Universita degli Studi di Padova
Università degli Studi di Napoli Federico II

Joint Research Doctorate in Fusion Science and Engineering
XXXVI Cycle

Model-free and data-driven approaches to the Vertical Stabilization problem in tokamak plasmas

Candidate: Sara Dubbioso

Supervisor: Prof. Gianmaria De Tommasi

Co-supervisor at Università di Napoli Federico II: Prof. Marcello Cinque

Napoli, December 22, 2023

Abstract

Nuclear fusion has long been considered the holy grail of energy production, offering the potential to provide abundant and safe power for the future. The achievement of controlled nuclear fusion has proven to be a significant scientific and engineering challenge, and in this framework, the tokamak technology appears to be the most promising approach. This type of reactor is used for magnetic-confinement thermonuclear fusion, where hydrogen fuel, in the form of plasma, is confined using a superposition of external and self-generated magnetic fields.

One of the major issues with tokamaks is the need to stabilize the plasma column vertically. This is essential to maintain the plasma stable and sustain the conditions for fusion. Although elongated plasmas can enhance the plasma performance, they require a magnetic field to shape the plasma column, which leads to an unstable equilibrium. Therefore, it is necessary to have feedback control of the plasma's vertical position.

This doctoral dissertation examines the development of vertical stabilization of the tokamak plasma to advance magnetic confinement in nuclear fusion research. It investigates the potential of developing magnetic controllers by combining traditional control engineering techniques with Artificial Intelligence. This thesis provides a basis for using model-free and data-driven approaches as an alternative to the commonly used model-based vertical stabilization controllers.

Model-based controllers have shown significant promise in the vertical stabilization of the tokamak plasma; however, their effectiveness can be limited by the complexity and uncertainty of plasma dynamics, potential model mismatch, and computational requirements. The work in this thesis addresses these limitations by developing control strategies that guarantee the required level of performance without relying on the knowledge of a plant model to improve the robustness of the overall plasma magnetic control system. To this aim, two model-free and data-driven approaches have been developed for vertical stabilization: the first one relies on the Extremum Seeking algorithm to achieve stabilization, while the second one is based on Reinforcement Learning. Most of the proposed controllers have been tested in simulation by considering the ITER plasma as a case study.

Specifically, a model-free Extremum Seeking algorithm for stabilization has been studied and deployed in the Vertical Stabilization system. To assess the robustness of the proposed approach, linear and nonlinear simulations were performed considering ITER and TCV tokamaks. In addition, automata for the adaptation of the real-time control gain adaptation and neural networks were included in the Extremum Seeking-based controller to enhance the robustness and generalization property of the ITER Vertical Stabilization even in the presence of significant model uncertainties. Indeed, the Extremum Seeking algorithm calls for a Kalman filter to compute the required Lyapunov function, making the corresponding approach only quasi-model-independent. To resolve this issue, neural networks have been trained to estimate the plasma unstable dynamic and replace the Kalman filter in the control scheme. With the use of neural networks, the controller becomes completely model-free and the operative space of the Vertical Stabilization system is also enlarged by making it possible to stabilize plasma equilibria that were not stabilized by the set-up based on a single Kalman filter.

Finally, Reinforcement Learning algorithms were considered to deploy an intelligent agent as a Vertical Stabilization system for the magnetic confinement of tokamak plasmas. A tabular Q-learning algorithm was first developed for the Vertical Stabilization of the EAST tokamaks. It was followed by a Deep Deterministic Policy Gradient algorithm, which exploits an actor-critic setup based on deep neural networks to approximate the optimal behavior of the agent. The latter was implemented considering the entire magnetic control system of ITER as an environment.

Keywords: nuclear fusion, tokamaks, magnetic confinement, Vertical Stabilization, model-free, data-driven, Extremum Seeking, Neural Network, Reinforcement Learning, Q-learning, DDPG

Abstract (Italiano)

La fusione nucleare è da tempo considerata il futuro della produzione energetica in quanto offrirebbe il potenziale per una produzione abbondante e sicura di energia. Il raggiungimento della fusione nucleare in maniera controllata si sta però dimostrando una sfida scientifica e ingegneristica significativa. In questo contesto, la tecnolgia dei tokamak sembra essere l'approccio più promettente. Questo tipo di reattore è infatti utilizzato per realizzare la fusione termonucleare a confinamento magnetico, dove l'idrogeno, usato come combustibile, viene trasformato in plasma e confinato utilizzando una sovrapposizione di campi magnetici, generati sia da fornti esterne e che autogenerati dal plasma stesso.

Uno dei problemi principali nei i tokamak è la necessità di stabilizzare verticalmente la colonna di plasma. Questa stabilizzazione verticale è essenziale infatti per mantenere il plasma stabile e sostenere le condizioni per la fusione. I plasmi che presentano una forma allungata possono migliorare le prestazioni ma richiedono che il plasma sia modellato con un campo magnetico che lo porta anche in un equilibrio instabile. Pertanto, risulta necessario un controllo in retroazione della posizione verticale del plasma.

In questa tesi di dottorato vengono studiate techine di stabilizzazione verticale del plasma nei tokamak per far progredire il confinamento magnetico nel campo della fusione nucleare. In particolare, si indaga la possibilità di implementare dei controllori magnetici combinando tecniche di controllo tradizionali con l'intelligenza artificiale. Questa tesi rappresenta un inizio per l'utilizzo di approcci *model-free* e *data-driven* come alternativa ai controllori comunemente utilizzati per la stabilizzazione verticale.

I controllori *model-based* hanno mostrato notevoli promesse nel campo della stabilizzazione verticale del plasma; tuttavia, la loro efficacia può essere limitata dalla complessità e dall'incertezza relativa alla dinamica del plasma, potenziali errori nel modello utilizzato come riferimento e dai requisiti computazionali di solito richiesti per ottenere i modelli necessari in real time. Il lavoro in questa tesi affronta queste limitazioni sviluppando strategie di controllo che garantiscono il livello di prestazioni richiesto senza però fare affidamento sulla conoscenza di un modello dell'impianto. In questo modo è possibile migliorare complessivamente la robustezza del sistema di controllo magnetico del plasma. Per questo scopo, sono stati sviluppati due approcchi *model-free* e *data-driven* per la stabilizzazione verticale. Il primo si basa sull'algoritmo noto come Extremum Seeking per ottenere la stabilizzazione, mentre il secondo è un'applicazione del Reinforcement Learning. La maggior parte dei controllori proposti in questa tesi sono stati testati in simulazione considerando il plasma di ITER come un caso studio. Nello specifico, l'algoritmo di Extremum Seeking è stato studiato e implementato per la stabilizzazione verticale. Per valutarne la robustezza sono state eseguite simulazioni lineari e non lineari considerando i tokamak ITER e TCV. Inoltre, una logica ad eventi per adattare in real-time il guadagno di controllo e una rete neurale sono stati inclusi nello schema di controllo. Quest'uiltima permette di migliorare la robustezza e le proprietà di generalizzazione del sistema di stabilizzazione verticale ad ITER anche in presenza di incertezze significative sul modello del plasma considerato. Infatti, l'algoritmo di Extremum Seeking richiede l'utilizzo di un filtro Kalman per calcolare la funzione Lyapunov nella legge di controllo. Questo rende il sistema di controllo solo quasi *model-indipendent*. Per risolvere questo problema, delle reti neurali sono state addestrate per stimare la dinamica instabile del plasma e sostituire il filtro di Kalman nello schema di controllo. Con l'uso delle reti neurali, il controllore diventa quindi completamente *model-free* e lo spazio operativo del sistema di stabilizzazione verticale viene ampliato rendendo possibile stabilizzare equilibri non stabilizzabili con il set-up basato su un singolo filtro Kalman.

Infine, due algoritmi di Reinforcement Learning sono stati considerati per ottenere un agente intelligente che si comporti come un sistema di stabilizzazione verticale. L'algoritmo di Q-learning, nella sua forma tabellare, è stato sviluppato per la stabilizzazione verticale per il tokamak EAST. Successivamente invece è stato sviluppato l'algoritmo Deep Deterministic Policy Gradient, che sfrutta la configurazione attore-critico con reti neurali del Reinforcement Learining per approssimare il comportamento dell'agente. Quest'ultimo algoritmo è stato implementato considerando l'intero sistema di controllo magnetico di ITER come *environment*.

Parole chiave: fusione nucleare, tokamaks, confinamento magnetico, Stabilizazzione Verticale, model-free, data-driven, Extremum Seeking, reti neurali, Reinforcement Learning, Q-learining, DDPG

Contents

List of Acronyms

The following acronyms are used throughout the thesis.

clf	candidate Lyapunov function
CNN	Convolutional Neural Network
DDPG	Deep Deterministic Policy Gradient
DEMO	Demonstration Power Plant
DRL	Deep Reinforcement Learning
DIII-D	Doublet III-D
EAST	Experimental Advanced Superconducting Tokamak
ELM	Extreme Learning Machine
ES	Extremum Seeking
ESS	Extremum Seeking for Stabilization
FP	First Plasma
FPO	Fusion Power Operation
GTM	Generative Topographic Mapping
IAE	Integral Absolute Error
ITAE	Integral Time-weighted Absolute Error

ITER	International Thermonuclear Experimental Reactor
JET	Joint European Torus
JPS	JET Protection System
JT-60SA	Japenese Torus-60 Super Advanced
LR	Linear Regressor
LSTM	Long Short Term Memory
MCS	Magnetic Control System
MD	Minor Disruption
MHD	Magnetohydrodynamic
MIMO	Multiple Inputs Multiple Outputs
ML	Machine Learning
MLP	Multilayer Perceptron
NN	Neural Network
PCS	Plasma Control System
PF	Poloidal Field
PFC	Poloidal Field Current
PID	Proportional–Integral–Derivative
PFPO	Pre-Fusion Power Operation
RCN	Reservoir Computing Networks
ReLU	Rectified Linear Unit
RF	Random Forests
RL	Reinforcement Learning

RMSE	Root Mean Squared Error
RNN	Recurrent Neural Networks
RUL	Remaining Useful Life
SVM	Support Vector Machines
TCV	Tokamak à Configuration Variable
TD	Temporal Difference
VDE	Vertical Displacement Event
VS	Vertical Stabilization
XSC	eXtreme Shape Controller

List of Symbols

The following symbols are used within the thesis

α,k	ES control law gains
α_u	Input weight scalar parameter in the ELM network model
$\beta_{p_{eq}}$	Nominal value of the poloidal β
β_p	Poloidal β
Δ	Interval of values
δ	Variation with respect to the equilibrium value
$\dot{Z}_c$	Plasma centroid vertical velocity
ϵ	RL ϵ-greedy policy probability
Γ	RL discount factor
γ	Plasma vertical instability growth rate
$\hat{\xi}$	Estimated movement of the plasma state along the vertically unstable dynamic
ι	Time in the states of the event-driven gain adaptive logic in Figurez 3.16 and 3.19.
κ	Plasma elongation
λ	Neural network learning rate

$\mathcal{Q}$	Fusion gain power
ω	ES dithering frequency
$\overrightarrow{B}_{1,2}$	Magnetic field vectors in the plasma vertically unstable filamentary model
$\overrightarrow{F}_{1p,2p}$	Force vectors in the plasma vertically unstable filamentary model
$\overrightarrow{I}_{1,2}$	Current vectors in the plasma vertically unstable filamentary model
π	RL policy
π^*	RL optimal policy
σ^2	Noise variance in the OUP noise model
σ^2_{dr}	Noise variance decay rate in the OUP noise model
τ	Energy confinement time
τ_1	Pure delay in the model of the ITER $VS3$ linear power supply
τ_2	Time constant of the first order dynamic in the model of ITER $VS3$ linear power supply
$\mathbf{A}$	Plasma linearized model dynamix matrix
$\mathbf{B}$	Plasma linearized model state-input matrix
$\mathbf{C}$	Plasma linearized model output-state matrix
$\mathbf{D}$	Matrix with column D_t in the ELM output equation (4.2)
$\mathbf{E}$	Plasma linearized model state-disturbance matrix
$\mathbf{F}$	Plasma linearized model output-disturbance matrix
$\mathbf{I}$	Identity matrix
$\mathbf{R}$	Matrix with column R_t in the ELM output equation (4.2)
$\mathbf{W}^{in}$	Input weight matrix in the ELM equation (4.1)

$\mathbf{W}^{out}$	Output weight matrix in the ELM equation (4.1)
θ	RL learning rate
ε	Regularization parameter in the ELM network model
ξ	Movement of the plasma state along the vertically unstable dynamic
a, A	RL action
a_{red}	Eigenvalue of the reduced averaged unstable system
b_{red}	State-input scalar value of the reduced averaged unstable system
c	Speed of light
D_t	*Readouts* desired values in the ELM network model
E	Energy
f,g	Generic function
f_{ref}	Non-linear activation function of the reservoir neuron in the ELM equation (4.1)
G	RL cumulated reward
I_e	Eddy currents
I_{IC}	Current in the VS dedicated circuit at EAST
$I_{p_{eq}}$	Nominal value of the plasma current
$I_{PF_{eq}}$	Nominal values of the current in the PF coils
I_{PF}	Currents in the PF coils
I_p	Plasma current
I_{SC}	Currents in the ex-vessel superconductive circuits
I_{VS}	Currents in the in-vessel copper-made circuit dedicated to the VS at ITER

K^{in}	Input sparsity parameter in the ELM network model
$l_{i_{eq}}$	Nominal value of the plasma internal inductance
l_i	Plasma interanal inductance
m	Mass
n	High-energy neutron
n_e	Plasma electron density
n_i	Plasma ion density
$Q(s,a)$	RL action value function
R	RL reward
R_0	Plasma major radius
R_a	Plasma minor radius
R_t	Reservoir output in the ELM equation (4.1)
s, S	RL state
S_i	States in the event-driven gain adaptive logic included in the VS control architecture
T	Episode duration in the DDPG training
t	Time
T_e	Plasma electron temperature
T_f	Final time
T_i	Ion temperature
T_i	Plasma ion temperature
T_s	Sampling time
u	Input vector

U_t	Reservoir input in the ELM equation (4.1)
u_{max}	Maximum absolute output voltage of the ITER $VS3$ linear power supply
u_{SC}	Voltages applied to the ex-vessel superconductive circuits at ITER
u_{VS}	Voltage applied to the in-vessel copper-made circuit dedicated to the VS ITER
$V(s)$	RL state value function
$V(x)$	Candidate Lyapunov function
V_{IC}	Voltage applied to the VS dedicated circuit at EAST
$VS3$	The copper-made in-vessel dedicated circuit for the VS at ITER
w	Plasma linearized model disturbance vector
x	State vector
y	Output vector
Y_t	*Readouts*, network output in the ELM equation (4.1)
Z_c	Plasma centroid vertical position
HL	Noise half-life in the OUP noise model

List of Figures

List of Tables

1 Introduction

Nuclear fusion has long been seen as a promising sustainable source of energy [7, 8], offering the potential for clean and limitless power without the harmful emissions associated with traditional fossil fuels. It is the same process that powers the Sun and other stars, and unlike nuclear fission, which produces radioactive waste, it yields only helium as a by-product.

Nuclear fusion involves the combination of atomic nuclei to form heavier elements, and this process can be controlled to extract energy. The most promising reaction to achieve controlled nuclear fusion is between two hydrogen isotopes, deuterium (D) and tritium (T). When the deuterium and tritium nuclei combine, they form a helium nucleus (He) and release a large amount of energy, according to the following reaction:

$$D + T \rightarrow He + n + E$$

where n represents an high-energy neutron, and E is the energy released. The amount of energy released during fusion reactions depends on the mass difference between the hydrogen isotope masses at the beginning of the reaction and the total mass at the end. Since the final mass is less than before, according to Einstein's famous equation, $E = mc^2$, this mass difference is converted into energy that can be used to generate electricity.

The achievement of controlled fusion reactions for power generation is a significant scientific and engineering challenge. This is due to the extremely high temperatures and pressures required to overcome the electrostatic repulsion between the atomic nuclei and initiate the fusion process. Scientists are actively researching and developing various fusion technologies to create conditions that can sustain controlled fusion reactions and extract usable energy from them. Among these various technologies, from the 50s-60s of the last century, tokamak [9] stands out as the most promising. A tokamak is a toroidal device (doughnut-shaped) in which a fully ionized gas of hydrogen ions, the *plasma*, is heated to extremely high temperatures and confined using magnetic fields.

The tokamak has been adopted as the most promising configuration of magnetic fusion devices and several machines are operating around the world; the Joint European Torus (JET) [10] in the United Kingdom, which was recently exceeded in size by the joint European-Japan project Japenese Torus-60 Super Advanced (JT-60SA) [11, 12], now the largest and most powerful superconductive tokamak in operation [11]. Furthermore, the Doublet III-D (DIII-D) tokamak [13] in the United State, Experimental Advanced Superconducting Tokamak (EAST) [14] in China and Tokamak à Configuration Variable (TCV) in Switzerland [15]. These devices are used to study plasma behavior and test new technologies for obtaining fusion power.

The largest and most ambitious tokamak project is the ITER tokamak [1], currently under construction in France. ITER is a collaboration between 35 countries and aims to demonstrate the feasibility of fusion power on a commercial scale. Once completed, it will be the largest tokamak ever built and a major milestone in the pursuit of sustainable and clean energy.

1.1 What is a tokamak?

The tokamak is an experimental device created to capture the energy of fusion. It was first developed by Soviet scientists in the late 1960s, and its name is derived from a Russian acronym meaning ”toroidal chamber with magnetic coils”. Indeed, the core of a tokamak is its doughnut-shaped vacuum chamber, in which, due to extreme heat and pressure, the gaseous hydrogen fuel is transformed into a plasma, a hot and electrically charged gas. These plasmas are the medium in which light elements can fuse and generate energy. To overcome the natural electromagnetic repulsion between positively charged atomic nuclei and enable fusion reactions to take place, the plasma is heated to extremely high temperatures in the range of tens of millions of degrees Celsius. Moreover, the vacuum chamber is encompassed by a set of coils that create a powerful magnetic field. This magnetic field is used to contain the plasma and prevent it from making contact with the walls of the chamber, which would lead to it cooling down and losing energy. Finally, the energy generated in the tokamak is absorbed as heat on the walls of the vessel. This heat is then utilized to create steam and subsequently electricity through turbines and generators, just as in traditional power plants.

The main difficulty in running a tokamak is sustaining the plasma's stability and containment while managing the energy it produces. This requires precise regulation of the magnetic fields and careful handling of the heat and particle flows inside the plasma.

1.1.1 The ITER tokamak

ITER, which is Latin for "The Way", is one of the most ambitious energy projects in the world today. It is set to be the largest tokamak ever built, with twice the size of the current largest machine and a plasma chamber volume that is ten times larger. The primary goal of ITER is to prove the feasibility of fusion as a large-scale and carbon-free source of energy, paving the way for future fusion power plants.

At ITER, it will be investigate and demonstrate the capability of *burning plasmas*, a condition in which the energy produced by the fusion reactions is sufficient to maintain plasma temperature and fusion conditions without the need for external heating. Additionally, the potential for the production of fusion energy in a tokamak is determined by the number of fusion reactions taking place at its core: the larger the vessel and the volume of the plasma, the more fusion reactions can occur and more energy can be generated. ITER is designed to produce a ten-fold return on the power, or specifically, 500 MW of fusion power from 50 MW of input heating power.

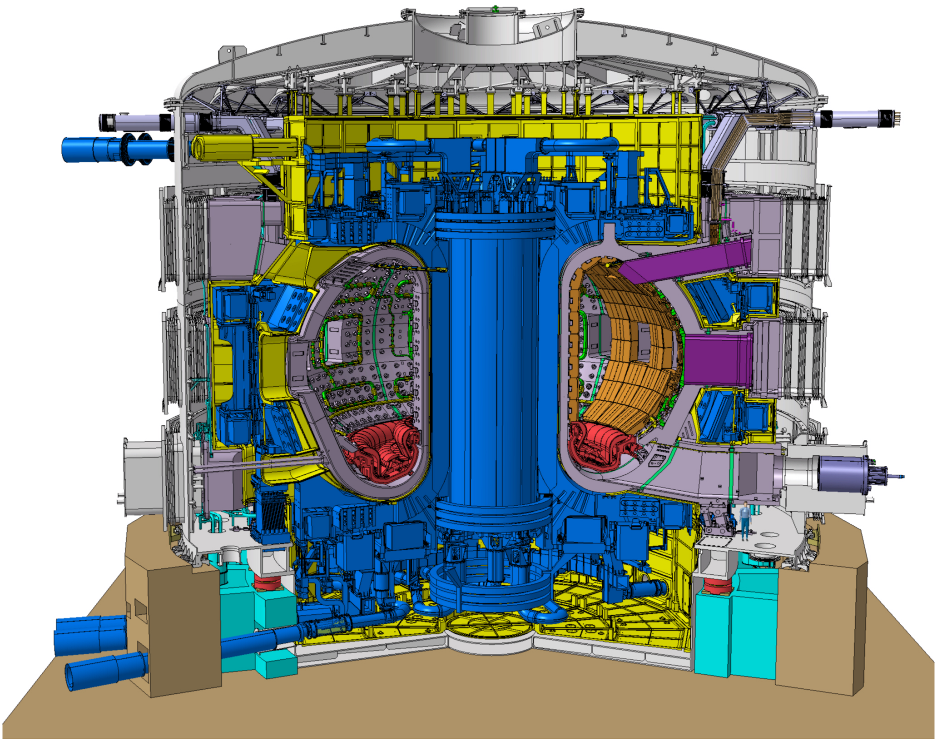


Figure 1.1: A 3D model of ITER, with the magnet system in blue which is composed of the toroidal and poloidal fields systems, the central solenoid, the correction coils, the magnet feeders, and the in-vessel coils. Moreover, the divertor is highlighted in red, the breeding blanket in orange, the vacuum vessel in yellow, and the cryostat in grey [1].

It is important to keep in mind that ITER is still an experimental device and, as such, it will not be able to convert the net heating power it produces into electricity. However, it will demonstrate the safety features of a fusion device and pave the way for machines to do so. Indeed, the ITER successor, Demonstration Power Plant (DEMO), which is an objective of the EUROfusion Fusion Technology Programme, will allow the transition of tokamaks to the world of industry-driven technology [16].

A 3D representation of the ITER design is shown in Figure 1.1, which includes a section of the toroidal vacuum chamber and the main compo-

Plasma volume	$840\, m^3$
Major radius, R_0	$6.2\, m$
Minor radius, R_a	$2.0\, m$
Toroidal field at R_0	$5.3\, T$
Plasma current, I_p	$15\, MA$
Triangularity, κ	[0.33, 0.48]
Elongation, δ	[1.7, 1.85]
Average electron density, n_e	$10.1\ \times 10^{19}\, m^{-3}$
Average electron temperature, T_e	$8.8\, keV$
Average ion temperature, T_i	$8.0\, keV$
NBI power	$33\, MW$
RF power	$40\, MW$
Fusion power	$400\, MW$
Fusion gain, $\mathcal{Q}$	10
Non inductive current	$28\, \%$
Burn time	$400\, sec$
Blanket thermal load	$736\, MW$
Divertor heat load	$20\, MW/m^2$
Power supply	[100, 620] MW

Table 1.1: ITER technical data specifications and main parameters [5].

nents of the tokamak [5]. Specifically, the vacuum vessel is highlighted in yellow, the breeding blanket in orange, the divertor in red, the magnet system in blue, and the cryostat in gray.

However, they are only a portion of the entire ITER facility, where a variety of auxiliary systems are being designed to ensure the safe and effective operation of the tokamak. One of these is the heating and current drive system, which supplies the extra energy needed to heat the plasma. This system incorporates technologies such as radio frequency heating, neutral beam injection, and electron cyclotron heating. In addition, other essential components are the plasma control system, which utilizes complex controllers to adjust the magnetic fields and other parameters to ensure stability and optimize performance, and the diagnostic system, which is equipped with advanced tools to monitor plasma behaviors and measure plasma properties. In reality, ITER will also be a test

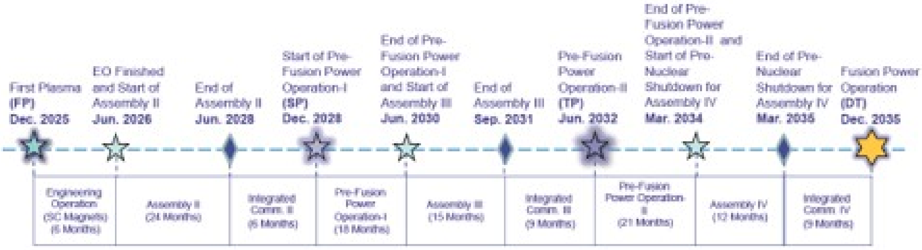

First Plasma (FP) Dec. 2025
EO Finished and Start of Assembly II Jun. 2026
End of Assembly II Jun. 2028
Start of Pre-Fusion Power Operation-I (SP) Dec. 2028
End of Pre-Fusion Power Operation-I and Start of Assembly III Jun. 2030
End of Assembly III Sep. 2031
Pre-Fusion Power Operation-II (TP) Jun. 2032
End of Pre-Fusion Power Operation-II and Start of Pre-Nuclear Shutdown for Assembly IV Mar. 2034
End of Pre-Nuclear Shutdown for Assembly IV Mar. 2035
Fusion Power Operation (DT) Dec. 2035
Engineering Operation (SC Magnets) (6 Months)
Assembly II (24 Months)
Integrated Comm. II (6 Months)
Pre-Fusion Power Operation-I (18 Months)
Assembly III (15 Months)
Integrated Comm. III (9 Months)
Pre-Fusion Power Operation-II (21 Months)
Assembly IV (12 Months)
Integrated Comm. IV (9 Months)

Figure 1.2: Timeline for the ITER operational campaigns (ITER Research Plan [2]).

for the accessibility and integration of all technologies essential for fusion reactors.

The ITER operational campaigns are specified in the ITER Research Plan [2]. This document defines the current plan for the exploitation of the available experimental capability to meet the ITER goal of nuclear operation in the late 2035. Specifically, it follows the intent to commission certain control capabilities, establish specific plasma scenarios, address uncertainties in the physics of plasmas at the ITER scale and in the burning plasma regime, and finally perform detailed scientific studies of burning plasmas while optimizing fusion performance.

The main stages of the operational schedule, illustrated in the timeline of Figure 1.2, are:

- **First Plasma (FP)**: achievement of the first plasma breakdown in hydrogen or helium with at least 100 kA of plasma current and for at least 100 ms.

- **Pre-Fusion Power Operation (PFPO)**: operations in hydrogen and helium and with diverted plasmas. These scenarios are supported by a program of plasma control, diagnostics, additional heating, and current drive commissioning. The development of the H-mode operations is also expected.

- **Fusion Power Operation (FPO)**: operations in deuterium and deuterium-tritium plasmas to demonstrate the production of a fusion power of several hundred MW for several tens of seconds at a Q value in the range of $5 - 10$.

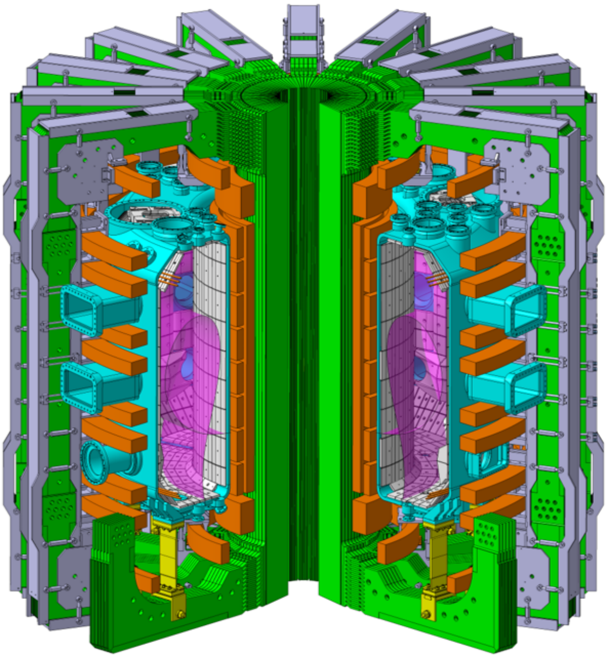


Figure 1.3: A 3D representation of TCV, showing the vacuum vessel in cyan and the carbon tiles inside the vessel in light gray, the toroidal field coils in green, the poloidal field and Ohmic coils in orange [3].

The technical data specifications and machine parameters are reported in Table 1.1.

1.1.2 The TCV tokamak

The TCV tokamak at the Swiss Plasma Center of the École Polytechnique Fédérale de Lausanne in Switzerland is a medium-sized device that is renowned for its high flexibility in creating a variety of poloidal magnetic configurations and its access to cutting-edge plasma diagnostics. It is equipped with sixteen Poloidal Field (PF) copper-made coils, also known as *shaping* coils (highlighted in orange in Figure 1.3), which can be powered independently, allowing the formation of a wide range of plasma shapes, including positive and negative triangularity, and even doublets.

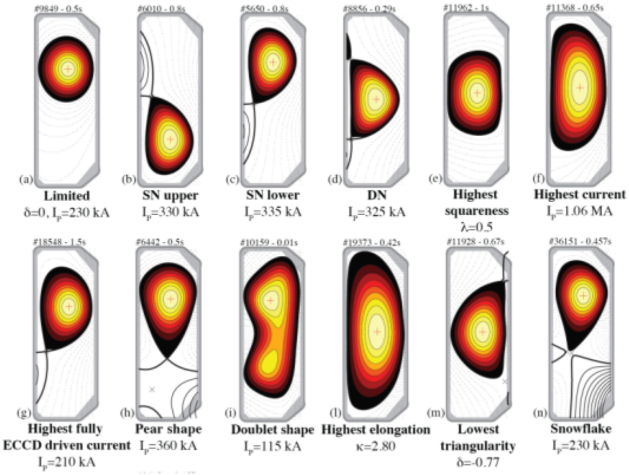


Figure 1.4: A series of plasma shapes realised in TCV [4].

In Figure 1.4 a series of plasma shapes obtained during TCV operations is reported. This highly versatile feature combined with a modern diagnostic system makes TCV an ideal device for investigating the physics of magnetically confined plasmas and assessing sophisticated real-time magnetic control systems.

Figure 1.3 shows a 3D representation of the TCV machine. The vacuum vessel is highlighted in cyan, the carbon tiles inside the vessel in gray, the toroidal field coils in green, and the poloidal field and Ohmic coils in orange [3]. The technical data specifications and machine parameters are provided in Table 1.2.

1.1.3 The EAST tokamak

The EAST machine is the first superconducting tokamak in the world [17]. Its superconducting magnet system and noncircular cross section are advantageous for the investigation of advanced steady-state plasma operation modes. The purpose of EAST is to achieve plasma pulses of up to 1000 *sec*. In addition, in recent years, plasma-facing components, plasma heating, diagnostics, and other systems have been upgraded, making EAST a significant test bench for studying long-pulse steady-state operations and

Major radius, R_0	$0.89\,m$
Minor radius, R_a	$0.25\,m$
Toroidal field at R_0	$1.54\,T$
Plasma current, I_p	$1\,MA$
Triangularity δ	$[-0.6,\ +0.9]$
Elongation κ	$[0.9,\ 2.8]$
Core electron density, n_e	$[1,\ 20]\ \times 10^{19}\,m^{-3}$
Core electron temperature, T_e	$\leq 1\,keV$ (ohmic) $\leq 15\,keV$ (ECH)
Core ion temperature, T_i	$\leq 1\,keV$
NBI power	$2.3\,MW$
RF power	$4.5\,MW$
Burn time	$3\,sec$

Table 1.2: TCV technical data specifications and main parameters [4].

Major radius, R_0	$1.85\,m$
Minor radius, R_a	$0.45\,m$
Toroidal field at R_0	$3.5\,T$
Plasma current, I_p	$1\,MA$
Triangularity δ	$[0.6,\ 0.8]$
Elongation κ	$[1.6,\ 2]$
LHCD power	$3\,MW$
ICRH power	$4\,MW$
Burn time	$[1, 1000]\ sec$

Table 1.3: EAST technical data specifications and main parameters [6].

conducting ITER-like advanced plasma science and technology research.

The technical data specifications and machine parameters are reported in Table 1.3.

1.2 Plasma magnetic control

The plasma magnetic control in a tokamak is a sophisticated system designed to monitor and optimize the magnetic field in real-time. It is the key to crate and precisely control the plasma inside the vacuum chamber, preventing it from coming into contact with the device walls.

Through magnetic confinement, the behavior of charged particles moving in a magnetic field is exploited: charged particles experience a force perpendicular to their motion, which causes them to spiral along the magnetic field lines, rather than moving in straight lines. In a tokamak, the magnetic field is carefully designed to create a helical path for the charged particles which confines the plasma within the chamber's central region. It is created by a combination of external magnets and a current flowing through the plasma itself. In particular, a toroidal field component is produced by a set of superconductive coils wrapped around the vacuum vessel (see the blue coils in Figure 1.5), while a poloidal one is generated by the presence of a plasma current induced in the ionized gas, and by a set of toroidally continuous coils (in gray in Figure 1.5), called PF coils. These two field components create a complex magnetic topology that is optimized for plasma confinement and must be precisely controlled to maintain stability, avoid plasma-wall interactions [18] and prevent disruptions [19, 20][1].

Feedback control of the poloidal magnetic field is one of the fundamental problems that need to be tackled to operate a tokamak. It is called *magnetic control problem* and includes the control of the current induced into the plasma, as well as the shape and position of the plasma by regulating the currents flowing in the PF coils [21].

In particular, the currents applied to the PF circuits are exploited to confine the hot plasma through the pulse phases, which define the so-called plasma scenario. Indeed, at the start of the discharge, the PF currents are used to generate the magnetic field that is necessary to achieve the conditions for plasma formation inside the vacuum chamber, the so-called breakdown and burn-through phases [22]. Immediately after the plasma is formed, the magnetic field generated by the PF coils needs to be controlled to induce a current in the plasma itself, increase it during the ramp-up phase, keeping it almost constant during the flat top, and then ramp it

[1] Plasma disruption in tokamaks refers to the sudden loss of plasma confinement and stability, leading to a rapid release of energy and potentially damaging effects on the tokamak. This can be caused by a variety of factors, including instabilities in the plasma or disruptions in the magnetic fields that control it.

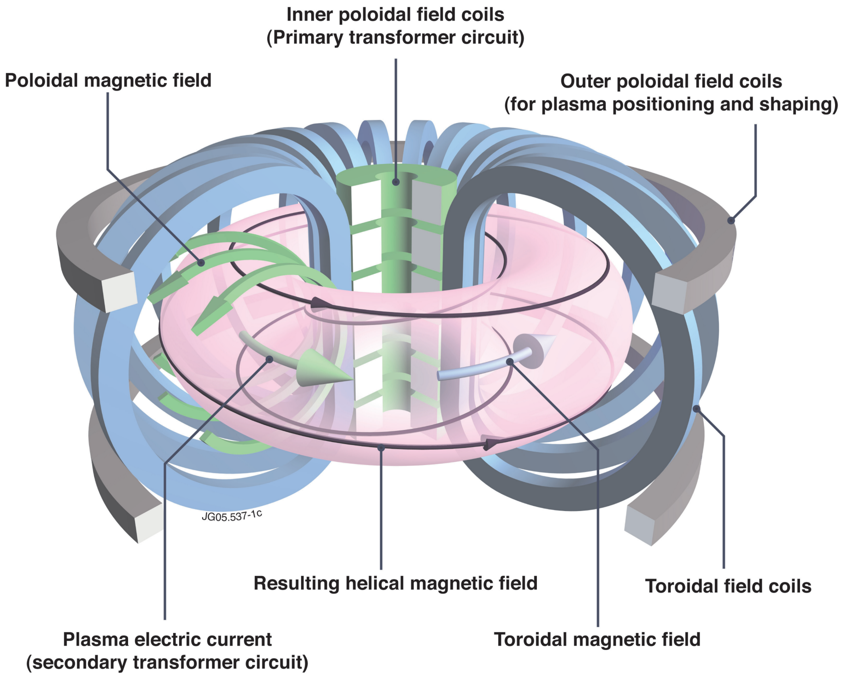

Inner poloidal field coils
(Primary transformer circuit)
Poloidal magnetic field
Outer poloidal field coils
(for plasma positioning and shaping)
JG05.537-1c
Resulting helical magnetic field
Toroidal field coils
Plasma electric current
(secondary transformer circuit)
Toroidal magnetic field

Figure 1.5: Simplified scheme of a tokamak fusion device.

down during the final phase of the discharge. Figure 1.7 illustrates and explains the different phases of a plasma pulse, based on the time behavior of the plasma current envisaged for the ITER tokamak. In addition to controlling the plasma current, the PF coils are used to adjust the shape and position of the plasma to maximize its performance and meet the desired experimental objectives.

Furthermore, to increase the energy confinement time, which is a vital criterion for obtaining fusion reactions, modern devices tailor vertically elongated plasma shapes [23] (as an example, see the elongated cross-section of the ITER plasma reported in Figure 1.6). The downside is that with this *diverted* shapes the plasma turns out to be vertically unstable. Active control of currents in some of the PF coils is mandatory to generate the radial field required to vertically stabilize the plasma column. In particular, an active feedback system, called Vertical Stabilization (VS),

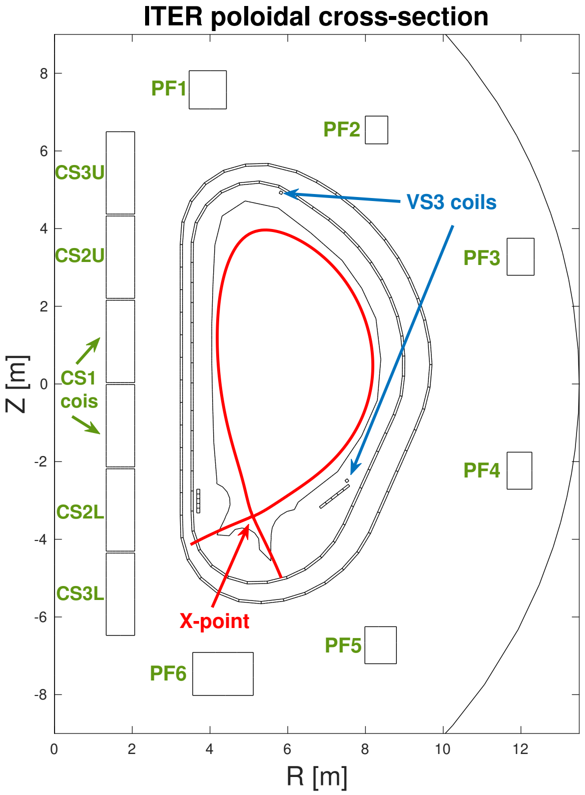


Figure 1.6: A poloidal cross-section of an elongated ITER tokamak plasma. The red line outlines the boundary of the elongated plasma. The green coils represent the superconductive PF circuits, while the blue coils are part of the VS3 circuit, which is the actuator for the ITER VS system.

became essential to run the machine. In Figure 1.6 the copper-made coils positioned inside the tokamak vessel usually dedicated to the VS, which at ITER form the so-named $VS3$ circuit, are highlighted.

