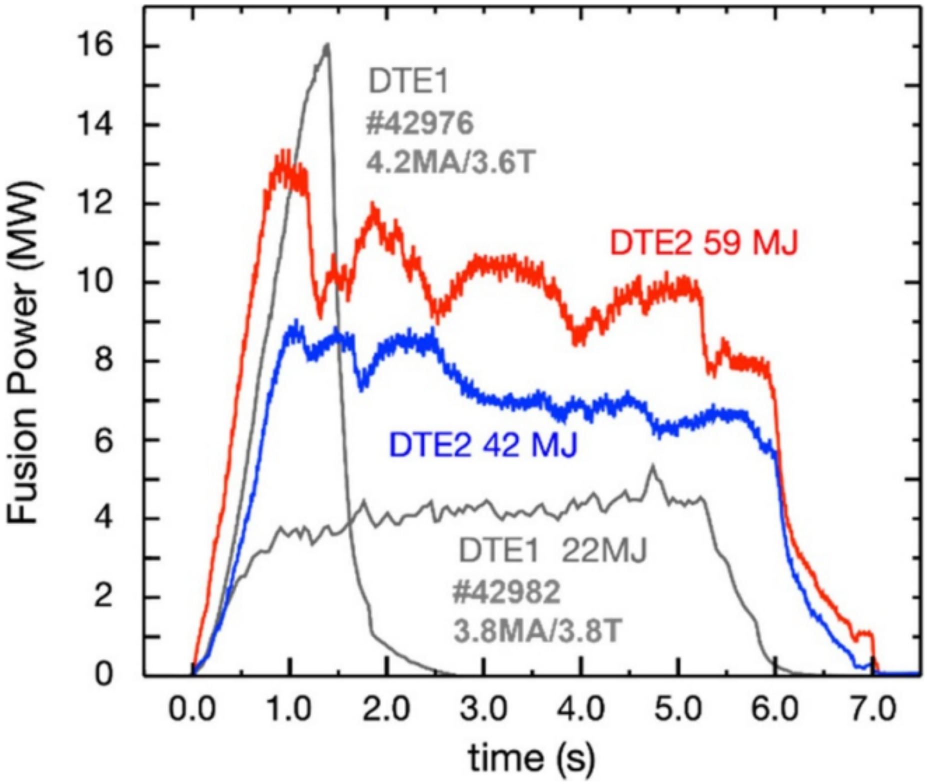


Figure 2.4: Time trace of the nuclear fusion power produced in the latest world record of nuclear fusion energy produced. Reproduced from [69]

Chapter 3

Theoretical framework

This chapter overviews the fundamental physical processes of classical and neoclassical tearing instability. The discussion covers all possible phases of the instability's life cycle, including onset, growth, and suppression. The analysis of the magnetic reconnection process has been approached considering the linear stability of a current sheet; Section 3 shows how such an ideal MHD approach is insufficient. In Section 3.0.3, the cylindrical tearing equation is introduced, with its derivation and both analytical and numerical solutions, to calculate the linear stability index Δ_0' [114]. Section 3.1 introduces the non-linear, neoclassical terms that drive the evolution of the tearing mode within the framework of Rutherford's theory. In this context, in Section 3.2, the stabilizing effects of current drive and heating, as consequences of electron cyclotron, are quantified. Section The linear theory: slab configuration Historically, [37], the challenge of magnetic reconnection has been approached by initially considering a simplified system: the slab configuration. This methodology offers valuable insights into the basic mechanisms involved in magnetic reconnection, aiming to model the magnetic configuration that leads to the trigger of a tearing mode. Considering an infinite plane current layers, the slab geometry is described by:

$$\vec{B}_0 = \hat{y}B_{y0}(x) + \hat{z}B_{z0}(x) \tag{3.1}$$

where B_{y0} and B_{z0} have the form:

$$B_{0y}(x) = \begin{cases} B_{0y}'x & -a < x < a \\ -B_{0y}'a & x < -a \\ B_{0y}'a & x > a \end{cases} \tag{3.2}$$

$$B_{0z}(x) = \text{constant} \tag{3.3}$$

The behavior of the electric and magnetic fields is defined by Maxwell's equations coupled with Ohm's law, where the ion pressure and inertia terms are neglected and quasi-

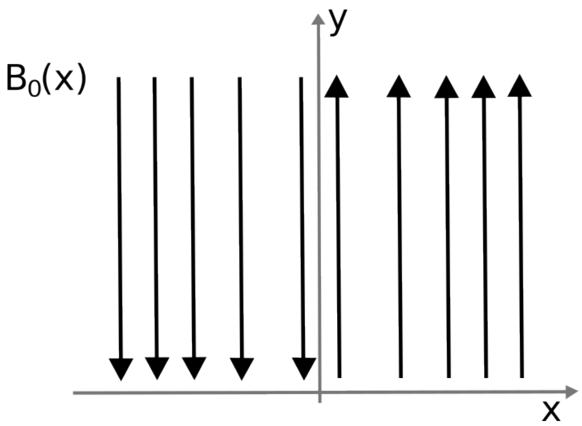


Figure 3.1: Representation of the magnetic field in slab configuration

neutrality is assumed (the charge density $\sigma = 0$):

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \tag{3.4}$$

$$\nabla \cdot \vec{E} = 0 \tag{3.5}$$

$$\nabla \times \vec{B} = \mu_0 \vec{J} \tag{3.6}$$

$$\nabla \cdot \vec{B} = 0 \tag{3.7}$$

$$\vec{E} + \vec{v} \times \vec{B} = \eta \vec{J} \tag{3.8}$$

Applying the curl to Ohm's law and using Equation 3.4 and 3.6, recalling that $\nabla \times \nabla \times \vec{B} = -\nabla^2 \vec{B}$ by the divergence-free property of the magnetic field:

$$\frac{\partial \vec{B}}{\partial t} = \nabla \times (\vec{v} \times \vec{B}) + \frac{\eta}{\mu_0} \nabla^2 \vec{B} \tag{3.9}$$

The equation can be simplified assuming an incompressible fluid, namely the time derivative of the density ρ in a certain volume, remains constant in time. According to the continuity equation 1.7, a null time derivative of the density implies a null divergence of the flow within a certain volume that, in the case of homogenous fluids, reduces to $\nabla \cdot \vec{v} = 0$. Eq. 3.9 becomes:

$$\frac{\partial \vec{B}}{\partial t} = \frac{\eta}{\mu_0} \nabla^2 \vec{B} + \left[(\vec{B} \cdot \nabla)\vec{v} - (\vec{v} \cdot \nabla)\vec{B} \right] \tag{3.10}$$

The physics of the equations used to derive Eq.3.10 describes the behaviour of the electric and magnetic fields. The MHD model must be employed to model the velocity $\vec{v}$. The curl of the momenta equation, neglecting the viscosity (Eq. 1.8) reads:

$$\nabla \times \left(\rho \frac{d\vec{v}}{dt} \right) = \nabla \times \left[\frac{1}{\mu_0} \nabla \times \vec{B} \times \vec{B} + \rho \vec{g} \right] \tag{3.11}$$

where ρ is the mass density, and $\rho \vec{g}$ is a term associated with a general acceleration, for example, the gravitational effects or the pressure gradients as a result of the curvature of the current layer. The total derivative of the resistivity η and the term $\rho\vec{g}$ is zero so that these functions are assumed to be modified only through convection:

$$\frac{\partial\eta}{\partial t} + \vec{v}\cdot\nabla\eta = 0 \tag{3.12}$$

$$\frac{\partial(\vec{g}\rho)}{\partial t} + (\vec{v}\cdot\nabla)(\vec{g}\rho) = 0 \tag{3.13}$$

Finally, the equilibrium is assumed to be static ($\vec{v}_0 = 0$). Eq. 3.10, assuming no change in times, reduces to:

$$\nabla\times(\eta_0\nabla\times\vec{B}_0) = 0 \tag{3.14}$$

Eqs. 3.10, 3.11, 3.12 and 3.13 are perturbed and linearized, obtaining dimensional equations describing the behavior of the perturbed magnetic field. By neglecting the resistivity, the ideal limit ($\eta = 0$) is investigated, underlining that the ideal MHD introduces a singularity that can be resolved only through the resistive MHD. The perturbed quantities are generally defined as:

$$f_1 = f_1(y)e^{i(k_y y + k_z z)+\omega t} \tag{3.15}$$

Perturbation is introduced in Eqs. 3.10, 3.11, 3.12 and 3.13, producing a linearized system of equations:

$$\omega\vec{B}_1 = \nabla\times(\vec{v}_1\times\vec{B}_0) - \frac{1}{\mu_0}\nabla\times\left[\eta_0\nabla\times\vec{B}_1 + \eta_1\nabla\times\vec{B}_0\right] \tag{3.16}$$

$$\omega\nabla\times\rho_0\vec{v}_1 = \nabla\times\left[\frac{1}{\mu_0}\left[(\vec{B}_0\cdot\nabla)\vec{B}_1 + (\vec{B}_1\cdot\nabla)\vec{B}_0\right] + (\vec{g}\rho)_1\right] \tag{3.17}$$

$$\nabla\cdot\vec{v}_1 = \nabla\cdot\vec{B}_1 = 0 \tag{3.18}$$

$$\omega\eta_1 + (\vec{v}_1\cdot\nabla)\eta_0 = 0 \tag{3.19}$$

$$\omega(\vec{g}\rho)_1 + (\vec{v}_1\cdot\nabla)(\vec{g}\rho)_0 = 0 \tag{3.20}$$

Considering a $\vec{v}_1 = (v_{1x}, 0, 0)$ and $\vec{B}_1 = (B_{1x}, 0, 0)$ with the spatial and temporal dependences defined by Eq. 3.15, the system of equations can be reduced to a set of two equations depending on the equilibrium and perturbed quantities v_{1x} and B_{1x}:

$$\omega B_{1x} = (B_{0y}k_y + B_{0z}k_z)v_{1x} - \frac{1}{\mu_0}\left((\nabla\eta_0\times(\nabla\times\vec{B}_1))_x + \eta_0\nabla^2 B_{1x} + \frac{1}{\omega}v_{1x}\frac{\partial\eta_0}{\partial x}\frac{\partial}{\partial x}(B_{0y}k_y + B_{0z}k_z)\right) \tag{3.21}$$

$$\omega\rho_0\vec{k}\times\vec{v}_1 = \frac{1}{\mu_0}\vec{k}\times\left[(B_{0y}k_y + B_{0z}k_z)\vec{B}_1 + B_{1x}\frac{\partial\vec{B}_0}{\partial x}\right] + \frac{1}{\omega}\vec{k}\times\left(v_{1x}\frac{\partial(\vec{g}\rho)_0}{\partial x}\right) \tag{3.22}$$

Following the notation introduced by Furth [37], Eqs. 3.21 and 3.22 are written in terms of adimensional quantities:

$$\begin{aligned}
\psi &= \frac{B_{1x}}{B} & W &= -iv_{1y}k\tau_r \\
F &= \frac{(k_yB_{0y}+k_zB_{0z})}{kB} & k &= \sqrt{k_y^2+k_z^2} \\
\alpha &= ka & \tau_r &= \frac{\mu_0 a^2}{<\eta>} \\
\tau_H &= \frac{a\sqrt{\mu_0 <\rho>}}{B} & S &= \frac{\tau_r}{\tau_H} \\
p &= \omega\tau_r & \bar{\eta} &= \frac{\eta_0}{<\eta>} \\
\bar{\rho} &= \frac{\rho_0}{<\rho>} & G &= \tau_H^2 A
\end{aligned}$$

where τ_H and τ_r represent respectively the Alfvenic and resistive times, a is the width of the current layer. The quantities $<\rho>$, $<\eta>$, and B refer to a measure of the density, the resistivity, and the magnetic field in the configuration. Eqs. 3.21 and 3.22 can be rewritten as:

$$\frac{\psi''}{\alpha^2} = \psi\left(1+\frac{p}{\bar{\eta}\alpha^2}\right)+\frac{W}{\alpha^2}\left(\frac{F}{\bar{\eta}}+\frac{\bar{\eta}'F'}{\bar{\eta}p}\right) \tag{3.23}$$

$$\frac{(\bar{\rho}W')'}{\alpha^2} = W\left[\bar{\rho}-\frac{S^2G}{p^2}+\frac{FS^2}{p}\left(\frac{F}{\bar{\eta}}+\frac{\bar{\eta}'F'}{\bar{\eta}p}\right)\right]+\psi S^2\left(\frac{F}{\bar{\eta}}-\frac{F''}{p}\right) \tag{3.24}$$

where the primes correspond to the derivation with respect to the adimensional quantity $\mu \equiv x/a$. The zero-order equilibrium condition 3.14 can be rewritten in terms of these quantities reads:

$$\frac{\bar{\eta}'}{\bar{\eta}} = -\frac{F''}{F'} \tag{3.25}$$

Using relation 3.25, Eqs. 3.24 and 3.23 can be written as:

$$pW\left(\frac{F}{\bar{\eta}}+\frac{\bar{\eta}'F'}{\bar{\eta}P}\right) = p\psi''-p\psi\left(\alpha^2+\frac{p}{\bar{\eta}}\right) \tag{3.26}$$

$$\frac{p^2}{\alpha^2S^2F}\left[(\bar{\rho}W')'+\alpha^2W\left(\frac{S^2G}{p^2}-\bar{\rho}\right)\right] = WF\left(PF'+\frac{\eta'}{\bar{\eta}}\frac{F'}{F}\right)+P\psi(PF-\bar{\eta}F') \tag{3.27}$$

The parenthesis in the left-hand side of Eq. 3.26 can be rewritten exploiting Eq. 3.24:

$$\left(\frac{F}{\bar{\eta}}+\frac{\bar{\eta}'F'}{\bar{\eta}p}\right) = \frac{p}{FS^2}\left(\frac{(\bar{\rho}W')'}{W\alpha^2}-\frac{\psi S^2}{W}\left(\frac{F}{\bar{\eta}}-\frac{F'}{p}\right)-\bar{\rho}+\frac{S^2G}{p^2}\right) \tag{3.28}$$

substituting in Eq. 3.26:

$$\frac{p^2}{\alpha^2S^2F}\left[(\bar{\rho}W')'+\alpha^2W\left(\frac{S^2G}{p^2}-\bar{\rho}\right)\right] = p\psi''-p\psi\left(\alpha^2+\frac{F'}{F}\right) \tag{3.29}$$

Then, for convenience, the quantity $<\eta>$ is taken to satisfy the relation $\bar{\eta}F' = 1$ and Eqs. 3.27 and 3.29 become:

$$\frac{p^2}{\alpha^2 S^2 F}\left[(\bar{\rho}W')' + \alpha^2 W\left(\frac{S^2 G}{p^2} - \bar{\rho}\right)\right] = p\psi'' - p\psi\left(\alpha^2 + \frac{F'}{F}\right) \tag{3.30}$$

$$\frac{p^2}{\alpha^2 S^2 F}\left[(\bar{\rho}W')' + \alpha^2 W\left(\frac{S^2 G}{p^2} - \bar{\rho}\right)\right] = (p\psi + WF)\left(pF' - \frac{F''}{F}\right) \tag{3.31}$$

Eqs. 3.30 and 3.31 describe the behaviour of v_{1y}, contained in W and B_{1y} contained in ψ. The infinite conductivity limit is investigated by taking $S \to \infty$. Eq. 3.31 becomes:

$$\frac{WG}{F} = (p\psi + WF)\left(pF' - \frac{F''}{F}\right), \tag{3.32}$$

that can be used to obtain an expression for W and ψ:

$$W = \frac{1}{F}\left(\frac{WG}{F}\frac{1}{\left(PF' - \frac{F''}{F}\right)} - p\psi\right) \tag{3.33}$$

$$\psi = \frac{1}{p}\left(\frac{WG}{F}\frac{1}{\left(PF' - \frac{F''}{F}\right)} - WF\right) \tag{3.34}$$

The shape of the magnetic field defined in Eq. 3.3 implies that $F' \to 0$ when $\mu \to \infty$, so the unique condition that prevents W and ψ to diverge is:

$$p\psi \simeq -WF \tag{3.35}$$

It is noteworthy that, considering a system where $B_{0z} = 0$, making explicit the dimensional quantities, the condition $p\psi \simeq -WF$ can be written as:

$$B_{1x} = i\xi_{1x}k_y B_{0y} \tag{3.36}$$

This relation between the perturbed magnetic field B_{1x} and the displacement ξ_{1x} can be read as the frozen-flux theorem. As described by Equation 3.36, the perturbed magnetic field $\tilde{B}_x$ is proportional to the displacement of the plasma column ξ_x, so that the magnetic field moves with the plasma and vice versa. Eq. 3.35 must hold in every point of the domain except a small region where $F = 0$, where the ideal MHD breaks and the resistivity effect becomes important. This region, which will be investigated in the next Section, is called the resistive layer.

3.0.1 Resistive Layer

The objective of investigating the physics of the resistive layer is to find a dispersion relation for the perturbation. The growth rate of the perturbation is shown to depend on an

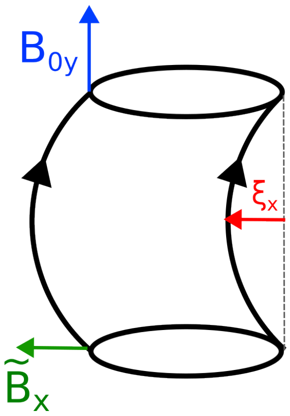


Figure 3.2: Illustration of Eq. 3.36. The picture reflects a well-known behavior in the framework of the ideal MHD: the frozen-flux theorem.

index Δ', which is connected to the jump of the gradient of the perturbed B_{1x} field. Initially, an expression for the width of the resistive layer is derived. A general perturbation in the plane $x-y$ is considered: $\vec{B} = \left[B_{1x}, B_{1y}, 0\right]$. By integrating the z-component of Amperes' law over the surface defined by the resistive layer, the following expression is obtained:

$$\int_S (\nabla \times \vec{B}_1)_z dS = \int_S \mu_0 j_{1z} dS \tag{3.37}$$

Considering that the current density j_{1z} is constant over the surface S, and $\partial/\partial x >> \partial/\partial y$ due to the thin resistive layer, applying Stoke's theorem to the left-hand side of the equation yields:

$$\oint_{\partial S} B_{1y} \cdot dl = \int_a^b B_{1y} dy + \int_c^d B_{1y} dy = \mu_0 j_{1z} L\delta \tag{3.38}$$

Equation 3.38 indicates that a discontinuity in the y-component of the perturbation of the magnetic field is associated with a current sheet j_{1z} flowing over the surface defined by the resistive layer. The relationship between j_{1z} and the discontinuity in B_{1y} can be determined by solving the integrals in Equation 3.38, assuming B_{1y} is constant along the integration path:

$$\lim_{\delta \to 0} \frac{B_{1y}(x+\delta) - B_{1y}(x-\delta)}{2\delta} = \mu_0 j_{1z} \tag{3.39}$$

Here, δ represents the half thickness of the resistive layer. Using the divergence-free property of the perturbed magnetic field, Eq. 3.39 can be expressed in terms of the discontinuity of the derivative of B_{1x} magnetic field:

$$\nabla \cdot \vec{B} = 0 = \frac{\partial B_{1x}}{\partial x} + kB_{1y} \tag{3.40}$$

So Equation 3.38 becomes:

$$\frac{1}{k\delta} \frac{\partial B_{1x}}{\partial x}\bigg|_{-\delta}^{\delta} = \mu_0 j_{1z} \tag{3.41}$$

Eq. 3.41 shows that there is a strict relation between the perturbed current flowing in the z-direction within the resistive layer and the jump of the derivative of the perturbed magnetic field B_{1x} across the resistive layer. This is closely related to the non-linear theory proposed by Rutherford and treated in Sec. 3.1. Equation 3.41 is written in terms of the resistive layer width δ. It is possible to obtain an expression for δ, examining the process from an energetics point of view [119]. Far from the rational surface, a displacement $\vec{\xi}$ of the plasma produces an electric field $\vec{E} = \vec{v}_1 \times \vec{B}_0$ with $\vec{v} = d\vec{\xi}/dt$. This induces a current $\vec{J} = \vec{v}_1 \times \vec{B}_0/\eta$ that causes a $\vec{J} \times \vec{B}_0$ restoring force:

$$\vec{F}_{J\times B} = \frac{1}{\eta}(\vec{v}_1 \times \vec{B}_0) \times \vec{B}_0. \tag{3.42}$$

where η is the resistivity. Close to the resistive layer, this corresponds to a force *external* to the resistive layer:

$$\vec{F}_{ext} = \frac{4}{\eta} v_{1x} (B'_{0y})^2 \delta^2 \hat{x} \tag{3.43}$$

The force defined in Eq. 3.43 corresponds to a power $P = F_{ext} v_{1x}$, and that must be equal to the rate of change of the kinetic energy within the layer. In principle, the kinetic energy depends on both the perturbed velocities:

$$\frac{d}{dt} E_k = \frac{d}{dt} \frac{1}{2} \rho |\vec{v}_1|^2 \tag{3.44}$$

however the divergence free condition imposes that $v_{1x} << v_{1y}$, because $\partial/\partial x \sim 1/\delta$:

$$\nabla \cdot \vec{v}_1 = \frac{1}{2\delta} v_{1x} + k v_{1y} = 0 \tag{3.45}$$

Recalling that the wavelength of the perturbation is much larger than the resistive layer width, we have that $\delta k << 1$ that implies $v_{1x} << v_{1y}$. In considering Eq. 3.44, we neglect v_{1x}:

$$\frac{d}{dt} E_k = \frac{d}{dt} \frac{1}{2} \rho v_{1y}^2 = \gamma \rho v_{1y}^2 \tag{3.46}$$

Eq. 3.46 can now be written in terms of v_{1x} by exploiting the divergence-free condition Eq. 3.45:

$$\frac{d}{dt} E_k = \gamma \rho \frac{v_{1x}^2}{(\delta k)^2} \tag{3.47}$$

and now Eq. 3.47 must be equated to the power flowing in the resistive layer:

$$P_{ext} = \vec{F}_{ext} \cdot \vec{v}_1 = \frac{4}{\eta} v_{1x}^2 (B'_{0y})^2 \delta^2 = \gamma \rho \frac{v_{1x}^2}{(2\delta k)^2} \tag{3.48}$$

obtaining an expression for the resistive layer width δ:

$$\delta = \frac{1}{2} \left(\frac{\eta \gamma \rho}{k^2 B'_{0y}} \right)^{1/4} \tag{3.49}$$

Using Faraday's law, we produce an equation relating the growth rate γ with the resistive layer δ:

$$j_z = \frac{1}{\eta k}\gamma B_{1x} \tag{3.50}$$

Using Eq. 3.41 can be used to produce an expression for γ:

$$\gamma = \frac{\eta}{2\mu_0\delta}\Delta' \tag{3.51}$$

where we have defined the quantity:

$$\Delta' = \frac{1}{B_{1x}}\frac{\partial B_{1x}}{\partial x}\bigg|_{-\delta/2}^{\delta/2} \tag{3.52}$$

Finally, putting Eq. 3.51 in Eq. 3.49 we obtain an expression for the resistive layer width:

$$\delta = \frac{1}{2}\left(\frac{\eta^2\Delta'\rho}{\mu_0 k^2 (B'_{0y})^2}\right)^{1/5} \tag{3.53}$$

The quantity Δ' is defined in terms of the perturbed magnetic field in the outer layer. The stability of the rational surface is dependent on the sign of this term. A positive Δ' corresponds to an unstable fluctuation, while the fluctuation is stable for $\Delta' < 0$. In general, as shown in Eq. 3.41 and 3.51, the effect of every phenomenon that contributes to change a current j_z can be expressed in terms of Δ', which modifies the stability of the rational surface. Regarding linear stability, the linear stability index Δ'_0 can be found by linearizing the ideal MHD equations in the outer layer and assessing the jump of the perturbed, linearized magnetic field across $x = 0$. In the next Section, the calculation for the derivation of the tearing equation in slab geometry will be carried out.

3.0.2 Tearing equation and linear stability in slab

The linear stability depends on the linear stability index, calculated as the logarithmic jump of B_{1x} evaluated on the boundary of the resistive layer. The so-called tearing equation can obtain an expression for the B_{1x}. The approximation obtained in the high conductivity limit (Eq.3.35) provides an expression for p:

$$p = -\frac{WF}{\psi} \tag{3.54}$$

that can be put in Eq. 3.30 in the limit of infinite conductivity $S \to \infty$:

$$-\frac{F}{\psi}\psi'' + F(\alpha^2 + \frac{F''}{F}) = \frac{WG}{F} \tag{3.55}$$

that is rearranged to obtain the *tearing equation*:

$$\psi'' - \psi\left(\alpha^2 + \frac{F''}{F} - \frac{G}{F^2}\right) = 0 \tag{3.56}$$

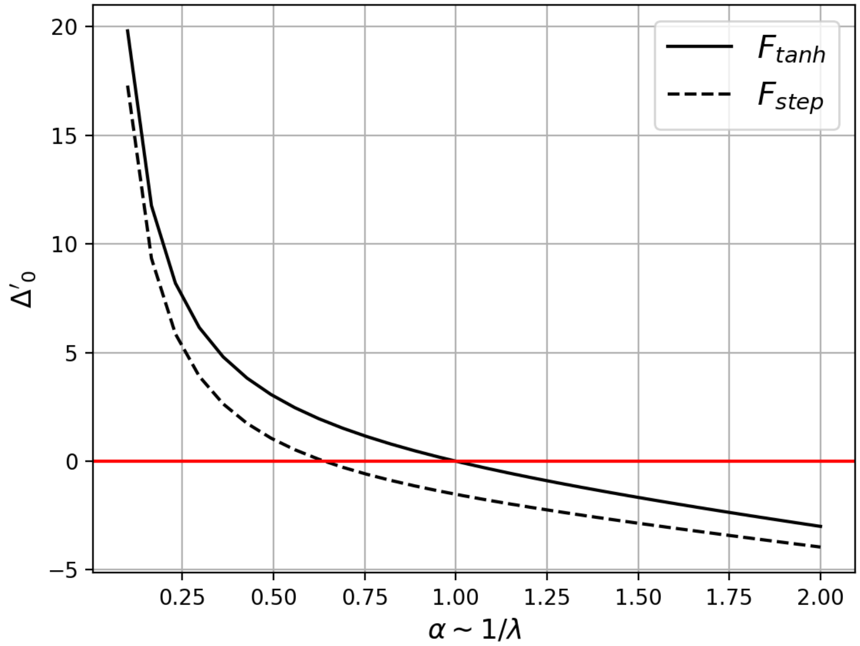


Figure 3.3: Plot of the analytical value of Δ'_0 calculated in slab geometry solving Eq. 3.56 for the cases illustrated in Eq. 3.66 and 3.67.

Eq. 3.56 is obtained by exploiting the infinite conductivity condition, which holds everywhere except for the resistive layer. Indeed, the function ψ, solution of 3.56, presents a discontinuity in $\mu = 0$, reflecting that ideal MHD breaks and resistive MHD analysis is needed. However, as shown in Sec. 3.0.1, the linear stability in $\mu \equiv x/\delta = 0$ (implicitly the resistive layer width δ is assumed to be equals to the current channel width a) depends on the parameter Δ'_0, which can be written in terms of ψ as :

$$\Delta'_0 = \frac{\psi'_2}{\psi_2} - \frac{\psi'_1}{\psi_1} \tag{3.57}$$

where $\psi'_1 = \psi'(\mu = -1)$ and $\psi'_2 = \psi(\mu = 1)$. Eq. 3.56 can be solved in the region $[-\infty, -1]$ and $[1, \infty]$ to obtain ψ'_1 and ψ'_2. In the limit $\alpha^2 >> 1$, the tearing equation reduces the Δ'_0 can be calculated exactly. The tearing equation reduces to:

$$\psi'' = \psi \alpha^2 \tag{3.58}$$

that can be integrated by imposing, as boundary conditions, that the functions ψ, ψ' and ψ'' go to zero exponentially as $\mu \to \infty$. To ensure this condition, the functions are multiplied by $e^{-\alpha|\mu|}$. The integration in the regions $[-\infty,-1]$ and $[1,\infty]$ reads:

$$\int_{-\infty}^{-1} e^{\alpha\mu}\psi'' d\mu = \alpha^2 \int_{-\infty}^{-1} e^{\alpha\mu}\psi d\mu \tag{3.59}$$

$$\int_{1}^{+\infty} e^{-\alpha\mu}\psi'' d\mu = \alpha^2 \int_{1}^{+\infty} e^{-\alpha\mu}\psi d\mu \tag{3.60}$$

Integrating by parts:

$$\psi' e^{\alpha\mu}\Big|_{-\infty}^{-1} - \alpha\int_{-\infty}^{-1} e^{\alpha\mu}\psi' d\mu = \alpha\left(e^{\alpha\mu}\psi\Big|_{-\infty}^{-1} - \int_{-\infty}^{-1} e^{\alpha\mu}\psi' d\mu\right) \tag{3.61}$$

$$\psi' e^{-\alpha\mu}\Big|_{1}^{+\infty} + \alpha\int_{1}^{+\infty} e^{-\alpha\mu}\psi' d\mu = \alpha\left(-e^{-\alpha\mu}\psi\Big|_{1}^{+\infty} + \int_{1}^{+\infty} e^{-\alpha\mu}\psi' d\mu\right) \tag{3.62}$$

The integrals cancel out obtaining:

$$-\frac{\psi_1'}{\psi_1} = -\alpha \tag{3.63}$$

$$\frac{\psi_2'}{\psi_2} = -\alpha \tag{3.64}$$

That corresponds to:

$$\Delta_0' = -2\alpha \tag{3.65}$$

In the case $\alpha^2 << 1$, the term F''/F cannot be neglected anymore, and an analytic expression for F must be specified to obtain an analytical expression for Δ_0'. Two cases can be investigated [37]:

$$F(\mu) = F_{tanh} = \tanh\mu \implies \Delta_0' = 2\left(\frac{1}{\alpha} - \alpha\right) \tag{3.66}$$

$$F(\mu) = F_{step} = \begin{cases} -1 & \mu < -1 \\ \mu & |\mu| < 1 \\ 1 & \mu > 1 \end{cases} \implies \Delta_0' = 2\alpha\left(\frac{(1-\alpha) - \alpha\tanh\alpha}{\alpha - (1-\alpha)\tanh\alpha}\right) \tag{3.67}$$

The values of Δ_0' for the two cases are plotted in Fig. 3.3. The behavior of Δ_0' is monotonic and decreases as α increases. Recalling that $\alpha = ka$, the current sheet becomes stable to small wavelength perturbations. In a heuristic picture, a sufficient large wavelength is necessary to bend the field line close to $\mu = 0$, allowing them to diffuse and reconnect, giving rise to the tearing instability.

3.0.3 Linear stability in a cylindrical tokamak

The equilibrium in slab configuration is represented by a magnetic field oriented solely along the y-direction, which varies linearly with x. In this simplified model, it has been

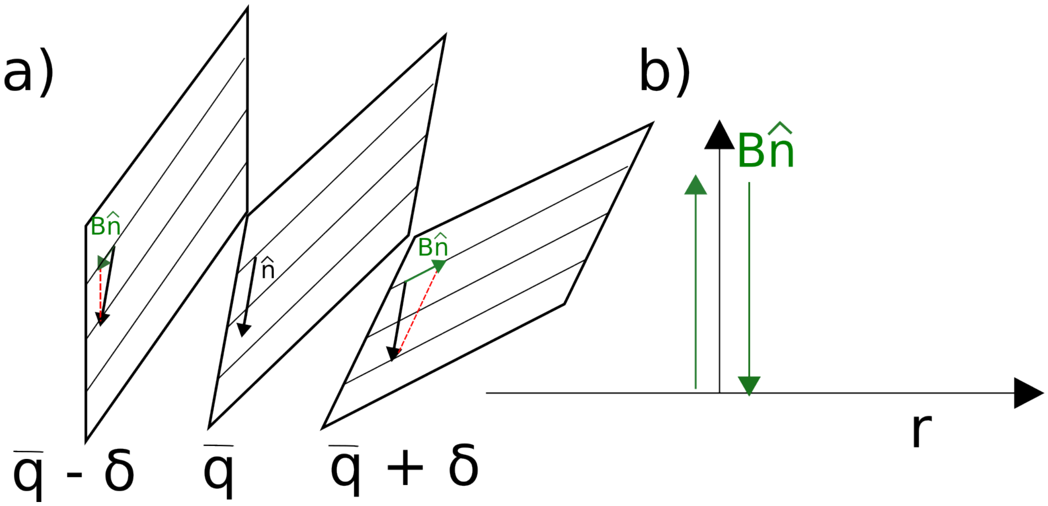


Figure 3.4: Illustration of how the slab configuration also occurs in a tokamak, where the plasma lies over closed nested magnetic surfaces with different winding steps.

shown that at the point where the magnetic field becomes null ($\mu = 0$), perturbations of the magnetic field can lead to the diffusion of the field line, causing magnetic reconnection depending on the sign of the linear stability index Δ_0'. In the context of a tokamak, the equilibrium is characterized by a combination of toroidal and poloidal magnetic fields, which produce closed nested magnetic surfaces with varying winding steps. These surfaces are described in a toroidal domain. Magnetic reconnection occurs when two oppositely directed magnetic field lines come into proximity, a configuration that, as seen in the slab approximation, also arises in tokamaks. The magnetic surface' winding steps, are defined by the safety factor q. Considering an increasing q profile, a magnetic surface described by $\bar{q}$ will be close to surfaces with $\bar{q}+\delta$ and $\bar{q}-\delta$ as depicted in Fig. 3.4(a). Then the projection of the magnetic field, along the direction $\hat{n}$, shown in Figure, perpendicular to the magnetic field lines in $\bar{q}$, will produce a magnetic field which is zero on $\bar{q}$, having opposite direction on $\bar{q}-\delta$ and $\bar{q}+\delta$. Fig.3.4(b) shows a picture of this magnetic field component, underscoring its similarity with the slab configuration we used in the previous Section. Now, the tearing equation in cylindrical geometry will be derived. Employing the cylindrical coordinates (r,θ,ϕ), where r is the radial coordinate, θ is the poloidal angular coordinate, and ϕ is the toroidal angular coordinate, the large aspect ratio approximation is exploited, assuming the ratio between the major radius R and the minor radius a to be large and using its inverse, ε, as a parameter describing the degree of smallness of a quantity. Physically, the condition $\varepsilon << 1$ means that the torus can locally be treated as cylindrical, so, from now on, the set of cylindrical coordinates (r,θ,z) will be used. Furthermore, axial symmetry is assumed so that the derivative with respect to the z direction $\partial_z = 0$. These approximations produce an ordering in the equilibrium and perturbed quantities:

$$B_\theta \sim \frac{1}{r}\frac{\partial B_z}{\partial r} \sim \varepsilon B_z \tag{3.68}$$

$$J_\theta \sim \varepsilon J_z \tag{3.69}$$

The first ordering reflects that the poloidal magnetic field and the radial variation of the toroidal magnetic field are smaller than the toroidal magnetic field. The second ordering highlights that the toroidal current primarily determines the equilibrium. It is noteworthy that a tokamak, by construction, cannot allow an equilibrium magnetic field or a current in the radial direction, so $B_r = J_r = 0$. We define an ordering also in the perturbed quantities:

$$B_{1z} \sim \varepsilon B_{1r} \sim \varepsilon B_{1\theta} \tag{3.70}$$

$$J_{1r} \sim J_{1\theta} \sim \varepsilon J_{1z} \tag{3.71}$$

So we will consider a perturbation lying mainly in the (r,θ) plane and a perturbed current that is consistent with Ampere's law $\nabla \times \vec{B}_1 = \mu_0 \vec{J}_1$. The MHD equilibrium in the outer region is described by:

$$\nabla p = \vec{J} \times \vec{B} \tag{3.72}$$

Following the same procedure of Sec. 3.0.2, we apply the curl and project along the z-axis, removing the pressure from the equation, we obtain:

$$(\vec{B}\cdot\nabla)J_z - (\vec{j}\cdot\nabla)B_z = 0 \tag{3.73}$$

The ordering defined before allows us to neglect the second term as $(\vec{B}\cdot\nabla)J_z >> (\vec{J}\cdot\nabla)B_z$ [112]. We can now perturb and linearize the equation:

$$(\vec{B}\cdot\nabla)J_{1z} + (\vec{B}_1\cdot\nabla)J_z = 0 \tag{3.74}$$

The scalar variable ψ is introduced:

$$\vec{B} = \hat{z} \times \nabla\psi \tag{3.75}$$

where $\hat{z}$ is the verse pointing on the z direction. The variable ψ is closely related to the poloidal magnetic flux, and this is a general definition that holds both for perturbed and equilibrium quantities. The Ampere's law can be rewritten in terms of ψ:

$$\nabla \times \vec{B} = \mu_0 \vec{J} = \nabla \times (\hat{z} \times \nabla\psi) \tag{3.76}$$

The right-hand side of the equation can be rewritten by exploiting the vector calculus identities:

$$\mu_0 \vec{J} = (\nabla^2\psi)\hat{z} - (\hat{z}\cdot\nabla)\nabla\psi \tag{3.77}$$

Now this expression can be used to rework the quantity J_{1z} contained in the first term of Eq. 3.74, assuming a perturbed quantity scaling as $e^{i(m\theta - n\frac{z}{R})}$:

$$(\vec{B}\cdot\nabla)J_{1z} = \frac{1}{\mu_0}\left(B_\theta\frac{\partial}{\partial\theta}+B_z\frac{\partial}{\partial z}\right)\left(\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial\psi_1}{\partial r}+\frac{1}{r^2}\frac{\partial^2\psi_1}{\partial\theta^2}\right) =$$
$$= \frac{1}{\mu_0}\left(B_\theta\frac{m}{r}-B_z\frac{n}{R}\right)\left(\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial\psi_1}{\partial r}+\frac{m^2}{r^2}\psi_1\right) \tag{3.78}$$

The scalar poloidal flux ψ is introduced in the second term of Eq. 3.74:

$$(\vec{B}_1\cdot\nabla)J_z = \left((\hat{z}\times\nabla\psi_1)_r\frac{\partial}{\partial r}\right)J_z = \left(-\frac{1}{r}\frac{\partial\psi_1}{\partial\theta}\frac{\partial}{\partial r}\right)J_z = -\frac{m}{r}\psi_1\frac{\partial J_z}{\partial r} \tag{3.79}$$

So, using Eq. 3.78 and 3.79, Eq. 3.74 becomes:

$$\frac{1}{\mu_0}\left(B_\theta\frac{m}{r}-B_z\frac{n}{R}\right)\left(\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial\psi_1}{\partial r}+\frac{m^2}{r^2}\psi_1\right)-\frac{m}{r}\psi_1\frac{\partial J_z}{\partial r}=0 \tag{3.80}$$

Recalling that the cylindrical definition of the safety factor profile is $q = rB_z/RB_\theta$, the tearing equation is obtained:

$$\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial\psi_1}{\partial r}+\frac{m^2}{r^2}\psi_1-\frac{\partial J_z}{\partial r}\frac{1}{\frac{B_\theta}{\mu_0}\left(1-\frac{nq}{m}\right)}\psi_1=0 \tag{3.81}$$

Eq. 3.81 presents a singularity on magnetic surfaces described by a safety factor profile $q_s(r_s)$ that can be written as the ratio between two integer numbers that match the poloidal and toroidal wave number of the perturbation. These surfaces are called rational surfaces and are intrinsically susceptible to the onset of the mode. The solution ψ_1 and its derivative is illustrated in Fig.3.5. The first derivative presents a singular behavior on the position of the rational surface r_s. The linear stability index Δ_0' is now defined in terms of the poloidal flux ψ_1 as the logarithmic jump of ψ_1 across the rational surface:

$$\Delta_0' = \frac{1}{\psi_1}\frac{\partial\psi_1}{\partial r}\bigg|_{r_s-\delta}^{r_s+\delta} \tag{3.82}$$

An estimation of the linear stability index Δ_0' can now be obtained by evaluating ψ_1 and its derivative close to the position of the rational surface r_s. The discontinuity makes dealing with the tearing equation difficult using numerical algorithms. In Ref. [74], a numerical method to solve the tearing equation, overcoming the limitation imposed by the discontinuity point, is proposed. The solving algorithm combines the 4-th order Runge Kutta method and a shooting algorithm. The solution's domain is divided into two parts, $[0, r_s-\delta)$ and $(r_s+\delta, 1]$. A solution is obtained in every domain imposing, as boundary conditions, $\psi_1(0) = 0$, $\psi_1(r_s-\delta) = \psi_1(r_s+\delta)$ and $\psi_1(1) = 0$. The limitation of the Runge-Kutta method is that the problem is posed as a Cauchy problem, specifying

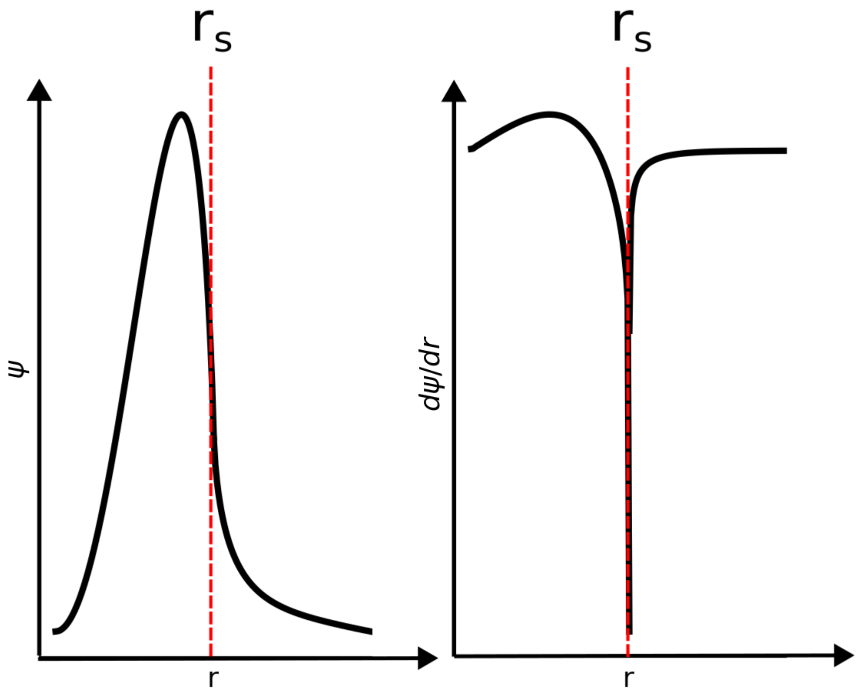


Figure 3.5: On the left the shape of the function $\tilde{\psi}$ obtained solving Eq. 3.81. On the right, the plot of the first derivative underscores the singular behavior on the rational surface r_s

the value of the function and its derivative at $x = 0$. In the context of the tearing equation, the boundary conditions are posed on the values of the solution at the boundaries of the regions, namely a Dirichlet problem. To overcome this issue, the shooting algorithm is employed. The first derivative of the solution $\psi_1'(0)$ is iteratively modified, minimizing the distance of the $\psi_1(r_s - \delta)$ and the desired value of the function at the boundary. Specifically, at every step, $\psi_1^{'(i)}(0)$ is used to calculate $\psi_1^{(i)}(r_s - \delta)$ using the Runge-Kutta algorithm. Then at the $i+1$-th step, ψ_1' is updated to minimize the function:

$$F(\psi_1'(0)) = \psi_1(r_s - \delta) - \psi_1^{(target)} \tag{3.83}$$

using a gradient descent method:

$$\psi_1^{'(i+1)} = \psi_1^{'(i)} - \alpha \frac{dF}{d\psi_1^{'(i)}} \tag{3.84}$$

where α is the step size (or learning rate in the context of machine learning). An illustration of how the algorithms work is shown in Fig. 3.6. The slope of the solution at the starting point is iteratively modified to reduce the distance of the solution's value to the target point. In this case, the algorithm can aim at the target point after three iterations. Combining these algorithms, the tearing equation can be solved in the first region, imposing $\psi_1(0) = 0$ and $\psi_1(r_s - \delta) = 1$ and in the second region, imposing $\psi_1(r_s + \delta) = 1$ and

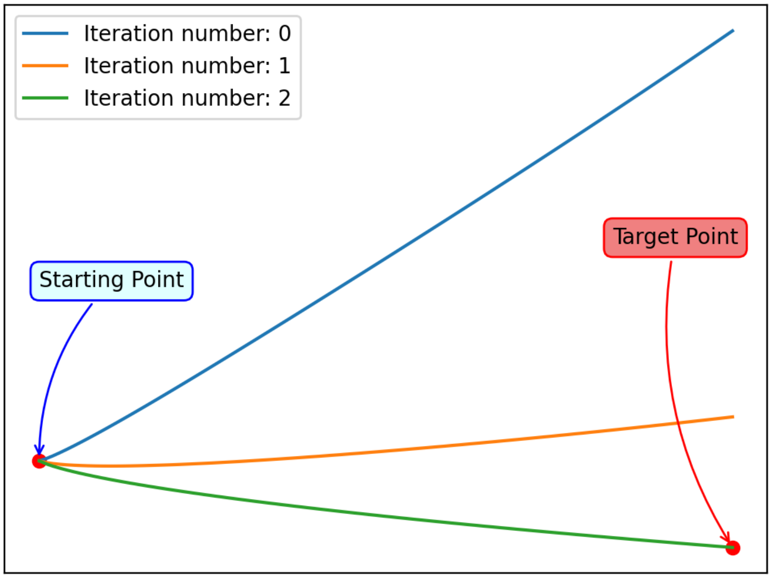


Figure 3.6: Illustration of the shooting method for solving the tearing equation posed as a Dirichlet problem. The equation is iteratively solved by adjusting the derivative of the solution at the starting point, reducing the distance to the target point.

