HQ-PIDT: A Hybrid Quantum Physics-Informed Decision Transformer for Resilient Distribution System Restoration Under Uncertainty

Hong Zhao[1], Jin Wei-Kocsis[1]*, Irfan Ullah[2], Karen L Butler-Purry[2]

[1]Purdue University, West Lafayette
[2]Texas A&M University, College Station
{zhao1211,kocsis0}@purdue.edu, {irfan,klbutler}@tamu.edu

Abstract

Distribution system restoration (DSR) requires sequential switching decisions that restore loads while satisfying physical and operational constraints. Deep reinforcement learning (DRL)-based methods have shown promise for DSR, but their data-intensive training process and limited long-horizon reasoning capability can restrict their effectiveness in large-scale and complex distribution systems. Decision-transformer (DT)-based methods provide an alternative by modeling restoration as a trajectory-conditioned sequence prediction problem. However, conventional transformer attention mechanisms still face challenges in extracting stable and informative trajectory representations from noisy and time-varying system observations. To address this challenge, this paper proposes a hybrid quantum physics-informed decision transformer (HQ-PIDT) for DSR under dynamic and uncertain operational conditions. The key innovation of HQ-PIDT is a hybrid quantum fidelity-based multi-head self-attention mechanism that embeds parameterized quantum circuits into query-key-value projection and uses quantum-state fidelity to compute attention similarity. This design enables restoration contexts to be represented and compared through quantum-state-overlap-based representations while preserving the autoregressive structure of causal transformers. Evaluations on three distribution-system test feeders show that HQ-PIDT achieves more reliable optimal or near-optimal restoration outcomes than a physics-informed DT baseline and two DRL baselines under constrained, dynamic, and noisy-input scenarios.

Introduction

Extreme events, such as natural disasters and severe weather events, have underscored the urgent need to enhance the resilience of modern power systems. Distribution systems play a critical role in delivering electricity to end users, and their ability to recover rapidly after disruptions is essential for maintaining overall grid resilience. A key resilience-enhancement task is distribution system restoration (DSR), which aims to restore out-of-service loads, especially critical loads, after disruptive events (Chen, Wang, and Ton 2017). One promising restoration strategy is to form microgrids (MGs) with dynamic boundaries by coordinating distributed generators (DGs), energy storage resources, renewable generation, and remotely controlled switches (Liang et al. 2022; Igder, Liang, and Mitolo 2022; Zhao and Wang 2022). Through sequential switching actions, a disrupted distribution system can be partitioned into multiple islanded MGs, thereby improving service restoration capability and maintaining power supply continuity to critical loads. However, practical DSR remains challenging because restoration decisions must satisfy physical and operational constraints while responding to time-varying load demand, renewable generation, and uncertain system observations.

Traditional DSR methods, such as mathematical programming (Yang et al. 2019; Bassey and Butler-Purry 2020), heuristic algorithms (Sedzro et al. 2019; Wang et al. 2023), and expert systems (Chen, Tsai, and Kuo 2005; Singh, Mehfuz, and Kumar 2016), have been widely studied. Although these methods have demonstrated effectiveness in many restoration settings, they often suffer from limited scalability and robustness, and generally lack the ability to adapt to dynamic environments. Deep machine learning methods, especially deep reinforcement learning (DRL), have emerged as powerful approaches for enabling more adaptive and efficient DSR decision making (Yao et al. 2020; Gao et al. 2020; Du and Wu 2022). However, DRL methods typically rely on Markov Decision Process (MDP) or partially observable MDP assumptions and require extensive training data, which limits their applicability to large-scale DSR scenarios that demand long-horizon temporal reasoning, especially when operating under data uncertainty which is common in general DSR settings.

Decision transformers (DTs) (Chen et al. 2021) provide an alternative learning paradigm for sequential decision-making problems by formulating control as trajectory-conditioned sequence modeling. Instead of learning a policy through online trial-and-error interaction, DT-based methods can learn from offline trajectories and use causal transformer architectures to capture dependencies among return-to-go (RTG) values, states, and actions (Fu et al. 2024; Hsu et al. 2024; Yao et al. 2025). This property makes DTs attractive for DSR, where restoration requires long-horizon sequential decision making under physical and operational constraints. In our prior work (Zhao et al. 2024), we introduced a physics-informed decision transformer (PIDT) for DSR and showed that DT-based modeling can improve restoration performance compared with conventional DRL baselines. However, conventional transformer atten-

*Corresponding author.

tion mechanisms rely primarily on classical query-key-value projection and dot-product similarity, which limit their ability to extract stable and informative trajectory representations when observed system states are noisy, corrupted, or time-varying. Quantum machine learning (QML), which integrates parameterized quantum circuits with trainable machine learning models, provides a promising direction for enriching representation learning in complex data-driven tasks (Qi et al. 2025). Parameterized quantum circuits can map classical inputs into high-dimensional Hilbert-space representations and process feature interactions through quantum circuit evolution. These properties motivate the use of hybrid quantum models for learning expressive trajectory representations under uncertainty. However, the integration of QML with decision-transformer-based DSR remains largely unexplored. To the best of our knowledge, this is the first study to investigate a hybrid quantum decision-transformer architecture for constrained DSR under time-varying operating conditions and observation uncertainty.

Motivated by these challenges and this research gap, this paper proposes a hybrid quantum physics-informed decision transformer (HQ-PIDT) for constrained DSR under time-varying operating conditions and observation uncertainty. HQ-PIDT formulates DSR as a trajectory-conditioned sequential decision-making problem while incorporating physical restoration objectives and constraints through physics-informed RTG conditioning and admissible-action checking. The key methodological innovation is a hybrid quantum fidelity-based multi-head self-attention mechanism. Instead of relying only on classical query-key-value projection and dot-product attention, the proposed attention module embeds parameterized quantum circuits into the representation projection stage and uses quantum-state fidelity to compute attention similarity between encoded query and key representations. This design enables restoration trajectory contexts to be represented and compared through quantum-state overlap while preserving the autoregressive structure and gradient-based trainability of causal transformers. The main contributions of this paper are threefold. First, we develop a hybrid quantum physics-informed decision-transformer framework for constrained DSR in dynamic and uncertain operational environments. Second, we design a hybrid quantum fidelity-based multi-head self-attention mechanism that introduces parameterized quantum circuits into representation projection and fidelity-based attention-score computation. Third, we conduct comprehensive evaluations on the IEEE 13-node test feeder, a modified IEEE 123-node test feeder, and a modified Iowa 240-node test feeder simulated in OpenDSS. Comparative studies against PIDT, PPO, and A2C demonstrate that HQ-PIDT achieves more reliable optimal or near-optimal restoration outcomes under constrained, dynamic, and noisy-input scenarios.

The remainder of this paper is organized as follows. Next section introduces the DSR problem setting and the physics-informed formulation. The following two sections describe the proposed HQ-PIDT framework and present the performance evaluation results on three distribution-system test feeders. Conclusions and future research directions are presented in the last section.

Problem Setting

In this work, we formulate DSR as a constrained sequential decision-making problem in which a control policy determines a sequence of switching actions to restore out-of-service loads after a disturbance (Zhao et al. 2024, 2025). At each decision step, the restoration controller observes the current operating condition of the distribution system, selects a feasible switching action, and updates the restoration trajectory. The objective is to maximize restored load over the restoration horizon while maintaining physical and operational feasibility.

We adopt a node-cell-based representation of the distribution system (Chen et al. 2019). A node cell is defined as a group of buses that are directly connected through non-switchable lines and therefore must be energized simultaneously. Under this representation, the DSR process can be viewed as the sequential energization of node cells rather than individual buses. During restoration, each available active energy source, such as a distributed generator (DG), is associated with a sequence of switching actions referred to as an energization branch. Each energization branch restores a set of node cells and forms an islanded microgrid. Different islanded microgrids are not electrically connected during the restoration process. In this formulation, the physical constraints include: 1) node voltage limits, 2) power-flow feasibility, 3) generation-capacity limits of controllable energy sources, and 4) source ramping limits. The operational constraints include: 1) prevention of loop formation and 2) ensuring that no node cell is visited more than once. The definitions of the parameters and variables used in this paper are summarized in Table 1.

Several key components of the DSR problem are formulated as follows.