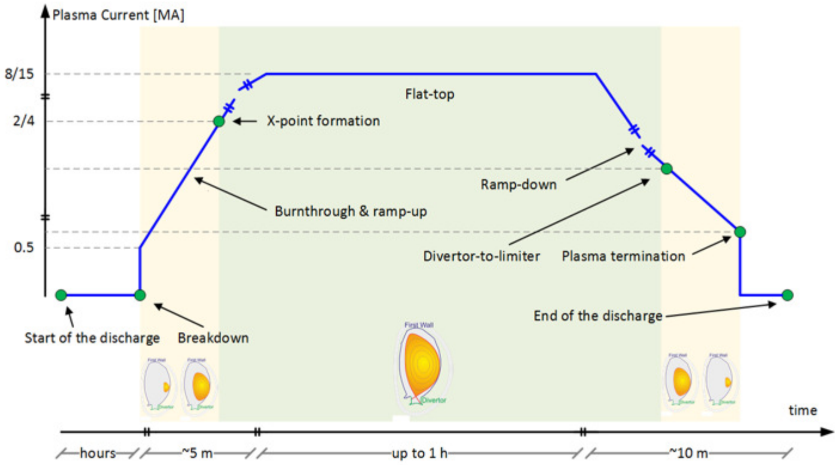


Figure 1.7: The ITER tokamak's plasma discharge is composed of several phases. The first is the breakdown and burn-through, when the hydrogen gas is heated to the point of ionization, forming a plasma. This is followed by the ramp-up phase, where more energy is added to the plasma, increasing the plasma current. During the flat-top, the plasma has reached a steady state and can be sustained for a longer period. Finally, the ramp-down phase begins, where the plasma loses energy and returns to its original state as a gas until the plasma termination.

1.2.1 Backgrounds

In general, the plasma in a tokamak is a complex, strongly nonlinear system; its equilibrium state must satisfy the Grad-Shafranov equation [24], which relates the pressure and current density in the plasma to the magnetic field. From the magnetic control point of view, a plasma equilibrium can be specified in terms of nominal values of the plasma current $I_{p_{eq}}$, of the currents in the PF circuits $I_{PF_{eq}}$, and of both the poloidal beta $\beta_{p_{eq}}$ and the plasma internal inductance $l_{i_{eq}}$[2]. Indeed, although the behavior of the overall plant is non-linear, around a given equilibrium, the behavior of the plasma and the currents in the surrounding coils can be conveniently

[2] For a given equilibrium, the two parameters β_p and l_i provide a synthetic measure of the plasma internal distributions of pressure and poloidal current.

described by the linear model [25]

$$\dot{x}(t) = \mathbf{A}x(t) + \mathbf{B}u(t) + \mathbf{E}\dot{w}(t) \tag{1.1a}$$

$$y(t) = \mathbf{C}x(t) + \mathbf{F}w(t)\,, \tag{1.1b}$$

where:

- the state vector $x = \begin{pmatrix} \delta I_{PF}^T & \delta I_e & \delta I_p \end{pmatrix}^T$ holds the variations of the PF currents, of the currents in the passive conductive structures, and of the plasma current.

- the input vector u consists of the voltage variations applied to the PF control circuits;

- $w = \begin{pmatrix} \delta\beta_p & \delta l_i \end{pmatrix}^T$ is the vector of the variations of β_p and l_i, that, in terms of magnetic control, can be regarded as exogenous disturbances.

In almost all present-day tokamaks the PF circuits can be further partitioned into two parts: $I_{PF} = \begin{pmatrix} I_{SC}^T & I_{VS} \end{pmatrix}^T$. The first part, I_{SC}, is the vector of currents in the ex-vessel superconductive circuits used for plasma current and shape control, while I_{VS} is the current in the in-vessel copper-made circuit dedicated to vertical stabilization. The same partition holds for the input voltages $u = \begin{pmatrix} u_{SC} & u_{VS} \end{pmatrix}^T$. As an example, for the ITER tokamak I_{SC} are the currents in the 11 superconductive PF circuits (the green-labeled coils in Figure 1.6) used as actuators by the Multiple Inputs Multiple Outputs (MIMO) plasma current and shape controller, while I_{VS} is the current in the in-vessel circuit, called $VS3$ (see coils labeled in blue also in Figure 1.6). Moreover, to accurately model the behavior of the eddy currents I_e, the passive structure of the ITER tokamak is discretized into 110 circuits. Therefore, the order of the model (1.1), in the case of ITER, is about 120.

The simplified block diagram of a possible plasma magnetic control architecture is reported in Figure 1.8. This architecture, in addition to being widely adopted in many operating tokamaks, such as JET [26] and EAST [27], is also the one currently considered for ITER [28, 29].

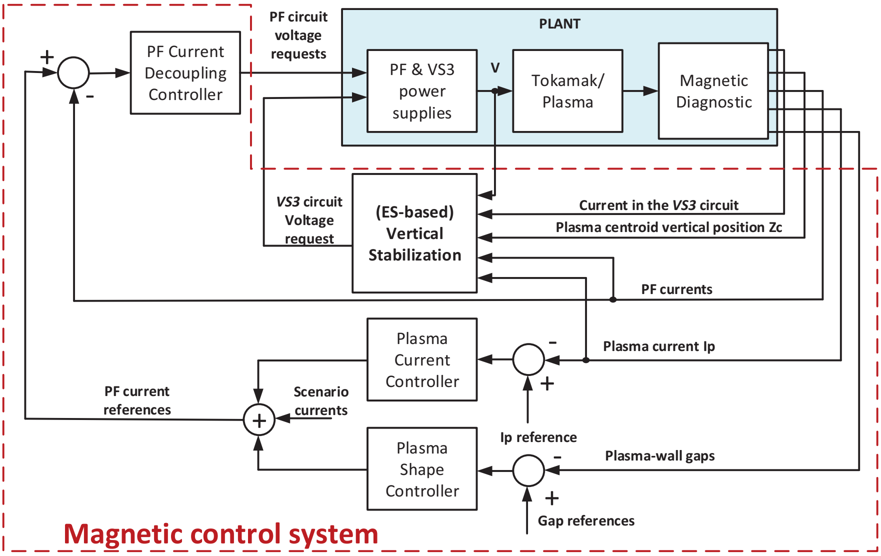


Figure 1.8: Block diagram of a typical ITER-like architecture for plasma magnetic control in tokamaks.

The main components of the plasma MCS shown in Figure 1.8 are[3]:

- The **Poloidal Field Current (PFC) Decoupling Controller**, this block acts as the inner control loop of a nested architecture that also includes the plasma current and shape controllers. By generating the required voltages to be applied to the superconductive coils, this block tracks the PF current references, which are a sum of the *scenario* (i.e., the nominal) currents and the corrections requested by the outer loops to track the desired plasma shape and current;

- The **Plasma Current Controller**, which tracks the plasma current reference by sending the correspondent requests to the PFC Decoupling Controller;

- The **Plasma Shape Controller**, which controls the shape of the

[3] For more details on the control algorithms implemented by the various blocks shown in Figure 1.8, the interested reader can refer to [21] or [30].

last closed flux surface within the vacuum chamber by tracking a set of plasma shape descriptors; this block also generates requests for the PFC Decoupling Controller.

- The **Vertical Stabilization** (VS) system, which is in charge of vertically stabilizing the plasma column.

1.2.2 The Vertical Stabilization problem

The inclusion of a VS system in any MCS is mandatory since the shapes commonly pursued in modern tokamaks are characterized by large elongation. In fact, the performance of controlled fusion devices can be measured by the triple product $n_i \tau T_i$, where n_i is the density of the ion, τ is the energy confinement time and T_i is the temperature of the ion. At the end of the last century, there was a four-fold increase in the maximum value of the triple product. This improvement was mainly due to the transition from a circular to a noncircular plasma cross section. Scientists discovered that plasma with an elongation κ[4] up to 2.5 comes with higher values of the plasma current I_p and nominal β_p[5], therefore increasing the maximum achievable efficiency [31].

However, because of the configuration of the magnetic field used to produce the elongated shapes, these plasmas are vertically unstable. Indeed, taking into account the cartoon picture reported in Figure 1.9 with a plasma filamentary model, the elongated shape can be obtained using the magnetic field generated by the currents $\overrightarrow{I}_1$ and $\overrightarrow{I}_2$ flowing in the PF coils. These two currents, flowing in the same direction as the plasma current, are able to stretch the plasma by generating the forces $\overrightarrow{F}_{1p}$ and $\overrightarrow{F}_{2p}$ in the direction specified in figure. With the plasma in equilibrium position, the two forces balance each other. However, since the magnetic field is proportional to the distance between the plasma and the coil, given a small vertical displacement δZ , the plasma is attracted towards one of the coils as the corresponding field becomes stronger. In Figure 1.9 the

[4] The plasma elongation κ is defined as the ratio between the vertical and horizontal axis of the plasma cross-section. The value of the plasma current scales as $I_p \sim (\kappa^2+1)/\kappa$

[5] The value of β_p represents the ratio of the kinetic plasma pressure energy density to the magnetic field energy density. It is a measure of the effectiveness with which the magnetic field confines the plasma.

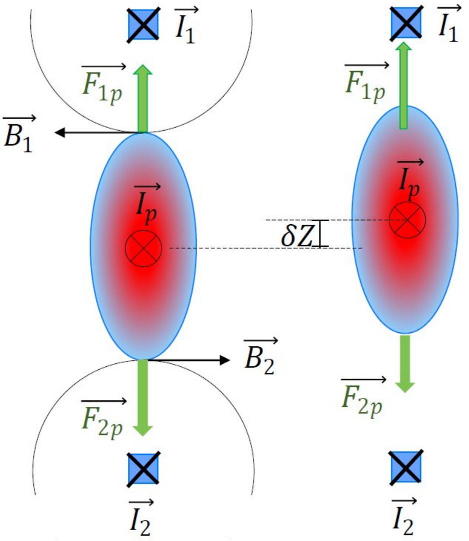


Figure 1.9: Illustration of plasma vertical instability

case of vertical displacement in the upper direction and the consequent distortion of the force balance are shown. Without active control, even slight vertical movements of the plasma column can lead to a loss of confinement and energy.

It follows that the VS block in Figure 1.8 is an essential component of the MCS to run tokamak discharges with elongated plasmas. Note that the vertical plasma mode is stabilized due to a combination of passive elements and active feedback coils. Indeed, the eddy currents flowing in the surrounding passive structures produce a stabilizing action that can counteract the vertical drift of the plasma, bringing the instability characteristic time to a scale that can then be controlled by an active stabilization circuit. Virtually all present-day tokamaks are built with a set of dedicated PF coils used mainly to produce a radial magnetic field that can again balance the forces in Figure 1.9 and stop the plasma column. Specifically, this field is generated by a current flowing in the opposite direction to the plasma current. For the ITER tokamak, the VS dedicated circuit is the copper made $VS3$ one (see Figure 1.6).

Nowadays, the VS control problem is usually solved using model-based control techniques. These techniques rely on control-oriented models that describe the response of the plasma and currents in the surrounding conductive structures for a considered magnetic configuration. Indeed, the vertical instability of the plasma is revealed in the presence of a real and positive eigenvalue in the dynamic matrix of the linearized plasma response model (well-known as *plasma equilibrium*). In the nuclear fusion community, this unstable eigenvalue is an estimation of the so-called *growth rate* γ of the instability, and the associated eigenvector describes the behavior of the plasma and the currents in the conductive structures along the unstable direction.

As matter of fact, the performance of any existing VS system strongly depends on the growth rate, which varies both based on the specific plasma configurations and also throughout a single plasma discharge. Specifically, it increases rapidly with elongation [32]. To achieve the required robustness, it is very common to resort to adaptive control strategies that take into account the features of the considered plasma scenario and of the specific experimental device. For example, in [27] a model-based ITER-like VS has been tested on the EAST tokamak; another option can be to adapt the VS parameters according to an empirical relationship between γ and some measurements [26], or tuning the VS parameters considering an envelope of possible plasma models [33] and exploiting multi-objective optimization [34]. More examples are [26] for the VS system at JET, [35] for the DIII-D system, [36] for ITER, and [37] for DEMO. Similar considerations hold when non-linear control approaches are considered, as in the case of JET [38] and TCV [39].

However, such model-based controllers have certain drawbacks, due to the complexity and uncertainty of plasma dynamics. Plasma behavior is highly nonlinear and can be affected by a wide range of factors, such as instabilities, turbulence, and disruptions. This makes it difficult to obtain an accurate and reliable model for the plasma dynamics, leading to potential inaccuracies in the controller's performance.

Additionally, the plasma's dynamic nature can cause a discrepancy between the model's predictions and the plasma's actual behavior. This can lead to inadequate control performance and can even destabilize the plasma, resulting in disruptions and energy loss. Therefore, algorithms with few control parameters are usually preferred [26, 40], as they allow one to use effective adaptive algorithms and guarantee robustness in various scenarios.

Furthermore, the complexity of tokamak systems with high-dimensional plasma dynamics can make the computational requirements for reliable control-oriented models and for the estimation of the relevant plasma parameters not feasible when compared to the time scale in which the controllers have to act. This can be a hindrance to the practical implementation of model-based controllers in real-world tokamak experiments and commercial fusion reactors.

Overall, model-based controllers have demonstrated potential in vertically stabilizing tokamak plasma, even if they require access to an accurate plant model to analytically implement the controller. However, the complexity and unpredictability of plasma dynamics, the possibility of model mismatch, and the computational demands for real-time models can restrict their effectiveness. Addressing these limitations will be essential to further the development of practical and scalable control systems for nuclear fusion reactors.

An alternative option to achieve the required level of robustness without requiring knowledge of a plasma model is represented by model-free and data-driven approaches. Indeed, with model-free approaches, it is possible to rely on the controller agnosticism with respect to the plant model to increase robustness, while the data-driven ones learn the desired plant behavior via extensive simulation campaigns and/or access large experimental data sets.

1.3 Contribution

This thesis contributes to the research on the VS problem in tokamak plasmas by developing control strategies that can guarantee the desired level of performance without needing to know the details of the plant model and aiming at improving the robustness of the overall plasma magnetic control system. In particular, it explores the possibility of resorting to model-free and data-driven approaches, such as:

- The design of model-free VS system based on the Extremum Seeking (ES) algorithm. Specifically, different versions of a ES-based VS system have been designed to provide the control voltage for the VS circuit able to vertically stabilize the plasma column. Starting from the crude ES control law, several features have been added to the control scheme, such as a switching power supply for the VS circuit and an automaton for an event-driven control gain adaptation logic, with the aim of increasing the robustness and generalization properties of the VS system. Most of the approaches have been tested in simulation considering the ITER plasma MCS as case study. Both linear and non-linear simulations have been performed to assess the effectiveness of the proposed control architecture and validate its capability to counteract the relevant disturbances that can occur during ITER operations. Moreover, the ES controller has been also enhanced with Neural Network (NN)s to improve its robustness even in the presence of significant model uncertainties. Specifically, the possibility of using either a Linear Regressor (LR), a Multilayer Perceptron (MLP), an Extreme Learning Machine (ELM), or a Long Short Term Memory (LSTM) network have been considered.

The ES-based algorithm has been also adapted for the VS of the TCV tokamak. Preliminary simulations have been carried out to validate the approach with the plan of testing it also in an experimental campaign next year.

- The setup of a Reinforcement Learning (RL) algorithms proposed as potential data-driven solutions for the VS problem. RL agents have been trained to counteract the vertical instability of the plasma by interacting with a linearized plasma model in simulation. To show the effectiveness of the proposed RL-based VS approaches, plasma models, different from those used for training, have been considered to test the obtained agents in various operational scenarios. More in detail, a tabular Q-learning algorithm was firstly developed for the VS system of the EAST tokamak. Subsequently, a more complex Deep Reinforcement Learning (DRL) algorithm, such as Deep Deterministic Policy Gradient (DDPG) was explored including the whole ITER MCS in the environment. DDPG exploits deep NNs to approximate the behavior of the agent, and because of this it is specifically designed to handle problems with continuous action and state spaces, which are essential in this case for a fair representation of plasma behavior. Furthermore, to better understand DRL algorithms, a sensitivity analysis was performed to choose the best-performing set of DDPG hyperparameters. Specifically, the effects of these hyperparameters and their tuning have been analyzed with respect to the convergence of rewards.

1.4 Thesis layout

The thesis is organized into two parts.

Chapter 1 introduced the reader throughout the field of nuclear fusion, illustrating the peculiarities of tokamak devices that will be considered in the remainder of the thesis. The plasma control problem was also addressed together with the plasma control-oriented model and the main components of a plasma MCS used, specifically, to design and assess the VS solutions proposed in this thesis. Particular attention was given to the VS problem discussing the motivation for research in the field of model-independent controllers for the VS of tokamak plasmas. A comprehensive review of the model-based techniques usually employed for the VS system was reported together with the thesis's main contributions.

The first part of the thesis deals with the development of a model-free VS system based on the ES algorithm for stabilization. First, in Chapter 2 the ES control algorithm studied in this thesis to tackle the VS problem is described in its unbounded and bounded versions. Then, Chapter 3 begins from the ES basic approach and then introduces the different features added to the control scheme to improve performance, ending with the most enhanced version in which a power supply characterized by switching behavior is used in combination with an even driven adaptation logic for the control gain of the ES bounded control law. In this chapter, the ITER tokamak has been considered as a test case and both linear and non-linear numerical simulations have been performed for the robustness assessment of the ES-based VS system. In Chapter 5, on the other hand, the VS system of the TCV tokamak is considered as a test case for the proposed ES approach.

Finally, in Chapter 4 is presented the final version of the ES-based VS system where the model-agnostic ES algorithm is combined with a NN. After an extensive review of the use of NN in the nuclear fusion community, a complete model-free control scheme for the VS system of the ITER tokamak is presented, highlighting how the use of a data-driven NN allows to increase the robustness of the approach.

The second part of this thesis treats the development of data-driven VS systems using RL algorithms. Specifically, Chapter 6 starts with a general introduction on RL properties and peculiarities and then presents the two algorithms considered in this thesis: Q-learning and DDPG algorithms. In Chapter 7 the Q-learning algorithm is applied to the VS system of the EAST tokamak as a case study. Since the EAST tokmak is in operation, this chapter also discusses the comparison of the results obtained using the Q-learning agent and the real experimental data. Chapter 8 presents the training of a DDPG agent to be used as a substitute of the VS system in the ITER MCS. In this context, a sensitivity analysis is also proposed to study the optimal tuning of the DDPG hyperparameters.

Finally, in Chapter 9 final remarks are given.

Part I

Model-free Vertical Stabilization of tokamak plasma via Extremum Seeking

2
Extremum Seeking for stabilization

THIS CHAPTER introduces the ES control algorithm with a particular focus on the versions that can be used to stabilize unknown unstable systems when used as a feedback controller [41].

The ES control is a widely used model-free online adaptive control algorithm that is especially beneficial in cases where the dynamics of the system are unknown or change over time. This control technique is used to stabilize and control systems by continuously searching for the optimal operating point or extremum of a performance metric. The performance metric can differ depending on the application, such as a cost function that needs to be minimized (e.g. energy consumption or production time), or a quality metric that needs to be maximized (e.g. product yield or system stability). The fundamental concept of ES is to adjust some input or parameters of the process based on the observed performance to drive the system toward the optimal operating point.

The ES does not require explicit knowledge of plant dynamics, as it uses online measurements and optimization techniques to achieve stabilization. This model-free characteristic is one of the primary advantages of ES compared to other optimization methods; it is applicable in situations where available modeling resources are too limited to create a sufficiently precise model for optimization, as well as in cases where fundamental difficulties, such as the presence of uncertainties, prevent accurate modeling. The foregoing makes the ES approach an appealing control method for the VS problem in tokamak plasmas (see 1.2.2 for the limitations introduced by standard model-based controllers). Additionally, as a feedback-based method, it also inherits the flexibility and robustness of this control architecture. Specifically, the feedback structure enables the method not only to locate an optimum solution, but also to track it if it changes over time (e.g., due to disturbances or process dynamics), demonstrating robustness to some of the common types of uncertainty that are present in most process control problems [42].

Since the first proposal in 1922 [43] where it was suggested as a method to optimize power transfer between a train and an overhead power line, ES has been successfully applied in various fields. Its simple and effective optimization scheme gained attention, especially in Russia during the 1940s[44, 45] until it gained a wider audience within the international control community due to the publication of Draper and Li[46]. Moreover, like all other forms of adaptive control, ES was a popular research topic in the 1950s and 1960s. In these decades, many variants of the ES algorithm were explored, in particular, based on specific applications and design issues[47]. Then during the next decades, research related to ES continued steadily[48], experiencing the biggest growth in the practical field of industrial application.

In 2000, the interest in ES sparked again with the breakthrough result by Krstić and Wang [49, 50]. In their works, Krstić and Wang, provided the first rigorous stability analysis of ES showing that it was possible to apply it to a much larger class of plants than had previously been considered.

As already mentioned, ES has been successfully implemented in many different engineering systems, and there have been thousands of publications on theory and application (see [51] for more detail about ES history and behavior characteristics).

This thesis focuses on the use of the ES algorithm to stabilize unknown unstable plants, as presented in [41] by Schienker and Krstić. The idea exploited is to use the ES method to minimize a candidate Lyapunov function (clf) for an unstable system using a Lie-bracket averaging technique. In fact, averaging is an important tool in the analysis of this type of ES controllers. Even if the ES approach does not rely on the system model, yet still achieves the same effect as a clf-based feedback laws (which employ the full modeling knowledge), but in a time-average sense. The combination of clf with the ES periodic perturbation signal leads to semi-global and practical asymptotic stability (instead of a classic global stability property), which can be a reasonable trade-off for model-free stabilization with a very simple control algorithm.

In particular, it has been shown that given the nonlinear systems affine in control

$$\dot{x}(t) = f(x,t) + g(x,t)u\,, \tag{2.1}$$

it is possible to employ the nonlinear time-varying control law

$$u = \alpha\sqrt{\omega}cos(\omega t) - k\sqrt{\omega}sin(\omega t)V(x)\,, \tag{2.2}$$

where $V(x)$ is a clf for (2.1). The scaling factor $\sqrt{\omega}$ is to amplify the mixing and dithering terms to appropriately increase them; otherwise, highly oscillatory components may have little or no influence on the overall dynamics of the system.

It follows, considering the corresponding Lie Bracket average system

$$\dot{\bar{x}} = f(\bar{x},t) - k\alpha\, g(\bar{x},t)g^T(\bar{x},t)\left(\frac{\delta V(\bar{x})}{\delta\bar{x}}\right)^T\,, \tag{2.3}$$

that the choice of a sufficiently high positive gain $k\alpha$ makes the gradient term dominant and the average system asymptotically stabilized. It can be demonstrated, by means of averaging arguments, that the trajectories of the initial system (2.1) can be kept arbitrarily close to those of the averaged system (2.3), provided that the frequency ω is chosen high enough. This guarantees that all the trajectories of the original system are confined to a neighborhood of the averaged ones, making the system semi-globally practically stabilized (more details can be found in [41]). Finally,although Extremum Seeking for Stabilization (ESS) does not require knowledge of the system model, it assumes that the value of $V(\cdot)$ is accessible.

The ESS is designed to work with unknown systems. Despite the theoretical progress and applications that have followed from the first definition of (2.2), there is still a limitation in terms of the uncertainty of the convergence rate and the control effort. This is because clf, which is an unknown function, is incorporated into the control scheme in an affine way. A new ES scheme is proposed in [52], in which the uncertainty is confined to the argument of a bounded function, resulting in guaranteed bounds on both the update rate in the minimum seeking and the control effort in stabilization. Therefore, in case of minimization of a measurable, but unknown clf a possible nonlinear time-varying bounded control law can be

$$u(t) = \sqrt{\alpha\omega} cos\left(\omega t + kV(x)\right) \ , \tag{2.4}$$

It yields the same Lie Bracket average system as in (2.3), which ensures the same stabilization properties discussed for (2.2).

2.1 ES-based Vertical Stabilization

A possible ES-based control scheme for the VS system of tokamak plasma is reported in Figure 2.1. The proposed architecture can be used to deploy the VS system in the plasma MCS.

As already mentioned, even though ES does not require a plant model, which in the case of magnetic control is a model that describes the behavior of the plasma and currents in the surrounding structures in the form of (1.1), it assumes that the state of the system can be accessed through $V(\cdot)$, implying that some knowledge of the behavior of the system is necessary. In a tokamak plasma, only a subset of the currents in the system state reported in (1.1) can be measured, namely the PF currents I_{PF} and the plasma current I_p, while the eddy currents cannot be easily estimated in real-time. As a matter of fact, it is not possible to define a candidate Lyapunov function solely based on the measurements of I_{PF} and I_p, since the eddy currents play a fundamental role in the dynamic of vertical instability. Without the passive effect of the eddy currents, the vertical instability would be too fast, making it impossible to be actively stabilized. Therefore, the clf for this application is constructed

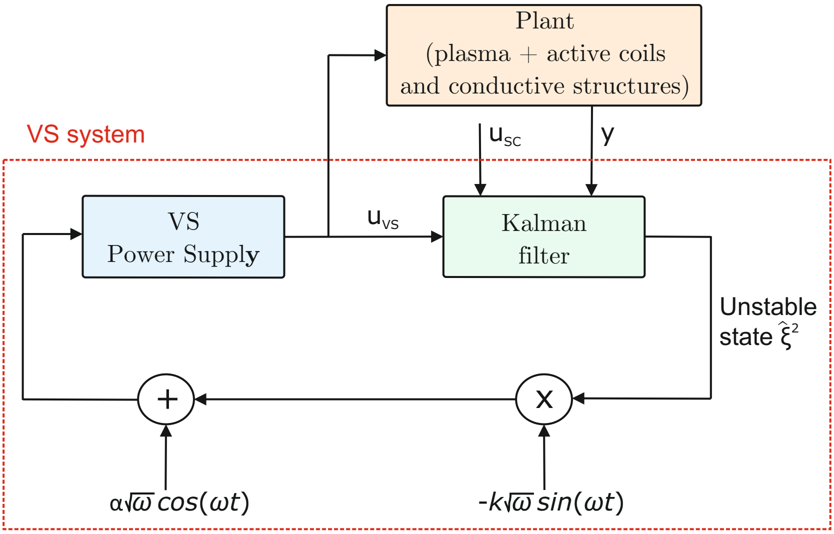


Figure 2.1: The VS system, based on the ESS algorithm, applied to the plant, i.e. to the plasma and the surrounding conductive structures and PF coils.

from the unstable mode of (1.1) and chosen as $V(x) = \xi^2$, where ξ is the movement along the unstable mode of the diagonalized plant state (which is well defined in the case of a linearized system; it is assumed that this definition holds "locally" for the fully nonlinear plant as well).

The value of ξ can be estimated using a Kalman filter. Although such a filter requires the knowledge of a plant model, it is proved that the proposed architecture can cope with relevant model uncertainties because it exploits the *model-agnostic* nature of the ES algorithm. Indeed, since there is no *a priori* guarantee that the estimate of the unstable dynamic provided by the Kalman filter is accurate enough to stabilize the plant, a robustness assessment is carried out *a posteriori* by means of numerical simulations that cover a wide range of plasma parameters and configurations.

As shown in Figure 2.1, the control output provided by the ES law (2.2) is the voltage applied to the dedicated VS circuit (the plant input u_{VS}).

Further, the plant input u_{SC} consists of the voltages applied to the superconductive coils provided by the plasma current and shape controllers. The Kalman filter receives as input, together with u_{VS}, also the u_{SC} vector. This additional input is necessary for the Kalman filter estimation but is not accountable in the VS problem. Furthermore, the Kalman filter receives as input the following plant outputs

$$y = \begin{pmatrix} \delta I_{SC}^T & \delta I_{VS} & \delta I_p & \delta Z_c \end{pmatrix}^T ,$$

where δI_{SC}^T, δI_{VS} and I_p are the current variations in the superconductive coils, in the VS circuit and in the plasma itself, while δZ_c is the variation of the vertical position of the plasma current centroid with respect to its nominal equilibrium value[1]. Finally, the filter is designed by assuming high confidence in the measurements, which is reflected in the choice of almost negligible covariance matrices. The estimation of the dynamic along the unstable mode $\hat{\xi}$ returned by the Kalman filter is then used to compute the clf $V(x) = \hat{\xi}^2$ that is minimized by the ES control algorithm.

The tuning of the control gains k and α can be carried out by means of numerical simulations. However, a first guess for the product $k \cdot \alpha$ can be derived by making negative the single eigenvalue of the average system obtained starting from the first-order reduced model that links the voltage applied to the VS circuit to the unstable state. Indeed, when the reduced first-order linear model is considered, from (2.3) it readily follows that the closed-loop average system is equal to

$$\dot{\overline{x}} = \left(a_{red} - k\alpha \, b_{red} b_{red}^T \right) \overline{x} \, .$$

Hence, if $a_{red} > 0$ is the unstable eigenvalue of the reduced system, the average reduced system is stable if the product $k\alpha$ is sufficiently high. Furthermore, the choice of a high value of the frequency ω can be limited due to the bandwidth of the power supply that feeds the vertical stabilization circuit.

[1] The position of the plasma centroid can be reconstructed starting from the magnetic field and flux measurements by the so-called *plasma magnetic diagnostic* system, which consists of a set of real-time algorithms that reconstruct relevant plasma parameters. The interested reader can refer to [53], for more details).

3 Application of the Extremum Seeking approach to the ITER case

This chapter introduces the different ES-based approaches that have been studied to design a model-free VS system. Some of these approaches have previously been discussed and presented in [54, 55, 56]. Specifically, in [54] the first attempt to apply ESS to the VS system of the ITER tokamak was made. Initially, the unbounded ES law is considered together with a linear power supply used as an actuator for the VS circuit. Subsequently, to improve performance, additional features were added to the VS system architecture. In [56] a switching power supply has been substituted to the VS circuit amplifier allowing the choice of a higher switching frequency in the ES control law. This helped reduce the amplitude of the oscillations in the system response compared to what was presented in [55]. In [56], a bounded version of the ESS control law is considered, which includes a gain adaptation logic to exploit all available voltage levels of the switching power supply. Linear and nonlinear simulations, performed considering the ITER tokamak as a case study, prove that the proposed ES-based VS schemes can satisfy a certain level of robustness regardless of the specific plasma configuration and that each added feature contributes to improving the general performance of the controller. Finally, a revised logic for the event-driven mechanism used to adjust the gain in the bounded ES control law allows to reduce the maximum voltage in the VS circuit by a factor of four compared to what is presented in [56], bringing it closer to the value envisaged for the ITER VS power supply.

Furthermore, in Chapter 4 a *complete* model-free VS control scheme will be developed by combining the ES algorithm with deep NNs.

To demonstrate the efficacy of the proposed approach, each considered VS scheme has been tested on a set of ITER operational scenarios. For this reason, two groups of ITER plasma equilibria described by the linearized models in Equation (1.1) were generated for the simulation purposes of this thesis. The equilibrium parameters of these two families of plasma equilibria are reported in Table 3.1. Both Group A and Group B models refer to the behavior of the plasma and currents in the surrounding structures during the flat-top phase of 15 MA ITER plasma discharges (see Figure 1.7), however, they were obtained through different methods:

- Group A is a collection of models that all have origins in the same plasma equilibrium. By altering the plasma parameters, a single plasma model can be used to generate new equilibria and explore different behaviors. Different values of plasma elongation κ and distributions of plasma internal kinetic profiles result in different values of β_p and l_i. The elongation κ, together with both β_p and l_i, have an impact on the plasma unstable mode through the growth rate γ. Thus, changing these parameters allows to model different behaviors as far as vertical stabilization is concerned. See Table 3.1a for the ranges of plasma equilibrium parameters considered for these linearized models.

- The models in Group B are three different snapshots taken from the same 15 MA scenario. In particular, *Equilibrium #1* and *#3* refer to the beginning of the flat-top, i.e. when I_p reaches 15 MA, while *Equilibrium #2* to the end, right before the beginning of the ramp-down (at ITER the current flat-top at 15 MA can last for minutes).

Group A

Equilibrium #	[1, 24]
$I_{p_{eq}}$	15 MA
$\beta_{p_{eq}}$	[0.1, 1]
$l_{i_{eq}}$	[0.8, 1.3]
κ	[1.77, 1.81]
γ	[2.6, 12.5] s^{-1}

(a) Group A: Family of 24 different ITER plasma equilibria, all at a plasma current of $15MA$, generated to cover the reported intervals for the value of the profile parameters $\beta_{p_{eq}}$, $l_{i_{eq}}$, elongation κ_{eq} and growth rate γ. The corresponding linearized models are the ones considered for the training of the NNs in Chapter 4.

Group B

ITER equilibrium	$I_{p_{eq}}$	$\beta_{p_{eq}}$	$l_{i_{eq}}$	κ	γ
Equilibrium #1	15 MA	0.66	0.88	1.77	4.9 s^{-1}
Equilibrium #2	15 MA	0.82	0.71	1.8	2.9 s^{-1}
Equilibrium #3	15 MA	0.08	0.92	1.86	9.1 s^{-1}

(b) Group B: Plasma equilibrium parameters for the three equilibria obtained as snapshots of the same ITER discharge. For these three equilibria, the values of the corresponding parameters $I_{p_{eq}}$, of $\beta_{p_{eq}}$, $l_{i_{eq}}$, κ_{eq} and γ are reported.

Table 3.1: The equilibrium parameters for the 27 available ITER models used within this work are reported.

These equilibria lead to different γ values due to the different conditions of the plasma. Plus, *Equilibrium #3* was collected before the additional heatings were turned on, making it a more difficult configuration to stabilize. The corresponding equilibrium parameters are reported in Table 3.1b.

It is important to point out that all the results presented in this thesis, to validate the new VS approaches, have been obtained considering not only the VS system, but the whole plasma MCS, whose scheme is reported in Figure 1.8. This means that the control architecture used for the simulations also includes the plasma shape and current controllers. This allowed to observe the behavior of the plasma MCS when the proposed VS solution is employed. Indeed, in this way, it is possible to verify that the plasma boundary does not come into contact with the first wall during transients, that the plasma current is not excessively perturbed, and that the desired plasma shape is successfully recovered at steady state.

In particular, the shape control algorithm adopted is the so-called eXtreme Shape Controller (XSC), which in this case controls 29 plasma wall distances, called *gaps*, with a settling time of about 10 s. To control a number of plasma shape descriptors, i.e. the 29 gaps, greater than the number of available actuators (i.e., the 11 currents in the superconductive coils), the XSC design is based on a Singular Value Decomposition of the static relationship between the control inputs and outputs. In this way, it is possible to minimize in the least mean-square sense the control error at steady state. For more details about the XSC the interested reader can refer to [57, 58].

3.1 Linear power supply

The first attempt to use the ESS algorithm for the VS of ITER involves the control law (2.2), as proposed in [54]. The behavior of the power supplies for both the PF and VS circuits is modeled as a straightforward amplifier. To be more precise, they have been designed as a saturation plus a pure time delay τ_1 in series with a first-order dynamic characterized by a pole at $1/\tau_2$ [40]. The parameters for the ITER power supply model dedicated to the VS circuit are listed in Table 3.2a.

The bandwidth of this type of power supply limits the rate of convergence of the ES algorithm and affects the choice of ω in (2.2). The frequency of the dithering and mixing terms in (2.2) has been set equal to $\omega = 20\pi\ rad/s$ ($10\ Hz$).

The values of the ES law (2.2) control parameters considered for this specific application are reported in Table 3.3.

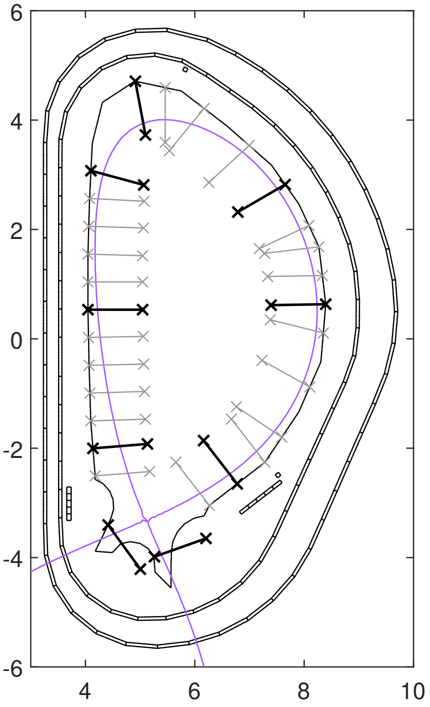


Figure 3.1: Gaps used for the shape control. The gaps shown in black are the ones whose behavior is reported in Figures 3.11.

Results

A set of operation scenarios for the ITER tokamak is considered to show the effectiveness of the proposed VS system.

Two different plasma equilibria from Group B (see Table 3.1b) have been considered for the simulations, namely *Equilibrium #1* and *Equilibrium #2*. The first corresponds to a linearized model obtained at the beginning of the *flat-top* phase of a 15 MA ITER discharge, while the latter is from the end of the flat-top. It follows that the behavior of the plant around the two considered equilibria is modeled by two different linearized models in the form (1.1), whose equilibrium values are reported in Table 3.1b. However, in all operational scenarios considered, the same

Linear power supply

τ_1	τ_2	$u_{\max}$
2.5 ms	7.5 ms	2.3 kV

(a) Parameters of the ITER $VS3$ linear power supply model. τ_1 is the pure delay, τ_2 is the time constant of the first-order response, while $u_{\max}$ is the maximum absolute output voltage (the $VS3$ power supply is a four quadrants one).

Switching power supply

Parameters	ERFA
Max output voltage	±12 kV dc
Max output current	±5 kA dc
Max output voltage step	±3 kV
Time for full ± voltage excursion	$\leq 100\mu$ s
Max switching frequency	1 kHz

(b) Main parameters of the fast switching power supply.