$\psi_1(1) = 0$. The parameter δ accounts for the closeness the solution reaches, and in the current work is taken to be equal to 10^{-8}. The solutions in the two regions are matched in a neighborhood of r_s imposing that $\lim_{x \to r_s} \psi_1 = 1$, meaning that the solution $\psi_1(r)$ is normalized with respect to the value $\psi_1(r_s)$. This can be done without losing generality in the context of the calculation of Δ_0', as, by definition, the linear stability index is not affected by the normalization of the eigenfunction ψ_1. The nature of the tearing instability becomes very quickly non-linear, with a superimposition of multiple phenomena affecting the evolution of the mode, so a different treatment is needed. The next Section will present the non-linear growth of the tearing instability through the Rutherford theory.

3.1 Non Linear Theory

This Section explores the non-linear evolution of the magnetic island within the framework of Rutherford's theory [85]. The nonlinearity arises from the interactions, depending on the island's width, between the tearing mode and phenomena connected to the complex nature of the plasma. We neglected the interaction between the magnetic island and the conductors outside the plasma, so the nonlinearity is given by the interaction with the bootstrap current, the pressure gradients that arises due to the curvature of the magnetic field lines and the differences in the electron-ion inertia that generates the so-called ion polarization current. Considering nonlinearity allows for describing typical non-linear behaviours like saturation of the island width and a critical width beyond which the mode becomes unstable. First, the helical quantities required to model the magnetic island in a tokamak are introduced. Then, the relationship between the island's growth rate and the non-linear factors is established through the presentation of Rutherford's equation. Finally, the non-linear contributions are quantified.

3.1.1 Helical description of the magnetic island

The helical quantities necessary to describe the island in a helical coordinate system are introduced. For reference, the definition of the scalar magnetic flux ψ, defined for the first time in the previous Section, is recalled here:

$$\vec{B} = \hat{z} \times \nabla\psi \tag{3.85}$$

The scalar magnetic flux is strictly connected to the poloidal magnetic flux Ψ:

$$\Psi = \int_S B_\theta \cdot \hat{n} dS = \int_0^{2\pi} d\phi \int_0^r dr \frac{\partial\psi}{\partial\theta} = \frac{\psi}{2\pi} \tag{3.86}$$

In the helical description, it is convenient to introduce a new angular variable ζ, defined as:

$$\zeta = m\theta - n\phi - \int^t \omega(t')dt' \tag{3.87}$$

This will be considered an independent variable that aggregates the poloidal angle θ and the toroidal angle ϕ, considering also the poloidal mode number m and the toroidal mode number n. The integral represents the correction due to the rotation frequency $\omega(t)$. Applying the cylindrical approximation to a torus, the helical angular variable ζ is written in terms of (r,θ,z) as [35]:

$$\zeta = m\theta - n\frac{z}{R_0} - \int^t \omega(t')dt' \tag{3.88}$$

The helical magnetic flux can be defined in the helical coordinate system in terms of the helical angle ζ as:

$$\chi = -\int_r^{r_s} \left(1 - \frac{q}{q_s}\right) B_\theta dr' - \psi(r)\cos(m\zeta) \tag{3.89}$$

Close to the rational surface, the helical poloidal magnetic flux χ can be expanded as:

$$\chi = -\int_r^{r_s} \left(1 - \left(1 + \frac{x}{q_s}\frac{\partial q}{\partial r}\right)\right) B_\theta dr' - \psi(r)\cos(m\zeta) = -\int_0^x x\frac{1}{q_s}\frac{\partial q}{\partial r} B_\theta dx - \psi(r)\cos(m\zeta) \tag{3.90}$$

where $q(r) = q(r_s + x)$ with $x = r - r_s$ and $|x| << 1$. Considering the derivative of the safety factor q and the poloidal magnetic field nearly constant close to the rational surface, the helical flux becomes:

$$\chi = \frac{s_s}{r_s} B_\theta \frac{x^2}{2} - \bar{\psi} \cos(m\zeta) \tag{3.91}$$

with $s_s = (r(\partial q/\partial r)/q)_{r_s}$ the magnetic shear on the rational surface. In the last derivation, the so called constant-ψ approximation [37, 54] has been applied, considering the function ψ to be constant within the island:

$$\lim_{r \to r_s} \psi(r) = \psi(r_s) = \bar{\psi} \tag{3.92}$$

The transformation defined in Eq. 3.89 yields an expression for the flux χ (Eq. 3.91) close to the magnetic island, composed of two terms. The second one refers to the perturbation of the equilibrium produced by the magnetic island. The first one represents the magnetic field at the equilibrium. It is noteworthy that the equilibrium part of Eq. 3.91 has the same form of the equilibrium flux ψ defined in slab geometry. Indeed, taking Eq. 3.85 and applying $\times \hat{z}$ yields:

$$B'_{y0} \hat{x} = \frac{\partial \psi}{\partial x} \tag{3.93}$$

Considering the equilibrium configuration defined in Eq. 3.3, the flux function ψ becomes:

$$\psi = B'_{y0} \frac{x^2}{2} \tag{3.94}$$

which has the same form of the helical flux function χ. In this sense, the helical flux χ can be seen as the yield of a coordinate transformation that allows to transform from cylindrical/toroidal coordinates to a system of coordinates where the flux has the same form as the flux in slab geometry. Using the cylindrical definition of $q = \frac{rB_z}{R_0 B_\theta}$, Eq.3.91 becomes:

$$\chi = s_s \frac{B_z}{R_0 q_s} \frac{x^2}{2} - \bar{\psi} \cos(m\zeta) \tag{3.95}$$

Defining the island width [35, 53] as:

$$W = 4\sqrt{\frac{R_0 q_s}{B_z s_s} \bar{\psi}} \tag{3.96}$$

the helical poloidal magnetic flux can be written as:

$$\chi = 8 \frac{x^2}{W^2} \bar{\psi} - \bar{\psi} \cos(m\zeta) \tag{3.97}$$

Finally, it is convenient to define the normalized helical magnetic flux as $\Omega = \dfrac{\chi}{\bar{\psi}}$:

$$\Omega = 8 \frac{x^2}{W^2} - \cos(m\zeta) \tag{3.98}$$

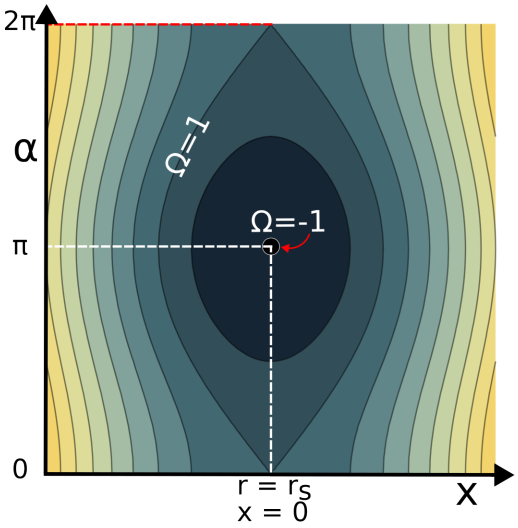


Figure 3.7: Illustration of the magnetic island in the helical plane defined by the helical angular coordinate α and $x = r - r_s$. The isoflux curves refer to the helical normalized flux Ω, defined in Eq. 3.98. $\Omega = -1$ refers to the O-point of the island, while $\Omega = 1$ correspond to the separatrix.

The island is described in terms of the helical coordinate system (x, Ω, ζ). Fig. 3.7 represents a plot showing the value of the variables in the helical coordinate system referred to the magnetic island. The normalized helical flux Ω is equal to -1 in the O-point and to 1 at the separatrix. The angular coordinate has a periodicity of 2π with $\zeta = 0$ and $\zeta = 2\pi$ corresponding to the x-point and $\zeta = \pi$ on the O-point.

3.1.2 The Generalized Rutherford Equation

The Generalized Rutherford Equation (GRE) describes the non-linear evolution of the tearing mode. The equation will be derived in the following, providing the general form of the equation without explicitly evaluating the contribution of the phenomena that affect the evolution of the island. The combination of the Ohm's law and the Faraday's law yields:

$$\frac{\partial \vec{B}}{\partial t} = \nabla \times (\vec{v} \times \vec{B} - \eta \vec{J}) \tag{3.99}$$

It is noteworthy that the electron inertia has been neglected. This assumption is invalid for $m = 1$ mode [115]. Writing the magnetic field $\vec{B}$ in terms of the scalar flux ψ defined in the Eq.3.85, Eq. 3.99 becomes: Putting 3.85 in 3.99:

$$\frac{\partial}{\partial t}(\hat{z} \times \nabla \psi) = \nabla \times (\vec{v} \times (\hat{z} \times \nabla \psi) - \eta \vec{J}) \tag{3.100}$$

that can be decomposed using the vector calculus identities:

$$\frac{\partial}{\partial t}(\hat{z} \times \nabla \psi) = \nabla \times (\hat{z}(\vec{v} \cdot \nabla \psi) - \nabla \psi(\vec{v} \cdot \hat{z}) - \eta \vec{J}) \tag{3.101}$$

Ignoring the curl of the gradient and recalling that for a generic function f and a generic vector $\vec{A}$:

$$\nabla \times (f\vec{A}) = f\nabla \times \vec{A} + \vec{A} \times \nabla f \tag{3.102}$$

given that the curl of a versor is null, the first term on the left-hand side can be rewritten as:

$$\hat{z} \times \nabla\psi = -\nabla \times (\psi\hat{z}) - \psi\nabla \times \hat{z} = -\nabla \times (\psi\hat{z}) \tag{3.103}$$

Eq. 3.101 becomes:

$$\nabla \times \frac{\partial}{\partial t}(\psi\hat{z}) = -\nabla \times (\hat{z}(\vec{v} \cdot \nabla\psi) + \eta\vec{J}) \tag{3.104}$$

So, projecting along the $\hat{z}$ axis we obtain:

$$\frac{\partial\psi}{\partial t} + (\vec{v} \cdot \nabla\psi) = \eta J_z \tag{3.105}$$

Introducing a perturbation such that $\psi = \psi_0 + \psi_1$, with the fluctuation developed in Fourier series as:

$$\psi_1 = \sum_{n=1}^{\infty} \tilde{\psi}_n \cos(nky) \tag{3.106}$$

It is possible to obtain an expression for the j_z in the rational surface by considering the diffusion Eq. 3.105 and averaging using the constant ψ approximation. It is noteworthy to say that in the derivation of the GRE, this assumption makes the mathematical model able to describe the physics within the island. Describing the magnetic island implies the constant-ψ approximation, and the constant-ψ approximation implies describing the island. This approximation is valid when $\psi >> \delta\psi$, i.e. when the magnitude of ψ is much larger than its change within the volume we are considering. This holds in a narrow layer close to the rational surface, namely the inner layer. The current density j_z can be expressed as (Eq. 11 of Ref. [85]):

$$j_z = j_{z0} + \frac{1}{\eta} \frac{\left\langle \frac{\partial\psi_1/\partial t}{[\psi-\psi_1]^{1/2}} \right\rangle_y}{\left\langle \frac{1}{[\psi-\psi_1]^{1/2}} \right\rangle_y} \tag{3.107}$$

At the interface of the rational surface, the model for the inner layer must be matched with the model for the outer layer, following the same procedure made for the linear theory. The Ampere's law describes the behavior of the perturbation ψ_1 for the flux:

$$\nabla^2\psi = \mu_0 j_z \tag{3.108}$$

Using equation 3.106, the Equation 3.108 becomes ($\nabla^2\psi = d^2\psi/dx^2$):

$$\frac{d\psi_1}{dx} = 4\pi \int_{-\infty}^{\infty} j_z dx \tag{3.109}$$

The fluctuation $\tilde{\psi}$ is expanded in a Fourier series, and same order component in Eq. 3.109 are matched:

$$\frac{d\tilde{\psi}_n}{dx}\cos(nky) = 4\pi \int_{-\infty}^{\infty} j_z dx \tag{3.110}$$

Multiplying for $\cos(nky)$ and introducing the average over the y direction, Eq. 3.110 becomes:

$$\frac{d\tilde{\psi}_n}{dx}\cos^2(nky) = 4\pi \int_{-\infty}^{\infty} \cos(nky)\, j_z dx \tag{3.111}$$

$$< f >_y = \frac{1}{2\pi}\int_0^{2\pi/k} fkdy \tag{3.112}$$

The average of the function $\cos^2(nky)$ can be exactly calculated:

$$< \cos^2(nky) > = \frac{1}{2\pi}\int_0^{2\pi/k} \cos^2(nky)kdy \tag{3.113}$$

$$= \frac{1}{2\pi}\int_0^{2\pi/k} \frac{1-\cos(2nky)}{2}kdy = \frac{1}{2} \tag{3.114}$$

By averaging along y direction, equation 3.111 becomes:

$$\frac{d\tilde{\psi}_n}{dx} = 8\pi \left\langle \cos(nky)\int_{-\infty}^{\infty} j_z dx \right\rangle_y \tag{3.115}$$

By introducing the quantity:

$$\Delta'_n = \left.\frac{\partial \ln \psi_n}{\partial x}\right|_{0-}^{0+} \tag{3.116}$$

Eq. 3.115 can be written as:

$$\Delta'_n \tilde{\psi}_n = 8\pi \left\langle \cos(nky)\int_{-\infty}^{\infty} j_z dx \right\rangle_y \tag{3.117}$$

An expression for j_z is provided by Eq. 3.107:

$$\Delta'_n \tilde{\psi}_n = \frac{16\pi}{\eta\sqrt{2B'_y}}\int_{\psi_{min}}^{+\infty} d\psi \left\langle \frac{\partial\psi_1/\partial t}{\sqrt{\psi-\psi_1}} \right\rangle_y \left\langle \frac{\cos nky}{\sqrt{\psi-\psi_1}} \right\rangle_y \frac{1}{< (\psi-\psi_1)^{-1/2} >_y} \tag{3.118}$$

Defining the rescaled variable $K \equiv \psi/\psi_1$ and writing $\psi_1 = \tilde{\psi}_n \cos nky$ (that corresponds to equate the same order terms of the Fourier expansion), Eq. 3.118 becomes a relation between the temporal derivative of $\tilde{\psi}_n$:

$$\Delta'_n \tilde{\psi}_n^{1/2} = \frac{16\pi C}{\eta(2B'_y)^{1/2}}\frac{\partial\tilde{\psi}_n}{\partial t} \tag{3.119}$$

with:

$$C = \int_{-1}^{\infty} dK \left\langle \frac{\cos ky}{(W-\cos ky)^{1/2}} \right\rangle^2 \frac{1}{< (W-\cos ky)^{-1/2} >} \simeq 0.7 \tag{3.120}$$

In general the island width scales as $W = k\psi_1^{1/2} = k\sum_n \tilde{\psi}_n^{1/2}$ where k is a proportionality constant. Eq. 3.119 can be read in terms of W:

$$D\frac{\partial W}{\partial t} = \sum_n \Delta'_n \tag{3.121}$$

with $D = \dfrac{32\pi C}{k\eta(2B'_y)^{1/2}}$. Eq.3.121, known as *Rutherford Equation*, is the key result of Rutherford's non-linear theory: the evolution of the island width depends on the sum of different phenomena, whose contribution to the growth rate is quantified in the index Δ'. The magnitude of the Δ'_n related to a certain physical phenomenon can be calculated by solving Eq. 3.117 and quantifying how much that phenomenon contributes to modifying the current j_z in the right-hand side of the equation. In its modern form, in a tokamak, Rutherford's Equation describing the free evolution of the tearing mode is commonly written as (neglecting the effect of the conducting vessel on the island evolution) [88]:

$$g_1\frac{\tau_r}{r_s}\frac{\partial W}{\partial t} = r_s(\Delta'_0 + a_c\Delta'_{bs} - a_c\Delta'_{ggj} + a_c\Delta'_{pol}) \tag{3.122}$$

where g_1 is a constant equal to 0.82, τ_r is the resistive time, r_s is the position of the rational surface. Δ'_0 is the linear stability index, calculated performing the linear analysis described in Sec. 3, Δ'_{bs} is the destabilizing effect given by the reduction of the bootstrap current within the island, Δ'_{ggj} (Green-Glasser-Johnson) is the stabilizing effect due to the curvature of the magnetic field lines and Δ'_{pol} is the effect of the ion polarization current, caused by the differences in the inertia of the electrons and ions. The coefficient a_c takes into account the degradation in the confinement time due to the growth of the island calculated through the belt model [20, 89]:

$$a_c = 1 - 4\left(\frac{r_s}{a}\right)^3\frac{W}{a} \tag{3.123}$$

where a is the minor radius, r_s the position of the rational surface, W the island width.

3.1.3 The bootstrap current: Δ'_{bs}

The bootstrap current is a self-generated current that flows in a tokamak due to the momentum transfer between the trapped and the passing particles. The interaction between the magnetic island and the bootstrap current reduces the bootstrap current within the island, which has a destabilizing effect. Initially, the underlying physics will be identified, followed by quantifying the index Δ'_{bs}. Taking a perturbed magnetic field $\vec{B}_1 = \hat{z}\times\nabla\psi_1$ with $\psi_1 = \tilde{\psi}_1\cos(m\zeta)$ with ζ the helical angle (Eq. 3.87), to account for the effect of the bootstrap current in the evolution of the mode, the current density j_z is written as [54]:

$$j_z = j_{ind} + j_{bs} = \frac{< E_\parallel >}{\eta} - \mu_e n_e m_e\frac{B_z R}{\psi'_s}\left\langle\frac{\partial p}{\partial x}\right\rangle \tag{3.124}$$

where j_{ind} represents the inductive part of the current, depending on the average of the parallel electric field on the rational surface. The term j_{bs}, representing the bootstrap current, depends on the electron viscosity μ_e, the electron density n_e, the electron mass m_e and the flux surface average of the gradient in $x = r - r_s$ of the pressure. The operator $< C >$ is the flux surface averaging operator, defined as:

$$< C >= \frac{\oint \frac{d\zeta}{2\pi} \frac{C(\Omega,\zeta)}{\sqrt{2\Omega + W^2 \cos(m\zeta)/8}}}{\oint \frac{d\zeta}{2\pi} \frac{1}{\sqrt{2\Omega + W^2 \cos(m\zeta)/8}}} \tag{3.125}$$

where Ω is the helical normalized flux defined in Eq. 3.98. Under the hypothesis that p is a flux function (i.e. $\vec{B} \cdot \nabla p = 0$), then: $< \partial p/\partial x >=< x > dp/d\Omega$. The modeling of the interaction between the bootstrap current and the island reduces to define the effect of the magnetic island on the gradient of the pressure profile within the island. Assuming that the dominant diffusion process is perpendicular to the magnetic field lines and that there is no source of pressure within the island, the pressure profile close to the rational surface can be written as:

$$\frac{dp}{d\Omega} = \left.\frac{dp}{dr}\right|_{eq} \frac{\Theta\left(\Omega - \frac{W^2}{16}\right)}{\oint \frac{d\zeta}{2\pi} \sqrt{2\Omega + W^2 \cos(\zeta)/8}} \tag{3.126}$$

where Θ is a step function and $dp/dr|_{eq}$ is the value of the gradient of the pressure profile at the equilibrium. This definition models the flattening of the pressure profile typical of the enhanced cross-field diffusion within the magnetic island. Due to the flattening of the pressure gradient profiles, the bootstrap current in the O-point is reduced. The bootstrap current flowing in the X-point generates a radial magnetic field $\delta B_r^{(bs)}$ that increases the radial magnetic field $B_{r1} = (\hat{z} \times \nabla \psi_1)_r$ consequence of the perturbation ψ_1, reinforcing the perturbation, providing an additional, neoclassical, drive for the instability. The index Δ'_{bs} can be calculated using Eq. 3.117, expressed in terms of the helical coordinates Ω and ζ and the average flux surface operator previously defined [46]:

$$\frac{q'_s}{q_s} \frac{B_\theta(r_s) W}{16\sqrt{2}} \Delta'_{bs} = - \int_{-1}^{\infty} d\Omega \oint \frac{d\zeta \, j_{bs} \cos(\zeta)}{\sqrt{2\Omega + W^2 \cos(m\zeta)/8}} \tag{3.127}$$

Exploiting the kinetic theory, the bootstrap current within the island can be written as:

$$j_{bs} = -\frac{1.46\sqrt{\varepsilon}}{B_\theta(r_s)\mu_0} \frac{\partial p}{\partial r} \tag{3.128}$$

with $\varepsilon = r/R$ the inverse aspect ratio. From Eqs. 3.127 and 3.128, the Δ'_{bs} can be calculated in two regimes, namely for small island and large island:

$$\Delta'_{bs} = \begin{cases} a_{bs} \frac{q_s}{q'_s} \beta_p \sqrt{\varepsilon} \frac{1}{T_e} \frac{\partial T_e}{\partial r} \frac{W}{W_d^2} & W < W_0 \\ a_{bs} \frac{q_s}{q'_s} \beta_p \sqrt{\varepsilon} \frac{1}{T_e} \frac{\partial T_e}{\partial r} \frac{1}{W} & W > W_0 \end{cases} \tag{3.129}$$

where β_p is the poloidal β and $a_{bs} = 3$ [91] is a constant. The term W_χ depends on the perpendicular transport coefficient $\chi_\perp$:

$$W_d = 3.16 q_s \sqrt{\frac{\sqrt{\chi_\perp \nu R}}{m q_s' v_{T_e}}} \tag{3.130}$$

A general expression, written in terms of the pressure profile, that asymptotically matches the two regimes is given by:

$$\Delta_{bs}' = a_{bs} \beta_p \left| \frac{L_q}{L_p} \right| \frac{W}{W^2 + W_d^2} \tag{3.131}$$

where the characteristic length $L_A = A/A'$ (the prime indicates the derivative with respect r) of the q and pressure profiles have been introduced.

3.1.4 The curvature: Δ_{GGJ}'

The effect of the curvature has been studied for the first time in the framework of the linear MHD theory by Green Glasser and Johnson [42]. Their theory highlights that the curvature's effect on the magnetic island's linear stability is stabilizing so that Δ_0' must exceed a critical value to trigger the instability. The critical width depends on the value of β, and for low β, the critical width approaches zero, and the mode is destabilized for $\Delta_0' > 0$. The linear theory of Green Glasser and Johnson can be extended through a quasi-linear model [62]. Here, a heuristic derivation will be presented. The detailed mathematical demonstration produces the same results. The slab configuration defined in Eq. 3.3 is considered. The magnetic flux ψ can be written as the sum of an equilibrium part (derived in Eq. 3.94) and a perturbed part due to the island:

$$\psi = B_{0y}' \frac{x^2}{2} + A \cos k_y y \tag{3.132}$$

where A is a function depending on the island width. In this heuristic representation, the effect of the curvature is modeled as an external acceleration $\vec{g}$ along the $\hat{x}$ direction. The equilibrium satisfies the equation:

$$\vec{J} \times \vec{B} = \rho \vec{g} \tag{3.133}$$

A thin magnetic island is considered to affect the equilibrium of the plasma. The Ampere's law reads:

$$\nabla^2 \psi = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \psi = J_z \tag{3.134}$$

the thin island approximation allows assuming that the equilibrium quantities close to the island vary more in the $\hat{x}$ direction than in $\hat{y}$, so that $\partial^2/\partial x^2 >> \partial^2/\partial y^2$. A Grad-Shafranov equation describing the shape of the flux surfaces close to the magnetic island can now be written:

$$\frac{\partial^2}{\partial x^2} \psi = F(\psi) + x g \frac{\partial \rho}{\partial \psi} \tag{3.135}$$

where $F(\psi)$ is a function that needs to be specified and represents the current on a flux surface. Averaging the current j_z on a flux surface, yields [62]:

$$\eta < J_z >= \left\langle \frac{\partial \psi}{\partial t} \right\rangle = \left\langle \frac{\partial A}{\partial t} \cos(k_y y) \right\rangle \tag{3.136}$$

where η is the resistivity. The time derivative of the first term of the magnetic flux ψ (Eq. 3.132) is null because it represents the equilibrium. The island's dynamic is described by Δ', which can be calculated from Eq. 3.117. In the notation adopted in this derivation, Eq. 3.117 can be rewritten as:

$$A\Delta' = \left\langle \int_{-\infty}^{+\infty} \delta j \cos(k_y y) dx \right\rangle \tag{3.137}$$

where δj is the difference between the equilibrium current (without magnetic island) and the perturbed current. The thin island assumption allows considering the relevant currents localized in the island region, so:

$$A\Delta' \simeq j_1 W \tag{3.138}$$

with $W \simeq \sqrt{A/B'_{0y}}$ the island width. The average over the flux surface of Eq. 3.135 produces :

$$< j_z >_y = F(\psi) + < x >_y g \frac{\partial \rho}{\partial \psi} \tag{3.139}$$

noting that, obviously, the average over the flux surface of a flux function is the function itself. Eq. 3.136 allows to find an expression for $F(\psi)$:

$$F(\psi) = \frac{1}{\eta} \left\langle \frac{\partial A}{\partial t} \cos(k_y y) \right\rangle_y - < x > g \frac{\partial \rho}{\partial \psi} \tag{3.140}$$

that can be put in Eq. 3.135 to obtain an expression for δj:

$$\delta j = \frac{1}{\eta} \frac{\partial A}{\partial t} + W g \frac{\partial \rho}{\partial \psi} \tag{3.141}$$

where $x - < x > \simeq W$ and $< \partial A \cos(k_y y)/\partial t > \simeq \partial A/\partial t$. The function $\partial \rho/\partial \psi$ can be written in the island region as:

$$\frac{\partial \rho}{\partial \psi} = \frac{\partial \rho/\partial x}{\partial \psi/\partial x} \simeq \frac{\rho}{W B'_{0y} L_\rho} \tag{3.142}$$

where L_ρ is the characteristic length of the density profile. Putting Eq. 3.142 in Eq. 3.141 and then Eq. 3.141 in Eq. 3.138:

$$\frac{A\Delta'}{W} \simeq \frac{1}{\eta} \frac{\partial A}{\partial t} + \rho g \frac{1}{L_n B'_{0y}} \tag{3.143}$$

Writing the constant A in terms of the island width W, Eq. 3.143 becomes:

$$\frac{1}{\eta}\frac{\partial W}{\partial t} \simeq \Delta_0' - \frac{\rho g}{W}\frac{L_s^2}{L_\rho B_{0y}^2} \tag{3.144}$$

where $B'_{0y} = B_{0y}/L_s$ with L_s the shear length. As previously pointed out, the curvature and the pressure effect is modeled by the ρg representing a generic acceleration along the $\hat{x}$ axis. In the framework of this heuristic derivation, in Eq. 3.144, the term ρg is replaced by the product κp with κ the curvature and p the pressure. Eq. 3.144 becomes:

$$\frac{1}{\eta}\frac{\partial W}{\partial t} \simeq \Delta_0' - \beta\frac{\kappa L_s^2}{L_\rho}\frac{1}{W} \tag{3.145}$$

A more detailed mathematical analysis produces an evolution equation that reads [62]:

$$\frac{dW}{dt} = \frac{1.22}{\tau_r}\left(\Delta_0' + a_{ggj}\left(1 - \frac{1}{q^2}\right)\beta_p\varepsilon^2\frac{L_q^2}{L_p r_s}\frac{1}{W}\right) \tag{3.146}$$

where τ_r is the resistive time, q is the safety factor profile, p is the pressure profile, $L_q \equiv q/q'$, $L_p \equiv p/p'$, r_s is the position of the rational surface, $\beta_p \equiv 2\mu_0 p/B_p^2$ is the poloidal β, ε^2 is the inverse aspect ratio and $a_{ggj} \simeq 6$ is a constant. Eq. 3.146 describes the evolution of the magnetic island in the non-linear phase, in the sense that the second term of the equation has been derived starting from an equilibrium describing a magnetic island with a certain width W. However, the stabilizing term introduced diverges in the limit of $W \to 0$, failing to describe the evolution of the island in its linear phase (when the island width W is comparable with the linear layer δ). This failure is caused by the fact that in the non-linear regime, the thermal transport within the island plays an important role in flattening the pressure profile, while in the linear regime, this phenomenon is not relevant. Modelling of the transition between linear and non-linear phases can be performed by introducing a characteristic diffusion length scale W_d defined as [68]:

$$W_d^4 = 64\frac{\chi_\perp}{\chi_\parallel}r_s^2\left(\frac{R}{ns}\right)^2 \tag{3.147}$$

where $\chi_\parallel$ and $\chi_\perp$ are respectively the parallel and perpendicular heat transport, $s = rq'/q$ is the magnetic shear at the position of the resonant surface r_s. Finally, the second term in Eq.3.146 is properly modified to take into account this transition as [68, 66]:

$$\Delta_{ggj} = a_{ggj}\left(1 - \frac{1}{q^2}\right)\beta_p\varepsilon^2\frac{L_q^2}{L_p r_s}\frac{1}{\sqrt{W^2 + 0.2W_d^2}} \tag{3.148}$$

3.1.5 The ion polarization current: Δ'_{pol}

The relative motion of the island with respect to the plasma frame of reference generates a current due to the different inertial responses of the ions and the electrons. This current, called ion polarization current, induces a parallel current (to preserve quasi-neutrality) that impacts the evolution of the island according to Eq. 3.117. The effect of the ion polarization current is quantified by the term Δ'_{pol} found in the Generalized Rutherford Equation. Smoliakov [98] identifies two regimes based on the ion Larmor radius: the small island ($W << \rho_i$) and large island ($W >> \rho_i$) regimes. The growth of the island, in the small island regime, is written as:

$$\Delta'_{pol} = G_2 \frac{2\tau}{1+\tau} \frac{r_s^2 L_s^2 (\omega - \omega_{*e})(\omega - \omega_{*i})}{m^2 d_i^2} \frac{1}{W} \tag{3.149}$$

where G_1 and G_2 are a constants equal to 0.39 and 0.38. $D_r = 1/(4\pi\sigma_{\|})$ where $\sigma_{\|}$ is the parallel conductivity, τ is defined as the ratio between the ion and electron temperatures. L_s is the magnetic shear length. The quantities ω, ω_{*e} and ω_{*i} refer respectively to the rotation of the island with respect to the plasma frame of reference, the electron diamagnetic frequency, and the ion diamagnetic frequency [17]:

$$\omega_{*j} = -\frac{m k_b T_j (d p_j / d\psi)}{e_j p_j q} \tag{3.150}$$

In the large island regime, the growth of the island reads:

$$\Delta'_{pol} = \frac{G_3 L_s^2 r_s^2}{m^2 W^3 v_a^2} \omega(\omega - \omega_{*pi}) \tag{3.151}$$

where $G_3 = 1.06$ is a constant [98], $\omega_{*pi} = \omega_{*i}(1+\eta_i)$ and $\eta_i = \partial \ln T_i / \partial \ln n$. The quantity $v_a = B^2/(4\pi n m_i)$ refers to the Alfven velocity. In this thesis, an expression that takes into account the transition between these two limits will be used [66]:

$$\Delta'_{\text{pol}}(W) = a_{\text{pol}} g(\varepsilon, \nu_{ii}) \beta_p \left(\frac{L_q}{L_p}\right)^2 \frac{\rho_{\theta i}^2 W}{W^4 + W_\rho^4} f(\omega), \tag{3.152}$$

$$f(\omega) = \frac{\omega(\omega - \omega_{*\text{i}} - \omega_{*\text{T}})}{\omega_{*e}^2}, \tag{3.153}$$

$$g = \begin{cases} \varepsilon^{3/2} \nu_{ii}^2/\omega^2 & \nu_{ii}/\varepsilon\omega < \varepsilon^{-3/4} \\ \varepsilon^{-3/4} & \varepsilon^{-3/4} < \nu_{ii}/\varepsilon\omega < \varepsilon^{-3/2} \\ 1 & \nu_{ii}/\varepsilon\omega > \varepsilon^{-3/2}, \end{cases} \tag{3.154}$$

where $\beta_p = 2\mu_0 p^2/B_\theta^2$ is the poloidal component of β, ε is the inverse aspect ratio, $L_{<*>} = \frac{<*>}{d<*>/dr}|_{r=r_s}$ is the characteristic length of the pressure p and the safety factor q profile. $\nu_{ii} = 4.810^{-8} Z^4 n_i \ln(\Lambda) T_i^{-3/2}$ is the ion-ion collision frequency, $\omega_{*T} = k_\theta (dT_i/dr)/(eB_0)$ takes into account the toroidal velocity of the island. Chapter 5 will provide a detailed explanation regarding the ion polarization current.

3.2 Island Suppression

As shown in the previous Section, the evolution of the magnetic island is mainly affected by the modification of the currents flowing parallel to the mode. The growth of a neoclassical tearing instability can lead to complete confinement loss. It is crucial to elaborate a technique to control and eventually suppress the magnetic island. The primary approach exploited to stabilize the mode is to replace the lack of bootstrap current caused by the flattening of the pressure profile within the island [19]. A current can be induced in the magnetic island through two physical mechanisms: a direct current driven by external sources and modifying the local value of the temperature that changes the local resistivity, inducing another current. In the following sections, the effect of the current drive and the heating will be quantified, considering the effects related to a misalignment with respect to the O-point and the modulation of the EC system.