- *Observed State:* The observation vector $\mathbf{s}_t \in \mathcal{S}$ represents the available observed state of the distribution system at time step t. In our problem formulation, $\mathbf{s}_t$ is constructed by concatenating two binary vectors, $\mathbf{s}_{1,t}$ and $\mathbf{s}_{2,t}$. The vector $\mathbf{s}_{1,t}$ indicates whether each node cell is selected for restoration at the current time step, while $\mathbf{s}_{2,t}$ indicates whether each node cell has already been restored by time step t. Therefore, $\mathbf{s}_t$ provides the decision model with both the current restoration choice and the accumulated restoration status of the distribution system.
- *Action:* The action vector $\mathbf{a}_t \in \mathcal{A}$ represents the control action applied to the operable switches at time step t. During DSR, the selected action determines which switching operation is executed to maximize load restoration while satisfying physical and operational constraints.
- *Time horizon:* The time horizon T defines the total number of decision-making steps in the restoration process.
- *Return-to-go:* The return-to-go (RTG) $\hat{R}_t$ is defined as

$$\hat{R}_t = \sum_{j=t}^{T} r_j, \tag{1}$$

which represents the cumulative reward from time step t to the end of the restoration horizon. The immediate re-

Table 1: Definition of variables and parameters

$s_t, \mathcal{S}$	State at time step t, the state space
$a_t, \mathcal{A}$	Action at time step t, the action space
R	Reward
$\mathcal{P}$	Transition probability
r_t	Reward at time step t
T	Time Horizon
L	Set of all the loads
N	Set of all buses
$C, \|C\|$	Set of all node cells, number of node cells
E	Set of all DGs (excluding PV systems)
PV	Set of all PV systems
$x_{l,t}^L, x_{i,t}^C$	Energization status of Load l at time step t, energization status of node cell i at time step t
$P_{l,t}^L, P_{i,t}^C, P_{l,t}$	Active power of Load l at time step t when restored, cumulative active power of all loads in node cell i at time step t when restored, nominal active power of Load l at time step t.
$Q_{l,t}^L$	Nominal Reactive power of Load l at time step t
$P_{pv,t}^{PV}$	Available active power of PV pv at time step t.
$P_{dg,t}^E$	Active output power of DG dg at time step t.
P_{dg}^{ramp}	Maximum absolute change of active power settings of DG dg
$H_{l,t}^L$	Squared voltage magnitude of Load l at time step t
$H_l^{\min}, H_l^{\max}$	Minimum and maximum squared nodal voltage of Load l
$V_{l,t}^{L,p}$	Voltage penalty function of Load l at time step t
$S_{dg,t}^{E,p}$	Ramp rate penalty function of DG dg at time step t
θ	Learnable parameters for the proposed HQ-PIDT

ward r_t evaluates the effectiveness of taking action $\mathbf{a}_t$ under the observed state $\mathbf{s}_t$. It is designed to support the objective of maximizing load restoration over the time horizon T while penalizing physical constraint violations. Operational constraints are enforced separately through an admissible-action checking function during decision making. Specifically, the reward is formulated as

$$r_t = R_t^A(\mathbf{s}_t, \mathbf{a}_t) + w_{p_1} R_t^V(\mathbf{s}_t, \mathbf{a}_t) + w_{p_2} R_t^E(\mathbf{s}_t, \mathbf{a}_t), \quad (2)$$

where $R_t^A(\mathbf{s}_t, \mathbf{a}_t)$ is the restored active-power reward, $R_t^V(\mathbf{s}_t, \mathbf{a}_t)$ is the voltage-violation penalty, $R_t^E(\mathbf{s}_t, \mathbf{a}_t)$ is the source ramp-rate penalty, and w_{p_1} and w_{p_2} are weighting coefficients used to balance the penalty terms with the restored-power objective.

The restored active-power reward is defined as

$$R_t^A(\mathbf{s}_t, \mathbf{a}_t) = \left(\sum_{l \in \mathcal{L}} x_{l,t}^L P_{l,t}^L \right) \Delta t. \quad (3)$$

By incorporating the node-cell representation, Eq. (3) can be rewritten as

$$R_t^A(\mathbf{s}_t, \mathbf{a}_t) = \left(\sum_{i \in \mathcal{C}} x_{i,t}^C P_{i,t}^C \right) \Delta t, \quad (4)$$

where $x_{i,t}^C$ denotes the energization status of node cell i, and $P_{i,t}^C$ denotes the cumulative active power of the loads within node cell i.

The voltage-violation penalty is defined as

$$R_t^V(\mathbf{s}_t, \mathbf{a}_t) = -\left(\sum_{l \in \mathcal{L}} x_{l,t}^L V_{l,t}^{L,p} \right) \Delta t, \quad (5)$$

where

$$V_{l,t}^{L,p} = \max\left(0, H_{l,t}^L - H_l^{\max}\right) + \max\left(0, H_l^{\min} - H_{l,t}^L\right). \quad (6)$$

where $V_{l,t}^{L,p}$ is the voltage-violation penalty associated with load l, $H_{l,t}^L$ is the squared voltage magnitude at the corresponding node, and $[H_l^{\min}, H_l^{\max}]$ defines the allowable squared-voltage range.

The source ramp-rate penalty is defined as

$$R_t^E(\mathbf{s}_t, \mathbf{a}_t) = -\left(\sum_{dg \in \mathcal{E}} S_{dg,t}^{E,p} \right) \Delta t, \quad (7)$$

where

$$\begin{aligned} S_{dg,t}^{E,p} &= \max\left(0, (P_{dg,t}^E - P_{dg,t-1}^E) - P_{dg}^{\text{ramp}}\right) \\ &\quad + \max\left(0, -P_{dg}^{\text{ramp}} - (P_{dg,t}^E - P_{dg,t-1}^E)\right). \end{aligned} \quad (8)$$

where $S_{dg,t}^{E,p}$ penalizes source-output changes that exceed the allowable ramping limit. Equivalently, the active-power change $P_{dg,t}^E - P_{dg,t-1}^E$ should remain within $[-P_{dg}^{\text{ramp}}, P_{dg}^{\text{ramp}}]$.

- *Operational feasibility:* Operational constraints are enforced through an admissible-action checking function during decision making. This function filters out switching actions that would violate operational requirements, such as loop prevention or repeated restoration of the same node cell. Therefore, the policy prediction is combined with an action-feasibility check to ensure that the executed restoration action remains operationally admissible.

Proposed HQ-PIDT Architecture

The proposed hybrid quantum physics-informed decision transformer (HQ-PIDT) is developed as an integrated framework for sequential and constraint-aware DSR under time-varying operating conditions and observation uncertainty.

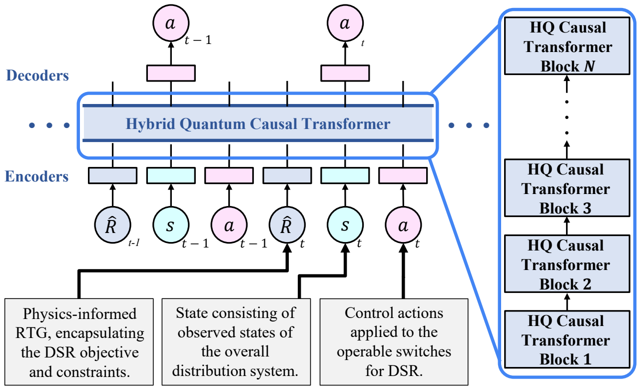


Figure 1: Overview of the architecture of our proposed HQ-PIDT framework.

HQ-PIDT models restoration as a trajectory-conditioned decision process over RTG, observed state, and action tokens, where the RTG encodes the physics-informed restoration objective and the causal transformer captures temporal dependencies among switching decisions. Beyond this decision-transformer formulation, HQ-PIDT redesigns the attention pathway by introducing hybrid quantum projection modules for query-key-value representation learning and fidelity-based quantum similarity for attention-score computation. This enables restoration contexts to be compared through quantum-state-overlap-based representations while preserving the autoregressive structure required for sequential decision making.

Hybrid Quantum Physics-informed Decision Transformer Architecture for DSR

Figure 1 illustrates the overall architecture of the proposed HQ-PIDT framework for DSR. As shown in the figure, HQ-PIDT consists of three main components: an encoder that maps trajectory elements into token embeddings, a hybrid quantum (HQ) physics-informed causal transformer that models temporal dependencies among restoration decisions, and a decoder that maps transformer outputs to control actions applied to the operable switches.