Table 3.2: Parameters of the linear amplifier and the switching power supply employed as actuators of the VS circuit for the ITER tokamak.

configuration of the VS system reported in Figure 2.1 is used. This implies that the same Kalman filter, as well as the same ES parameters reported in Table 3.3 were used in all simulations. In particular, the Kalman filter was designed considering a reduced linearized model of order 25 for *Equilibrium #1*.

The test scenarios considered refer to the counteraction of relevant disturbances that can occur during ITER operations. In particular, the following cases were considered:

- the rejection of a Vertical Displacement Event (VDE) [40] of 5 cm;
- the response to a Minor Disruption (MD);

k	α	ω
$7.2 \cdot 10^{-3}$	50	20π rad/s

Table 3.3: Control parameters for the proposed VS system based on the ES control law (2.2) when a linear amplifier is used as actuator for the VS circuit.

- the response to an H to L transition.

VDE rejection A VDE is an uncontrolled growth of the unstable vertical mode of the plasma. Although the plasma is always vertically controlled, in practice these uncontrolled growths can occur for various reasons, such as fast disturbances acting on a time scale that is outside the VS bandwidth, unforeseen delays in the control loop, or wrong control action due to measurement noise when the plasma velocity is almost zero. Since it models various types of disturbances the VDE became a standard benchmark to assess VS performance [59].

From the magnetic control point of view, a VDE is equivalent to an almost instantaneous change in the position of the plasma. Indeed, it can be modeled as an instantaneous change of the state vector in (1.1) along the unstable mode, scaled to produce the prescribed vertical displacement of the plasma centroid (see also [40]). After a VDE, the VS system must be able to stop the vertical plasma motion and, together with the position and shape controller, bring the vertical position of the plasma centroid back to the equilibrium value.

To evaluate the behavior of the proposed VS approach in the case of VDEs, the response of the overall MCS to a 5 *cm* VDE for *Equilibrium #2* is performed. The simulation results are shown in Figure 3.2, where the displacement with respect to the equilibrium position of the vertical position of the plasma centroid δZ_c, the voltage in the $VS3$ circuit u_{VS} and the plasma current I_p are shown. The proposed VS system stabilizes the plasma column and recovers the original position for Z_c while minimizing the induced variation on I_p. By including the shape controller in the simulation scheme, it was possible to verify that the plasma boundary does not touch the vessel during the transient and that the desired

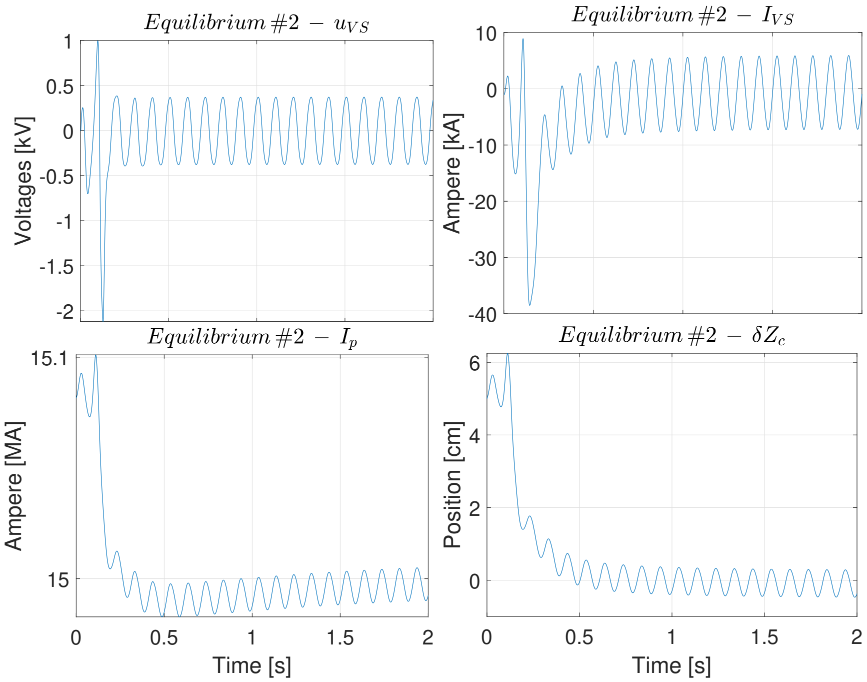


Figure 3.2: The Response to a 5 *cm* VDE applied to the *Equilibrium #2* plasma. The displacement of the plasma current centroid vertical position δZ_c, the voltage applied to the $VS3$ circuit u_{VS}, the corresponding current I_{VS} and the plasma current I_p are shown. The Kalman filter used in the VS system has been designed using the reduced order linear model corresponding to *Equilibrium #1*.

shape is recovered at steady state, as shown by the two plasma boundary snapshots shown in Figure 3.3.

Moreover, the oscillations induced in all control and controlled variables by the proposed ES-based approach are acceptable. Indeed, in many cases, these oscillations are within the expected noise range[1] (e.g., the relative I_p variation is about 0.15 %), while in the case of δZ_c the oscillations

[1] In the presented results the noise on the reconstructed plasma parameters has not been considered.

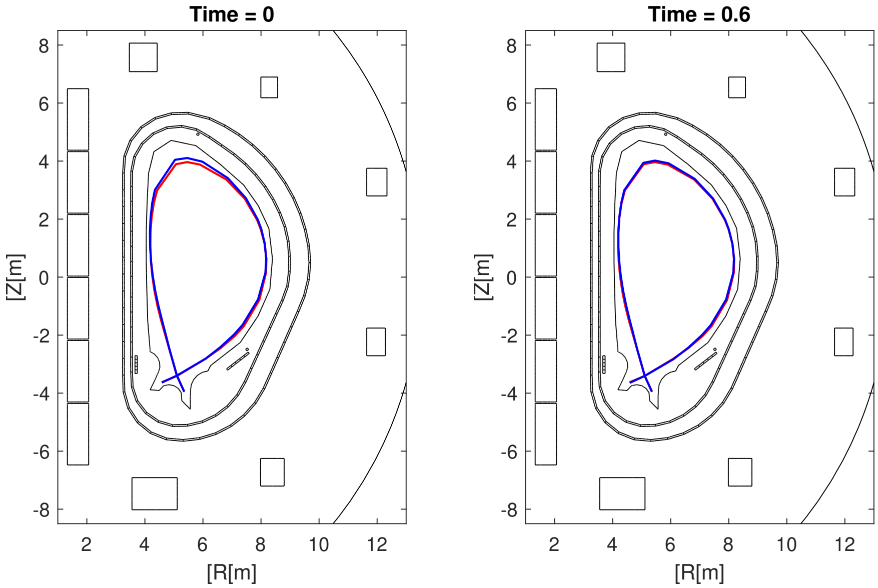


Figure 3.3: Plasma boundary shapes for the case of a VDE applied to pulse *Equilibrium #2*. The red curve shows the desired plasma boundary, while the black one is the simulated one. Two snapshots at $t_1 = 0\ s$ and $t_2 = 0.6\ s$ are shown.

are kept within the ± 1 cm range, which is compatible with ITER operation.

Minor disruption For the minor disruption scenario, the closed-loop response of *Equilibrium #1* is considered.

During a MD, a fraction of the plasma thermal energy is lost due to the uncontrolled growth of some plasma instability [60]. For magnetic control, a minor disruption can be modeled as a variation of the two disturbances $\delta\beta_p$ and δl_i in (1.1); Figure 3.6a reports the time traces of these variations for a MD that may occur during an ITER discharge. The closed-loop response for the *Equilibrium #1* is shown in Figure 3.4. Also in this case, the proposed VS system can guarantee stability while keeping the oscillations on δZ_c within ± 1 cm.

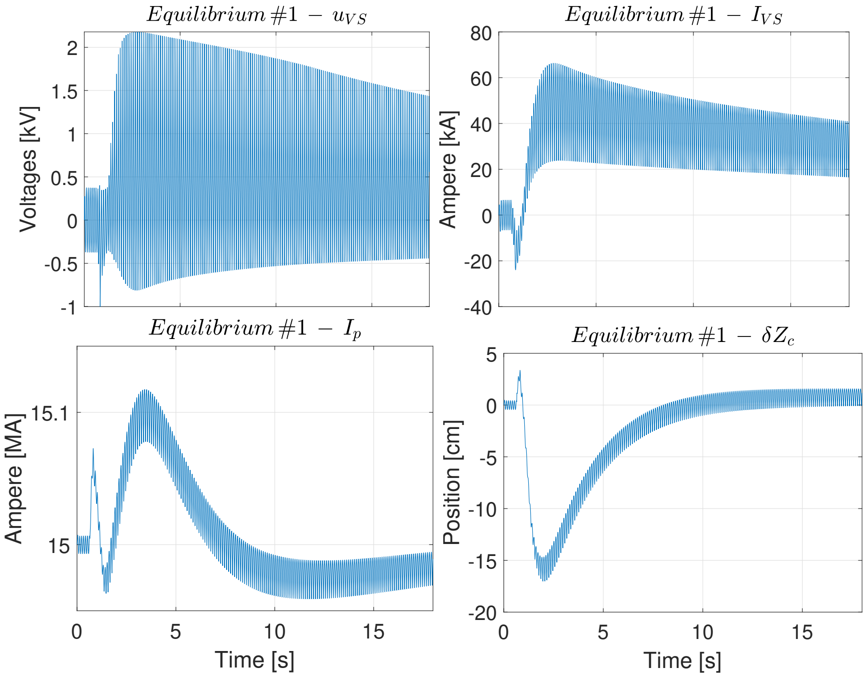


Figure 3.4: Response to the minor disruption modeled by the time traces of the disturbance variables δw shown in Figure 3.6a, when *Equilibrium #1* is considered. The displacement of the plasma current centroid vertical position δZ_c, the voltage applied to the $VS3$ circuit u_{VS}, the corresponding current I_{VS} and the plasma current I_p are shown. The Kalman filter employed in the VS system has been designed using the reduced-order system of the same linear model.

H to L transition At the end of an ITER discharge, when the additional heating systems are turned off, a transition from a high-confinement (H) to a low-confinement (L) regime occurs [61]. Similarly to the case of minor disruption, this transition represents a disturbance that can be modeled by the $\delta\beta_p$ and δl_i time traces shown in Figure 3.6b and therefore needs to be rejected by the plasma magnetic control.

The case of the H to L transition has been assessed on *Equilibrium #2*,

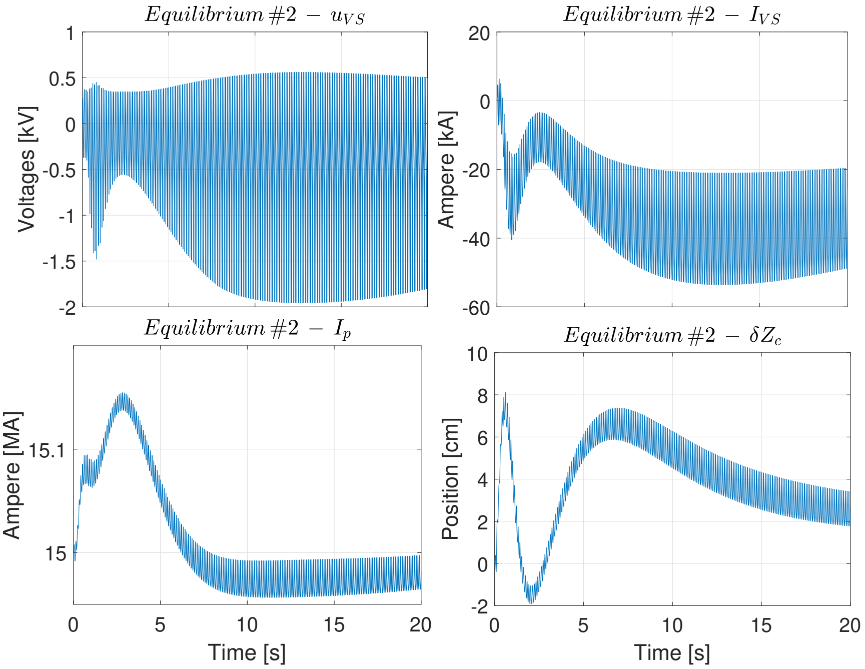
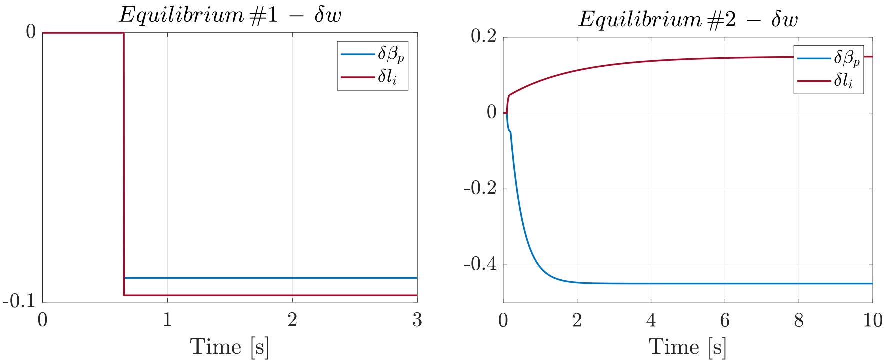


Figure 3.5: Response to a H to L transition modeled by the time traces of $\delta\beta_p$ and δl_i shown in Figure 3.6b, when *Equilibrium #2* is considered. The displacement of the plasma current centroid vertical position δZ_c, the voltage applied to the $VS3$ circuit u_{VS}, the corresponding current I_{VS} and the plasma current I_p are shown. The Kalman filter used in the VS system has been designed using the reduced order linear model corresponding to *Equilibrium #1.*

since it refers to the plasma state at the end of the discharge, before the plasma current ramp-down. The results are shown in Figure 3.5. Similar comments to those for the previous cases apply.

3.2 Switching power supply

For the averaging arguments leading to (2.3) to be valid, the switching frequency ω in (2.2) must be chosen “high enough”. This is a common



(a) $\delta\beta_p$ and δl_i time traces that model the minor disruption in Figure 3.4.

(b) $\delta\beta_p$ and δl_i time traces that model the H to L transition reported in Figure 3.5.

Figure 3.6: Time traces of the disturbances δw considered for simulating the MD and H to L transition.

requirement in averaging analyses, and the resulting system exhibits intrinsic time-scale separation. However, when a power amplifier is used, as in [54], the value of ω is limited due to the bandwidth of the power supply. To address this problem, in [55], a switching power supply was started to be considered for the $VS3$ circuit of ITER. Specifically, a power supply similar to the one used for the JET VS system, based on integrated gate commuted thyristors [62] is employed.

The availability of a faster actuator enables the choice of a higher switching frequency ω for the mixing and dithering terms in the ES control law (2.2), leading to an improvement in performance. A higher dithering frequency leads to a reduction in the amplitude of the induced oscillations in the system response with respect to what is presented in [54].

The characteristic of this type of power supply, which exhibits a multilevel hysteresis, is reported in Figure 3.7 while the setting parameters are the same that were used at JET with a maximum voltage of 12 kV and steps of 3 kV (see Table 3.2b).

The values of the ES parameters in the control law (2.2) used in this case are reported in Table 3.4.

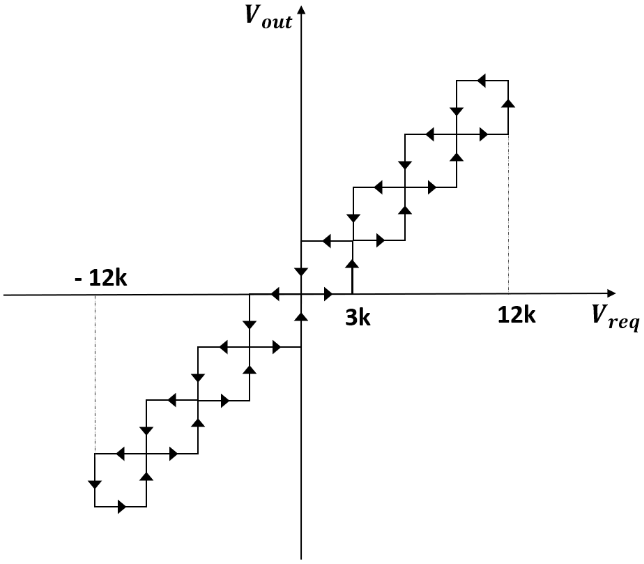


Figure 3.7: Characteristic of the fast switching power supply.

k	α	ω
$2.7 \cdot 10^{-3}$	1	$250 \cdot 2\pi$ rad/s

Table 3.4: Control parameters for the proposed VS system based on the ES control law (2.2) when a switching power supply is used as actuator for the VS circuit.

Results

The proposed ES-based VS is tested using linear and nonlinear simulations to demonstrate its validity and evaluate its robustness. Linear simulations show that the system can stabilize a wide range of plasma models, specifically all Group A equilibria in Table 3.1b, although the embedded Kalman filter is always the same. Nonlinear simulations, which involve solving a free boundary evolutionary problem, are used to demonstrate the robustness of the ES-based VS throughout the entire ITER discharge, starting with the Group B equilibria in Table 3.1a.

Linear Validation The purpose of the linear simulation was to demonstrate that the designed ES-based approach can stabilize a wide range of plasma models, even though the embedded Kalman filter remains the same. Indeed, the Group A family of plasma models consists of 24 different plasma equilibria, all at a plasma current of $15MA$, generated to cover the interval

$$(l_i, \beta_p) \ \in \ [0.8,\ 1.3] \times [0.1,\ 1] \ ,$$

with two different plasma shapes characterized by two slightly different elongations $\kappa = 1.81, 1.76$, respectively, as reported in Table 3.1a.

The unique Kalman filter adopted for linear simulations was obtained considering a reduced linearized model, of order 25, for the equilibrium characterized by $l_i = 1.3$, $\beta_p = 1$ and a growth rate $\gamma = 7.6\ s^{-1}$. The operational scenario considered for the linear simulations is a rejection of a VDE of 5 cm. The results of the simulations are shown in Figures 3.8, 3.9 and 3.10, where the displacement from the equilibrium of the plasma centroid position δZ_c, the current and voltage in the VS coils, $IVS3$ and $VS3$ respectively, and the behavior of the main gaps (chosen according to Figure 3.1) are reported for the family of models considered.

The results demonstrate that the proposed architecture can deal with relevant model uncertainties due to the model-agnosticism of the ES algorithm. A single Kalman filter is sufficient to stabilize all the plasma configurations studied, even if it is designed to estimate a simplified and reduced-order dynamic. This implies that the proposed VS system can guarantee a satisfactory degree of robustness and flexibility. Indeed, as can be seen in Figure 3.8, the vertical position variation δZ_c is rejected very rapidly in most of the cases considered. It should also be noted that in all cases considered, the maximum in-vessel current is on the order of a few kA.

Since plasma current and shape controllers were also included in the simulation scheme, it was possible to verify that the plasma current remains practically unchanged (variations are on the order of a few kA for a plasma current of 15 MA), while the plasma boundary does not touch the vessel during the transient in any of the scenarios considered. Figure 3.11 shows that the gaps are always positive, that is, the plasma boundary

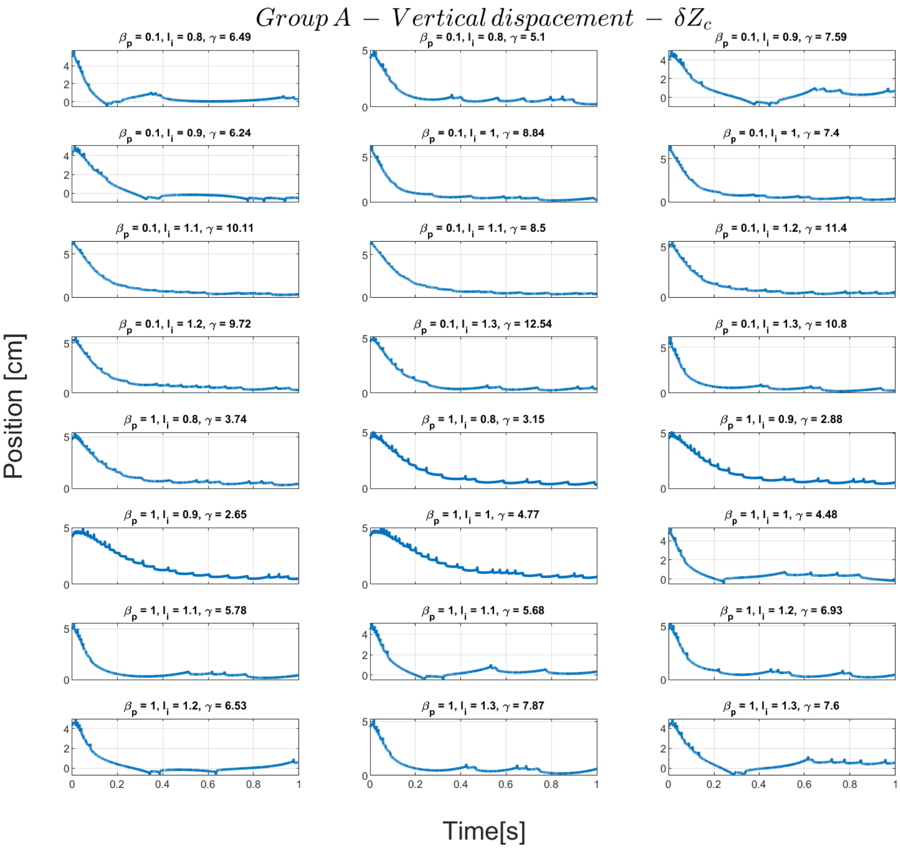


Figure 3.8: Response to a VDE of 5 cm for the family of different plasma models in Group A (Table 3.1a) in terms of the displacement from the equilibrium value of the plasma vertical position δZ_c.

never collides with the surrounding walls.

Nonlinear simulations In addition to the linear simulation discussed in the previous paragraph, nonlinear numerical simulations were conducted using the CREATE-NL+ free boundary evolutionary code. These simulations allow to validate the proposed VS control approach taking into account significant nonlinearities and a more realistic representation of

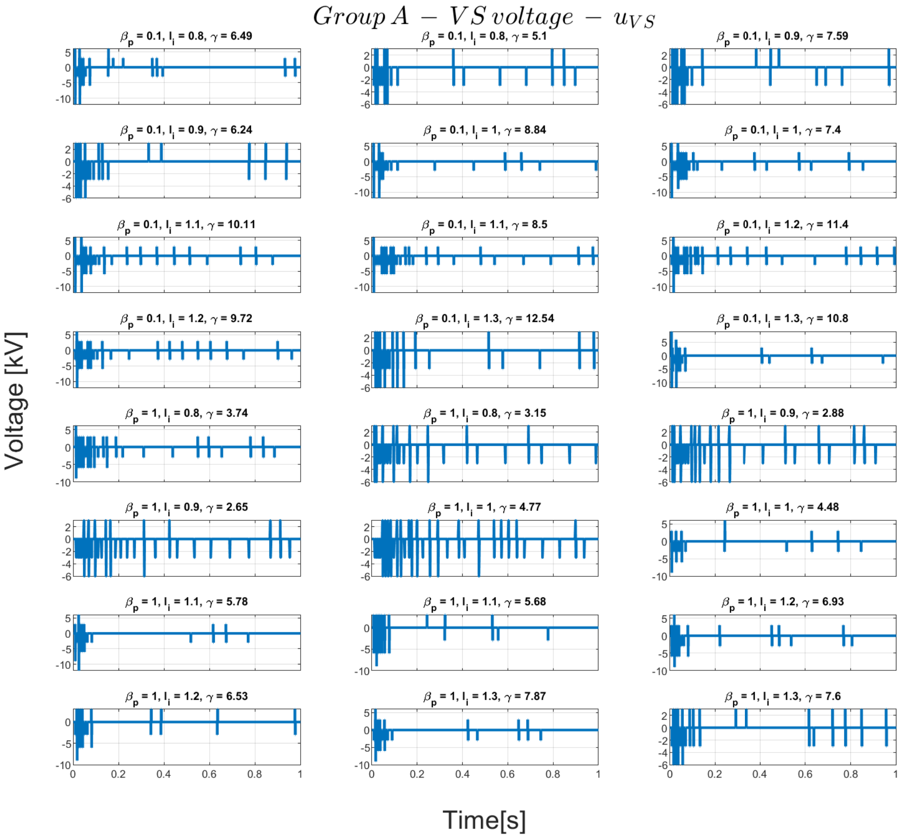


Figure 3.9: Response to a VDE of 5 cm for the considered family of different plasma model in Group A (Table 3.1a. The figure reports the voltages applied to the VS circuit, u_{VS}.

the ITER plasma. In the past, the CREATE-NL+ code has been validated against experimental data from several tokamaks, including JET [25]. For the simulations, the three equilibria from Group B (whose main parameters are summarized in Table 3.1b) have been chosen as starting equilibria.

The simulations have been performed considering a set of operational scenarios concerning:

- the rejection of a VDE of 5 cm;

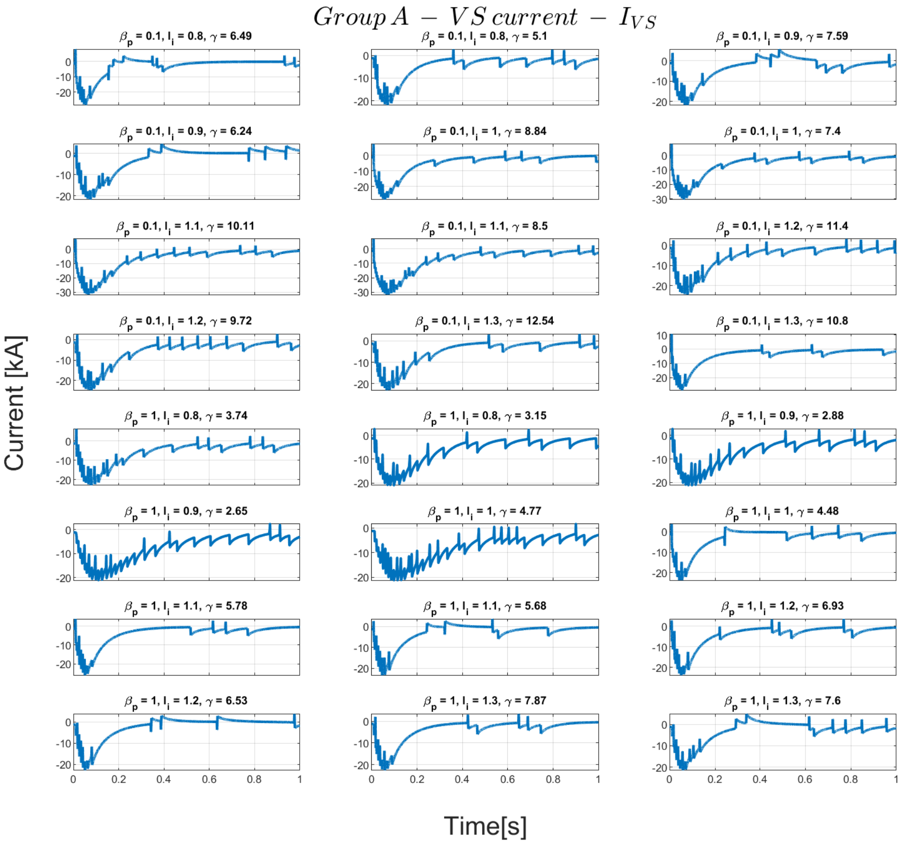


Figure 3.10: Response to a VDE of 5 cm for the Group A family of different plasma (Table 3.1a). The time behaviour of the currents in the VS system coils, I_{VS} have been reported.

- the response to a MD.

The MD has been modeled as a drop of $\Delta = 0.1$ for both disturbance parameters β_p and l_i. It should be noted that while the disturbance of a VDE has been applied to all three equilibria (see Figure 3.12, the MD has been considered only for *Equilibrium #1* and *Equilibrium #2* (see Figure 3.13 and 3.14, respectively). This is because MD is a phenomenon that can arise when the plasma energy content is high, which in the plasma

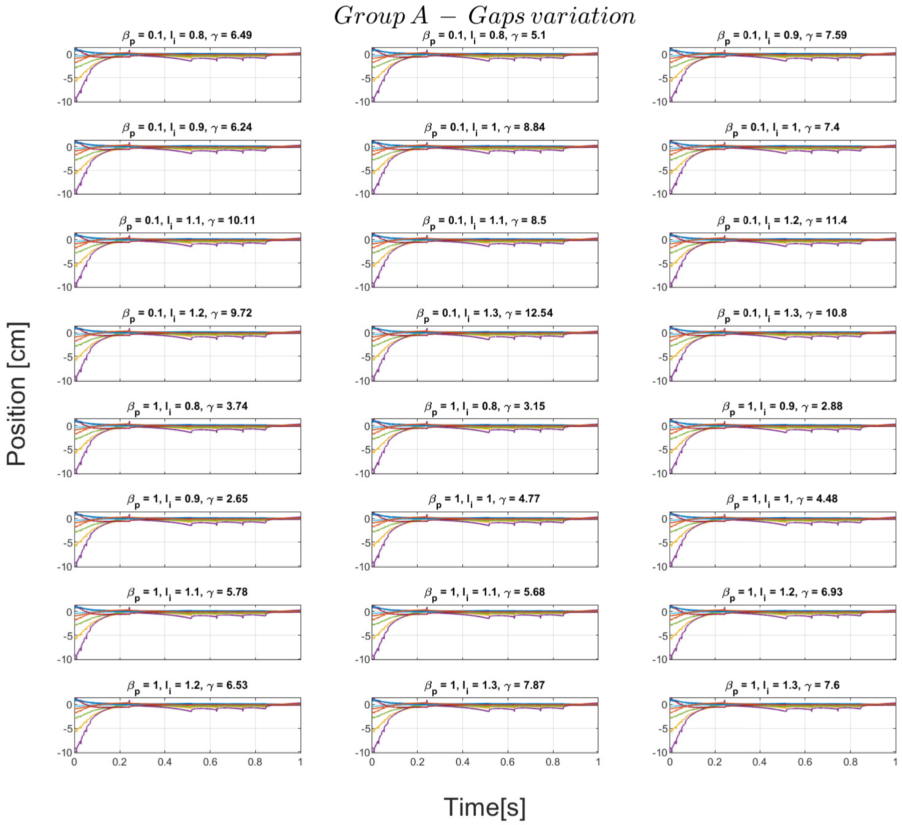


Figure 3.11: Response to a VDE of 5 cm for the Group A family of different plasma (Table 3.1a). The time behavior of the main controlled gaps highlighted in Figure 3.1 has been reported.

model is represented by the value β_p. Since *Equilibium #3* refers to the beginning of the flat-top phase, it is characterized by a value of β_p which is lower than $\Delta = 0.1$ considered int the MD simulation.

As for the linear case, the nonlinear simulations have been performed including all blocks of the plasma MCS shown in Figure 1.8, thus taking into account the interaction of the VS with the plasma current and shape controls. Furthermore, in all simulations the Kalman filter employed in the VS scheme (Figure 2.1) is always the one obtained from the reduced 25 order linearized model of *Equilibrium #1* (already considered in [54]) while the ES control parameters in the control law (2.2) are the same ones reported in Table 3.4.

The displacement of the plasma centroid from the equilibrium position δZ_c, the voltage u_{VS} and current I_{VS} in the VS circuit and the behavior of some of the controlled gaps are reported for the scenario considered. Specifically, the gaps considered are those highlighted in Figure 3.1. It can be seen that the controller is able to reject both the considered disturbances starting from the proposed equilibria. The worst case, in terms of δZ_c overshoot is that of *Equilibrium #2* in the case of the MD. Moreover, as expected, the highest current is reached again for *Equilibrium #2* in the case of MD. Lastly, the evolution of the gaps shows how the initial shape is restored after the appearance of the disturbance.

These simulations once again underline the robustness of the proposed model-free architecture.

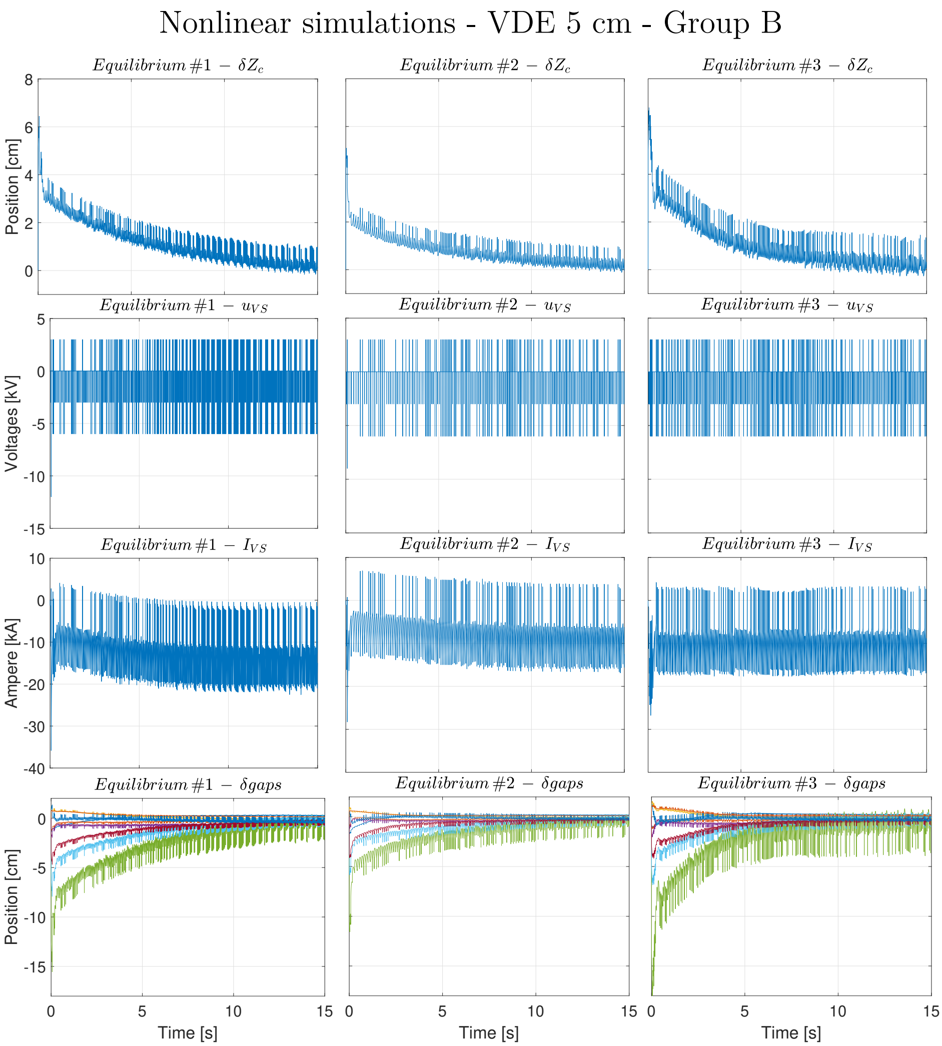


Figure 3.12: The nonlinear response to a VDE of $5\,cm$ in terms of the time traces of the vertical displacement of the plasma centroid δZ_c, of the voltages u_{VS3} andcurrent I_{VS3} in the in-vessel circuit and of some of the controlled gaps (highlighted in Figure 3.1) are shown for the equilibria in Group B.

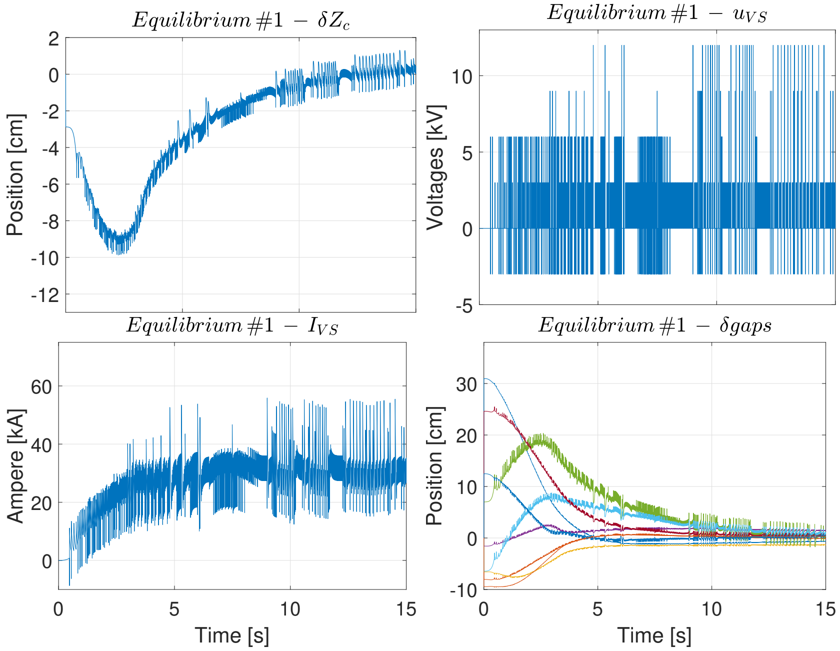


Figure 3.13: The nonlinear response to a MD in terms of the time traces of the vertical displacement of the plasma centroid δZ_c, of the voltages u_{VS3} andcurrent I_{VS3} in the in-vessel circuit and of some of the controlled gaps (those highlighted in Figure 3.1) is shown for *Equilibrium #1*.

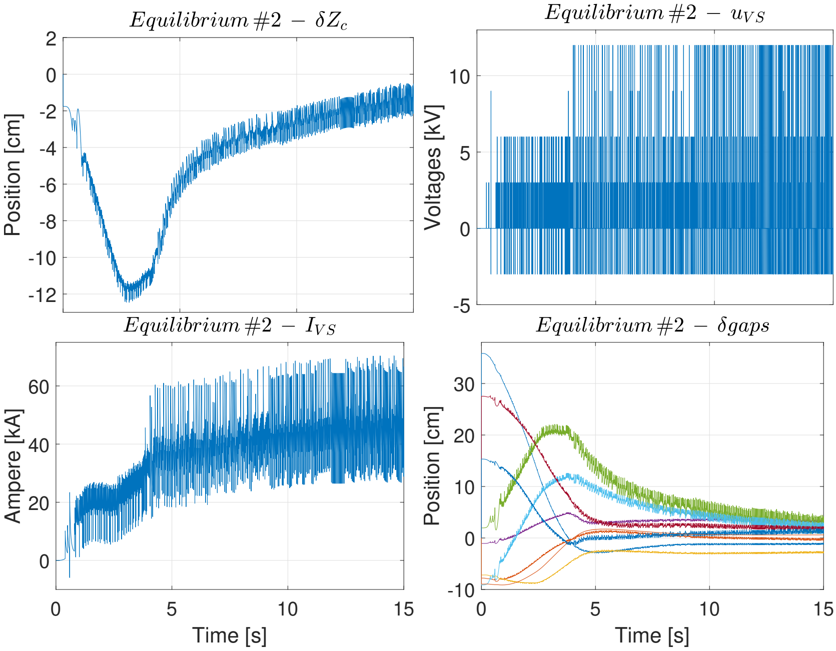


Figure 3.14: The nonlinear response to a MD in terms of the time traces of the vertical displacement of the plasma centroid δZ_c, of the voltages u_{VS3} andcurrent I_{VS3} in the in-vessel circuit and of some of the controlled gaps (those highlighted in Figure 3.1) is shown for *Equilibrium #2*.