3.2.1 The current drive: Δ'_{cd}

The effect of the current drive in the stabilization of the mode is quantified in the term Δ'_{cd}, which is calculated according to Rutherford's theory (Eq. 3.117) [65]:

$$\Delta'_{CD} = \frac{16\mu_0 L_q}{B_p \pi W^2} \int_{-\infty}^{+\infty} dx \oint d\zeta \, j_{CD} \cos(m\zeta) \tag{3.155}$$

Eq. 3.155 can be written in terms of the helical normalized flux Ω defined in Eq. 3.98:

$$x = \frac{W}{2\sqrt{2}}\sqrt{\Omega + \cos(m\zeta)} \qquad \frac{dx}{d\Omega} = \frac{W}{4\sqrt{2}} \frac{1}{\sqrt{\Omega + \cos(m\zeta)}} \tag{3.156}$$

$$\Delta'_{CD} = \frac{16\mu_0 L_q}{B_p \pi w^2} \int_{-\infty}^{+\infty} \frac{W}{4\sqrt{2}} \frac{1}{\sqrt{\Omega + \cos(m\zeta)}} d\Omega \oint d\zeta \, j_{CD} \cos(m\zeta) \tag{3.157}$$

The term j_{cd} can be written [65] as:

$$j_{cd} = 2\pi R \eta_{CD} \frac{< p_{EC} >}{< 1 >} \tag{3.158}$$

with $\eta_{CD} = I_{cd}/P_{tot}$ and the operator $< F >$ defined as:

$$< F(\sigma,\zeta,\Omega) > = \begin{cases} m \oint \frac{d\zeta}{2\pi} \frac{W}{4\sqrt{2}} \frac{F(\sigma,\zeta,\Omega)}{\sqrt{\Omega+\cos(m\zeta)}} & \Omega > 1 \\ m \int_{-\hat{\zeta}}^{\hat{\zeta}} \frac{d\zeta}{2\pi} \frac{W}{4\sqrt{2}} \frac{\frac{1}{2}[F(-\sigma,\zeta,\Omega)+F(\sigma,\zeta,\Omega)]}{\sqrt{\Omega+\cos(m\zeta)}} & \Omega \leq 1 \end{cases} \tag{3.159}$$

with $\sigma = \text{sign}(x)$ and $\hat{\zeta} = \arccos(-\Omega)/m$. The operator defined in Eq. 3.159 represents a flux surface average. Note that every quantity in the domain $\Omega \leq 1$ is averaged in the island topology. In this region, the operator is redefined as the mean between the value of the quantity F for a certain x and its symmetric with respect $x = 0$, corresponding to the vertical axis passing in the O-point, in Fig. 3.7. The term p_{EC} in Eq. 3.158 is defined as:

$$p_{EC} = P_{tot}\tilde{p} \tag{3.160}$$

with:

$$\tilde{p} = \frac{1}{2\pi^{5/2} w_{dep} R r_s} e^{-4(x-x_{dep})^2/W_{dep}^2} \tag{3.161}$$

here x_{dep} is the deposition location with respect to the rational surface. The coefficient W_{dep} is the full e^{-1} power density width. P_{tot} is the total injected power. In the definition 3.160 we neglected the term related to the modulation. A more complete definition is provided by De Lazzari [65]. Using Eqs. 3.158 and 3.159, Eq. 3.157 can now be rewritten as:

$$\Delta'_{CD} = \frac{16\mu_0 L_q}{B_p \pi W^2} \int_{-\infty}^{+\infty} 2\pi R \eta_{CD} \frac{< p_{EC} >}{< 1 >} < \cos(m\alpha) > d\Omega \tag{3.162}$$

Eq. 3.162 can be rearranged as:

$$r_s \Delta'_{CD} = \frac{16\mu_0 L_q}{B_p \pi} \frac{\eta_{CD} P_{tot}}{w_{dep}^2} F_{CD}(w^*, x_{dep}) \tag{3.163}$$

we have introduced the variable $w^* \equiv W/W_{dep}$ and the function $F_{CD}(w^*, x_{dep})$. Following De Lazzari[65], the function $F_{CD}(w^*, x_{dep})$ can be factorized into three figures of merit:

$$F_{CD}(w^*, x_{dep}) = N_{CD}(w^*) G_{CD}(w^*, x_{dep}) M_{CD}(w^*) \tag{3.164}$$

The function G_{CD} considers the misalignment of the current drive with respect to the rational surface, the function M_{CD} refers to the modulation of the current drive. De Lazzari et. al [65] writes the functions N_{CD} in terms of polynomial fraction that agrees with Sauter et. al [90]:

$$N_{CD}(w^*) = \frac{0.25 + 0.24 w^*}{1 + 0.64 w^{*3} + 0.42 w^{*2} + 1.5 w^*} \tag{3.165}$$

The effect of the misalignment is modeled as:

$$G_{CD}(w^*, x_{dep}) = 1 - \frac{x_{dep}}{g(w^*)} \int_0^{x_{dep}/g(w^*)} \mathrm{d}t\ \mathrm{e}^{(t^2 - (x_{dep}/g(w^*))^2)} \tag{3.166}$$

$$g(w^*) = \frac{0.38 w^{*2} + 0.26 w^* + 0.5}{w^* + 1} \tag{3.167}$$

while the function $M_{CD}(w^*)$ taking into account the modulation, is written as:

$$M_{CD}(w^*, \mathscr{D}) = \frac{1}{w^{*3}} \left(m_1(\mathscr{D}) w^{*2} + m_2(\mathscr{D}) \right) + m_3(\mathscr{D}), \tag{3.168}$$

with:

$$m_1(\mathscr{D}) = 2.26\mathscr{D}^4 - 3.44\mathscr{D}^3 - 0.99\mathscr{D}^2 + 2.2\mathscr{D} - 0.02, \tag{3.169}$$

$$m_2(\mathscr{D}) = 10^{-2}(0.34\mathscr{D}^5 - 1.02\mathscr{D}^4 + 0.87\mathscr{D}^3 - 0.28\mathscr{D}^2 + 0.1\mathscr{D}), \tag{3.170}$$

$$m_3(\mathscr{D}) = (1.34\mathscr{D}^4 - 3.54\mathscr{D}^3 + 1.1\mathscr{D}^2 + 2.09\mathscr{D} + 0.01). \tag{3.171}$$

where $\mathscr{D}$ represents the power on-time fraction. Summarizing, Eqs. 3.163-3.168 describe the stabilizing effect of the current drive in the Generalized Rutherford Equation, also considering the effect of the modulation and the misalignment. In this thesis, an analytical expression derived in Ref. [61] will be used:

$$G_{CD}(w^*, x_{dep}) = (1+0.6)\frac{1-\tanh\left(\frac{0.75\frac{x_{dep}}{W_{max}}-0.3}{0.2}\right)}{1-\tanh(1.5)+2\left(\frac{x_{dep}}{W_{max}}\right)^3} - 0.6e^{\left(-\left(\frac{x_{dep}}{W_{max}}\right)^2\right)} \tag{3.172}$$

The expression for Δ'_{CD} becomes:

$$\Delta'_{CD} = 16\mu_0 \frac{L_q P_{ec} I_{cd}}{\pi B_{pol} r_s} \frac{1}{w_{cd}^2} G_{CD}(w^*, x_{dep}) M_{CD}(w^*) \tag{3.173}$$

with I_{cd} the driven current, P_{ec} the injected power and $W_{max} = \max(W, W_{cd})$. The function for the modulation $M_{CD}(w^*, x_{dep})$ is formulated to consider a continuous wave injection (CW) or a 50% on - 50% off modulated power (50-50) [90]:

$$M_{CD}^{(CW)}(w^*) = \frac{0.25}{(1+2(w^*)^{2/3})} \tag{3.174}$$

$$M_{CD}^{(50-50)}(w^*) = \frac{0.45\tanh\left(\frac{w^*}{2.5}\right)}{(w^*)^2} \tag{3.175}$$

In Chapter 4, the expression 3.173 will be exploited to simulate the stabilization of NTM starting from the equilibrium scenario of DTT.

3.2.2 The heating: Δ'_h

The effect of the heating, in terms of the evolution of the NTM, is to change locally the value of the resistivity. This contributes to drive a current parallel to the island and modifying the growth rate of the island according to Eq. 3.117. The strength of the effect of the heating is quantified in the index Δ'_H defined as [65][55]:

$$r_s \Delta'_H = \frac{16\mu_0 L_q r_s}{B_p \pi W^2} \int_{-\infty}^{\infty} \mathrm{d}x \oint \mathrm{d}\xi \, j_H \cos(m\xi), \tag{3.176}$$

where the term j_H is the current induced by the change in the resistivity within the island defined as:

$$j_H = \frac{3}{2}\frac{\delta T_e}{T_{e0}} j_{\parallel 0} \tag{3.177}$$

where $j_{\parallel 0} = E_{\parallel}/\eta_0$ is the current induced by the loop voltage and η_0 the resistivity at T_{e0}. The term δT_e represents the temperature fluctuation due to the external heating system. It is worth noting that, thanks to symmetry-related arguments, we can focus solely on the temperature fluctuations within the separatrix [55]. Modified temperature can be found using a diffusion model for the temperature:

$$\nabla n \chi_{\perp} \nabla \delta T_e = -S \tag{3.178}$$

where n is the density, $\chi_\perp$ and S is the source of energy. From Eq. 3.178 an expression for δT_e is obtained [65]:

$$\delta T_e = \frac{P_{\text{tot}} w}{8\pi^2 R r_s \chi_\perp n_e k_B} \delta \tilde{T}_e, \tag{3.179}$$

with

$$\delta \tilde{T}_e \equiv \int_\Omega^1 \mathrm{d}\Omega \frac{\tilde{P}}{\langle |\nabla \Omega|^2 \rangle} \frac{8\pi^2 R r_s}{w}, \qquad \tilde{P} = \frac{P(\Omega)}{P_{\text{tot}}}. \tag{3.180}$$

Substituting Eq.3.179 in Eq. 3.176 an expression for Δ_H is obtained:

$$r_s \Delta'_H \approx \frac{16\mu_0 L_q \eta_H P_{\text{tot}}}{B_p \pi w^2_{\text{dep}}} F_H(w^*, x_{\text{dep}}, \mathscr{D}) \tag{3.181}$$

with

$$F_H(w^*, x_{\text{dep}}, \mathscr{D}) = \frac{1}{2\pi W} \int_{-1}^1 \mathrm{d}\Omega\, \delta \tilde{T}_e \int \mathrm{d}\xi \frac{W}{4\sqrt{2}\sqrt{\Omega + \cos(m\zeta)}} \cos(m\zeta) \tag{3.182}$$

where $x_{dep} = r - r_{\text{dep}}$ with r_{dep} the position of the deposition, $\mathscr{D}$ the power on-time fraction. W_{dep} the beam width, η_H represents the efficiency at which the power is converted into perturbative inductive current written as[8]:

$$\eta_H = \frac{3 w^2_{\text{dep}}}{8\pi R n_e \chi_\perp k_B} \frac{j_{\text{sep}}}{T_{\text{sep}}}. \tag{3.183}$$

Eq. 3.181 has the same form as Eq. 3.163 with the function $F_H(w^*, x_{\text{dep}}, \mathscr{D})$ that can be factorized in three functions:

$$F_H(w^*, x_{\text{dep}}, \mathscr{D}) = N_H(w^*) G_H(w^*, x_{\text{dep}}) M_H(w^*, \mathscr{D}). \tag{3.184}$$

with $N_H(w^*)$ representing the normalization to the geometrical function, $G_H(w^*, x_{\text{dep}})$ representing the misalignment and $M_H(w^*, \mathscr{D})$ the modulation of the heating source:

$$N_H(w^*) = \frac{0.077 w^{*2} + 0.088 w^*}{w^{*2} + 0.8 w^* + 2.17}, \tag{3.185}$$

$$G_H(w^*, x_{\text{dep}}) = \exp\left(- \left(\frac{x_{\text{dep}}}{g(w^*)} \right)^2 \right) \tag{3.186}$$

$$g(w^*) = 0.00035 w^{*4} - 0.008 w^{*3} + 0.07 w^{*2} + 0.02 w^* + 0.5 \tag{3.187}$$

$$M_H(D) = 1.2D^3 - 3.5D^2 + 3.3D - 0.06. \tag{3.188}$$

Finally, an expression for Δ'_h is given by:

$$\Delta'_h = \frac{16\mu_0 P_{ec} L_q}{8\pi^2 B_{pol} r_s} \frac{3 j_{sep}}{R_0 n_e(r_s) \chi_\perp T_{sep}} G_H(w^*, x_{\text{dep}}) M_H(D) \tag{3.189}$$

here $D = 1$ in continuous injection and $D = 0.5$ in 50% on - 50% off modulation.

Chapter 4

Modelling of NTM in DTT

The Divertor Tokamak Test (DTT) is an experimental tokamak designed to explore advanced divertor configurations for the future DEMO reactor. Its primary aim is to address heat and particle exhaust challenges in fusion devices. The DTT operates at high power to simulate conditions expected in larger fusion reactors, providing critical insights into plasma behavior and divertor design. This work focuses on the detailed modeling of NTMs within DTT full-power scenarios to prevent disruptions and ensure the stability of the plasma during operation. The NTM modeling in the full-power scenario design allows for a comprehensive understanding of how these instabilities evolve under high-performance conditions in DTT.

4.1 JETTO-Jintrac

The simulations here presented were carried out by integrating a module solving the Generalized Rutherford Equation, described in Section 3.1.2, in the integrated modeling framework JETTO-Jintrac, which is a comprehensive simulation toolset designed for modeling a wide range of plasma scenarios within tokamak devices. It is made up of several key modules, each tailored to specific aspects of tokamak physics and control. The primary modules within JETTO-Jintrac include an equilibrium module, which solves for the magnetic equilibrium of the plasma, and a dedicated module for the Electron Cyclotron Resonant Heating (ECRH), simulated by the code GRAY [77]. For transport simulations, two options are available: the QualiKitz (QLKZ) code [23], which offers high precision but slow computational performance, and Bohm-Gyro-Bohm (BgB) [30], which provides a significantly faster but less accurate alternative. This work typically preferred the latter due to the need for rapid simulations across multiple full-power scenarios. In addition to these modules, JETTO-Jintrac includes several other specialized modules, such as impurity transport, neutral beam injection (NBI), and various MHD instability models. For the purposes of this research, the focus was on the NTM module within JETTO-Jintrac [12]. The NTM module was used to simulate the behaviour of tearing mode in DTT's full-power scenarios, providing critical indications regarding their saturation width and potential mitigation strategies. The following section will describe further details on the NTM module and its specific application to the scenarios.

4.1.1 NTM module

The NTM module is designed to be integrated with the transport models, allowing for a comprehensive analysis of magnetic island dynamics over time. The NTM module operates by solving the island's evolution equation, which describes the growth of tearing modes within the plasma. It calculates the amplitude of the magnetic island as a function of time, providing detailed insights into the behavior of these instabilities under various operational conditions. JETTO-Jintrac runs simulations from time t_0 to t_1, with the time step being adaptive, determined by the code itself to balance computational efficiency with accuracy. The magnetic surface equilibrium is recalculated for each time loop, along with the kinetic profiles through transport and the ECRH deposition profile using the GRAY code. When the mode is triggered, the NTM module is incorporated into this loop, calculating the island width and growth rate at each time step according to the Generalized Rutherford equation. Once the initial seed width, W_{seed}, is defined, all the terms of the Rutherford equation are assembled using the equilibrium quantities to compute the growth rate, and the island width is updated using a finite difference method. It is important to highlight that, in calculating the mode evolution, the simulation does not account for the effect of the island on transport and, consequently, on equilibrium. This means that transport phenomena within the island, such as the flattening of temperature and pressure profiles, cannot be observed. As described in Chapter 3, the Generalized Rutherford Equation is composed of different terms referring to different physical effects. The linear stability index, denoted as Δ'_0, quantifies the susceptibility of the rational surface to linear perturbations. It indicates whether the surface is linearly stable or unstable. If Δ'_0 is negative, the rational surface is considered linearly stable, while a positive Δ'_0 signifies linear instability. The NTM module, allows for a calculation of the linear stability index from the temperature profile [16], providing an analytic estimation of the linear stability of the equilibrium. However, a reliable and accurate calculation of Δ'_0 can be challenging due to its dependence on various plasma parameters and conditions. Furthermore, in principle, there could be phenomena, such as impurities penetration, that break the equilibrium condition, possibly increasing the value of the Δ'_0. In the methodology used in this study, the value of Δ'_0 is manually specified to explore different plasma scenarios, including those where the surface is either linearly stable or unstable. In the simulation, the terms related to the destabilizing effect of the bootstrap current Δ'_{bs} and the stabilizing effect of the curvature Δ'_{ggj} are included. Both of them generally depend on quantities from the equilibrium, such as the safety factor and pressure profile. The ion polarization term, whose possible impact on the trigger of a TM is discussed in Chapter 5, is usually neglected due to its scaling as $1/W^3$. This makes it relatively small when the island reaches measurable sizes that are of interest for the evolution of the island. Furthermore, it is possible to simulate the effect of ECRH power injection: Δ'_{CD} and Δ'_h. As shown in Chapter 3, the Δ'_h term accounts for the effects of the ECRH heating that contributes to the total current induced by altering the plasma resistivity and has a stabilizing effect on the tearing mode evolution. Based on the explanation provided in Chapter 3, the effects of the current drive and heating depend on parameters, some of which are related to the plasma configuration, while others depend on the launcher. This allows us to evaluate the impact of different mitigation strategies in various scenarios. In particular, the NTM module allows defining the beam width (i.e. how wide the beam is when it reaches the island), through the W_{cd} parameter, the effects of misalignment with the O-point of the island specifying x_{dep} parameter and modulation (50% on and 50% off), that can be set by a flag parameter provided as input. The free parameters in the code related to the island are two: the seed island width, which represents the initial width of the island at time zero, and the W_d parameter, contained in the denominator of the Δ'_{bs} and Δ'_{GGJ} terms. In Chapter 3, the asymptotic behaviors of Δ'_{bs} and Δ'_{GGJ} for small island widths were highlighted, showing that $\Delta'_{bs} \sim W$ and $\Delta'_{GGJ} \sim$ constant. From a modeling perspective, the parameter W_d defines the threshold below which the small island regime becomes significant. Operationally, W_d is set to be equal to the marginal width [66], corresponding to the minimal width the magnetic island can assume to maintain a marginal equilibrium. During the simulations, the seed island width W_{seed} is fixed at 1.75 cm, and the parameter W_d is set to 1.8 cm.

4.2 Methodology

The simulations are based on a reference run of the full-power scenario (E1) [18] during the flat-top phase of the shot. The transport code used is QLKZ, taking equilibrium data from SANCO. Furthermore, the settings of this reference simulation allow to define the power and toroidal and poloidal angles of each beam launcher. The choice of the transport code plays a key role, and a preliminary study was conducted to assess whether it is more advantageous to use a faster but less accurate code, such as Bohm-Gyro-Bohm, or to maintain consistency with the reference simulation by utilizing the more precise QLKZ. Directly switching the transport code is not feasible, as doing so would introduce discontinuities in calculating the kinetic profiles. When altering the transport code, it is essential to allow the simulation to stabilize and attain a new equilibrium profile that reflects the updated model. Therefore, to compare BgB and QLKZ, we first conducted a transient run with the new transport code, allowing the profiles to stabilize before proceeding with further analysis. Figure 4.1 illustrates the relative average difference in the kinetic profiles and safety factor between time t and $t+dt$. The kinetic profiles stabilize after approximately 2 seconds. The transient phase concludes after 6 seconds, with the average difference in the kinetic profiles approaching zero. The simulation at $t=0$ will be the reference for subsequent runs with BgB. Now, it is possible to compare the kinetic profiles with

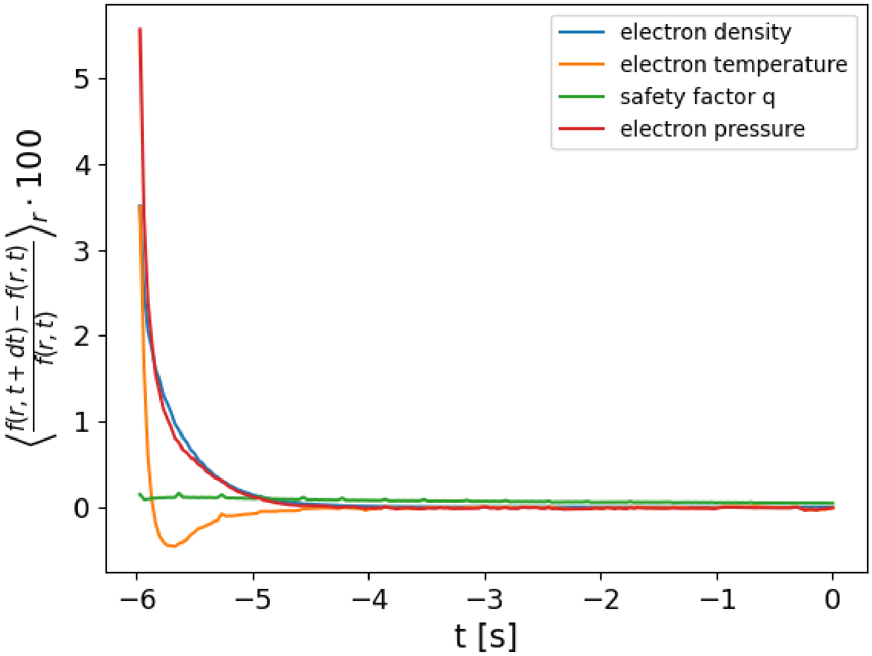


Figure 4.1: Average relative difference of the kinetic profiles between the profile a time t and time $t+dt$, as a function of the time, for the transient simulation.

QLKZ and BgB: Fig. 4.2 shows a comparison between the kinetic profiles. The vertical red line marks the location of the $q = 2$ rational surface, where equilibrium quantities are calculated to solve the GRE. The kinetic profiles differ primarily in the plasma core, with deviations ranging between 10% and 20%. These differences lead to slight variations in the deposition profile of the EC in the plasma center (note that, the deposition on the $q = 2$ surface is the same for QLKZ and BgB). However, on a rational surface, the discrepancies are negligible, and the results from BgB align closely with those from QLKZ. Despite the differences near the magnetic axis, BgB offers sufficient accuracy at the rational surface. As a result, this simplified model will be employed in the simulations to prioritize computational speed. Given that, as outlined in the previous section, the code does not account for the effect of the island on transport (thus, the island does not alter the kinetic profiles), it is possible to consider both the equilibrium and transport as fixed. The only element that varies throughout the simulation is the EC power deposition profile, which is modified by the movement of the upper launchers to align with the island o-point. As a result, significant computational speed gains can be achieved by decoupling the NTM module from the JETTO-Jintrac time loop. This can be done by developing an offline code that operates independently, taking the necessary parameters to solve the Rutherford equation as input and keeping them fixed throughout the simulation. This approach assumes that the plasma remains in equilibrium and that transport does not affect this equilibrium.

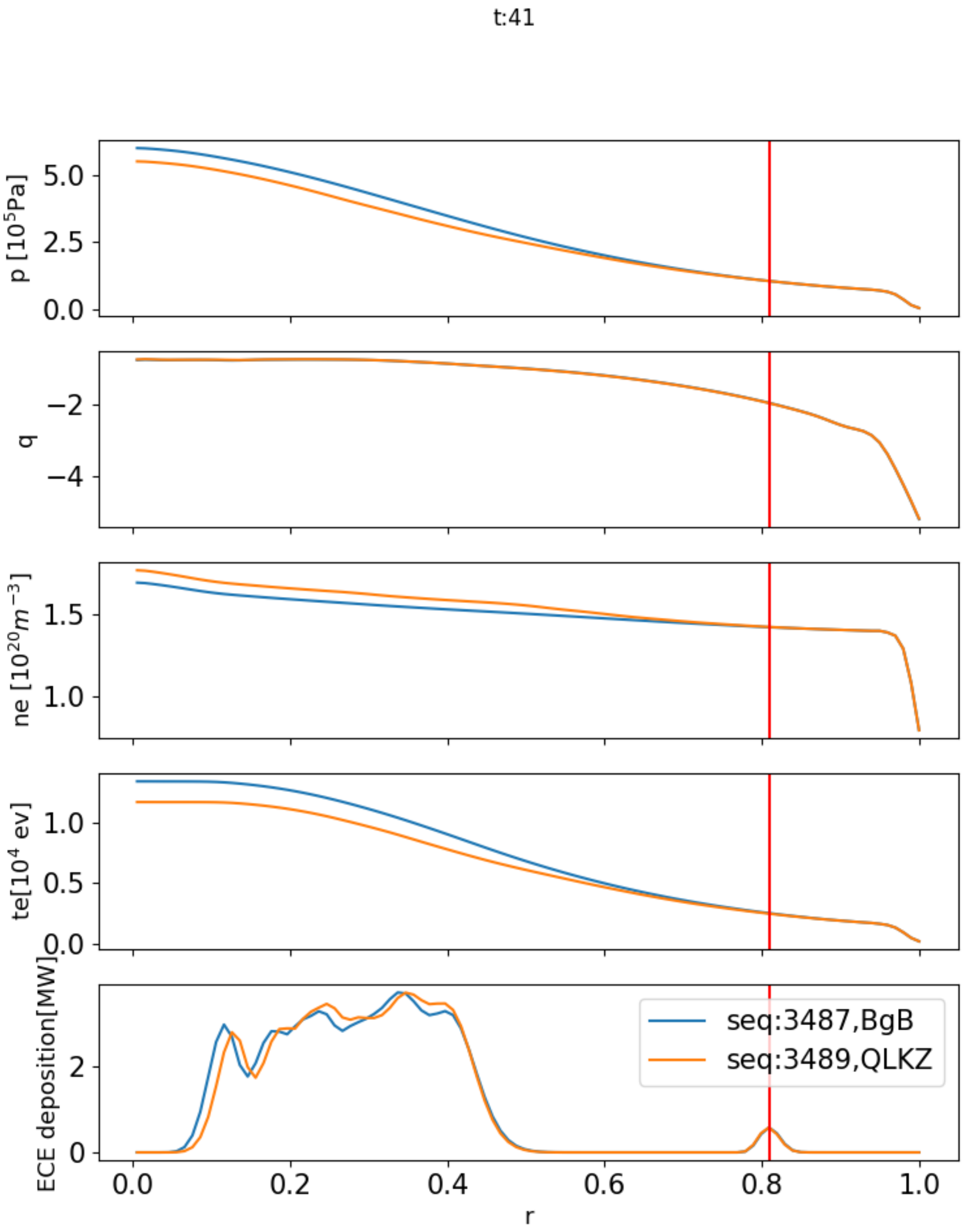


Figure 4.2: Comparison between kinetic profiles for the Bohm-Gyro-Bohm (BgB) model and the Quasilinear Gyrokinetic model Qualikiz (QLKZ) as a function of the normalized radius. Starting from the top, the electron pressure, the safety factor, the electron density, the electron temperature, and the EC power deposition profile. The blue line represents the Bohm-Gyro-Bohm, and the orange line represents QLKZ. The red vertical line corresponds to the position of the rational surface.

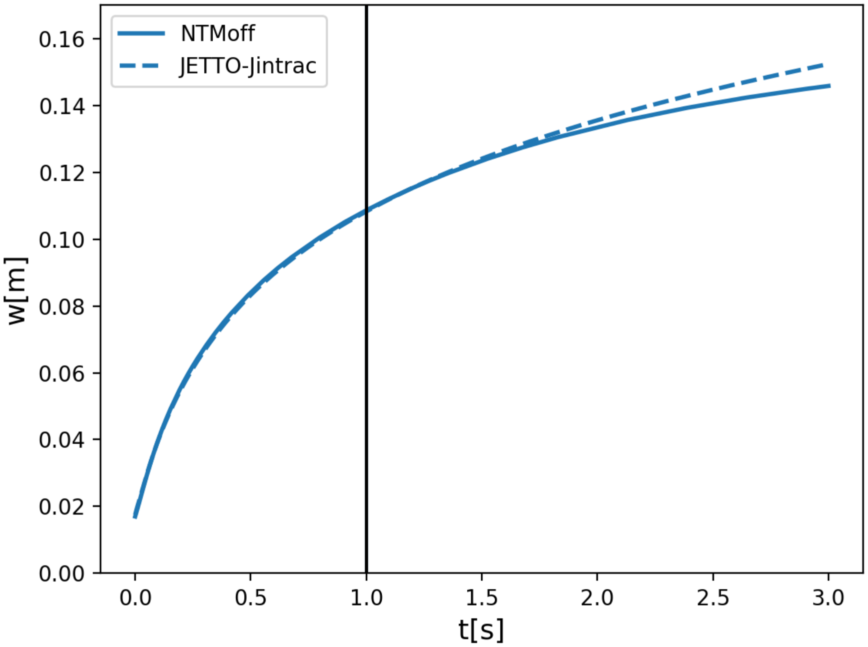


Figure 4.3: Comparison between the mode evolution calculated by the integrated code JETTO-Jintrac and the offline NTM module. The accordance is good till $t = 2$ s, one second after turning on the ECRH (black vertical line). After three seconds, there is an acceptable difference of less than 10%.

These assumptions enable a significant increase in speed, as the evolution of the NTM is no longer tied to transport and equilibrium. In particular, it becomes possible to rapidly conduct studies by scanning the parameter space defined by Δ_0' and W_{cd}, the beam width. This approach necessitates having a reference simulation from which to extract inputs related to the kinetic profiles and the EC. It is important to emphasize that since we are not considering impurities or other external phenomena affecting the plasma, the only source of profile changes is the EC heating. Therefore, JETTO-Jintrac is used to perform reference simulations with various deposition profiles, from which the input parameters for the offline NTM module are extracted. GRAY consistently computes the parameters that describe the EC deposition, specifically W_{cd} (the beam width) and the induced current J_{cd}. When performing simulations using the offline NTM code, the beam width W_{cd} is treated independently from the GRAY code, allowing us to modify it without the need to rerun the reference simulation of the JETTO-Jintrac code. This is based on the assumption of a linear relationship between the EC deposition profile and the beam width: if we double the beam width, J_{cd} will be halved to conserve the induced I_{cd}. In this context, variations of W_{cd} are related solely to the launcher configuration and do not depend on the plasma physics. In particular, changing the beam width, can be seen as a change of the ECH

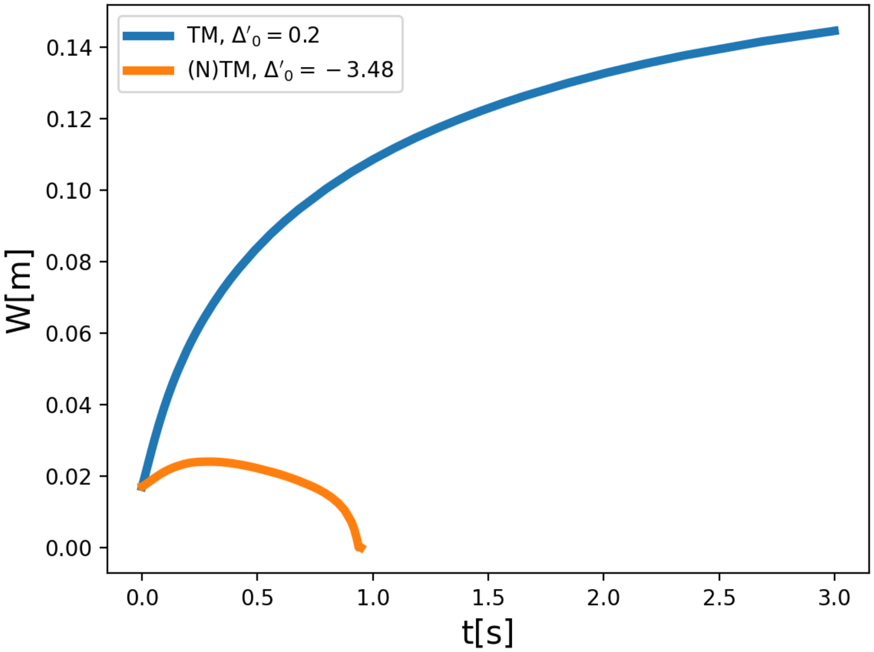


Figure 4.4: Comparison between the free evolution of the mode with the Δ'_0 (green line) calculated by JETTO-Jintrac and $\Delta'_0 = 0.2$ (blue line)

launching antenna. The inputs provided to the offline NTM module, derived from the reference simulation, include the gradient values of the kinetic profiles in the rational surface, essential for constructing the various terms of the Generalized Rutherford Equation (GRE). Additionally, only the current density J_{cd} induced on the rational surface by the upper launchers is considered, as other current drives are not relevant in this context since they do not directly influence the evolution of the magnetic island. Figure 4.3 compares the evolution of the NTM calculated by JETTO-Jintrac and the offline NTM module using the same simulation parameters. EC heating is initiated at $t = 1$ (indicated by the black vertical line), injecting 1 MW of power. This reduces the growth rate, although the power is insufficient for complete stabilization. The simulation results closely match for the first two seconds; an acceptable discrepancy of approximately 10% is observed. The results presented in the following section are obtained using the NTM offline module, considering a fixed equilibrium, calculated from a reference integrated simulation employing a BgB model for the transport according to previous considerations.

4.3 Results

4.3.1 Free evolution of the mode

As pointed out before, it is possible to calculate the Δ'_0 using an analytical approach from equilibrium data [16]. A computation of this index provides a value of -3.48, indicating a linear stability state for the 2/1 rational surface. In this condition, the mode is self-healing (orange line in Fig. 4.4), namely, even with a seed island of $W_{seed} = 1.75$cm, the mode width decrease and after $1s$ disappears. This result suggests that the equilibrium of the full-power scenario in DTT will be linearly stable. However, in principle, an increase of the Δ'_0 as a consequence of external phenomena, such as impurities [80], should be possible. A comparison with the mode evolution corresponding to a linearly unstable case ($\Delta'_0 = 0.2$) is performed and represented by the blue line in Fig. 4.4. In this configuration, the mode evolves and saturates at an amplitude of $\sim$ 14 cm after 3s. This represents the worst-case scenario and will be used as a reference value in the following results.

4.3.2 Perfect alignment

The study of the stabilization follows the analysis of the free evolution of the mode. The effect of the EC is considered in the terms Δ'_{cd} and Δ'_h of the Generalized Rutherford Equation. Initially, the case of perfect stabilization alignment with the magnetic island was considered. In this case, the Electron Cyclotron (EC) power is deposited exactly on the resonant surface. A typical deposition profile for a perfect alignment scenario is shown in Fig.4.2 (the last plot at the bottom), where the deposition of the upper launchers corresponds to the rational surface (red line). A perfectly aligned 5 MW EC injection has been considered, starting after 1 s when the island width is approximately 10 cm. The results are shown in Fig. 4.5. The grey line represents the free evolution of the mode. The deposition profile, computed by GRAY in the integrated simulation, produces a beam width of $W_{cd} = 2.27$ cm (orange line). This configuration allows a complete stabilization of the mode after 2.2s. As specified before, the beam width is changed, corresponding to the different focus of the launcher antenna; in particular, the W_{cd} computed by JETTO has been halved (blue line) and doubled (green line). As shown in Fig. 4.5, the general behavior is that the larger the beam width, the less effective the stabilization. Indeed, when the W_{cd} is halved, maintaining the same power, the mode is stabilized after 1.9 s, while when the W_{cd} is doubled, the mode reaches a metastable condition where $W_{sat} = 4$ cm. Then the modulation effect is assessed with a $W_{cd} = 2.27$ cm (the reference from JETTO) and 5 MW of stabilizing power. A modulation 50% on and 50% off has been considered here. In Fig. 4.6 the evolution with and without the effect of the modulation is compared, along with the values of Δ'_{cd} and Δ'_h. As anticipated by experimental observations and literature, modulation has a beneficial effect on stabilizing the mode: the island is stabilized earlier when modulated power is applied. According to the model used, the physical reason is

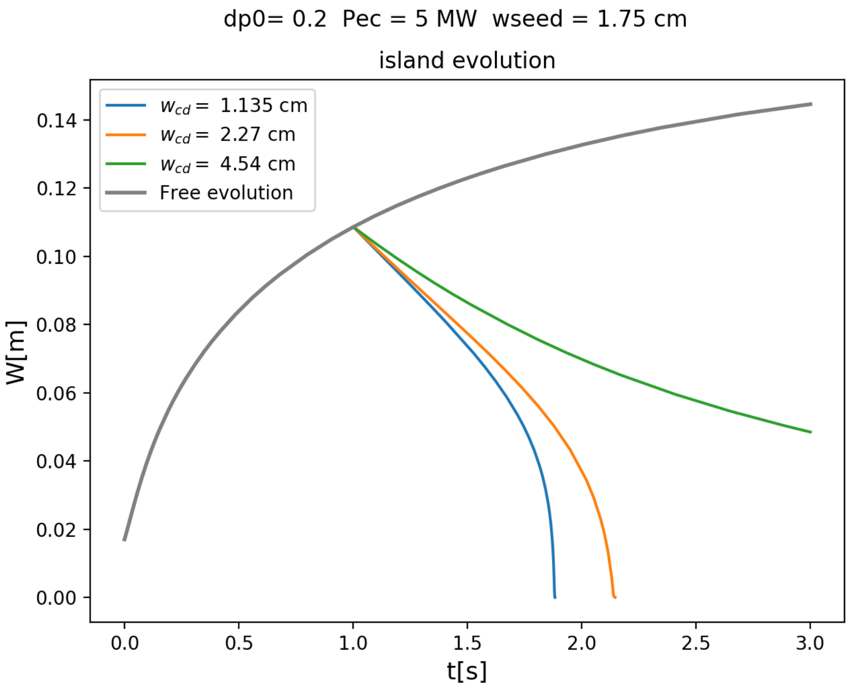


Figure 4.5: Evolution of the mode for different value of W_{cd}. The orange line represents the beam width computed self-consistently by GRAY. The blue and the green lines represent, respectively, the cases where the beam width is halved and doubled. The grey line, for reference, shows the free evolution of the mode. The starting island $W_{seed} = 1.75$ cm, the $\Delta_0' = 0.2$ (simulating the worst case scenario), and the stabilizing power is 5 MW.

that, despite a reduction in the stabilizing contribution from heating (see the second plot, red line), modulation increases the stabilizing contribution of $\Delta'_{\rm cd}$. This effect is particularly evident as the island reduces its size. Indeed, the trajectory of the island width over time is quite similar with and without modulation during the initial phase of stabilization. The discrepancy becomes more pronounced when the island width falls below 8 cm, due to the increased influence of $\Delta'_{\rm cd}$ in this phase.

4.3.3 Misalignment

The effect of misalignment was studied for two control cases. In the late intervention case, stabilization of the magnetic island begins 1 s after the trigger, with the island width at the start of stabilization being approximately 11 cm. In the early intervention case, stabilization begins 200ms after the trigger, when the island width is about 6 cm. Specific reference points on the magnetic island were selected to evaluate the effect of misalignment, and the poloidal launch angle of the upper launchers was adjusted so that power was deposited at positions other than the O-point. Specifically, positions were chosen symmetrically with respect to the island axis at the half-width of the island (± midW), at the separatrix (±

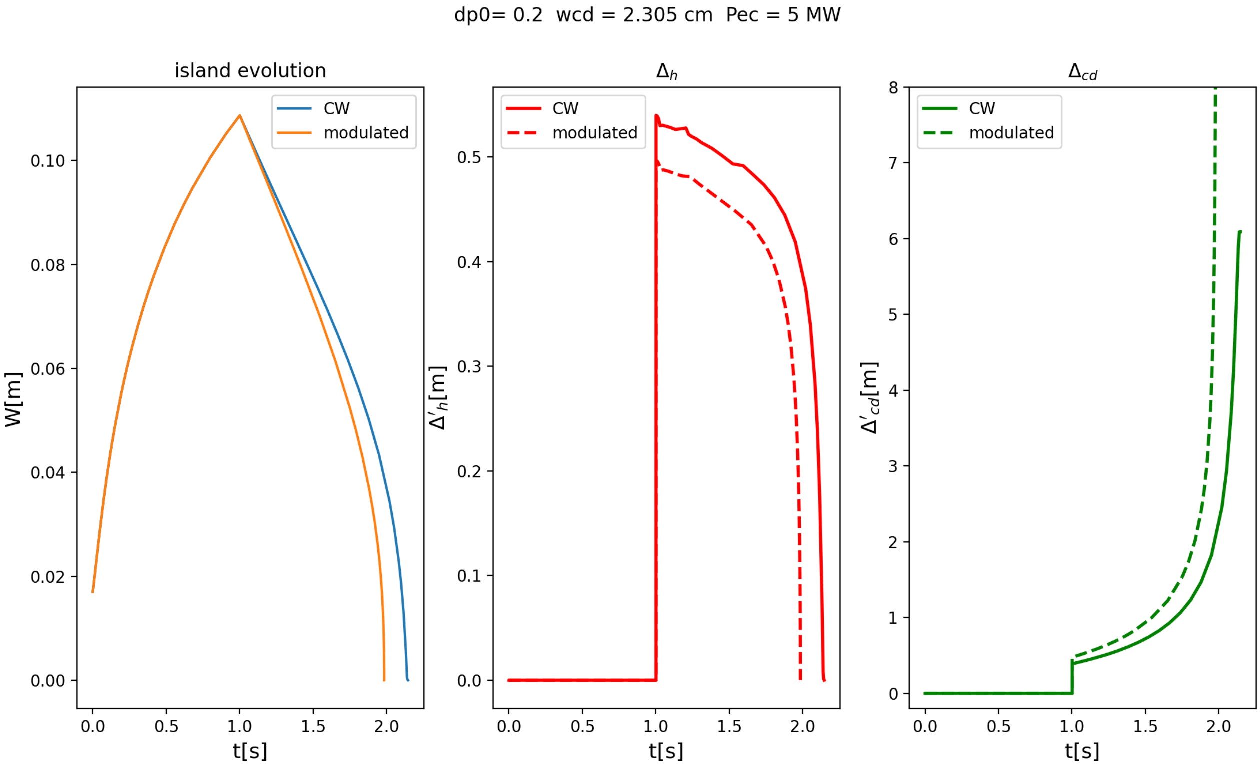


Figure 4.6: Comparison between stabilizing the island with modulated and unmodulated power. Starting from the left, a comparison of the width evolution for continuous wave (CW), blue line, and modulated orange line. On the center and on the right, a comparison of Δ'_h and Δ'_{cd}, for CW (solid line) and modulated (dotted).

sep), and at a location outside the separatrix (± ex). Considering both the inboard and outboard sides, this approach resulted in seven different deposition profiles. To evaluate each poloidal launch angle in both the late and early intervention scenarios,

integrated simulations were performed using the JETTO-Jintrac code. In each case, GRAY self-consistently calculated the deposition profile. Fig. 4.7 and 4.8 show the depositions profiles for the reference position defined before. The black vertical lines represent the radial projection of the magnetic island. For completeness also the case perfectly aligned is shown. It is noteworthy that for every position, the shape of the profile changes according to the calculation performed by GRAY. For every poloidal angle, the evolution of the island has been studied, changing the beam width, following the same approach of the case perfectly aligned.