The encoder takes a trajectory of RTG values, observed states, and actions as input and maps them into token embeddings for the hybrid quantum causal transformer. At time step t, the input trajectory is represented as

$$\hat{\tau}_{t-K:t} = \{\hat{R}_{t-K}, \mathbf{s}_{t-K}, \mathbf{a}_{t-K}, \ldots, \hat{R}_{t-1}, \mathbf{s}_{t-1}, \mathbf{a}_{t-1}, \hat{R}_t, \mathbf{s}_t\}, \tag{9}$$

where K denotes the context length. Building on the DSR formulation described previously, HQ-PIDT treats the RTG, observed state, and action as distinct token types in the restoration trajectory. The RTG token provides physics-informed conditioning because it is computed from the reward function, which incorporates the restored-power objective and physical-constraint penalties. The state and action tokens encode the restoration context and previous switching decisions used for autoregressive action prediction.

The HQ physics-informed causal transformer is the central sequence-modeling component of HQ-PIDT. As shown in Fig. 1, it consists of a stack of HQ causal transformer blocks. Each block preserves the residual and feed-forward structure of a GPT-style causal transformer while incorporating the proposed hybrid quantum attention mechanism into the attention pathway. The internal structure of each hybrid quantum causal transformer block is detailed in the next subsection. This design allows HQ-PIDT to retain the autoregressive sequence-modeling capability of causal transformers while using fidelity-based hybrid quantum attention to capture restoration trajectory dependencies under observation uncertainty.

After the trajectory elements in $\hat{\tau}_{t-K:t}$ are projected into the embedding space, the HQ causal transformer processes the token sequence and produces contextualized trajectory representations. The decoder then maps the transformer output associated with the current decision step to the original switching-action space. The resulting policy is represented as

$$\mathbf{a}_t = \hat{\pi}_\theta(\hat{\tau}_{t-K:t}), \tag{10}$$

where $\hat{\pi}_\theta$ denotes the policy parameterized by the proposed HQ-PIDT model. During training, the decoder output is used to compute the cross-entropy loss between the predicted action distribution and the ground-truth action labels from sampled offline restoration trajectories. During inference, the predicted action is treated as the candidate switching action and is combined with the admissible-action checking function to ensure operational feasibility.

Hybrid Quantum Causal Transformer Block

We now introduce the hybrid quantum causal transformer block used in the HQ physics-informed causal transformer. The complete structure is shown in Fig. 2. The block follows a pre-layer-normalized transformer architecture and consists of a HQ multi-head self-attention module, residual connections, and a position-wise feed-forward network. The HQ multi-head self-attention module is the key quantum component of the block and is detailed in the next subsection. Given an input sequence representation $\mathbf{X} \in \mathbb{R}^{B\times S\times E}$, where B denotes the batch size, S denotes the sequence length, and E denotes the embedding dimension, the input is first normalized and then passed through the hybrid quantum self-attention module. A residual connection is applied as

$$\hat{\mathbf{X}} = \mathbf{X} + \mathrm{HQSA}(\mathrm{LN}_1(\mathbf{X})), \tag{11}$$

where $\mathrm{HQSA}(\cdot)$ denotes the hybrid quantum self-attention operation and $\mathrm{LN}_1(\cdot)$ denotes the first layer-normalization operation.

A second layer-normalization operation is then applied before the position-wise feed-forward network, followed by another residual connection:

$$\mathbf{Y} = \hat{\mathbf{X}} + \mathrm{MLP}(\mathrm{LN}_2(\hat{\mathbf{X}})), \tag{12}$$

where $\mathbf{Y}$ is the output of the hybrid quantum causal transformer block. The feed-forward network follows a GPT-style transformer block design and consists of a linear expansion from E to $4E$, a GELU activation function, a linear projection back to E, and dropout regularization.

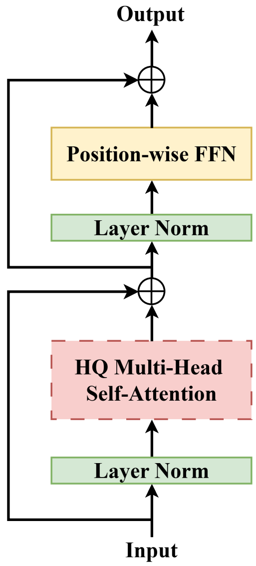


Figure 2: Architecture of the proposed HQ causal transformer block. The hybrid quantum multi-head self-attention module is marked by dashed borders.

This design preserves the standard causal transformer block structure while incorporating the proposed hybrid quantum attention mechanism into the attention pathway. As a result, the block remains compatible with autoregressive sequence modeling and gradient-based training, while enabling fidelity-based hybrid quantum representation learning for restoration trajectories under observation uncertainty.

Hybrid Quantum Fidelity-Based Multi-Head Self-Attention

This subsection details the hybrid quantum self-attention operation $\mathrm{HQSA}(\cdot)$ used in the HQ causal transformer block. The attention mechanism is the central component through which HQ-PIDT identifies relevant information from a restoration trajectory. In dynamic and uncertain DSR operational environments, system conditions may evolve over time due to time-varying load demand and renewable generation, while the measured or estimated system states may also deviate from the true operating states due to sensor noise, measurement errors, communication-induced data corruption, or potential malicious data manipulation. These factors make it challenging for a sequential decision model to extract stable and informative trajectory features. To address this challenge, we draw inspiration from the high-dimensional Hilbert-space representations induced by parameterized quantum circuits, which provide a rich nonlinear feature space for learning robust feature representations from noisy and time-varying classical trajectory inputs.

As shown in Fig. 3, we integrate classical neural processing and quantum circuit operations to realize a HQ-classical multi-head self-attention module. The proposed module follows the overall workflow of classical multi-head self-attention, including query-key-value projection, attention-score computation, attention-weight normalization, and value aggregation. Within this hybrid workflow, parameterized quantum circuits are embedded into both the representation projection stage and the attention-score computation stage. In the representation projection stage, classical trajectory features are processed through neural preprocessing, parameterized quantum circuit evolution, quantum measurement, and neural postprocessing to generate hybrid quantum query, key, and value representations. In the attention-score computation stage, the query and key representations are further encoded into quantum states, and their fidelity is used to define the attention similarity. The resulting fidelity scores are row-normalized to obtain attention weights, so that the attention weights are governed by the overlap between encoded quantum states rather than by classical dot-product similarity. These attention weights are then applied to the value representations for weighted value aggregation. In this way, HQ-PIDT learns trajectory representations through a fidelity-based HQ attention mechanism while remaining compatible with gradient-based training.

Hybrid Quantum Query-Key-Value Projection The proposed HQ multi-head self-attention module introduces a HQ-classical projection mechanism to enrich the representation projection stage. Specifically, classical token features are first processed by neural preprocessing and then passed through a parameterized quantum circuit (PQC), which consists of angle encoding, trainable variational quantum transformation, and Pauli-Z measurement. The measured quantum features are then mapped through neural postprocessing to generate the query (Q), key (K), and value (V) representations.

Let $\mathbf{x} \in \mathbb{R}^{d_{\text{in}}}$ denote the classical input vector associated with one input token. The hybrid quantum projection module maps $\mathbf{x}$ into an output vector in $\mathbb{R}^{d_{\text{out}}}$. To achieve this, the module first performs preprocessing by applying a feed-forward neural network (FFN) $f_{\text{pre}}(\cdot)$ to transform the input vector into a vector whose dimension equals the number of qubits.

$$\mathbf{z} = f_{\text{pre}}(\mathbf{x}). \tag{13}$$

The resulting vector is then bounded and rescaled into an angle range $(-\pi/2, \pi/2)$ so that each component can be interpreted as a quantum rotation angle:

$$\tilde{\mathbf{z}} = \frac{\pi}{2}\tanh(\mathbf{z}), \tag{14}$$

where $\tanh(\cdot)$ is applied element-wise. The bounded range is important because each component of $\tilde{\mathbf{z}}$ is used as a rotation angle in the quantum circuit.