3.3 Bounded control law

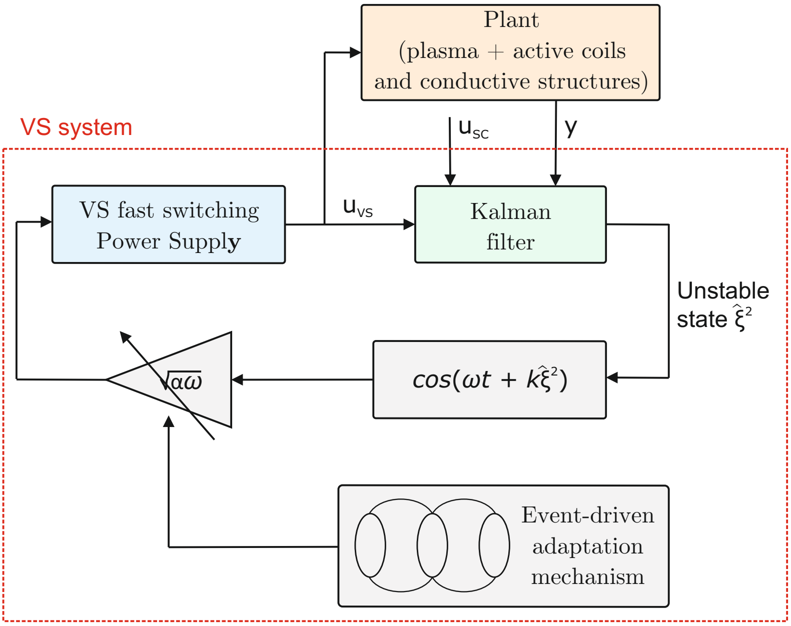


Figure 3.15: The block diagram of the VS system based on the ES bounded stabilization algorithm (2.4) extended with an event-driven adaptive mechanism and the switching power supply.

The bounded version of the ESS algorithm applied to the VS system of ITER was first proposed in [56]. In this case, the function to be minimized is introduced as an argument of a cosine or sine term, thus ensuring a bounded control effort. However, when the bounded control law (2.4) is combined with a power supply characterized by an intrinsic switching behavior (see Figure 3.7), different control gain thresholds are needed to activate the available voltage levels. Therefore, to take full advantage of both the available control input range and the bounded control action provided by the considered ES algorithm, an event-based control gain adaptation logic is introduced. A complete scheme of the bounded ES-based VS

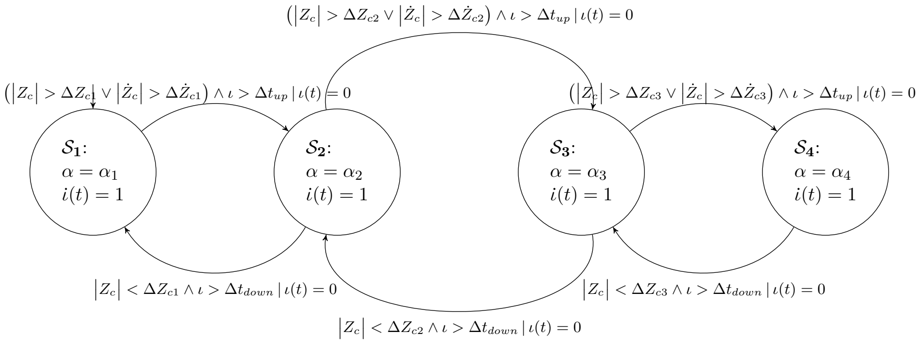

$(|Z_c| > \Delta Z_{c2} \vee |\dot{Z}_c| > \Delta \dot{Z}_{c2}) \wedge \iota > \Delta t_{up} \mid \iota(t) = 0$
$(|Z_c| > \Delta Z_{c1} \vee |\dot{Z}_c| > \Delta \dot{Z}_{c1}) \wedge \iota > \Delta t_{up} \mid \iota(t) = 0$
$(|Z_c| > \Delta Z_{c3} \vee |\dot{Z}_c| > \Delta \dot{Z}_{c3}) \wedge \iota > \Delta t_{up} \mid \iota(t) = 0$
$\mathcal{S}_1$: $\alpha = \alpha_1$ $i(t) = 1$
$\mathcal{S}_2$: $\alpha = \alpha_2$ $i(t) = 1$
$\mathcal{S}_3$: $\alpha = \alpha_3$ $i(t) = 1$
$\mathcal{S}_4$: $\alpha = \alpha_4$ $i(t) = 1$
$|Z_c| < \Delta Z_{c1} \wedge \iota > \Delta t_{down} \mid \iota(t) = 0$
$|Z_c| < \Delta Z_{c3} \wedge \iota > \Delta t_{down} \mid \iota(t) = 0$
$|Z_c| < \Delta Z_{c2} \wedge \iota > \Delta t_{down} \mid \iota(t) = 0$

Figure 3.16: Event-driven gain adaptive logic included in the control architecture of Figure 3.15.

controller explored in [56] is reported in Figure 3.15

	α_1	α_2	α_3	α_4
Maximum allowed voltage	$\pm 3kV$	$\pm 6kV$	$\pm 9kV$	$\pm 12kV$

Table 3.5: Correspondence between the values of the ES control parameter α and the level of voltages in the switching power supply characteristic activated by the control gain.

From (2.4) it can be seen that only the levels below $\sqrt{\alpha\omega}$ can be activated by the control law. Choosing $\sqrt{\alpha\omega}$ to activate only the lowest voltage level, i.e. $\pm 3\ kV$, the control action is not robust enough to counteract the relevant disturbances that can affect the plasma during typical tokamak operations. On the other hand, when the power supply is used at its maximum capability, i.e. $\pm 12\ kV$, the control effort is often unnecessarily high, which reflects in a higher request of current in the $VS3$ circuit.

Therefore, an event-driven mechanism is implemented in the control architecture to use all available voltage levels. The adaptation logic can be modeled with the hybrid automaton [63] shown in Figure 3.16. It adjusts the gain α of the control law (2.4) based on the real-time behavior of the system.

The proposed machine has four states, each of which is related to

	Δt_{up}	Δt_{down}
Values	$20\,ms$	$50\,ms$

(a)

	ΔZ_{c1}	ΔZ_{c2}	ΔZ_{c3}
Values	$0.02\,m$	$0.03\,m$	$0.04\,m$

(b)

	$\Delta \dot{Z}_{c1}$	$\Delta \dot{Z}_{c2}$	$\Delta \dot{Z}_{c3}$
Values	$0.2\,m/s$	$0.3\,m/s$	$0.4\,m/s$

(c)

Table 3.6: Values of the parameters introduced in the states and transitions of the event-driven state machine of the control gain adaptive logic shown in Figure 3.16.

one of the four distinct (and specular) voltage levels of the power supply characteristic (Figure 3.7). Each state is associated with a value α_i, as indicated in Table 3.5.

The transitions between different states are subordinate to variations of the vertical position Z_c and velocity $\dot{Z}_c$ of the plasma centroid with respect to the nominal values. The farther these variables move from the equilibrium condition, the more effort is needed to control as the machine explores states where the value of α is incremented. On the opposite, as Z_c approaches zero the control gain is reduced. The values of ΔZ_{ci} and $\Delta \dot{Z}_{ci}$ used in the event-driven logic shown in Figure 3.16 are reported in Tables 3.6b and 3.6c, respectively.

Moreover, the transitions are time-inhibited, in the sense that the system must remain in a state for a certain period for the control action to affect the behavior of the plasma before the state machine is allowed to transit to another state and hence to a different value of α. The values of Δt are listed in Table 3.6a. In particular, the value of Δt_{up} used for the transitions to states that correspond to a higher control gain is shorter than Δt_{down} considered in the transition to lower values of α. This is because the controller should act faster and with a higher effort when the plasma is moving away from the desired position.

Results

In this section, the results obtained with the VS architecture proposed in Figure 3.15 are presented and discussed. The simulations have been carried out considering meaningful plasma operational scenarios, i.e.:

- the rejection of a VDE of 5 cm;
- the response to a MD.

Two different equilibria, corresponding to different time instants of a 15 MA ITER discharge, were taken into account. Specifically, *Equilibrium #1* and *Equilibrium #2* from Group B. The nominal values of the plasma parameters for the two equilibria are reported in Table 3.1b.

For all scenarios considered, the same configuration of the VS system has been used. Specifically, the parameters of the control law (2.4) have been chosen as in [55] equal to $k = 2.7 \cdot 10^{-3}$ and $\omega = 250 \cdot 2\pi$ (see Table 3.4) while the values α_i is chosen according to Table 3.6. The same Kalman filter, used also in the previously proposed applications, has been adopted; it is designed by referring to a reduced model of order 25 for *Equilibrium #1*.

VDE rejection A VDE rejection for *Equilibrium #2* is considered. Figure 3.17 reports a comparison between different configurations of the ES algorithm: the former VS system presented in Section 3.2 based on the unbounded ES control law (2.2) (yellow line), the bounded ES architecture proposed in this Section (red line) and the same bounded control law where the parameter α has been fixed at the highest value to allow all available voltage levels. The vertical position of the plasma centroid Z_c, the u_{VS} voltage, the I_{VS} current, and the plasma current I_p are shown. These results show that even if all three algorithms can stabilize the plasma column, the proposed bounded ES is more effective than the original ES control law in stabilizing the plasma column. The bounded ES recovers in a shorter time the original Z_c position while requiring a lower current in the VS circuit. Furthermore, the oscillations induced on both the centroid vertical position Z_c and on the current in the actuator I_{VS} are reduced with respect to the case where all voltage levels are enabled.

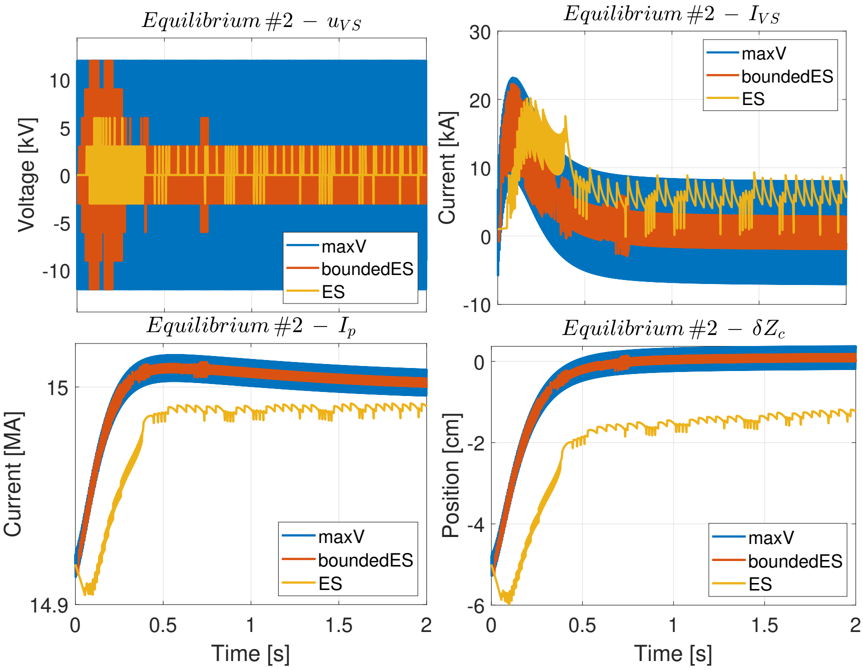


Figure 3.17: The Response to a 5 *cm* VDE applied to the *Equilibrium #2* plasma. The traces refer to 3 different configurations of the ES algorithm: the bounded ES algorithm with the event-driven adaptation logic introduced in this section (red line), the unbounded ES algorithm presented in Section 3.2 (yellow line), and the bounded ES version with all available power supply levels enabled at all times (blue line). The u_{VS} voltage, the I_{VS} current, the plasma current I_p, and the vertical position of the plasma current centroid Z_c are shown. The Kalman filter used in the VS system has been designed considering the reduced order linear model corresponding to *Equilibrium #1*.

Minor disruption The MD is modeled as an instantaneous drop of the plasma internal profile parameters, which in this case corresponds to a drop of 0.11 for β_p and of 0.12 for l_i with respect to their nominal value. The comparison between the bounded (2.4) and the unbounded (2.2) versions of the ES control law is reported in Figure 3.18. In particular, the red traces refer to the bounded ES-based VS, while the yellow traces refer to the solution with the unbounded control law discussed in Section 3.2. Comments similar to those made for the previous case apply. It results that the bounded ES algorithm by requesting a higher voltage in the first phase, achieves a better transient response by minimizing the effect of the disturbance on both Z_c and I_p while requesting a slightly lower I_{VS} current.

3.4 Gain adaptation logic

The switching power supply used in [55, 56] exhibits a characteristic with multilevel hysteresis with a maximum voltage of $12\,kV$ and steps of $3\,kV$. However, these voltage levels appear to be quite high compared to the maximum voltage that can be supplied by the linear amplifier envisaged for the ITER VS system. On the other hand, coupling a switching power supply with a control algorithm such as the ES brings clear advantages compared to the use of a linear amplifier (see results in [54]). The faster time response of the former enables the choice of a higher switching frequency ω in the ES control law (2.4) reducing the amplitude of oscillations in the system response and improving overall performance.

For this application, the switching power supply was retained along with a modified version of the control gain adaptation logic that allows lower voltage levels to be employed. In particular, steps of 250 V with a maximum allowed voltage of 3 kV are considered. This allowed a four-fold decrease in voltage level in the VS circuit with respect to the use of the hybrid automaton previously considered (Figure 3.16).

In this case, the event-driven adaption logic is modeled with the automaton shown in Figure 3.19. This automaton has five states highlighted in green that are associated with five different voltage levels allowed by the corresponding values of α_i, which are reported in Table 3.7a. The novelty in comparison to the one implemented in [56] (see Figure 3.16) is that the

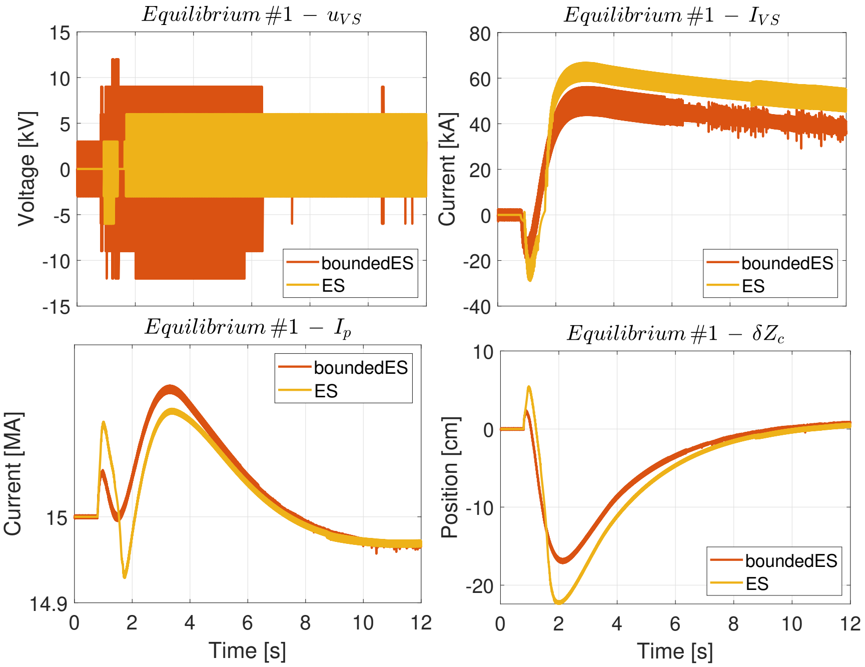


Figure 3.18: Response to an MD when *Equilibrium #1* is considered. The behavior with the bounded ES algorithm with the event-driven gain adaptation logic presented in this section (red line) and with the unbounded ES VS presented in Section 3.2 (yellow line) are shown. The u_{VS} voltage, the I_{VS} current, the plasma current I_p, and the vertical position of the plasma current centroid Z_c are shown.

transitions between these states are subordinated only to the variation of the vertical position of the plasma centroid. The further away from equilibrium Z_c moves, the more effort the controller must expend, as the machine investigates states where the value of α_i is increased. On the contrary, as the variation of Z_c approaches zero, the control gain diminishes. These transitions are no longer time-inhibited to allow the controller to act faster, enhancing, or reducing, the control effort as soon as the plasma is moving away from, or getting closer to, the controlled position.

Another improvement is the addition of the states highlighted in or-

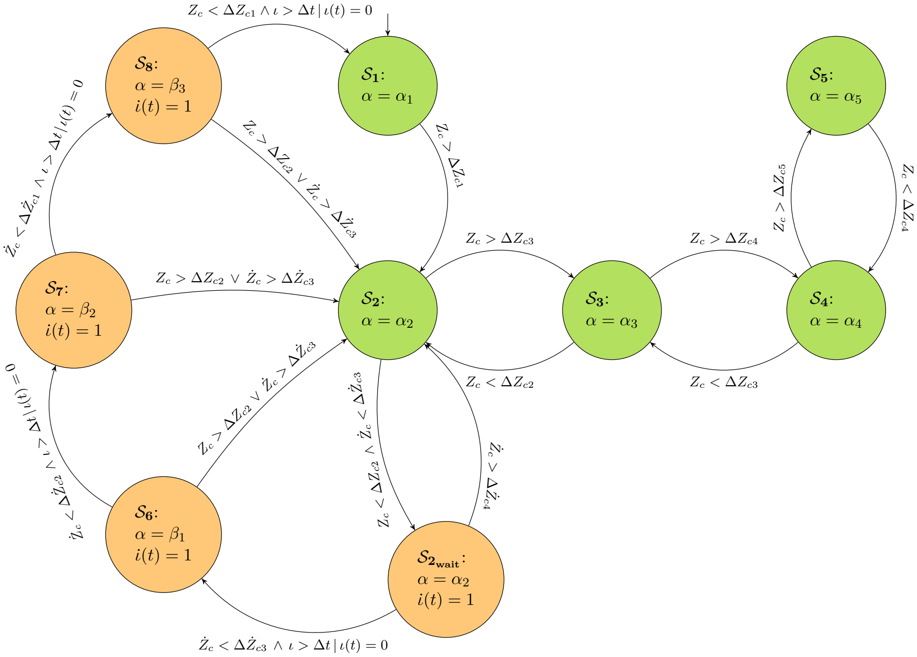

$Z_c < \Delta Z_{c1} \wedge \iota > \Delta t \,|\, \iota(t) = 0$
$\mathcal{S}_8$: $\alpha = \beta_3$ $i(t) = 1$
$\mathcal{S}_1$: $\alpha = \alpha_1$
$\mathcal{S}_5$: $\alpha = \alpha_5$
$\dot{Z}_c < \Delta \dot{Z}_{c1} \wedge \iota > \Delta t \,|\, \iota(t) = 0$
$Z_c > \Delta Z_{c2} \vee \dot{Z}_c > \Delta \dot{Z}_{c3}$
$Z_c > \Delta Z_{c1}$
$Z_c > \Delta Z_{c5}$
$Z_c < \Delta Z_{c4}$
$Z_c > \Delta Z_{c3}$
$Z_c > \Delta Z_{c4}$
$\mathcal{S}_7$: $\alpha = \beta_2$ $i(t) = 1$
$Z_c > \Delta Z_{c2} \vee \dot{Z}_c > \Delta \dot{Z}_{c3}$
$\mathcal{S}_2$: $\alpha = \alpha_2$
$\mathcal{S}_3$: $\alpha = \alpha_3$
$\mathcal{S}_4$: $\alpha = \alpha_4$
$Z_c < \Delta Z_{c2}$
$Z_c < \Delta Z_{c3}$
$Z_c > \Delta Z_{c2} \vee \dot{Z}_c > \Delta \dot{Z}_{c3}$
$\dot{Z}_c < \Delta \dot{Z}_{c2} \wedge \iota > \Delta t \,|\, \iota(t) = 0$
$Z_c < \Delta Z_{c2} \wedge \dot{Z}_c < \Delta \dot{Z}_{c3}$
$\dot{Z}_c > \Delta \dot{Z}_{c4}$
$\mathcal{S}_6$: $\alpha = \beta_1$ $i(t) = 1$
$\mathcal{S}_{2_{\text{wait}}}$: $\alpha = \alpha_2$ $i(t) = 1$
$\dot{Z}_c < \Delta \dot{Z}_{c3} \wedge \iota > \Delta t \,|\, \iota(t) = 0$

Figure 3.19: The event-driven gain adaptive logic included in the VS control architecture reported in Figure 4.1.

ange in Figure 3.19. These states are used to control the transition between the states S_2 and S_1, thus to smooth the switching from a control voltage of $2.25\,kV$ to the lowest level of $500\,V$. These *orange-states* are associated with the voltage levels allowed by the corresponding values of the control gain β_i (see Table 3.7b). The transitions between these states are also subordinate to the vertical velocity of the plasma centroid. When the variations of Z_c or $\dot{Z}_c$ increase the machine moves back to S_2 asking for a higher control effort. On the other hand, after an idle step in the $S_{2_{wait}}$ state, where the machine waits for both position and velocity to decrease, the control gain values can be further reduced as the transition between the *orange-states* progresses until the initial state S_1 when the equilibrium is recovered. These latter transitions are time-inhibited. The state machine is allowed to transit to another state only after an interval of time $\Delta t = 150\ ms$. This ensures that the control action affects the

	α_1	α_2	α_3	α_4	α_5
Maximum voltage	$\pm 500\ V$	$\pm 2.25\ kV$	$\pm 2.5\ kV$	$\pm 2.75\ kV$	$\pm 3\ kV$

(a)

	β_1	β_2	β_3
Maximum voltage	$\pm 1.75\ kV$	$\pm 1.5\ kV$	$\pm 1.25\ kV$

(b)

	ΔZ_{c1}	ΔZ_{c2}	ΔZ_{c3}	ΔZ_{c4}	ΔZ_{c5}
Values	$0.01\,m$	$0.02\,m$	$0.03\,m$	$0.04\,m$	$0.05\,m$

(c)

	$\Delta \dot{Z}_{c1}$	$\Delta \dot{Z}_{c2}$	$\Delta \dot{Z}_{c3}$
Values	$0.1\,m/s$	$0.5\,m/s$	$0.6\,m/s$

(d)

Table 3.7: The parameters of the event-driven state machine designed for the adaptation logic of the control gains in Figure 3.19 are reported. (a) and (b) report the correspondence between the values of the ES control parameter α in the different states and the voltage levels that can be activated. (c) and (d) report the values of the variation of Zc and $\dot{Z}_c$ used fpr the state transitions.

plasma behavior before decreasing the value of α.

Results

In this section, the result obtained considering the automaton in Figure 3.19 for the gain adaptation of the bounded ES control law (2.4) in the VS scheme in Figure 3.15 are presented and discussed. In particular, this approach is compared to the one obtained considering the previous version of the adaptation logic presented in Section 3.3. The main benefit obtained with the new automaton is the reduction of the allowed voltage in the VS circuit using steps of $250\,V$ (instead of $3kV$) for a maximum value of $3\,kV$ (instead of $12\,kV$).

For this purpose, the rejection of a VDE of $5\,cm$ for *Equilibrium #2* and the response to a MD while considering *Equilibrium #1* have been carried out. Recall that these equilibria correspond to different time instants of a 15 MA ITER discharge and Table 3.1b reports the nominal values of their plasma parameters.

Moreover, the same ES control parameters and the same Kalman filter as in Section 3.3 have been considered. Specifically, the parameters of the control law (2.4) have been chosen again equal to $k = 2.7 \cdot 10^{-3}$ and $\omega = 250 \cdot 2\pi$ (see Table 3.4 while the values α_i chosen according to Table 3.7a-3.7b and the same Kalman filter, designed by referring to a reduced model of order 25 for *Equilibrium #1*, has been adopted.

VDE rejection A comparison between the VS scheme with the adaptation logic performed by the automaton in Figure 3.16 and Figure 3.16 is reported in Figure 3.20. The comparison considers the rejection of a $5\,cm$ VDE for *Equilibrium #2.*The u_{VS} voltage, the I_{VS} current, the vertical position of the plasma centroid Z_c, and the plasma current I_p are shown. The time traces show that it is possible to reject the disturbance and recover the equilibrium position by requesting lower voltage and current in the VS circuit. Moreover, this allows to reduce even more the amplitude of the oscillation introduced by the ES approach.

Minor Disruption The case of a MD is analyzed considering *Equilibrium #1*. The comparison between the performance of the ES-based VS scheme when considering the even-driven adaptation logic in Figure 3.16 and Figure 3.16 is shown in Figure 3.21. The u_{VS} voltage, the I_{VS} current, the vertical position of the plasma centroid Z_c, and the plasma current I_p are reported. The same consideration applies as in the VDE rejection case. It is possible to counteract the same disturbance while requesting a lower effort to the VS in terms of both values of u_{VS} and I_{VS}.

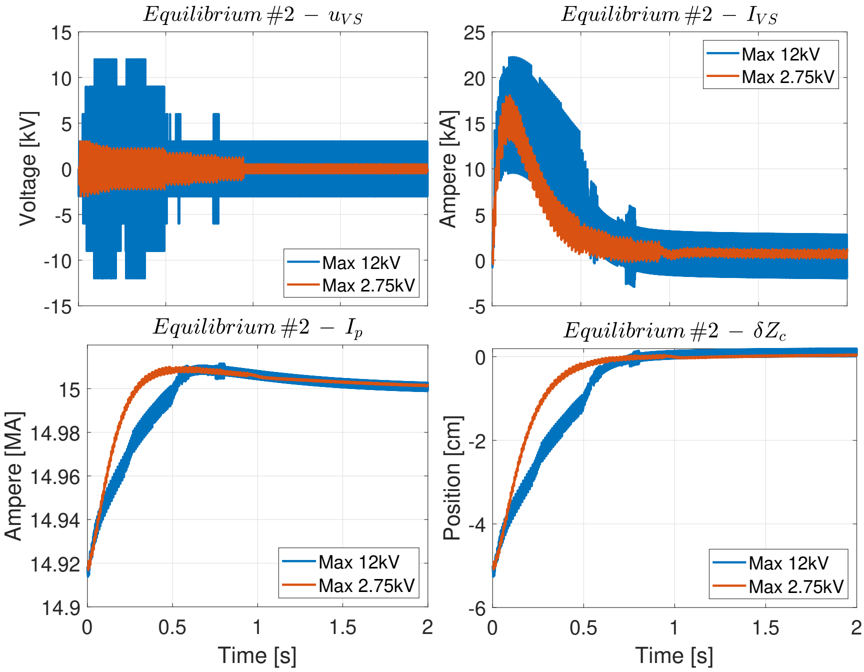


Figure 3.20: The Response to a 5 *cm* VDE applied to the *Equilibrium #2* plasma. The u_{VS} voltage, the I_{VS} current, the plasma current I_p, and the vertical position of the plasma current centroid Z_c are shown.

Summary

In this chapter, the plasma VS problem in the ITER tokamak has been addressed by means of the ESS algorithm. The proposed technique is almost model-free in the sense that the only knowledge of the plant dynamics lies in a single, reduced-order Kalman filter. Moreover, the simulations have been performed taking into account the whole ITER plasma MCS to verify that the interaction between the proposed VS system and the plasma current and shape controllers does not result in a degradation of the overall plasma magnetic control performance.

By a posteriori assessment it is shown that the proposed VS scheme can practically stabilize the plasma column by bringing to zero the *mo-*

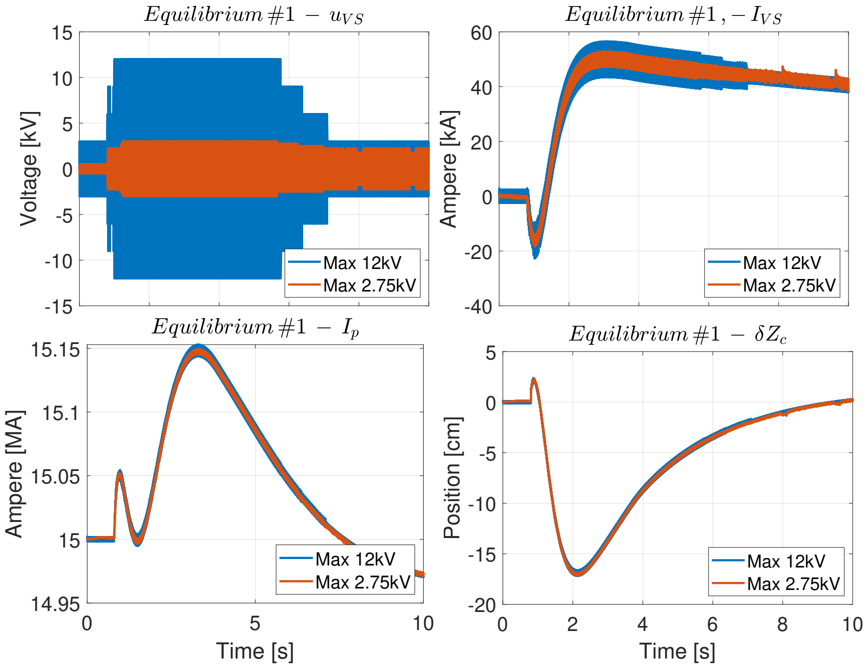


Figure 3.21: Response to an MD when *Equilibrium #1* is considered. Drop of 0.11 for β_p and of 0.12 for l_i with respect to their nominal value The u_{VS} voltage, the I_{VS} current, the plasma current I_p, and the vertical position of the plasma current centroid Z_c are shown.

tion of the plasma along the unstable mode while counteracting relevant plasma disturbances. Moreover, thanks to the *model-agnostic* nature of the ES algorithm the proposed approach allow also to cope with relevant model uncertainties. Indeed, the simulations presented, even if always employing the same Kalman filter, have been performed considering different plasma equilibria to cover a variety of plasma parameters and configurations. These results have also been confirmed by nonlinear simulations performed on a whole ITER discharge and using equilibrium codes [25].

Moreover, the bounded ES control algorithm is incorporated with an event-driven control gain adaptation mechanism to fully exploit the characteristics of the considered switching power supply. The bounded solution is shown to be more effective than the unbounded ES control algorithm when dealing both with VDEs and MDs. Finally, by a proper design of the hybrid automaton performing the gain adaptation, it is possible to reduce the required voltage in the VS circuit bringing it closer to the value actually envisaged for the ITER VS system.

It is worth stressing once again that the main advantage of the proposed technique resides in the fact that it can be adapted rather easily to different plasma configurations. In fact, this usually requires low or no effort, provided that the considered observer is capable of describing, at least roughly, the unstable dynamic of the plant and that suitable controller gains are chosen. This is not the case for standard VS techniques, which usually need to be tuned based on the specific plasma configuration, a task that usually requires some significant modeling and testing effort. This opened an interesting perspective for the development of model-free VS stabilization techniques. Along this line of research, a fully model-free VS system in which the residual model dependence embedded in the Kalman filter is removed is presented in the next Chapter 4.

4

Neural Network-based Extremum Seeking

THIS CHAPTER starts bringing forward the role that neural networks are playing in the nuclear fusion community. Subsequently, an entirely model-free ES-based solution for VS system of the ITER tokamak is presented. The ES-based VS system proposed in Chapter 3 has been enhanced including NNs in the control scheme.

The ES based approaches presented in [54, 55, 56], even if showing promising results, can only be considered *quasi* model-independent since the need for a Kalman filter to compute the required Lyapunov function. In this chapter, a new solution is presented in which this issue is tackled and resolved. Indeed, NNs have been studied to replace the Kalman filter, as shown in the block diagram, with the novel architecture reported in Figure 4.1. Specifically, the possibility of using either a LR, a MLP, an ELM, or a LSTM networks i considered for this application.

As will be shown by simulations of relevant ITER test cases, such a control approach allows to enlarge the operative space of the overall VS system, making it possible to stabilize plasma equilibria that are not sta-

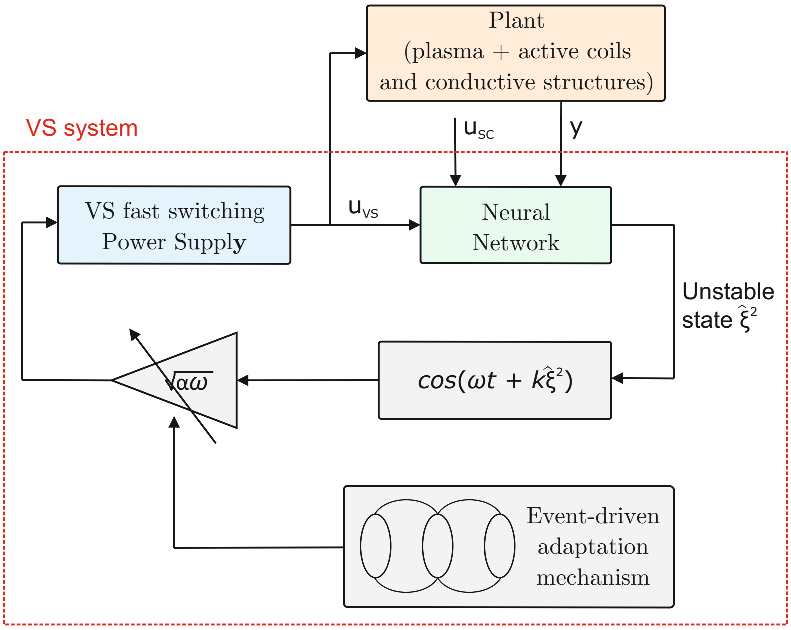


Figure 4.1: The block diagram of the proposed VS system is based on the ES bounded stabilization algorithm, the event-driven adaptive mechanism, and a NN that reconstructs the movement along the unstable mode.

bilized by the setup based on a single Kalman filter. Indeed, a single Kalman filter could not achieve the accuracy for the estimation of the plasma unstable mode that is needed to stabilize a wide range of plasma configurations. In this chapter, it will be shown how such limitation is overcome by replacing the Kalman filter with a NN and thus removing the residual model-dependence from the VS scheme.

4.1 Neural networks for nuclear fusion

The scientific community for nuclear fusion is raising awareness about the potential of artificial NNs to analyze large amounts of data collected from experiments and simulations. Data-driven techniques have been studied and applied in various aspects of fusion research. In particular, NNs have been extensively studied to predict plasma instabilities and disruption, as well as to monitor the evolution of plasma profiles. Different data-driven algorithms and network architectures, such as unsupervised learning algorithms [64], k-nearest neighbor technique [65], Binary Classification [66], MLP [67, 68], Recurrent Neural Networks (RNN) [69], Support Vector Machines (SVM) [64, 70], Self-Organizing Maps [67], Generative Topographic Mapping (GTM) [69, 71], Classification and Regression Trees [72] and Random Forests (RF) [73] algorithms have been used to predict and classifying disruption in various operating tokamaks. Moreover, beyond being trained considering real-time data collected from machine experiments, several of these approaches have also been tested in real time. As an example [74] discuss the result of an RF disruption predictor used to run than DIII-D 900 discharges (more than 4 months of operation) as part of the Plasma Control System (PCS). Moreover, since for ITER the development of a disruption training database may be infeasible, in [75] a RF learning method was trained on a large database that included data from different machines (Alcator C-Mod, DIII-D, and EAST) for cross-machine validation. Indeed, since on a machine of the size such as ITER, disruptions should be avoided by any means necessary, different from what is done on relatively small existing experimental machines, the authors of [75] proved the generalization property of NNs. While in [64] was shown how an SVM predictor can outperform the JET Protection System (JPS) of about 1800 ms before the occurrence of a disruption. Recently, deep Convolutional Neural Network (CNN)s have become attractive also in disruption prediction, as they allow imaging to be included directly in the training database, such as electron cyclotron emission imaging diagnostic [76] and spatio-temporal information obtained from plasma profile diagnostics [77]. Furthermore, [78] reports a performance comparison between CNN disruption prediction system at JET [79] with those of a classical MLP and a more sophisticated GTM model [71].

Moreover, starting from a first isolated attempt in the '90s [80], a big effort has recently been made to use MLP for plasma equilibrium reconstruction from magnetic measurements [81, 82, 83]. Deep learning algorithms have been implemented for Magnetohydrodynamic (MHD) model identification [84] and tearing modes and disruption avoidance [85]. NNs have been trained also to classify Alfvén eigenmodes [86, 87] and predict the evolution of plasma profiles [88], performance, and tearability [89].

It is worth noting that the solutions mentioned above are not a direct matter of feedback control. Most of them are operated offline for the reconstruction or estimation of plasma quantities from experimental data. When they are used online, and as in the case of some disruption predictors included in the tokamak PCS, they are not incorporated into any feedback control loop. The main function is exception handling. The predictions obtained can be used to trigger specific events, such as the occurrence of disruptions or instabilities, which can then be handled by a safety system.

4.2 Estimation of the plasma unstable mode using neural networks

In this section, the training of NNs to estimate the movement of the plasma along the unstable mode ξ for replacing the Kalman filter in the ES-based VS system is presented. The section starts by explaining how the data used to train and validate the NNs have been generated, i.e., the variables considered as inputs, how they have been sampled, and the plasma linear model used to generate them, are detailed in what follows. Subsequently, the NN models considered are introduced with their architectures and training strategies. Indeed, for each network, different trainings were performed to evaluate the generalization property of the different networks and the robustness of the VS control system when the data-driven NNs are used in the closed loop instead of the Kalman filter.

The ES stabilization algorithm, albeit being model-independent, requires some access to the system's state to compute $V(x)$ in (2.4). In [55], ξ is obtained by a unique Kalman filter, which requires the access to a plasma linear model (1.1) to be computed. In particular, in [55] a single Kalman filter was employed to stabilize a family of 24 ITER plasma linearized models (whose equilibrium parameters are reported in Table 3.1a, Group A), while another filter was used to stabilize other three ITER equilibria (reported in Table 3.1b, Group B) and was proven to work with linear and non-linear simulations. However, none of the considered

Kalman filters could be used to stabilize plasmas from both the considered groups of equilibria.