Figs. 4.9(left) and 4.9(right) show the evolution of the island for every scenario, respectively, in early intervention and late intervention. For every different position of the deposition, the evolution of the island is evaluated, considering the beam width computed by JETTO-Jintrac (orange), its double (green) and its half (blue). The general behavior that the larger the beam width, the less effective the modulation is holds. According to the sim-

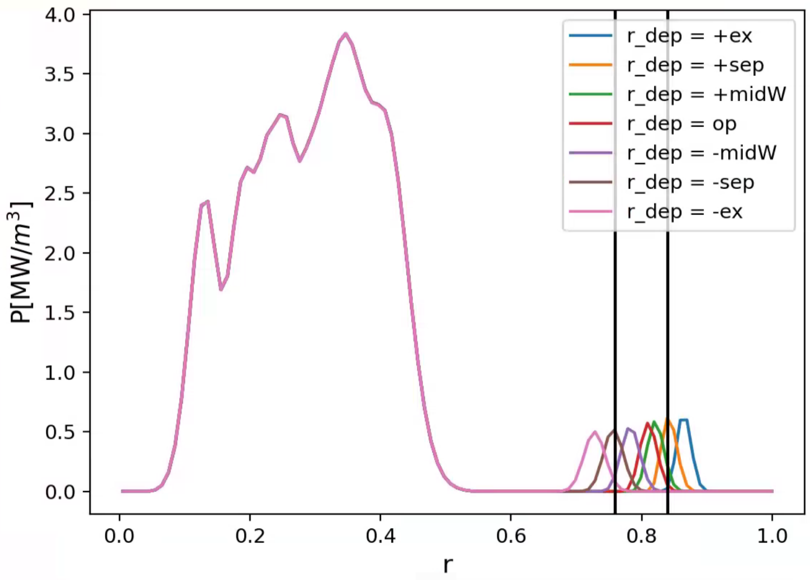


Figure 4.7: Profile depositions for six different poloidal angle represented by six different labels: *ex*, corresponding to shooting externally to the island, *sep*, the separatrix, *midW*, the middle width of the island and *op*, the O-point. The sign before the label corresponds to the high field side (-) or the low field side (+). The vertical black lines represent the radial projection of the island width. The plot shows the different cases of the deposition profile for the early intervention case

ulations, complete stabilization can be obtained only by hitting the island perfectly aligned with the O-point, either in late or early intervention. A slight misalignment (±midW) prevents the system from completely stabilizing the island, generating a metastable saturated island of around 9 cm in late intervention and 4 cm in early intervention. The effectiveness of stabilization rapidly decreases as the deposition moves away from the O-point, and the EC becomes destabilizing when injected outside the island. This effect is more pronounced when injecting power on the high-field side. Specifically, in the *-ex* configuration shown in Fig.4.9(right), the saturated island width ranges between 12 cm and 14 cm, depending on the beam width. In contrast, the corresponding behavior on the low-field side (see the last plot on the right in Fig.4.9(right)) results in a saturated island width ranging between 8 cm and 12 cm. Generally speaking, a misalignment oriented towards the low-field side tends to be more effective than the corresponding deposition profile on the high-field side. This could be attributed to the fact that, as shown in Figs. 4.7 and 4.8, injecting EC on the low-field side results in a more focused deposition, thereby enhancing its stabilizing effect. Modulation rapidly loses effectiveness as the deposition point moves away from the O-point. The rate at which it becomes ineffective depends on the timing of the intervention. Specifically, when the magnetic island is small, a slight misalignment from the O-point (±midW) still allows the beneficial effects of modulation to be observed.

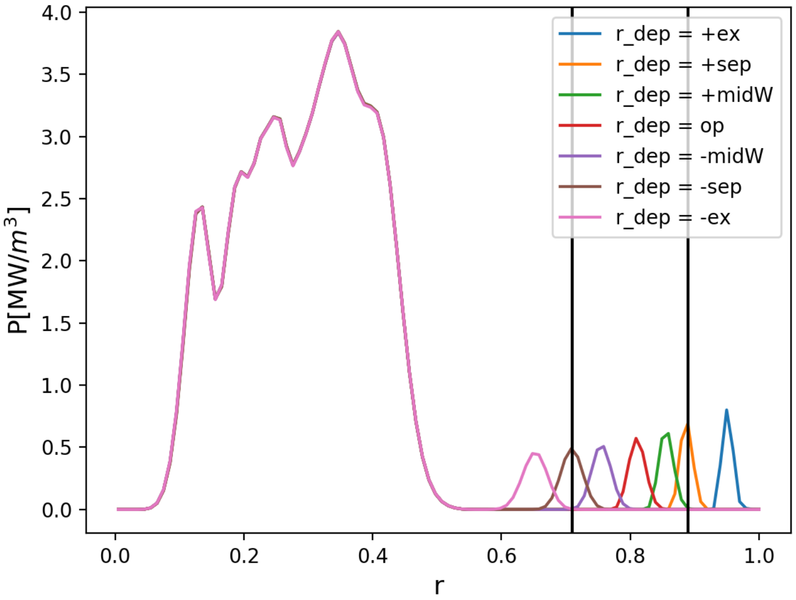


Figure 4.8: Profile depositions for six different poloidal angle represented by six different labels: *ex*, corresponding to shooting externally to the island, *sep*, the separatrix, *midW*, the middle width of the island and *op*, the O-point. The sign in front of the label corresponds to the high field side (-) or the low field side (+). The vertical black lines represent the radial projection of the island width. The plot shows the different cases of the deposition profile for the late intervention case.

In contrast, during late intervention, modulation quickly loses its efficacy, becoming irrelevant as soon as any misalignment from the O-point is introduced. This is evident from Fig. 4.10, which shows the evolution of the island in four different conditions. The modulation conserved effectiveness both in late and early intervention when perfectly aligned (top and bottom, right), leading to a more rapid and complete stabilization of the mode. Introducing a slight misalignment, the modulation remains effective in the early intervention case (top, left plot), while becomes ineffective if the island is larger (bottom, left plot). The effectiveness of the modulation is quantified by an effectiveness parameter defined as:

$$\mathrm{E}(W_{\mathrm{cd}}) = \int_0^{t_{\max}} W_{\mathrm{CW}}(t')\,dt' - \int_0^{t_{\max}} W_{\mathrm{modulated}}(t')\,dt' \tag{4.1}$$

This parameter corresponds to the difference between the integrals of the curves representing the continuous-wave (CW) and modulated evolution of the island. The larger the index defined in Eq. 4.1, the greater the difference between the curves. If the index is positive, modulation results in faster stabilization (since the integral is smaller), making it beneficial. Furthermore, the benefit increases as the index becomes larger. Fig. 4.11a and 4.11b represent the value of the parameter $E(W_{cd})$ as a function of the beam width W_{cd}.

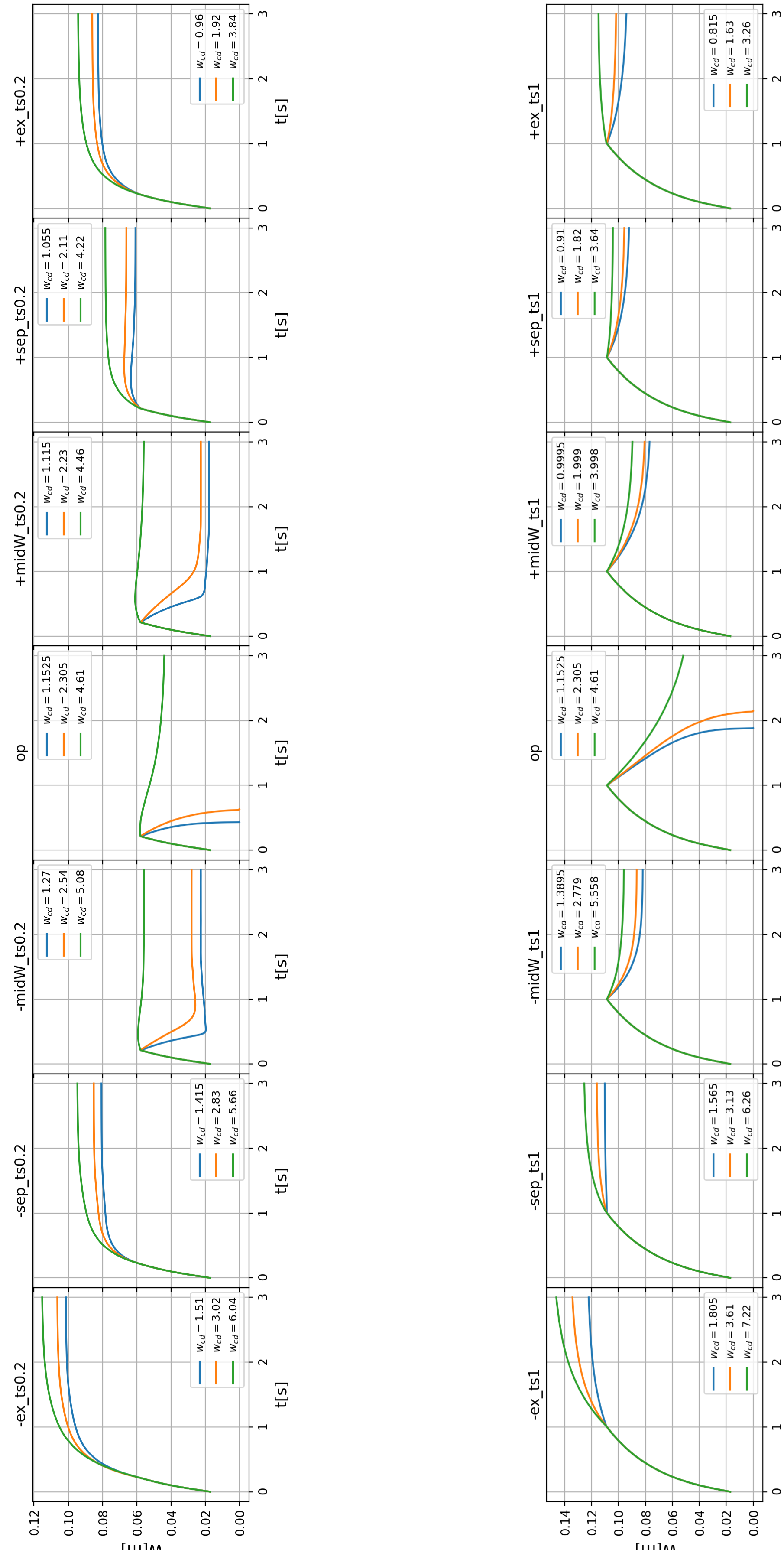


Figure 4.9: evolution of the island for the different cases of deposition profile. The left panel shows early intervention, while the right panel shows late intervention. Each line in the plots corresponds to a different W_{cd} value. The orange line represents the evolution computed with the beam width W_{cd} calculated self-consistently by GRAY. The blue and green curves represent a halved and double beam width.

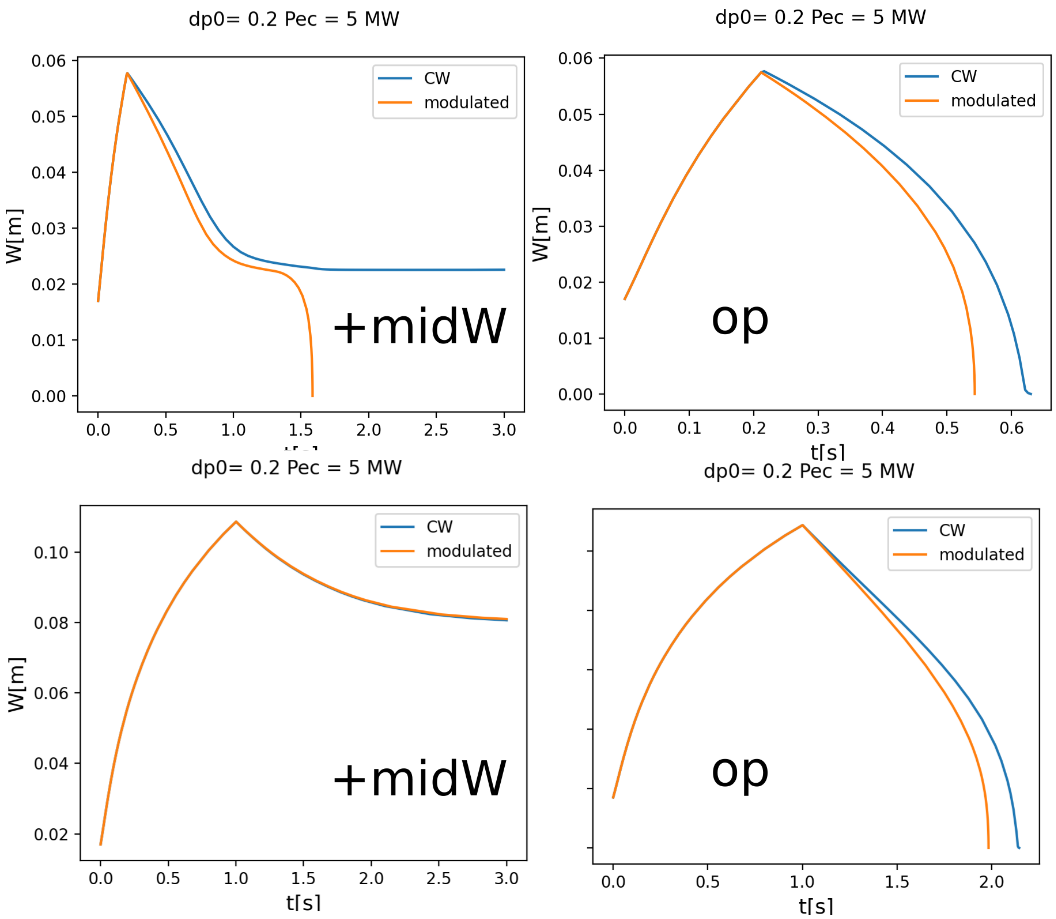
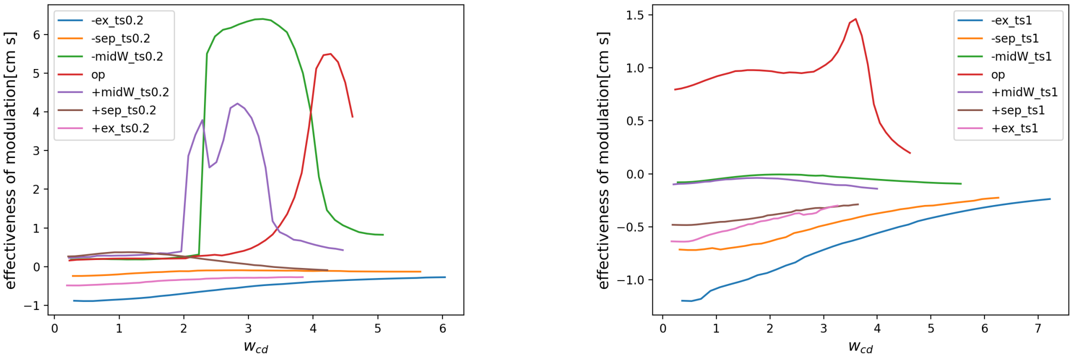


Figure 4.10: plot of the evolution of the mode, for the case continuous wave and modulated for different deposition profiles using different control schemes. The orange line corresponds to a modulated stabilization and the blue line to continuous wave (CW) stabilization. The plots on the left represent the evolution of the mode with a misalignment at half width in early intervention (top) and late intervention (bottom). On the right, a perfectly aligned stabilization is shown in early intervention (top) and late intervention (bottom)

In the case of early intervention (Fig. 4.11a), the island is smaller, and the modulation is more effective even for small misalignments (purple and green curves). Injecting power at the separatrix, both on the low-field side and the high-field side, during both early and late interventions, drastically reduces the effectiveness of modulation. The effectiveness parameter $E(W_{\text{cd}})$ approaches zero or may even become negative, indicating that modulation does not have a beneficial effect under these conditions. In the case of late intervention (Fig. 4.11b), the index $E(W_{cd})$ is positive only when there is perfect alignment, and it decreases rapidly with the introduction of even slight misalignment. Notably, the value of $E(W_{cd})$ is smaller in this scenario, indicating that it is generally more advantageous to apply modulation to a small island.



(a) Early intervention

(b) Late intervention

Figure 4.11: Value of the effective parameter $E(W_{cd})$ as function of the beam width W_{cd} for different deposition profiles.

4.4 Conclusion

To summarize, a study on the behavior of the Neoclassical Tearing Mode (NTM) in the DTT full-power scenario was conducted. Initially, the free evolution of the mode was analyzed using a Δ_0' calculated through an analytical approach and by fixing Δ_0' to simulate a worst-case scenario where the linear stability index increases due to external phenomena. Using the computed Δ_0', the plasma was found to be linearly stable, resulting in the mode evolving and disappearing within approximately 1 s. In contrast, for a fixed $\Delta_0' = 0.2$, the mode saturates at an amplitude of approximately 14 cm after 3 s. The stabilization has been simulated, firstly assuming a perfect alignment condition and then introducing a misalignment. Considering the effects of misalignment, it is reasonable to contemplate a trade-off between accurately targeting the center of a large island during late intervention or targeting with less precision but at an earlier stage when the island is small. Figures 4.9(left) and 4.9(right) demonstrate that the results obtained by depositing power at the separatrix during early intervention are comparable to those achieved by targeting the mid-width of the island during late intervention. This indicates that, in early intervention, a greater degree of misalignment is acceptable to achieve similar stabilization outcomes, meaning that the less accuracy in the detection of the island, due to the small size of the island, is balanced by the less accuracy required to suppress the island. On the other hand, shooting accurately (on the O-point) when the island is large (late intervention) would ensure a complete stabilization of the mode in around 2 s. Finally, the effectiveness (how much it is evident the effect of the modulation) is evaluated, defining the index $E(W_{cd})$. Figures 4.11a and 4.11b illustrate this parameter as a function of the beam width. In the case of late intervention (right panel), only the red curve, representing the perfect alignment scenario, remains visibly above zero. The other curves rapidly approach zero and may even become negative. In contrast, during early intervention, even with a small misalignment (at half the island width), the effect of modulation remains positive. This indicates that modulation is more effective in stabilizing small islands than larger ones, implying that it can be useful to compensate the lack of accuracy in the detection in the early intervention case.

Chapter 5

Ion Polarization Current

The neoclassical effect caused by the favorable curvature of the magnetic field lines, modeled by Green-Glasser-Johnson (Δ'_{GGJ}), stabilizes the NTM even during the linear phase. For this reason, linear stability ceases to be a sufficient condition for triggering the instability, and a critical Δ'_{0c} must be overcome to start the non-linear growth of the tearing mode. During plasma operation, the equilibrium can be altered by the effect of the impurities that penetrate the plasma, providing a channel of energy dissipation through irradiation. As explained in Chapter 6, a possible disruption path starts with the edge cooling, where a collapse of the temperature profile at the edge of the plasma occurs. The current density profile shrinks, which is known to have a destabilizing effect, increasing the linear stability index Δ'_0 [106, 79], which triggers the mode, leading to disruption. The linear stability analysis focuses only on the equilibrium current density profile, neglecting all non-linear phenomena that arise in the interaction between the magnetic island and the plasma. Furthermore, in edge cooling, the current density profile modification and the linear destabilization are a consequence of the change of the temperature, so it is important to put attention to the processes that depend on the temperature and that could cover an important role in the development of the magnetic island in the very first phase of its development. The ion polarization current arises due to the different inertial response between the electrons and ions. The published work, Ref. [13] and reported in Appendix A provides a detailed physical description. As shown in Chapter 3, in the Generalized Rutherford Equation, the effect of the ion polarization current is modeled in the terms of Δ'_{pol}, which depends on the rotation frequency ω of the island with respect to the plasma frame of reference. In the small island regime ($W < \rho_i$, with ρ_i the ion Larmor radius), the island is dragged by the motion of the electrons, and ω corresponds to the diamagnetic electron frequency $\omega_{*e}(T)$. The collapse of the temperature at the edge suddenly reduces $\omega_{*e}(T)$, enhancing the destabilizing effect of the polarization current. Furthermore, $\Delta'_{pol} \sim 1/W^3$, so that, for a small island, the effect of the ion polarization current is dominant, and in the published paper, an analysis on a database of disrupting JET pulses is carried out to model the effect of the ion polarization current. The work shows that when the flattening of the temperature profile reaches the rational surface, there is an increase in the destabilizing effect of the ion polarization current, and that moment corresponds to the onset of the mode. The destabilizing effect of the ion polarization itself is not sufficient to trigger the mode. However, it can strongly contribute to putting the rational surface in a metastable instability condition, bypassing the stabilizing effect of the curvature and favoring the onset of the mode when a small perturbation bends the field line close to the rational surface. The ion polarization current makes the rational surface susceptible to small, fast fluctuations of the plasma's rotation frequency. Statistical analysis over a dataset of disruptions in DIII-D [10] recently shows that the onset of the modes follows a Poisson distribution, suggesting a stochastic behavior of the onset of the tearing mode that can be physically explained with the ion polarization current.

Chapter 6

Detection of disruptive precursors

When uncontrolled, the tearing instability grows, slows down, and locks with the tokamak wall. At this stage, the confinement is highly compromised, the plasma current rapidly decreases, and all the energy stored is released. Avoiding and mitigating this behavior, known as disruption, is a key challenge for future power plants. Disruptive events can occur due to various phenomena and following different patterns. Deep learning, through data-driven approach, offers a promising way to unveil the causal chain of processes that eventually lead to disruptions, allowing for the detection of disruptive precursors. This chapter provides a phenomenological description of the disruption, describing different possible disruptive patterns. Then, a general introduction to deep learning algorithms is presented, focusing in particular on neural networks and convolutional neural networks and their application in disruption prediction. The chapter ends with describing the explainable artificial intelligence (XAI) analysis applied to a convolutional neural network trained to classify disruptive plasma. The analysis focuses on particular disruptive patterns, edge cooling, and temperature hollowing. A detailed explanation is provided in Ref. [14], and the paper is reported in Appendix B.

6.1 Disruption

Tokamak operations are intrinsically unstable and susceptible to events that can disrupt the plasma confinement, quickly destroying the magnetic equilibrium. A disruption is a very fast phenomenon in which different processes interact non-linearly. Phenomenologically, the disruption occurs in two stages. The first phase is where the internal energy is lost, typically due to radiation. The radiated power leads to a drop in the temperature profile, known as the *thermal quench*, which lasts a few milliseconds. This rapid cooling significantly increases the plasma resistivity (which scales as $\sim T^{-3/2}$). In the first stage, this reduction results in an increased power injection into the plasma through Ohmic heating, arising from the release of magnetic energy due to the decrease in the poloidal magnetic field. The low temperature of the plasma favors the dissipation of this energy in the form

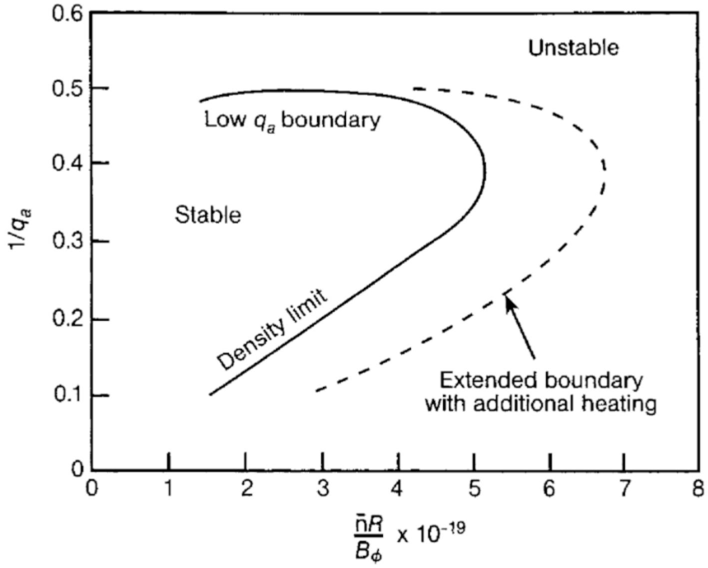


Figure 6.1: Illustration of the operational limit in the Hugill's diagram. The hard low-q limit is represented by the horizontal line at $1/q_a = 0.5$, while the threshold due to the density limit can be moved by introducing additional heating. [104]

of impurities radiation, which prevents the electrons from regaining energy and restoring temperature. The current decay is very rapid, with $dI/dt \sim 200\text{MA/s}$. [113]. This phase, known as *current quench*, ends with the complete loss of the current and the confinement. The very short timescale at which disruption happens makes it difficult to avoid it once the thermal quench starts. For this reason, disruption avoidance is closely related to the ability to predict disruptions. A statistical, empirical analysis of disruptions, allows for identifying instability thresholds that involve global plasma parameter, thereby helping to prevent disruption. The Hugill's diagrams plots the pulses in the plane defined by the inverse of the q at the plasma edge, $1/q_a$ with respect to the Murakami's parameter $\bar{n}R/B_\phi$. This parameter space defines two different kinds of disruptions by revealing two limits the plasma cannot overcome to maintain stability: the low-q and the density limits. The first limit, known as the ideal MHD low-q hard limit, is characterized by the onset of an external kink instability, which occurs when the $q = 2$ rational surface reaches the plasma edge (namely $1/q_a = 0.5$ in the Hugill's diagram). In these cases, the plasma rapidly disrupts, and this constraint on the value of q at the edge limits the achievable plasma current (as $q_a \sim 1/I$) . The second limit is imposed by the so called density limit, and occurs when increasing the line average density at the edge, with respect to the toroidal magnetic field, moving to the right in the Hugill's diagram. An increase in the density, which is not balanced by an increment of the injected power, reduces the temperature at the edge, and the plasma undergoes the so-called edge cooling. This favors the impurities radiation and leads to a contraction of the temperature profile. The contraction itself is not dramatic, and the plasma in principle could produce a new equilibrium with a decreased internal energy.

However, reducing the current channel moves the $q = 2$ surface outward, destabilizing the

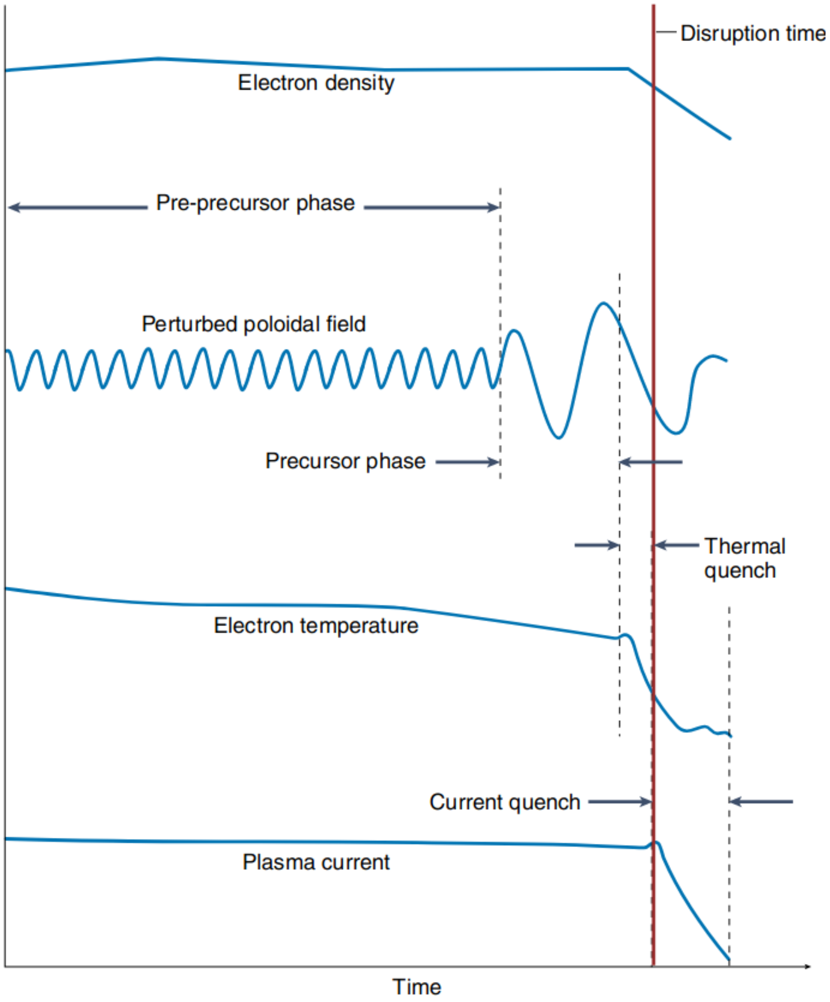


Figure 6.2: Illustration [108] of the different phase of a disruption. Perturbation of the equilibrium can make the plasma unstable, starting the pre-precursor phase. This example shows a sketch of the perturbed poloidal field, connected to MHD activities. The pre-precursor phase is followed by the precursor phase, where the phenomena directly responsible for disruption occur. Then starts the thermal quench, with the rapid decrease of temperature, that induces the current quench and the complete loss of confinement at the disruption time (red line)

$(2,1)$ tearing mode. The mode slows down during its growth until it locks with the machine wall. At this point, the thermal and current quenches start and the plasma disrupts, terminating the discharge. While the low-q limit is a hard limit, namely, it intrinsically affects the plasma, as shown in Fig. 6.1, the density limit can be moved if the decrease of temperature due to the increase of density at the edge is balanced by additional heating that prevents the contraction of the current channel. While in the density limit, edge cooling is caused by the increase in the density at the edge, another possible disruption path is given by impurities, which can penetrate the plasma and modify the temperature profile through radiation. Depending on the penetration depth, the impurities can alter different areas of the temperature profile, causing a cooling of the edge or a hollowing of the temperature on the plasma axis. Both the edge cooling and the temperature hollowing are shown to be linearly destabilizing [80] for the $(2, 1)$ rational surface, triggering an NTM leading to disruption. These limits establish operational boundaries, defining the plasma parameters required for a safe discharge. However, disruptions may also arise from other causes (e.g., interaction between instabilities, error fields, and impurity accumulation) that go beyond these defined limits and are more difficult to identify in a parameter space as Hugill's diagram. For this reason, an effective strategy in disruption prediction involves the detection of phenomena that precede the disruption. Fig.6.2 illustrates the disruption that starts with the current quench, is preceded by the thermal quench, and by a phase where the precursor perturbs measurable quantities. In JET the most frequent precursors are MHD instabilities that can occur even when the plasma is in a stable operational region, particularly $(2, 1)$ tearing mode [25]. Taking into account significant perturbed quantities, such as the poloidal magnetic field measured by the Mirnov's coils, can allow elaborating indexes that anticipate disruptions, triggering disruptions mitigation systems. This is the case, for example, of the mode locking index, which accounts for the locking of the $(2, 1)$ mode as a precursor of disruption. Such systems allow for a rapid and safe discharge shutdown when critical events occur. However, effective disruption avoidance systems must provide extended warning times, enabling the anticipation of precursor onset and moving the plasma in a safe operational region without the need to shut down the discharge [105]. Developing such a system is highly non-trivial due to the pronounced nonlinearity of the sequence of events (which begins with the pre-precursor phase (Fig. 6.2) and culminates in a disruption), which makes the causal chain difficult to reconstruct. A promising approach to addressing the disruption avoidance problem lies in applying deep learning algorithms. These algorithms have the potential to uncover relationships in the data that are difficult to detect through data-driven analysis and challenging to predict with physical modeling. In the next section, a description of the basic deep learning techniques is provided, focusing in particular on deep neural networks and convolutional neural networks. Then, its application in the prediction of disruption is shown.

6.2 Deep learning

Machine learning comprehends a range of algorithms, enabling software to 'learn'. Here, 'learning' carries a precise meaning, specifically:

> A program is said to learn from experience E with respect to a class of tasks T and a performance measure P, if its performance at tasks in T, as measured by P, improves with experience [72]

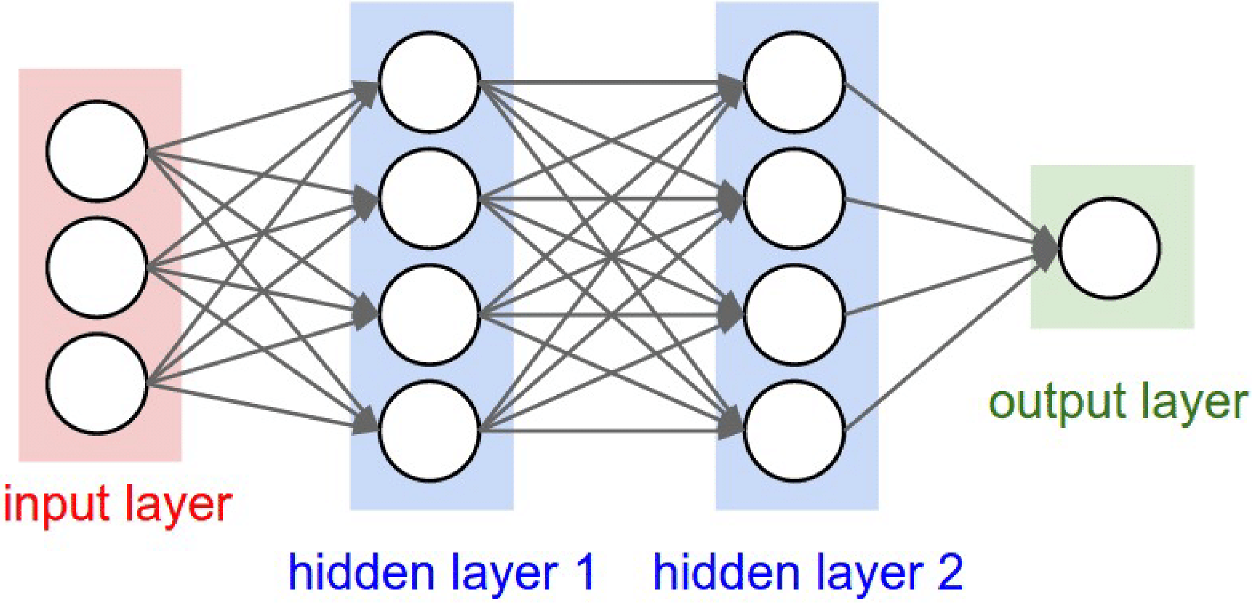


Figure 6.3: An example of deep neural network is shown. The input layer accepts three-dimensional inputs. Every node is connected to each node of the following layer and a weight characterizes this connection. The neural network in this example produces a monodimensional output in the output layer. Every layer between the input and the output layers is called *hidden layer*. [2]

Machine learning algorithms involving the specific use of deep neural networks is known as deep learning. In this section, the neural network (NN) architecture is first defined, followed by a general explanation of its functionalities. The convolutional neural network (CNN) is then presented as a generalization of the neural network that allows to elaborate images. The architecture of CNN is then applied to the context of nuclear fusion, illustrating their use in disruption prediction.

6.2.1 Deep Neural Networks

A NN can be visualized as a network of interconnected nodes, referred to as neurons, each characterized by an activation value, $y(x)$, which is dependent on the input x. These neurons are organized into layers, starting with the input layer and concluding with the output layer. Between these two, the architecture is further composed of a variable number of hidden layers, forming the core of the network's structure. Each neuron in the $(i-1)$-th layer is connected to all neurons in the subsequent i-th layer. This connection implies that the value of the n-th neuron in the i-th layer is determined by the values of the neurons in the $(i-1)$-th layer, following the relation:

$$y_n^{(i)} = A^{i-1}\left(\sum_{k=0}^{N^{(i)}} w_{kn} y_k^{(i-1)} + b_k\right) \tag{6.1}$$

The weights w_{kn} correspond to how much the value of the neuron in a previous layer determines the value of the neurons in the layer. The coefficients b_i are called bias.

Activation functions Particular attention must be posed on the function $A(x)$, called the activation function. It determines how the values of the neurons in the previous layer "activate" the neuron in the current layer. The form of the activation function can be chosen within different functions. One of the most used activation functions is the Rectified Linear Unit (ReLU) defined as:

$$A(x) = \max(0,x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{6.2}$$

The function is continuous, but its first derivative is a piecewise constant, exhibiting a discontinuity at $x = 0$. Higher-order derivatives (beyond the first) are not well-defined. The logistic function is defined as:

$$A(x) = \frac{1}{1+e^{-x}} \tag{6.3}$$

The function is infinitely smooth and monotonic. Its range is (0,1), meaning it is useful for representing probabilities. The neural network can generally be composed of layers with different activation functions. For a given generic input x, the output of the neural network composed by N_l layers can be expressed in terms of nested matrix multiplications:

$$\phi(x,W) = A^{(N_l)}\left(W^{(N_l)}A^{(N_l-1)}\left(W^{(N_l-1)}\cdots A^{(1)}\left(W^{(1)}x\right)\cdots\right)\right) \tag{6.4}$$

where $W^{(i)}$ is a weight matrix referring to the i-th layer of the neural network. The matrix element w_{kn} refers to the weight between the k-th neuron in the $i-1$-th layer and the n-th neuron in the i-th layer. A neural network is characterized by the number of hidden layers, the activation function, and the number of neurons in each layer; these features define the network's structure. Typically, they are fixed and remain constant. Beyond the structure, a neural network is also defined by the weight matrices $W^{(i)}$ associated with each layer.

Universal approximation theorem The most important feature of a neural network is demonstrated in the so-called *Universal Approximation Theorem* [57] of the neural network. The theorem states that for every function $f(x)$ belonging to a function space where a norm is defined (Banach space), always exists a **non-linear** neural network $\phi(x,W)$ such that $\phi(x,W) \to f(x)$. Technically, the demonstration focuses on proving that in a defined Banach space X, every function $f \in X$ is arbitrarily closed to a neural network $\phi \in \Phi$ where $\Phi \subset X$ represents the set of all the neural networks contained in X. Φ is said to be dense in X. A crucial condition for neural networks is nonlinearity. While the activation function can, in principle, be linear, a linear activation would result in the network producing an output that is a linear combination of the inputs, limiting the network to a linear approximation of the function $f(x)$. Consequently, the set Φ would no longer be dense in the Banach space X. The theorem does not suggest how a certain network must be built to perform this approximation; however, different algorithms allow for that.

Training phase Given a NN with a certain structure, the neural network weights represent the degree of freedom that can be tuned to find the network that best approximates a certain function. The phase where the network's weights are tuned to approximate a function is called the training phase. In this phase, the weights of the neural network are iteratively adjusted to minimize a cost (also known as loss) function that takes into account the distance of the output of the neural network $\phi(x,W)$ to the value of the target function $f(x)$. The cost function can be expressed in different way, but its effectiveness depends on the context. For a certain set of input data $(x_j, y_j = f(x_j))$ with $j = 1,2,...,N_{training\ sample}$ and a neural network output $\phi(x_i,W)$, a typical cost function is represented by the squared deviation:

$$C(x,\phi) = \sum_{i=1}^{N_{sample}} ||y_i - \phi(x_i,W)||^2 \tag{6.5}$$

The set of weights W that minimizes the cost function $C(x,\phi)$, defines the neural network that approximates the target function $f(x)$, as stated by the universal approximation theorem. In this context, the learning phase of a neural network is formulated as an optimization problem, requiring the computation of the derivatives of the network's weights with respect to the cost function. A minimization algorithm can then be employed to minimize the cost function. Gradient descent is an iterative algorithm that finds the value of a parameter that minimizes a certain function. The goal of the gradient descent is to minimize a given ($C(x,\phi)$) function with respect to certain parameters (the weights $w_{kn}^{(i)}$, for every layer of the NN) by iteratively adjusting the parameters in the direction opposite to the gradient. For a generic weight w:

$$w_{t+1} = w_t - \eta \frac{\partial C(x,\phi)}{\partial w_t} \tag{6.6}$$

the parameter η is called the learning rate and can be adequately tuned to optimize the descent.

