Reinforcement Learning-Based Controlled Switching Approach for Inrush Current Minimization in Power Transformers

Jone Ugarte-Valdivielso[a,*], Jose I. Aizpurua[b], Manex Barrenetxea[a] and Brian G. Stewart[c]

[a]Mondragon University, Electronics & Computer Science Department, Loramendi 4, Mondragon, 20500, Spain
[b]University of the Basque Country (UPV/EHU), Manuel de Lardizabal, 1 Donostia, 20018, Spain
[c]University of Strathclyde, Glasgow, Glasgow, G11RD, UK

ARTICLE INFO

Keywords:
Transformer
Inrush current
Reinforcement Learning
Proximal Policy Optimization
Deep Q-Network
Circuit Breaker

ABSTRACT

Transformers are essential components for the reliable operation of the power grid and the connection of renewable energy sources (RES) to them. The transformer core is constituted by a ferromagnetic material and depending on the magnetization state of the core, the energisation of the transformer can lead to high magnetizing inrush currents. Such high amplitudes can shorten the transformer life expectancy and cause power quality issues in the power grid, thereby challenging the reliable integration of RES to the grid. Various machine learning methods have been proposed to classify inrush currents. However, minimizing inrush current implies learning a sequence of actions considering the dynamic operation of the transformer. This is different from a classification problem and entails substantial challenges associated with magnetization properties of materials and power transformer operation dynamics. This paper introduces a novel Reinforcement Learning (RL) approach for inrush current minimization, capable of generating sequential decision-making strategies adapted to the dynamic operation of power transformers. The proposed method employs the Proximal Policy Optimization (PPO) algorithm and is trained and evaluated using an equivalent duality-based model of a real 7.4 megavolt-ampere (MVA) power transformer. The PPO-based strategy is compared with two benchmarks: Deep Q-Network (DQN) and the classical circuit breaker (CB) switching method used in laboratory tests. Results demonstrate that the PPO approach can minimise inrush current and achieves a reduction of 14.29% compared to DQN and 79.78% compared to laboratory measurements.

1. Introduction

Transformers are key assets for power system reliability and renewable energy integration (Ramirez et al., 2024). However, their operation can present challenges, particularly during energisation, when high magnetizing currents, known as inrush currents, can appear with amplitudes several times greater than the nominal current (Cigre, 2014). These inrush currents arise from core saturation and are influenced by the magnetic properties of the core, the remanent flux, and the transformer energisation instant, which is determined by the circuit breaker (CB) operation (Alassi et al., 2022).

Inrush currents are directly linked to mechanical stress and can contribute to a reduction in transformer lifetime. They can also cause power quality issues in the power grid, such as false tripping (Stipetic et al., 2023) and harmonic distortion (Sanati et al., 2024). The increasing decentralization of renewable energy generation has led to a proliferation of transformer-connected renewable energy sources (RES) located far from consumption centres, as well as their connection to distribution grids. This situation further challenges the power quality and grid stability, and thus inrush mitigation techniques are becoming increasingly important (Fang et al., 2016).

1.1. Problem Statement

Inrush currents in power transformers occur when the magnetic core saturates during energisation. Saturation is influenced by the magnetization curve, which is governed by the nonlinear hysteresis characteristics of the core material. Fig. 1 shows the relationship between the magnetizing flux (Φ), the magnetization curve, and the magnetizing current (i). An offset in the magnetizing flux can shift the operating point on the magnetization curve beyond the knee point,

*Corresponding author.
jugarte@mondragon.edu (J. Ugarte-Valdivielso); joxe.aizpurua@ehu.eus (J.I. Aizpurua); mbarrenetxeai@mondragon.edu (M. Barrenetxea); brian.stewart.100@strath.ac.uk (B.G. Stewart)
ORCID(s):

leading to high magnetizing currents and core saturation. In this condition, the permeability of the core decreases, increasing its reluctance, and significantly reducing the magnetization impedance. This results in a high magnetizing current, known as inrush current (Jiles, 1991).

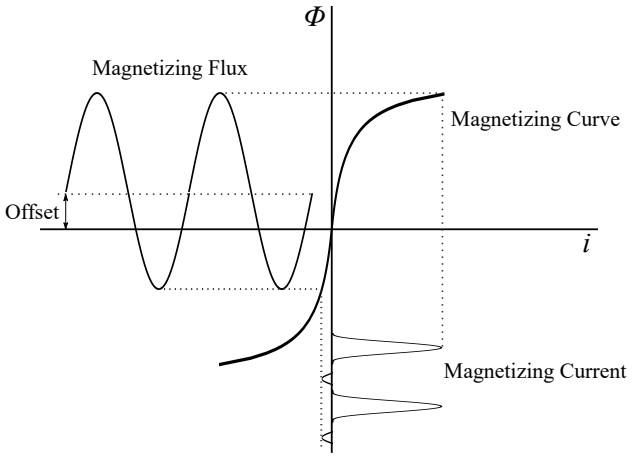


Fig. 1: Example of inrush current due to core saturation (Chiesa, 2010).

The origin of the flux saturation lies in the non-ideal transformer energisation instant, determined by the CB closing angle, and results from a mismatch between the remanent flux in the core and the magnetic flux due to the applied grid voltage. Fig. 2 illustrates an improper connection instant (denoted t_0), resulting in a flux increase beyond the expected flux (u denotes voltage [v] and Φ magnetizing flux [Wb]) .

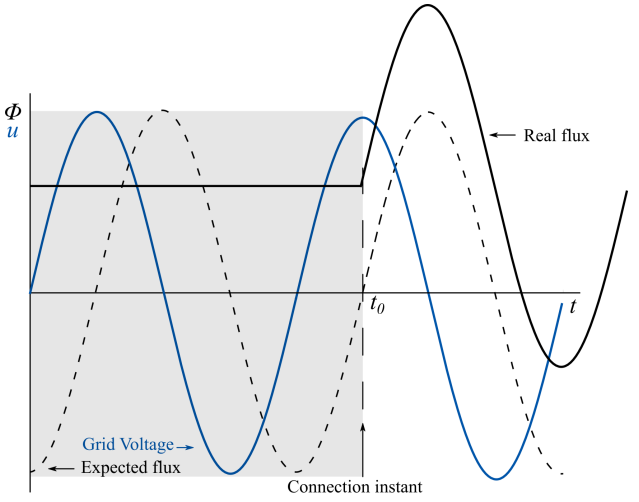


Fig. 2: Mismatch between the remanent and expected flux in the connection instant (Cano-González, 2015).

To minimise the amplitude of the magnetizing current, the difference between the remanent and expected flux at the connection instant must be reduced. This highlights the importance of precise CB switching and accurate estimation of the remanent flux. The remanent flux and the expected flux values are typically obtained by integrating the primary-side voltage at the moment of disconnection and the grid voltage, respectively. This principle forms the foundation of controlled switching, an inrush current mitigation technique that relies on coordinated disconnection and reconnection of the CB located between the power transformer and the grid (Cano-González, 2015).

In order to develop an effective inrush current minimization technique, a precise and realistic model of the system under study is necessary. Accurately modelling the transformer to reproduce real inrush current transients presents several challenges. The ferromagnetic core must be modelled to capture hysteresis behaviour, which can be achieved

using the Jiles–Atherton (JA) model (Jiles and Atherton, 1984, 1986). This model is expressed as a partial differential equation with five physical parameters that must be estimated (Ugarte et al., 2024; Silveyra et al., 2024; Garrido et al., 2026). The electrical model of the transformer also plays an essential role. The duality-based transformer model has shown satisfactory results in various inrush current simulation studies (Chiesa et al., 2010; Mork et al., 2007). Appendix A presents the duality-based electrical circuit of a three-phase, three-legged transformer with two windings, while further details about the simulation and validation environment can be found in (Ugarte et al., 2024, 2025).

Although controlled switching can suppress inrush current across many conditions, CB switching instants are traditionally determined through analytical methods that rely on deterministic system parameters and predictable operating conditions. In practice, however, the effectiveness of such approaches can be compromised by factors such as parameter measurement errors, remanent flux deviations (see Appendix B), and CB switching delays.

A critical source of uncertainty arises from the remanent flux estimation. The remanent flux is commonly estimated by integrating the terminal voltage at the transformer windings. However, in practice, this calculation can be challenging since voltage measurement errors directly affect the estimated flux, and the interactions with the electrical grid to which the transformer is connected can further distort the waveform. In particular, transient ferroresonance phenomena can arise when transformer impedance and power system impedance interact, producing distorted voltages that affect the accuracy of the remanent flux estimation (Valverde et al., 2007). These inaccuracies may lead to an incorrect determination of the optimal switching instant, causing a mismatch between the remanent and expected flux and resulting in undesired inrush currents, as shown in Fig. 1. These limitations highlight the need for alternative strategies that can adapt to system-specific uncertainties and measurement inaccuracies.

The CB and its insulating medium (typically vacuum or SF_6) can introduce additional uncertainties that affect the energisation process. During closing, if the voltage across the breaker exceeds the dielectric strength of the medium before full contact closure, a pre-strike arc can occur, causing current to flow earlier than intended. During opening, re-strike events may occur depending on the dielectric recovery of the insulating medium. Furthermore, the mechanical operating time of the breaker may vary due to several factors such as ambient temperature, inactivity periods, and mechanical wear caused by repeated operations. Although manufacturers provide nominal timings, the actual switching behaviour of the CB may gradually deviate from its nominal characteristics, challenging the determination of the optimal transformer energisation instant (Cano-González et al., 2017).

Both remanent flux estimation and CB operation exhibit uncertainties and time-varying behaviour. Accordingly, the optimal closing instant can be formulated as a dynamic, sequential decision-making problem. Each switching action provides feedback through the resulting magnetizing current, which can be used to improve future switching decisions. In this context, static rules can fail to accurately capture the evolving behaviour of the environment, particularly when system parameters change over time or are partially unknown.

These limitations motivate the development of adaptive control strategies capable of learning through interaction with the system. In this regard, Machine Learning (ML) techniques have been successfully applied to various power system problems, including battery health prediction (Li et al., 2025; Chen et al., 2024, 2025) and state-of-charge estimation (Mehta et al., 2024). In the context of transformer inrush currents, prior ML applications have primarily focused on classification and prediction tasks. In contrast, Reinforcement Learning (RL) is inherently suited to sequential decision-making problems. RL agents can continuously observe the system state, evaluate the outcomes of previous switching actions, and update future decisions accordingly to optimise a control policy. This capability enables the learning of adaptive switching strategies that progressively minimise inrush currents under uncertain and time-varying operating conditions, while accounting for both transformer dynamics and CB behaviour.