The residual model dependence embedded in the Kalman filter limits the generalizability of the approach. The main objective of this part of the thesis work is to turn the ES-based VS system completely model-free. Namely, a NN has been incorporated into the architecture proposed in the previous chapter. A LR, a MLP, an ELM [90, 91], and a LSTM networks were explored as a substitute for the Kalman filter. The above-mentioned NNs have been trained to estimate, starting from magnetic measurements, the dynamic of the plant along the unstable mode, which is then used to compute the Lyapunov function to be minimized by the ES control algorithm.

In what follows, it will be shown how the use of a NN can help overcome the above-mentioned limitation. Indeed, data-driven NNs can enhance the generalization property of the ES control law, ensuring more robustness with respect to model uncertainties or changes in the plasma configuration and behavior. In particular, the different NNs will be trained to estimate the plasma unstable mode ξ considering only a subgroup of linearized models in Group A and then tested in closed-loop also on the equilibria of Group B.

4.2.1 Synthetic data-set for NNs training and validation

Since ITER is not yet in operation, in this work models have been used to generate training data by running synthetic plasma simulation, referred to as plasma *shots*. More in detail, data were generated by performing linear simulations using the models in Group A. The ES-based VS architecture reported in Figure 3.15 (embedded with the event-driven gain adaptation logic of Figure 3.19) has been used to generate the synthetic data. Although different plasma linear models have been considered, all shots have been performed using the same Kalman filter, that is, the one considered in [55] for the equilibria in Group A. In particular, such Kalman filter is obtained by exploiting the linear model characterized by $\gamma = 7.6\, s^{-1}$, $\kappa = 1.8$, $\beta_p = 1$ and $l_i = 1.3$.

It is worth to remark that, once the NNs have been trained and used to replace the Kalman filter, the dependence on the specific plasma model is also removed from the controller. However, for closed-loop testing of

NNs inputs

Signals name	Units
PF circuit currents δI_{PF}	kA
Plasma current δI_p	MA
Plasma vertical position δZ_c	cm
PF circuit voltages u_{PF}	kV
Gaps position	cm
X-point position	cm

Table 4.1: Variables used as NNs inputs for the estimation of movement of plasma state along the unstable mode $\hat{\xi}$. The data were collected performing synthetic ITER pulses and sampled at a frequency of 20 kHz, and each variable was normalized by subtracting out the median value and dividing by the interquartile range.

the NNs a model of the plasma is still required, driven by the need to mimic the plant behavior. Consequently, different plasma models have been considered to perform closed-loop validation, contributing to assess the generalization property of the proposed architecture.

Futhermore all simulations, both for generating the training database and for validating the NNs have been carried out considering not only the VS system but the whole MCS as reported in Figure 1.8, i.e. the plasma shape and current controllers have also been included in the simulation scheme. This allowed to also take into account the interaction of the proposed ES-based VS with the other control loops.

The operational scenario considered to generate the shots is the rejection of a VDE, whose amplitude was randomly chosen between $[-5\,,\,5]$ cm.

The variables considered as inputs for the NNs are reported in Table 4.1. They include the variation with respect to the nominal values of currents in the PF coils δI_{PF}, of the plasma currents δI_p, of the plasma centroid vertical position δZ_c, which were also used as input for the Kalman filer. Moreover, the variations of plasma shape descriptors such as the X-point position δXP and of the controlled *gaps*[1] have also been considered as inputs for the NNs. Note that, as far as magnetic control is concerned, the considered input variables can be used as a suitable representation of the plasma state at each time step.

These variables need to be sampled at a frequency $f_s = 20\,kHz$ to achieve the required accuracy for the estimation of the plasma unstable mode ξ using NNs. However, it should be noted that once included in the VS feedback loop, the NNs do not need to run at 20 kHz.

Finally, the data were preprocessed by normalizing each variable or group of variables by subtracting the median value and dividing by the interquartile range. The training set (80 %), the validation set (10 %), and the test set (10 %) have been randomly chosen.

4.2.2 Training and validation

Here, the NNs considered in this application are introduced. Specifically, two feed-forward networks, a *vanilla* MLP and a single-layer ELM, and a recurrent one such as the LSTM, have been taken into account.

The ELM network belongs to the family of efficient and fast learning Reservoir Computing Networks (RCN)s [94, 95], which use random, nonlinear projections of inputs into a high-dimensional feature space. Their input nodes are randomly connected to a single hidden layer, the so-called *reservoir*, consisting of nonlinear neurons (see Figure 4.4 for the ELM learning architecture). Only the connections from the hidden layer to the output are trained, typically using linear regression. This approach eliminates the need for iterative training, which is required in traditional neural networks, resulting in faster learning time. They have been proven to perform well in many applications, among which there are the forecasting of Remaining Useful Life (RUL) of mechanical components [96], the radar signal processing [97], the diagnosis of faults for industrial systems [98], the noise-robust speech processing [99] and, in medicine, they have been used for cancer classification [100].

[1] In fusion *jargon* the *gaps* refer to a finite number of distances between the first wall and the plasma boundary computed along a given set of segments. They are used to control the plasma boundary [92], while the *X-point* is the point at the plasma boundary where the poloidal magnetic field is zero, and can be observed in the so-called diverted or X-shaped plasmas (see Tutorial 10 in [93]).

LSTM networks have been considered to assess the performance of recurrent NNs in the proposed architecture. However, compared to feed-forward solutions, the RNN introduces a delay in estimation that may lead to loss of control when inserted into a feedback stabilization loop. Indeed, as will be shown in Section 4.2.2, by using LSTMs it is not possible to satisfy the VS requirements, being negligible the maximum rejectable disturbance.

The following subsections will introduce the considered NNs with their architectures and optimized hyperparameters used for the training. Moreover, different trainings performed on each network will be presented and compared, while in Section 4.3 the comparison between the different NNs will be discussed. Indeed, to evaluate the generalization property of the NNs different trainings have been carried out by expanding the training database while keeping fixed the NNs architectures. This gives an idea of how much data need to be shown to the considered NNs to enlarge the operational space of the ES-based VS algorithm. A summary of the numerical experiments considered is reported in Table 4.3 while more details will be given in the next subsections.

The results shown in this section and the next refer to the closed-loop assessment of the NNs when included in the VS feedback loop of the ITER MCS (see schemes in Figure 1.8). Moreover, the simulations have been performed considering the counteraction of VDEs of 6 cm. The rejection of such VDEs, also in closed-loop, already shows the generalization property of the NN since only smaller disturbances were shown to the NNs during the training. It is also worth remarking that for the equilibria of Group A the Kalman filter considered in [55] is not able to reject a 6 cm VDE, while the considered NNs can stabilize such disturbance. Therefore, when the response to such a VDE is presented for this group of models, the Kalman filter performance is not reported. On the contrary, the Kalman filter designed exploiting one of the Group B equilibria in [55] is able to reject the 6 cm VDEs for the three Group B models and, therefore, the performance of the Kalman filter has been compared with obtained NNs. Time traces of the vertical displacement of the plasma centroid with respect to the equilibrium position, δZ_c, of the voltage u_{VS3} and the current I_{VS3} in the VS circuit have been reported for all the test cases considered.

MLP hyper-parameters

Activation function	ReLu
Optimizer	Adam
Learning rate	0.001
Hidden layers	2
Hidden layers dimension	400
Batch size	2000
Dropout	10%
Number of epochs	20

(a) Set of optimized hyper-parameters used for training the MLP described in Section 4.2.2.

ELM hyper-parameters

Input weight scaler α_U	1.5
Leaking rate λ	0.3
Regularization ϵ	0.0001
Reservoir size	2000
Activation function	*htan*
Input sparsity K^{in}	5

(b) Set of optimized hyper-parameters used for training the ELM described in Section 4.2.2.

Multilayer Percepton

In this section, the MLP network is considered. The closed-loop results for the MLP models obtained from the experimental trainings reported in Table 4.3b, show how a richer training database can improve both the generalization property of the network and the overall VS performance.

MLPs are a type of feed-forward neural network that is commonly used for classification and regression tasks. It consists of multiple hidden layers of fully interconnected nodes, or neurons, that process input data to generate output predictions. The weights and biases of these neurons are learned via backpropagation. Each neuron in the hidden layers applies a nonlinear activation function to its inputs, allowing the network to model complex relationships between the input and output data. They are known for their ability to model complex nonlinear relationships between input and output data, making them a powerful tool for solving a wide range of problems.

In Table 4.2a the MLP hyper-parameters used for the training are reported. A random search, by executing different trainings using different values of the NNs hyperparameter, was carried out to find the best-performing perceptron model on the validation and test data sets. Starting from the backpropagation parameters (optimizer algorithm and learning rate), the network architecture was optimized with the choice of a wider than deep network: two hidden layers of 400 neurons each, with Rectified Linear Unit (ReLU) activation functions, defined as $ReLU(x) = \max(x, 0)$ .. The batch size of 2000 for data sampled with a frequency of 20 kHz, which implies that each batch carries information of 0.1 s of the total simulation time. The use of smaller batches, while significantly increasing the computation time, did not lead to relevant performance improvements. The same consideration holds for the number of epoch runs for training the network. The addition of dropout layers (one after each fully connected layer) improved the generalization property of the network, especially when further tested in closed-loop.

Furthermore, during the training process of the MLP model, it appeared that the synthetic shots performed to collect the training data needed to last at least 5 s. Indeed, training performed on databases with shots of shorter lengths leads to MLPs that in closed-loop are not able to produce a good estimation of the plasma unstable movement in the long term.

Table 4.3b reports the characteristics of three different training strategies of the MLP. Training A was the first performed with a training database that counts 100 shots obtained considering only one equilibrium from Group A (see Table 3.1a). The chosen model is the same one that was considered in [55] to build the Kalman filter for the equilibria in Group A. What matters is that the MLP obtained from Training A is able to stabilize 26 models among the 27 available ones. Indeed, it is possible to use

MLP trainings

	Shots	Equilibria	Shots T_f	Stabilization rate
MLP Training A	100	1	$5\,s$	$^{26}/_{27}$ equilibria
MLP Training B	200	5	$5\,s$	$^{27}/_{27}$ equilibria
MLP Training C	200	10	$5\,s$	$^{27}/_{27}$ equilibria

(a)

ELM trainings

	Shots	Equilibria	Shots T_f	Stabilization rate
ELM Training A	200	5	$1\,s$	$^{27}/_{27}$ equilibria
ELM Training B	400	10	$1\,s$	$^{27}/_{27}$ equilibria
ELM Training C	900	24	$1\,s$	$^{27}/_{27}$ equilibria

(c)

Table 4.3: A summary of the training performed for the two considered NNs is reported in these tables. Training databases were changed to evaluate the generalization property of the different networks and the robustness of the VS feedback loop when data-driven NNs are used instead of the Kalman filter. (b) refers to three different trainings of the MLP model while (d) of the ELM. The number of shots collected for each training database, the number of equilibria used to generate the synthetic simulations, and the duration of the shot T_f (that is, the duration of the single simulation) are reported. The stabilization rate is related to the potential of NN to produce an accurate estimation of the plasma unstable mode ξ such that it is possible to stabilize VDEs of $6\,cm$. The tests are performed including the NNs in the VS closed-loop of the ITER plasma MCS and considering all the 27 available ITER models.

the same network to stabilize not only Group A models but also *Equilibrium #1* and *Equilibrium #2* from Group B (see Table 3.1b). This is already an enhancement with respect to the use of the Kalman filter, where it is not possible to stabilize with the same filter models from a different group. To increase the robustness, the next trainings were performed taking into account additional equilibria. Indeed, for both Training B and Training C, 200 shots have been collected by respectively considering 5 and 10 equilibria from Group A, respectively.

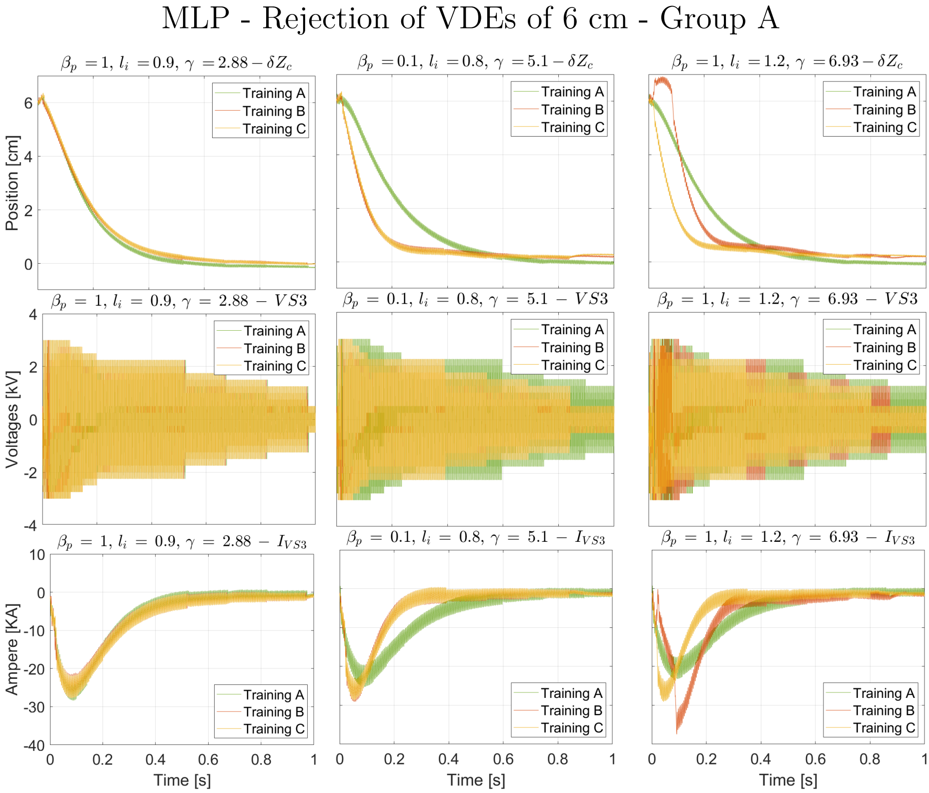


Figure 4.2: The response to a VDE of 6 cm when the MLPs models, obtained from the trainings in Table 4.3b, are substituted to the Kalman filter in the VS control loop is considered. In particular, the time trace of the vertical displacement of the plasma centroid δZ_c, of the voltages u_{VS3} and current I_{VS3} are shown. Three models selected from Group A, whose corresponding equilibrium parameters are reported above the figures, have been considered. These samples were selected to give an idea of the overall performance of the MLPs for the whole family of models in Group A.

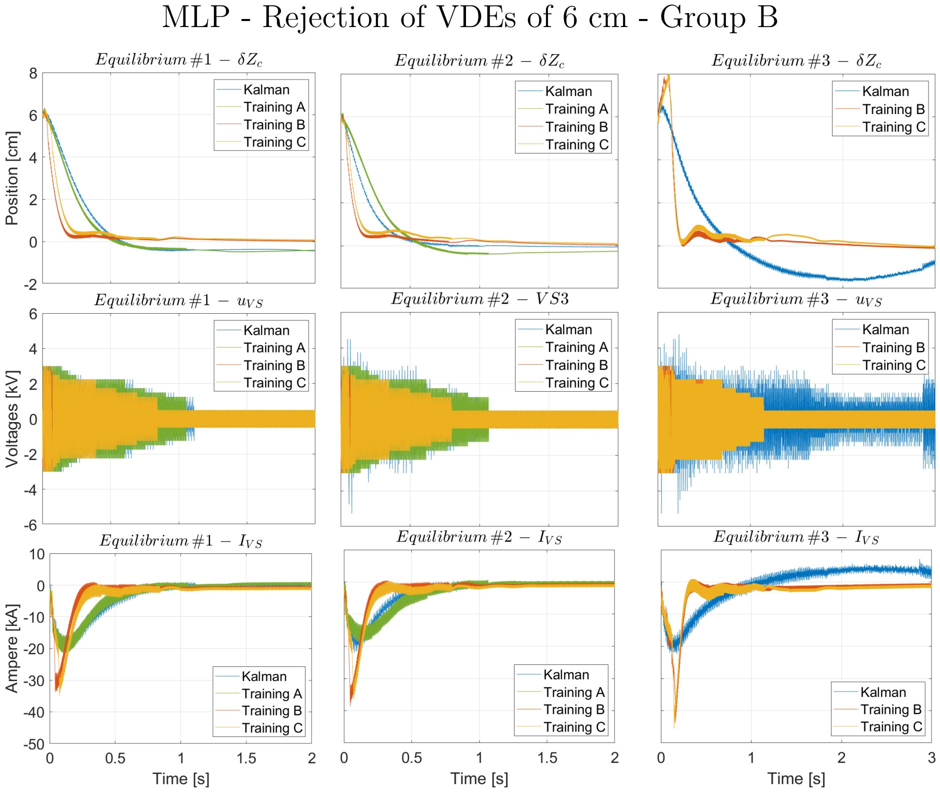


Figure 4.3: The response to a VDE of 6 cm when the MLPs models obtained from the training in Table 4.3b, replace the Kalman filter in the VS control loop is considered. In particular, the time trace of the vertical displacement of the plasma centroid δZ_c, of the voltages u_{VS3} and current I_{VS3} in the in-vessel circuit are shown for the equilibria in Group B. For these equilibria it was possible to stabilize the same VDE also with the Kalman filter. Therefore, the performance of the MLP models is compared with that obtained before introducing the NN in the VS loop.

A comparison between the MLP models obtained from these trainings when an initial VDE of 6 cm is simulated is reported in Figures 4.2 and 4.3. In particular, the results shown in Figure 4.2 refer to equilibria in Group A, whose corresponding equilibrium parameters are reported in the title of each figure. These equilibria have been selected among the 24 available ones to give an idea of the overall performance of the MLPs for the whole family of models in Group A. Indeed, concerning the first one, the MLP models have a comparable performance, while in the second and third it is clear that the network that has seen more equilibria during the training (Training C) generalizes better and outperforms the others. The voltage and current requests in the VS coils are always similar or lower for the MLP obtained from Training C.

A similar consideration holds also for the three equilibria of Group B shown in Figure 4.3. As expected, the MLP obtained with Training A is not capable to reject the VDE for *Equilibrium #3*, while the other two MLPs, trained with richer databases, can assure the stability for all the models considered.

Extreme Learning Machine

The ELM networks are introduced highlighting the characteristics that make this family of NN attractive. Indeed, as a result of this work, compared to MLP, it turned out that they required fewer training data and time to achieve similar or better performance (see Section 4.3).

ELMs belong to the class of single-layer feedforward neural networks. The input nodes are connected to the so-called *reservoir*, a single layer consisting of nonlinear neurons. The basic ELM architecture is shown in Figure 4.4. The idea behind ELM is to randomly initialize the input weights and biases for the neurons of the hidden layer and then optimize only the output weights to solve the target problem.

In the case of ELMs, the neurons in the hidden layer are not interconnected and the basic ELM is thus closely related to a MLP with a much simpler and faster training algorithm. The ELM equation is

$$R_t = (1-\lambda)R_{t-1} + \lambda f_{res}(\mathbf{W}^{in}U_t) \tag{4.1a}$$

$$Y_t = \mathbf{W}^{out}R_t\,, \tag{4.1b}$$

where $\lambda \in (0,1]$ is the leak rate, U_t, R_t and Y_t represent, respectively, the reservoir inputs, the reservoir outputs and the network outputs which

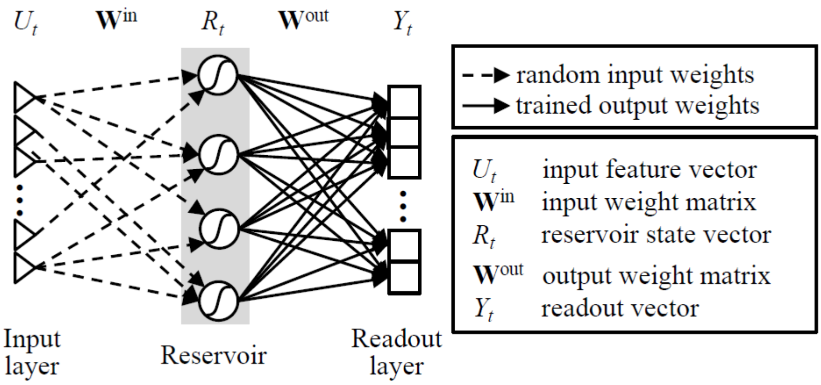

U_t
$\mathbf{W}^{\text{in}}$
R_t
$\mathbf{W}^{\text{out}}$
Y_t
random input weights
trained output weights
U_t input feature vector
$\mathbf{W}^{\text{in}}$ input weight matrix
R_t reservoir state vector
$\mathbf{W}^{\text{out}}$ output weight matrix
Y_t readout vector
Input layer
Reservoir
Readout layer

Figure 4.4: The basic ELM consists of a reservoir and a readout layer. The reservoir is composed of non-linear neurons which are randomly connected to the inputs. The readout layer consists of linear neurons with trained weights.

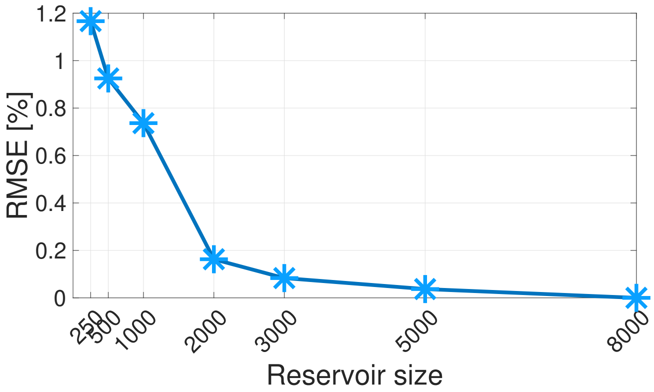


Figure 4.5: Performance of ELM on the validation set as a function of the reservoir size (number of nodes). The RMSE is reported in percentage and normalized with respect to the minimum obtained when using 8000 neurons.

in the RCN context is referred to as *readouts*, at time t. The f_{res} is the non-linear activation function of the reservoir neurons, which, in our case, is taken equal to a hyperbolic tangent. Moreover, $\mathbf{W}^{in}$ and $\mathbf{W}^{out}$ are the input and output weight matrices (see Figure 4.4). The ELM equation provides some sort of output memory *smoothing* the update of the reservoir output by a factor given by the value of the leak rate λ. Indeed, at each time step t, the reservoir output is calculated using a linear combination of the input at time t and of the same output at the previous time step $t-1$.

The weights of the hidden neurons are fixed using a random process that is characterized by two parameters. The first, α_U allows specifying the maximum absolute eigenvalue of the input weight matrix $\mathbf{W}^{in}$ and controls the relative importance of the inputs in the activation of the reservoir neurons. The latter K^{in} defines the number of inputs that drive each reservoir neuron and can be used to control the sparsity of the input weight matrix. The values chosen for these parameters for the ELM network considered in this work are reported in Table 4.2b.

During the training, the ELM algorithm tries to find a set of output weights that minimizes the mean square difference between the readouts Y_t and their desired values D_t across the available training examples. Given the $\mathbf{R}$ and $\mathbf{D}$ matrices with columns R_t and D_t, respectively, the output weight can be found, in closed form, as

$$\mathbf{W}^{out} = (\mathbf{R}\mathbf{R}^T + \varepsilon \mathbf{I})^{-1}(\mathbf{D}\mathbf{R}^T)\,. \tag{4.2}$$

The considered value of the regularization parameter ε is reported in Table 4.2b, together with the other optimized ELM hyperparameters used in the proposed application. Figure 4.5 shows the performance of ELMs on the validation set as a function of the reservoir size (number of nodes). The Root Mean Squared Error (RMSE) for each training is shown in percentage and normalized with respect to the minimum encountered at a reservoir size of 8000 neurons. However, the optimal parameter was chosen as a reservoir size of 2000 nodes. Higher numbers of nodes lead to a significant increase in the model complexity and therefore training time while bringing only marginal performance improvement.

Compared to *vanilla* MLPs, in the case of ELMs the training data set cannot be limited to synthetic data generated with only one equilibrium. Indeed, to obtain a suitable estimation of the plasma unstable mode ξ,

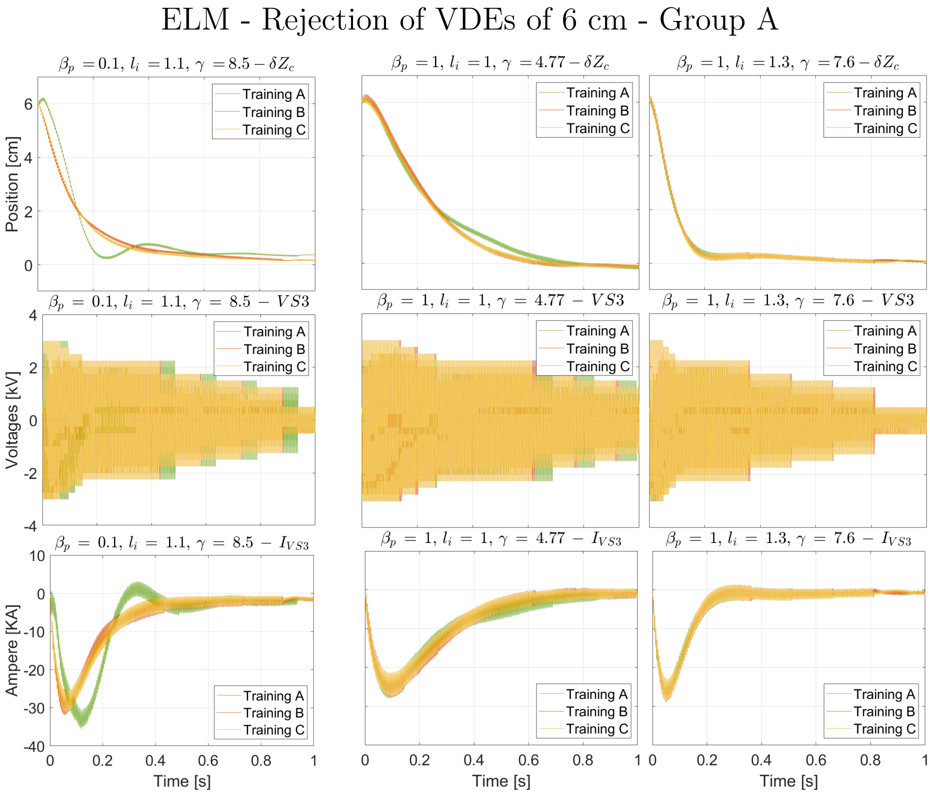


Figure 4.6: The response to a VDE of 6 cm when the ELMs models, obtained from the trainings in Table 4.3d, are substituted to the Kalman filter in the VS control loop is considered. In particular, the time trace of the vertical displacement of the plasma centroid δZ_c, of the voltages u_{VS3} and current I_{VS3} are shown. Three models selected from Group A, whose corresponding equilibrium parameters are reported above the figures, have been considered. These samples were selected to give an idea of the overall performance of the MLPs for the whole family of models in Group A.

data from more than one model must be shown to the network during training. Table 4.3d refers to three trainings for which 5, 10, and all 24 models of Group A were taken into account to generate the data. Each

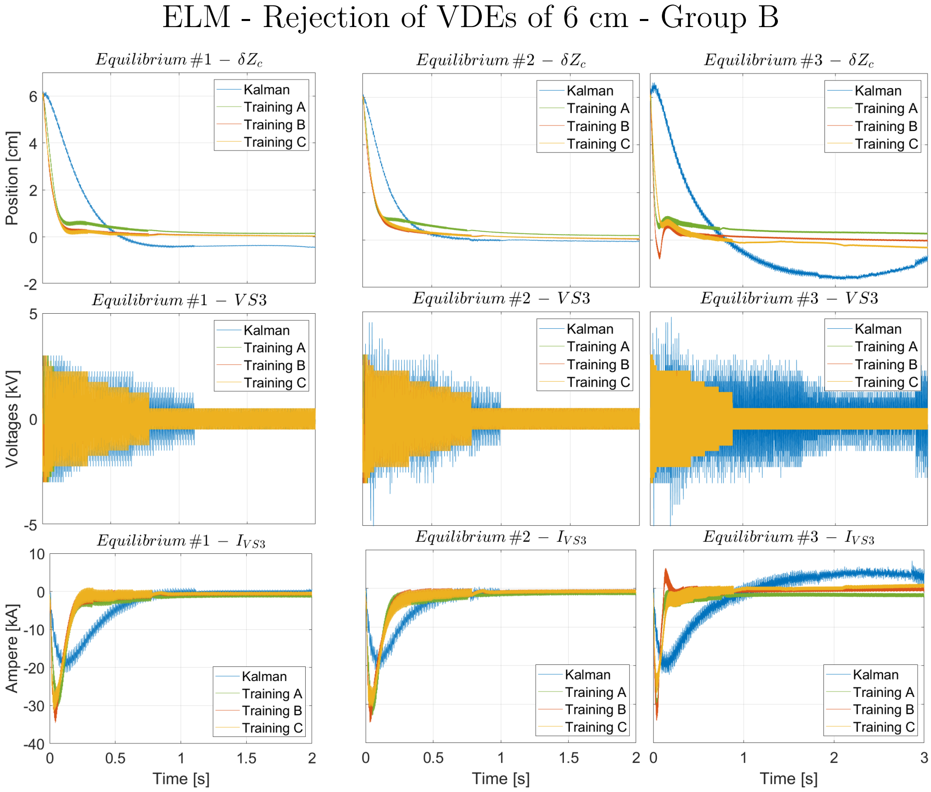


Figure 4.7: The response to a VDE of $6\,cm$ when the MLPs models, obtained from the training in Table 4.3b, are substituted to the Kalman filter in the VS control loop is considered. In particular, the time trace of the vertical displacement of the plasma centroid, δZ_c, of the voltages u_{VS3} and current I_{VS3} in the in-vessel circuit are shown for the equilibria in Group B. For these equilibria it was possible to stabilize the same VDE also with the Kalman filter. Therefore, the performance of the MLP models is compared with that obtained before introducing the NN in the VS loop

training database is also characterized by an increasing number of shots. This allows the network to have access to enough information about the different equilibria that it is seeing. On the other hand, it is sufficient to consider shots that last just 1 s to successfully train ELM networks. The performance of the ELM models obtained from the different trainings in Table 4.3d are shown in Figures 4.6 and 4.7.

Figure 4.6 reports the few cases where the NNs obtained with Training A show slightly different behavior than the others. In general, the three ELM models tested show similar performance in terms of settling time and power request. The same appears in Figure 4.7, where for the three equilibria of Group B the NNs outperform the results obtained when the Kalman filter is employed.

Therefore, in terms of the use of feedforward NNs to estimate ξ in the set-up considered, it is possible to conclude that while for the MLP models the training on richer databases can bring a performance improvement, in the case of ELM networks a smaller group of equilibria for generating the training data is already enough to obtain networks with good generalization property.

Recurrent Neural Network

This section briefly discusses why LSTMs are not a suitable choice for a NN to be included in the VS system considered.

LSTM is a type of RNN that is specifically designed to address the problem of long-term dependency. In classic RNNs, neurons in the layers are allowed to remember the previous sequence of input data by means of recurrent connections: the output of a neuron at a given step is provided alongside the input in the next step. However, over time, as more information piles up, RNNs become less effective at learning new things.

The key feature of LSTM networks is their ability to selectively remember or forget information over long periods of time. This is achieved by including an internal state in the LSTM node, i.e. specialized memory cells that are connected to each other through a series of gates. Gates control the flow of information into and out of cells, allowing the network to selectively remember or forget information based on its relevance to the current task. They are widely used in applications with time-dependent or sequence input data, such as speech recognition, natural language processing, and image captioning.

Machine learning models

	Training	Training	Training	Validation	Training	Stabilization
	data	parameters	RMSE	RMSE	time	rate
MLP	10 models	2400	0.31	0.43	$1\,h\,7\,min$	$^{27}/_{27}$ equilibria
ELM	200 shots	2000	0.4	0.43	$25\,min\,42\,sec$	$^{27}/_{27}$ equilibria
LR	$T_f = 5\,s$	58	1.47	1.37	$35\,sec$	$^{21}/_{27}$ equilibria

Table 4.4: Comparison between the NN models. The number of training parameters, the RMSE obtained on the training and validation data, the training time (measured on an Intel Skylake CPU), and the stabilization rate on the whole available models are reported as performance indicators for the MLP, the ELM, and the LR models. The stabilization rate refers to the capability to reject a VDE of $6\,cm$ when the NN models are included in the VS feedback loop and tested over the whole equilibria available. The training database used was the same for all models.

Since in plasma control the available data are time series data, the use of LSTM seemed a good starting point. However, as a matter of fact, in the proposed study the delay introduced by LSTMs that were successfully validated and tested on the corresponding data sets, was such that stabilization in closed-loop cannot always be guaranteed. Indeed, the use of LSTMs in the VS control loop, with their unavoidable initial delay, results in a loss of control for relatively small VDEs. The maximum VDE that can be counteracted while using the LSTM is of maximum $1\,cm$.

4.3 Results

In this section, the feed-forward NNs (MLP and ELM) studied in this work are compared. The same training database has been used for both models. The data of Training B used for the MLP model in Table 4.3b have also been used for training the ELM model. Specifically, 10 equilibria of Group A (see Table 3.1a) were considered to generate a database of 200 shots simulating the rejection of VDEs between $[-5\,,\,5]$ cm and lasting $5\,s$. Table 4.4 reports the details of the training data together with the performance indicators for the different models. The RMSE obtained on the training and validation data, the number of training parameters for each

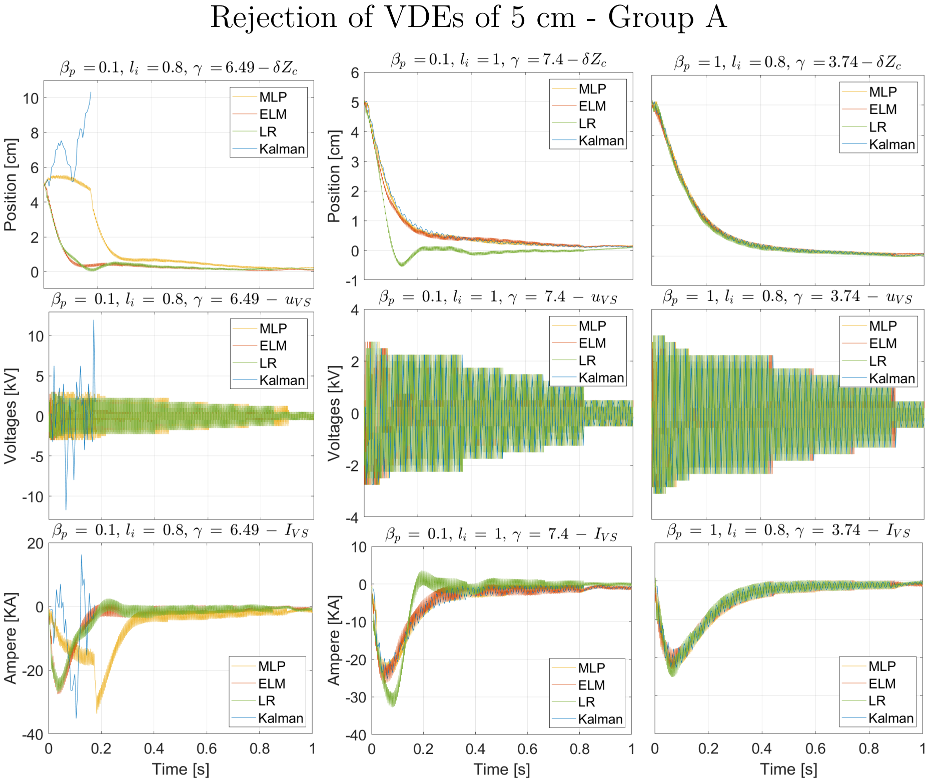


Figure 4.8: Comparison between the MLP, the EML, and the LR models reported in Table 4.4 and the Kalman filter when used in the VS control loop to estimate the plasma unstable mode ξ is reported. In particular, the time traces of the vertical displacement of the plasma centroid δZ_c, of the voltages u_{VS3} and current I_{VS3} in the in-vessel circuit are shown for the equilibria in Group A. The response to an initial VDE of 5 cm has been considered for the simulations.

network, the time required to train the networks, and the stabilization rate for VDEs of 6 cm are reported.

Furthermore, the obtained networks are also compared with a LR model. A simple LR tries to predict the future applying linear regression directly to input features, bypassing the reservoir in the ELM architecture.

It can be considered as a baseline for other more complex networks.

Figure 4.8 reports the rejection of a VDE of $5\,cm$ for three equilibria selected from Group A. The NNs are compared with the simpler LR and the model-based Kalman filter. The latter was able to stabilize such VDE for all minus one equilibrium of Group A while the LR has a stabilization rate of $^{21}/_{27}$ equilibria when a VDE of $5\,cm$ is considered. In fact, the LR model cannot be used to stabilize Group B equilibria (Table 3.1b). In Figure 4.9 the rejection of a higher VDE of $6\,cm$ is considered for the equilibria of Group B and the data-driven NNs are compared only with the Kalman filter and with the model-based ITER-like VS proposed in [27].

The three examples selected from the Group A in Figure 4.8, refer to different results. In the first, NN models are able to stabilize the initial disturbance while the Kalman filter cannot, and the MLP model does not guarantee the same transient performance as the others. For the second equilibrium, when the simplest LR model leads to a slightly different time trace for δZ_c, and lastly the NN models and the Kalman filter yield similar performance. However, from an overall comparison between the NNs studied as substitutes for the Kalman filter, it follows that the ELM model can outperform the MLP for about 5 equilibria over the 24 in Group A. This is especially in terms of the plasma centroid vertical position Z_c recovery time. The same conclusions come when considering the result for the Group B equilibria in Figure 4.9, where the NNs have been also with a classic ITER-like VS system [27].