Backpropagation A crucial task is to efficiently calculate derivatives, and the backpropagation algorithm provides an effective approach to this. This algorithm consists of two main steps: the forward pass and the backward pass. First, for each training data pair (x_i, y_i), the cost function $C(x_i, \phi(x_i, W_t))$ is evaluated by passing data through the network in a forward direction. Then, the derivatives of the cost function with respect to each weight in the network are computed through backpropagation, using the chain rule applied to Eq. 6.4. With these computed gradients, the weights W_t are updated according to the gradient descent algorithm Eq. 6.6. This process is repeated for all training data pairs (x_i, y_i) until the cost function is minimized.

6.2.2 Image processing

The convolutional neural network (CNN) [48] is an extension of the neural network, particularly suitable for analyzing input images. The problems that such a network can face are manifold, for example, object tracking [22], face recognition [9], object recognition and classification [117]. In the fusion context, a promising application for neural networks applied to image processing is the disruption prediction [49, 6, 5]. A brief overview of the methods applied to the neural network architecture to analyze images is now presented. A greyscale image, with $M \times N$ pixels, is a $M \times N$ matrix where every element represents the grey intensity (-1 is black and 1 is white). In principle, an image can be used as input of a generic neural network as described in the previous section, modifying the input layer so that it takes an input $x \in \mathbb{R}^{M \cdot N}$, however, all spatial relations in the image would be lost, resulting in a less effective neural network.

Convolutional layer A CNN introduces in the structure of the network a convolutional layer (CL), where a convolutional operation between the input image x and a kernel function g is performed. The kernel g is represented by a matrix $m \times n$ acting over a patch of the input image, which globally is $M \times N$. The convolution operation will produce an output o of dimensions $(M-m)+1 \times (N-n)+1$. The discrete convolution transformation can be formalized as:

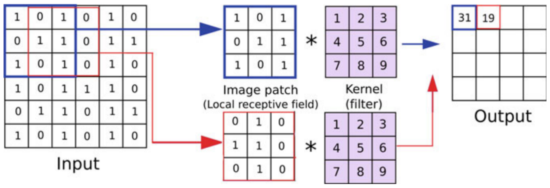


Figure 6.4: Illustration of convolution operation between an input matrix 6×6 and a convolutional kernel 3×3. The kernel acts over a patch of input matrix, whose dimensions are defined by the dimension of the kernel. The convolution produces an output 4×4. [107]

$$o_{i,j} = \sum_{m'=1}^{\bar{N}} \sum_{n'=1}^{\bar{N}} g_{(m',n')} x_{(i+m',j+n')} \tag{6.7}$$

An illustration of the convolution transformation on an input matrix 6×6, with a kernel 3×3 is shown in figure 6.4. In this case $\bar{N} = 4$. The convolution operation transforms the input matrix of the convolutional layer (CL) into a feature matrix that highlights various geometric features of the input. Each convolutional layer performs multiple convolutions using different kernels, with each convolution producing a distinct feature map that emphasizes different geometric aspects of the input image. In general, the kernel matrix is a learnable parameter so that the network learns what are the best features of an image that allow it to minimize the cost function. An example is shown in Fig. 6.5, where various

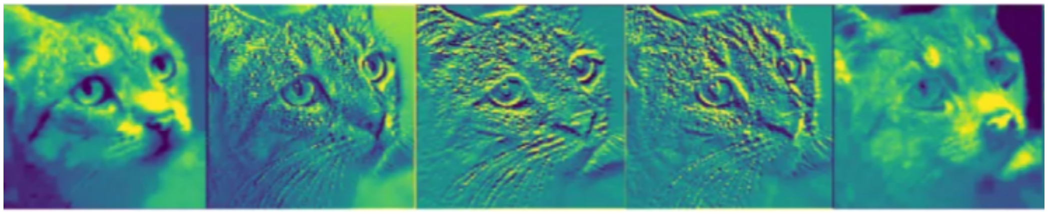


Figure 6.5: example of features extraction performed by different layers of the convolutional unit. Every image refers to a different convolutional kernel that is able to highlight various geometrical features of the image. [75]

geometric features of a cat's head are highlighted. Each image corresponds to a different convolutional kernel, emphasizing distinct characteristics of the cat. For instance, different images highlight the horizontal and vertical lines of the fur, the eyes, and the overall shape of the head.

Pooling The generalization of the NN in order to elaborate images introduces a new way the layer interacts with the input image. In particular, in the convolutional layer, the input image undergoes a convolution transformation driven by a kernel that acts on an image patch with a specific operation defined in Eq.6.7. The same mechanism can be applied with different operations, as in the pooling layers. There are two kinds of pooling layers: the average and the max pooling layer. Figure 6.6 shows an example of max pooling, where a 2×2 pooling operation is applied to a 4×4 matrix. The filter moves with a stride of 2 both horizontally and vertically. Max pooling selects the highest value from the portion of the matrix covered by the filter. Similarly, average pooling applies the same process but takes the average value instead. Pooling layers do not aim to extract features from the input image as convolutional layers do. They are used to reduce the

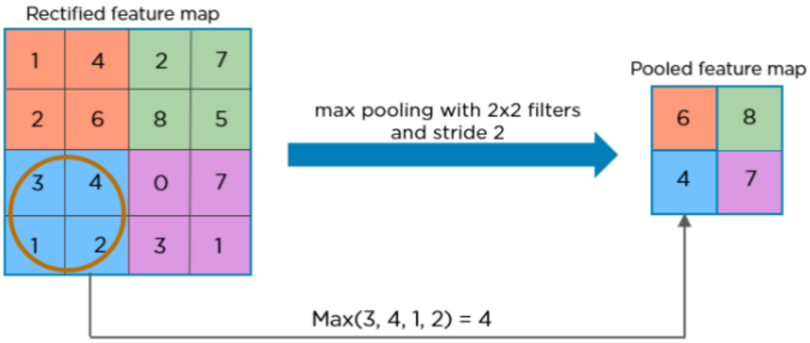


Figure 6.6: Illustration of how a max pooling layer acts. The filter has 2×2 dimension and is moved on the input matrix with a horizontal and vertical stride of 2. The max pooling select the maximum of the values covered by the filter, reducing the dimensions of the input matrix. The same process applies for the average pooling.[39]

dimensionality by filtering out the image.

6.3 Deep learning for disruption prediction

The disruption prediction is a promising application of deep learning in nuclear fusion. In this problem, the relevant diagnostics are organized into an input matrix I of dimensions $M \times N$, where M represents the number of diagnostic channels, and N represents the number of time points at which the data are collected. The input matrix can be normalized appropriately, and I is treated as a greyscale image representing a snapshot of the shot's history. The problem of disruption prediction is then reduced to a binary classification problem, where the CNN is trained to distinguish between disruptive and non-disruptive shots. In this approach, the key task is to select the data that carries most information related to the disruption. This is not a trivial challenge, as disruptions are highly non-linear phenomena that occur over very short timescales. The optimal combination of data for a

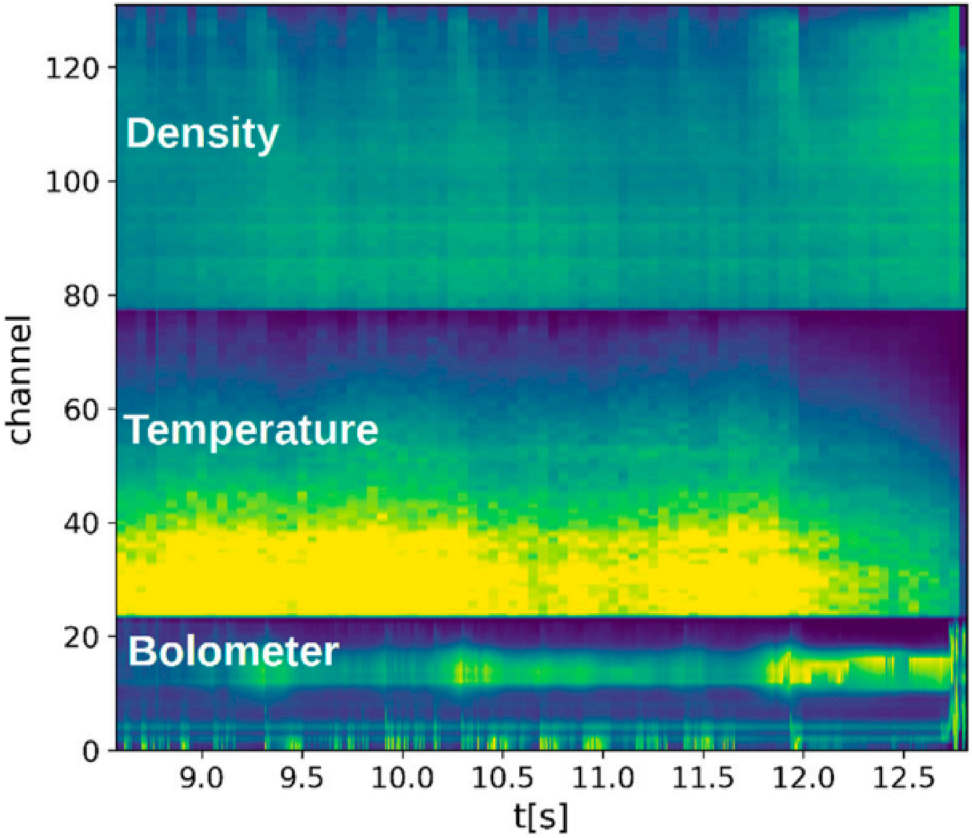


Figure 6.7: The channels of the selected quantities, measured at different times are stacked in the y direction, producing an input image, where the x direction corresponds to the temporal dimension. In this example, the signal from the bolometer, the temperature, and the density are shown.

disruption predictor can be found following the knowledge built by the physical modeling of disruption and its precursors or with a data-driven, empirical analysis to maximize the predictor's performance. In the Ref. [6] the authors proposed a CNN due to their ability to extract spatiotemporal features analysing plasma 1-D and 0-D quantities. An example is proposed in Fig. 6.7 where the data from the horizontal bolometer, the temperature, and the density profiles are stacked to create a temporal snapshot of the image. Then, a database of disrupting and non-disrupting pulses is created, and the training process described in 6.2 is performed. The CNN produces an output between 0 and 1 for a given input, representing the likelihood of disruption. The data used to perform the classification are the electron density, the electron temperature, the horizontal bolometer, the mode locking signal, and the internal inductance. This choice is noteworthy because it is connected to physical quantities that are closely related to the disruptions or their precursors. For example, the density is related to the density limit, the temperature to the temperature quench but also to the flattening due to the effect of the magnetic island, which is a precursor of disruption. The radiation detected by the bolometer is related to the contraction of the temperature profile in the first phase of the disruptions due to a density limit. Finally, the internal inductance l_i and the ML signal are particularly relevant, the first one as emerged by a statistical data-driven analysis [50], and the second as precursor index.

The one proposed in Ref. [6]is not the only possible approach. Recent work has obtained encouraging results using more complex neural network architectures, implementing a memory able to account for temporal relation between phenomena [50] or very recent results built on transformers [102].

6.4 eXplainable Artificial Intelligence

In this thesis, an extension of the work carried out by E. Aymerich et al. [6] will be presented. The work addresses the problem of the neural network's reliability, investigating the patterns it follows to classify disruptions. This is part of an essential problem regarding how a neural network learns and how this can be investigated. The process of analysing and explaining how a neural network generates its output is referred to as eXplainable Artificial Intelligence (XAI). An XAI analysis is required to be adapted to the system, and a standardized approach does not exist. In case of Ref. [6], and in general, in a CNN, the information is passed in a manner that, in linguistics, is referred to as ostensive learning, meaning that information is provided directly (ostensed), accompanied by signals or labels (as in case of deep learning), that makes the association explicit. In this type of learning, the most straightforward way to determine whether information has been effectively absorbed is to evaluate if the subject can recognize the object by focusing on it when presented with a specific input. In the case of CNN for disruption predictions, an assessment of how it performs predictions is to investigate what are the most relevant parts of the input image and compare the results with the physical models to assess whether they are consistent with each other. Two techniques are exploited. The occlusion consists in perturbing the system by removing a part of the input image and studying how the output changes. The saliency map aims to compute the derivatives of the output with respect to each pixel of the input image. Here, the gradient is computed using the backpropagation algorithm, the same used for the training of the NN, described in Sec. 6.2. These analyses produce sensitivity maps, represented as matrices in which high values indicate areas where the output is more sensitive. These areas highlight the part of the input that the neural network has learned to be significant for performing classification. Comparing these parts with physical modelling can be used to assess whether the CNN learns to predict disruption following the same patterns as scientific investigation. Specifically, a subset of the test dataset, containing disruptions that follow a known disruption path, has been selected. The disruption path is related to modification of the temperature profile as consequence of impurities radiation that causes a collapse of the temperature profile at the edge, or an hollowing of the temperature profile at the centre. These chains of events are particularly suitable for analysis as they are easily detectable from temperature changes and mark the onset of the pre-precursor phase. Furthermore, this phase is a significant indicator in the triggering of disruptions and represents a common disruption pathway, allowing the database to contain a satisfactory number of related pulses. The work in the Appendix shows evidence of the fact that the CNN changes the way the part of the input related to the temperature affects the output when performing a classification of an edge cooling or a temperature hollowing. Specifically, the CNN is more sensible to the inner part of the temperature profile when a temperature hollowing occurs. On the other hand, the sensibility in the outer part of the temperature profile is increased during edge cooling. This demonstrates that, without any prior hints during the training phase about edge cooling and temperature hollowing, the CNN learns to maximize prediction accuracy by changing the sensibility to different parts of the profile, depending on the regions affected by the occurring phenomena. The work does not aim to provide indications to improve CNN's performance but just to check whether CNN performs its classification following common patterns, such as those in physics. This can contribute to increasing knowledge about how CNN performs its classifications, enhancing the trustworthiness and credibility of such a predictor. In principle, wider analyses could produce useful results to optimize the choice of signals, showing which are the most relevant to perform the classification. Furthermore, applying XAI analysis can help unveil hidden relationships between data that the CNN leverages to perform the classification. This could suggest causal connections between phenomena that affect the input quantities, guiding the physical modelling. The detailed work is reported in Appendix B.

Conclusion

The problem of the neoclassical tearing mode instabilities has been addressed in this thesis. Specifically, three different works have been presented, each one guided by a general question: “what is the phenomena involved in the trigger of the (N)TM?”, “what is the best strategy to control the (N)TM by injecting ECW waves after the detection?” and in case of uncontrollable magnetic island “how the neural network disruption predictors learn to predict disruptions?”. The problem of the controllability of the NTM is faced in Chapter 4, where predictive work is performed on DTT. The evolution of the NTM is simulated in the full power scenario, first assuming a free mode evolution and then introducing a stabilizing system based on the current design specifics. The simulations highlight different stabilizing strategies: in early stabilization, the island detection is harder due to the small island size, but even without perfect alignment, stabilization is possible. On the other hand, with a late intervention, island detection is easier, but small displacements of the absorption position of the EC waves from the O-point reduce their stabilization. This work will contribute to the design of the NTM stabilizing system of DTT.

This thesis addresses two experimental cases involving changes in the temperature profile, focusing on the relationships between (N)TM triggers and disruptions: edge cooling and temperature hollowing. The former is connected to a decrease in the value of the temperature profile at the edge, while the latter refers to a hollow of the temperature on the plasma axis.

At the end of JET discharges, edge cooling events can trigger a 2/1 tearing instability, which likely leads to plasma disruption. The conditions for triggering 2/1 have been analyzed. A detailed analysis shows that, in such cases, the linear stability term does not overcome the stabilizing curvature effect (Green-Glasser-Johnson term), and a critical island width must be overcome to trigger the mode. In Chapter 5, the ion polarization current is presented as a phenomenon that can favor the onset of the mode by reducing the island’s critical width, making the rational surface susceptible to fast, sudden fluctuations of the rotation frequency of the island with respect to the plasma. The analysis considers the specific disruption path of the edge cooling.

In Section 6.4, the explainable artificial intelligence (XAI) is introduced in the context of disruption predictors based on convolutional neural networks (CNN). The work in Appendix B analyses a database of disruptions connected with the deformation of the electron temperature profiles caused by edge cooling or temperature hollowing with XAI techniques. The results show that the CNN is more sensitive to the inner part of the temperature profile in the case of TH and to the outer part of the temperature profile in the case of EC. The interpretation for this behavior is that, in analogy to the physical investigation, the CNN learns to change its focus according to the phenomena occurring in the plasma to maximize the prediction performances. The work demonstrates for the first time the feasibility of applying XAI methods to disruption predictors, contributing to the enhancement of the reliability of deep learning algorithms in disruption avoidance. In conclusion, this thesis contributes to understanding neoclassical tearing modes, focusing on their onset, control, and disruptive behavior. The ultimate goal is to enhance the performance of future fusion power plants and reduce the risks connected to the unstable nature of the plasma.

Appendix A

Analysis of the role of the ion polarization current on the onset of pre-disruptive magnetic islands in JET

A.1 Introduction

The sudden loss of confinement of the plasma, which occurs during disruptions, may result from the onset of a low-mode-number helical magnetic instability, which is eventually responsible for a full collapse of the magnetic equilibrium topology in the plasma. However, the operation of modern tokamaks has been improved to the point that discharges can proceed unperturbed even in high-performance regimes and instabilities can be avoided by keeping the safety factor q large enough, up to exhaustion of the inductive flux. Still occasionally without easily identifiable causes (such as impurities influx and enhanced radiation), an apparently quiescent equilibrium condition, collapses on a short timescale in clear association with footprints of tearing modes onset, both classical and neoclassical (at finite β_p), occurring at q rational magnetic surfaces. The fast decrease in the magnetic field can induce strong currents that lead to mechanical stresses on the metallic vessel, which can damage the machine itself. As a consequence, the tokamak operability is usually limited in order to avoid disruptions. Disruptions occur following different sequence of events,whose details depend on the primary cause and, when a plasma control system is active, on the response to such events. The chain of primary causes and subsequent events is an object of active research Recently, Pucella et al. [80] identified disruption paths in JET related to the connection between heavy impurity dynamics and triggering of tearing modes. In particular, they identified two possible paths, connected to the change in the temperature profile. The first one (temperature hollowing) is related to the penetration of impurities, which cools down the plasma core by radiation, producing a hollowing of the temperature profile on the plasma axis. A second path (edge cooling) is connected to a local flattening of the temperature profile at the edge of the plasma caused by a loss of radiated energy associated with impurities. Sozzi et al. [101] analyzed a database of JET disruptions, creating a solid statistic that confirms that most pulses follow the two paths previously identified. As highlighted by De Vries et al. [110], accidental disruptions are in most cases preceded by the onset, the development, and the locking of a 2/1 magnetic island. A deep understanding of the phenomena that lead to a disruption must involve the study of the appearance of a 2/1 tearing mode (TM), which, possibly in conditions of overcritical β_p, leads to the trigger of a neoclassical tearing mode (NTM). Pucella et al. [80] found out that the change in the temperature due to the hollowing of the temperature or to the edge cooling (EC) leads to a linear destabilization of the 2/1 tearing mode. Despite this, the problem related to the trigger of the mode is not clear at all, as the linear stability does not completely determine the behavior of the instability. The non-linear Rutherford's theoty [86], predicts the onset of the mode when helical perturbation of the magnetic flux produces a so-called seed island for which the sum of destabilizing effects overcomes the sum of stabilizing effects. Within the different destabilizing effects, the ion polarization current (IOPC) is thought to be particularly crucial in the problem of the onset of the mode [82, 38]. A perpendicular current is produced as a result of the island's relative motion with respect to the ions. Then, because of the closure condition $\nabla \cdot J = 0$, a parallel current, called ion polarization current, appears, affecting the evolution of the island [78, 51].Following the generalized Rutherford's theory, the contribution of the polarization current can be either stabilizing or destabilizing according to the propagation frequency ω of the island in the plasma frame with respect to the electrons and ions diamagnetic frequencies [66]. Furthermore, the effect of the ion polarization current is found in the evolution of an island with width W scales as $1/W^3$. This makes it a key actor able to strongly influence the evolution of an NTM in the initial part of its development. While in the low collisionality conditions, typical of the steady phase of operation, the effect of the ion polarization current is reduced [71] by a factor that scales the inverse aspect ratio ε as $\varepsilon^{3/2}$, so it is negligible; according to some recent kinetic approaches [28], destabilizing event such as the EC, could induce a decrease of the temperature on the rational surface, increasing the collisionality and the contribution of the ion polarization current, which could become destabilizing. In this work, an analysis on the non-linear phenomena, which affects the evolution of the magnetic island according to the generalized Rutherford theory, is performed over a database of JET disrupting pulses. In Sec. A.2.1, a description of the model employed is given. In Sec.A.4 the results of the analysis are shown. The first part of the analysis is focused on the destabilizing effect of the bootstrap current and on the stabilizing effect of the curvature. Then, the ion polarization current is taken into account for a subset of pulses. Eventually, in Sec. A.5, the analysis is summarized in order to identify a mechanism able to explain the trigger of the mode and a stability criterion.

A.2 Physical picture

In general, the ion polarization current is a current generated by the differences in the inertial response of the ions and electrons. Here, we provide a physical picture connected to the neoclassical geometry.The intrinsic $E \times B$ motion of the magnetic island contributes to generate an electric potential, which influences the motion of the particles outside the island [78, 51]. Fig. A.1 represents a cartoon of what the particles experience: in the blue region, the potential due to the island motion generates an $E \times B$ drift in the opposite direction of the island rotation, while in the green region, the drift velocity points in the same direction. The interaction between the trapped and passing particles in this field generates the neoclassical polarization current. The width of the banana orbit is proportional to the Larmor radius $\rho_{i,e}$ so that the velocity of a particle lying in the green region during the bouncing back will decrease due to the $E \times B$ drift, reducing the Larmor radius and then the width of the banana. On average, this corresponds to an outward drift caused by the electric potential generated by the movement of the island. The same, but in the opposite direction, will happen for the particles lying in the blue region. Collision between the trapped and passing particles will generate a current perpendicular to the magnetic island, pointing outward in the green region and inward in the blue region. Collision between trapped and passing particles will generate a current perpendicular to the magnetic island,

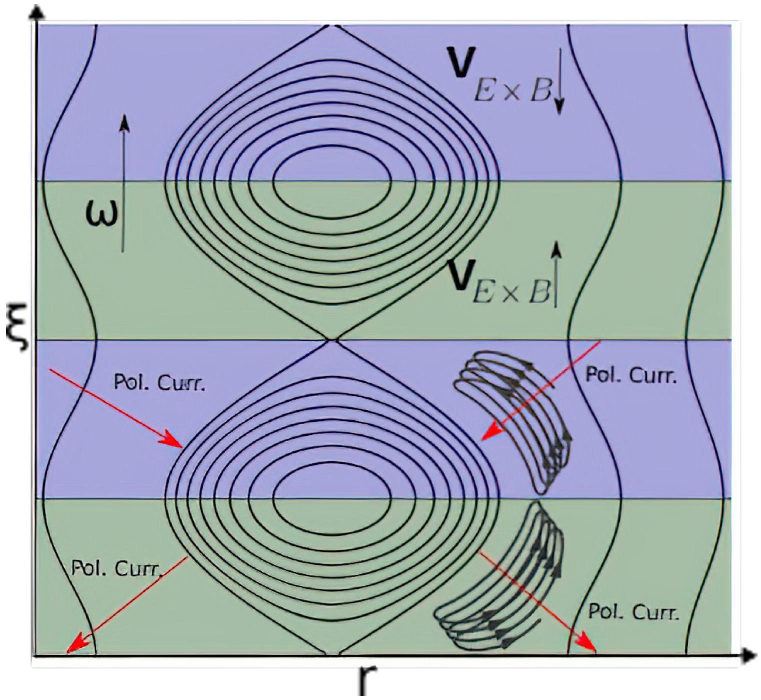


Figure A.1: Illustration of the physical picture that leads to the onset of the ion polarization current. The motion of the island generates a potential that affects the neoclassical transport, producing the ion polarization current.

pointing outward in the green region and inward in the blue region. This current, called *polarization current*, comes from the movement of trapped particles, proportional to the Larmor radii $\rho_{i,e}$. Given that $\rho_e << \rho_i$, the contribution of electrons is very small, so this current is called *ion polarization current.* A current parallel to the magnetic field is generated in order to preserve the local quasi-neutrality close to the island. This parallel current affects the evolution of the island.According to the physical picture we provided, the strength of the IOPC must depend on the collisionality and rotation frequency of the island in the frame of reference of the plasma. The former contributes to increasing the efficiency with which the trapped particles transfer momentum to passing particles. The latter is proportional to the electric field and the $v_{E\times B}$ that causes the drift of the trapped particle.

A.2.1 Generalized Rutherford Equation

The evolution of the magnetic island is described by the Rutherford theory[52, 12]. The non-linear behavior of a magnetic island of width W is related to the currents parallel to B. Performing a change of coordinates, from a tokamak frame of reference Performing a change of coordinates, from a tokamak frame of reference $(r(\psi), \theta, \phi)$ to a helical coordinate system $(x, \zeta(\theta, \phi))$ such that:

$$\zeta = \theta - \phi/q_s \tag{A.1}$$

$$x = r - r_s \tag{A.2}$$

where q is the field line winding number. We can integrate the Ampere's law around the island obtaining [28]:

$$\frac{g_1 \tau_r}{r_s^2}\frac{dW}{dt} = \Delta_0' + \int_0^{2\pi} d\zeta \int_{-\infty}^{+\infty} J_{\|}(\psi) dx \tag{A.3}$$

Here g_1 is a constant, $\tau_r = \mu_0 r_s^2/(1.22\eta)$ is the resistive timescale, r_s is the position of the rational surface $q = m/n$ (with m, n integer). The term Δ_0' is the linear stability term, calculated as the logarithmic jump of the perturbed linearized flux $\tilde{\psi}$ across the rational surface, computed using the current density profile. The second term on the right-hand side represents the non-linear component of the evolution of the magnetic island. The integral is performed across the rational surface. The function $J_{\|}(\psi)$ can be split into three contributions: the bootstrap current, the Green–Glasser–Johnson (GGJ) term, related to the curvature of the plasma, and the ion polarization current. By computing the integral in Eq. A.3 for every contribution to the total $J_{\|}(\psi)$, we obtain a form of the Generalized

Rutherford Equation [66, 71, 52, 12, 94, 97, 88]:

$$\frac{g_1 \tau_r}{r_s^2}\frac{dW}{dt} = \Delta'_0 + \Delta'_{boot} + \Delta'_{GGJ} + \Delta'_{pol} \tag{A.4}$$

$$\Delta'_{boot}(W) = a_{boot}\beta_p\sqrt{\varepsilon}\left|\frac{L_q}{L_p}\right|\frac{W}{W^2+W_d^2} \tag{A.5}$$

$$\Delta'_{GGJ}(W) = -a_{GGJ}\beta_p\varepsilon^2\frac{L_q^2}{r_s|L_p|}\left(1-\frac{1}{q^2}\right)\frac{1}{\sqrt{W^2+kW_d^2}} \tag{A.6}$$

$$\Delta'_{pol}(W) = a_{pol}g(\varepsilon,\nu_{ii})\beta_p\left(\frac{L_q}{L_p}\right)^2\frac{\rho_{\theta i}^2 W}{W^4+W_\rho^4}f(\omega) \tag{A.7}$$

$$f(\omega) = \frac{\omega(\omega-\omega_{*i}-\omega_{*T})}{\omega_{*e}^2} \tag{A.8}$$

$$g = \begin{cases} \varepsilon^{3/2} & \nu_{ii}/\varepsilon\omega < \varepsilon^{-3/4} \\ \varepsilon\nu_{ii}^2/\omega^2 & \varepsilon^{-3/4} < \nu_{ii}/\varepsilon\omega < \varepsilon^{-3/2} \\ 1 & \nu_{ii}/\varepsilon\omega > \varepsilon^{-3/2} \end{cases} \tag{A.9}$$

where $\beta_p = 2\mu_0 p^2/B_\theta^2$ is the poloidal component of β, ε is the inverse aspect ratio and $L_{<*>} = \frac{<*>}{d<*>/dr}|_{r=r_s}$ is the characteristic length of the pressure p and the safety factor q profile. Equation A.4 represents the Generalized Rutherford Equation [66, 12, 88]. The coefficients a_{boot}, a_{GGJ} and a_{pol} are constants, which can be set in order to model the data. In this work, the values of $a_{boot} = 1.7$, $a_{GGJ} = 6$ are chosen following Zohm [120] and Sauter [91]. They also proposed a value for $a_{pol} \sim O(1)$, however, it comprehends the dependence on the rotation frequency. In our model, the rotation frequency is taken into account in the function $f(\omega)$, which varies in a wide range ($f(\omega) = 10-200$) as shown in Fig. A.8. To model suitably the IOPC effect, we choose $a_{pol} = 0.01$, so that $a_{pol}f(\omega) \sim O(1)$ consistently with Zohm and Sauter. The term [66, 52] defined in Eq A.5, represents the destabilizing effect caused by the reduction of the bootstrap current inside the magnetic island. As the island grows, the heat diffusion tends to produce a flattening of the temperature and pressure profiles in the part overlapping the island. The flattening of the pressure reduces the bootstrap current, thus producing a destabilizing effect that boosts the growth of the island [91]. The electron temperature evolves on a fast parallel temperature diffusive equilibration timescale, with a diffusivity $\chi_\parallel$ much larger than $\chi_\perp$, characterizing the slower perpendicular diffusive transport. In the island region, the balance of the parallel and perpendicular heat flows [35, 88] leads to a temperature diffusion equation characterized by a scale length $W_d \propto (\chi_\perp/\chi_\parallel)^{1/4}$. The term [66, 43] Δ'_{GGJ}, defined in Eq A.6, represents the stabilizing term related to the curvature of the plasma inside the torus. This term is modelled heuristically so that it goes as $1/(kW_d)$ for $W << kW_d$ avoiding the singularity for $W \to 0$. The constant k is introduced in order to allow Δ'_{GGJ} being greater than Δ'_{boot} in the limit of small island. This is important because $\Delta'_{boot} \sim \sqrt{\varepsilon}$ while $\Delta'_{GGJ} \sim \varepsilon^2$, in which the former is intrinsically larger than the latter and the constant k yields models where the islands become unstable for larger values of width, according to the experiments. The term[66, 97, 95], in Eq A.7, Δ'_{pol}, represents the effect of the IOPC. The magnitude of its contribution depends on the regime of collisionality, according to the function $g(\omega, \varepsilon, \nu_{ii})$ with $\nu_{ii} = 4.810^{-8} Z^4 n_i \ln(\Lambda) T_i^{-3/2}$ [15] and $\ln(\Lambda)$ the Coulomb logarithm. The formulation, proposed by Mikhailovskii[71] (Eq. A.9), admits three regimes: a low collisionality regime, where the Δ'_{pol} is reduced by a factor $\varepsilon^{3/2}$ due to the reduced interaction between ions flowing outside the island and the plasma, which inhibits the generation IOPC. A mid-collisionality regime followed by a high-collisionality regime where the IOPC is increased by the strong interaction between the ions and the plasma. The limits of the regimes depend on the collisionality ion–ion ν_{ii}, the inverse aspect ratio ε and the rotation frequency of the island ω in the plasma frame of reference. The effect of the ion polarization current can be both stabilizing and destabilizing according to the sign of the function $f(\omega)$, shown in Eq. A.8 and defined by Smolyakov [94]. ω_{*e} and ω_{*i} are respectively the electron and ion diamagnetic frequencies defined as [17]:

$$\omega_{*j} = -\frac{mk_b T_j (dp_j/d\psi)}{e_j p_j q} \tag{A.10}$$

with $j = e, i$. n is the poloidal mode number, p'_j,e_j and n_j are respectively the radial derivative of the pressure, the charge and the density of the species j. The coefficient [95] $\omega_{*T} = k_\theta (dT_i/dr)/(eB_0)$ takes into account the toroidal velocity of the island in the $f(\omega)$, connected to the radial derivative of the ion temperature profile T_i and which is non-negligible at an early stage of development of the island. If $0 < \omega < \omega_{*i} + \omega_{*T}$ then $\Delta'_{pol} < 0$ and its contribution is stabilizing, otherwise it is destabilizing. Finally, the parameter $w_\rho = \mathrm{O}(\rho_{\theta i})$ is introduced in order to let the Δ'_{pol} vanish as $W << W_\rho$. This behavior can be explained heuristically by thinking that a current of ions in a magnetized plasma must have at least a width of $2\rho_i$, so that its contribution to the total longitudinal current of an island with $W << \rho_i$ can be neglected. This is also confirmed by theoretical models developed by Smolyakov et al. [94] where he investigates the contribution of the ion polarization current in the limit of $W << \rho_i$, showing that:

$$\Delta'_{pol} \sim \frac{1}{W}(\omega - \omega_{*e})(\omega - \omega_{*i}) \tag{A.11}$$

Other models [36, 34, 58] showed that in this limit $\omega \to \omega_{*e}$, so that the overall contribution of the ion polarization current cancels out. An acceptable transition formula from the small to large width limit can be expressed as [66]:

$$\Delta'_{pol} \sim \frac{W}{W^4 + W_\rho^4} \tag{A.12}$$

A.3 The role of the linear index Δ'_0

According to the model described in the previous Section, the stabilizing contribution of the curvature does not vanish when the island width goes to zero. If the Δ'_0 is positive, then the reconnection of a rational surface is energetically favorable, but this does not imply the onset of an instability. Indeed, the destabilizing strength of the linear term must overcome the stabilizing contribution of the curvature; otherwise, the magnetic island is non-linearly stable. Figure A.2 shows the behavior of the non-linear contribution with respect to the island width W, for an unstable pulse, at the onset of the mode. If the island is smaller than a critical width W_{crit} then the stabilizing contribution of the curvature prevents the trigger of the mode. In this picture, a linearly unstable rational surface reconnects,producing a magnetic island with $W < W_{crit}$ and then is non-linearly stable and the mode will not grow. The critical width defines how big the reconnection must be in order to produce an instability so that it depends on the linear combination of the linear and non-linear contribution in Eq. A.4. The value of W_{crit} is calculated by computing the smallest zero of the Generalized Rutherford Equation (Eq. A.4). In general, the value of the critical width depends also on the strength of the stabilization due to the curvature, the destabilizing contribution of the bootstrap, and the contribution of the ion polarization current. In order

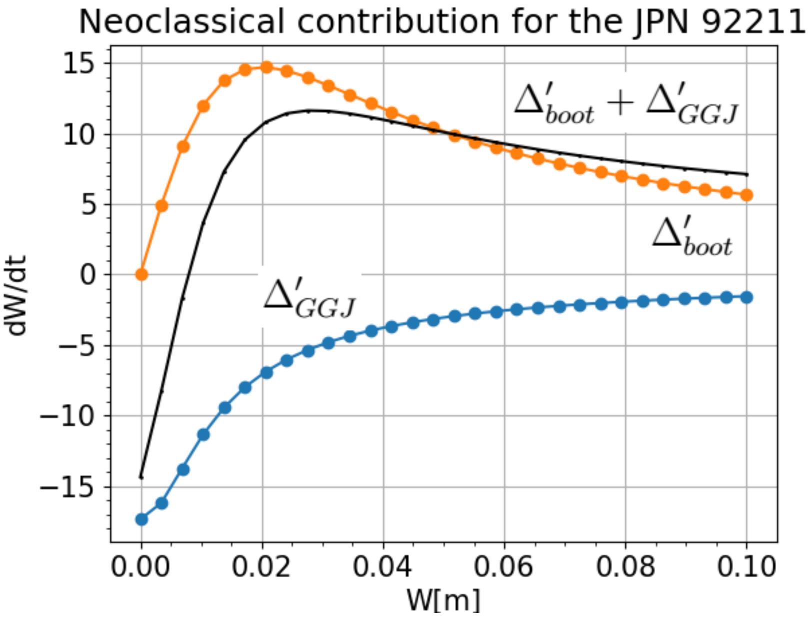


Figure A.2: The Figure shows how the value of Δ' changes for the different terms in the Generalized Rutherford Equation with respect to the island width for a real pulse (JPN 92211) where the coefficients are calculated at the onset of the mode. The orange trace represents Δ'_{boot},the blue trace Δ'_{GGJ} ad the black, their sum $\Delta'_{boot} + \Delta'_{GGJ}$ goes to zero for small width, so that there is a region $W < W_{crit}$ where the sum of the contributions is negative, so an island is stable.

to gain a deeper insight about the mechanism that leads to the onset of the mode, the analysis is extended to these terms.

A.4 Analysis of non-linear terms

In this section, a detailed comparison of the competing non-linear terms of the equation governing the growth of magnetic islands is presented, to identify the reason for the onset of the mode in a specific set of JET pulses. The Green–Glasser–Johnson and the bootstrap terms depend on equilibrium quantities (e.g., pressure, safety factor, and current density), so their contributions vary over an equilibrium timescale. On the other hand, the ion polarization current depends on the equilibrium quantities, but its dependence on the island rotation frequency causes the Δ'_{pol} to be susceptible to faster fluctuation, which could suddenly change the strength and the sign of the term in the Rutherford's equation. Although the evolution of the magnetic island is affected by both linear and non-linear phenomena, their effects sum up linearly, so that it is possible to study them separately. First, the analysis is performed by considering only the GGJ and the bootstrap current, and then, it is extended considering also the ion polarization current. The safety factor profile q and the value of β are taken from an EFIT equilibrium. The electron density and temperature are measured by the interferometer and radiometer diagnostic. The pressure is calculated as the product of the density and the temperature. In the termination phase, data from charge exchange diagnostic are not available because of the lack of neutral beam injection. Ion temperature is estimated from D-D reaction rate in order to evaluate the ions diamagnetic frequency ω_{*i}. The ion pressure is calculated from the temperature and the density, assuming $n_e = n_i$.