Next, $\tilde{\mathbf{z}}$ is processed by the PQC. The PQC consists of three main operations: angle encoding, trainable variational quantum transformation, and Pauli-Z measurement. First, $\tilde{\mathbf{z}}$ is encoded into a quantum state using angle encoding. Specifically, each component $\tilde{z}_r$ is applied to the corresponding qubit through a rotation around the Y-axis of the Bloch sphere, as illustrated in Fig. 4. This process can be represented as

$$|\phi(\mathbf{x})\rangle = \bigotimes_{r=1}^{n_1} R_Y(\tilde{z}_r)|0\rangle, \tag{15}$$

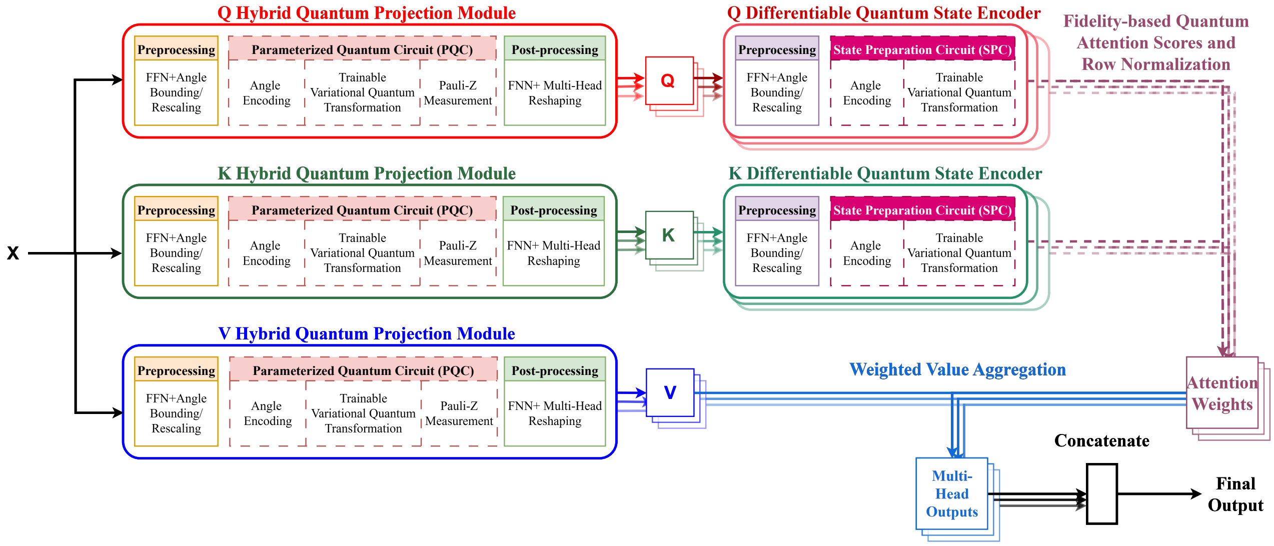


Figure 3: Overview of the architecture of the proposed hyrbid quantum fidelity-based multi-head self-attention. Dashed borders indicate quantum circuit components or quantum modules.

where $\bigotimes$ denotes the tensor product of single-qubit states, and n_1 is the number of qubits. For a single qubit, the R_Y rotation is given by

$$R_Y(\alpha) = e^{-i\frac{\alpha}{2}Y} = \begin{bmatrix} \cos(\alpha/2) & -\sin(\alpha/2) \\ \sin(\alpha/2) & \cos(\alpha/2) \end{bmatrix}. \quad (16)$$

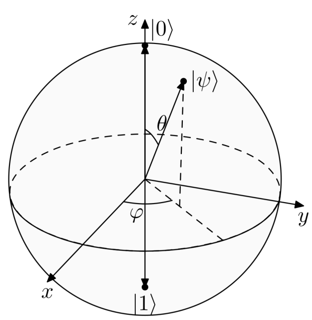


Figure 4: Bloch sphere illustration of single-qubit angle encoding through rotation gates.

After angle encoding, the projection module applies a trainable variational quantum transformation. In the implementation, this transformation is realized using L strongly entangling layers. Each layer contains parameterized single-qubit rotation gates, such as R_X, R_Y, and R_Z, together with entangling gates, such as CNOT or CZ. Let Θ denote the collection of trainable quantum parameters, the variational transformation can be written as

$$|\psi(\mathbf{x};\Theta)\rangle = U_\Theta|\phi(\mathbf{x})\rangle, \quad (17)$$

where U_Θ denotes the quantum transformation parameterized by Θ. The strongly entangling structure enables interactions among feature representations encoded across different qubits. The projection module then measures Pauli-Z observables on each qubit. Each measurement produces a scalar feature, resulting in an n_1-dimensional real-valued feature vector $\mathbf{h} = [h_1, h_2, \ldots, h_{n_1}]$. For the r-th qubit, the measured feature is

$$h_r = \langle\psi(\mathbf{x};\Theta)|Z_r|\psi(\mathbf{x};\Theta)\rangle. \quad (18)$$

Finally, the measured quantum feature vector is passed through a neural postprocessing network, $\mathbf{y} = f_{\text{post}}(\mathbf{h})$. The postprocessing network is implemented as a two-layer FFN and maps the n_1-dimensional measurement vector to the desired output dimension, i.e., $\mathbf{y} \in \mathbb{R}^{d_{\text{out}}}$. The resulting output representations are then reshaped into multiple attention heads for subsequent attention-score computation and value aggregation.

The above formulation describes the hybrid quantum projection for a single input token. When the input is a sequence tensor $\mathbf{X} \in \mathbb{R}^{B\times S\times d_{\text{in}}}$, where B is the batch size and S is the sequence length, the hybrid quantum projection module applies the same preprocessing, PQC transformation, measurement, and postprocessing procedure to each token. In our implementation, the input tensor is first reshaped into a flattened token batch for quantum-circuit evaluation and is then reshaped back to recover the original batch and sequence dimensions. The resulting postprocessed representation has shape $B \times S \times d_{\text{out}}$.

For the sequence tensor input $\mathbf{X}$, three independent hybrid quantum projection modules are instantiated to generate the query, key, and value representations:

$$\mathbf{Q} = \Phi_Q(\mathbf{X}), \qquad \mathbf{K} = \Phi_K(\mathbf{X}), \qquad \mathbf{V} = \Phi_V(\mathbf{X}), \quad (19)$$

where each projection module has its own neural preprocessing parameters, variational quantum parameters, and neural postprocessing parameters. The resulting $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ representations each have shape $B \times S \times d_{\text{out}}$ before multi-head reshaping. For H attention heads, each representation is then reshaped as

$$B \times S \times d_{\text{out}} \rightarrow B \times H \times S \times D, \qquad D = \frac{d_{\text{out}}}{H}. \quad (20)$$

Through this hybrid quantum projection design, the proposed attention module introduces quantum circuit dynamics into the representation projection stage of multi-head self-attention while preserving compatibility with gradient-based neural-network training.

Fidelity-Based Quantum Attention Scores After the query, key, and value representations are generated, the proposed module computes attention scores through a fidelity-based quantum similarity mechanism. Specifically, the $\mathbf{Q}$ and $\mathbf{K}$ representations generated by the hybrid quantum projection modules are further mapped into parameterized quantum states using differentiable quantum state encoders. The overlap between the encoded query and key states is then evaluated through quantum-state fidelity and used to define the attention similarity. In this way, query-key relevance is evaluated in a hybrid quantum representation space, where attention scores are determined by quantum-state overlap.

The differentiable quantum state encoder has a structure analogous to the hybrid quantum projection module described in the previous subsection, but it outputs a quantum state rather than measured classical features. Specifically, for a vector $\mathbf{u} \in \mathbb{R}^D$ generated by the hybrid quantum projection module, where $\mathbf{u}$ denotes either a query or key representation from one attention head, the encoder first applies neural preprocessing and angle rescaling, following Eq. (14), to obtain the corresponding angle vector $\tilde{\mathbf{z}}$. The angle vector is then processed by a state preparation circuit (SPC), where it is encoded into a quantum state using angle encoding, as in Eq. (15), and transformed by trainable variational quantum layers, as in Eq. (17). The resulting quantum state is denoted by $|\psi(\mathbf{u};\theta)\rangle$, where θ represents the trainable parameters of the quantum state encoder. For N input vectors processed in parallel, the encoder produces a complex-valued state-vector tensor with shape $N \times 2^{n_2}$, where n_2 is the number of qubits used in the quantum state encoder.