1.2. Related Work: Inrush Current and Sequential Decision-Making Frameworks

Related research on intelligent inrush current minimization can be categorized into two main areas: (i) ML solutions for inrush current challenges, and (ii) sequential decision-making methodologies.

1.2.1. Machine Learning for Inrush Current Challenges

Apart from power system related problems, ML has addressed challenges related to inrush current in power transformers, including classification, detection, and prediction. A significant focus has been on distinguishing inrush transients from fault currents, which is critical for reliable differential protection. For example, in (He et al., 2023), a Convolutional Neural Network (CNN) is proposed to improve differential protection accuracy and reliably differentiate inrush currents from faults. A variation known as Signal Localized CNN (SLCNN) is introduced in (Murugan et al., 2022) for the same purpose. Deep Neural Networks (DNNs) are applied in (Afrasiabi et al., 2020) to distinguish inrush

from fault currents, while (Afrasiabi et al., 2022) employs a Recurrent Neural Network (RNN) for the same task. In (Thomas and Shihabudheen, 2023), a differential architecture search (DARTS) algorithm is used to optimise network structures for distinguishing between fault and inrush conditions. Moreover, (Civelek et al., 2025) explores Gaussian Process Regression (GPR) and boosting methods to detect anomalies such as high inrush currents in smart grids.

Beyond classification, ML methods have also been applied to predict and estimate the magnitude and waveform of inrush currents. In (Singh and Chaturvedi, 2015), Neural Networks (NNs) are used to predict the maximum inrush current peak for large power transformers. Similarly, (Gopalakrishnan et al., 2024) proposes an NN-based estimation approach. A recent study (Peng et al., 2025) evaluates multiple algorithms, including Random Forest (RF), Principal Component Analysis (PCA), eXtreme Gradient Boosting (XGBoost), NNs, and CNNs, to predict magnetic field behaviour under inrush conditions in single-phase transformers. The study reports high accuracy and significant computational time reductions compared to finite element analysis (FEA) methods.

1.2.2. Sequential Decision-Making Solutions

With the increasing complexity of modern power grids, RL has become a promising framework for sequential decision-making, enabling agents to dynamically interact with their environments and learn optimal policies that maximize cumulative rewards (Sutton and Barto, 2015; Vaish et al., 2021; Yin et al., 2020; Bui et al., 2025). RL has been widely adopted for optimal control of complex systems, for instance, Degrave et al. [2022] introduced a maximum a posteriori policy optimization (MPO) method for plasma control in Tokamak reactors. In power systems, RL techniques have been applied to diverse tasks, including frequency regulation with Deep Deterministic Policy Gradient (DDPG) (Khooban and Gheisarnejad, 2021), voltage control via Least-Squares Policy Iteration (LSPI) (Xu et al., 2020), power converter control using Deep Q-Networks (DQN) and DDPG (Zandi and Poshtan, 2023), and electric vehicle demand management with DQN (Choudhary et al., 2025). More recently, (Wang et al. [2025]) proposed an RL-based Early Classification Proximal Policy Optimization (ECPPO) approach for distinguishing inrush currents from faults.

Considering the reviewed literature, Table 1 summarizes representative ML and RL methods relevant to this study, indicating whether they address inrush phenomena, employ RL strategies, their core methods, objectives, and specific use cases.

Table 1
Comparison of inrush studies and RL methods, including relevant features for this work.

Method	Inrush	RL	Method	Objective	Use Case
(He et al., 2023)	✓	✗	CNN	Classification	Power Transformer
(Murugan et al., 2022)	✓	✗	SLCNN	Classification	Power Transformer
(Afrasiabi et al., 2020)	✓	✗	DNN	Classification	Power Transformer
(Afrasiabi et al., 2022)	✓	✗	RNN	Classification	Power Transformer
(Thomas and Shihabudheen, 2023)	✓	✗	DARTS	Classification	Power System
(Civelek et al., 2025)	✓	✗	GPR and boosting	Classification	Power System
(Singh and Chaturvedi, 2015)	✓	✗	NN	Prediction	Power Transformer
(Gopalakrishnan et al., 2024)	✓	✗	NN	Prediction	Power Switch Network
(Peng et al., 2025)	✓	✗	RF, PCA, XGBoost, NN, CNN	Prediction	Power Transformer
(Degrave et al., 2022)	✗	✓	MPO	Control	Tokamak Reactor
(Khooban and Gheisarnejad, 2021)	✗	✓	DDPG	Control	Frequency
(Xu et al., 2020)	✗	✓	LSPI	Control	Voltage
(Zandi and Poshtan, 2023)	✗	✓	DQN, DDPG	Control	Power Converter
(Choudhary et al., 2025)	✗	✓	DQN	Control	Electric Vehicle
(Wang et al., 2025)	✓	✓	ECPPO	Classification	Power Transformer
Proposed Method	✓	✓	DQN, PPO	Minimization	Power Transformer

In summary, existing ML studies on inrush currents have primarily focused on classification, identification, and prediction. Within the broader RL community, various optimization and control problems have been addressed. However, to the best of the authors' knowledge, no studies have yet addressed inrush current minimization using either ML or RL approaches.

1.3. Contribution, Objectives & Impact

The main contribution of this paper is the development of a novel inrush current minimization framework based on RL. The proposed approach employs the Proximal Policy Optimization (PPO) algorithm to learn the optimal CB closing instant from the transformer environment.

A further contribution is the development of a transformer environment model capable of accurately reproducing inrush current transients. This contribution is particularly relevant given the limited availability of data-driven transformer models at high power ratings for inrush studies. The model is validated using a 7.4 MVA, 30/20 kV transformer (MU, 2025). While RL training and validation in a real laboratory would be ideal, medium-voltage inrush current tests are often impractical due to high costs and operational risks. Simulation-based studies therefore offer a safer, more cost-effective, and efficient alternative.

In this context, the main objective of this research is to develop an inrush current minimization strategy capable of adapting to the fluctuating operational and environmental conditions of power transformers. Conventional energisation strategies typically rely on analytical rules or deterministic point-on-wave switching, which assume ideal knowledge of system states such as residual flux and CB conditions. In practice, however, these quantities are uncertain, time-varying, or difficult to measure accurately, which can significantly degrade the performance of rule-based switching approaches.

To address these limitations, transformer energisation is formulated as a sequential decision-making problem in which switching actions interact with the transformer dynamics and the resulting current response provides feedback for improving future decisions. RL offers a natural framework for this setting, as it enables the learning of adaptive control policies directly through interaction with the transformer environment. By leveraging this capability, the proposed RL-based strategy learns an optimal switching policy that can respond to uncertainty and variability in operating conditions. Furthermore, to mitigate transformer degradation and potential grid stability issues associated with high magnetising currents, the proposed method aims to limit the inrush current amplitude below 1 per unit (pu) of the transformer nominal current.

The proposed PPO-based RL model is benchmarked against (i) two DQN-based models that employ different decay functions (linear and exponential) to balance exploration and exploitation, and (ii) classical CB switching strategies used in real inrush current tests. The results demonstrate that the proposed approach not only effectively minimises inrush current but also provides a more efficient and adaptive solution compared to both the DQN-based models and conventional minimization strategies.

1.4. Organization

The remainder of the paper is organized as follows. Section 2 introduces the basics of RL. Section 3 explains the methodology of the proposed approach, including the transformer modelling, laboratory experiments, and the development and evaluation of the RL strategy. Then Section 4 presents the numerical results of the transformer model validation along with the training, evaluation, and validation of the proposed inrush current minimization strategy. A discussion in Section 5 provides a comparative assessment of the study, addresses practical implementations, and outlines its limitations. In Section 6 the conclusions are presented.

2. Reinforcement Learning Fundamentals

RL algorithms solve sequential decision-making problems that require a series of actions to obtain an optimal solution. The algorithm follows a Markov Decision Process (MDP) and, as such, is typically represented through the tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P}, \gamma \rangle$, where $\mathcal{S}$ represents the state of the environment, $\mathcal{A}$ represents the set of actions that RL can take, and $\mathcal{P}$ represents the transition probability function. The reward function is typically represented as a function of the next state and the current action $\mathcal{R}(s_{t+1}, a_t)$. The function generates the reward due to the action induced by the state-transition from s_t to s_{t+1}. Finally, $\gamma \in [0, 1]$ is the discount factor that determines the relevance of the rewards (Sutton and Barto, 2015).

RL is based on the decision-making of an agent in a dynamic environment to maximize a cumulative reward. At each time step (t), the agent takes an action (a_t) depending on the state (s_t) and its past knowledge. This action is applied to the environment, providing the agent with a reward for the next state (r_t) based on the quality of the action and the transition to the next state (s_{t+1}). The RL agent-environment interaction process is shown in Fig. 3.

In this study, the PPO method is selected for its stable learning ability and is compared against a DQN-based algorithm and a conventional CB closing strategy. The selection of DQN is based on its extensive use in the literature (Chen et al., 2022).

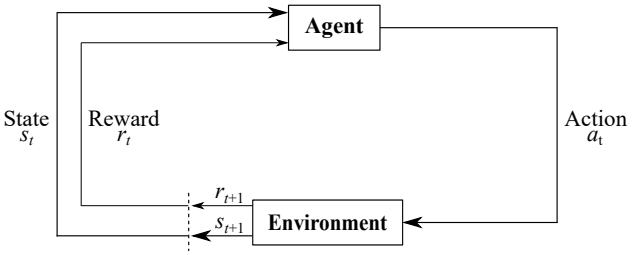
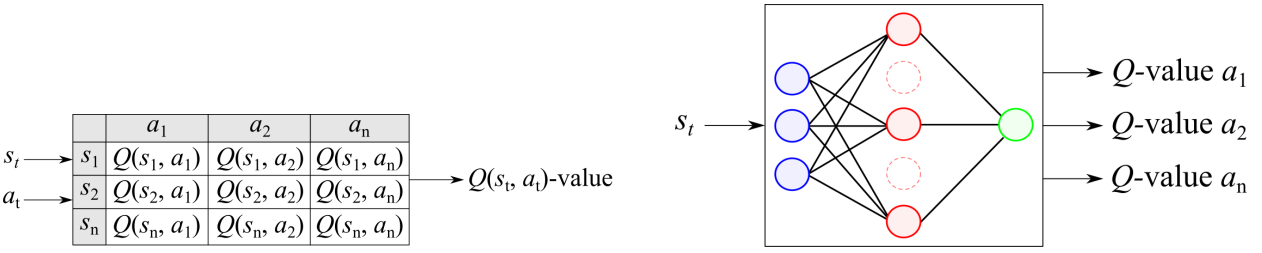


Fig. 3: Agent-environment interaction in RL (Sutton and Barto, 2015).