The ITER-like controller computes the voltage requests for the in-vessel circuit u_{VS3} and for a superconductive circuit, at ITER called $VS1$, usually employed to assist in the stabilization of the plasma column. A combination of the plasma centroid vertical speed $\dot{Z}_c$ and of the in-vessel current I_{VS} are used to obtain control actions, such as:

$$u_{VS3} = \frac{1+s\tau_1}{1+s\tau_2}\left(K_v I_{p_{eq}} \frac{s}{1+s\tau_z} Z_c(s) + K_1 I_{VS}(s)\right) \tag{4.3a}$$

$$u_{VS1} = K_2 I_{VS} \tag{4.3b}$$

where K_v is the plasma speed gain, K_1 and K_2 are the current gains, τ_1 and τ_2 are the time constants of the lead compensator used to adjust the stability margins and τ_z the time constant of the derivative filter (for more

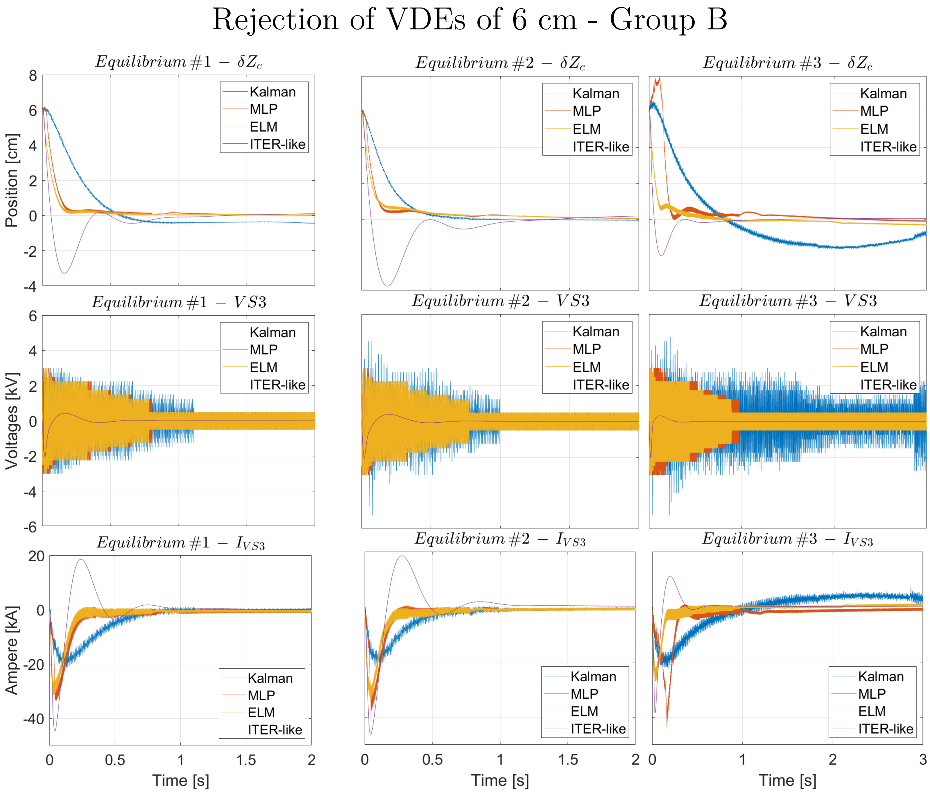


Figure 4.9: The response to a VDE of 6 cm when the MLP and ELM models, obtained from the training setup described in Table 4.4, and the Kalman filter are used in the VS control loop to estimate the plasma unstable mode ξ is reported. In particular, the time traces of the vertical displacement of the plasma centroid δZ_c, of the voltages u_{VS3} and current I_{VS3} in the in-vessel circuit are shown for the equilibria in Group B. For a baseline comparison, the results of a classic ITER-like VS algorithm are also reported.

details, the reader can refer to [27, 40]).

The MLP and ELM models outperform the Kalman filter in terms of both δZ_c settling time and voltage u_{VS} and current I_{VS} request for

Rejection of a VDEs of 6 cm - Group B

	Equilibrium #1		*Equilibrium #2*		*Equilibrium #3*	
	IEA	ITAE	IEA	ITAE	IEA	ITAE
Kalman filter	3.1×10^{-2}	5.8×10^{-2}	1.5×10^{-2}	2.2×10^{-2}	7×10^{-2}	2×10^{-1}
MLP	8.3×10^{-3}	7.7×10^{-3}	9.6×10^{-3}	9.5×10^{-3}	2.3×10^{-2}	3.5×10^{-2}
ELM	8.1×10^{-3}	1.6×10^{-2}	8.9×10^{-3}	7×10^{-3}	2.1×10^{-2}	7.5×10^{-2}
ITER-like	1.4×10^{-2}	2×10^{-2}	9.6×10^{-3}	9.5×10^{-3}	7.7×10^{-3}	9.6×10^{-3}

Table 4.5: Performance indices for the VS controllers for the rejection of a VDE of 6 cm when considering the models in Group B. The corresponding simulation time traces are those reported in Figure 4.9

all three equilibria. However, they assure similar performance in case of *Equilibrium #1* and *Equilibrium #2* while the ELM model seems to generalize better for *Equilibrium #3*. As already seen in Section 4.2.2, the MLP models obtained both from Training B and Training C are able to stabilize the latter equilibrium but with an initial overshoot in the vertical position.

Furthermore, the effectiveness of the proposed VS controllers is evaluated by considering the Integral Absolute Error (IAE) and Integral Time-weighted Absolute Error (ITAE) performance indexes. Given the specific objective of the vertical stabilization task, which aims at bring to zero the plasma vertical displacement, these indexes are computed as follows:

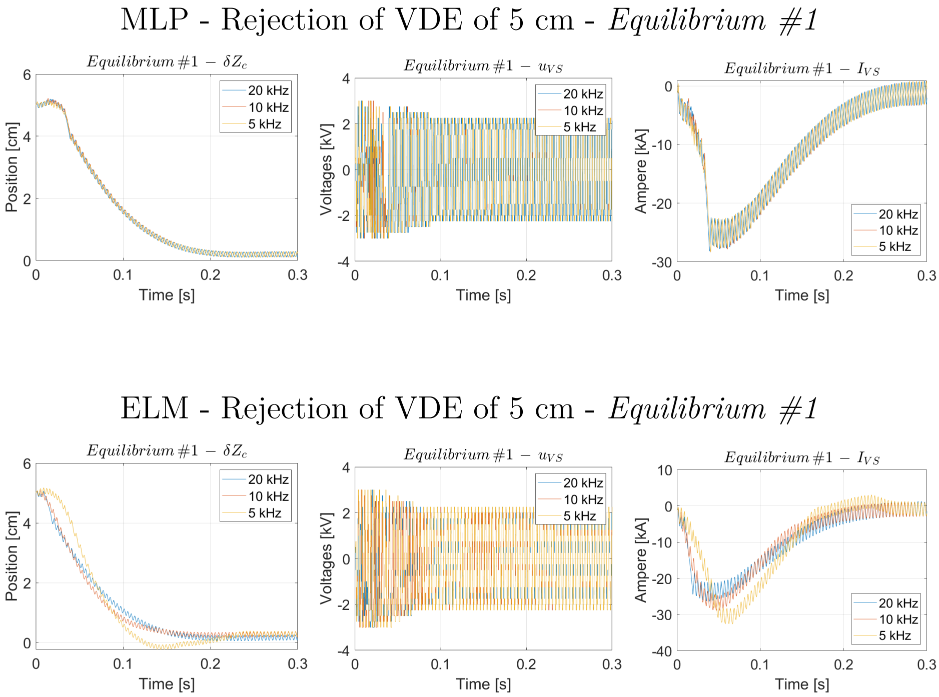
$$IAE = \int_0^{t_f} |\delta Z_c(t)| \, dt \,, \tag{4.4}$$

$$ITAE = \int_0^{t_f} t|\delta Z_c(t)| \, dt \,, \tag{4.5}$$

where $t_f = 8\,s$. Table 4.5 reports these performance indices for the VS systems compared in Figure 4.9 for Group B models. It follows that, if the NNs are included in the VS loop then it is possible to improve the overall performance, since both the IAE and ITAE indices assume values smaller than those assumed when the Kalman filter is used, or when the model-based ITER-like VS is considered; in some cases, the improvement is almost an order of magnitude. This improvement comes only partially at the expense of a greater control effort due to the ES switching control law. Indeed, there is an improvement between the ES architecture with the Kalman filter and the one with NN, although both rely on a switching control law.

Finally, the two NNs can also be compared in terms of computational complexity. From Table 4.4 it is clear that even if the training and validation losses, computed as RMSE between the sampled data and the predicted ones are similar for the two models, the training times are quite different. On the same database and on the same hardware (one core of Intel(R) Xeon(R) Gold 6226R), the ELM model is almost three times faster than the MLP one. This latter result is making RCNs more appealing compared to the traditional deep learning architecture. This is due to that while for the MLP model the number of parameters to be trained is given by the interconnection between all the network layers (input, hiddens, and output), for the ELM the only trained parameters are those that connect the reservoir to the output layer.

Lastly, both NNs have been trained on data sampled at the frequency of $20\,kHz$. Therefore, when the trained NNs were included in the VS closed loop during the test phases, they were executed at the same fixed step. All reported results are obtained considering the NNs computation rate at the frequency of 20 kHz. Further analysis of the NNs prediction property, when executed at a different rate, has been performed. In Figures 4.10a and 4.10b the simulations of the rejection of an initial VDE of $5\,cm$ for the MLP and ELM model defined in Table 4.4 when performed at different rates are reported. The results obtained with the baseline frequency of 20 kHz are compared to those obtained with the slower frequency of 10 kHz and 5 kHz, respectively. It should be noted that the slightly worsening of the performance obtained at 5 kHz is not related to the use of a specific NNs, but to the fact that the estimation of ξ is updated less frequently, hence inducing a delay in the closed-loop. It follows that the MLP and ELM models can also run with an execution time slower than the sample time used to collect the training data.



MLP - Rejection of VDE of 5 cm - *Equilibrium #1*

ELM - Rejection of VDE of 5 cm - *Equilibrium #1*

Figure 4.10: (a) The response to an initial VDE of $5\,cm$ when the MLP model from Table 4.4 is executed at different sampling rates is reported. The time behavior of the plasma centroid vertical position δZ_c, of the voltage u_{VS} and current I_{VS} in the VS circuit are reported. The MLP model can be run at a frequency around $1\,kHz$ and retain its generalization property. While at lower values of computational rate, the stabilization capability of the proposed VS system is lost.(b) The ELM model defined in Table 4.4 has been performed at different frequencies during the simulation of the rejection of a VDE of $5\,cm$. The lower computational rate for this model that still allows for satisfactory performance is associated with $4\,kHz$.

Summary

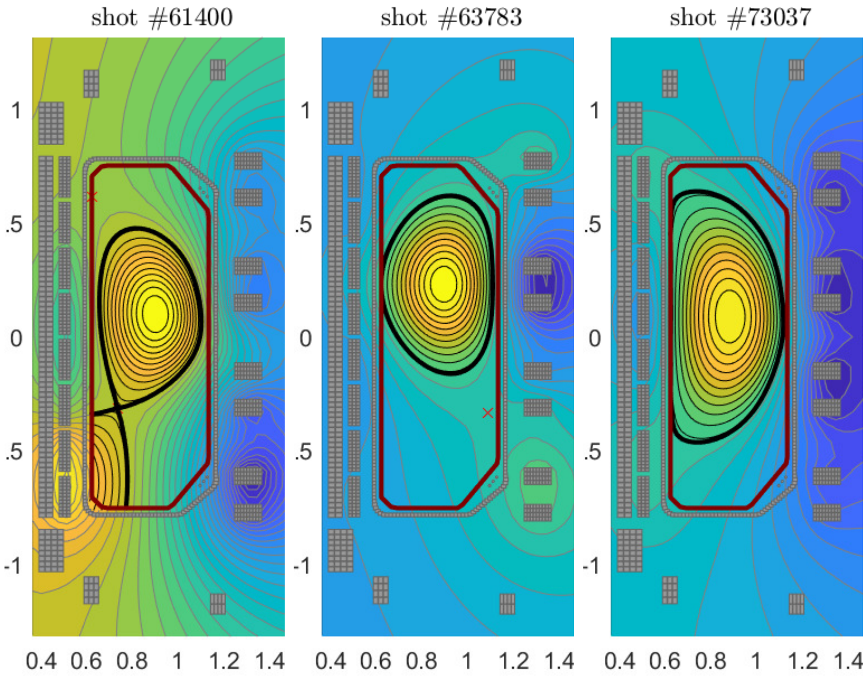
In this chapter, the possibility of replacing the Kalman filter in the bounded ITER ES-based VS proposed in [56, 55] with a NN has been investigated.

The ES control law requires knowledge of the state of the system, which was previously ensured by the use of the model-based Kalman filter. Even if in [55] the same filter model was employed in different scenarios, this approach suffers from the dependency on the models used to design the Kalman filter. Therefore, the main contribution of this part of the thesis work is to show how the use of a NN can improve the robustness and the generalization property of the ES approach, making it fully model-free. Two feed-forward NNs have been studied and suit the purpose: a classic MLP and an ELM network. For both models, different training has been performed especially to understand their generalization capability when tested in closed-loop on plasma equilibria not seen during the training. The ELM network proved to perform better, while also ensuring shorter training times.

5 Application of the Extremum Seeking approach to the TCV case

This chapter treats the research project to validate in TCV the model-free ES-based VS system originally proposed in [55, 56]. In fact, given the flexibility of both the machine and the control system [101, 102], TCV has often been the machine where cutting-edge control techniques have been tested for the first time, such as H-infinity [103] and DRL [104], and therefore represents a good candidate to evaluate the capacity of the proposed ES-based VS system.

A preliminary assessment of the proposed VS architecture for TCV has been carried out considering the bounded ES control law (2.4). The latter algorithm is an appropriate application given the switching power supply currently adopted at TCV to feed the G coils used as actuators by the VS system [105]. Furthermore, the version based on the Kalman filter has been taken into account since it does not require a training simulation campaign for the NNs.



(a) (b) (c)

Figure 5.1: TCV plasma equilibria considered for the preliminary studies. The corresponding main plasma parameters are reported in Table 5.1.

The TCV equilibria considered are reported in Figure 5.1. They correspond to different plasma shapes and different growth rates γ, as reported in Table 5.1.

The simplified Simulink scheme of the TCV magnetic control system shown in Figure 5.2 has been built considering only the VS system. As a consequence of this choice, the Kalman filter used to estimate plasma movement along the unstable dynamic $\hat{\xi}$ considers as input only the voltage applied to the G coils used for the VS, V_{a_G}, and the corresponding current I_{a_G}, the plasma current I_p and the plasma centroid vertical position zI_p. The Kalman filter has been designed considering a reduced

TCV plasma equilibria parameters

Pulse	Time instant	$I_{p_{eq}}$	γ
#61400 (Single null)	$t = 1.5\,s$	$210.6\,kA$	$1420\,s^{-1}$
#63783 (Negative triangularity)	$t = 1\,s$	$252.3\,kA$	$620\,s^{-1}$
#73037 (Elongated)	$t = 0.4\,s$	$554.7\,kA$	$2034\,s^{-1}$

Table 5.1: Main plasma parameters for the equilibria reported in Figure 5.1.

k	α	ω
$2.3 \cdot 10^{-2}$	50	2000π rad/s

Table 5.2: Control parameters for the proposed model-free VS system based on the bounded ES control law (2.4).

plasma linear model of order equal to 25 of the equilibrium obtained at the time instant $t = 1.5\,s$ from pulse *#61400*, whose shape is reported in Figure 5.1a. It is worth remarking that, despite the equilibria considered for the assessments being different, the same Kalman filter has been used for all the considered cases.

The considered ES control parameters are reported in Table 5.2. The power supply that feeds the G coils at TCV has been modeled according to the non-linear model described in [105]. Therefore, although simplified, the simulation scheme considered includes the non-linear behavior of the power supply.

The preliminary assessment have been performed by running the following two test cases:

- rejection of a VDE when only the ES-based VS system is considered in closed-loop;
- rejection of a VDE when the plasma current and plasma centroid position control loop currently adopted at TCV are included. This allows to perform a preliminary evaluation of possible unwanted coupling between the considered VS system and the remaining plasma

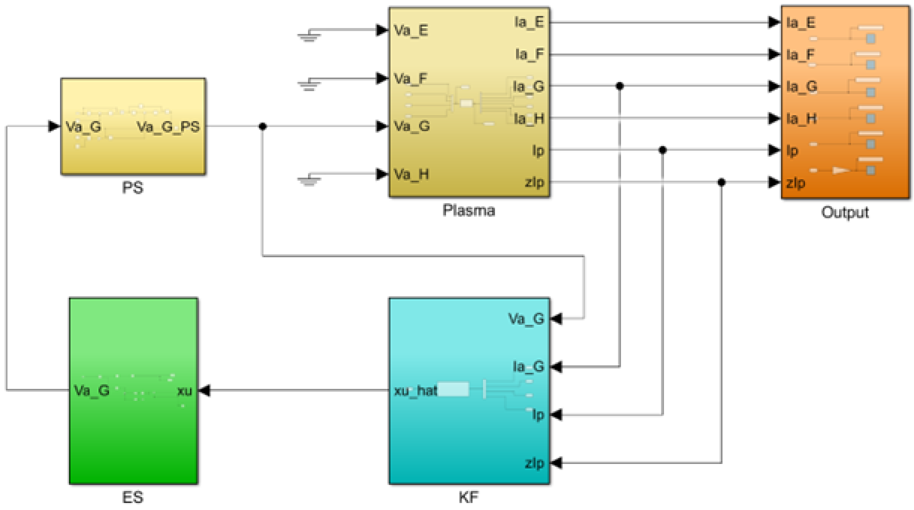
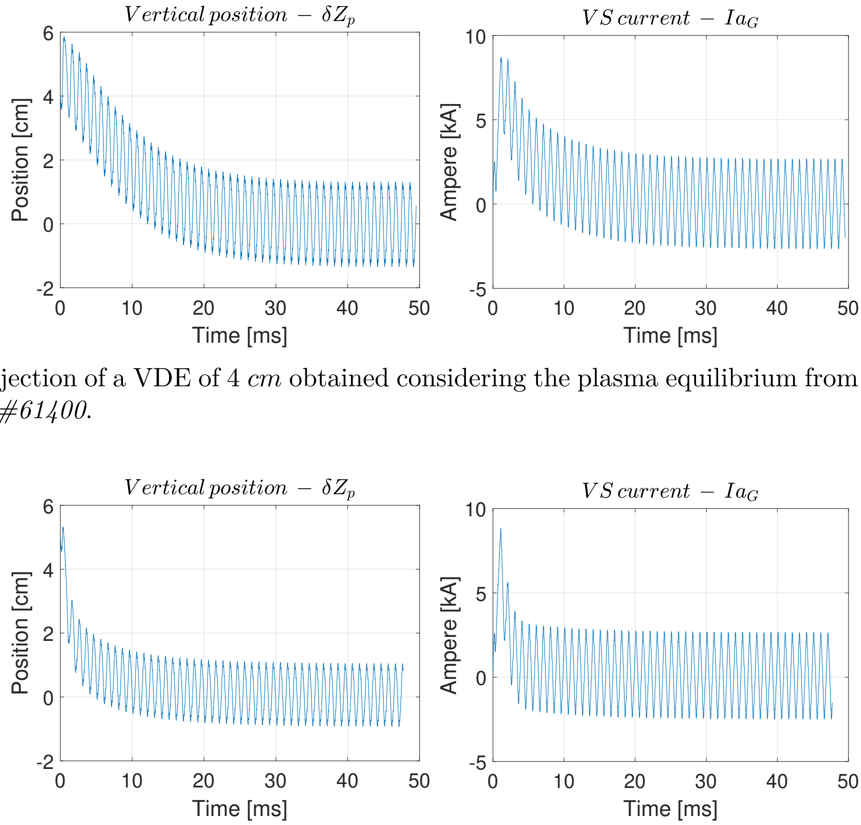


Figure 5.2: Simulink scheme of the TCV ES-based VS, including the nonlinear model of the switching power supply for the G coils.

magnetic control loops.

Rejection of VDEs The response to a VDE of 4 cm for the TCV plasma equilibrium obtained from pulse *#61400* is reported in Figure 5.3a, while Figure 5.3b shows the rejection of a VDE of 5 cm when the equilibrium from pulse *#63783* is considered. The time traces of the plasma vertical position displacement δZ_p and of the current in the G coils $I_a G$ have been reported.

Assessment with the other control loops Currently, the TCV plasma magnetic control system is the so-called *hybrid controller* that includes the plasma shape and current controllers. Specifically, it is a MIMO-Proportional–Integral–Derivative (PID) controller where the plasma centroid radial and vertical positions R_p and Z_p, respectively, are controlled in feedback together with the plasma current I_p, while orthogonal PF coil currents are controlled in feedforward based on pre-computed waveforms. A mutual decoupling and resistive compensation term is then employed to obtain the voltages to be applied to the coils starting from the current



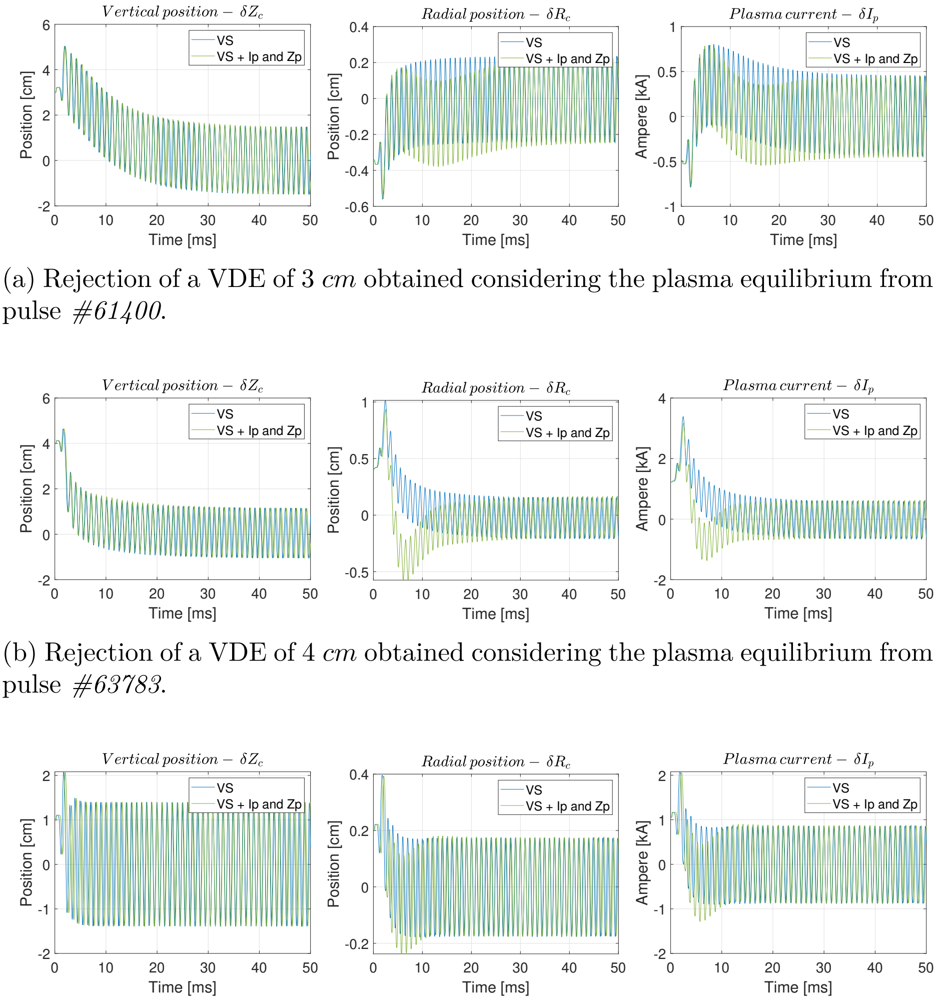
(a) Rejection of a VDE of 4 cm obtained considering the plasma equilibrium from pulse *#61400*.

(b) Rejection of a VDE of 5 cm obtained considering the plasma equilibrium from pulse *#63783*.

Figure 5.3: Rejection of VDEs when only proposed the ES-based VS scheme is considered. The Simulink model used to obtain these results is reported in Figure 5.2.

requests provided by the hybrid controller.

For this preliminary work assessment, only the plasma vertical position and current loops have been considered, to evaluate the interaction of the proposed ES-based VS system with the other magnetic control tasks. Moreover, the power supply for the G coils has been modeled as a time delay $\tau = 0.05\ ms$ (more details can be found in [105]).



(a) Rejection of a VDE of 3 *cm* obtained considering the plasma equilibrium from pulse *#61400*.

(b) Rejection of a VDE of 4 *cm* obtained considering the plasma equilibrium from pulse *#63783*.

(c) Rejection of a VDE of 1 *cm* obtained considering the plasma equilibrium from pulse *#73037*.

Figure 5.4: Rejection of VDEs when only proposed the ES-based VS scheme is coupled with the plasma vertical position Z_p and current I_p loops.

In Figure 5.4 is reported the response to VDEs of different amplitude for the TCV plasma equilibria considered (see Table 5.1). Specifically, for each equilibrium, the comparison between the rejection of the same VDE when only the proposed ES-based VS system is considered and when also the plasma position and current feedback loops are closed is shown. The time traces of the variation of the plasma vertical position δZ_p, the plasma radial position δR_p and current I_p around the given equilibrium are also reported. As expected, there are induced oscillations in all the controlled variables, due to the sinusoidal perturbation injected by the ES-based VS. However, as far as the I_p control is concerned, the amplitude of such oscillations is negligible, while, in the worst case, the oscillations on the position of the plasma centroid are within $\pm 2.5\ mm$ in the case of R_p, while are within $\pm 1.5 .cm$ for Z_p, for the considered VDEs. These results are considered encouraging.

Part II

Data-driven Vertical Stabilization of tokamak plasma via Reinforcement Learning

6 Reinforcement Learning

This chapter starts with a brief introduction on RL followed by a review of the application of RL algorithms in the field of nuclear fusion. Subsequently, the RL algorithms considered in this thesis to develop data-driven VS, the Q-learning and DDPG specifically, are presented. More details about RL can be found in Appendix A.

RL [106] is a branch of Machine Learning (ML) dedicated to developing intelligent agents capable of making autonomous decisions in dynamic and uncertain environments. It can be especially beneficial in tackling goal-directed and decision-making problems where the optimal course of action is not known beforehand and the agent must learn by trial and error, i.e., making mistakes along the way.

At its core, RL is a process of successive interactions between an agent and an environment. At each moment, t, the agent takes an action, a_t, and receives feedback from the environment in the form of rewards or penalties, R_t. This feedback helps the agent learn the optimal behavior based on the corresponding outcome of its actions. Indeed, the goal of the agent is to maximize, over time, the cumulative reward it receives from the environment by learning the optimal policy that maps states to actions. In this framework, the agent does not need prior knowledge of the environment; instead, it continuously learns from its interactions with it. This allows the implementation of efficient model-free adaptation algorithms without the need for ad hoc solutions.

In recent years, considerable progress has been made especially in the area of DRL [106] where RL algorithms are combined with deep NNs. This combination has enabled remarkable success in solving complex tasks such as playing Atari games, mastering the game of Go, and controlling robotic systems. Indeed, DRL has recently become a promising approach in the nuclear fusion community. So far, few attempts have been made to apply these techniques to avoid tokamak tearing instability [107], to control plasma internal profiles [108]), or to design optimal feedforward reference waveforms [109] by exploiting a virtual tokamak environment (also based on NN [110]), and constraining the plasma state in terms of stability or magnetic structure [111]. DRL has also been considered to address the plasma magnetic control problem. A full demonstration of a DRL agent capable of solving the whole magnetic control problem at TCV has been successfully tested for a diverse set of plasma configurations [104].

This thesis presents the design of two RL-based VS systems as a potential data-driven solution for the VS problem in tokamak plasmas. To begin with, a Q-learning algorithm is applied to the VS system of EAST tokamak [112]. Additionally, the possibility of using DRL with the International Thermonuclear Experimental Reactor (ITER) plasma and MCS as the environment for a DDPG algorithm is explored [113].

6.1 Q-learning

Q-learning is a model-free, value-based, off-policy algorithm that is used for systems with discrete action and state spaces. It stores the expected cumulative rewards for each state-action pair, referred to as Q-values, in a table known as a Q-table. The Q-values, which are calculated from the action-value function $Q(s, a)$, indicate the quality of taking a certain action in a given state.

The Q-learning algorithm utilizes an iterative approach to refine its tabular approximation of the action-value function $Q(s, a)$. It does this by excuting the following steps:

1. Begin by initializing the Q-table with random values or zeros for all state-action pairs.

2. Select the action to take based on the current state using an exploration-exploitation trade-off strategy. Specifically, the agent follows an ϵ-greedy policy to explore new actions and gather more information about the environment at the beginning and then gradually exploits the learned knowledge to make optimal decisions, choosing actions that correspond to the highest Q-values.

3. After performing the selected action, observe the next state of the environment and the corresponding reward.

4. Update the Q-value for the current state-action pair using the Bellman equation, which is a mathematical formula used to calculate the expected cumulative reward. This equation makes use of Temporal Difference (TD) learning to update the optimal action-value function following the law:

   $$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \theta(R_t + \Gamma Q(S_{t+1}, a) - Q(S_t, A_t)) \tag{6.1}$$

   where $Q(S_t, A_t)$ is the Q-value for the current state-action pair, $Q(S_{t+1}, a)$ is the Q-value for the subsequent state, τ is the discount factor, and $\theta \in [0, 1)$ is a step size parameter.

5. Repeat steps 2-4 until convergence or for a predefined number of iterations.

Through this process, the agent can effectively navigate complex environments and learn optimal decision-making strategies, which are the sequences of actions that lead to the maximum cumulative reward.

6.2 Deep Deterministic Policy Gradient

DDPG is an off-policy, online, model-free RL algorithm that is especially suitable for tasks with a continuous action space. Indeed, it is an actor-critic algorithm that combines the advantages of both policy-based and value-based methods. It uses a neural network (the *actor network*), to approximate the policy and select actions based on the current state, and another neural network, (the *critic network*), to estimate the action-value function Q. The use of an actor network to directly output continuous actions allows the DDPG agent to be effective also in the case of plasma magnetic control and accurately represents the plasma behavior.

Additionally, since an off-policy algorithm is employed, the critic and actor networks are replicated in the so-called target networks. The weights of these networks are updated slowly over time towards the weight of the original networks using a soft update mechanism. This helps to provide a more stable and accurate estimation of the Q-values while also ensuring a more consistent and reliable learning process.

Generally, the DDPG training algorithm performs the following steps:

1. Initialize the actor and critic networks with random weights.

2. Copy the weights of the original networks to the target networks for both actor and critic networks.

3. Select an action based on the current state using the actor network and add noise to encourage exploration.

4. Execute the chosen action and observe the subsequent state and the related reward.

5. Store the experience tuple (state, action, reward, next state) in a replay buffer.

6. Draw a mini-batch of experiences from the replay buffer.

7. Use the data in the mini-batch to update the critic network by minimizing the mean squared TD error between the predicted Q-value and the target Q-value .

8. Use the data in the mini-batch to update the actor network by computing the gradient of the expected return with respect to the actor's parameters and performing gradient ascent.

9. Gradually blend the weights of the target networks with the weights of the original networks.

10. Repeat steps 3-9 until convergence or for a predefined number of iterations.

The replay buffer is a memory structure that stores the experiences (state, action, reward, next state) the agent has encountered while interacting with the environment. This allows more efficient use of the data by breaking the temporal correlations between consecutive experiences and providing the agent with a diverse set of transitions to learn from. Indeed, during the training process, mini-batches of experiences are randomly sampled from the replay buffer and used to update the actor and critic networks.

The leverage of both the reply buffer and target networks in the DDPG algorithm offers a more reliable learning process, as it reduces bias and variance in updates, thus resulting in better convergence and performance.

Finally, similar to Q-learning, DDPG uses an exploration-exploitation trade-off strategy, by introducing decaying noise in selecting actions. This allows the agent to balance the initial exploration of new actions with the exploitation of the acquired knowledge to make the best decisions.

7 Q-learning based Vertical Stabilization for the EAST tokamak

This chapter shows how an agent was trained using the Q-learning approach to act as an alternative VS system for the EAST tokamak MCS. The objective of the training is to make the mentioned controller learn how to counteract the plasma vertical instability by interacting with a linearized plasma model in simulation, which plays the role of the RL environment.

A common choice for the feedback quantities of a tokamak VS system is represented by the in-vessel current and the vertical velocity of the plasma centroid. For this reason, the couple $(I_{IC}\,, \dot{Z}_c)$[1] has been chosen as the environment observed quantity. The RL-based VS agent will then implement a policy to select the voltage request to the in-vessel coils V_{IC} based on such observations. Note that the plasma current and shape

[1] In the case of the EAST tokamak the in-vessel circuit dedicated to the VS is referred to as IC

	V_{IC}	I_{IC}	$\dot{Z}_c$
Points number	17	21	21
Max absolute value	300 V	6 kA	30 m/s
Bonus assignation threshold	–	50 A	0.5 m/s

Table 7.1: Number of discretization points and ranges for both states and action of the proposed VS agent. The state thresholds for the assignment of the bonus are also reported.

controllers are not included in the training environment to reduce the computational burden. However, validation of the VS agent is carried out taking into account the overall EAST plasma MCS.

The state and action spaces have been both linearly discretized. The number of points used for the discretization together with the corresponding variable ranges used in the Q-learning algorithm are reported in Table 7.1. The following reward function has been adopted:

$$R(s\,,a) = -k_1 \cdot \left(\frac{\dot{Z}_c}{\dot{Z}_{c_{\max}}}\right)^2 - k_2 \cdot \left(\frac{I_{IC}}{I_{IC_{\max}}}\right)^2 - k_3 \cdot \left(\frac{V_{IC}}{V_{IC_{\max}}}\right)^2 , \quad (7.1)$$

being the state $s = (I_{IC},\, \dot{Z}_c)$ and the action $a = V_{IC}$, and where $\dot{Z}_{c_{\max}}$, $I_{IC_{\max}}$ and $V_{IC_{\max}}$ refer to the maximum values specified in Table 7.1. The former two terms in (7.1) reflect the main objective of the VS system, i.e. to stop the vertical unstable motion of the plasma, while keeping the in-vessel current as low as possible. On the other hand, the latter term penalizes high values of the control action V_{IC}; indeed, a low voltage limits the control power and further contributes to keeping the in-vessel current low. For the case of the EAST tokamak, the reward gains have been set as $k_1 = 3\,, k_2 = 1$ and $k_3 = 0.2$.

In Algorithm 1 the process of training the VS agent for the EAST tokamak while following the Q-learning algorithm is reported. The initial value of $Q(s, a)$ for each tuple $(\dot{Z}_c,\, I_{IC},\, V_{IC})$ is set equal to $R(s, a)$ at the beginning of the training procedure (see **Step 2** in Algorithm 1), while the cumulative reward G is initialized to 0 at the beginning of each episode

```
Input:
  • state $s = (\dot{Z}_c \, , I_{IC})$ and action $a = V_{IC}$ discretized spaces
  • Maximum state values $\dot{Z}_{c_{\max}} \, , I_{IC_{\max}}$
  • Bonus assignation threshold $\dot{Z}_b \, , I_{IC_b}$
  • Bonus $b$
  • step size $\theta \in [0, 1)$
  • discount factor $\Gamma \in [0, 1)$
  • initial exploration parameter $\epsilon \in [0 \, , 1)$
  • $\epsilon$-decay factor $\delta \in [0 \, , 1)$
Specify reward function $R(s \, , a)$ according to (7.1)
Initialize $Q(s \, , a) \leftarrow R(s \, , a)$ for all state-action pairs
foreach episode do
    Initialize $s_0$ with a random VDE in the range $[-5 \, , 5]$ cm
    Initialize the cumulative reward $G \leftarrow 0$
    foreach step $t$ in an episode do
        Choose $a_t \in \mathcal{A}$ given $s_t \in \mathcal{S}$ according to the $\epsilon$-greedy policy applied
          on current $Q$ table
        Simulate plasma linearized model starting from state $s_t$ applying
          action $a_t$
        Observe the new state $s_{t+1}$
        Update $Q(s_t \, , a_t)$ according to (6.1)
        Update $s_t \leftarrow s_{t+1}$
        (* Evaluate bonus and episode terminating condition *)
        Initialize the current reward $r \leftarrow R(s_t \, , a_t)$
        if $|\dot{Z}_{c_t}| < \dot{Z}_b$ and $|I_{IC_t}| < I_{IC_b}$ then
            $r \leftarrow r + b$
        end
        else if $|\dot{Z}_{c_t}| >= \dot{Z}_{\max}$ or $|I_{IC_t}| >= I_{IC_{\max}}$ then
            $r \leftarrow r - 10 * b$
            terminate the episode
        end
        Update the episode cumulative reward $G \leftarrow G + \Gamma^t r$
        Update $\epsilon \leftarrow \delta\epsilon$
    end
end
```

Algorithm 1: Q-learning algorithm for the training of the RL-based VS system of the EAST tokamak.

(see **Step 5**). Each training episode simulates the reaction to a VDE over a time interval of 400 ms, with a time step of 1 ms. For each episode, the initial state of the plasma linearized model is randomly chosen along the unstable eigenvector of the A matrix to result in a VDE with a possibly maximum amplitude of 5 cm (see **Step 4**).

Since the purpose of the VS system is not only to bring the vertical velocity of the plasma to zero but also to keep it close to zero throughout the entire training episode, a positive bonus is added to the current step reward $R(S_t, A_t)$ every time the agent manages to keep the state within a prescribed bound specified by the thresholds reported in Table 7.1 (see **Step 14**). On the other hand, an episode is terminated, and a penalty is assigned to the reward when at least one of the two states reaches the maximum allowed value (see **Step 17**).

Both the cumulative reward G and the exploration parameter ϵ are updated after each time step (see **Steps 21** and **22**).

The VS system used in the simulations has been trained using the model (1.1) of the equilibrium corresponding to the EAST pulse #78289 at the time instant $t = 3$ s (see the plasma equilibrium parameters in Table 7.2). The training lasted $\approx$ 9000 episodes, after which the cumulative reward settled down to a "regime" value, as shown in Figure 7.1, suggesting that at least a locally optimal policy had been learned. The result is the action-value function Q, represented as a table that associates the expected cumulative reward for every state $(\dot{Z}_c, I_{IC})$ and action V_{IC}. The obtained RL-based VS system has been then validated on a set of different equilibria, by including also the other components of the EAST plasma MCS (for details on the simulation environment, see [114]).