A.4.1 Curvature and Bootstrap

The analysis is performed over a database composed of 70 disrupting pulses, where the disruption is preceded by the trigger of a magnetic island after an edge cooling. The pulses are classified according to their value of q on the plasma axis. A reliable indicator of the value of q on the plasma axis is the sawtooth instability, so that its presence implies that $q_0 < 1$. Figure A.3 shows the position of the database in the plane q_{95} - l_i. Here l_i represents the internal inductance and q_{95} the value of the safety factor q at the 95% of the minor radius. This allows us to identify the pulses of the database as density limits [100]. The colormap shows that discharges terminating with sawtooth activity ($q_0 < 1$) can be subject to a disruption in the final phase of the current ramp down ($I_{onset}/I_{ft} < 0.6$). As pointed out before, the curvature and the bootstrap terms depend on equilibrium quantities, so, in order to evaluate their effect on the trigger of the mode, it is important to compare, for a certain equilibrium, their respective strength. To do that, the ratio of $k_{boot} = \beta_p \sqrt{\varepsilon} |L_q/L_p|$ to $k_{GGJ} = \beta_p \varepsilon^2 L_q^2/(r_s|L_p|)(1-1/q^2)$ is used as a figure of merit. The

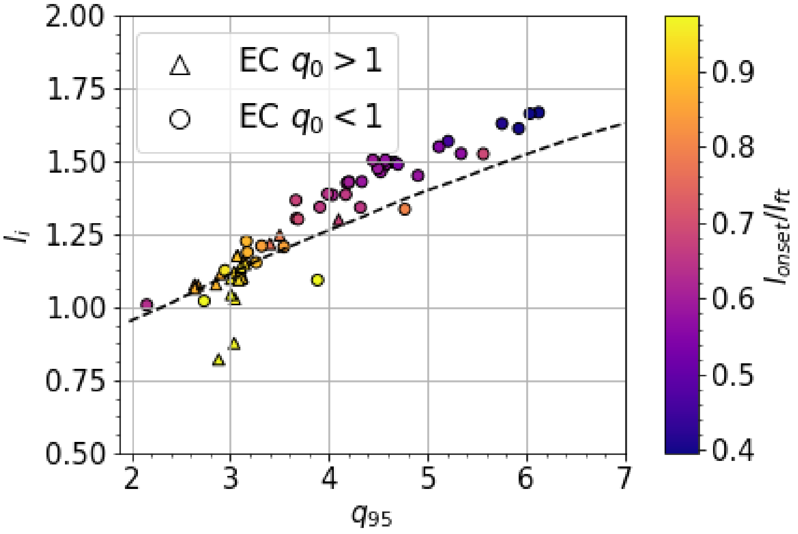


Figure A.3: Position of the pulses of the database in the plane $l_i - q_{95}$. The dotted line represents the Wesson density limit. The colorbar shows the ratio between the current at the onset and the current during the pulse. This is a measure of how far the onset happens from the stable plasma operation

parameter $\Theta \equiv k_{GGJ}/k_{boot}$ quantifies how much the coefficient for the stabilizing Green-Glasser-Johnson term is smaller than the destabilizing term of the bootstrap current. A

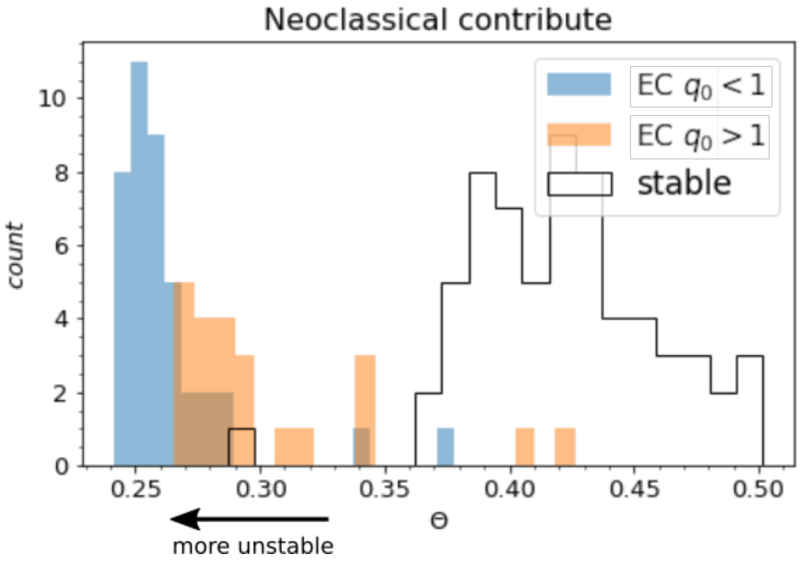


Figure A.4: The Figure shows the distribution of the value of Θ for the pulses contained in the database. The blue and the orange bars represent pulses with edge cooling, with the value of q on axis respectively less and greater than one. The white bars represent the distribution of Θ for the disrupting pulses, calculated during a stable phase of the pulse.

smaller ratio means a more unstable plasma, while a larger ratio means a more stable plasma. It is important to highlight that we are not taking the width of the island into account. A reduction in the ratio does not imply the onset of the mode, but rather a plasma that is more susceptible to the development of instability. The profiles are evaluated at the position of the rational surface, and Θ is computed for every pulse at the onset of the mode. The Figure A.4 shows a histogram of the distribution of the ratio for the pulses in the database. The distribution Θ examined during a steady phase of the pulses is larger than the same values tested before the onset of the mode. This is consistent with the physical explanation we provided for the parameter Θ: as the parameter decreases, the destabilizing bootstrap term gets larger in comparison with the stabilizing curvature term, increasing the likelihood of the onset of the mode. The two populations of pulses are clearly separated in the figure, showing that for the pulses with q_0 less than 1, the equilibrium configuration corresponds to a lower ratio and more unstable plasma. The magnetic islands are triggered in a more unstable condition, suggesting that the sawtooth contributes to stabilizing the mode. This is consistent with Fig. A.3 where the modes in a plasma with $q_0 < 1$ are triggered later in the termination.

A.4.2 Polarization current

The change in the temperature profile due to an edge cooling is found to provide a slight linear destabilization [80], which is related also to a reduction of the stabilizing contribution of the curvature (GGJ) with respect to the destabilizing bootstrap contribution as shown in the previous Section. In the present section, the possible effect of the edge cooling on the ion polarization current is investigated. The impact of the edge cooling changes the equilibrium profiles, resulting in a contribution of the curvature (GGJ) and bootstrap terms favouring instability. On the other hand, the change in the temperature profile directly modifies the contribution of the IOPC by changing the diamagnetic frequencies contained in the $f(\omega)$. In the termination phase, the decrease of the temperature implies an increase in the collisionality, which is taken into account in the Generalized Rutherford Equation by means of the function $g(\omega, \varepsilon, \nu_{ii})$, reported in Eq A.4. The Figure A.5 (bottom) represents the value of the function $g(\omega, \varepsilon, \nu_{ii})$, close to the onset of the mode, for the pulse 91977. Just before the onset, the collisionality (Figure A.5 top) modifies the value of Δ'_{pol} increasing its contribution. The exact value of the Δ'_{pol} depends on the value of the rotation ω of the island with respect to the plasma. A measure of the ω in the laboratory frame of reference, can in principle be obtained using the Mirnov coils signals. Then, in order to obtain the rotation frequency with respect to the plasma, we have to subtract the intrinsic $E \times B$ [56, 51, 17, 96] motion of the plasma, so that:

$$\omega = \omega_{Mirnov} - \omega_{E\times B} \tag{A.13}$$

$$\omega_{E\times B} = \frac{-nv_{\phi_j}}{2\pi R} + \frac{n\nabla p_j}{2\pi Z_j n_j R B_\theta} \tag{A.14}$$

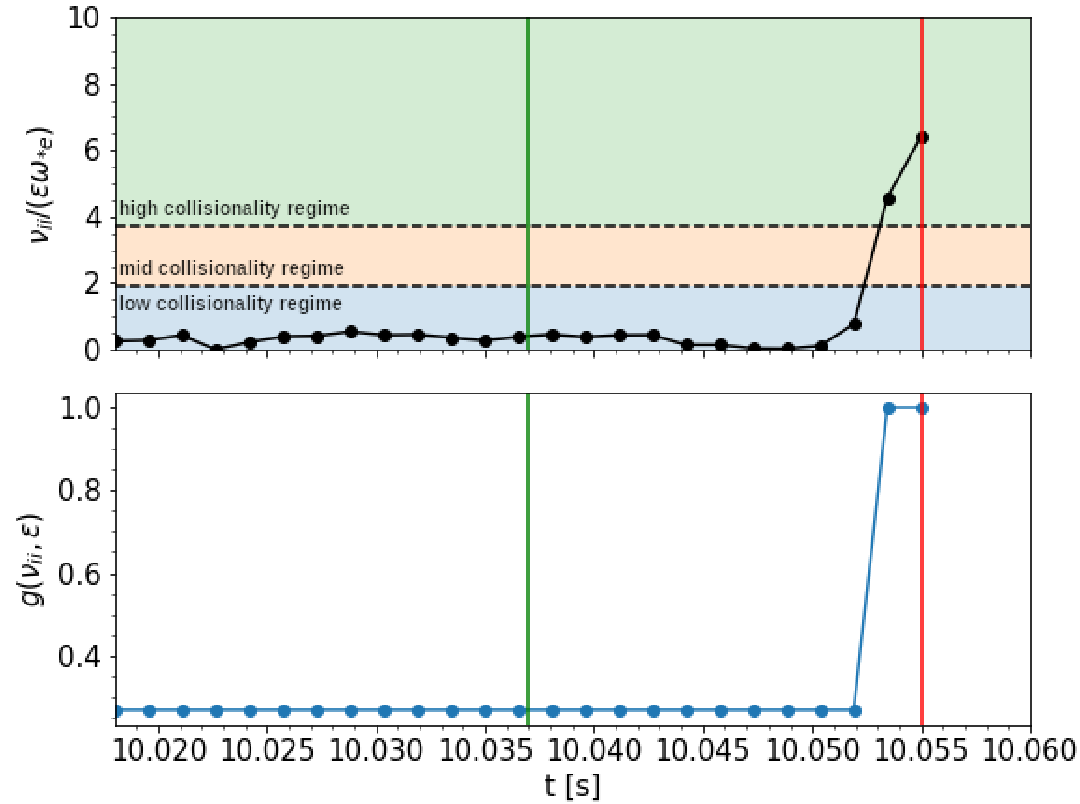


Figure A.5: (Top) Time trace of the function $\nu_{ii}/\varepsilon\omega$. This function determines the value of $g(\mu_{ii},\omega)$ (Bottom).

where $v_{\phi j}$ is the toroidal velocity, n is the poloidal mode number, ∇p, e_j and n_j are the radial derivative of the pressure. The index j refers to the ion specie which is used as reference for the measure of the toroidal velocity by the charge exchange diagnostic [51]. During the current ramp down, the shutdown of the NBI makes the charge exchange diagnostic not available, so that a theoretical approach is needed in order to evaluate the island rotation. According to the theory [111, 36], the island moves in the plasma close to a value which is the viscosity weighted mean of the unperturbed electron and ion velocities. Then it has been shown that, for small island, the island frequency tends to be close to the electron diamagnetic frequency [34] and this is also confirmed by numerical models [58]. So the island rotation will be assumed $\omega = \omega_{*e}$ and the analysis is focused on the factor $f(\omega = \omega_{*e})$. Under this assumption, the nature of the IOPC is always destabilizing for small island. This is consistent with some recent theoretical works [29]. The diamagnetic frequencies are calculated using the ionic and electronic pressure profiles measured by the diagnostics. In the majority of the pulses in the database, the ion temperature is assumed to be equal to the electron temperature. Generally speaking, this is a reasonable approximation while the species are at equilibrium, but during the termination phase, it may be a rough approximation that hides the differences between the diamagnetic frequencies. For this reason, the pulses of the database are filtered in order to neglect the pulses where

pulse	B_t[T]	I_{onset}/I_{ft}	$\rho_{\theta i}$[cm]	$min(w_{measured})$[cm]
91977	2.69	0.95	0.88	2.92
92296	2.78	0.90	1.10	5.76
96745	2.77	0.95	0.73	4.72
91970	2.71	0.89	0.95	2.10

Table A.1: The toroidal magnetic field on the axis, the ratio between the current at the onset of the mode and the current at the flat top and the value of the poloidal ionic Larmor radius are reported to allow a comparison between the pulses. Furthermore, we reported the first measured width

$T_e = T_i$. From the starting database, we selected four disrupting pulses. The Table A.1 reports the characteristic quantities of the selected pulses. In particular, it is noteworthy that the resolution of the Mirnov coils allows us to detect the island only when the island W is larger than the ion Larmor radius $\rho_{\theta i}$, so that the Mirnov measures cannot be exploited to estimate the island rotation frequency at the very beginning of the island evolution. Along with these, we also selected three pulses, which develop an edge cooling without developing a tearing mode. In order to provide a complete physical picture of the pulses that we analyzed, we present here a description of the behavior of the pulse 91977. The plasma current is 3 MA and the toroidal magnetic field 2.69 T. At the beginning of the ramp down of the current ($I_{onset}/I_{flat-top} \simeq 0.95$), the pulse experiences an edge cooling, that starts at 10.037s. The onset of the mode occurs at 10.055s. The spectrogram of the Mirnov coils signals (Figure A.6) shows the magnetic perturbation caused by the 2/1 mode and by the 1/1 mode (connected to the sawtooth crashes). Figure A.7 shows the evolution of the island evaluated using the magnetic perturbation measured by the magnetic coils, according to the expression [79]:

$$W = 4\sqrt{\frac{R_0 r_s}{n B_T s_s}\frac{1}{2}\left(\frac{r_c}{r_s}\right)^{m+1}\tilde{B}_\theta} \tag{A.15}$$

where n is the toroidal mode number, B_T, r_s, R_0 and s_s are the toroidal magnetic field, the position of the rational surface, the major radius and the magnetic shear evaluated by the EFIT equilibrium. $\tilde{B}_\theta$ is the poloidal perturbed magnetic field. The expression A.15 is valid for modes with $m \geq 2$ and under the assumption that the position of the coils r_c coincides with that of the wall. The first measure of the amplitude of the magnetic field perturbation provides an island width of 2.92 cm. During its growth, the mode slows down, as it is evident by the reduction of the frequency in the spectrogram, which is reported in Figure A.7 (grey line). The mode grows till 30 cm, when locks at 10.064 s and disrupts the plasma at 10.112 s. The Figure A.8 represents the time traces of the factor $f(\omega_{*e})$ for the selected pulses. The behaviour of $f(\omega_{*e})$ is similar for the safe pulses. In particular, we have $f(\omega_{*e}) \simeq 2$ meaning that $|\omega_{*e}| \simeq |\omega_{*i}|$. This represents the thermal equilibrium value, when $T_e \simeq T_i$. In the unstable pulses, the function $f(\omega_{*e})$ shows strong

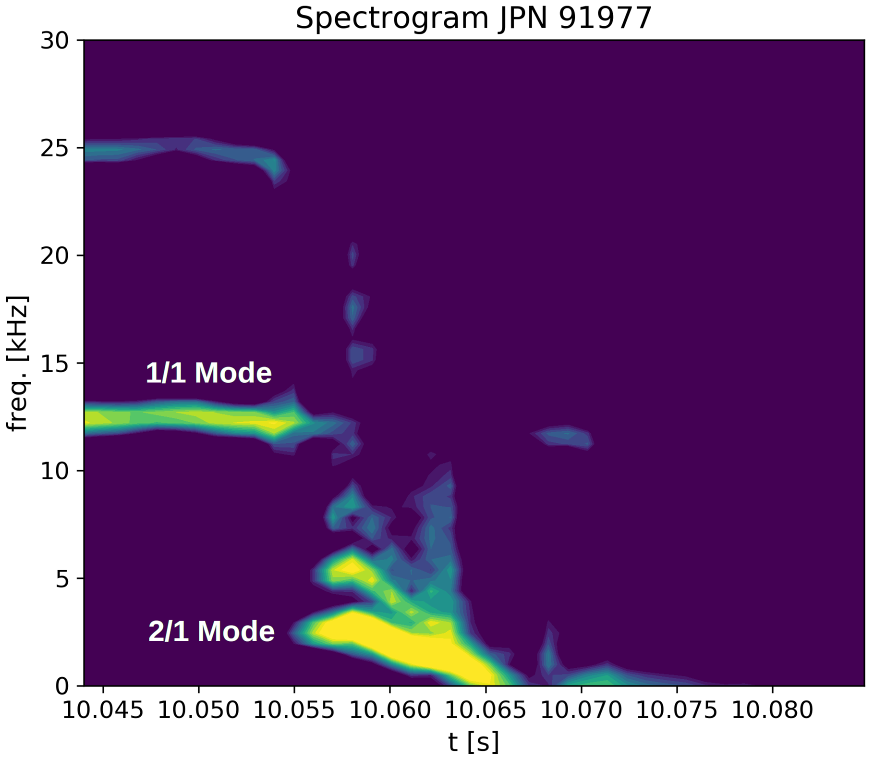


Figure A.6: Frequency spectrogram of the signal of the Mirnov coils versus time. The 2/1 mode can be seen starting from $t \simeq 10.05$ s, when it slows down till the locking at $t \simeq 10.064$. It is also possible to distinguish the magnetic activity of the $1/1$ mode, connected to the sawtooth crashes.

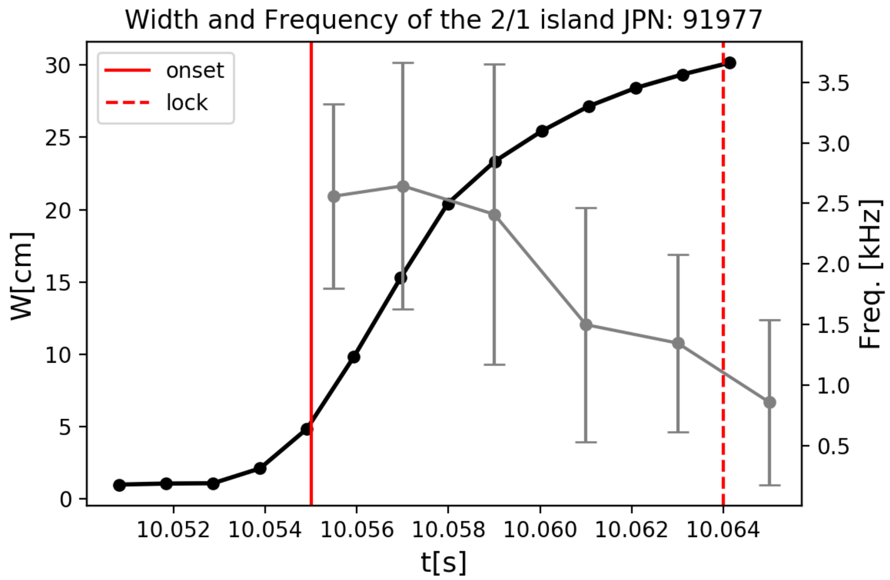


Figure A.7: The Figure shows the evolution of the island width (black line) for the disrupting pulse 91977. The mode starts at 10.055 s leading to disruption. The grey line represents the rotation frequency of the mode measured by the Mirnov coils.

increment of the function before the onset. The peak of the function $f(\omega_{*e})$, corresponds to a stronger destabilizing effect of the IOPC at that time. The pulse 91970 is the only one which does not follow this behaviour: the value of $f(\omega_{*e})$ oscillates and there is no particular increment between the edge cooling and the onset of the mode. The values of $f(\omega_{*e})$ at the onset are higher with respect to the equilibrium value for three pulses out of four. The reason why the parameter $f(\omega_{*e})$ is so high is the reduction of the electron diamagnetic frequency compared to the ion diamagnetic frequency as a consequence of the edge cooling. The results suggest that the effect of the edge cooling is to increase the destabilizing effect of the ion polarization current. It is important to highlight that any kind of fluctuations of the temperature profile could correspond to a fluctuation of the IOPC, which could pass very quickly from stabilizing to destabilizing. In general, this analysis suggests that the ion polarization current represents a key contribution, which leads to a more unstable configuration and suddenly increases the probability of onset of the mode.

A.5 Critical width

The theory of the (neoclassical) tearing mode introduces the idea of critical width: given a seed island of width W_{seed}, the W_{crit} is that width such that if $W_{seed} > W_{crit}$, then the seed island becomes unstable. W_{crit} can be seen as an indicator of the stability condition of the plasma because the smaller the critical width, the simpler it is for a perturbation to cause an instability. The critical width is calculated as the zero of the generalized Rutherford

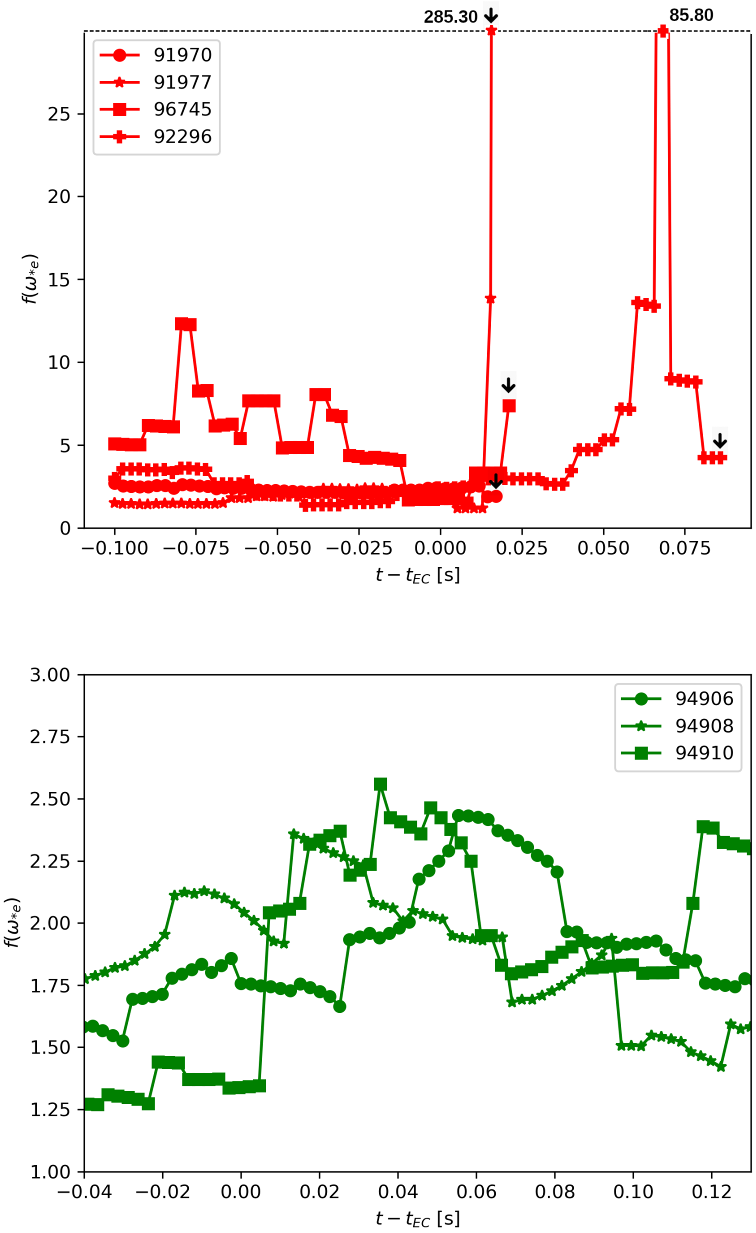


Figure A.8: Comparison between the $f(\omega_{*e})$ versus the time distance from the edge cooling, for the selected pulses. Top: the red traces show the behaviour for the unstable pulses. The arrow shows the time of the onset of the mode. Bottom: the green traces represent the behaviour of the $f(\omega_{*e})$ for the stable pulses. It is interesting to highlight that before the onset, we have a peak of the function $f(\omega_{*e})$, corresponding to an increment of the destabilizing effect of the ion polarization current, while for the stable pulses $f(\omega_{*e} \sim 2)$

equation, evaluated by summing the contributions of the curvature (Δ'_{GGJ}), the bootstrap (Δ'_{boot}), the ion polarization current (Δ'_{pol}) and the linear contribution (Δ'_0). For a given

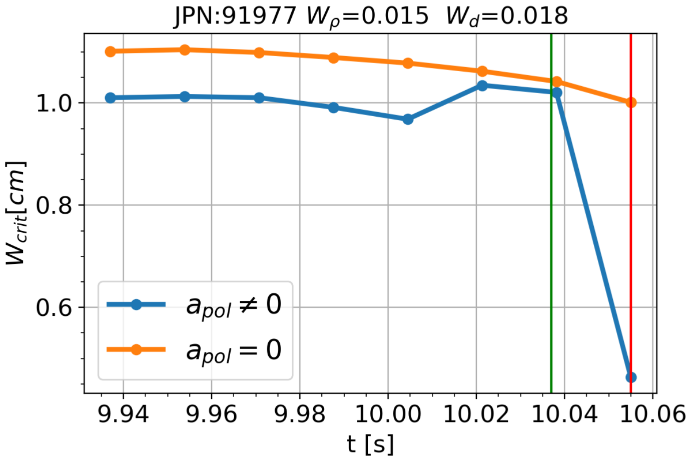


Figure A.9: The Figure shows the time trace of the critical width of the pulse 91977, taking in to consideration the effect of the ion polarization current (blue trace) and neglecting it (orange trace).

time t, the critical width $W_{crit}(t)$ is calculated by evaluating the dW/dt for every width and taking the width such that $dW/dt = 0$.

In order to highlight the effect of the ion polarization current on the critical width, Fig. A.9 shows the value of the critical width vs time,for a real pulse, considering and neglecting the ion polarization current. The effect of the ion polarization current is to reduce the critical width just before the onset of the mode. Figure A.10 shows the trace of every single contribution for a fixed width ($W = 0.005 \sim W_{crit}$ in this case). The sign of the sum (black line) determines if the mode is stable or unstable. If the sum is positive, then the mode is unstable, otherwise is stable. The linear contribution Δ_0' can be calculated using the equilibrium current density profile. However, the time resolution needed in this phase of the pulse is high, so the data are not resolved enough to follow the evolution of the mode in the edge cooling timescale. In order to highlight the contribution of the non-linear terms, the plasma is taken to be linearly stable so that $r_s\Delta_0' = -m$. The trigger of the mode is determined by an increase in the destabilizing effect of the IOPC and a decrease in the stabilizing curvature effect with respect to the bootstrap current. The island rotation is fixed to be equal to the electron diamagnetic frequency, as we are focusing in the first phase of the evolution of the mode; therefore, the island can be considered in the small width limit described in Section A.4.2. The ion polarization current dominates the curve for small W ($W_\rho \sim W < W_d$) and its effect is to reduce the W_{crit} from $W_{crit}(a_{pol} = 0) = 1.01$cm to $W_{crit}(a_{pol} \neq 0) = 0.47$cm. The same analysis is carried out over the selected pulses and the results are summarized in Table A.2.

The reduction of the critical width is more evident for higher values of $f(\omega_{*e})$, while it is small for the 91970, where the destabilizing contribution of the ion polarization current is weaker. The analysis is performed over a restricted subset of pulses due to lack of data, but some general behavior can be inferred. The model shows that for fluctuations of the

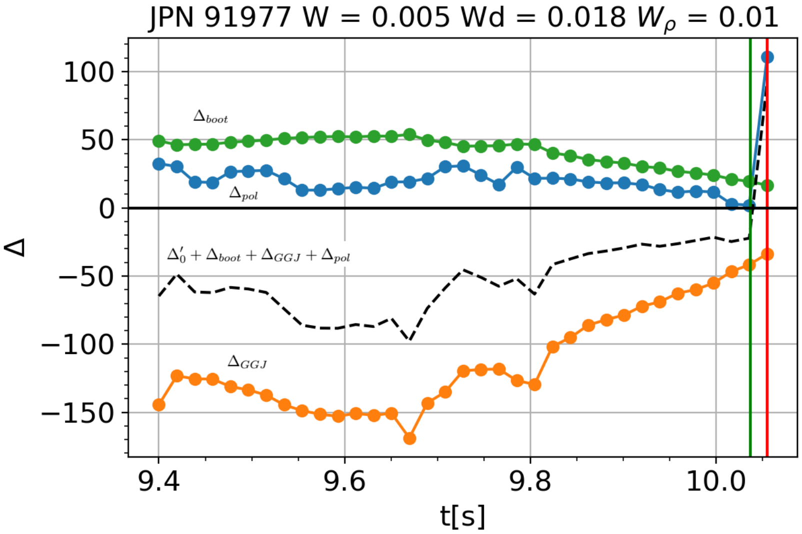


Figure A.10: The Figure represents the time trace of the contributions of the Generalized Rutherford Equation for the 91977. The vertical green and red lines represent respectively the time of the edge cooling and the time of the onset of the mode.

pulse	$w_{crit}(a_{pol}=0)$ $[cm]$	$w_{crit}(a_{pol}\neq 0)$ $[cm]$
91977	1.01	0.47
92296	1.18	0.60
96745	1.00	0.35
91970	0.99	0.89

Table A.2: Summary of the critical width, calculated with and without the contribution of the ion polarization current.

ion polarization current, a consequence of the flattening of the temperature profile due to the edge cooling is that the value of the W_{crit} decreases, leading to a more unstable condition for the 2/1 rational surface. The effect of the fluctuation of the IOPC is higher when the flattening due to the EC gets closer to the rational surface, as Δ'_{pol} depends on local quantities, such as the rotation and the collisionality. This can be verified for the other pulses, despite the lack of data in the termination phase. Defining a threshold κ and a flattening width such as:

$$r_{flattening} = min(r \in (0,1) \text{ s.t. } \left|\frac{dT_e}{dr}\right| < \kappa) \tag{A.16}$$

we can infer an onset criteria based on the fluctuation of the ion polarization current from that analysis. Figure A.11 shows the position of the rational surface (red line) and the width of the flattening (black lines) as defined in Eq A.16 at the onset of the mode, for the pulse 91977. In particular, it has been shown that a flattening of the temperature profile reduces the value of the electron diamagnetic frequency, increasing the destabilizing contribution of the Δ'_{pol}. This happens when the flattening of the rational surface causes a consistent fluctuation of the temperature, namely, when $r_s > \gamma r_{flattening}$. Here $\gamma \sim O(1)$ is a constant that takes into account that the flattening does not necessarily have to reach the rational surface for the temperature fluctuation to be strong enough to modify the contribution of the ion polarization current and then trigger the mode.

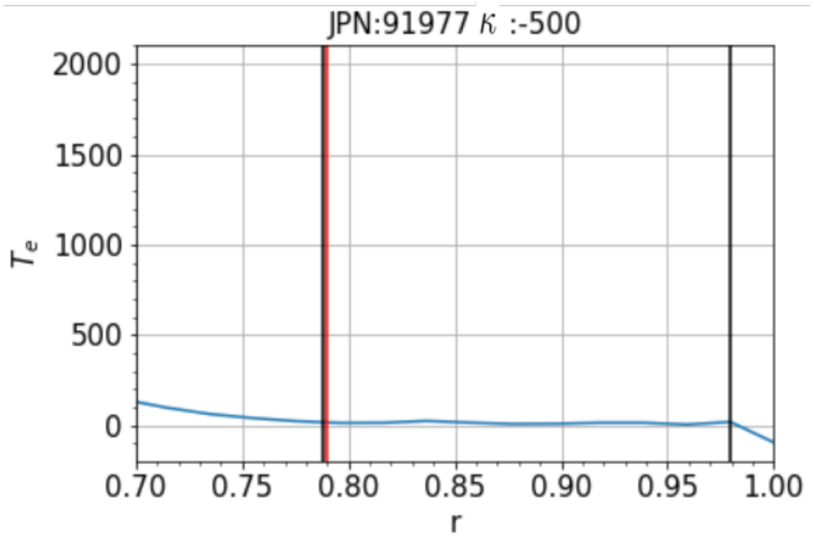


Figure A.11: Example of the algorithm used to calculate the position of the flattening compared to the position of the rational surface. Here, the temperature profile at the onset of the mode for the pulse 91977 is shown. The black vertical lines show the flattening width. The red vertical line represents the position of the rational surface.

Figure A.12 shows the position in this plane for the pulses contained in the database. The dashed black line represents $r_s = r_{flattening}$,so, according to the interpretation given

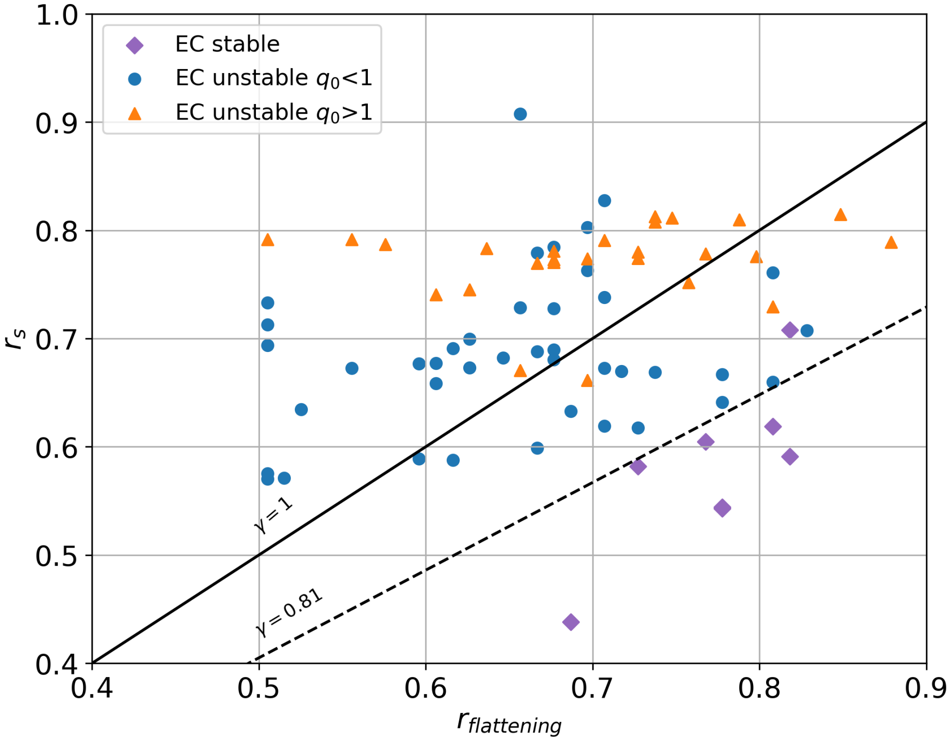


Figure A.12: Position in the plane $(r_{flattening}, r_s)$ of the edge cooling pulses of the database. The purple dots represent the edge cooling which do not trigger a mode. The blue and orange dots represents respectively the edge cooling with $q_0 < 1$ and $q_0 > 1$. The position is calculated at the onset of the mode for the disrupting pulses and by taking the temperature profile with the wider flattening within 200 ms (median time of onset according to ref. [80]) after the edge cooling for the purple dots.

by this model, we can see that the most of the pulses stay in the region $r_s > r_{flattening}$. The threshold obtained for $\gamma = 0.81$, splits the plane in two parts, so that every analyzed pulses overcomes the threshold and then develops a neoclassical tearing mode. The stable pulses (purple dots) do not overcome the threshold except for one pulse that slightly exceeds the dashed line. Figure A.12 shows that, in these cases, the flattening of the temperature profile does not reach the position of the resonant surface, reducing the fluctuation of the ion polarization contribution and, according to the analysis, preventing the onset of the mode.

A.6 Summary and conclusion

The stability of a tokamak equilibrium with respect to the tearing perturbations depends on the linear stability index Δ'_0 and the non-linear terms (the curvature and the bootstrap term) in the Rutherford equation. The equilibrium configuration determines the trigger of the mode affecting the Δ_0', Δ_{GGJ} and Δ_{boot}. As shown in Section A.3,the linear stability index does not completely describe the mechanism of triggering a mode because the stabilizing effect of the curvature could prevent the onset of the mode also for positive value of Δ_0'. It is more suitable to think in terms of the critical width, which can be seen as a quantity proportional to the likelihood of triggering a mode. In this picture, the role of the Δ_0' is to reduce the value of the critical width to increase the probability of onset of a mode. On the equilibrium timescale, the value of the critical width is represented by a balance between the Green–Glasser–Johnson term, the bootstrap term, and the linear term. However, the value of the critical width can be affected also by faster fluctuations, which change the rotation frequency of the island, leading to a more unstable plasma due to the ion polarization current. Here, a database of disrupting pulses is considered. The analysis of the neoclassical phenomena (GGJ and bootstrap), in Sec. A.4.1, pointed out that, for the analysed pulses, the ratio $\Theta \equiv k_{GGJ}/k_{boot}$ is lower at the onset of the mode than during a stable phase (Fig. A.4). This is consistent because it corresponds to a decrease in the stabilizing effect of the curvature with respect to the destabilizing effect of the bootstrap current, thus raising the probability of triggering a mode. Then, the analysis performed over a subset of selected pulses in Sec. A.4.2 highlights that fluctuations of the temperature profiles, in consequence of the edge cooling, lead to an increase in the destabilizing effect of the IOPC by increasing the value of $f(\omega)$. The interpretation proposed for a small subset of pulses has been generalized and used to predict a general behavior for every pulse of the database, which is consistent with the results. In fact, it has been shown in Fig. A.12 that, after an edge cooling, a pulse tends to develop a mode whenever the flattening, as a consequence of the edge cooling, reaches the position of the rational surface. At this point, the function $f(\omega)$ increases due to the reduction of the electron diamagnetic frequency, confirming that the ion polarization current has a key role in the onset of the mode. The analysis has been carried out by treating every single contribution of Eq. A.4 separately. This can be done due to the form of the model we are using: the effect of every contribution, though non-linear, sums up linearly. Even though the validity of these results is restricted to a specific case of onset (i.e., in termination phase, after an edge cooling), the analysis suggests a new general explanation for the trigger of the mode. A noisy fluctuation of the $f(\omega)$ could lead to a reduction of the critical width at some time. This corresponds to a condition where it is more likely for a random resonant helical perturbation to produce a seed island larger than the critical width, leading to an instability. The IOPC introduces randomness [11] in the problem of the onset of the mode, such that it is more suitable to define the stability using a probabilistic approach.

Acknowledgments

This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grant

Agreement No 101052200 EUROfusion). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

Data availability statement

The data that support the findings of this study are available from JET-Eurofusion. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the authors upon reasonable request and with the permission of JET-Eurofusion.

Appendix B

eXplainable artificial intelligence applied to algorithms for disruption prediction in tokamak devices

B.1 Introduction

Tokamaks facilities rely on a combination of magnetic fields to confine the plasma. An important role is played by the magnetic field generated by a net current toroidally flowing in the plasma. To achieve efficient energy production in a fusion reactor, the plasma must be maintained for a sufficient amount of time, much larger than the characteristic energy confinement time. The plasma is sensitive over different spatial and time scales to perturbations that can give rise to instabilities that destroy the magnetic configuration on very small timescale. These phenomena, called disruptions, cause a sudden interruption of the plasma current that in turn induces strong electromagnetic forces in the metallic vessel and in the surrounding structures. Furthermore, the disruption process generates non-thermal relativistic electrons, called run away electrons, that can damage the first wall of the machine. Due to the intrinsic non-linearity of the phenomena involved in a disruption, it is difficult to model the interactions that lead to the termination of the plasma discharge, however, it is possible to study processes that precede a disruption to identify and avoid disruptions before they happen. ([26]) perform a complete survey over a database of JET disruptions, identifying the chains of events, such as human errors in pulse management, MHD instabilities, like internal kink modes or most importantly, neoclassical tearing modes (NTM). In metallic wall machines, it is possible to identify the chain of events ([80]), related to the ingress of impurities or loss of density control, that determine the onset of an NTM and leads to disruption. Due to its fast, non-linear nature and the big range of phenomena that can trigger a disruption, it is difficult to set up a system which is accurately able to predict and avoid a disruption. Various studies have highlighted promising applications of deep learning in the field of nuclear fusion research ([76], [31], [32]). The

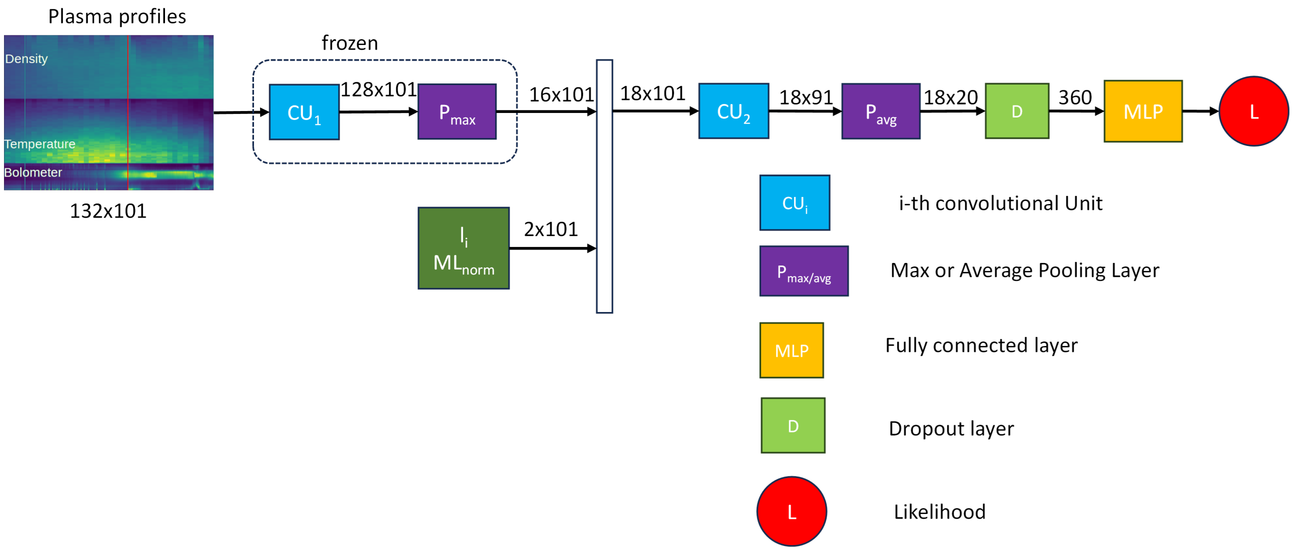


Figure B.1: Final CNN architecture, where the internal inductance (li) and the normalized locked mode (MLnorm) serve as inputs to the second convolutional unit. These inputs are concatenated with the output image generated by the max-pooling layer.

use of convolutional neural network (CNN) architectures have shown great potential for disruption prediction. This can be exploited to monitor phenomena that lead to disruption, (e.g. the locked modes ([33])) or trained specifically to predict disruptions both using raw data from a specific tokamak ([21]), or that can be multi-machine ([60, 118]). The use of deep CNNs proves to be especially well-suited for the analysis of plasma profiles. In ([6]) ([5]), the authors propose the use of deep CNN for the early detection of disruptive events at JET, utilizing both images constructed from 1D plasma profiles and 0D time signals. The predictors exhibit high performance also comparing them with those of other machine learning algorithms ([7]). The use of CNN allows learning relevant spatio-temporal information straight from 1D plasma profiles, avoiding hand-engineered feature extraction procedures. The CNN from ([6]) is adopted in the present paper to showcase the ability of eXplainable AI (XAI) methods to interpret the network prediction, and its architecture is detailed in the following Section B.2. The spread of deep learning algorithms depends on the trust that the scientific community has in these tools. One of the main causes of scepticism is that it is not possible to provide an explanation, neither in the testing phase nor in the training phase, of why a neural network produces a certain output. This issue becomes even more important when we are dealing with algorithms that are responsible for preventing and mitigating disruptions. The eXplainable Artificial Intelligence algorithms aim at providing an interface between humans and AI, producing results that explain the behaviour of the neural network in a comprehensible way to humans ([41, 3]). An XAI analysis is a very flexible tool that strongly depends on the algorithm used. There are XAI approaches, the agnostic algorithms, that are generally applicable independently on the kind of AI and other techniques that requires to be built ad hoc on the AI system. In dealing with CNN, the XAI algorithms provide a visual explanation, by producing heatmaps, related to the input image, that show the most relevant part of the input in order to classify the image. This work aims at addressing the problem of explaining how a neural network classifies a disruption, trying to fill the knowledge gap between CNN prediction and physical insights/interpretations. The application of XAI algorithms to CNNs in the problem of disruption prediction offers three main advantages. The first one is that there must be consistency between the explanations offered by XAI and the physical models. This consistency is essential to assert that the algorithm is genuinely learning to predict disruptions. So an analysis showing that the reason why an algorithm predicts a disruption is the same as the physical models, contributes to increase the trustworthiness of the NN. The second reason is that XAI might be able to provide indications about which signals are more useful for prediction, suggesting how to improve the performance of the CNN itself. The third reason lies in the unveiling of the CNN's prediction process, enabling the identification of novel data patterns that may have eluded conventional physical investigations. This, in turn, offers valuable insights for the development of new physical models. In this paper, we will start analysing a CNN trained to distinguish between disruptive and not disruptive input data frames and compare the results with physical classification of the disruptions, comparing how the CNN handles different disruptive paths. The description of the physical paths is provided in B.2.1. In Section B.2.1, the CNN and training and test database are explained. The database has been analysed and reduced distinguishing

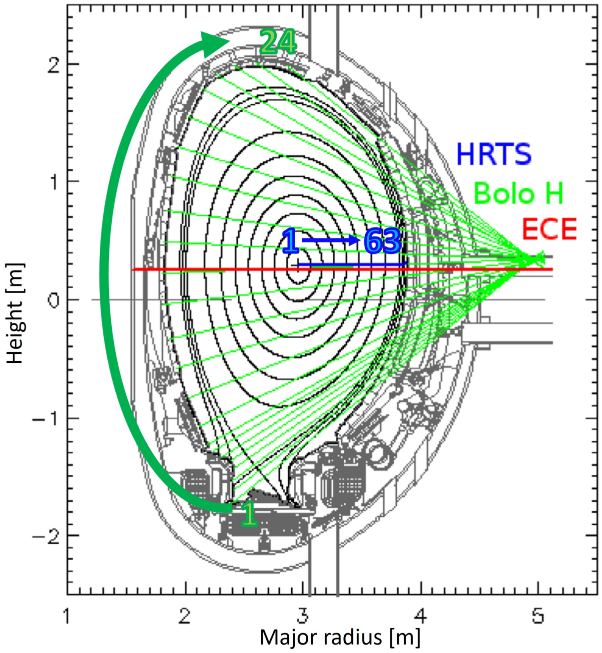


Figure B.2: JET poloidal cross section: the 24 chords of the horizontal bolometer camera (in green) and the 63 lines of sight of the HRTS (in blue) are numbered according to the order in which they are taken to construct the image of the profiles in input to the CNN.

between discharges following the two different paths. Two XAI methods are introduced in Section B.3. Section B.4 reports about the results provided by the two methods, and the results will be discussed and compared in Section B.5.