In the proposed design, separate quantum state encoders are used for the query and key representations. Given a query vector $\mathbf{q}_i$ from $\mathbf{Q}$ and a key vector $\mathbf{k}_j$ from $\mathbf{K}$, the corresponding encoded quantum states are denoted by $|\psi_Q(\mathbf{q}_i)\rangle$ and $|\psi_K(\mathbf{k}_j)\rangle$, respectively. The attention similarity between the two tokens is defined using quantum-state fidelity. The fidelity score is defined as the squared magnitude of the inner product between two normalized quantum states:

$$F_{ij} = \left|\langle \psi_Q(\mathbf{q}_i) | \psi_K(\mathbf{k}_j) \rangle\right|^2, \quad i, j = 1, 2, \ldots, S. \quad (21)$$

For normalized quantum states, $F_{ij} \in [0, 1]$, where a value close to 1 indicates that the encoded query and key states are highly similar, while a value close to 0 indicates that they are nearly orthogonal. In this way, the differentiable quantum state encoders map classical query and key representations into quantum states whose geometric overlap defines the attention similarity. The resulting fidelity scores are then used to compute the attention weights.

In this work, the fidelity scores are computed using a quantum simulation backend with access to statevectors, which enables direct evaluation of quantum-state overlap. On physical quantum hardware, full statevectors are not directly observable. Therefore, implementing the same fidelity-based attention mechanism on hardware would require overlap-estimation protocols, such as swap tests, destructive swap tests, or related fidelity-estimation circuits.

Attention-Weight Normalization and Value Aggregation After the fidelity-based quantum attention scores are obtained, they are normalized to produce attention weights. Instead of applying a softmax function, the proposed module uses row-wise normalization of the fidelity-score matrix. This choice is motivated by the nature of quantum-state fidelity. For normalized pure quantum states, the fidelity score can be interpreted as the probability of obtaining one quantum state when the other state is measured in a basis containing the target state. Therefore, the fidelity scores are already nonnegative, bounded, and overlap-based similarity measures.

Applying a softmax function, as commonly done in standard scaled dot-product attention, would introduce an additional exponential transformation to these bounded similarity values, which may alter their relative scale and weaken their direct connection to quantum-state fidelity. In contrast, row-wise normalization preserves the proportional relationship among the fidelity scores while converting them into valid attention weights that sum to one across the key positions. This choice allows the attention distribution to remain directly governed by the quantum-state overlap structure: tokens whose encoded quantum states have larger overlap receive proportionally larger weights, while dissimilar or nearly orthogonal states contribute less. In this sense, row-wise normalization provides a natural bridge between the fidelity-based quantum similarity measure and the probabilistic weighting required by the proposed HQ attention mechanism.

For each attention head, let $\mathbf{F} \in \mathbb{R}^{S \times S}$ denote the fidelity-score matrix and $\mathbf{A} \in \mathbb{R}^{S \times S}$ denote the corresponding attention-weight matrix. The (i, j)-th element of $\mathbf{A}$ is computed as

$$A_{ij} = \frac{F_{ij}}{\sum_k F_{ik}}. \quad (22)$$

When an attention mask is applied, masked positions are assigned zero similarity before normalization. The value aggregation follows the standard multi-head self-attention operation. The normalized attention-weight matrix is applied to the value representation:

$$\mathbf{O} = \mathbf{A}\mathbf{V}. \quad (23)$$

The outputs from all attention heads are then concatenated and passed through a final linear transformation to obtain the output of the proposed HQ attention module.

Offline Training and Sequential Restoration Inference

HQ-PIDT is trained using offline DSR trajectories. Each trajectory consists of a sequence of RTG values, observed states, and switching actions. These trajectories can be generated from domain-expert restoration strategies, optimization-based restoration solutions, or simulated restoration rollouts. During training, minibatches of trajectory segments with context length K are sampled from the offline dataset and fed into HQ-PIDT. The model predicts the switching action at each decision step, and the learnable parameters are updated by minimizing the cross-entropy loss between the predicted action distribution and the ground-truth action labels. The overall training procedure is summarized in Algorithm 1.

During inference, the trained HQ-PIDT model generates switching actions sequentially. At each time step, the current trajectory context is used to predict a candidate switching action. The action is then checked by the admissible-action function to ensure operational feasibility. After the action is executed, the reward, observed state, and return-to-go are updated, and the trajectory context is refreshed by keeping only the most recent K time steps. The overall inference procedure is summarized in Algorithm 2.

Algorithm 1: Training Procedure of HQ-PIDT

Initialize: Dataset D, HQ-PIDT model with learnable parameter θ, sample size L, maximum number of episodes M, and minibatch size b.

for episode $m = 1$ to M **do**
  Sample a random minibatch B of b sequences of length K from D.
  Obtain predictions using HQ-PIDT from the batch B.
  Calculate the cross-entropy loss between predicted and ground-truth values of action in sequence.
  Update θ.
end for

During inference, the trained HQ-PIDT model generates switching actions sequentially. At each time step, the current trajectory context is used to predict a candidate switching action. The action is then checked by the admissible-action function to ensure operational feasibility. After the action is executed, the reward, observed state, and RTG are updated, and the trajectory context is refreshed by keeping only the most recent K time steps. The overall inference procedure is summarized in Algorithm 2.

Algorithm 2: Inference of the Trained HQ-PIDT Model

Initialize: HQ-PIDT model with trained parameter θ, initial return-to-go $\hat{R}_0$, and initial state of the distribution system s_0.

Set $\hat{R}_1 \leftarrow \hat{R}_0, s_1 \leftarrow s_0, \tau \leftarrow (s_1, \hat{R}_1)$
for time step $t = 1$ to $T - 1$ **do**
  Obtain $a_t \leftarrow HQ - PIDT(\tau)$ and $r_t \leftarrow R_t(s_t, a_t)$
  Observe the next state s_{t+1} after taking switching action a_t, and obtain $\hat{R}_{t+1} \leftarrow \hat{R}_t - r_t$.
  Append $(a_t, s_{t+1}, \hat{R}_{t+1})$ to τ, and keep the last K time steps of τ (i.e., $\tau \leftarrow \tau_{-K:t+1}$).
end for

Performance Evaluation

In this section, we evaluate the performance of the proposed HQ-PIDT framework using three distribution system testbeds simulated in the Open Distribution System Simulator (OpenDSS): the IEEE 13-node test feeder (IEEE 2014; OpenDSS Project 2017), a modified IEEE 123-node test feeder (IEEE 2014; Chen et al. 2019), and a modified Iowa 240-node test feeder (Bu et al. 2019; Ullah et al. 2025). To assess its effectiveness in DSR tasks, we compare HQ-PIDT with one DT-based baseline and two DRL baselines. The DT-based baseline is the physics-informed decision transformer (PIDT) developed in our previous work (Zhao et al. 2024). PIDT is selected as the DT-based baseline because it addresses the same DSR task using a physics-informed DT structure, but does not include the proposed hybrid quantum attention mechanism. The DRL baselines include Proximal Policy Optimization (PPO) (Schulman et al. 2017) and Advantage Actor-Critic (A2C) (Mnih et al. 2016).

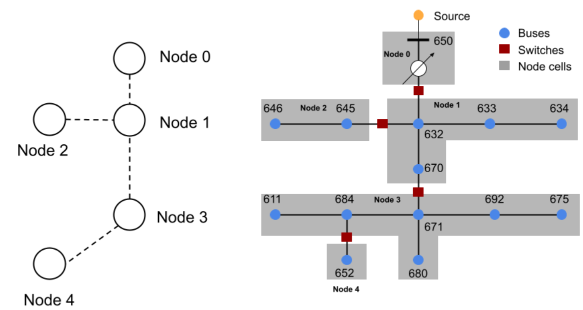
Table 2 summarizes the hyperparameters and model settings used for HQ-PIDT, PIDT, PPO, and A2C. For A2C and PPO, we use three-layer MLPs with a hidden-layer size of 64. The PIDT, PPO, and A2C models, as well as the classical components of HQ-PIDT, are implemented using PyTorch. The quantum circuits and quantum operations in HQ-PIDT are simulated using the PennyLane default qubit device (Xanadu 2026).

Table 2: Hyperparameters and model settings used in implemented methods.