2.1. Deep Q-Network

Conventionally, for small state-action spaces, the Q-training is carried out with the help of a Q-table. Each row in the Q-table corresponds to a state-action pair Q-value. At each time step, the Q-table is updated by training a Q-function. Therefore, depending on the selected state-action pair by the agent at each time step, the Q-function searches the Q-table for the corresponding Q-value. Initially, the Q-table does not provide any information about the state-action selection, and is gradually refined through experience. Depending on the study, the state space can be too large, and the use of a Q-table can become inefficient and computationally infeasible. To address this limitation, the present work employs NNs as function approximators in place of Q-tables (Wilson and Riccardi, 2023). This approach is referred to as the DQN method. To highlight the difference between these two methods, Fig. 4 illustrates the basic architectures of the conventional Q-table and DQN.



(a) Q-table with states, actions, and Q values. **(b)** Basic DQN structure with a NN model.

Fig. 4: Basic structures of Q-table and DQN model.

To improve the current policy, the agent explores the environment through actions. The action of the agent for each state is chosen according to a ε-greedy policy. The ε-greedy method gives a 1-ε probability of taking the action with the highest value for a given state, which is known as the exploitation phase. In contrast, the agent has an ε probability of selecting a random action and performing the investigation, known as the exploration phase. Normally, the $\varepsilon_{\text{initial}}$ value is set to 1 and the ε value decays with each training time step. The decaying function can be linear or exponential, and the decaying rate depends on the hyperparameter $exploration fraction$. The $\varepsilon_{\text{final}}$ value is another hyperparameter that can be tuned. The selection of an adequate ε decay function is essential for a proper exploration/exploitation trade-off in the training (Huang et al., 2019).

The value of a state-action pair $Q\left(s_t, a_t\right)$ is updated in each decision step following (Silva et al., 2020):

$$\begin{aligned} Q\left(s_t, a_t\right) \leftarrow & Q\left(s_t, a_t\right) \\ & + \alpha\left(r_t + \gamma \max_a Q\left(s_{t+1}, a\right) - Q\left(s_t, a_t\right)\right) \end{aligned} \tag{1}$$

where s_t and a_t are the state and action in time step t, respectively, α is the learning rate, which determines how fast the NN is trained, r_t is the reward in time step t, γ is a discounting factor and $\max_a Q(s_{t+1}, a)$ is the best state-action pair value of the next state s_{t+1} and a given action a.

2.2. Proximal Policy Optimization

PPO is an *actor* − *critic* RL algorithm, in which two NNs are trained simultaneously. The first NN, the *actor*, is a policy that gives the probability of taking an action over a state. The second NN, the *critic*, learns to approximate a value-based function similar to DQN, and assists the policy update of the *actor* function by providing feedback on the quality of the taken action. This technique reduces the variance in training for the same starting point compared to other RL algorithms (Gu et al., 2021). The basic structure of PPO is shown in Fig. 5.

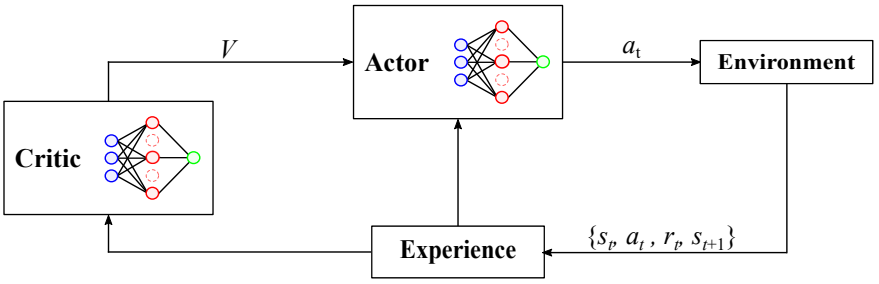


Fig. 5: Basic structure of PPO.

Moreover, PPO also incorporates a mechanism to improve training stability by limiting the size of policy updates. This is achieved through a clipping function applied to the probability ratio between the new and old policies at each time step defined as follows (Fraija et al., 2024):

$$\text{clip}\left(p_t(\theta), 1-\epsilon, 1+\epsilon\right) = \begin{cases} (1-\epsilon) & \text{if } p_t(\theta) < (1-\epsilon) \\ (1+\epsilon) & \text{if } p_t(\theta) > (1+\epsilon) \\ p_t(\theta) & \text{else} \end{cases} \tag{2}$$

where ϵ is the clipping range and $p_t(\theta)$ is the probability ratio between the old and new policies, with θ being the trainable parameters of the NN weights and biases ($\theta = \{w, b\}$). The probability ratio is defined as follows:

$$p_t(\theta) = \frac{\pi_{\theta(a_t|s_t)}}{\pi_{\theta_{\text{old}}(a_t|s_t)}} \tag{3}$$

where $\pi_{\theta(a_t|s_t)}$ is the probability of taking action a_t in a state s_t in the current policy and $\pi_{\theta_{\text{old}}(a_t|s_t)}$ is the same in the old policy. The old policy, $\pi_{\theta_{\text{old}}}$, denotes a frozen copy of the policy used to generate the trajectories for the current update, prior to any gradient-based modifications.

The clipped objective function L^{CLIP} is calculated by means of (2) and (3) and is defined as follows:

$$L^{\text{CLIP}} = \hat{\mathbb{E}}_t\left[\min\left(p_t(\theta)\cdot A_t, \text{clip}\left(p_t(\theta), 1-\epsilon, 1+\epsilon\right)\right)\cdot A_t\right] \tag{4}$$

where $\hat{\mathbb{E}}_t$ is the expectation over an episode and A_t is the advantage estimate, which quantifies how much better or worse taking action a_t in state s_t is compared to the expected value of that state.

In contrast, the value-based function, the *critic* network, is trained by minimizing the difference between the predicted state value $V_\omega\left(s_t\right)$ and its target state value V_t^{target}. The value-based function L^{V} is defined as follows:

$$L^{\text{V}} = \hat{\mathbb{E}}_t\left[\left(V_\omega\left(s_t\right) - V_t^{\text{target}}\right)^2\right] \tag{5}$$

Moreover, exploration and exploitation are handled differently compared to DQN. Exploration is encouraged by adding an entropy term, H, which promotes stochastic action selection and prevents premature convergence to suboptimal policies. Exploitation is reinforced through the clipped surrogate objective L^{CLIP}, which directly maximizes expected rewards based on the collected trajectories. The overall combined objective function is defined as follows:

$$L^{\text{CLIP+H+V}} = L^{\text{CLIP}} + hH + \upsilon L^{\text{V}} \tag{6}$$

where $L^{\text{CLIP+H+V}}$ is the overall objective function and h and υ are hyperparameters used as weighting factors for the entropy and the value-based functions, respectively.

3. Proposed Approach

The proposed approach is based on the combination of RL and controlled switching to minimise the inrush current in power transformers. Fig. 6 presents the main steps of the proposed inrush current minimization framework.

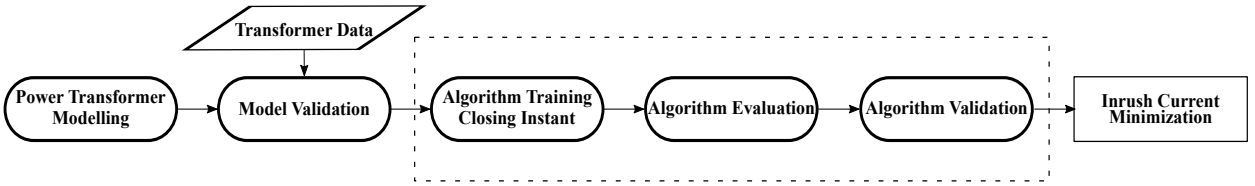


Fig. 6: Block diagram of the proposed approach.

3.1. Power Transformer Modelling

The first step is to develop an accurate representation of the system under study. Therefore, the study starts with the `Power Transformer Modelling` based on the duality transformation presented in Appendix A. The main `Transformer Data` and the values of the parameters shown in Fig. A.1 are presented in Table 2. The transformer core is made of an M120-27S material.

Table 2
Main parameters of the transformer under study.

Parameter	Symbol	Value	Unit
Apparent Power	S	7.4	MVA
Primary Voltage	V_p	30	kV
Secondary Voltage	V_s	20	kV
Primary Current	I_p	142.4	A
Frequency	f	50	Hz
Primary Turns	N	824	-
High-Voltage Winding Resistance	R_{HV}	1.28	Ω
Low-Voltage Winding Resistance	R_{LV}	0.26	Ω
Low-Voltage to Core Leakage Inductance	L_{LC}	55.58	mH
High- to Low-Voltage Leakage Inductance	L_{HL}	81.12	mH
Zero-Sequence Inductance	L_0	25.37	μH

3.2. Model Validation

The `Model Validation` is performed by comparing simulation inrush current results with real measurements. The real inrush current tests are conducted in the medium-voltage laboratory of Mondragon University (MU, 2025). The single line diagram of the laboratory test setup is presented in Fig. 7. The tests are divided into two parts: the disconnection of the power transformer and its subsequent energisation through the secondary CB switching.