B.2 The architecture of the neural network

The increasing use of deep learning in research is driven by improved computer processing power, allowing for the analysis of large datasets. Deep neural networks, known for their high accuracy even without complex feature extraction of the input data, play a key role in this. In image processing, Convolutional Neural Networks (CNNs) are widely favoured for their effectiveness in handling complex image data. Supported by these significant capabilities, Aymerich et al. ([6]) proposed the use of CNNs for extracting spatiotemporal features from JET 1D plasma profiles (density, temperature, and plasma radiation), by converting them into 2D images. Particularly, density and electron temperature from high-resolution Thompson scattering (HRTS) are pre-processed to synchronize time scales and eliminate outliers. Furthermore, in reference to Figure B.2 showcasing the HRTS's 63 lines of sight in blue, lines from the 54th to the 63rd were excluded because of their inclination to generate unreliable data, due to the outboard position. Concerning plasma radiation, in the same Figure B.2 the 24 channels of the JET bolometer horizontal camera are depicted in green. Also, data from these chords undergoes the pre-processing steps mentioned for HRTS data. Three spatiotemporal images are created, with each pixel representing the measurement at the corresponding line of sight and time sample. These images are vertically stacked and normalized based on the signal ranges in the training set, producing an ultimate image. Figure B.3 reports two showcases referring to two JET pulses. The generated images present, in a top-to-bottom sequence, density and temperature data from 54 lines of sight measured by the HRTS, along with radiation data from the 24 chords of the horizontal bolometer camera. In total, there are 132 channels, and the data is presented over time. The final image is segmented using an overlapping sliding window of 200 ms, yielding individual image slices of size 132x101. As the CNN operates as a supervised algorithm, we explicitly assigned labels to slices in the training dataset. Those belonging to regularly terminated discharges were labelled as 'stable.' In contrast, for disruptive discharges, the 'unstable' label was automatically assigned by detecting the pre-disruptive phase through the algorithm proposed in [4]. For balancing the two classes, the stable phases of disrupted pulses were not included in the network training set and the overlap durations of the sliding window were different for regularly terminated and disrupted discharges. Conversely, during the testing phase, a 2 ms stride was employed for all discharges, covering both regularly terminated and disrupted pulses. Leveraging these diagnostics, which often exhibit behaviours linked to the onset of destabilizing physical mechanisms like MHD precursors, a straightforward CNN disruption prediction model is firstly deployed. In addition to the aforementioned plasma profiles, ([6]) takes into account

0D diagnostic signals commonly used in the literature, specifically internal inductance and locked mode signals, as inputs for the disruption predictor. The internal inductance is indeed a crucial parameter because it provides information about the current profile within the plasma and it is known to be connected to the density limit ([99]). A higher internal inductance suggests a more peaked current profile, concentrated towards the plasma core, while a lower internal inductance indicates a more distributed or flat current profile. Moreover, at JET, mode locking indicates when a rotating (neoclassical) tearing mode locks with the external wall, which is closely followed by the disruption typically manifesting in the later stages of the disruptive process. JET provides a real-time mode locking signal. In ([6]) this signal has been normalized by the plasma current, as already done for disruption mitigation purposes. The normalized locked mode signal contributes significantly to the successful prediction of faster disruptions. The CNN architecture shown in Figure B.1, comprises a series of interconnected convolutional (CU) and pooling (P) blocks, linked by a nonlinear activation layer with a ReLU function. These blocks filter the input image both vertically (along the spatial dimension) and horizontally (along the temporal dimension), extracting essential features. These resulting features are fed into a fully connected multilayer perceptron neural network (FC), where the final SoftMax layer determines the likelihood of the input image slice belonging to either a regularly terminated or a disrupted discharge. To incorporate the two 0-D signals, the CNN architecture underwent modifications by introducing them downstream of the initial filter block. It's noteworthy that the first filter block was initially trained exclusively with 1-D diagnostic data, and its weights were subsequently frozen. In a subsequent training phase, both the second convolutional block and the FC block were trained using all plasma parameters. Note that the network architecture enables the separation of the two dimensions, spatial and temporal. Specifically, the first two blocks (CU1 and Pmax) filter solely across the spatial direction, while the subsequent two (CU2 and Pavg) filter exclusively across time. This facilitates the seamless concatenation of the 0-D signals (li and MLnorm) with the image features processed by the initial convolutional and pooling blocks, thereby preserving temporal synchronization. Figure B.1 illustrates the ultimate CNN architecture as presented in ([6]), showcasing the dimensions of input features for various blocks. Meanwhile, Table 1 provides a comprehensive overview of the corresponding parameters. The vertical kernel size for the convolutional and pooling blocks was designed considering a few constraints: a kernel size equal or larger than 24 would have been larger than the bolometer number of lines of sight, and a small size kernel would reduce the effect of the discontinuity between the stacked diagnostic images. The small kernel size (5x1) allows the network to still identify changes in the spatial dimension of the HRTS scattering profile. Regarding the time filtering, a similar operation was performed, due to the different time resolution of the diagnostics employed, the filter size has been chosen to mainly process the highest frequency signals (the bolometer data). To determine the pooling type, two networks were trained: one with only average pooling and another with only max-pooling. Analyzing their performances on both the training and validation sets, it was observed that average pooling exhibited lower performance compared to max-pooling. However, the max-pooling response proved to be overly sensitive to transient changes in the data time traces. Consequently, the max-pooling layer was retained for spatial processing (vertical pooling), while average pooling was chosen for temporal pooling (horizontal pooling). Testing of discharges that were not included in the training phase underscores the predictor's applicability across diverse operational scenarios.

CNN block	Dimension	Output feature size
CU1 (Filter size)	5×1	128×101
P_{max} (Pool size; Stride)	$8 \times 1; 8 \times 1$	16×101
CU2 (Filter size)	1×11	18×91
P_{avg} (Pool size; Stride)	$1 \times 12; 1 \times 4$	18×20
MLP (input layer)	$18 \times 20 = 360$	-
MLP hidden layer	360	-
MLP output layer (likelihood)	2	2

Table B.1: CNN architecture

In this paper, the described CNN predictor is considered to demonstrate the application of Explainable AI, aiming to enhance understanding and confidence in the decision-making process of the CNN predictor.

B.2.1 Database

The local balance of energy flowing into and out of a system determines its temperature profile. Impurities can break this balance by increasing the amount of energy that escapes as radiation. As a result, the temperature profile becomes susceptible to impurity penetration. The bolometer can provide an integrated measure of the radiative emission of the plasma. Strong radiation is associated with a loss of energy and so a decrease of the temperature. Usually, changes in the temperature profile are preceded by radiative losses measured by the bolometer, they depend on the distribution of impurity and density inside the plasma and can be categorized in two different ways: edge cooling (EC) and temperature hollowing (TH). Edge cooling is a collapse of the temperature profile at the edge, while temperature hollowing is a decrease of the central value of the temperature, often due to impurity accumulation on the plasma axis. A sketch of typical shapes of the temperature profile during an edge cooling and a temperature hollowing is depicted in Figure B.4. These events are known to linearly destabilize the 2/1 mode ([80]), creating a magnetic island that rotates with the plasma. As it grows, the island experiences drag forces that tend to slow down its motion until the island locks onto the walls, leading to a disruption. We analysed a database composed by 87 pulses, divided both in safe and disrupting, belonging to the train/test database of the neural network presented in Table 1

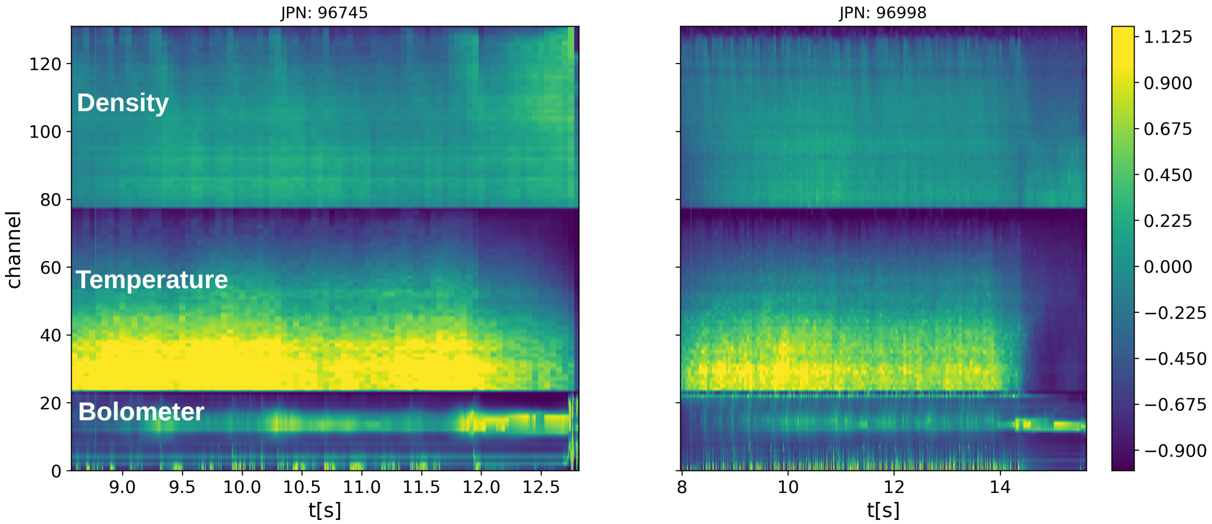


Figure B.3: The figure shows two illustrative cases contained in the database. On the left an edge cooling and a temperature hollowing on the right. The values on the colorbar are expressed in terms of normalized unit $[-1, 1]$

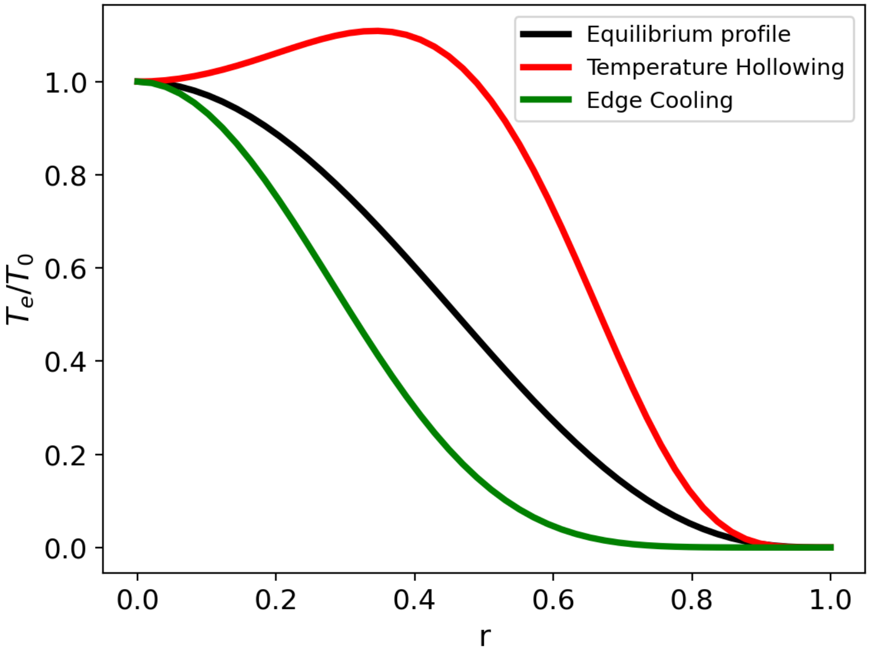


Figure B.4: Sketch of typical shapes of the electron temperature profile after an edge cooling (green line) and a temperature hollowing (red line), compared with an equilibrium profile (black line).

Type	Total	EC	TH	THEC
Disruptive	26	17	2	7
Safe	61	–	–	–

Table B.2: The table represents a summary of the different phenomena occurring in the pulses: the edge cooling (EC), the temperature hollowing (TH) and a temperature hollowing followed by and edge cooling (THEC). The safe pulses do not exhibit any of these phenomena.

of [6]. Our database is composed by pulses presented a specific disruption path, in particular, the disrupting pulses are preceded by an Edge Cooling (EC), a temperature hollowing (TH) or a combination of TH followed by an EC (THEC). Table B.2 shows the distribution in EC, TH and THEC. For our purpose, the pulses THEC are considered as pure TH. No indication regarding the EC and TH has been provided to the CNN in the training phase. The input data are composed by radiation profiles of the horizontal chords of the bolometer diagnostic and the radial profile of the electron temperature from the High Resolution Thompson Scattering and the electron density. Data from the different channels are converted in images and vertically stacked. It is possible to define the time $t_{EC/TH}$ at which the EC/TH starts by introducing indexes related to the shape of the temperature profile ([84]), measured using the radiometer diagnostic, and defining the start of the event by introducing a conventional threshold. For every pulse of the database $t_{EC/TH}$ has been measured. Two examples are shown in Figure B.3. In temperature and density, higher channels correspond to a more external part of the profile. The temperature profile is obtained using the Thompson scattering, which provides a measure of the electron temperature integrated along different lines of sight placed on the radial dimension. Examples of the different behaviour for EC and TH is shown in Figure B.3. The vertical red line represents the time at which the EC/TH occurs. On the left side, an edge cooling is presented. The collapse of the temperature at the edge is visible in the plot by the increase of the darker points in the region between the channels 60 and 80, which represent the outer part of the profile. The edge cooling starts with an increment in the radiated power, measured by the bolometer.

The plot on the right represents the input image for the neural network in case of a temperature hollowing. Here the hollowing of the temperature profile on the plasma axis occurs at $t = 54\ s$ and it is evident by looking at the channels between 20 and 40 that are the part of the temperature profile on the plasma axis.

B.3 The XAI techniques

In general, an XAI algorithm is an additional layer of analysis, built by the user on a given AI, in order to produce an explanation of the output, for a certain input. In this work, the XAI analysis is built over an existing CNN, described in Section B.2, trained to predict disruptions. The input of the CNN is composed by physical quantities, and the aim of XAI is to interpret which part of the input contributes the most in the classification of the image as disruptive or safe. This allows us to not only build a hierarchy of the most relevant physical quantities, but also to understand which part of the profile of a certain physical input quantity matters most. Various methods can be used to explore this issue. One approach is to analyse the sensitivity of the output when a perturbation is introduced at a certain point in the classification chain. This type of analysis, known as sensitivity analysis, produces a heat map ([109]) that shows which part of the input has the greatest impact on the output. We can employ this approach in two modes: agnostic or non-agnostic. In the agnostic mode, we don't delve into the behavior of the CNN's internal components (weights, gradients). Instead, we directly perturb the input and analyze the resulting output changes. Conversely, the non-agnostic mode involves analyzing the output's sensitivity with respect to the weights within the network's hidden layers. In Sections B.3.1 and B.3.2 we will explain the methods adopted, briefly presenting an example of the output produced.

B.3.1 Occlusion

The most straightforward agnostic approach is the occlusion ([40])([44]), where, as a perturbation, a constant value patch is applied in a certain part of the input and the effect of the patch on the output is analysed. We then interpret the fluctuation of the output as how much the part covered by the patch is important for the classification. Each input image for the CNN is made up of 132×101 pixels, as reported in Section B.2. Adopting the overlapping window approach to perform the occlusion is too computational demanding, because every time slice must be analysed for every possible position of the patch. Therefore, the global input is divided into M non overlapping temporal slices of dimension 132×101. To split the complete input in M sub-images, a zero padding p is introduced in order to ensure that $N_t = p + 101 \times M$ where N_t is the time length of the pulse. A patch of dimensions $W \times H$ is introduced in every slice, producing a perturbed image which is the same as the original except for the area covered. The patch replaces the value of the pixel, with a constant value V. The patch is moved with a horizontal s_h and vertical s_v step. The width, the height, and the vertical and horizontal steps define the number of positions that the patch can assume to perturb the output. N perturbed sub-images $I_{occl,k}$, with $k = 1, \ldots, k$, are obtained, where every image contains the patch in a different position. The occluded input $I_{occl,k}$ is passed to the neural network, resulting in an output $f_{NN}(I_{occl,k})$ with $k \in [1,N]$ where $f_{NN} : \mathbb{R}^{132\times101} \to \mathbb{R}$ is a function that represents the CNN. Then we define $\Delta_k = f_{NN}(I_{occl,k}) - f_{NN}(I)$ (I the original input) as the difference between the output of the occluded input and the original input. We define the fluctuation

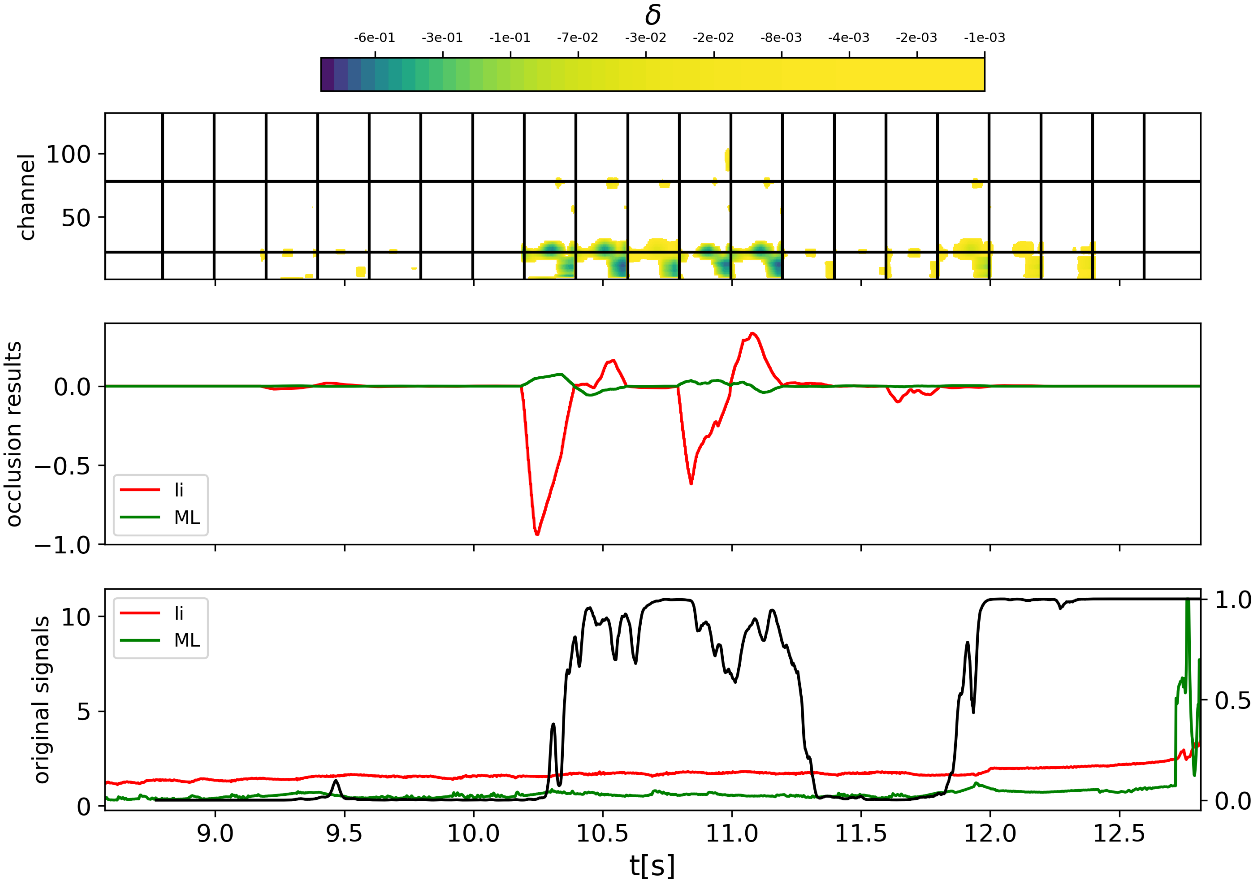


Figure B.5: Occlusion analysis for the JPN 96745. The heatmap as result of the occlusion technique for the 1-D signals (top) and for the 0-D signals (middle). The red line refers to the fluctuation as a consequence of the occlusion of the internal inductance, while the green line is for the ML signal. The third plot (bottom) represents the output of the neural network (black line), the signal of the internal inductance (red line) and the ML signal (green line) and the counting tensors δ_k and c_k as follows:

$$c_k = \begin{pmatrix} 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & 1 & \dots & 1 & \dots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & 1 & \dots & 1 & \dots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 0 \end{pmatrix} \tag{B.1}$$

$$\delta_k = \Delta_k c_k \tag{B.2}$$

These are 132×101 matrices having the same dimensions (number of pixels) of the original image, where the non-zeros elements have the same positions as the patch. The non-zero values of δ_k are the values of the fluctuation Δ_k. The matrix δ_k represents, for a given position of the patch, the pixels that, if occluded, produce the fluctuation Δ_k. The matrix c_k is built so that the sum over all the possibles k (positions of the patch), returns the number of times that a certain pixel (i, j) is covered by the patch:

$$N_{(i,j)} = \sum_{k=1}^{N} c_{(i,j)k} \tag{B.3}$$

Finally, we define the matrix δ as that matrix where every element (i, j) is the fluctuation Δ_k averaged over all the possible positions of the patch:

$$\delta_{(i,j)} = \frac{1}{N_{(i,j)}} \sum_{k=1}^{N} \delta_{(i,j)k} \quad \text{with } i \in (1,132), j \in (1,101) \tag{B.4}$$

The matrix $\delta_{(i,j)}$ represents the occlusion heat map for a single sub-image 132×101 . The occlusion depends on 5 free parameters related to the patch: the size (width W and height H), the value V and the stride (horizontal s_h and vertical s_v). These parameters define how the input is perturbed by the occlusion method. By applying this method to all the slices the occlusion produces an $132 \times N_t$ output. Since the monodimensional signals are treated as distinct inputs, their occlusion is also performed independently. A patch of constant value V is applied to the 0-D signal region, leaving the 1-D signals unaffected. This patch is moved along the horizontal axis with step size, s_h following the same algorithm as for the 1-D signals generating a $2 \times N_t$ matrix. An example is provided in Figure B.5: at the top, the heatmap for the 1-D signals is shown. Starting from the bottom, the image refers to the radiation, the temperature, the density. The colour intensity corresponds to the degree to which occluding a particular input feature affects the neural network's output. A fluctuation of -1 indicates that occluding that input feature reduces the network's output by 1, from 1 to 0. This matrix highlights the importance of each input feature for disruption classification. For visualization purpose, the fluctuation related to the signals is plotted in the plot in the middle, with the red line that refers to l_i and the green line that refers to the ML signal. The areas where the occlusion produces the strongest fluctuation are related to the bolometer and the central part of the temperature profile. The monodimensional signals, on the other hand, becomes important only in the transient phase, when the output of the neural network (black trace at the bottom) changes from stable to unstable or viceversa. The occlusion heatmap brought out an interesting behaviour: the neural network seems to be sensitive mostly to the right part of the input. This behaviour is shown in Figure B.5 (top), where the areas highlighted in the bolometer are asymmetric, with a reverse d-shape.

B.3.2 Saliency map

The previous method is coupled with a non-agnostic method to provide a more complete and general insight into the interpretation of the neural network. There is a wide variety of non-agnostic methods. Following ([93]), we define the saliency map as a matrix, made up of the derivative of the output of the neural network backpropagated to every single pixel of the input. Due to the particular architecture of the neural network we are studying, the gradient of the output will be backpropagated till the second convolutional unit, as indicated in Figure B.1. The second convolutional unit will produce as output a matrix $A_{(\alpha,\beta)}$. A backpropagation to the input of the neural network is not possible as the network is interrupted to add the monodimensional signals before the second convolutional layer. The first convolutional layer reduces the image size, but the ratio between the distances remains the same. Therefore, we can understand which areas of the input the saliency map output refers to by simply rescaling it. The saliency map will have the same dimension as the network layer, where every element (α,β) will be the partial derivative of the output with respect to $A_{\alpha\beta}(I)$ and is calculated as:

$$g_{\alpha\beta} = \max\left(\frac{\partial f_{NN}(I)}{\partial A_{\alpha\beta}}, 0\right) \tag{B.5}$$

where we have introduced the operator $\max(\bullet, 0)$, known as ReLU (Rectified Linear Unit), in order to filter out the negative values. The derivative is calculated with a guided backpropagation algorithm ([103]) that reduces the fluctuation of the gradient in presence of non-linear activation layer (e.g. ReLu). The definition in Equation B.5 must be adapted to the structure of the neural network that we are trying to analyse. In this case, the global input $I_{tot} \in M^{132\times N_t}$ is sliced in a set of 132×101 images by means of an overlapping sliding window. So that every pixel of I_{tot} appears in the sliding window 101 times. Given a certain pixel $a_{ij} \in I_{tot}$ with $1 \leq i \leq 132$ and $1 \leq j \leq N_t$ we define the set of all the slices containing a certain time j as:

$$I_j = \{I_k | a_{ij} \in I_k\} \tag{B.6}$$

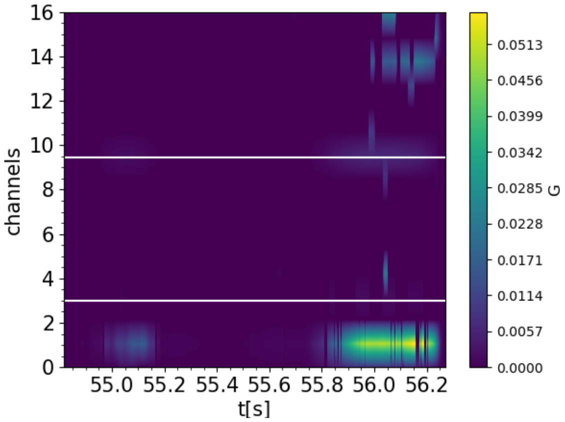


Figure B.6: Saliency map for the JPN 94966. The plot shows the matrix G as defined in Equation B.7. Larger value of G are connected to larger values of the gradient of the output with respect to the output of the neuron in the second convolutional layer.

where k is the time of the right edge of the overlapping time window. For a certain time j, we have $j \leq k \leq j+101$. Finally, we define the saliency map as:

$$G_{\alpha\beta} = \left\langle g_{\alpha\beta} \right\rangle_{I_j} \tag{B.7}$$

We averaged the single-frame saliency maps for all possible saliency maps that involve the pixel (i, j). The output of the saliency map is a heat map $18 \times N_t$. An example is provided in Figure B.6. Specifically, channels at the bottom are related to the bolometer $(0-3)$, then the temperature $(4-10)$ and the density $(11-16)$, while at the top there is the gradient with respect to the monodimensional signals $(17-18)$. Figure B.7, shows the function g of the saliency map for different phases of the pulse: a stable, a transient, an unstable phase. The value of the function g in the different regimes is noteworthy: when the output of the neural network changes, passing from stable to unstable, the sensitivity of the output becomes larger of around three order of magnitudes. The saliency map shows that the most sensitive part of the input are the radiation and the central part of the temperature, while the density seems to be of secondary importance and the monodimensional signals only relevant close to the trigger of the alarm.

B.4 Results

The sensitivity map and occlusion provide consistent results: there is a strong indication that the neural network relies mainly on the bolometer signal to make its predictions. The central part of the temperature profile is the second most important feature for the neural network, while 0-D signals play a role in the classification only near the alarm. The density seems to be of secondary importance for the disruption prediction. The saliency

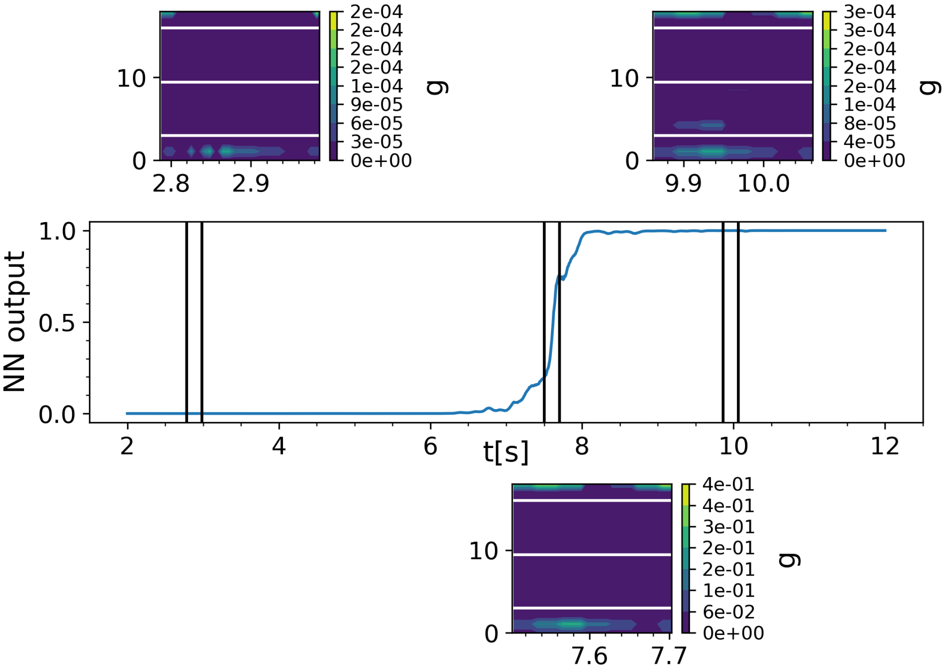


Figure B.7: Example of the matrix g as defined in Equation B.5, for different time. The plot in the middle represents the output of the neural network. The images show saliency maps in different times.

map tends to show a significant gradient in the bolometer even when there is no relevant signal in that area of the input. A comparison between different methods is shown in Figure B.6 and B.8. The heat maps turn on in the same moment, but in the saliency map the area of the bolometer is much more important than the occlusion. The temperature is relevant for the occlusion, even though a peak in the inner part of the temperature profile can be seen also in the saliency map. In both heat maps, there is an increase in sensitivity in the part connected to the density near the alarm, that anyway remains less relevant than the temperature and the radiation. The network does not recognize the change in the temperature profile that characterizes the EC/TH as the first event in the chain of phenomena that leads to disruption, as the alarm is often triggered before the EC/TH. This is the reason why the network usually is able to predict the disruption before the event of EC/TH. However, it is interesting to understand whether the NN is sensitive to the change in the temperature that physically is responsible for triggering the instability, as described in Section B.2.1. Since the occlusion technique includes several free parameters, it is suitable for analysing individual discharges, but there is the risk of not being able to obtain a uniform procedure when comparing different discharges. For this reason, a local systematic analysis has been carried out using only the saliency map approach near the

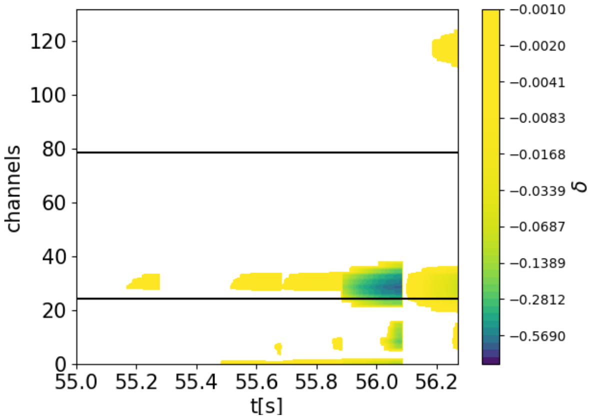


Figure B.8: Occlusion analysis for the JPN 94966. The colorbar represent the fluctuation δ as defined in Equation B.4. The blue corresponds to a stronger fluctuation as a consequence of the occlusion of that part of input, which is interpreted as more importance given to that part of input. White areas do not produce any fluctuation of the output. The colour are in logarithmic scale.

time of EC/TH, measured as reported in Section B.2.1. Saliency maps are calculated close to the event of edge cooling and temperature hollowing. The maps are superimposed, and the gradient is averaged for every pixel. The result is shown in Figure B.9. The plots show the aggregated heat map for the EC (left) and TH (right). The two plots exhibit different behaviours: the edge cooling highlights multiple areas (in the red circles) in the outer part of the profile where there are peaks in the gradient. On the other hand, the temperature hollowing exhibits multiple peaks at the centre of the profile, with a reduced value of the gradient at the edge. Also, the plot for the EC shows an important gradient at the centre of the profile, but it has a more continuous behaviour and lights up close to the peaks at the edge. The Figure B.9 shows that the gradient increase occurs in an interval 200 ms before the event of EC/TH. This is also confirmed in Figure B.11 where it is plotted the distribution of the temporal differences between the closest peak of the gradient and the time of EC/TH for all the analysed pulses. The distribution is peaked around $t - t_{EC/TH} = 0$, confirming the strong correlation between the EC/TH and the gradient increase. Furthermore, the distribution is strongly asymmetric, reflecting the tendency of the neural network to anticipate the EC/TH. The Figure B.12 shows the position of every pulse of the database in the space composed by the average of the gradient in the inner and outer half of the profile. The average gradient is calculated as the arithmetic mean of the elements of the matrix G around the time of the edge cooling/temperature hollowing. The inner region refers to the temperature profile with $r/a \in (0, 0.5)$ and the outer refers to the region $r \in (0.5, 1)$. The Figure B.12 shows that when analyzing edge cooling, the neural network tends to produce a heat map with a non-zero gradient in the outer region, indicating that it maintains its focus on the edge in the presence of a physical phenomenon

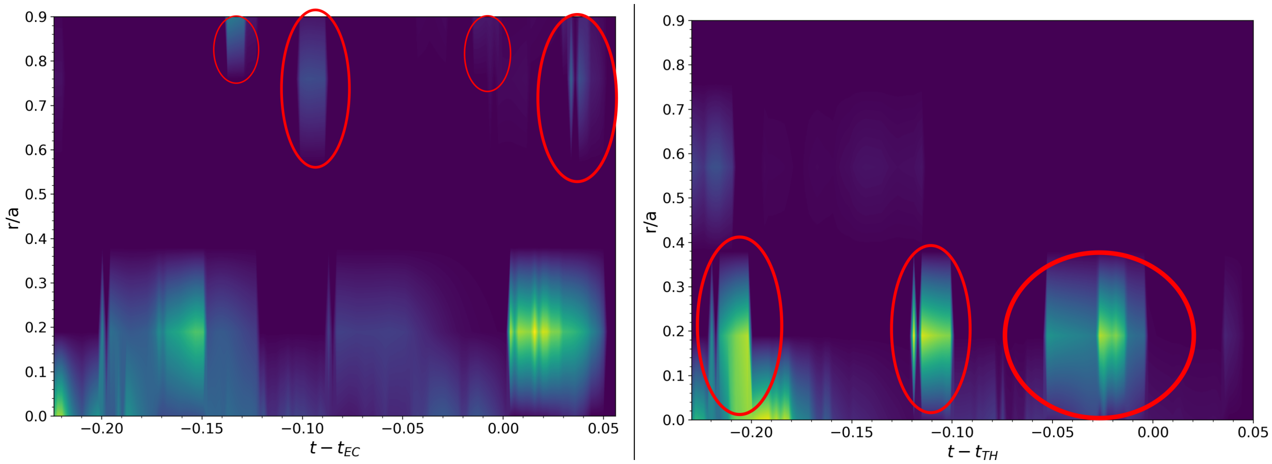


Figure B.9: The plots show the average of the saliency maps of every pulse of the database for the edge cooling (left) and temperature hollowing (right)

that affects that portion of the profile. On the other hand, temperature hollowing produces a heatmap with a zero average gradient on the edge, indicating that the neural network does not consider the outer region of the temperature profile to be important for classifying the disruption. Finally, we analysed the safe pulses. Figure B.10 shows the average of the sensitivity maps of all safe pulses around a reference time in the stable phase of the discharge. The plot does not show any peaks in the gradient, but rather a continuous area in the radiation and the 0-D signals. This indicates that the neural network does not focus on any particular phenomenon, but maintains its attention on the radiation and the one-dimensional signals, waiting for some event that could represent a precursor to the disruption.