Source	A2C	PPO	PIDT	HQ-PIDT
Number of qubits used	-	-	-	PQC: 4 SPC: 3
Number of strongly entangling layers	-	-	-	PQC: 1 SPC: 1
Number of transformer blocks	-	-	6	6
Number of attention heads	-	-	8	8
Transformer embedding dimension	-	-	128	128
Activation function	Tanh	Tanh	Tanh	GELU/Tanh
Optimizer	RMSprop	RMSprop	AdamsW	AdamsW
Discount factor	0.99	0.99	-	-

IEEE 13-Node Test Feeder

The system topology and its corresponding node-cell-based graph representation are shown in Fig. 5 (a) and (b), respectively. As illustrated in Fig. 5 (a), this system includes a substation located at source bus 650, with Bus 670 that connects with Bus 632.



(a) (b)

Figure 5: (a): The physical topology of the IEEE 13-node test feeder; and (b): The node-cell-based graph representation of the IEEE 13-node test feeder.

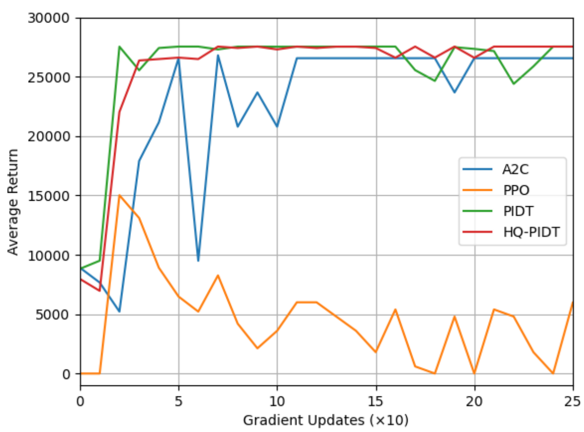


Figure 6: Learning curves for the average returns of the four methods for the DSR operations in the 13-node test feeder. The curves are updated every ten gradient updates.

In this test case, we use the energization-path version of the operational constraints to study the DSR problem. Due to the relatively simple topology of this test feeder, the ground-truth energization path is determined as Node Cell 0 $\rightarrow$ Node Cell 1 $\rightarrow$ Node Cell 3 $\rightarrow$ Node Cell 4. The corresponding final restored power demand is 3006.509 kW, which is used as the restoration objective. We set the time horizon to $T = 10$ for this test case. The evaluation results of HQ-PIDT, PIDT, A2C, and PPO across 50 independent inference trials are stated in Table 3. The methods are compared using average return, standard deviation of returns (Std. Return), and the number of optimal solutions (# Opt. Sols.). In addition, Fig. 7 presents the distribution of restored power levels across the 50 independent trials, providing further insight into the restoration performance of each method.

From Figs. 6 and 7 and Table 3, we can observe that, during the learning process, our HQ-PIDT method shows a faster convergence rate and achieves a higher convergence average return than the two DRL methods, while showing similar convergence rate and similar convergence average return than the PIDT method. However, our HQ-PIDT method has more stable learning behavior than PIDT method.

Table 3: Performance comparison of HQ-PIDT, PIDT, A2C, and PPO for the DSR operation in the IEEE 13-node test feeder.

Method	Average Return	Std. Return	# Opt. Sols.
A2C	26557.891	0.000	0
PPO	5771.989	396.000	0
PIDT	24511.205	6416.281	42
HQ-PIDT	25294.790	6070.172	44

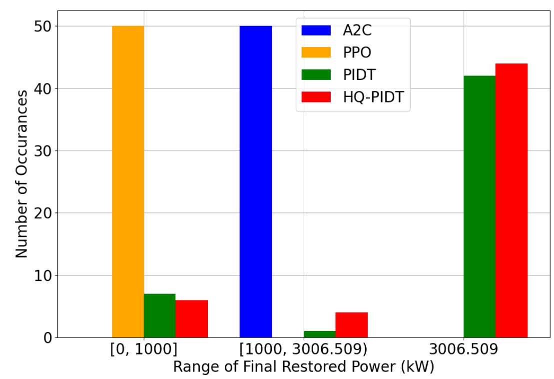
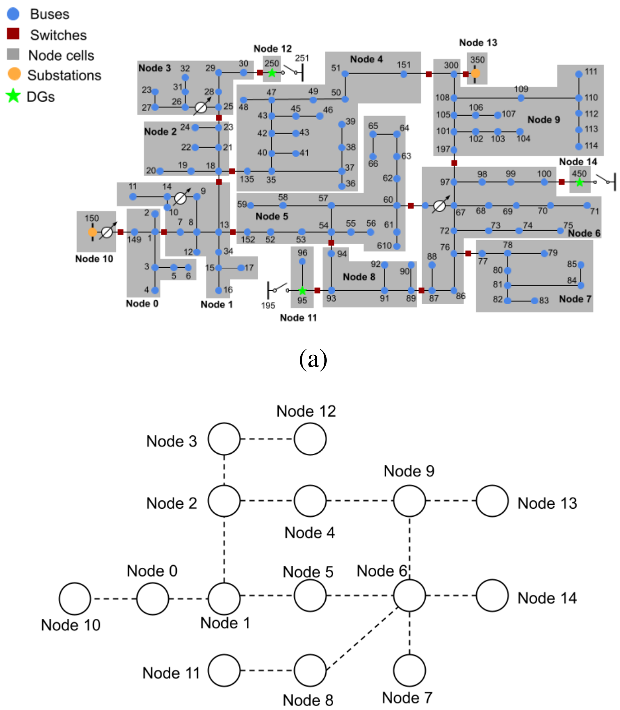


Figure 7: Distribution of power restoration levels across 50 independent inference trials using the four methods in the 13-node test feeder.

Additionally, from Figs. 7 and Table 3 we can observe that A2C consistently converges to suboptimal solutions ranging between 1000 kW and 3006.509 kW across the 50 trials, while PPO shows convergence to suboptimal solutions ranging below 1000 kW, and therefore fails to achieve any optimal solutions in all 50 trials. Furthermore, when comparing HQ-PIDT and PIDT methods in the number of optimal solutions, we see that our HQ-PIDT method achieves more optimal solutions than the PIDT method.

Modified IEEE 123-Node Test Feeder

The system topology and the corresponding node-cell-based graph representation are shown in Fig. 8 (a) and (b), respectively. As shown in Fig. 8(a), the modified IEEE 123-node test feeder includes five energy sources: two substations located at Buses 150 and 350, and three DGs located at Buses 95, 250, and 450. To evaluate the constraint-handling capability of the methods, we modify the operating capability of the DG at Bus 250 so that it is allowed to fully energize only Node Cells 2 and 3. If this DG attempts to energize any other node cell beyond these two node cells, it is assumed to shut down automatically, causing the voltage of the corresponding energization branch to drop to zero. In this test case, the time horizon is set to $T = 50$.



(a)

(b)

Figure 8: (a): The physical topology of the modified IEEE 123-node test feeder; and (b): The node-cell-based graph representation of the modified IEEE 123-node test feeder.

The total load demand of the system is approximately 3350 kW when all loads are properly energized, which is used as the restoration objective. The learning curves over the first 2000 gradient updates for HQ-PIDT, PIDT, A2C, and PPO are shown in Fig. 9. Further evaluation results across 50 independent inference trials are presented in Table 4. In these evaluations, we adopt five metrics: average return, standard deviation of returns (Std. Return), average power restoration (APR), standard deviation of power restoration (SDPR), and the number of optimal solutions (# Opt. Sols.). As shown in Fig. 9, for the modified IEEE 123-node test feeder with a more complex and larger-scale topology, HQ-PIDT achieves a convergence behavior comparable to PIDT, although both DT-based methods converge more slowly than the two DRL baselines during early training. However, the final inference results in Table 4 show that the proposed HQ-PIDT achieves substantially more optimal solutions than PIDT, PPO, and A2C across the 50 independent trials.

As shown in Fig. 9, for the modified IEEE 123-node test feeder with a more complicated and larger-scale topology, our method has convergence speed comparable to the PIDT, even though our method shows slower convergence speed to the two DRL methods. Despite the convergence speed, from Table 4, our proposed HQ-PIDT achieves much more optimal solutions in 50 independent trials than PIDT, PPO and A2C.