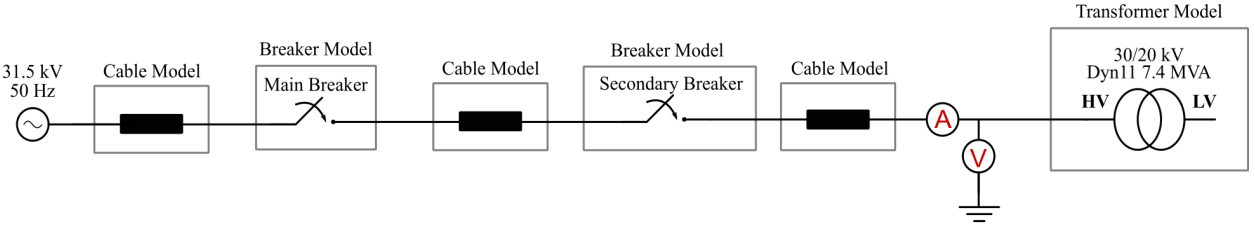


Fig. 7: Laboratory setup for the inrush current tests.

In the laboratory tests, a total of 48 CB opening cases have been recorded. As the secondary CB follows a classical switching, openings can occur at any angle of the voltage waveform. The three-phase remanent fluxes, ϕ_1, ϕ_2, and ϕ_3, are obtained by integrating the primary-side voltage of the transformer when it is disconnected from the grid. The obtained remanent flux values and data are presented in Appendix B.

On the other hand, 21 different inrush current measurements have been recorded. To replicate the inrush current transient, it is necessary to model not only the transformer, but also the remaining components of the laboratory setup. The component models and additional materials, are publicly available at (Ugarte, 2025) and further explained in (Ugarte et al., 2025). These supplementary materials provide comprehensive support for the further exploration and implementation of different inrush minimization strategies in the same environment.

It is worth noting that the availability of laboratory tests conducted at such high power ratings are scarce in the literature. The remanent flux values obtained from these measurements enabled accurate modeling of the remanent flux behaviour as a function of the CB opening instant, thereby providing valuable data for future research. Additionally, the real inrush current tests on medium-voltage power transformers offer important insights into the phenomenon and serve as a useful reference for subsequent minimization studies.

3.3. Algorithm Training

Once the transformer model is validated, the next stage of the study is focused on the development of an inrush current minimization strategy based on RL and controlled switching. The three stages covered by the dashed box in Fig. 6 are part of the RL inrush current minimization strategy framework. The RL environment is a power transformer together with laboratory cables, CBs, and the power grid (cf. Fig. 7). The RL states correspond to the opening angle of the secondary CB and the three remanent fluxes. The RL action is the closing angle of the secondary CB. Thus, the agent learns the best CB closing instant (action) depending on the opening angle and the remanent fluxes (state) of the last CB opening.

The `RL training` phase is carried out using the opening angles and remanent fluxes obtained from the validated transformer simulation model. These cases are generated by simulating the secondary breaker opening for 360° with 1° steps and obtaining the remanent fluxes for each step. Note that the opening and closing angles are defined with respect to the positive zero-crossing phase-to-ground voltage of phase a. For each iteration of the training process, the remanent fluxes at that state (ϕ_1, ϕ_2, ϕ_3), and the selected closing angle (action) by the RL algorithm are used to feed the transformer model (environment). The model responds with the maximum peak inrush current value (I_{max}) in [pu] for that state-action pair, and with this data, the reward is calculated. This process and the inputs and output of the environment are shown in Fig. 8.

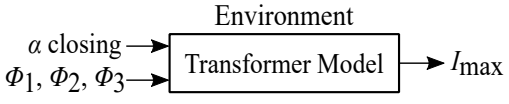


Fig. 8: Diagram of the environment.

The objective of the RL algorithm is to maximize the cumulative reward, hence, the inrush current minimization strategy should be able to minimise the peak at least below the rated current or 1 pu. The reward function is defined as:

$$r = \begin{cases} -I_{\text{max}}, & \text{if } I_{\text{max}} > 1, \\ (1 - I_{\text{max}}), & \text{otherwise.} \end{cases} \tag{7}$$

where r is the reward.

The piecewise linear reward function is selected for its simplicity, interpretability, and direct alignment with the operational objective of limiting the inrush current amplitude. In particular, the function assigns stronger penalties when the peak current exceeds the rated value (1 pu), since this operating condition is undesirable and may lead to protection trips, equipment stress, or other operational issues. Conversely, when the peak current remains below 1 pu, the reward increases as the inrush current decreases, encouraging the agent to identify switching strategies that further reduce the peak current amplitude.

Alternative reward formulations were also considered during preliminary experimentation, including purely negative rewards proportional to the peak current and quadratic penalty functions. However, these formulations provided

weaker learning signals near the desired operating region and produced larger reward magnitudes that slowed training convergence. The adopted piecewise linear formulation provided a stable learning signal while explicitly distinguishing between acceptable ($I_{\text{max}} < 1$ pu) and undesirable ($I_{\text{max}} > 1$ pu) operating regimes.

The I_{max} parameter is expressed in pu rather than amperes to normalise the reward scale and improve the numerical stability of the learning process. Preliminary tests showed that using current values expressed in amperes resulted in larger reward magnitudes, which led to slower convergence and less stable training of the RL algorithm.

The selected RL algorithm based on PPO is compared against DQN. In DQN, two types of ε-greedy policies are tried to obtain the best strategy for a proper exploration/exploitation trade-off. The compared ε-greedy policies have linear and exponential decays. Both DQN variants use a multilayer perceptron (MLP) model with two hidden layers of 64 neurons and ReLU activation functions. The PPO model adopts a stochastic policy with an actor-critic architecture, where both the actor and the critic are implemented as fully connected networks, each with two hidden layers of 64 neurons and ReLU activations.

The learning process of these RL algorithms is controlled by hyperparameters, which control the learning steps, the training speed, and the proper convergence of the algorithm. Therefore, setting the hyperparameters appropriately is an important step in the modelling process. In this study, hyperparameter selection is conducted by Bayesian optimization, which is implemented through the Optuna library (Akiba et al., 2019). The different hyperparameters used for the PPO and DQN algorithms, their tuning range, and their selected values are presented in Table 3.

Table 3
Hyperparameter ranges for PPO and DQN tuning.

Hyperparameter	PPO		DQN		
	Range	Select	Range	Linear Select	Exponential Select
Batch Size	[32, 256]	256	[32, 256]	256	256
Clipping Parameter ($\varepsilon_{\text{clip}}$)	[0.1, 0.4]	0.14	-	-	-
Discount Factor (γ)	0.99	0.99	0.99	0.99	0.99
Entropy Coefficient (h)	$[10^{-6}, 10^{-1}]$	5.53×10^{-2}	-	-	-
Exploration Initial Epsilon ($\varepsilon_{\text{initial}}$)	-	-	1	1	1
Exploration Final Epsilon ($\varepsilon_{\text{final}}$)	-	-	$[10^{-5}, 10^{-2}]$	9.19×10^{-4}	8.67×10^{-4}
Exploration Fraction	-	-	$[10^{-2}, 1]$	0.22	0.49
Learning Rate (α)	$[10^{-4}, 10^{-2}]$	2.42×10^{-3}	$[10^{-4}, 10^{-2}]$	2.74×10^{-3}	1.15×10^{-3}
Replay Buffer Size	-	-	$[10^{3}, 10^{5}]$	1000	100000
Value-Based Coefficient (h)	[0.1, 0.9]	0.89	-	-	-

The training stop criterion is set to 70,000 iterations. The selection of this value is justified from the search space of the problem, which consists of 360 opening angles with 360 closing angles, resulting in a total of 129,600 different cases. Consequently, the maximum iteration number is set to a value close to half of the total search space. In addition, preliminary experiments were conducted with a larger number of training iterations. It was observed that, while the learning performance initially improved with additional iterations, the accumulated reward reached a plateau beyond a certain point. Based on these observations, 70,000 iterations were selected as the stopping criterion, as further increasing the number of iterations would result in significantly higher computational time without meaningful performance improvement.

Moreover, to account for the stochastic nature of RL algorithms, including random initialization and exploration/exploitation rates, each algorithm is trained 10 times with different random seeds. By performing multiple runs, the mean and standard deviation of the cumulative rewards are reported. This procedure ensured that the results reflect the actual potential training of the algorithm, instead of relying on single run results (Henderson et al., 2018).

3.4. Algorithm Evaluation

The `evaluation` of each trained NN is carried out using as inputs the laboratory-measured remanent fluxes (see Appendix B) and their corresponding opening angles, which were not used during the training process (see Fig. 9). The NN gives the closing angle of the secondary CB as the output. The remanent fluxes and the closing angle are used to run the transformer environment simulation model, and hence, the inrush current peak for these settings is obtained. The evaluation process is repeated for the three RL methods (DQN linear, DQN exponential, and PPO), and their results

are compared and analysed. If the inrush current peak values are below 1 pu, training the RL algorithm is considered successful.

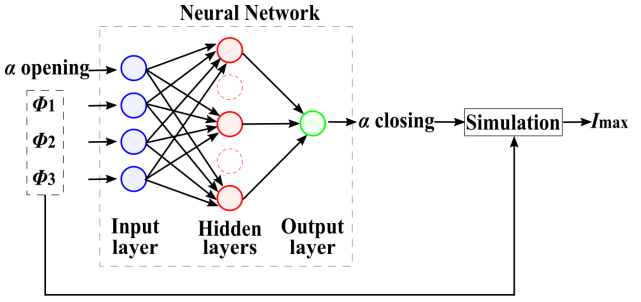


Fig. 9: Visual representation of the `Evaluation` stage, including opening and closing stages, simulation model and inrush current calculation

.

3.5. Algorithm Validation

Finally, the `RL algorithm validation` is performed by comparing the inrush current results obtained with the trained RL algorithms and the laboratory tests performed with classical CB closings. By fulfilling this comparison, the proposed approach is validated as an alternative `inrush current minimization` technique for power transformers.

4. Numerical Results

The validation of the RL inrush current minimization strategy is performed by first validating the transformer model under study. The validation is carried out by comparing the inrush current values of the measurement with the simulation results of the transformer model. This environment is then used to train the agents of the selected RL algorithms. The RL strategy is analyzed in three parts; training, evaluation, and validation. The simulation and algorithms presented in this section are executed with an Intel(R) Core(TM) i5-8265U CPU @ 1.60GHz processor.