7.1 Agent validation

To show the effectiveness of the proposed RL-based VS approach, two EAST plasma models different from the one used for training have been taken into account. Plasma parameters of the validation equilibria are reported in Table 7.2. In particular, for the case of the EAST pulse #92141, the response of the VS to the experimental disturbance that corresponds to the injection of 1 MW of lower hybrid [115] is presented. Subsequently, the rejection of a 2 cm VDE applied during the plasma current flat-top of

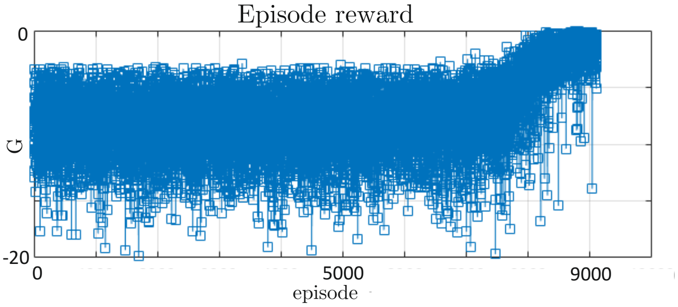


Figure 7.1: Cumulative reward G over the training of the VS agent.

EAST equilibrium	$\beta_{p_{eq}}$	$l_{i_{eq}}$	Growth rate γ
Pulse #78289 at t=3 s (training)	0.39	1.06	224 s^{-1}
Pulse #92141 at t=3 s (validation)	0.17	1.49	156 s^{-1}
Pulse #79289 at t=3 s (validation)	0.33	1.19	177 s^{-1}

Table 7.2: Values of the plasma parameters for the equilibria used for both the training and validation of the proposed RL-based VS system. For all the considered cases $I_{p_{eq}} = 250\ kA$.

the EAST pulse #79289 is considered.

It should be noted that the VS agent has been invoked at a frequency of 10 kHz, which is consistent with the sampling frequency of the EAST plamsa MCS.

Minor disruption Figure 7.3 shows the time traces of the disturbance parameters β_p and l_i that corresponds to the injection of $\approx 1\ MW$ of lower hybrid into the plasma during pulse #92141 at $t \approx 4\ s$. The simulation results are shown in Figure 7.2 as red traces and are compared with the one obtained during the experiment (blue traces). In particular, the voltage and current in the in-vessel coil V_{IC} and I_{IC}, respectively, the plasma vertical speed $\dot{Z}_c$, as well as the plasma current I_p are reported. It can be seen that despite a slight increase in terms of control effort (i.e., of both the voltage and the current in the IC circuit), the plasma vertical motion is stabilized and the behavior of the plasma current is improved. Furthermore, comparing the behavior of the differences in poloidal flux at the control points shown in Figure 7.4, it can be seen that a better decoupling between the VS and the shape control is achieved with the RL agent. Indeed, when the latter is used, no oscillations arise when the disturbance occurs.

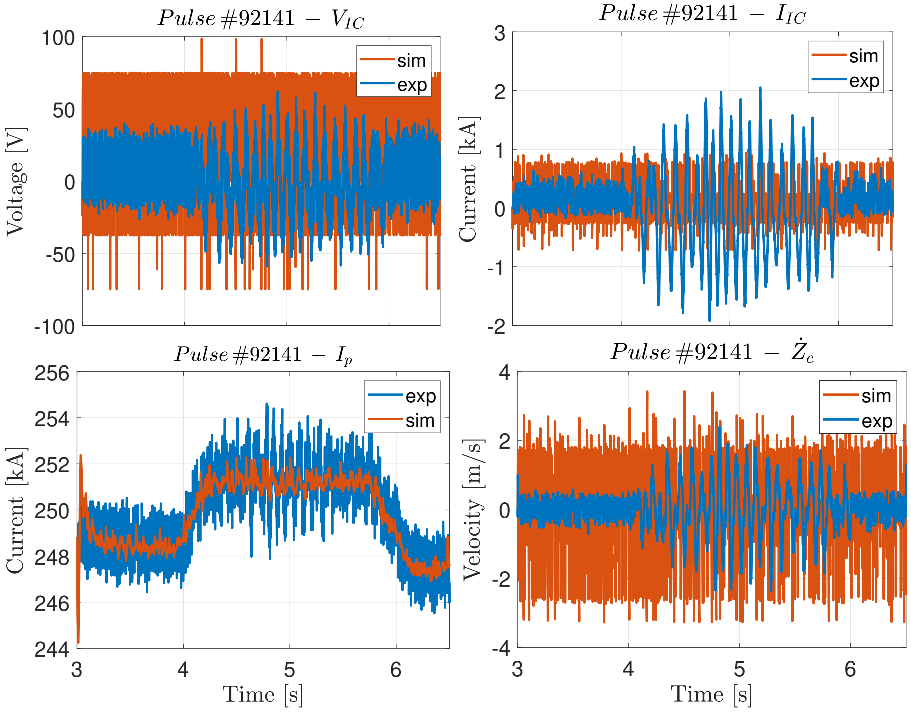


Figure 7.2: Rejection of external disturbances due to lower hybrid power injection. The control voltage V_{IC}, the corresponding current I_{IC}, the plasma vertical speed $\dot{Z}_c$ and current I_p are shown. The simulated results (red traces) are compared to the ones obtained during the EAST pulse #92141.

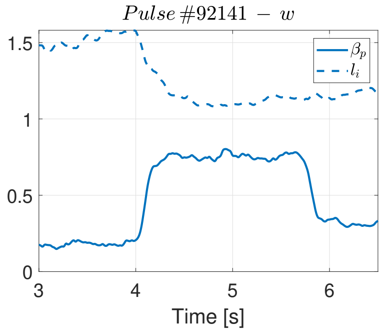


Figure 7.3: Time traces of the disturbances β_p and l_i for pulse #92141.

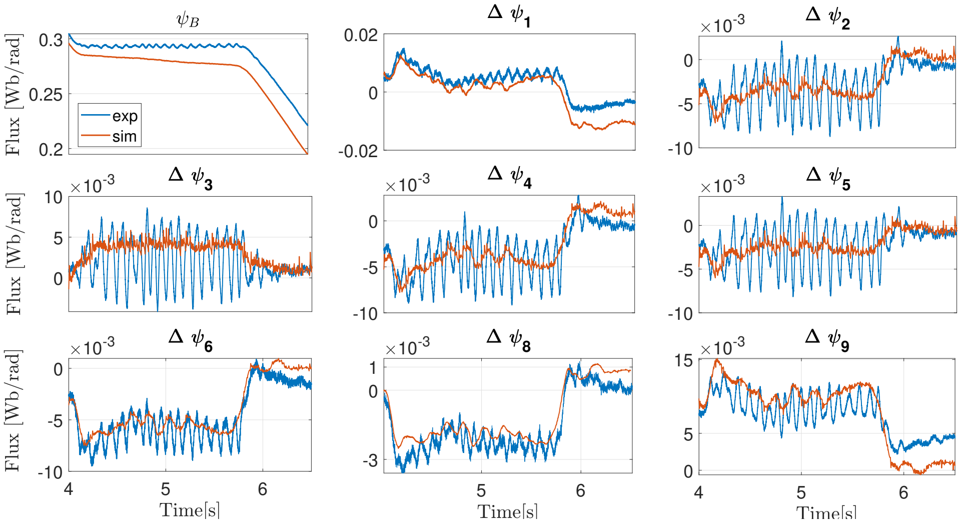


Figure 7.4: Poloidal flux error at the control points for pulse #92141. The comparison between the simulation result obtained with the proposed RL based VS system (red traces) and the experimental data (blue traces) is shown.

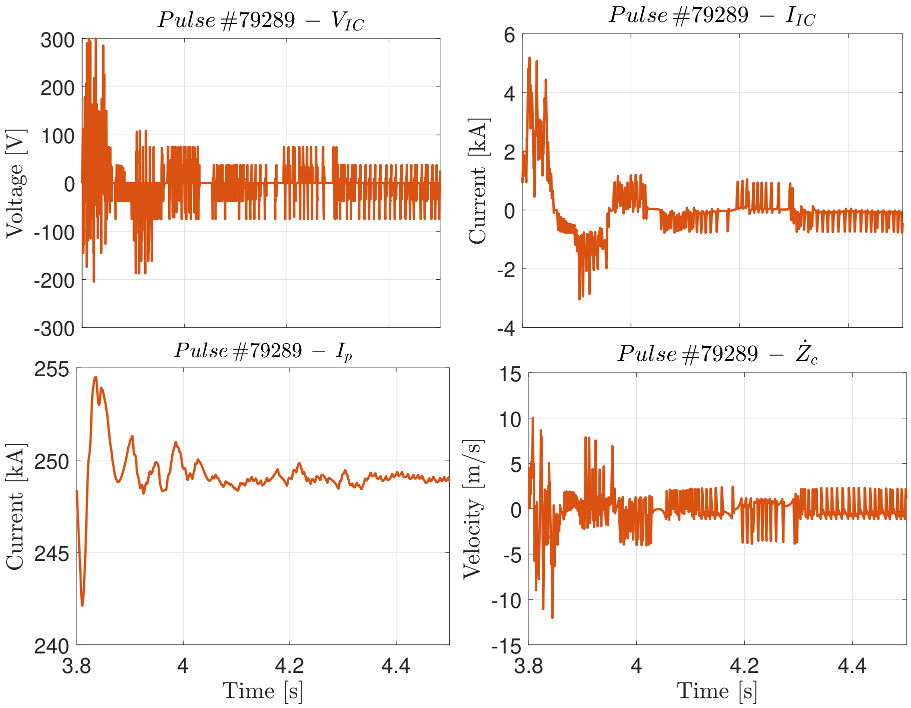


Figure 7.5: Response to a 2 *cm* VDE for pulse #79289. The control voltage V_{IC}, the corresponding current I_{IC}, the plasma vertical speed $\dot{Z}_c$ and current I_p are shown.

VDE rejection The response of the overall EAST plasma MCS to a VDE of 2 *cm* for pulse #79289 is shown in Figure 7.5. In this case, the proposed RL-based VS system proves to be able to stop the plasma vertical motion, recovering from the initial VDE. Also in this case the interaction of the RL-based VS with the overall EAST plasma MCS does not affect the performance of the plasma shape controller. Figure 7.6 shows two snapshots of the plasma cross-section during the simulation.

As a final remark of this section, it is worth noting that, although a single equilibrium with a growth rate of ≈ 220 s^{-1} has been used for the training, the VS agent has proved to be robust enough to cope with plasma with different growth rates (≈ 155 s^{-1} and ≈ 180 s^{-1}) and scenarios not

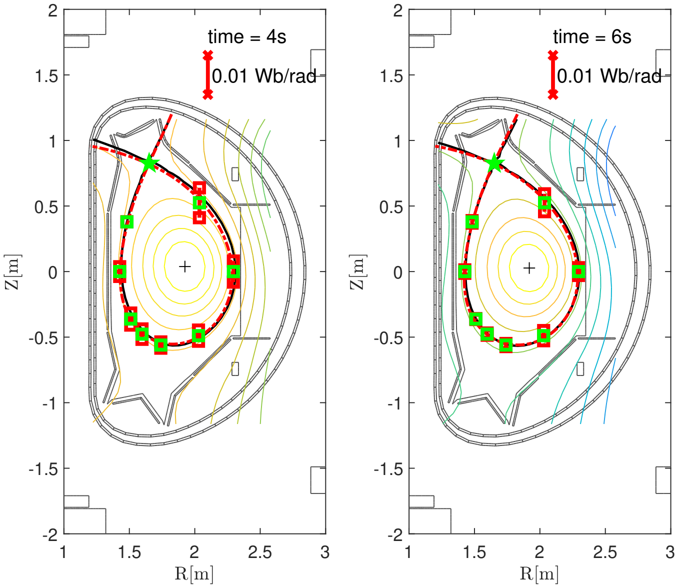


Figure 7.6: Plasma boundary shapes for the case of a VDE applied to pulse #79289. The red curve shows the desired plasma boundary, while the black one is the simulated one. Two snapshots at $t_1 = 4\ s$ and $t_2 = 6\ s$ are shown. The bars show the poloidal flux error at the control points.

considered during training.

8 Deep Reinforcement Learning based Vertical Stabilization for the ITER tokamak

In this chapter, the applicability of DDPG approach to the VS problem is analyzed. The effectiveness of the proposed solution is shown in the test case of the ITER tokamak. In addition to what has been proposed in the previous section the whole ITER plasma MCS has been taken into account during the training process and a sensitivity analysis has been performed that allowed the choice of the best-performing set of training DDPG hyperparameters.

A scheme showing how the DDPG agent has been applied to the ITER VS is shown in Figure 8.1. The agent interacts with a RL environment consisting of a linear model representing the plasma dynamics and power supply, and includes the interaction of the VS system with the other magnetic control loops, i.e. the plasma current and shape con-

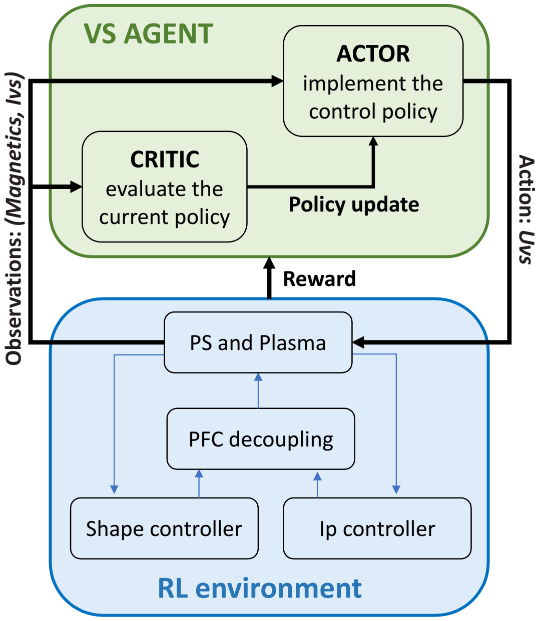


Figure 8.1: DDPG scheme of interaction between the RL environment of plasma magnetic control and the actor-critic VS agent.

trollers.

The actuator considered for the VS agent is the in-vessel circuit, so for the VS agent the control action is chosen as $a = u_{VS}$. Feedback signals are organized in the observed vector $s = (I_{VS}\,, y_{mag})$, and are, respectively, the current in the VS circuit I_{VS}, and the vector of magnetic diagnostic signals y_{mag}. The latter signals are part of the output of the linearized plasma model (1.1) and are usually used to reconstruct the plasma centroid vertical position Z_c and velocity $\dot{Z}_c$. Specifically, the in-vessel Mirnov coil measurements and flux sensors have been included in y_{mag}.

The simulation scheme used in this part of the thesis work as a training and validation environment for the proposed VS agent is shown in Figure 8.2.

The actor and critic neural networks are feed-forward networks with fully connected layers and have been implemented as reported in Fig-

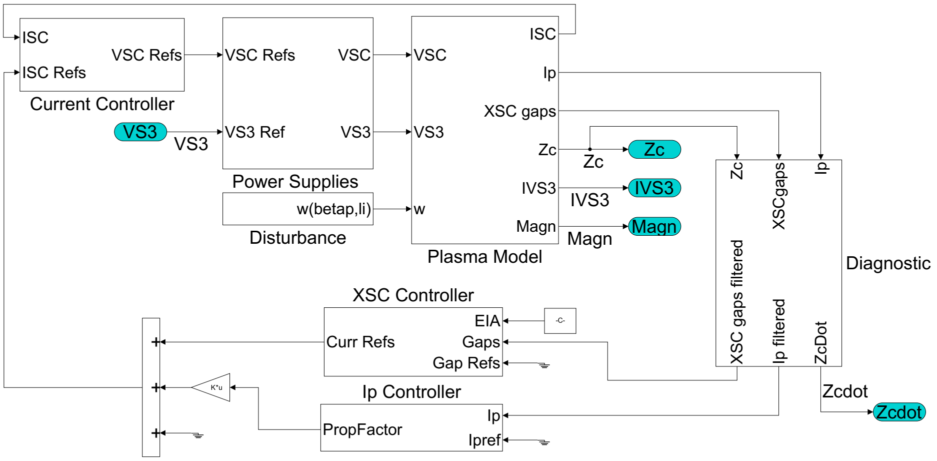


Figure 8.2: Simulation scheme used as environment for both training and validation of DDPG-based VS agent. The scheme includes the plasma-coils model, the model of the power supplies, and the controllers for both plasma current and shape.

ure 8.3. ReLU activation functions have been chosen, defined as $ReLU(x) = \max(x, 0)$.

The reward function has been chosen as a function of the agent state s and action a, and is given by:

$$R(s\,,a) = -k_1 \cdot \left(\frac{\dot{Z}_c(y_{mag})}{\dot{Z}_{c_{\max}}}\right)^2 - k_2 \cdot \left(\frac{I_{VS}}{I_{VS_{\max}}}\right)^2 - k_3 \cdot \left(\frac{u_{VS}}{u_{VS_{\max}}}\right)^2 , \quad (8.1)$$

where $\dot{Z}_{c_{\max}}$, $I_{VS_{\max}}$ and $u_{VS_{\max}}$ refer to the maximum values specified for the vertical speed of the plasma centroid and the current and voltage of the vessel coils, respectively.

The reward function (8.1) reflects the main objective of a VS system i.e. to stop the unstable vertical motion of the plasma to avoid disruption while keeping the in-vessel current as low as possible and limiting the control voltage. An additional penalty is then added to the reward (8.1) and the training episode is terminated if the centroid position variation

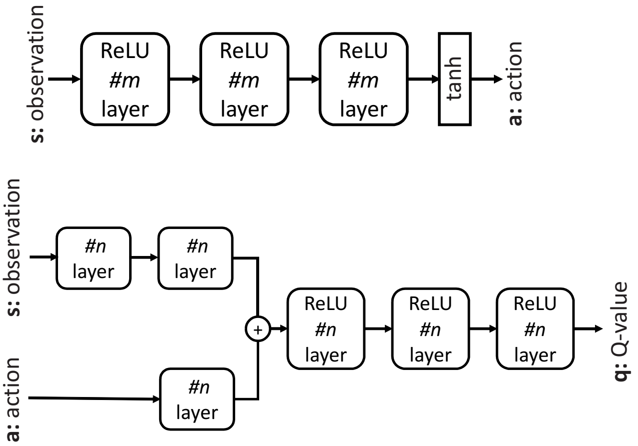


Figure 8.3: DDPG Actor and Critic networks architecture. The networks are only feed-forward, with no recurrent element, and have been implemented using fully connected layers.

with respect to the equilibrium value δZ_c exceeds a threshold beyond which disruption cannot be avoided. Furthermore, a +2 bonus is added to (8.1) at each simulation time step if the agent manages to keep δZ_c within the prescribed bound. These bonuses, summed over all the episode time samples, turn into a maximum value for the cumulative reward that becomes positive. In particular, given the sampling time $T_s = 2.5\ ms$, the possible maximum cumulative reward could be 4000 for the episodes whose duration is equal to 5 s (see Figures 8.4, 8.6 and 8.7), while if the episode duration is 20 s the maximum could reach 16000 (see Figure 8.5).

Moreover, at each time step of the DDPG training process the expected reward y_i is computed as

$$y_i = R_i + \Gamma Q(s_{i+1}, a_{i+1}), \tag{8.2}$$

where R_i is the experienced reward at the i-th step, Γ is the discount factor and $Q(s_{i+1}, a_{i+1})$ is the action-value function predicted by the critic network.

The described setup has been implemented in MATLAB® by using the

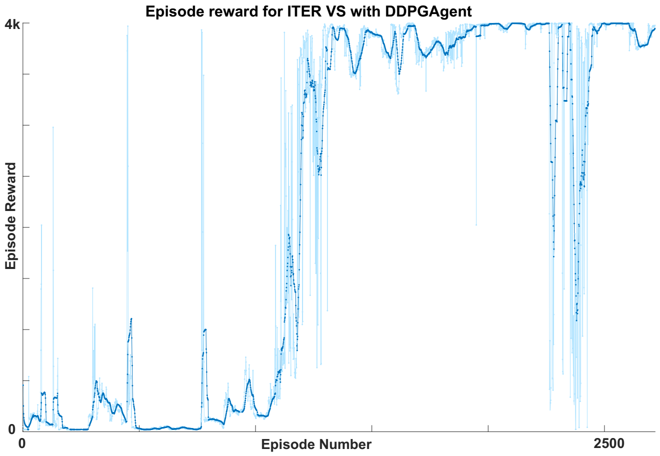


Figure 8.4: Episodes cumulative rewards obtained with the best choice of the considered DDPG hyper-parameters and for the reward function coefficients set equal to $k_1 = 1\,, k_2 = 2\,, k_3 \;=\; 1$ (which also correspond to the setup considered for *Training A* in Section 8.2).

Reinforcement Learning Toolbox® [116], to take advantage of Simulink® to integrate the VS agent with the other components of the ITER plasma magnetic control, already available in this environment.

8.1 Sensitivity analysis for the DDPG training hyper-parameters

The effects of some hyperparameters and their tuning are analyzed with respect to reward convergence. This study allowed to find the set of parameters that led to the successful training of the VS agent. Specifically, it has been considered the effect of the following hyperparameters: episode duration, number of hidden layers for both the actor and critic networks, and action-noise variance decay rate. Table 8.1 reports the setting of all the DDPG hyper-parameters and the range of variation for those that have been changed during our analysis.

Figure 8.4 reports the *training graph*, i.e. the trace of the cumulative reward as a function of the *i*-th episode, when the coefficients in (8.1) are

Hyper-parameter	**Considered Values**	
Sampling time T_s	$2.5\,ms$	
Episode duration T	**5 s**	$20\,s$
Actor learning rate	5×10^{-4}	
Critic learning rate	10^{-3}	
Actor hidden layers $\#m$	**64**	128
Critic hidden layers $\#n$	**32**	128
Discount factor Γ	0.99	
Batch size	256	
OUP variance	1840	
OUP variance decay rate	$\mathbf{8.66 \times 10^{-6}}$	3.5×10^{-6}

Table 8.1: Set of the DDPG hyper-parameters. The range of variations exploited during the sensitivity analysis is specified for those parameters whose setting was changed. When multiple values are specified, those reported in **boldface** are the ones chosen to obtain the results reported in Figure 8.4.

set equal to $k_1 = 1\,, k_2 = 2\,, k_3 \;=\; 1$, and the *optimal choice* for the three considered hyperparameters has been made. In particular, the latter has been set equal to the values reported in bold in Table 8.1.

In the following, for each hyperparameter variation considered, the corresponding training graph is reported and a brief discussion is made to motivate the final choice. Notice that in the training graph not only the time trace of the cumulative reward as a function of the i-th episode (lighter trace) is reported but also the average cumulative reward (darker trace) over the 20 most recent episodes.

Episode duration Initially, the duration of an episode has been set equal to $20\,s$; the corresponding training is shown in Figure 8.5. When comparing this training with the one shown in Figure 8.4, it can be seen that a longer interaction between the DDPG agent and the plasma envi-

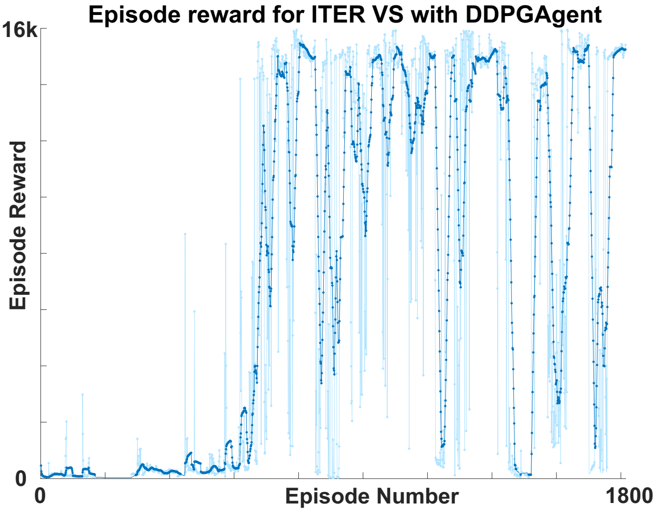


Figure 8.5: DDPG episodes cumulative rewards obtained with an episode duration of 20 s

ronment leads to higher rewards, but does not ensure convergence toward an optimum. Furthermore, when the duration of the episode is set equal to 20 s, the obtained agents focus more on satisfying the performance in steady state, rather than during the transient. Therefore, an episode duration of 5 s has been chosen for the agent training procedure.

OUP Variance Decay Rate The agent uses the Ornstein-Uhlenbeck action noise model for exploration. The noise variance and its decay rate are computed as

$$\sigma^2 \cdot \sqrt{T_s} = (1\% \, to \, 10\%) \, of \, \Delta A$$

while its half-life, in time steps, is given by

$$HL = \frac{\ln{(0.5)}}{\ln(1-\sigma^2_{dr})}$$

where σ^2 is the noise variance, ΔA is the range of the action variable, HL is the half-life of the noise and σ^2_{dr} is the decay rate of the variance

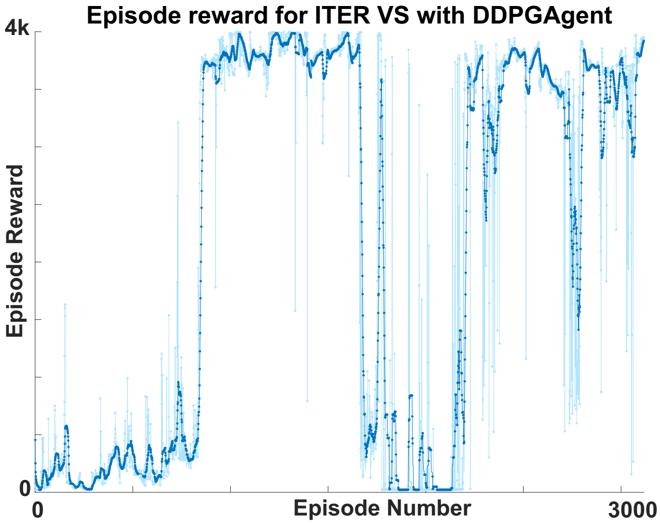


Figure 8.6: DDPG episodes cumulative rewards obtained with an agent noise decay rate of 3.5×10^{-6}.

of the noise. In this analysis, two values of the decay rate have been considered. This parameter has been first set equal to 3.5×10^{-6}, which is equivalent to about 2×10^5 time samples. The resulting training, reported in Figure 8.6, shows that even if exploration seems to terminate after about 1000 episodes, there is a sudden drop in the reward value between episodes 1700 and 2200, after which the agent does not fully recover. On the other hand, when the decay rate is set equal to 8.66×10^{-6}, corresponding to a variance half-life of about 8×10^4 time samples, the results shown in Figure 8.4 are obtained. In this case, once the plateau is reached, the behavior of the reward oscillates less up to episode 2500. Therefore, in our setup, we set the decay rate equal to 8.66×10^{-6}.

Critic and Actor hidden layers size The first choice of size for fully connected layers in the architecture of the critic and actor networks was equal to 128, as shown in Figure 8.3. From the training reported in Figure 8.7, it appears that network architectures can significantly affect results and convergence and that considering a simpler network can produce better results.

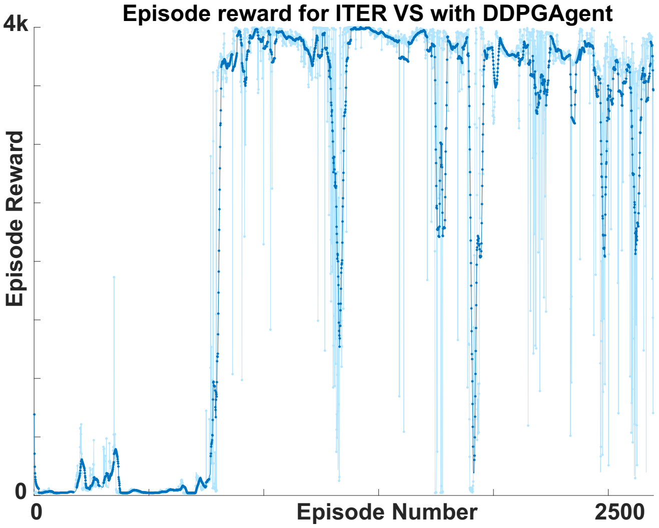


Figure 8.7: DDPG episodes cumulative rewards obtained with, both actor and critic networks implemented using fully connected layers with a size 128.

Therefore, in the training reported in Figure 8.4, 64 layers were chosen for the actor, and 32 for the critic.

8.2 Agents validation

For agent validation, two equilibria different from those used during training have been taken into account. These two validation equilibria correspond to*Equilibrium #2* and *Equilibrium #3* from Group B of ITER plasma linear models previously introduced and whose nominal values are reported in Table 3.1b. While *Equilibrium #1* has been used for training.

In addition to an agent obtained from the training shown in Figure 8.4 (hereafter referred to as *Training A*), two more agents have been selected; these correspond to two choices of the reward function aimed at improving the VS performance, and, in particular, reducing the steady state current in the in-vessel coils. Namely, *Training B*, corresponds to a higher value of k_2, while *Training C* was obtained by adding the term $-k_4 \cdot \left(\frac{\bar{I}_{VS}}{\bar{I}_{VS_{\max}}}\right)^2$

	$\mathbf{k_1}$	$\mathbf{k_2}$	$\mathbf{k_3}$	$\mathbf{k_4}$
Training A	1	2	1	0
Training B	1	2000	1	0
Training C	1	10	1	20

Table 8.2: Values of the penalty parameters in the reward function (8.1) for the different considered DDPG trainings.

to the reward function. This additional term allows penalizing also the integral value $\overline{I}_{VS} = \frac{1}{T}\int_0^T I_{VS}(\tau)d\tau$ of the current in the in-vessel coils. The parameters of the reward function for the trainings considered are summarized in Table 8.2. A preliminary performance assessment of the three agents considered is reported in the following.

Validation without disturbances The model corresponding to *Equilibrium #2* is used to assess the ability of the various agents to stabilize a plasma different from that used for training when no external disturbances are applied. At the beginning of the considered simulation, the plasma starts from the considered equilibrium. The RL-based agent, however, can lead to small initial displacements in the plasma position that must be actively compensated for. Figure 8.8 shows the in-vessel circuit voltage u_{VS} and current I_{VS}, the plasma current I_p, and the variation in the vertical position of the plasma centroid δZ_c for the training options considered. It can be seen that although all agents considered achieve the stabilization objective, the ones corresponding to *Training B* and *Training C* are preferable. They require a lower steady-state in-vessel coil current since they bring the centroid closer to its equilibrium value.

Validation in the presence of a VDE Further validation is performed by applying a VDE of 5 *cm* to *Equilibrium #3* and Figure 8.9 shows the comparison between the results obtained using the three considered DDPG VS agents. Also in this case the agents corresponding to *Training B* and *Training C* allow minimizing the steady-state current in the in-vessel coil, confirming the results obtained with *Equilibrium #2*. Moreover, the agent corresponding to *Training C* shows a smoother be-

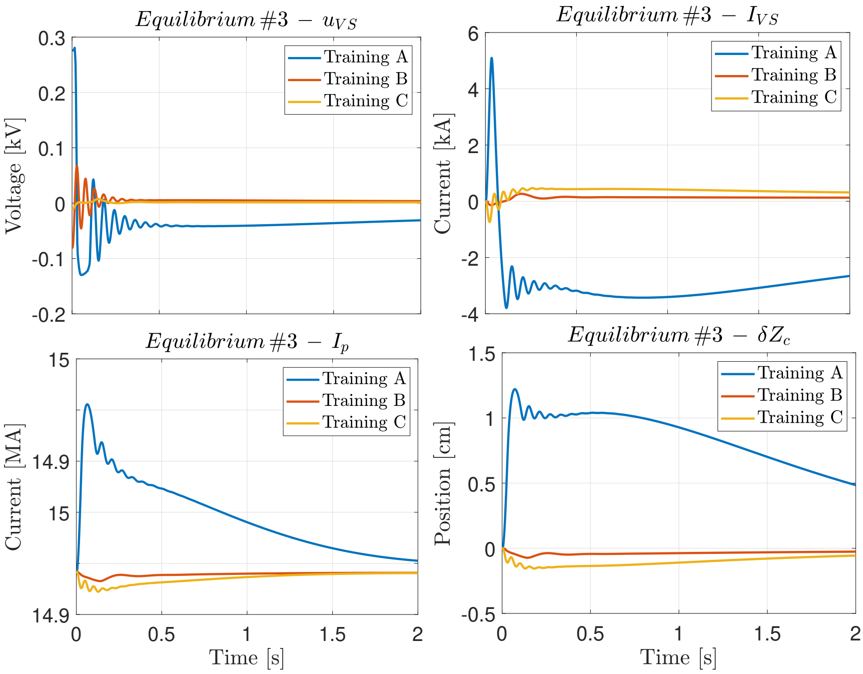


Figure 8.8: Simulation results obtained with *Equilibrium #2* for the DDPG VS agents corresponding to *Training A* (blue trace), *Training B* (blue trace) and *Training C* (red trace). The u_{VS} voltage, the I_{VS} current, the plasma current I_p, and the vertical displacement of the plasma centroid δZ_c with respect to the equilibrium are shown.

havior in terms of variations in plasma current and vertical displacement.

Comparison with a linear VS A comparison between a model-based linear VS algorithm and the validated VS agents is reported in Figure 8.10. The former controller is an ITER-like VS system and computes the voltage u_{VS} to be applied to the in-vessel coils as a combination of the plasma vertical speed $\dot{Z}_c$ and of the current I_{VS} flowing in the VS circuit. The interested reader can refer to [27] and [34] for more details. To compare the two approaches considered, here we report the results of the simulation of a 5 *cm* VDE applied to *Equilibrium #3*. For this case considered, Figure 8.10 shows that all the RL agents have a performance similar to

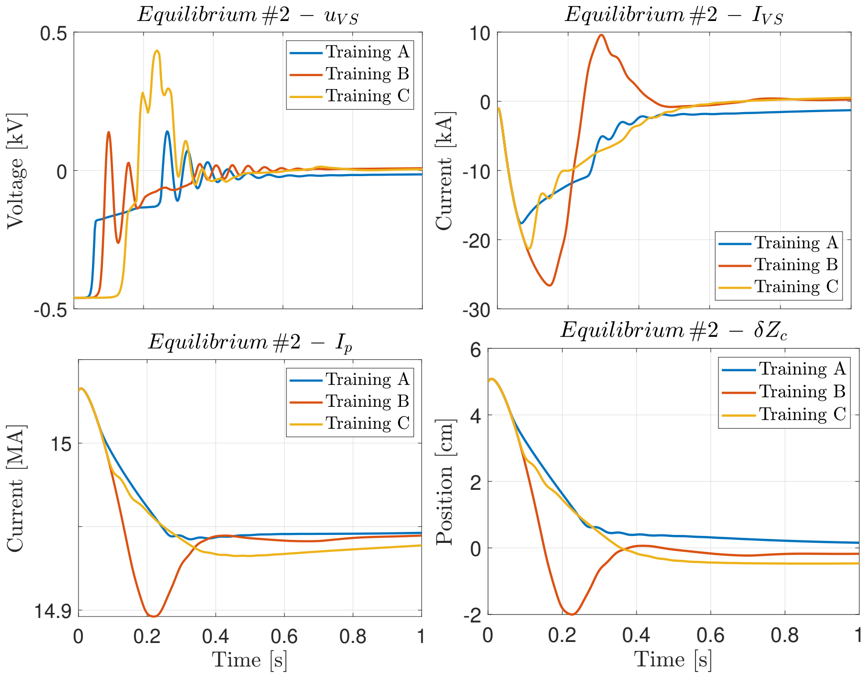


Figure 8.9: Rejection of a 5 *cm* VDE applied to the linear model corresponding to *Equilibrium #2*. The time traces of the u_{VS} voltage, the I_{VS} current, the plasma current I_p, and the vertical displacement of the plasma centroid δZ_c with respect to the equilibrium are shown.

the model-based VS in terms of setting time. Moreover, RL approaches require a lower control effort in terms of applied voltage, during the initial phase of the simulation, when the plasma displacement with respect to the equilibrium value is maximum.

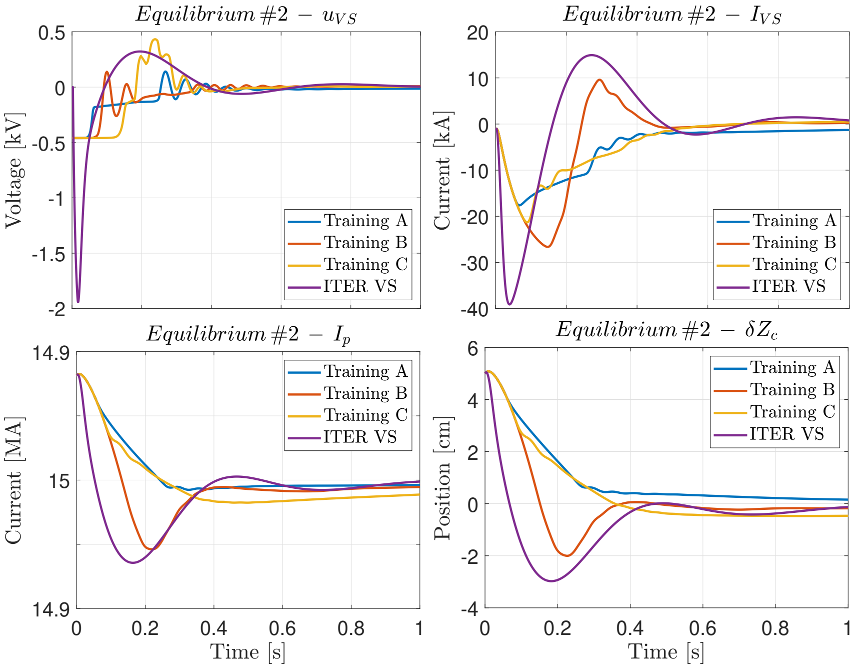


Figure 8.10: Comparison between the model-based linear VS algorithm (purple trace) and the RL agents corresponding to *Training A* (blue trace), *Training B* (blue trace), *Training C* (red trace) in case of rejection of a VDE of 5 *cm* applied to *Equilibrium #2*.

9 Conclusions and future activities

In this doctoral the current state of tokamak technology and especially the challenges and limitations of plasma confinement when it comes to plasma vertical stabilization are presented and questioned. The work done in this thesis addresses the limitations introduced by commonly used model-based VS systems by developing control strategies that guarantee the required level of performance without relying on the knowledge of a plant model. Indeed, model-free and data-driven approaches to the problem of VS in tokamak plasma were pursued. With model-free approaches, it is possible to rely on the controller agnosticism with respect to the plant model to increase robustness, while the data-driven ones learn the desired plant behavior via extensive simulation campaigns and/or access to large experimental data sets.