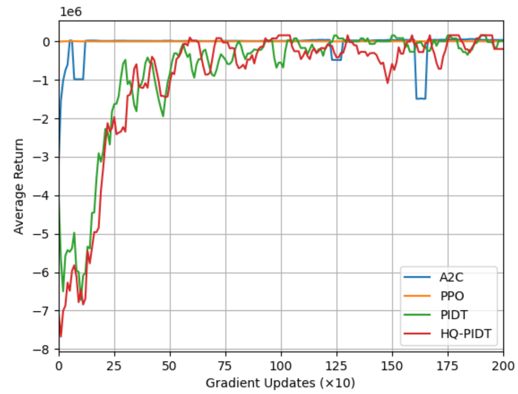


Figure 9: Learning curve for the average return of the first 2000 gradient updates using the four methods for the DSR operations in the modified 123-node test feeder. The curves are updated every ten gradient updates and smoothed using a moving average window of five data points.

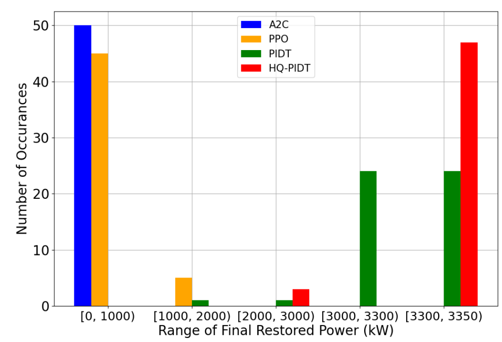


Figure 10: Distribution of power restoration levels across 50 independent trials using the four methods in the modified 123-node test feeder.

Table 4: Performance comparison between the four methods for the DSR operation in the modified IEEE 123-node test feeder.

Method	Average Return	Std. Return	APR (kW)	SDPR (kW)	# Opt. Sols.
A2C	45260.175	836.397	910.083	17.069	0
PPO	21079.395	12362.029	531.343	309.130	0
PIDT	158639.402	11846.575	3227.241	227.081	24
HQ-PIDT	162326.169	7852.889	3309.601	160.290	47

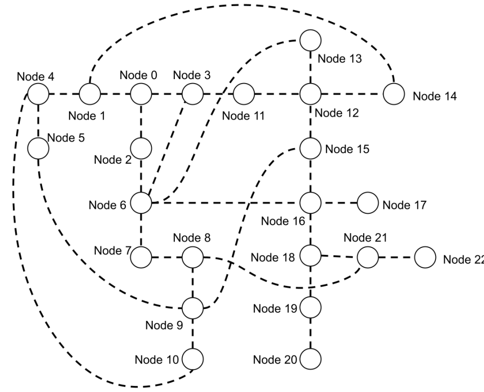
Furthermore, Fig. 10 provides additional insight into the simulation results shown in Table 4 by illustrating the distribution of final restored power levels across 50 independent inference trials. The proposed HQ-PIDT achieves optimal power restoration in almost all trials. In contrast, A2C predominantly produces solutions below 1000 kW, while PPO mainly produces solutions below 1000 kW and only occasionally reaches the range between 1000 kW and 2000 kW. PIDT performs substantially better than the two DRL baselines, with most of its solutions falling in the range between 3000 kW and 3350 kW. However, compared with PIDT, HQ-PIDT achieves more solutions in the near-optimal range between 3300 kW and 3350 kW. These findings highlight the advantage of the proposed hybrid quantum attention mechanism in improving restoration performance, enabling HQ-PIDT to achieve more reliable near-optimal restoration outcomes than PIDT and substantially outperform the two DRL baselines on the modified IEEE 123-node test feeder.

Modified Iowa 240-Node Test Feeder

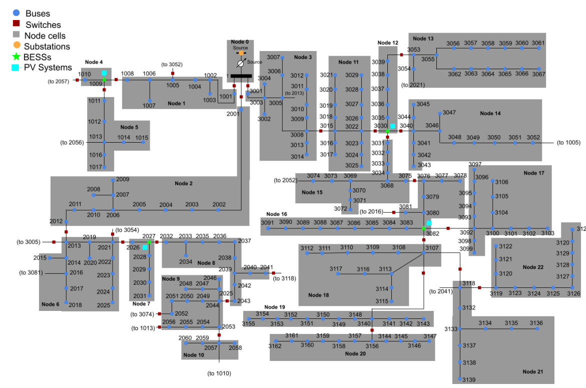
The system topology and the corresponding node-cell-based graph representation are shown in Fig. 11 (a) and (b), respectively. As shown in Fig. 11 (a), the modified Iowa 240-node test feeder includes five energy sources: one substation located at the source bus and four battery energy storage systems (BESSs) located at Buses 1009, 3030, 2027, and 3082. This modified system also includes four solar PV systems, with each BESS paired with one PV system.

Unlike the IEEE 13-node and modified IEEE 123-node test feeders, which only contain static loads and DGs, the modified Iowa 240-node test feeder includes hourly profiles for both loads and PV generation. Therefore, both power demand and available PV generation vary over the restoration horizon. This dynamic setting provides a more realistic test case for evaluating the ability of HQ-PIDT to handle input uncertainty and time-varying operating conditions. Due to this temporal variability, the total restorable power cannot be represented by a single fixed value unless the restoration horizon, time interval between time steps, and starting time are specified. Therefore, as the base case, we set the time horizon to $T = 20$, the starting time to the 1st hour of the profiles, and the time interval between two consecutive decision steps to 1 hour. Under this setting, the final total restored power is approximately 1565 kW.

The learning curves over the first 4000 gradient updates for HQ-PIDT, PIDT, A2C, and PPO are shown in Fig. 12. Additional evaluation results across 50 independent inference trials are presented in Table 5. As shown in Fig. 12, HQ-PIDT achieves faster convergence than the two DRL baselines and exhibits convergence behavior comparable to PIDT. The results in Table 5 further show that HQ-PIDT achieves substantially more optimal solutions than the two DRL baselines and a higher number of optimal solutions than PIDT. Furthermore, Fig. 13 provides additional insight into the simulation results reported in Table 5 by illustrating the distribution of final restored power levels across 50 independent inference trials using the four methods. HQ-PIDT and PIDT both produce restoration solutions in the near-optimal range of 1550-1600 kW in most trials. Compared with PIDT, HQ-PIDT achieves more solutions in this near-optimal range. In contrast, A2C consistently produces solu-



(a)



(b)

Figure 11: (a): The physical topology of the modified Iowa 240-node test feeder; and (b): The node-cell-based graph representation of the modified Iowa 240-node test feeder.

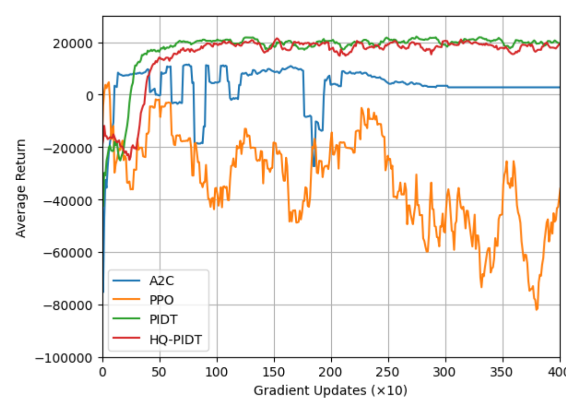


Figure 12: Learning curve for the average return of the first 5000 gradient updates using the four methods for the DSR operations in the modified Iowa 240-node test feeder. The curve is updated per ten gradient updates and is smoothed using moving average per 10 data points.

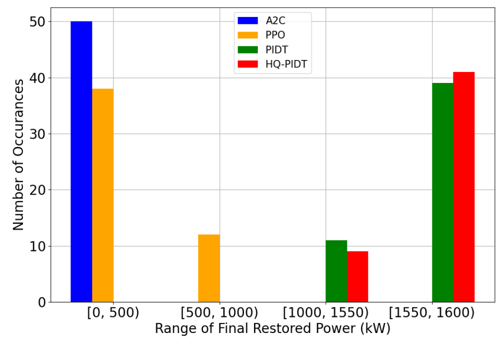


Figure 13: Distribution of power restoration levels in the 50 independent trials using the three methods, respectively, in the modified Iowa 240-bus test feeder.

tions below 500 kW, while PPO mainly produces solutions below 500 kW and occasionally reaches the 500-1000 kW range. These findings highlight the effectiveness of HQ-PIDT on the dynamic Iowa 240-node test feeder, where it maintains strong restoration performance and achieves more reliable near-optimal restoration outcomes than PIDT while substantially outperforming the two DRL baselines.