4.1. Transformer Model Validation

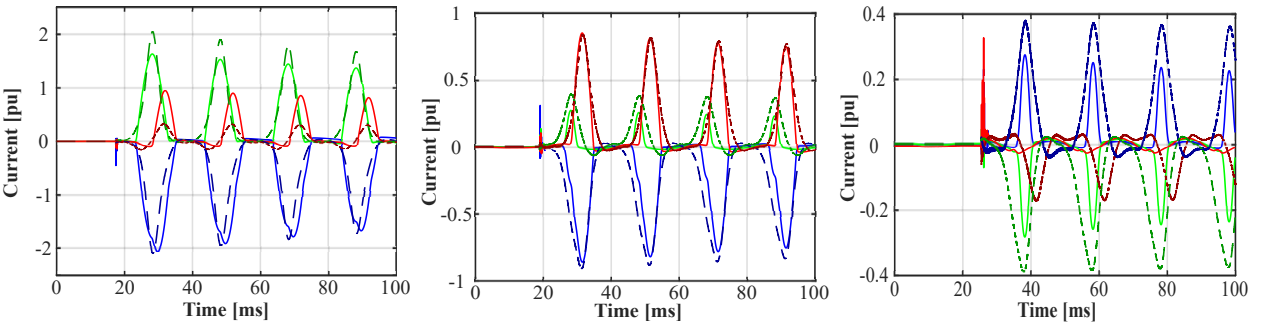
The transformer selected to test and validate the RL inrush current minimization strategy is a 7.4 MVA 30/20 kV Dyn11, three-legged core transformer, presented in Table 2. The transformer is modelled by the duality transformation approach, as detailed in Appendix A. The nominal primary current is 142.4 A rms. Considering this value, three different inrush current peak levels, $I_{\max}$, are chosen for the transformer model validation. Case 1 corresponds to the maximum expected inrush current specify by the transformer manufacturer, 411 A or 2.04 pu. Case 2 has an amplitude of 172 A or 0.85 pu and Case 3 has a maximum amplitude of 56 A or 0.28 pu. These cases are simulated by considering the remanent flux of the prior CB opening angle, which can be found in Appendix B. The CB opening angles related to these scenarios are 180° for Case 1, 257° for Case 2, and 37° for Case 3.

Fig. 10a, Fig. 10b, and Fig. 10c compare the inrush current waveforms obtained from experimental measurements and the developed transformer model for Cases 1, 2, and 3, respectively.

The continuous lines represent the three-phase inrush currents measured in the laboratory, with phases *a*, *b*, and *c* shown in blue, green, and red, respectively. The dashed lines correspond to the simulated three-phase inrush current waveforms from the model, following the same color scheme as the measurements. Note that the results are presented in pu format to consider the severity of the inrush current. The results show that the measured and simulated waveforms exhibit the same overall pattern, indicating that the transformer model can accurately replicate the inrush current transients of the transformer under study.

The results demonstrate that the implemented transformer model accurately reproduces the inrush current peak values observed in the laboratory. The peak errors between simulation and measurement are 0.04 pu or 7.27 A for Case 1, 0.05 pu or 11.04 A for Case 2 and 0.11 pu or 21.19 A for Case 3. The largest error occurs for Case 3, when the inrush current is minimal, which corresponds to only a 0.11 pu deviation from the measured value. Since the primary objective of this study is to train an RL algorithm to reduce inrush currents below 1 pu, this level of error is considered acceptable. In all the presented cases, the simulated inrush current values are slightly higher (0.11 pu maximum) than

the measured ones. This suggests that the transformer model is conservative, making more challenging to reduce the inrush in the simulation environment than in the experimental setup.



(a) Case 1. **(b)** Case 2. **(c)** Case 3.

Fig. 10: Comparison between measurement and transformer model for different inrush current cases.

4.2. Inrush Current Minimization Strategy Results

This section presents the results of the training and evaluation stages, comparing the learning performance and outcomes of PPO with two DQN-based methods. In the final stage, corresponding to validation, the proposed PPO approach is assessed against the inrush current results obtained from the DQN methods and classical switching.

4.2.1. Training

The training is conducted for PPO and two variants of DQN. Fig. 11 illustrates the evolution of the average episode reward with a shaded confidence interval (CI) of 95% across training iterations. The episode reward represents the cumulative sum of rewards within a single episode, where the individual rewards are calculated as described in (7). As shown in Fig. 11, PPO achieves the highest average episode reward and the lowest variance at the end of training. DQN with ϵ exponential decay exhibits only a 2.5% decrease in the average episode reward, however, the variance at the end of the training is higher than for PPO. DQN with ϵ linear decay performs the worst, with a 5.5% lower average episode reward than PPO and the highest variance at the end of the training. Beyond the final values, PPO also demonstrates the highest average episode reward and lowest variance at the beginning of training, whereas both DQN variants show considerably higher variance in the early stages. Although DQN with ϵ exponential decay occasionally surpasses PPO in average reward, PPO maintains a more stable learning trajectory throughout the training process. The high variance observed for both DQN methods underscores their relative instability during learning.

The computational time required for training each RL method is 348 seconds for DQN with linear ϵ decay, 305 seconds for DQN with exponential ϵ decay, and 201 seconds for PPO. One contributing factor to the faster training of PPO is its lower number of trainable parameters. Specifically, both DQN models comprise 55,630 parameters, whereas the PPO architecture contains 32,360 parameters in total.

Although PPO employs two neural networks (actor and critic), each network is smaller than the single DQN network. The higher parameter count in DQN is primarily due to its output layer, which scales with the number of discrete actions in the problem. In contrast, PPO uses separate actor and critic networks with more compact output representations, resulting in a lower overall parameter count and improved training efficiency .

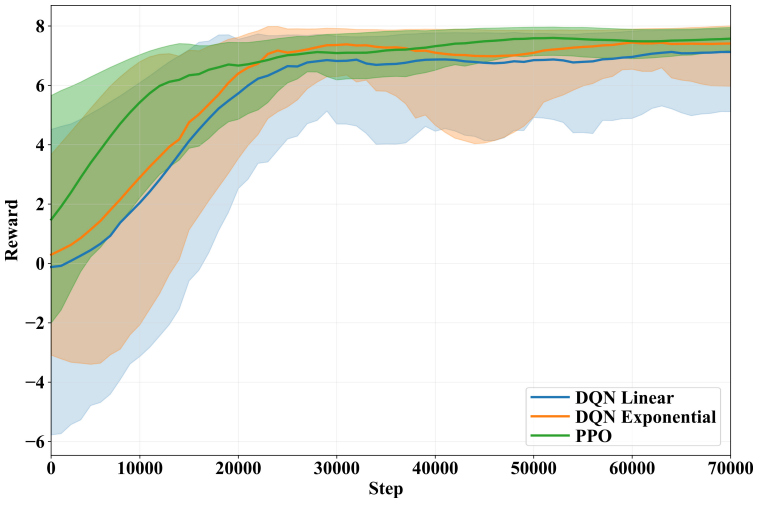


Fig. 11: Average episode reward for each RL algorithm during training, shaded areas indicate 95% CI.

4.2.2. Evaluation

The inrush current peak results obtained in the evaluation stage for the different RL algorithms are illustrated in Fig. 12 using box plots.

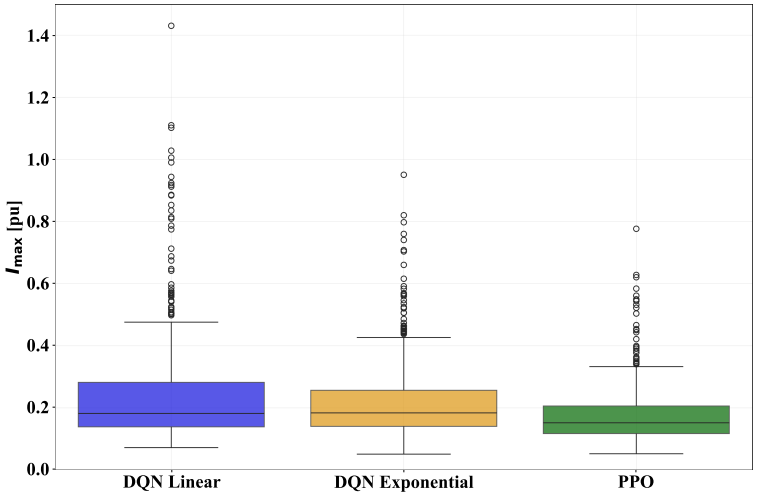


Fig. 12: Box plot results of the evaluation for each RL algorithm.

In the box plot, DQN with ϵ linear decay exhibits outliers exceeding the acceptable inrush current limit of 1 pu. Nevertheless, its overall mean value remains within an acceptable range at 0.25 pu. The other two algorithms do not exceed 1 pu at any case and achieve lower mean values: 0.22 pu for DQN with ϵ exponential decay and 0.18 pu for PPO. The most notable difference between DQN with ϵ exponential decay and PPO is their maximum values, with PPO being higher. However, since both maximum inrush current peaks remain within the 1 pu limit, both strategies are considered acceptable. Key statistics from the box plot in Fig. 12 are summarized in Table 4. The parameters Q1 and Q3 correspond to the first and third quartiles, representing the points below which 25% and 75% of the data fall, respectively.

The evaluation with new data indicates that the training is satisfactory for DQN with ϵ exponential decay and PPO. However, since the goal of the RL training is to minimise inrush current below 1 pu and DQN with ϵ linear decay exceeds this limit, it is discarded from further analyses.

Table 4
Detailed values of the inrush currents in the box plot in Fig. 12.

Parameter	DQN Linear	DQN Exponential	PPO	Units
Mean	0.25	0.22	0.18	pu
Median	0.18	0.18	0.15	pu
Q1	0.14	0.14	0.12	pu
Q3	0.28	0.26	0.21	pu
Minimum	0.07	0.05	0.05	pu
Maximum	1.43	0.95	0.78	pu

4.2.3. Validation

The validation results of the proposed PPO method, along with the DQN and classical switching inrush current methods, are shown in Fig. 13. Each bar represents one opening angle, with the bar height indicating the mean and the lines showing the minimum and maximum inrush current values. Detailed numerical results are provided in Table 5. Note that the inrush values of the classical switching method are experimental test results.