The ES algorithm for stabilization has been applied to the VS system of the ITER tokamak and validated on a *palette* of linearized equilibria showing that even if always employing the same Kalman filter, the model-agnostic nature of the ES algorithm allows to cope with large model uncertainties. The simulation results show that the proposed VS scheme achieves a satisfactory level of robustness during the overall flat-top phase of an ITER discharge and for different plasma parameters and configurations. Indeed, by means of linear and nonlinear simulations, it was proved that the proposed control architecture can practically stabilize the plasma column, by keeping the system state in a bounded set, while counteracting relevant plasma disturbances.

The introduction of a switching power supply enhanced with an adaptation logic for the bounded ES control gain yields an improvement in overall performance. It was possible to minimize the oscillation induced in the controlled variables by the ES approach and reduce the maximum voltage in the VS circuit bringing the control effort closer to the value currently envisaged for the ITER tokamak.

The main advantage of the proposed ES-based technique lies in its rather easy adaptation to different plasma configurations. Indeed, this usually requires low or no effort, provided that the considered observer is capable of describing, at least roughly, the unstable dynamic of the plant and that controller gains need to be suitably chosen just once. This is not the case for standard model-based VS techniques, which usually need to be tuned on the basis of the specific plasma configuration, a task that requires some significant modeling and testing effort.

This latter characteristic of ES approach made the proposed VS a suitable candidate for the WPTE (Work Package Tokamak Exploitation) at TCV tokamak. A preliminary evaluation for applying the ES control technique has already been carried out, drawn also by the switching power supply employed for the VS system of TCV. However, during this Ph.D. there was no possibility to test the proposed approach during experiments, but there are plans to validate it in an experimental campaign in 2024.

To remove the residual model dependence embedded in the Kalman filter, NNs have been trained to estimate the movement of the plasma along the unstable dynamic, This allowed to turn the proposed ES-based VS in a completely model-free control approach. A LR, a MLP, an ELM, and a LSTM networks were trained on synthetic shots performing ITER discharges and their generalization property has been tested in closed-loop considering scenarios not seen during training. It was shown that the use of NNs can enhance the operational space and generalization property of the ES control law, ensuring more robustness to model uncertainties or changes in the plasma configuration and behavior. This makes it possible to stabilize plasma equilibria that are not stabilized by the set-up based on a single Kalman filter.

Nevertheless, the replacement of the Kalman filter and the introduction of NNs in the VS control loop is not a trivial extension of the previous method in [55, 56]. As reflected by the results presented, the Kalman filter generalizes less than the NNs considered but yet it can be easily tuned on the basis of the controller experience. In contrast, NNs significantly increase the generalizability of the controller, but their training requires more effort. In fact, data needed to be collected to generate the required datasets, and a significant amount of time was required to train, validate and test the models obtained since the performance of NNs depends on the specific data seen during training.

However, the main advantage is that thanks to NNs generalization property, the proposed approach can be adapted even more easily to different plasma configurations, being the NNs, once properly trained, capable of reconstructing the unstable dynamic of the plant, while the ES controller ensures model-free adaption, without the need to tailor the gains for the specific scenario. This represents a potential improvement if compared with standard model-based VS techniques.

The presented results pave the way for further developments. For example, to increase robustness, the training database could be collected by performing synthetic shots on different scenarios and disturbances that can occur during ITER operations. The trained models can also take into account the variation of the disturbance β_p and l_i, or of shape parameters such as triangularity and elongation, which have an impact on the dynamic of the vertical instability. Moreover, the availability of operating tokamak, such as TCV, can be exploited not only to test the proposed approach in an experimental campaign but also to train the NNs on experimental data.

Finally, the possibility of applying RL to the VS problem was investigated. Both RL strategies pursued, the Q-learing and DDPG algorithms, allowed the implementation of a single VS agent which is robust enough to handle different plasma operating conditions without the need to adapt the controller parameters. The proposed VS agents were also shown to have similar or even improved performance compared to other classical controllers.

Also in this case, it is possible to further improve the robustness of the VS agent by enlarging the operational space seen during its training. Indeed, training could be carried out considering a palette of different plasma equilibria and configurations so that the final agent can also adapt to the value of the disturbance and shape parameters. Specifically, these parameters could be included in the observed quantities provided to the agent from RL environment.

A
Reinforcement Learning

THIS CHAPTER provides a more comprehensive introduction to the RL framework. It delves into its key features and explains the iterative optimization process that is the basis of all RL algorithms.

RL is one of the three main paradigms of ML, alongside with supervised and unsupervised learning. Unlike the other two, RL is used to deal with sequential decision-making problems, in which the action to be taken is contingent on the current state of the system and has an effect on its future.

The learner or decision maker in RL problems is referred to as the agent and interacts with what is called the environment. This interaction is continuous: the agent selects an action, and the environment responds to this action by altering its state and providing a scalar reward. The agent is not given instructions beforehand about which actions to take but instead must discover which actions result in the highest long-term cumulative reward by trying them. Through a trial-and-error process, it decides what to do based on the feedback it receives from the environment. Moreover, since the aim ofRL is to maximize the cumulated reward, i.e., the sum of the rewards over a period of time, the choice of actions affects not only the immediate reward but also all future ones. Trial-

and-error searches together with delayed rewards are the most important distinguishing features of RL.

In addition to the agent and the environment, four main components are necessary for the successful implementation of RL algorithms: a *policy*, a *reward signal*, a *value function*, and a *model* of the environment. These four elements are essential for the successful application of RL techniques. Specifically:

- A *policy* $\pi(s)$ is a way of determining how an agent will act in a given situation. It is a mapping from states to actions that the agent follows to make decisions. The policy can be either deterministic, meaning that it always chooses the same action for a given state, or stochastic, meaning that the agent chooses actions based on probability. Agents try to find the optimal policy that leads to the highest cumulative reward.

- The *reward signal* is a scalar value that provides the agent with feedback on the desirability of its actions. It is usually represented as $R(s_t, a_t)$, where t is the current time step, and s_t and a_t are the state and action at that time. The programmer designs the function $R(s_t, a_t)$ based on the goal of the problem to be solved and uses it to guide the agent's learning; by giving higher rewards to desirable actions and lower rewards to undesirable actions, the agent learns to prefer actions that lead to higher cumulative rewards.

- The *value function* evaluates the desirability of different states or actions in the long run. It estimates the expected cumulative reward an agent will receive from a particular state or state-action pair and is used to guide the agent's decision-making process. The value function can be expressed in two forms: the *state-value function* $V(s)$ and the *action-value function* $Q(s, a)$. The state-value function estimates the expected cumulative reward from a given state, while the action-value function estimates the expected cumulative reward from a particular state and when a specific action is taken.

- The *model* of the environment is a representation or approximation of the way the environment works, which the agent uses to make decisions. This model can predict the next state and reward given

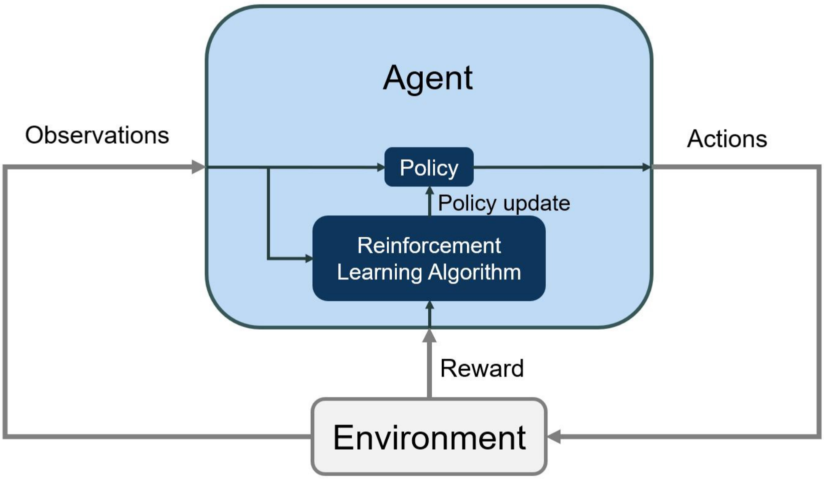


Figure A.1: Agent-environment interaction in the RL framework.

a certain state and action. It allows the agent to simulate the environment and make decisions without actually interacting with it. However, it is important to note that not all RL algorithms require a model. Model-free algorithms learn directly from the interaction with the real environment, without needing an explicit representation of its dynamics.

An RL algorithm is a method to formulate an optimization problem in which the agent's control objective is expressed in terms of a scalar reward function $R(s, a)$. The aim of RL agents is to identify the optimal behavior policy $\pi^*(s)$ for the given problem by maximizing the total reward received during the interaction with the environment. In the optimization process, the RL algorithms usually refer to the discounted cumulated reward, which is defined as:

$$G_t = \sum_{t=0}^{N} \Gamma^t R(s_t, a_t) \, , \tag{A.1}$$

with $\Gamma \in [0, 1)$ the so-called *discount factor* that determines the importance of future rewards with respect to the immediate one. The discount-cumulated reward, G_t allows RL algorithms to handle non-deterministic environments since it can account for future uncertainty. The agent can learn the optimal policy by making decisions based on a forecast of the total reward, which is obtained by using value functions. In RL, this process is referred to as the Bellman Optimality Principle and can be expressed in the Bellman equations:

$$V_\pi(s) = \mathbb{E}_\pi[R(S_t, a) + \Gamma V_\pi(S_{t+1})|S_t = s]\,, \tag{A.2a}$$

$$Q_\pi(s, a) = \mathbb{E}_\pi[R(S_t, a) + \Gamma Q_\pi(S_{t+1}, A_{t+1})|S_t = s, A_t = a]\,. \tag{A.2b}$$

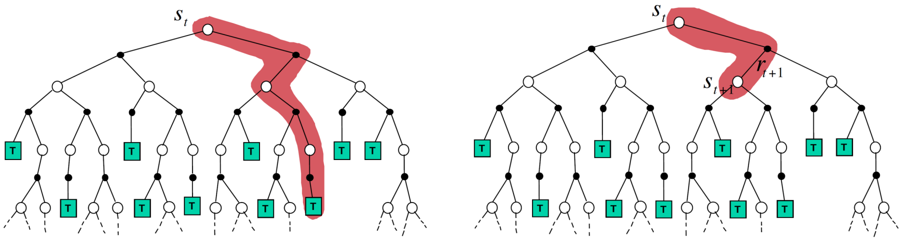
These equations express the expected value of a state when an agent takes an action prescribed by a policy π. The value of a state S_t is calculated by adding the immediate reward $R(S_t, a)$ to the discounted expected value of the subsequent state S_{t+1}.

The Bellman Optimality Principle states that the optimal course of action in any given situation can be decomposed into two parts: the most advantageous action to take in the current state, which yields $R(S_t, a)$, and the optimal policy to follow from the subsequent state given by $v_\pi(S_{t+1})$ (or $q_\pi(S_{t+1}, A_{t+1})$). This enables the Bellman equation to be used to progressively refine the value function. The update process, which may differ depending on the specific RL algorithm employed, generally follows the principle of updating towards a target value. For instance, for the Monte-Carlo algorithm, the value function is updated toward the actual cumulated return G_t:

$$V(S_t) \leftarrow V(S_t) + \theta\left(G_t - V(S_t)\right)\,, \tag{A.3}$$

where θ is the learning rate and is used to determine the importance given to new information and how much the value function is updated at each step. Indeed, higher learning rates lead to faster convergence but potentially less stability, while lower learning rates lead to slower convergence but potentially more stability.

The Monte Carlo approach utilizes complete episodes of the agent's interaction with the environment to generate samples and modify the agent's strategy (see Figure A.2a for the related backup diagram). Conversely, algorithms, like TD, modify the value function at each state-action transition (see Figure A.2 for the backup diagram). For TD, the update equation



(a) Monte Carlo

(b) Temporal Difference

Figure A.2: Backup diagram for the value function update process of Monte Carlo (a) and TD learning (b) algorithms.

for the value function is:

$$V(S_t) \leftarrow V(S_t) + \theta \left(R_t + \Gamma V(s_{t+1}) - V(S_t)\right) . \tag{A.4}$$

The target value is the sum of the current reward and the estimated value of the next state and the update is based on the so called TD errors: the difference between the observed value of a state and the estimated value of that state.

The Bellman Optimality Principle is often employed in conjunction with a greedy policy. After each update of the value function, the agent will always select the action that leads to the state with the highest estimated value. Therefore, the optimal policy is obtained by maximizing the (expected) state-action value function Q over all the possible actions, such as

$$\pi^*(s) = \underset{a \in \mathcal{A}}{\operatorname{argmax}} \; Q(s,a) . \tag{A.5}$$

In this way, the agent is always exploiting current knowledge and making the most advantageous decisions based on it. However, relying only on a greedy policy can lead to suboptimal solutions, as the agent may not consider other possible actions that could potentially bring higher rewards in the long run. To continue exploring the environment, it is possible to occasionally allow for random exploratory moves. This creates a trade-off between environmental exploration, with the aim of collecting more information to improve $\pi(s)$. and the exploitation of the already available information. A common exploration strategy is the ϵ-greedy policy, which allows the agent to occasionally choose a random action, with a small probability ϵ, instead of always selecting the greedy one. As the iteration progresses, the value of ϵ is gradually reduced so that the probability of taking random moves instead of optimal ones decays over time.

Finally, the RL algorithms can be divided into different categories. Depending on how the agent is allowed to learn optimal behaviors, there are value-based methods and policy-based methods. Value-based methods focus on estimating the state-value function $V(s)$ or the action-value function $Q(s, a)$ to allow the agent to select the action that yields the highest expected rewards. These techniques usually involve Q-learning or SARSA to learn the optimal value function. On the other hand, policy-based methods, instead of estimating the value of each state, aim to learn the optimal policy that maps states to actions directly. These techniques usually involve REINFORCE or Actor-Critic architecture to learn the optimal policy.

Furthermore, the RL algorithms can be classified according to the way the value functions are updated. On-policy algorithms update their value functions based on the current policy being followed. This means that the agent uses the same policy to both select actions and update the value function based on the rewards received from those actions. In contrast, off-policy algorithms update their value functions based on a policy different from the one that is being followed. The agent follows one policy, known as the *behavior policy*, to select actions, but updates its value function based on the maximum expected value of the next state, regardless of the action taken. How the value function is updated has implications for exploration and exploitation. On-policy algorithms are more conservative in their exploration, as they only modify their value function based on the actions they actually take. Conversely, off-policy algorithms can explore more extensively as they update their value function based on the highest expected value, regardless of the action taken.

B
Publications

Extremum Seeking

- **S. Dubbioso** and A. Jalalvand and J. Wai and G. De Tommasi and E. Kolemen, *“Vertical Stabilization of Tokamak Plasmas via a Model-free Neural Networks-based Architecture”* submitted to **Engineering Applications of Artificial Intelligence**.

- **S. Dubbioso** and L.E. di Grazia and G. De Tommasi and M. Mattei and A. Mele and A. Pironti, *“Vertical stabilization of tokamak plasmas via extremum seeking”* in **IFAC Journal of Systems and Control**, vol. 21, pp. 2468-6018, 2022.

  DOI: 10.1016/j.ifacsc.2022.100203
  URL: `https://www.sciencedirect.com/science/article/pii/S2468601822000116`

- G. De Tommasi and **S. Dubbioso** and A. Mele and A. Pironti, *“Event-driven adaptive Vertical Stabilization in tokamaks based on a bounded Extremum Seeking algorithm”* in Proceeding of **6th IEEE Conference on Control Technology and Applications (CCTA)**, Trieste, Italy, 2022, pp. 831-836.

DOI: 10.1109/CCTA49430.2022.9966100

- G. De Tommasi and **S. Dubbioso** and A. Mele and A. Pironti, *“Stabilizing elongated plasmas using extremum seeking: the ITER tokamak case study”* in Proceeding of **29th Mediterranean Conference on Control and Automation (MED)**, Bari, Italy, 2021, pp. 472-478.

  DOI: 10.1109/MED51440.2021.9480302

Reinforcement Learning

- **S. Dubbioso** and G. De Tommasi and A. Mele and G. Tartaglione and M. Ariola and A. Pironti, *“A Deep Reinforcement Learning approach for Vertical Stabilization of tokamak plasmas”* in **Fusion Engineering and Design**, vol. 194, pp. 113725, 2023.

  DOI: 10.1016/j.fusengdes.2023.113725
  URL: https://www.sciencedirect.com/science/article/pii/S0920379623003083

- G. De Tommasi and **S. Dubbioso** and Y. Huang and Z. P. Luo and A. Mele and B. J. Xiao, *“A RL-based Vertical Stabilization System for the EAST tokamak”* in Proceeding of **American Control Conference (ACC)**, Atlanta, GA, USA, 2022, pp. 5328-5333.

  DOI: 10.23919/ACC53348.2022.9867499

Others

- D. Ottaviano and M. Cinque and G. Manduchi and S. Dubbioso, *‘Virtualization of accelerators in embedded systems for mixed-criticality: RPU exploitation for fusion diagnostics and control”*, in **Fusion Engineering and Design**, Vol. 190, pp. 113518, 2023.

  DOI: 10.1016/j.fusengdes.2023.113518.
  URL: https://www.sciencedirect.com/science/article/pii/S0920379623001023

- M. Cinque and G. De Tommasi and **S. Dubbioso** and D. Ottaviano, *“RPUGuard: Real-Time Processing Unit Virtualization for Mixed-Criticality Applications”* in Proceeding of **18th IEEE European Dependable Computing Conference (EDCC)**, Zagaroza, Spain, 2022, pp. 97-104.

  DOI: 10.1109/EDCC57035.2022.00025

- M. Cinque and G. De Tommasi and **S. Dubbioso** and D. Ottaviano, *“Virtualizing Real-Time Processing Units in Multi-Processor Systems-on-Chip”* in Proceeding of **6th IEEE International Forum on Research and Technology for Society and Industry (RTSI)**, Naples, Italy, 2021, pp. 329-333.

  DOI: 10.1109/RTSI50628.2021.9597281

Bibliography

[1] "ITER website," https://www.iter.org/.

[2] "ITER Research Plan within the Staged Approach," https://www.iter.org/doc/www/content/com/Lists/ITER%20Technical%20Reports/Attachments/9/ITER_Research_Plan_within_the_Staged_Approach_levIII_provversion.pdf, 2018.

[3] M. L., "Spectroscopic characterisation of tcv divertor towards a detached regime," Ph.D. dissertation, Lausanne, EPFL, 2023.

[4] "Fusion Plasmas - TCV Tokamak," https://www.epfl.ch/research/domains/swiss-plasma-center/research/tcv/.

[5] "ITER - FusionWiki," http://fusionwiki.ciemat.es/wiki/ITER.

[6] J. Li and Y. Wan, "The experimental advanced superconducting tokamak," *Engineering*, vol. 7, no. 11, pp. 1523–1528, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2095809921003933

[7] "European Research Roadmap to the Realisation of Fusion Energy," https://www.euro-fusion.org/fileadmin/user_upload/EUROfusion/Documents/2018_Research_roadmap_long_version_01.pdf, 2018.

[8] "A Community Plan for Fusion Energy and Discovery Plasma Sciences," https://arxiv.org/ftp/arxiv/papers/2011/2011.04806.pdf, 2020.

[9] J. Wesson and D. Campbell, *Tokamaks*. Oxford University Press, 2011.

[10] J. Wesson, “The Science of JET,” https://www.euro-fusion.org/fileadmin/user_upload/Archive/wp-content/uploads/2012/01/the-science-of-jet-2000.pdf, 2000.

[11] “First Tokamak plasma for JT-60SA,” https://fusionforenergy.europa.eu/news/first-tokamak-plasma-for-jt-60sa/, 2023.

[12] “JT-60SA is officially the most powerful Tokamak.” https://fusionforenergy.europa.eu/news/jt-60sa-is-officially-the-most-powerful-tokamak/, 2023.

[13] J. Luxon, “A design retrospective of the diii-d tokamak,” *Nuclear Fusion*, vol. 42, 2002.

[14] S. Wu, “An overview of the east project,” *Fusion Engineering and Design*, vol. 82, 2007.

[15] H. Reimerdes *et al.*, “Overview of the tcv tokamak experimental programme,” *Nuclear Fusion*, vol. 62, no. 4, p. 042018, 2022.

[16] “DEMO - EUROfusion,” https://euro-fusion.org/programme/demo/.

[17] “EAST,” http://east.ipp.ac.cn/.

[18] G. De Tommasi *et al.*, “Shape control with the extreme shape controller during plasma current ramp-up and ramp-down at the JET tokamak,” *J. Fus. Energy*, vol. 33, 2014.

[19] F. C. Schuller, “Disruptions in tokamaks,” *Plasma Physics and Controlled Fusion*, vol. 37, 1995.

[20] P. de Vries, M. Johnson *et al.*, “Survey of disruption causes at jet,” *Nuclear Fusion*, vol. 51, 2011.

[21] M. Ariola and A. Pironti, *Magnetic Control of Tokamak Plasmas*, 2$^{\text{nd}}$ ed. Springer, 2016.

[22] G. Jackson, D. Humphreys, A. Hyatt, and J. Leuer, “Control issues related to start-up of tokamaks,” *Fusion Science and Technology*, vol. 59, 2011.

[23] M. Walker and D. Humphreys, “On feedback stabilization of the tokamak plasma vertical instability,” *Automatica*, vol. 45, 2009.

[24] V. Shafranov, “Plasma equilibrium in a magnetic field,” *Reviews of Plasma Physics*, vol. 2, 1966.

[25] R. Albanese, R. Ambrosino, and M. Mattei, “CREATE-NL+: A robust control-oriented free boundary dynamic plasma equilibrium solver,” *Fus. Eng. Des.*, vol. 96–97, 2015.

[26] F. Sartori, G. De Tommasi, and F. Piccolo, “The Joint European Torus,” *IEEE Control Sys. Mag.*, vol. 26, 2006.

[27] R. Albanese, R. Ambrosino *et al.*, “Iter-like vertical stabilization system for the east tokamak,” *Nuclear Fusion*, vol. 57, 2017.

[28] R. Ambrosino *et al.*, “Design and nonlinear validation of the ITER magnetic control system,” in *2015 IEEE Conference on Control Applications (CCA)*, 2015.

[29] M. Cinque *et al.*, “Management of the ITER PCS design using a system-engineering approach,” *IEEE Trans. Plasma Sci.*, vol. 48, 2020.

[30] G. De Tommasi, “Plasma magnetic control in tokamak devices,” *J. Fus. Energy*, vol. 38, 2019.

[31] F. Hoffman, O. Sauter *et al.*, “Experimental and thoretical stability limits of highly elongated tokamak plasmas,” *Physical Review Letters*, 1998.

[32] J. P. Freidberg, A. Cerfon, and J. Lee, “Tokamak elongation–how much is too much? Part 1. Theory,” *J. Plasma Phys.*, vol. 81, no. 6, p. 515810607, 2015.

[33] G. Ambrosino, M. Ariola, G. De Tommasi, and A. Pironti, “Robust vertical control of ITER plasmas via static output feedback,” in *2011 IEEE International Conference on Control Applications (CCA)*, 2011.

[34] G. De Tommasi, A. Mele, and A. Pironti, “Robust plasma vertical stabilization in tokamak devices via multi-objective optimization,” in *Int. Conf. on Optimization and Decision Science*, 2017.

[35] E. Schuster, M. Walker, D. Humphreys, and M. Krstić, “Plasma vertical stabilization with actuation constraints in the DIII-D tokamak,” *Automatica*, vol. 41, 2005.

[36] S. Gerkšič and G. De Tommasi, “Vertical stabilization of ITER plasma using explicit model predictive control,” *Fus. Eng. Des.*, vol. 88, 2013.

[37] W. Biel *et al.*, “Development of a concept and basis for the DEMO diagnostic and control system,” *Fus. Eng. Des.*, vol. 179, 2022.

[38] L. Scibile and B. Kouvaritakis, “A discrete adaptive near-time optimum control for the plasma vertical position in a tokamak,” *IEEE Trans. Contr. Sys. Tech.*, vol. 9, 2001.

[39] N. Cruz *et al.*, “An optimal real-time controller for vertical plasma stabilization,” *IEEE Trans. Nucl. Sci.*, vol. 62, 2015.

[40] G. Ambrosino, M. Ariola, G. De Tommasi, and A. Pironti, “Plasma Vertical Stabilization in the ITER Tokamak via Constrained Static Output Feedback,” *IEEE Trans. Contr. Sys. Tech.*, vol. 19, no. 2, pp. 376–381, 2011.

[41] A. Scheinker and M. Krstić, *Model-free stabilization by extremum seeking.* Springer, 2017.

[42] O. Trollberg, “On real-time optimization using extremum seeking control and economic model predictive control,” Ph.D. dissertation, KTH School of Electrical Engineering, 2017.

[43] M. Leblanc, “Sur l’electrication des chemins de fer au moyen de courants alternatifs de frequence elevee,” *Revue Generale de l’Electricite*, 1922.

[44] V. Kazakevich, “Technique of automatic control of different processes to maximum or to minimum,” *Avtorskoe svidetelstvo,(USSR Patent)*, no. 66335, 1943.

[45] K. V. V., “On extremum seeking,” Ph.D. dissertation, Moscow High Technical University, 1944.

[46] C. S. Draper and Y. T. Li, “Principles of optimalizing control systems and an application to the internal combusion engine,” *(No Title)*, 1951.

[47] K. J. Astrom, “Adaptive control around 1960,” *IEEE Control Systems Magazine*, vol. 16, no. 3, pp. 44–49, 1996.

[48] J. Sternby, “Extremum control systems?? an area for adaptive control?” in *Joint Automatic Control Conference*, no. 17, 1980, p. 8.

[49] M. Krstic and H.-H. Wang, “Stability of extremum seeking feedback for general nonlinear dynamic systems,” *Automatica-Kidlington*, vol. 36, no. 4, pp. 595–602, 2000.

[50] H.-H. Wang and M. Krstic, “Extremum seeking for limit cycle minimization,” *IEEE Transactions on Automatic Control*, vol. 45, no. 12, pp. 2432–2436, 2000.

[51] Y. Tan, W. H. Moase, C. Manzie, D. Nešić, and I. M. Mareels, “Extremum seeking from 1922 to 2010,” in *Proceedings of the 29th Chinese control conference.* IEEE, 2010, pp. 14–26.

[52] A. Scheinker and M. Krstić, “Extremum seeking with bounded update rates,” *Syst. & Contr. Lett.*, vol. 63, 2014.

[53] T. Bellizio *et al.*, “Control of elongated plasma in presence of ELMs in the JET tokamak,” *IEEE Trans. Nucl. Sci.*, vol. 58, no. 4, pp. 1497–1502, 2011.

[54] G. De Tommasi, S. Dubbioso, A. Mele, and A. Pironti, “Stabilizing elongated plasmas using extremum seeking: the ITER tokamak case study,” in *29th Mediterranean Conference on Control and Automation (MED)*, 2021.

[55] S. Dubbioso, L. E. d. Grazia, and others., “Vertical stabilization of tokamak plasmas via extremum seeking,” *IFAC Journal of Systems and Control*, vol. 21, 2022.

[56] G. De Tommasi, S. Dubbioso, A. Mele, and A. Pironti, “Event-driven adaptive vertical stabilization in tokamaks based on a bounded extremum seeking algorithm,” in *2022 IEEE Conference on Control Technology and Applications (CCTA)*, 2022.

[57] M. Ariola and A. Pironti, “The design of the eXtreme Shape Controller for the JET tokamak,” *IEEE Control Sys. Mag.*, vol. 25, no. 5, pp. 65–75, 2005.

[58] R. Albanese *et al.*, “Design, implementation and test of the xsc extreme shape controller in jet,” *Fus. Eng. Des.*, vol. 74, no. 1-4, pp. 627–632, 2005.

[59] R. Ambrosino *et al.*, “Sweeping control performance on DEMO device,” *Fus. Eng. Des.*, vol. 171, p. 112640, 2021.

[60] G. Ambrosino and R. Albanese, “Magnetic Control of Plasma Current, Position and Shape in Tokamaks,” *IEEE Control Sys. Mag.*, vol. 25, no. 5, pp. 76–92, 2005.

[61] F. Wagner *et al.*, “Regime of improved confinement and high beta in neutral-beam-heated divertor discharges of the ASDEX tokamak,” *Physical Review Letters*, vol. 49, no. 19, p. 1408, 1982.

[62] V. Toigo *et al.*, “Conceptual design of the enhanced radial field amplifier for plasma vertical stabilisation in JET,” *Fus. Eng. Des.*, vol. 82, no. 5-14, pp. 1599–1606, 2007.

[63] E. M. Navarro-López and R. Carter, “Hybrid automata: an insight into the discrete abstraction of discontinuous systems,” *Int. J. Systems Sci.*, vol. 42, no. 11, pp. 1883–1898, 2011.

[64] G. Rattá and J. o. Vega, “An advanced disruption predictor for JET tested in a simulated real-time environment,” *Nucl. Fus.*, vol. 50, 2010.

[65] B. Cannas *et al.*, “Automatic disruption classification in jet with the iter-like wall,” *Plasma Phys. Control. Fus.*, vol. 57, 2015.

[66] J. Vega, S. Dormido-Canto *et al.*, “Results of the JET real-time disruption predictor in the ITER-like wall campaigns,” *Fus. Eng. Des.*, vol. 88, 2013.

[67] B. Cannas *et al.*, “A prediction tool for real-time application in the disruption protection system at JET,” *Nucl. Fus.*, vol. 47, 2007.

[68] B. Cannas, A. Fanni, E. Marongiu, and P. Sonato, “Disruption forecasting at JET using neural networks,” *Nucl. Fus.*, vol. 44, 2004.

[69] D. R. Ferreira, P. J. Carvalho, and H. Fernandes, “Deep learning for plasma tomography and disruption prediction from bolometer data,” *IEEE Trans. Plasma Sci.*, vol. 48, 2020.

[70] A. Murari *et al.*, “Adaptive predictors based on probabilistic SVM for real time disruption mitigation on JET,” *Nucl. Fus.*, vol. 58, 2018.

[71] 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,” *Nucl. Fus.*, vol. 61, 2021.

[72] A. Murari, J. Vega *et al.*, “Unbiased and non-supervised learning methods for disruption prediction at JET,” *Nucl. Fus.*, vol. 49, 2009.

[73] C. Rea *et al.*, “Disruption prediction investigations using Machine Learning tools on DIII-D and Alcator C-Mod,” *Plasma Phys. Control. Fus.*, vol. 60, 2018.

[74] C. Rea, K. Montes, K. Erickson, R. Granetz, and R. Tinguely, “A real-time machine learning-based disruption predictor in DIII-D,” *Nucl. Fus.*, vol. 59, 2019.

[75] K. Montes *et al.*, “Machine learning for disruption warnings on Alcator C-Mod, DIII-D, and EAST,” *Nucl. Fus.*, vol. 59, 2019.

[76] R. M. Churchill, B. Tobias, Y. Zhu, and D.-D. team, “Deep convolutional neural networks for multi-scale time-series classification and application to tokamak disruption prediction using raw, high temporal resolution diagnostic data,” *Physics of Plasmas*, vol. 27, 2020.

[77] J. Kates-Harbeck, A. Svyatkovskiy, and W. Tang, "Predicting disruptive instabilities in controlled fusion plasmas through deep learning," *Nature*, vol. 568, 2019.

[78] E. Aymerich *et al.*, "Performance Comparison of Machine Learning Disruption Predictors at JET," *Appl. Sciences*, vol. 13, 2023.

[79] ——, "Disruption prediction at JET through deep convolutional neural networks using spatiotemporal information from plasma profiles," *Nucl. Fus.*, vol. 62, 2022.

[80] E. Coccorese, C. Morabito, and R. Martone, "Identification of noncircular plasma equilibria using a neural network approach," *Nucl. Fus.*, vol. 34, 1994.

[81] S. Joung *et al.*, "Deep neural network Grad–Shafranov solver constrained with measured magnetic signals," *Nucl. Fus.*, vol. 60, 2020.

[82] J. Wai, M. Boyer, and E. Kolemen, "Neural net modeling of equilibria in NSTX-U," *Nucl. Fus.*, vol. 62, 2022.

[83] L. L. Lao *et al.*, "Application of machine learning and artificial intelligence to extend EFIT equilibrium reconstruction," *Plasma Phys. Control. Fus.*, vol. 64, 2022.

[84] A. Bustos, E. Ascasíbar, A. Cappa, and R. Mayo-García, "Automatic identification of MHD modes in magnetic fluctuation spectrograms using deep learning techniques," *Plasma Phys. Control. Fus.*, vol. 63, 2021.

[85] Y. Fu, D. Eldon *et al.*, "Machine learning control for disruption and tearing mode avoidance," *Physics of Plasmas*, vol. 27, 2020.

[86] V. Škvára *et al.*, "Detection of Alfvén Eigenmodes on COMPASS with Generative Neural Networks," *Fus. Sci. Technol.*, vol. 76, 2020.

[87] A. Jalalvand *et al.*, "Alfvén eigenmode classification based on ECE diagnostics at DIII-D using deep recurrent neural networks," *Nucl. Fus.*, vol. 62, 2022.

[88] A. Jalalvand, J. Abbate *et al.*, "Real-time and adaptive reservoir computing with application to profile prediction in fusion plasma," *IEEE Trans. Neural Net. Learn. Sys.*, vol. 33, 2022.

[89] J. Seo, R. Conlin *et al.*, "Multimodal prediction of tearing instabilities in a tokamak," in *2023 International Joint Conference on Neural Networks (IJCNN).* Gold Coast, Australia: IEEE, 2023.

[90] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, 2006.

[91] S. Ding, X. Xu, and R. Nie, "Extreme learning machine and its applications," *Neural Computing and Applications*, vol. 25, 2014.

[92] A. Pironti and A. Portone, "Optimal choice of the geometrical descriptors for tokamak plasma shape control," *Fus. Eng. Des.*, vol. 43, 1998.

[93] M. Ariola and A. Pironti, "Plasma shape control for the JET tokamak," *IEEE Control Systems*, vol. 25, 2005.

[94] M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," *Computer Science Review*, vol. 3, 2009.

[95] G. Tanaka *et al.*, "Recent advances in physical reservoir computing: A review," vol. 115, 2019.

[96] Z. Pan, Z. Meng *et al.*, "A two-stage method based on extreme learning machine for predicting the remaining useful life of rolling-element bearings," *Mechanical Systems and Signal Processing*, vol. 144, 2020.

[97] A. Jalalvand, B. Vandersmissen, W. De Neve, and E. Mannens, "Radar signal processing for human identification by means of reservoir computing networks," in *2019 IEEE Radar Conf.* Boston, MA, USA: IEEE, 2019.

[98] Z. Chen, K. Gryllias, and W. Li, "Mechanical fault diagnosis using convolutional neural networks and extreme learning machine," *Mechanical Systems and Signal Processing*, vol. 133, 2019.

[99] A. Jalalvand, F. Triefenbach, K. Demuynck, and J.-P. Martens, "Robust continuous digit recognition using reservoir computing," *Computer Speech & Language*, vol. 30, 2015.

[100] A. Ghoneim, G. Muhammad, and M. S. Hossain, "Cervical cancer classification using convolutional neural networks and extreme learning machines," *Future Generation Computer Systems*, vol. 102, 2020.

[101] F. Felici, H. Le, J. Paley, B. Duval, S. Coda, J.-M. Moret, A. Bortolon, L. Federspiel, T. Goodman, G. Hommen *et al.*, "Development of real-time plasma analysis and control algorithms for the tcv tokamak using simulink," *Fusion Engineering and Design*, vol. 89, no. 3, pp. 165–176, 2014.

[102] H. Le, F. Felici, J. Paley, B. Duval, J.-M. Moret, S. Coda, O. Sauter, D. Fasel, P. Marmillod *et al.*, "Distributed digital real-time control system for tcv tokamak," *Fusion Engineering and Design*, vol. 89, no. 3, pp. 155–164, 2014.

[103] M. Ariola, G. Ambrosino, A. Pironti, J. B. Lister, and P. Vyas, "Design and experimental testing of a robust multivariable controller on a tokamak," *IEEE Transactions on Control Systems Technology*, vol. 10, no. 5, pp. 646–653, 2002.

[104] J. Degrave *et al.*, "Magnetic control of tokamak plasmas through deep reinforcement learning," *Nature*, vol. 602, 2022.

[105] F. Pesamosca, F. Felici, S. Coda, C. Galperti, and T. Team, "Improved plasma vertical position control on tcv using model-based optimized controller synthesis," *Fusion Science and Technology*, vol. 78, no. 6, pp. 427–448, 2022.

[106] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction.* MIT press, 2018.

[107] E. Kolemen, J. Seo *et al.*, *Avoiding tokamak tearing instability with artificial intelligence*, Jul 2023.

[108] T. Wakatsuki, M. Yoshida *et al.*, “Simultaneous control of safety factor profile and normalized beta for JT-60SA using reinforcement learning,” *Nucl. Fus.*, 2023.

[109] J. Seo *et al.*, “Feedforward beta control in the kstar tokamak by deep reinforcement learning,” *Nucl. Fus.*, vol. 61, 2021.

[110] “KSTAR tokamak simulator (KSTAR-NN),” https://github.com/jaem-seo/KSTAR_tokamak_simulator.

[111] J. Seo *et al.*, “Development of an operation trajectory design algorithm for control of multiple 0D parameters using deep reinforcement learning in KSTAR,” *Nucl. Fus.*, vol. 62, 2022.

[112] G. De Tommasi *et al.*, “A RL-based Vertical Stabilization System for the EAST tokamak,” in *2022 American Control Conf.* Atlanta, GA, USA: IEEE, 2022.

[113] S. Dubbioso *et al.*, “A deep reinforcement learning approach for vertical stabilization of tokamak plasmas,” *Fus. Eng. Des.*, vol. 194, 2023.

[114] A. Castaldo *et al.*, “Simulation suite for plasma magnetic control at east tokamak,” *Fus. Eng. Des.*, vol. 133, pp. 19–31, 2018.

[115] R. Ambrosino *et al.*, “Model-based MIMO isoflux plasma shape control at the EAST tokamak: experimental results,” in *2020 IEEE Conference on Control Technology and Applications (CCTA)*, 2020, pp. 770–775.

[116] “Matlab Reinforcement Learning Toolbox,” https://www.mathworks.com/help/reinforcement-learning.html, 2022.