B.5 Discussion of results

The XAI analyses provide insight about what the neural network considers important in the classification of a disruption given a certain input. One of the objectives of this work is to understand if the neural network considers important a class of phenomena (edge cooling and temperature hollowing) that involve areas of the input and not only single points. For this reason, we decided to perform the analysis using sensitivity approaches, that tend to produce results that highlight extensive areas rather than single pixels. Other approaches, as the decomposition, can in principle be used (an example is the layer-wise relevance propagation ([73])). These methods assign a “relevance score” to every pixel of the input by decomposing the output of the network in series, generating heatmaps that pinpoint the most critical pixels, offering a granular view of the input’s impact on the final result. Between the algorithms following the sensitivity approach, we developed the saliency map and the occlusion. More complicated algorithms are available, even if more suitable for larger CNN. For example the GRAD-CAM algorithm ([92]) foresees

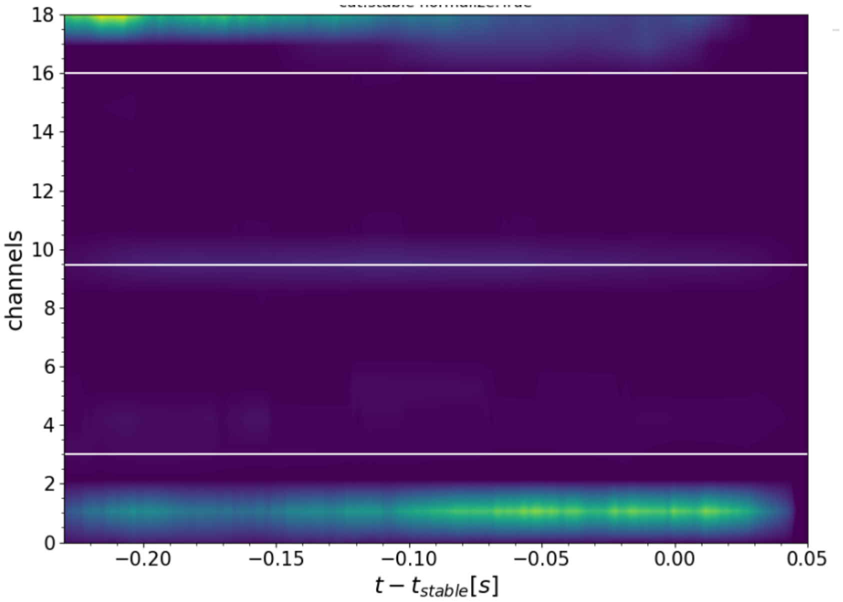


Figure B.10: The plot show the average saliency map for the safe pulses around a reference time during the stable phase.

to make a sum of the output of every feature map in a convolutional layer, weighted on an average pooling of the gradient of the final output with respect to the output of the feature map. However, this is thought to be applied on CNN composed by convolutional layer with many feature maps. The CNN we used is relatively simple, containing two convolutional layers with only one feature map, so GRAD-CAM would not provide any additional insights. The occlusion method is an agnostic method, easy to develop and to interpret, but intrinsically it depends on different free parameters such as the size of the occluded region. An in-depth analysis of the effect of occlusion parameters is beyond the scope of this work. The primary goal in using two methods was to compare them and find a set of parameters for which the results obtained are consistent with each other.

B.5.1 General results and comparison

As explained in Section B.2, the CNN is fed with inputs composed by physical data measured by the diagnostics. At first, the analysis is performed over the entire input, taking into consideration all the quantities in the input. The comparison between the two methods (Figures B.6 and B.8) allows us to identify the radiation as the most relevant part of the input for the classification. The central part of the temperature profile is also found to be particularly crucial. The monodimensional signals are important only close to the alarm time, and the density seems not to be important in the classification. The comparison produces consistent results, even though some differences should be discussed. The occlusion method seems to produce maps that are more sensitive to the right part of an input image, as shown in Figure B.5. Since the neural network is trained on temporally

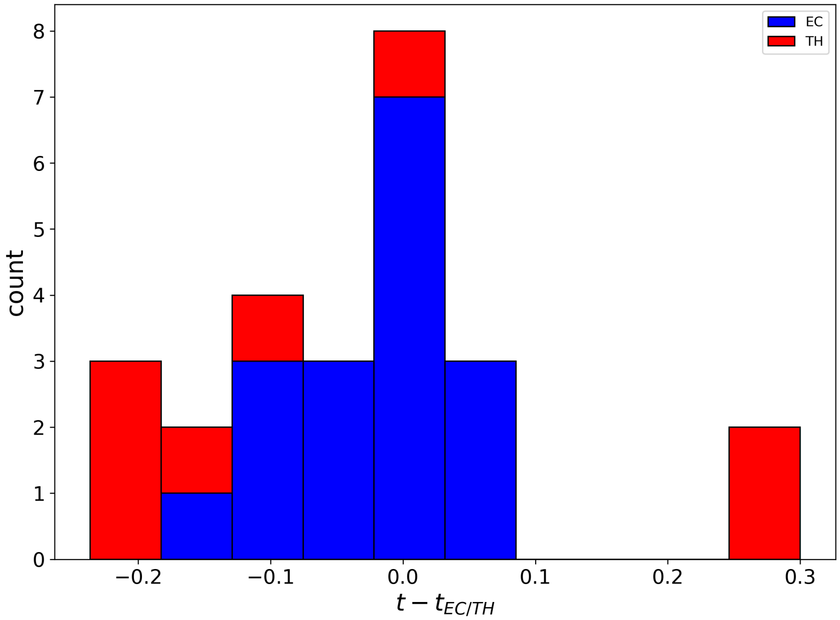


Figure B.11: The figure shows the temporal distance between the TH (red)/EC (blue) and the closest peak in the gradient in the inner/outer part of the temperature profile

ordered input images, when a disruption occurs, its evidences appear at first in the right part of the input. As a result, in the training phase, the CNN learns to be more sensitive to the right side of the input. This right-side bias is particularly evident in the occlusion technique because, in the saliency map approach, the final heatmap is the average of multiple heat maps produced with the sliding window, so the effect eventually averages out. The saliency map can be applied systematically to the data, but often produces biased results. In particular, the saliency map tends to have a strong gradient on the radiation, even when there is no significant signal. This could be because the CNN, in the training phase, adjusts its weights to give more importance to the radiation, since it has learned that it is an important feature. This implies that the gradient of the radiation, when backpropagated, are stronger with respect to the gradient coming from other diagnostics. It also reflects in the XAI analysis as the gradient in the radiation part of the input are highlighted with respect to the gradient of the other diagnostics, even if no relevant signals are present in the input. So the second part of the analysis focuses specifically on the temperature profile.

B.5.2 Analysis on the temperature

The main result of this paper is that there are strong evidences that the neural network is able to identify the edge cooling and temperature hollowing. This is shown in Figure B.9 where the average of the gradient matrix G (as defined in Section B.3.2), close to the

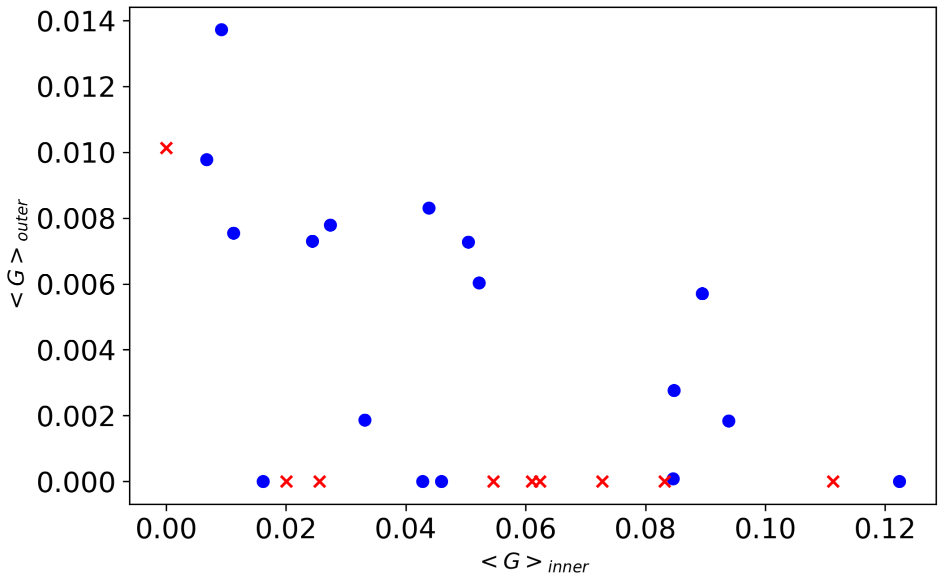


Figure B.12: Plot of the points of the database in the space composed by the average of the gradient as defined in equation B.7 in the inner and outer part of the temperature profile

EC/TH, is shown for all the disruptions. This is confirmed also in Figure B.11 and Figure B.12. In the latter, it is also evident that the gradient's average is greater in the inner region of the profile than in the outer. This confirms that, in general, the neural network places more emphasis on the temperature profile on the axis than on the edge. Furthermore, in Figure B.12 there are five points that represent edge cooling, but the gradient in the outer region is zero. Figure B.13 shows that the network gives the alarm close to the edge cooling ($\sim$ 100ms before). In this phase, there is a strong gradient on the radiation, meaning that the NN is keeping its attention on that part of the profile. So the reason why the outer average gradient of the edge coolings in Figure B.12 is zeros is that the edge cooling happens when the NN is focusing on the radiation. When the edge cooling starts, the output is already 1 and the neural network has already triggered the alarm. This is consistent to the fact that the neural network does not consider the temperature profile to be the most relevant feature to identify the disruption, and it gives more importance to the radiation. When not close to a significant event connected to the radiation, the neural network shows an increment of the gradient in the area interested in the edge cooling/temperature hollowing. This is a strong and interesting indication that the CNN learns to consider the pattern linked to the EC/TH as relevant for the classification of the disruption.

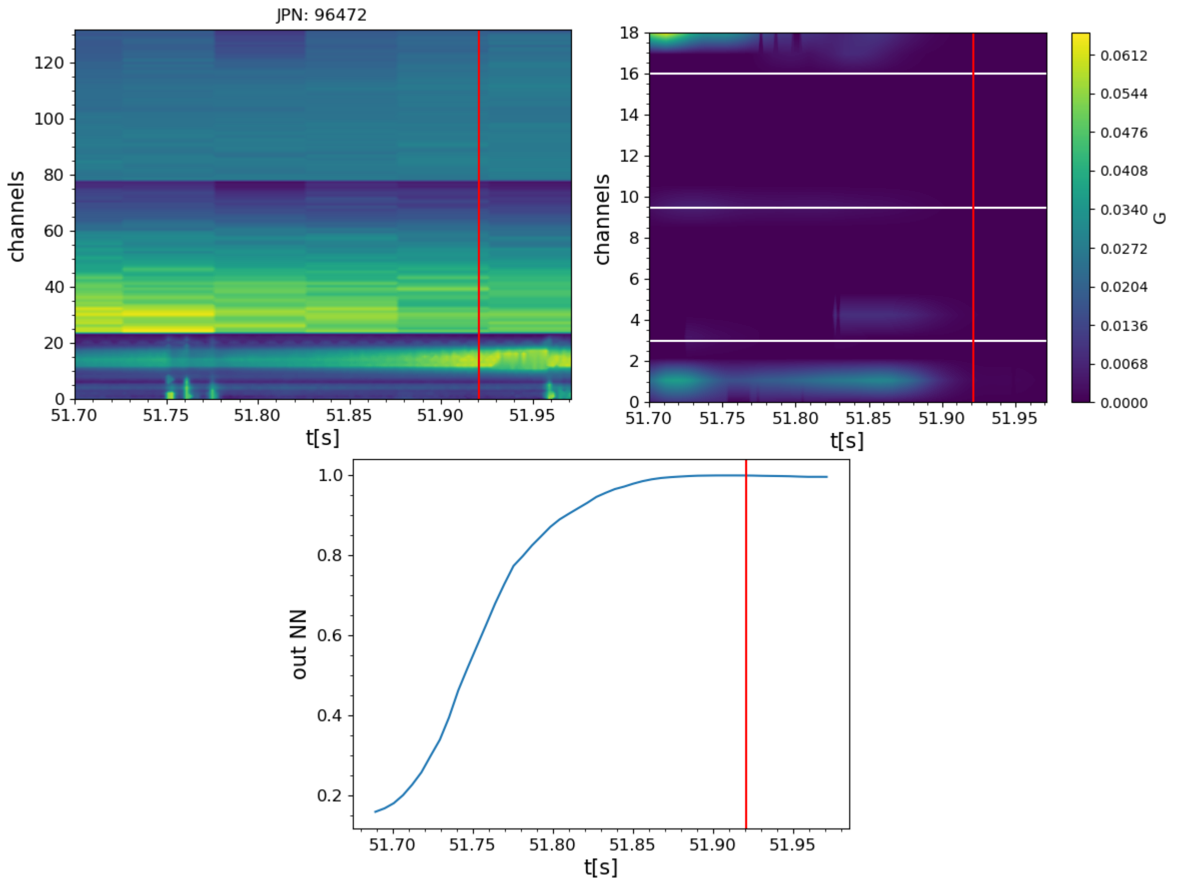


Figure B.13: The plot shows an explanation for the behaviour of the points in Figure B.12. The vertical red line represents the time of the edge cooling. The top left plot represents the input image of the neural network. On the top right there is the saliency map that corresponds to that input. At the bottom there is the output of the neural network.

B.6 Conclusion

This work shows the potential of XAI analysis on explaining the output of a CNN trained for disruption prediction. Regarding disruptions having edge cooling and temperature hollowing as precursors, the CNN behaves consistently with what we know from the physics, without providing any hint in the training phase. This could contribute to enhance the reliability of the neural network and promote its use in a disruption avoidance system. Furthermore, in principle, this could indicate that it is possible to investigate the physics by interpreting the way a neural network produces its output.

Acknowledgments

This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grant

Agreement No 101052200 EUROfusion). Views and opin- ions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

In this thesis different aspects of the tearing instabilities have been treated. A detailed theoretical framework is developed in Chapter 3, showing the problem of the linear stability of a sheared magnetic field, introducing the phenomena of the magnetic reconnection and the change of topology that gives rise to magnetic islands.

Bibliography

[1] R Ambrosino et al. "DTT-Divertor Tokamak Test facility: A testbed for DEMO". In: *Fusion Engineering and Design* 167 (2021), p. 112330.

[2] Hisham El-Amir and Mahmoud Hamdy. "Deep Learning Fundamentals". In: *Deep Learning Pipeline: Building a Deep Learning Model with TensorFlow*. Berkeley, CA: Apress, 2020, pp. 279–343. ISBN: 978-1-4842-5349-6. DOI: `10.1007/978-1-4842-5349-6_9`. URL: `https://doi.org/10.1007/978-1-4842-5349-6_9`.

[3] Plamen P Angelov et al. "Explainable artificial intelligence: an analytical review". In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 11.5 (2021), e1424.

[4] E. Aymerich et al. "A statistical approach for the automatic identification of the start of the chain of events leading to the disruptions at JET". In: *Nuclear Fusion* 61 (3 2021). ISSN: 17414326. DOI: `10.1088/1741-4326/abcb28`.

[5] E. Aymerich et al. "CNN disruption predictor at JET: Early versus late data fusion approach". In: *Fusion Engineering and Design* 193 (2023). ISSN: 09203796. DOI: `10.1016/j.fusengdes.2023.113668`.

[6] E. Aymerich et al. "Disruption prediction at JET through deep convolutional neural networks using spatiotemporal information from plasma profiles". In: *Nuclear Fusion* 62 (6 2022). ISSN: 17414326. DOI: `10.1088/1741-4326/ac525e`.

[7] Enrico Aymerich et al. "Performance Comparison of Machine Learning Disruption Predictors at JET". In: *Applied Sciences (Switzerland)* 13 (3 2023). ISSN: 20763417. DOI: `10.3390/app13032006`.

[8] MR de Baar et al. "Physics and real time control of tearing modes in TEXTOR". In: *Fusion Energy 2008 (Proc. 22nd Int. Conf., Geneva, Swizerland, 2008)*. 2008.

[9] Stephen Balaban. "Deep learning and face recognition: the state of the art". In: *Biometric and surveillance technology for human and activity identification XII* 9457 (2015), pp. 68–75.

[10] L Bardoczi et al. “The root cause of disruptive NTMs and paths to stable operation in DIII-D ITER baseline scenario plasmas”. In: *Nuclear Fusion* 64.12 (2024), p. 126005.

[11] Laszlo Bardoczi, NJ Richner, and NC Logan. “The onset distribution of rotating tearing modes and its consequences on the stability of high-confinement-mode plasmas in DIII-D”. In: *Nuclear Fusion* 63.12 (2023), p. 126052.

[12] V Basiuk et al. “Towards self-consistent plasma modelisation in presence of neoclassical tearing mode and sawteeth: effects on transport coefficients”. In: *Plasma Physics and Controlled Fusion* 59.12 (2017), p. 125012.

[13] L Bonalumi et al. “Analysis of the role of the ion polarization current on the onset of pre-disruptive magnetic islands in JET”. In: *Physics of Plasmas* 31.4 (2024).

[14] Luca Bonalumi et al. “eXplainable artificial intelligence applied to algorithms for disruption prediction in tokamak devices”. In: *Frontiers in Physics* 12 (2024), p. 1359656.

[15] D. L. Book. *NRL (Naval Research Laboratory) plasma formulary, revised*. Ed. by L. S. Wagner, J. A. Goldstein, and E. A. Chapman. 2007.

[16] D Brunetti et al. “Interpretation of tearing mode physics from ECE temperature fluctuations associated with magnetic islands and Mirnov signals”. In: *The European Physical Journal D* 64 (2011), pp. 405–411.

[17] P. Buratti et al. “Diagnostic application of magnetic islands rotation in JET”. In: *Nuclear Fusion* 56 (7 2016). ISSN: 17414326. DOI: `10.1088/0029-5515/56/7/076004`.

[18] Irene Casiraghi et al. “Core integrated simulations for the Divertor Tokamak Test facility scenarios towards consistent core-pedestal-SOL modelling”. In: *Plasma Physics and Controlled Fusion* 65.3 (2023), p. 035017.

[19] V Chan and G Guest. “Stabilization of tearing modes in tokamaks using electron cyclotron heating”. In: *Nuclear Fusion* 22.2 (1982), p. 272.

[20] Z Chang and James Donald Callen. “Global energy confinement degradation due to macroscopic phenomena in tokamaks”. In: *Nuclear Fusion* 30.2 (1990), p. 219.

[21] R. M. Churchill, B. Tobias, and Y. Zhu. “Deep convolutional neural networks for multi-scale time-series classification and application to tokamak disruption prediction using raw, high temporal resolution diagnostic data”. In: *Physics of Plasmas* 27 (6 2020). ISSN: 10897674. DOI: `10.1063/1.5144458`.

[22] Gioele Ciaparrone et al. “Deep learning in video multi-object tracking: A survey”. In: *Neurocomputing* 381 (2020), pp. 61–88.

[23] J Citrin et al. "Flux-driven multi-channel simulations with the quasilinear gyrokinetic transport model QuaLiKiz". In: *43rd EPS Conference on Plasma Physics*. European Physical Society. 2016.

[24] Nuclear Science Commission of the European Communities and Technology (Fusion). *The JET Project: Design Proposal for the Joint European Torus*. Tech. rep. EUR 5516e. Previously issued as EUR-JET-R5. Luxembourg: Commission of the European Communities, 1976.

[25] PC De Vries et al. "Survey of disruption causes at JET". In: *Nuclear fusion* 51.5 (2011), p. 053018.

[26] P. C. DeVries et al. "Survey of disruption causes at JET". In: *Nuclear Fusion* 51 (5 2011). ISSN: 00295515. DOI: `10.1088/0029-5515/51/5/053018`.

[27] Tony Donné and William Morris. *European Research Roadmap to the Realisation of Fusion Energy (Long Version)*. Tech. rep. Published by the European Commission. EUROfusion, 2018. URL: `https://www.euro-fusion.org/`.

[28] A. V. Dudkovskaia et al. "Drift kinetic theory of neoclassical tearing modes in a low collisionality tokamak plasma: Magnetic island threshold physics". In: *Plasma Physics and Controlled Fusion* 63 (5 2021). ISSN: 13616587. DOI: `10.1088/1361-6587/abea2e`.

[29] AV Dudkovskaia et al. "Drift kinetic theory of neoclassical tearing modes in tokamak plasmas: polarisation current and its effect on magnetic island threshold physics". In: *Nuclear Fusion* 63.12 (2023), p. 126040.

[30] M Erba et al. "Validation of a new mixed Bohm/gyro-Bohm model for electron and ion heat transport against the ITER, Tore Supra and START database discharges". In: *Nuclear Fusion* 38.7 (1998), p. 1013.

[31] Gonzalo Farias et al. "Applying Deep Learning for Improving Image Classification in Nuclear Fusion Devices". In: *IEEE Access* 6 (2018). ISSN: 21693536. DOI: `10.1109/ACCESS.2018.2881832`.

[32] Diogo R. Ferreira, Pedro J. Carvalho, and Horacio Fernandes. "Deep Learning for Plasma Tomography and Disruption Prediction from Bolometer Data". In: *IEEE Transactions on Plasma Science* 48 (1 2020). ISSN: 19399375. DOI: `10.1109/TPS.2019.2947304`.

[33] Diogo R. Ferreira et al. "Explainable deep learning for the analysis of MHD spectrograms in nuclear fusion". In: *Machine Learning: Science and Technology* 3 (1 2022). ISSN: 26322153. DOI: `10.1088/2632-2153/ac44aa`.

[34] R. Fitzpatrick and F. L. Waelbroeck. "Drift-tearing magnetic islands in tokamak plasmas". In: *Physics of Plasmas* 15 (1 2008). ISSN: 1070664X. DOI: `10.1063/1.2829757`.

[35] Richard Fitzpatrick. “Helical temperature perturbations associated with tearing modes in tokamak plasmas”. In: *Physics of Plasmas* 2.3 (Mar. 1995), pp. 825–838. ISSN: 1070-664X. DOI: `10.1063/1.871434`. eprint: `https://pubs.aip.org/aip/pop/article-pdf/2/3/825/19326142/825\_1\_online.pdf`. URL: `https://doi.org/10.1063/1.871434`.

[36] Richard Fitzpatrick and François L Waelbroeck. “Two-fluid magnetic island dynamics in slab geometry. I. Isolated islands”. In: *Physics of plasmas* 12.2 (2005).

[37] Harold P. Furth, John Killeen, and Marshall N. Rosenbluth. “Finite-Resistivity Instabilities of a Sheet Pinch”. In: *The Physics of Fluids* 6.4 (Apr. 1963), pp. 459–484. ISSN: 0031-9171. DOI: `10.1063/1.1706761`. eprint: `https://pubs.aip.org/aip/pfl/article-pdf/6/4/459/12485401/459\_1\_online.pdf`. URL: `https://doi.org/10.1063/1.1706761`.

[38] X Garbet, F Mourgues, and A Samain. “Non-linear self consistency of microtearing modes”. In: *Plasma physics and controlled fusion* 30.4 (1988), p. 343.

[39] H Gholamalinezhad and H Khosravi. “Pooling methods in deep neural networks, a review. arXiv 2020”. In: *arXiv preprint arXiv:2009.07485* (2009).

[40] Leonida Gianfagna and Antonio Di Cecco. *Explainable AI with python*. 2021. DOI: `10.1007/978-3-030-68640-6`.

[41] Leilani H. Gilpin et al. “Explaining Explanations: An Overview of Interpretability of Machine Learning”. In: *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*. 2018, pp. 80–89. DOI: `10.1109/DSAA.2018.00018`.

[42] A. H. Glasser, J. M. Greene, and J. L. Johnson. “Resistive instabilities in general toroidal plasma configurations”. In: *The Physics of Fluids* 18.7 (July 1975), pp. 875–888. ISSN: 0031-9171. DOI: `10.1063/1.861224`. eprint: `https://pubs.aip.org/aip/pfl/article-pdf/18/7/875/12357480/875\_1\_online.pdf`. URL: `https://doi.org/10.1063/1.861224`.

[43] AH Glasser, John M Greene, and JL Johnson. “Resistive instabilities in a tokamak”. In: *The Physics of Fluids* 19.4 (1976), pp. 567–574.

[44] Prashant Gohel, Priyanka Singh, and Manoranjan Mohanty. “Explainable AI: current status and future directions”. In: *CoRR* abs/2107.07045 (2021). arXiv: `2107.07045`. URL: `https://arxiv.org/abs/2107.07045`.

[45] R.J. Goldston and P.H. Rutherford. *Introduction to Plasma Physics*. Introduction to Plasma Physics. Taylor and Francis, 1995. ISBN: 9780750303255. URL: `https://books.google.it/books?id=Ji5HAQAAIAAJ`.

[46] NN Gorelenkov et al. "A threshold for excitation of neoclassical tearing modes". In: *Physics of Plasmas* 3.9 (1996), pp. 3379–3385.

[47] G Granucci et al. "The DTT device: System for heating". In: *Fusion Engineering and Design* 122 (2017), pp. 349–355.

[48] Jiuxiang Gu et al. "Recent advances in convolutional neural networks". In: *Pattern recognition* 77 (2018), pp. 354–377.

[49] BH Guo et al. "Disruption prediction using a full convolutional neural network on EAST". In: *Plasma Physics and Controlled Fusion* 63.2 (2020), p. 025008.

[50] Bihao H Guo et al. "Disruption prediction on EAST tokamak using a deep learning algorithm". In: *Plasma Physics and Controlled Fusion* 63.11 (2021), p. 115007.

[51] R. J. La Haye et al. "Propagation of magnetic islands in the Er=0 frame of co-injected neutral beam driven discharges in the DIII-D tokamak". In: *Physics of Plasmas* 10 (9 2003). ISSN: 1070664X. DOI: `10.1063/1.1602452`.

[52] C. C. Hegna and J. D. Callen. "Interaction of bootstrap-current-driven magnetic islands". In: *Physics of Fluids B* 4 (7 1992). ISSN: 08998221. DOI: `10.1063/1.860039`.

[53] C. C. Hegna and James D. Callen. "On the stabilization of neoclassical magnetohydrodynamic tearing modes using localized current drive or heating". In: *Physics of Plasmas* 4 (1997), pp. 2940–2946. URL: `https://api.semanticscholar.org/CorpusID:122070852`.

[54] CC Hegna. "The physics of neoclassical magnetohydrodynamic tearing modes". In: *Physics of Plasmas* 5.5 (1998), pp. 1767–1774.

[55] Chris C Hegna and James D Callen. "On the stabilization of neoclassical magnetohydrodynamic tearing modes using localized current drive or heating". In: *Physics of Plasmas* 4.8 (1997), pp. 2940–2946.

[56] FL Hinton and JA Robertson. "Neoclassical dielectric property of a tokamak plasma". In: *The Physics of fluids* 27.5 (1984), pp. 1243–1247.

[57] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. "Multilayer feedforward networks are universal approximators". In: *Neural networks* 2.5 (1989), pp. 359–366.

[58] A. Ishizawa et al. "Magnetic island evolution in hot ion plasmas". In: *Physics of Plasmas* 19 (7 2012). ISSN: 1070664X. DOI: `10.1063/1.4739291`.

[59] E. Joffrin. *Development of the 'Hybrid' scenario in JET*. Tech. rep. INIS-XA–08N0893. International Atomic Energy Agency (IAEA), 2008.

[60] Julian Kates-Harbeck, Alexey Svyatkovskiy, and William Tang. "Predicting disruptive instabilities in controlled fusion plasmas through deep learning". In: *Nature* 568.7753 (2019), pp. 526–531.

[61] Mengdi Kong. "Towards integrated control of tokamak plasmas: physics-based control of neoclassical tearing modes in the TCV tokamak". en. PhD thesis. Lausanne: EPFL, 2020. DOI: `10.5075/epfl-thesis-7510`. URL: `https://infoscience.epfl.ch/handle/20.500.14299/166264`.

[62] M Kotschenreuther, RD Hazeltine, and PJ Morrison. "Nonlinear dynamics of magnetic islands with curvature and pressure". In: *Physics of Fluids* 28.1 (1985), p. 294.

[63] Martin David Kruskal and RM Kulsrud. "Equilibrium of a magnetically confined plasma in a toroid". In: (1958).

[64] James Ladyman, James Lambert, and Karoline Wiesner. "What is a complex system?" In: *European Journal for Philosophy of Science* 3 (2013), pp. 33–67.

[65] D. De Lazzari and E. Westerhof. "On the merits of heating and current drive for tearing mode stabilization". In: *Nuclear Fusion* 49.7 (May 2009), p. 075002. DOI: `10.1088/0029-5515/49/7/075002`. URL: `https://dx.doi.org/10.1088/0029-5515/49/7/075002`.

[66] E. Lazzaro et al. "Physics conditions for robust control of tearing modes in a rotating tokamak plasma". In: *Plasma Physics and Controlled Fusion* 60 (1 2018). ISSN: 13616587. DOI: `10.1088/1361-6587/aa8be7`.

[67] Di S. Li et al. "Optimal Tracking for a Divergent-Type Parabolic PDE System in Current Profile Control". In: *Abstract and Applied Analysis* (2014). CC BY 4.0. DOI: `10.1155/2014/940965`. URL: `https://commons.wikimedia.org/w/index.php?curid=74679708`.

[68] Hinrich Lütjens, Jean-François Luciani, and Xavier Garbet. "Curvature effects on the dynamics of tearing modes in tokamaks". In: *Physics of Plasmas* 8.10 (2001), pp. 4267–4270.

[69] Costanza F Maggi. "Overview of T and DT results in JET with ITER-like wall". In: *Nuclear Fusion* (2023).

[70] M Maslov et al. "JET DT scenario with optimized non-thermal fusion". In: *Nuclear Fusion* 63.11 (2023), p. 112002.

[71] A. B. Mikhailovskii. *Theory of magnetic islands in tokamaks with accenting neoclassical tearing modes*. 2003. DOI: `10.1002/ctpp.200310013`.

[72] Tom M Mitchell and Tom M Mitchell. *Machine learning*. Vol. 1. 9. McGraw-hill New York, 1997.

[73] Grégoire Montavon et al. "Layer-Wise Relevance Propagation: An Overview". In: *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Ed. by Wojciech Samek et al. Cham: Springer International Publishing, 2019, pp. 193–209. ISBN: 978-3-030-28954-6. DOI: `10.1007/978-3-030-28954-6_10`. URL: `https://doi.org/10.1007/978-3-030-28954-6_10`.

[74] Y Nishimura, James D Callen, and Chris C Hegna. "Tearing mode analysis in tokamaks, revisited". In: *Physics of Plasmas* 5.12 (1998), pp. 4292–4299.

[75] Hamid Palangi. *Applied Deep Learning - Part 4: Convolutional Neural Networks*. `https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2`. 2018.

[76] A. Pavone et al. "Machine learning and Bayesian inference in nuclear fusion research: an overview". In: *Plasma Physics and Controlled Fusion* 65 (5 2023). ISSN: 13616587. DOI: `10.1088/1361-6587/acc60f`.

[77] E. Poli et al. "Electron-cyclotron-current-drive efficiency in DEMO plasmas". In: *Nuclear Fusion* 53.1 (Dec. 2012), p. 013011. DOI: `10.1088/0029-5515/53/1/013011`. URL: `https://dx.doi.org/10.1088/0029-5515/53/1/013011`.

[78] E. Poli et al. "Kinetic calculation of the polarization current in the presence of a neoclassical tearing mode". In: *Nuclear Fusion* 45 (5 2005). ISSN: 00295515. DOI: `10.1088/0029-5515/45/5/009`.

[79] G Pucella et al. "Beta-induced Alfvén eigenmodes and geodesic acoustic modes in the presence of strong tearing activity during the current ramp-down on JET". In: *Plasma Physics and Controlled Fusion* 64.4 (Mar. 2022), p. 045023. DOI: `10.1088/1361-6587/ac4ade`. URL: `https://dx.doi.org/10.1088/1361-6587/ac4ade`.

[80] G. Pucella et al. "Onset of tearing modes in plasma termination on JET: The role of temperature hollowing and edge cooling". In: *Nuclear Fusion* 61 (4 2021). ISSN: 17414326. DOI: `10.1088/1741-4326/abe3c7`.

[81] Paul-Henri Rebut. "The joint european torus (jet)". In: *The European Physical Journal H* 43.4 (2018), pp. 459–497.

[82] PH Rebut and M Hugon. "Magnetic turbulence self-sustainment by finite larmor radius effect". In: *Plasma physics and controlled fusion* 33.9 (1991), p. 1085.

[83] P Rindt et al. "Conceptual design of a liquid-metal divertor for the European DEMO". In: *Fusion Engineering and Design* 173 (2021), p. 112812.

[84] Riccardo Rossi et al. “Development of robust indicators for the identification of electron temperature profile anomalies and application to JET”. In: *Plasma Physics and Controlled Fusion* 64 (4 2022). ISSN: 13616587. DOI: `10.1088/1361-6587/ac4d3b`.

[85] P. H. Rutherford. “Nonlinear growth of the tearing mode”. In: *The Physics of Fluids* 16.11 (Nov. 1973), pp. 1903–1908. ISSN: 0031-9171. DOI: `10.1063/1.1694232`. eprint: `https://pubs.aip.org/aip/pfl/article-pdf/16/11/1903/12262797/1903\_1\_online.pdf`. URL: `https://doi.org/10.1063/1.1694232`.

[86] P. H. Rutherford. “Nonlinear growth of the tearing mode”. In: *Physics of Fluids* 16 (11 1973). ISSN: 10706631. DOI: `10.1063/1.1694232`.

[87] Sandro Sandri et al. “A review of radioactive wastes production and potential environmental releases at experimental nuclear fusion facilities”. In: *Environments* 7.1 (2020), p. 6.

[88] O Sauter et al. “Marginal β-limit for neoclassical tearing modes in JET H-mode discharges”. In: *Plasma Physics and Controlled Fusion* 44.9 (2002), p. 1999.

[89] O Sauter et al. “On the requirements to control neoclassical tearing modes in burning plasmas”. In: *Plasma Physics and Controlled Fusion* 52.2 (2010), p. 025002.

[90] O. Sauter. “On the contribution of local current density to neoclassical tearing mode stabilization”. In: *Physics of Plasmas* 11.10 (Oct. 2004), pp. 4808–4813. ISSN: 1070-664X. DOI: `10.1063/1.1787791`. eprint: `https://pubs.aip.org/aip/pop/article-pdf/11/10/4808/19190856/4808\_1\_online.pdf`. URL: `https://doi.org/10.1063/1.1787791`.

[91] O. Sauter et al. “Beta limits in long-pulse tokamak discharges”. In: *Physics of Plasmas* 4 (5 1997). ISSN: 1070664X. DOI: `10.1063/1.872270`.

[92] Ramprasaath R. Selvaraju et al. “Grad-CAM: Why did you say that? Visual Explanations from Deep Networks via Gradient-based Localization”. In: *CoRR* abs/1610.02391 (2016). arXiv: `1610.02391`. URL: `http://arxiv.org/abs/1610.02391`.

[93] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. “Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps”. In: (2014). arXiv: `1312.6034 [cs.CV]`.

[94] A. I. Smolyakov. “Nonlinear evolution of tearing modes in inhomogeneous plasmas”. In: *Plasma Physics and Controlled Fusion* 35 (6 1993). ISSN: 07413335. DOI: `10.1088/0741-3335/35/6/002`.

[95] A. I. Smolyakov et al. “Rotating nonlinear magnetic islands in a tokamak plasma”. In: *Physics of Plasmas* 2 (5 1995). ISSN: 1070664X. DOI: `10.1063/1.871308`.

[96] AI Smolyakov. “Drift magnetic islands in a tokamak”. In: *Fizika Plazmy* 15.10 (1989), pp. 1153–1159.

[97] AI Smolyakov. “Drift magnetic islands in a tokamak”. In: *Fizika Plazmy* 15.10 (1989), pp. 1153–1159.

[98] Andrei Ivanovich Smolyakov. “Nonlinear evolution of tearing modes in inhomogeneous plasmas”. In: *Plasma physics and controlled fusion* 35.6 (1993), p. 657.

[99] J. A. Snipes et al. “Large amplitude quasi-stationary MHD modes in JET”. In: *Nuclear Fusion* 28 (6 1988). ISSN: 17414326. DOI: `10.1088/0029-5515/28/6/010`.

[100] JA Snipes et al. “Large amplitude quasi-stationary MHD modes in JET”. In: *Nuclear fusion* 28.6 (1988), p. 1085.

[101] C. Sozzi et al. “Termination of Discharges in High Performance Scenarios in JET.” In: *Poster presented at 28th IAEA Fusion Energy Conference (FEC 2020), Virtual*. 2021.

[102] Lucas Spangher et al. “Autoregressive Transformers for Disruption Prediction in Nuclear Fusion Plasmas”. In: *arXiv preprint arXiv:2401.00051* (2023).

[103] Jost Tobias Springenberg et al. “Striving for Simplicity: The All Convolutional Net”. In: (2015). arXiv: `1412.6806 [cs.LG]`.

[104] WM Stacey. “A survey of thermal instabilities in tokamak plasmas: Theory, comparison with experiment, and predictions for future devices”. In: *Fusion science and technology* 52.1 (2007), pp. 29–67.

[105] EJ Strait et al. “Progress in disruption prevention for ITER”. In: *Nuclear Fusion* 59.11 (2019), p. 112012.

[106] Henry R Strauss. “Models of resistive wall tearing mode disruptions”. In: *Physics of Plasmas* 30.11 (2023).

[107] Eva Tuba et al. “Convolutional neural networks hyperparameters tuning”. In: *Artificial intelligence: theory and applications*. Springer, 2021, pp. 65–84.

[108] Jesús Vega et al. “Disruption prediction with artificial intelligence techniques in tokamak plasmas”. In: *Nature Physics* 18.7 (2022), pp. 741–750.

[109] Bas H.M. Van Der Velden. “Explainable AI: current status and future potential”. In: *European Radiology* (2023). ISSN: 14321084. DOI: `10.1007/s00330-023-10121-4`.

[110] P. C. De Vries et al. "Survey of disruption causes at JET". In: *Nuclear Fusion* 51 (5 2011). ISSN: 00295515. DOI: `10.1088/0029-5515/51/5/053018`.

[111] F. L. Waelbroeck. "Natural velocity of magnetic islands". In: *Physical Review Letters* 95 (3 2005). ISSN: 00319007. DOI: `10.1103/PhysRevLett.95.035002`.

[112] J. Wesson and D.J. Campbell. *Tokamaks*. International series of monographs on physics. Clarendon Press, 2004. ISBN: 9780198509226. URL: `https://books.google.it/books?id=iPlAwZI6HIYC`.

[113] JA Wesson et al. "Disruptions in JET". In: *Nuclear Fusion* 29.4 (1989), p. 641.

[114] John Wesson and D. J. Campbell. *Tokamaks*. Google-Books-ID: XJssMXjHUr0C. OUP Oxford, Oct. 13, 2011. 828 pp. ISBN: 978-0-19-959223-4.

[115] Roscoe B White et al. *Saturation of the tearing mode*. Tech. rep. Princeton Plasma Physics Lab.(PPPL), Princeton, NJ (United States), 1976.

[116] Wikipedia contributors. *Fusion Power*. `https://en.wikipedia.org/wiki/Fusion_power`. 2024.

[117] Wenzhi Zhao, Shihong Du, and William J Emery. "Object-based convolutional neural network for high-resolution imagery classification". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10.7 (2017), pp. 3386–3396.

[118] J. X. Zhu et al. "Integrated deep learning framework for unstable event identification and disruption prediction of tokamak plasmas". In: *Nuclear Fusion* 63 (4 2023). ISSN: 17414326. DOI: `10.1088/1741-4326/acb803`.

[119] Hartmut Zohm. "Current Driven ('classical') Tearing Modes in Tokamaks". In: *Magnetohydrodynamic Stability of Tokamaks*. John Wiley and Sons, Ltd, 2014. Chap. 9, pp. 141–158. ISBN: 9783527677375. DOI: `https://doi.org/10.1002/9783527677375.ch9`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/9783527677375.ch9`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527677375.ch9`.

[120] Hartmut Zohm. "Stabilization of neoclassical tearing modes by electron cyclotron current drive". In: *Physics of Plasmas* 4 (9 1997). ISSN: 1070664X. DOI: `10.1063/1.872487`.

Acknowledgements

This thesis testify my scientific research, which I conducted over three years of PhD studies. None of this would have been possible without the support of several people, whom I would like to acknowledge in this section.
I want to thank my supervisor, Carlo Sozzi, for wisely guiding my PhD journey and for his constant availability over these three years. His ability to skillfully untangle the thread of thought that shapes an idea through scientific discussion will always remain a model of inspiration for me.
I am particularly grateful to Edoardo Alessi for the continuous advice, support, and engaging conversations that frequently inspired fresh ideas. His physical insight and knowledge of the engineering aspects of the tokamak and the NTM control system have always provided me with original perspectives, pushing me to continuously improve my work with scientific rigor.
I wish to express my sincere gratitude to Silvana Nowak for sharing her expertise in modeling NTM evolution and stabilization with dedication and precision. Her support has been crucial in introducing me to the use of the JETTO-Jintrac code, providing detailed explanations of the Neoclassical Tearing Mode evolution module.
Finally, my heartfelt thanks to Enzo Lazzaro, an immense source of knowledge whose deep passion has guided me throughout. I will always cherish the long discussions about the theory of the tearing modes, which at times moved me with their elegance and depth of thought.