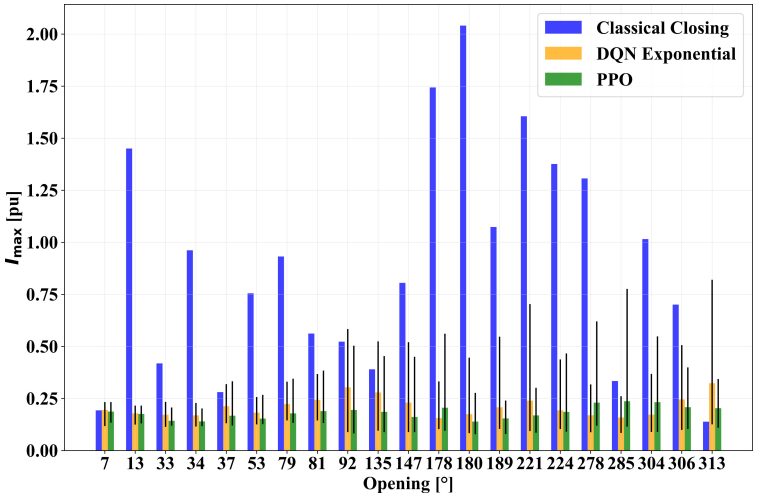


Fig. 13: Inrush current peak comparison between the classical closing and RL results, the black lines indicate the minimum and maximum for each case.

The results indicate that classical CB closing leads to inrush currents exceeding 1 pu in 38.09% of the cases. For DQN with ϵ exponential decay, the maximum inrush current remains below 0.82 pu, while for PPO it stays below 0.78 pu. The maximum mean inrush current is 0.41 pu for DQN and 0.39 pu for PPO. Moreover, in 95.24% of cases, the average inrush current peak obtained with PPO is lower than the corresponding classical switching value, demonstrating the effectiveness of the proposed minimization strategy.

In addition to the individual values, the average inrush current peak is calculated for all cases. For classical CB closing based strategy, the average inrush current mean is 0.89 pu. DQN with ϵ exponential decay and PPO achieve average values of 0.21 pu and 0.18 pu, respectively. This corresponds to an 14.29% reduction in the average inrush current peak for PPO compared to DQN and a 79.78% reduction compared to classical switching.

It must be highlighted that, according to the transformer model validation presented in Section 4.1, the maximum error in the inrush current amplitude between measured and simulated values is 0.11 pu. The model also exhibits a conservative behaviour, as the simulated inrush currents are slightly higher than those measured experimentally.

All the inrush current amplitudes obtained with the DQN exponential decay and PPO approaches in Fig. 13 are below the target limit of 1 pu. This value corresponds to the nominal current of the transformer, which is the rated operating point for which the device is designed. Therefore, currents below 1 pu are within the normal operating capability of the device. Considering the conservative nature of the model, the actual inrush currents are expected to be

equal to or lower than the simulated values. Hence, the validation error of 0.11 pu should not significantly affect the interpretation of the results presented in Fig. 13 and Table 5.

Table 5
Mean, minimum and maximum of the inrush current results presented in Fig. 13.

Opening [°]	Classical Closing [pu]	DQN Exponential [pu]			PPO [pu]		
		Mean	Minimum	Maximum	Mean	Minimum	Maximum
7	0.19	0.19	0.12	0.23	0.19	0.13	0.23
13	1.45	0.18	0.13	0.22	0.18	0.13	0.22
33	0.42	0.17	0.11	0.23	0.14	0.12	0.21
34	0.96	0.17	0.12	0.23	0.14	0.12	0.20
37	0.28	0.21	0.13	0.32	0.17	0.12	0.33
53	0.76	0.18	0.13	0.26	0.15	0.13	0.27
79	0.93	0.22	0.15	0.33	0.18	0.13	0.35
81	0.56	0.24	0.14	0.37	0.19	0.13	0.38
92	0.52	0.30	0.09	0.58	0.20	0.08	0.50
135	0.39	0.28	0.10	0.52	0.19	0.09	0.45
147	0.81	0.23	0.09	0.52	0.16	0.09	0.45
178	1.74	0.16	0.10	0.33	0.21	0.10	0.56
180	2.04	0.17	0.08	0.45	0.14	0.08	0.28
189	1.07	0.21	0.10	0.55	0.15	0.08	0.24
221	1.61	0.24	0.09	0.70	0.17	0.09	0.30
224	1.38	0.19	0.10	0.44	0.19	0.09	0.47
278	1.31	0.17	0.09	0.32	0.23	0.12	0.62
285	0.33	0.16	0.08	0.26	0.24	0.12	0.78
304	1.02	0.17	0.09	0.37	0.23	0.09	0.55
306	0.70	0.25	0.10	0.51	0.21	0.10	0.40
313	0.14	0.32	0.13	0.82	0.20	0.11	0.34
Mean	0.89	0.21	0.11	0.41	0.18	0.11	0.39

4.2.4. Model Performance Comparison

As shown in Fig. 13 and Table 5, the PPO model consistently achieves lower inrush current peaks and exhibits lower variance than the DQN-based approaches. This behaviour can be attributed to the different learning mechanisms of the two algorithms. PPO is a policy-gradient method that directly optimises the control policy through an actor–critic framework. In addition, the clipped surrogate objective (see Eq. (2)) constrains policy updates during training, preventing large policy shifts and promoting a smoother and more stable learning process (Fraija et al., 2024). As a result, PPO tends to converge more steadily, which explains the lower variability observed in the obtained switching strategies.

In contrast, DQN is a value-based algorithm that learns an action-value function and derives the policy indirectly through an ε-greedy exploration strategy. During the early stages of training, this mechanism requires frequent random actions to explore the environment, which often results in slower convergence and larger variability in the learned behaviour. Furthermore, the performance of DQN can be more sensitive to approximation errors in the value function, which may lead to unstable learning dynamics.

Another relevant factor is the size of the action space considered in the transformer energisation problem. In this study, the environment contains 360 possible switching actions corresponding to different closing instants. While DQN is typically effective in problems with small discrete action spaces, policy-gradient methods such as PPO are generally better suited for environments with larger action spaces, as they optimise the policy directly rather than evaluating each action separately (Tan, 2025). This characteristic enables PPO to explore the action space more efficiently and identify switching policies that achieve lower inrush current amplitudes.

In addition to the analytical results, statistical tests are conducted to further compare the performance of the PPO and DQN models. Specifically, a non-parametric Wilcoxon signed-rank test is employed (Colas et al., 2022), using 21 samples from each model. The test yielded a test statistic of $W = 43$ with a corresponding p-value of 0.0192. Here,

W represents the sum of ranks of the signed differences between paired observations, taken as the smaller of the positive and negative rank sums. A smaller W indicates that the observed differences are less likely to have arisen by chance. Moreover, the p-value falls below the conventional significance threshold of 0.05, suggesting that the observed difference is statistically significant. These results indicate that the superior performance of PPO is not only evident analytically but also statistically robust, providing evidence that PPO outperforms DQN in this study."

5. Discussion

The experimental results show that the proposed RL-based controlled switching strategy can significantly mitigate transformer inrush currents during energisation. In particular, the PPO controller reduces the peak inrush current by 79.78% compared with the classical uncontrolled switching, limiting the peak to approximately 0.18 pu. These results indicate that learning-based control can effectively identify favourable energisation instants under varying operating conditions and system states.

5.1. Comparative Assessment

To place the obtained results in context, the performance of the proposed strategy can be compared with other inrush mitigation approaches reported in the literature. These comparisons typically consider metrics such as inrush current reduction, maximum inrush current amplitude, and hardware requirements associated with each method.

Previous controlled-switching studies report that independent-pole CBs can reduce inrush current peaks to levels close to the steady-state no-load current near zero in 80% of the analysed cases. In contrast, three-pole controlled switching strategies generally limited peaks to below approximately 0.3 pu (Cano-González et al., 2017). Pre-fluxing techniques based on thyristor-controlled DC injection could maintain inrush current amplitudes below approximately 0.5 pu regardless of the initial residual flux condition. However, these approximations required additional hardware and more complex control systems (Sung and Kim, 2023).

Soft energisation approaches based on grid-forming inverters can further reduce inrush currents by gradually ramping the applied voltage (Alassi et al., 2022). However, these solutions are typically applicable only in converter-dominated systems and rely on specialised inverter control capabilities.

In comparison, the proposed RL-based strategy achieves competitive inrush mitigation while relying solely on the optimisation of switching instants of standard three-pole CBs. This reduces implementation cost and avoids the need for auxiliary equipment.

5.2. Practical Implications

Reducing inrush currents directly decreases the electromagnetic forces acting on transformer windings during energisation. These forces induce mechanical stresses and vibrations that contribute to cumulative insulation and structural degradation. Therefore, limiting current peaks is expected to reduce mechanical stress during switching events and support improved long-term transformer reliability.

From a system perspective, inrush current mitigation can also contribute to improved grid stability, particularly in weak networks. Large energisation currents may cause voltage dips and harmonic distortion, which can adversely affect nearby loads or distributed generation units. In the laboratory setup used in this study, the transformer is connected to a strong grid; therefore, the observed inrush events have a negligible impact on the overall system power quality. However, in weaker distribution networks, similar energisation events could result in more pronounced voltage disturbances.

The proposed approach also offers potential economic advantages. Traditional mitigation techniques often require hardware modifications, such as transformer oversizing or the installation of auxiliary devices, which could include pre-insertion resistors or pre-fluxing systems. In contrast, controlled switching optimises the operation of existing CBs without additional equipment, providing a simpler and potentially more cost-effective solution.

5.3. Limitations

Despite the promising results, certain limitations should be acknowledged. The proposed RL controller requires an offline training phase, which depends on the availability of a sufficiently accurate transformer model capable of replicating energisation dynamics.

While the results demonstrate significant inrush current mitigation, further evaluation under a wider range of grid conditions, transformer ratings, and network configurations is necessary to assess the robustness and generalisability of the proposed method.

In addition, the generalisation capability of the learned policy across different transformer designs and operating conditions has not been fully explored. Although the proposed framework can be retrained for new scenarios, further work is required to assess the transferability of the learned policy and reduce the need for retraining.

Finally, the performance of the RL controller may also be affected by model mismatch between the simulation environment used for training and real transformer behaviour. Although the model exhibits conservative behaviour, further investigation could help to quantify the impact of modelling inaccuracies on the learned policy.

6. Conclusions

This study presents an inrush current minimization strategy for power transformers based on Proximal Policy Optimization (PPO) Reinforcement Learning (RL) algorithm combined with controlled switching. The training objective is to learn the optimal circuit breaker (CB) closing angle based on the remanent fluxes at the transformer after its disconnection from the grid. The training of the algorithm is carried out by a power transformer model validated with inrush current tests performed in a real power transformer. The resulting PPO strategy is compared against two Deep Q-Network (DQN) variants (ϵ linear and exponential decays) and the typically used classical CB closing. The results show that both DQN with ϵ exponential decay and PPO can maintain inrush currents below 1 pu. On average, PPO reduces the inrush current peak by 14.29% compared to DQN with ϵ exponential decay and by 79.78% compared to classical switching. Overall, PPO provides the most stable, time-efficient, and accurate inrush minimization solution, achieving the lowest inrush current while requiring less computational time than DQN.

The proposed approach offers a novel, efficient, and adaptive RL-based solution for inrush current minimization. The suitability of the strategy has been evaluated with a focus on CB switching angles and remanent fluxes. While this scope has been demonstrated to be sufficient for the presented laboratory analysis, further research could investigate the impact of additional factors, such as different transformer ratings and CB characteristics, wider range of grid conditions, and uncertainties in remanent flux and hysteresis. The laboratory test results presented in this study could serve as a foundation for such future studies. Expanding the RL-based inrush current minimization approach therefore remains a task for future work.

Extending the proposed minimization technique to electrical power system-level studies introduces additional challenges. Although the transformer model used in this study accurately captures inrush current transients, its computational cost is high. Therefore, system-level analysis of inrush currents and other transients requires a computationally efficient and versatile transformer model. This challenge could be addressed in the future through the development of an appropriate transformer surrogate model.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was funded by the Spanish State Research Agency (grant No. CPP2021-008580 and grant No. PID2024-156284OA-I00) and the Department of Education of the Basque Government, Research Group Program (grant No. IT1634-22 and IT1504-22). In addition, J. I. Aizpurua is funded by the Ramón y Cajal Fellowship, Spanish State Research Agency (grant No. RYC2022-037300-I), co-funded by MCIU/AEI/10.13039/501100011033 and FSE+. Part of this work was undertaken by J. Ugarte Valdivielso at the University of Strathclyde, Scotland, UK, through an ERASMUS+ Student Mobility Award.

Code Availability

The code and simulation material will be available on the website: `https://doi.org/10.48764/zng9-ww32`.

References

Afrasiabi, S., Afrasiabi, M., Parang, B., Mohammadi, M., 2020. Designing a composite deep learning based differential protection scheme of power transformers. Applied Soft Computing Journal 87. doi:10.1016/j.asoc.2019.105975.

Afrasiabi, S., Afrasiabi, M., Parang, B., Mohammadi, M., Samet, H., Dragicevic, T., 2022. Fast grnn-based method for distinguishing inrush currents in power transformers. IEEE Transactions on Industrial Electronics 69, 8501–8512. doi:10.1109/TIE.2021.3109535.

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A next-generation hyperparameter optimization framework, in: KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2623–2631. doi:10.1145/3292500.333070.

Alassi, A., Ahmed, K.H., Egea-Alvarez, A., Foote, C., 2022. Transformer inrush current mitigation techniques for grid-forming inverters dominated grids. IEEE Transactions on Power Delivery doi:10.1109/TPWRD.2022.3218923.

Bui, V.H., Mohammadi, S., Das, S., Hussain, A., Hollweg, G.V., Su, W., 2025. A critical review of safe reinforcement learning strategies in power and energy systems. Engineering Applications of Artificial Intelligence 143. doi:10.1016/j.engappai.2025.110091.

Cano-González, R., 2015. Aportaciones a la Conexión Controlada de Transformadores de Potencia. Ph.D. thesis. Universidad de Sevilla.

Cano-González, R., Bachiller-Soler, A., Rosendo-Macías, J.A., Álvarez Cordero, G., 2017. Controlled switching strategies for transformer inrush current reduction: A comparative study. Electric Power Systems Research 145, 12–18. doi:10.1016/j.epsr.2016.11.018.

Chen, K., Li, J., Liu, K., Bai, C., Zhu, J., Gao, G., Wu, G., Laghrouche, S., 2024. State of health estimation for lithium-ion battery based on particle swarm optimization algorithm and extreme learning machine. Green Energy and Intelligent Transportation 3. doi:10.1016/j.geits.2024.100151.

Chen, K., Luo, Y., Long, Z., Wang, H., Li, Y., Gao, G., Wu, G., 2025. Battery state of health estimation using deep transfer learning on short-term charging data. Measurement: Journal of the International Measurement Confederation 256. doi:10.1016/j.measurement.2025.118233.

Chen, X., Qu, G., Tang, Y., Low, S., Li, N., 2022. Reinforcement learning for selective key applications in power systems: Recent advances and future challenges. IEEE Transactions on Smart Grid 13, 2935–2958. doi:10.1109/TSG.2022.3154718.

Chiesa, N., 2010. Power Transformer Modeling for Inrush Current Calculation. Ph.D. thesis. Norwegian University of Science and Technology: Thesis for the degree of Philosophiae Doctor.

Chiesa, N., Mork, B.A., Høidalen, H.K., 2010. Transformer model for inrush current calculations: Simulations, measurements and sensitivity analysis. IEEE Transactions on Power Delivery 25, 2599–2608. doi:10.1109/TPWRD.2010.2045518.

Choudhary, D., Mahanty, R.N., Kumar, N., 2025. Demand management of plug-in electric vehicle charging station considering bidirectional power flow using deep reinforcement learning. Engineering Applications of Artificial Intelligence 139. doi:10.1016/j.engappai.2024.109585.

Cigre, 2014. Transformer Energization in Power Systems: A Study Guide. Technical Report. Working Group C4.307.

Civelek, O., Gormus, S., İbrahim Okumus, H., Yilmaz, H., 2025. Addressing anomalies in smart grids: Methods for detecting and localizing high-current loads. Engineering Applications of Artificial Intelligence 159. doi:10.1016/j.engappai.2025.111223.

Colas, C., Sigaud, O., Oudeyer, P.Y., 2022. A hitchhiker's guide to statistical comparisons of reinforcement learning algorithms. URL: http://arxiv.org/abs/1904.06979.

Degrave, J., Felici, F., Buchli, J., Neunert, M., Tracey, B., Carpanese, F., Ewalds, T., Hafner, R., Abdolmaleki, A., de las Casas, D., Donner, C., Fritz, L., Galperti, C., Huber, A., Keeling, J., Tsimpoukelli, M., Kay, J., Merle, A., Moret, J.M., Noury, S., Pesamosca, F., Pfau, D., Sauter, O., Sommariva, C., Coda, S., Duval, B., Fasoli, A., Kohli, P., Kavukcuoglu, K., Hassabis, D., Riedmiller, M., 2022. Magnetic control of tokamak plasmas through deep reinforcement learning. Nature 602, 414–419. doi:10.1038/s41586-021-04301-9.

Fang, S., Ni, H., Lin, H., Ho, S.L., 2016. A novel strategy for reducing inrush current of three-phase transformer considering residual flux. IEEE Transactions on Industrial Electronics 63, 4442–4451. doi:10.1109/TIE.2016.2540583.

Fraija, A., Henao, N., Agbossou, K., Kelouwani, S., Fournier, M., Nagarsheth, S.H., 2024. Deep reinforcement learning based dynamic pricing for demand response considering market and supply constraints. Smart Energy 14. doi:10.1016/j.segy.2024.100139.

Garrido, J.M.C., Valdivielso, J.U., Aizpurua, J.I., Iñarra, M.B., Silveyra, J.M., 2026. Blind efficient method for optimizing jiles–atherton model parameters. IEEE Transactions on Magnetics 62. doi:10.1109/TMAG.2025.3632479.

Gopalakrishnan, V., Wu, B., Chhabria, V.A., 2024. Ml-insight: Machine learning for inrush current prediction and power switch network improvement, in: ISLPED '24: Proceedings of the 29th ACM/IEEE International Symposium on Low Power Electronics and Design, pp. 1–6. doi:10.1145/3665314.3670807.

Gu, Y., Cheng, Y., Chen, C.L.P., Wang, X., 2021. Proximal policy optimization with policy feedback. IEEE Transactions on Systems, Man, and Cybernetics: Systems 52, 4600 – 4610. doi:10.1109/TSMC.2021.3098451.

He, A., Jiao, Z., Li, Z., Liang, Y., 2023. Discrimination between internal faults and inrush currents in power transformers based on the discriminative-feature-focused cnn. Electric Power Systems Research 223. doi:10.1016/j.epsr.2023.109701.

Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., Meger, D., 2018. Deep reinforcement learning that matters, in: Proceedings of the AAAI conference on artificial intelligence, pp. 3207–3214. doi:10.48550/arXiv.1709.06560.

Huang, Q., Huang, R., Hao, W., Tan, J., Fan, R., Huang, Z., 2019. Adaptive power system emergency control using deep reinforcement learning. IEEE Transactions on Smart Grid 11, 1171 – 1182. doi:10.1109/TSG.2019.2933191.

Jiles, D., 1991. Introduction to Magnetism and Magnetic Materials. Springer US. doi:10.1007/978-1-4615-3868-4.

Jiles, D.C., Atherton, D.L., 1984. Theory of ferromagnetic hysteresis (invited). Journal of Applied Physics 55, 2115–2120. doi:10.1063/1.333582.

Jiles, D.C., Atherton, D.L., 1986. Theory of ferromagnetic hysteresis. Journal of Magnetism and Magnetic Materials 61, 48–60. doi:10.1016/0304-8853(86)90066-1.

Khooban, M.H., Gheisarnejad, M., 2021. A novel deep reinforcement learning controller based type-ii fuzzy system: Frequency regulation in microgrids. IEEE Transactions on Emerging Topics in Computational Intelligence 5, 689–699. doi:10.1109/TETCI.2020.2964886.

Li, Y., Gao, G., Chen, K., He, S., Liu, K., Xin, D., Luo, Y., Long, Z., Wu, G., 2025. State-of-health prediction of lithium-ion batteries using feature fusion and a hybrid neural network model. Energy 319. doi:10.1016/j.energy.2025.135163.

Mehta, C., Sant, A.V., Sharma, P., 2024. Optimized ann for lifepo4 battery charge estimation using principal components based feature generation. Green Energy and Intelligent Transportation 3. doi:10.1016/j.geits.2024.100175.

Mork, B.A., Gonzalez, F., Ishchenko, D., Stuehm, D.L., Mitra, J., 2007. Hybrid transformer model for transient simulation - part i: Development and parameters. IEEE Transactions on Power Delivery 22, 248–255. doi:10.1109/TPWRD.2006.883000.

MU, 2025. Mondragon university medium voltage laboratory. URL: https://www.mondragon.edu/en/research-transfer/engineering-technology/research-and-transfer-groups/medium-voltage-laboratory.

Murugan, S.K., Simon, S.P., Eapen, R.R., 2022. A novel signal localized convolution neural network for power transformer differential protection. IEEE Transactions on Power Delivery 37, 1242–1251. doi:10.1109/TPWRD.2021.3080927.

Peng, Q., Du, H., Zheng, Z., Zhu, H., Fang, Y., 2025. Prediction of magnetic fields in single-phase transformers under excitation inrush based on machine learning. Sensors 25. doi:10.3390/s25134150.

Ramirez, I., Aizpurua, J.I., Lasa, I., del Rio, L., 2024. Probabilistic feature selection for improved asset lifetime estimation in renewables. application to transformers in photovoltaic power plants. Engineering Applications of Artificial Intelligence 131, 107841. doi:https://doi.org/10.1016/j.engappai.2023.107841.

Sanati, S., Ahmadi, S.A., Sanaye-Pasand, M., 2024. An innovative harmonic-based approach for inrush current detection in low-loss transformers. IET Generation, Transmission and Distribution 18, 1303–1316. doi:10.1049/gtd2.13099.

Silva, F.L.D., Nishida, C.E., Roijers, D.M., Costa, A.H., 2020. Coordination of electric vehicle charging through multiagent reinforcement learning. IEEE Transactions on Smart Grid 11, 2347–2356. doi:10.1109/TSG.2019.2952331.

Silveyra, J.M., Gonzalez, M.I., Gonzalez, T.F., Garrido, J.M., 2024. Maganalyst: A matlab toolbox for anhysteretic magnetization analysis. IEEE Transactions on Magnetics 60. doi:10.1109/TMAG.2024.3408681.

Singh, P.K., Chaturvedi, D., 2015. Neural network based modeling and simulation for estimation of maximum transformer inrush current. International Journal of Computer Applications 123. doi:10.5120/ijca2015905331.

Stipetic, N., Filipovic-Grcic, B., Perkovic, M., 2023. Impact of autotransformer inrush currents on differential protection operation. Electric Power Systems Research 220. doi:10.1016/j.epsr.2023.109309.

Sung, B.C., Kim, S., 2023. Accurate transformer inrush current analysis by controlling closing instant and residual flux. Electric Power Systems Research 223. doi:10.1016/j.epsr.2023.109638.

Sutton, R.S., Barto, A.G., 2015. Reinforcement Learning: An Introduction. Second ed., A Bradford Book.

Tan, C., 2025. Comparative study of reinforcement learning performance based on ppo and dqn algorithms. Applied and Computational Engineering 175, 30–36. doi:10.54254/2755-2721/2025.ast24879.

Thomas, J.B., Shihabudheen, K.V., 2023. Neural architecture search algorithm to optimize deep transformer model for fault detection in electrical power distribution systems. Engineering Applications of Artificial Intelligence 120. doi:10.1016/j.engappai.2023.105890.

Ugarte, J., 2025. Transformer model data for inrush current minimization using rl. https://github.com/joneuga/Transformer-Model-Inrush.

Ugarte, J., Aizpurua, J.I., Barrenetxea, M., 2024. Uncertainty distribution assessment of jiles-atherton parameter estimation for inrush current studies. IEEE Transactions on Power Delivery 39, 2275–2285. doi:10.1109/TPWRD.2024.3398790.

Ugarte, J., Barrenetxea, M., Torres, A., G., B., Aizpurua, J.I., 2025. Power transformer modelling for the evaluation of a reinforcement learning-based nrush current minimization strategy, in: IEEE 8th International Advanced Research Workshop on Transformers (ARWtr), pp. 24–29.

Vaish, R., Dwivedi, U.D., Tewari, S., Tripathi, S.M., 2021. Machine learning applications in power system fault diagnosis: Research advancements and perspectives. doi:10.1016/j.engappai.2021.104504.

Valverde, V., Mazón, A.J., Zamora, I., Buigues, G., 2007. Ferroresonance in voltage transformers: Analysis and simulations. Renewable Energy and Power Quality Journal 1, 465–471. doi:10.24084/repqj05.317.

Wang, X., He, A., Li, Z., Jiao, Z., Lu, N., 2025. Reinforcement learning based early classification framework for power transformer differential protection. Expert Systems with Applications 292. doi:10.1016/j.eswa.2025.128632.

Wilson, C., Riccardi, A., 2023. Improving the efficiency of reinforcement learning for a spacecraft powered descent with q-learning. Optimization and Engineering 24, 223–255. doi:10.1007/s11081-021-09687-z.

Xu, H., Domínguez-García, A.D., Sauer, P.W., 2020. Optimal tap setting of voltage regulation transformers using batch reinforcement learning. IEEE Transactions on Power Systems 35. doi:10.1109/TPWRS.2019.2948132.

Yin, L., Gao, Q., Zhao, L., Zhang, B., Wang, T., Li, S., Liu, H., 2020. A review of machine learning for new generation smart dispatch in power systems. Engineering Applications of Artificial Intelligence 88. doi:10.1016/j.engappai.2019.103372.

Zandi, O., Poshtan, J., 2023. Voltage control of dc–dc converters through direct control of power switches using reinforcement learning. Engineering Applications of Artificial Intelligence 120. doi:10.1016/j.engappai.2023.105833.

A. Transformer Model

Fig. A.1 shows the duality-based electrical circuit of a three-phase, three-legged transformer with two windings.

R_{HV} and R_{LV} are the resistances of the high- and low-voltage windings. L_{HL} and L_{LC} are the leakage inductance between the high- and low-voltage windings and between the low-voltage winding and the core. L_0 is the zero-sequence inductance. L_{leg} and L_{yoke} are the non-linear inductances that represent the magnetic characteristics of the core legs and core yokes with the JA model. The implementation of the JA model consists of five physical parameters that need to be estimated. A more detailed explanation of the parameter estimation and JA model implementation into the simulation environment can be found in (Ugarte et al., 2024) and (Ugarte et al., 2025). The Δ and y connections, the phase shift, and the parasitic capacitance network are made externally connecting them to the transformer terminals identified with *a*, *b*, and *c* letters (uppercase for the primary winding and lowercase for the secondary winding).

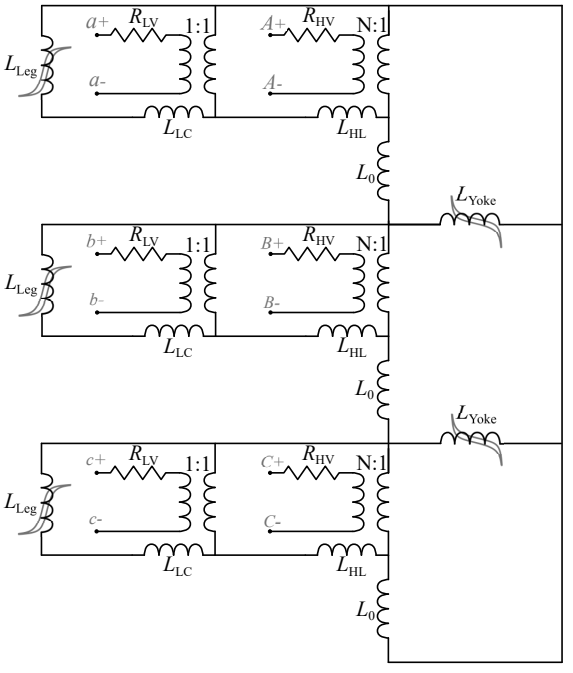


Fig. A.1: Duality-based transformer electrical circuit of a three-phase transformer (Chiesa et al., 2010).

B. Remanent Flux Data

The remanent flux values obtained from the voltage measurements at the laboratory are presented in Fig. B.1 as data points with respect to their opening angle. These data are approximated by curve-fitting sines, and their 95% Confidence Interval (CI) is also shown.

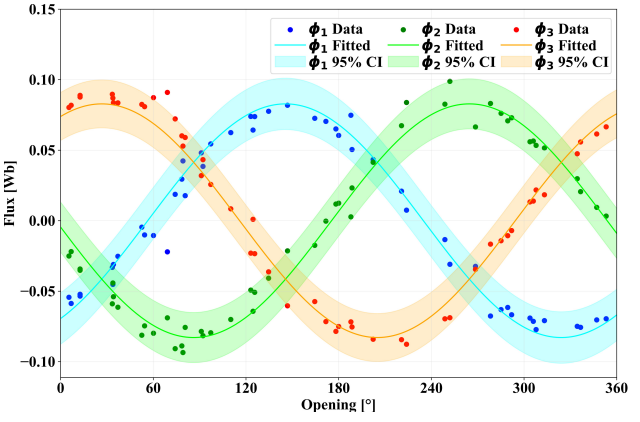


Fig. B.1: Remanent flux data obtained from the laboratory tests, its fitted curve, and 95% CI.

The equation that describes the evolution of the remanent fluxes concerning the opening angle of the CB is expressed as follows:

$$\phi = A \sin(\omega t + \psi) \pm \delta \tag{B.1}$$

where ϕ is the remanent flux in [Wb], and A is its amplitude in [Wb]. ω corresponds to $2\pi f$, where f is the 50 Hz frequency in [rad/s], t corresponds to time in [s], and ψ corresponds to the phase shift in [rad]. δ represents the tolerance between the data points and the fitted curve in [Wb]. The values of these parameters for the three remanent flux curves presented in Fig. B.1 are given in Table B.1.

Table B.1
Parameters for the fitted remanent flux curves.

Parameter	**Symbol**	ϕ_1	ϕ_2	ϕ_3	**Unit**
Amplitude	A	0.083	0.083	0.083	Wb
Phase Shift	ψ	-0.995	1.010	3.194	rad
Tolerance	δ	0.009	0.009	0.009	Wb