Table 5: Performance comparison between the four methods for the DSR operation in the modified Iowa 240-Node test feeder under the base case.

Method	**Average Return**	**Std. Return**	**APR (kW)**	**SDPR (kW)**	**# Opt. Sols.**
A2C	2775.228	0.000	156.561	0.000	0
PPO	-33967.531	70725.625	387.470	182.802	0
PIDT	21375.159	9319.341	1534.477	65.840	39
HQ-PIDT	21433.040	9478.040	1528.882	88.445	41

Moreover, we further evaluate the ability of HQ-PIDT to handle observation uncertainty during inference. Specifically, uncertainty is introduced into load demand and PV generation measurements to emulate noisy observed inputs. The uncertainty models are defined as follows.

- Gaussian noise is introduced into the active and reactive power of all loads in the system:

$$\begin{cases} P_{l,t}^{\text{obs}} = P_{l,t}^{\text{true}}(1+\epsilon_{l,t}), \\ Q_{l,t}^{\text{obs}} = Q_{l,t}^{\text{true}}(1+\epsilon_{l,t}), \end{cases} \tag{24}$$

  where $\epsilon_{l,t} \sim \mathcal{N}(\mu, \sigma^2)$, and $l \in \mathcal{L}$;
- Gaussian noise is introduced into the available active output of all PV systems. Mathematically,

$$P_{pv,t}^{PV\text{obs}} = P_{pv,t}^{PV,\text{true}}(1+\epsilon_{pv,t}), \tag{25}$$

  where $\epsilon_{pv,t} \sim \mathcal{N}(\mu, \sigma^2)$, and $pv \in PV$.

Based on these uncertainty models, we define three additional scenarios to evaluate the robustness of HQ-PIDT under observation uncertainty:

1. **Case 1**: Load demand uncertainty modeled by Eq. (24) with $\mu = 0$ and $\sigma = 0.20$.
2. **Case 2**: Load demand uncertainty with $\mu = 0$ and $\sigma = 0.20$, together with PV generation uncertainty modeled by Eq. (25) with $\mu = 0$ and $\sigma = 0.05$.
3. **Case 3**: PV generation uncertainty is kept at $\mu = 0$ and $\sigma = 0.05$, while load demand uncertainty is modeled with $\mu > 0$ and $\sigma = 0.05$.

For these uncertainty scenarios, the starting time is changed to the 9th hour of the profiles, and the time interval between consecutive decision steps is set to 15 minutes while keeping the time horizon at $T = 20$. Cubic spline interpolation is applied to generate smoother transitions between hourly load and PV profiles.

Table 6 reports the evaluation results for the three additional uncertainty scenarios, with each case tested across 50 independent inference trials. Since PIDT is the closest DT-based baseline, we compare HQ-PIDT with PIDT to evaluate the effect of the proposed hybrid quantum attention mechanism under observation uncertainty. Across all three cases, HQ-PIDT achieves a higher number of optimal solutions than PIDT. These findings highlight the effectiveness of the proposed hybrid quantum attention mechanism in improving restoration success in uncertain DSR operational environments.

Table 6: Performance comparison between HQ-PIDT and PIDT for the DSR operation in the modified Iowa 240-Node test feeder under three additional uncertainty scenarios.

Case 1

Method	**Average Return**	**Std. Return**	**APR (kW)**	**SDPR (kW)**	**# Opt. Sols.**
PIDT	24749.777	1011.019	1385.922	50.097	41
HQ-PIDT	24957.844	1910.380	1373.725	74.542	43

Case 2

Method	**Average Return**	**Std. Return**	**APR (kW)**	**SDPR (kW)**	**# Opt. Sols.**
PIDT	24744.385	1060.829	1386.055	46.061	41
HQ-PIDT	24938.649	1905.889	1377.095	82.171	43

Case 3

Method	**Average Return**	**Std. Return**	**APR (kW)**	**SDPR (kW)**	**# Opt. Sols.**
PIDT	26777.291	3222.052	1524.640	48.931	41
HQ-PIDT	26185.060	2000.137	1445.945	85.729	43

Conclusions

This paper proposed HQ-PIDT, a hybrid quantum physics-informed decision transformer for constrained distribution system restoration under time-varying operating conditions and observation uncertainty. The main methodological contribution is a hybrid quantum fidelity-based multi-head self-attention mechanism. Instead of relying solely on classical query-key-value projection and dot-product attention, HQ-PIDT embeds parameterized quantum circuits into the representation projection stage and uses quantum-state fidelity to compute attention similarity between encoded query and key representations. This design enables restoration trajectory contexts to be represented and compared using quantum-state overlap while preserving the autoregressive structure and gradient-based trainability of causal transformers.

Evaluations on three distribution-system test feeders show that HQ-PIDT achieves more reliable optimal or near-optimal restoration outcomes than a physics-informed DT baseline and two DRL baselines under constrained, dynamic, and noisy-input scenarios. These findings highlight the potential of hybrid quantum attention to enhance decision-transformer-based DSR in dynamic and uncertain operational environments. Future work will investigate larger-scale DSR systems, additional sources of observation and communication uncertainty, hardware-aware implementations of the fidelity-based attention mechanism, and extensions of HQ-PIDT to other sequential decision-making problems in power system operations.

References

Bassey, O.; and Butler-Purry, K. L. 2020. Black Start Restoration of Islanded Droop-Controlled Microgrids. *Energies*, 13(22): 5996.

Bu, F.; Yuan, Y.; Wang, Z.; Dehghanpour, K.; and Kimber, A. 2019. A Time-Series Distribution Test System Based on Real Utility Data. In *2019 North American Power Symposium (NAPS)*, 1–6. Wichita, KS, USA: IEEE.

Chen, B.; Ye, Z.; Chen, C.; and Wang, J. 2019. Toward a MILP Modeling Framework for Distribution System Restoration. *IEEE Transactions on Power Systems*, 34(3): 1749–1760.

Chen, C.; Wang, J.; and Ton, D. 2017. Modernizing Distribution System Restoration to Achieve Grid Resiliency Against Extreme Weather Events: An Integrated Solution. *Proceedings of the IEEE*, 105(7): 1267–1288.

Chen, L.; Lu, K.; Rajeswaran, A.; Lee, K.; Grover, A.; Laskin, M.; Abbeel, P.; Srinivas, A.; and Mordatch, I. 2021. Decision Transformer: Reinforcement Learning via Sequence Modeling. *arXiv preprint arXiv:2106.01345*.

Chen, W.-H.; Tsai, M.-S.; and Kuo, H.-L. 2005. Distribution System Restoration Using the Hybrid Fuzzy-Grey Method. *IEEE Transactions on Power Systems*, 20(1): 199–205.

Du, Y.; and Wu, D. 2022. Deep Reinforcement Learning From Demonstrations to Assist Service Restoration in Islanded Microgrids. *IEEE Transactions on Sustainable Energy*, 13(2): 1062–1072.

Fu, J.; Long, Y.; Chen, K.; Wang, W.; and Dou, Q. 2024. Multi-objective Cross-task Learning via Goal-conditioned GPT-based Decision Transformers for Surgical Robot Task Automation. *arXiv preprint arXiv:2405.18757*.

Gao, Y.; Wang, W.; Shi, J.; and Yu, N. 2020. Batch-Constrained Reinforcement Learning for Dynamic Distribution Network Reconfiguration. *IEEE Transactions on Smart Grid*, 11(6): 5357–5369.

Hsu, H.-L.; Bozkurt, A. K.; Dong, J.; Gao, Q.; Tarokh, V.; and Pajic, M. 2024. Steering Decision Transformers via Temporal Difference Learning. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 7477–7483. Abu Dhabi, United Arab Emirates: IEEE.

IEEE. 2014. IEEE Recommended Practice for Electric Power Distribution System Analysis.

Igder, M. A.; Liang, X.; and Mitolo, M. 2022. Service Restoration Through Microgrid Formation in Distribution Networks: A Review. *IEEE Access*, 10: 46618–46632.

Liang, X.; Saaklayen, M. A.; Igder, M. A.; Shawon, S. M. R. H.; Faried, S. O.; and Janbakhsh, M. 2022. Planning and Service Restoration Through Microgrid Formation and Soft Open Points for Distribution Network Modernization: A Review. *IEEE Transactions on Industry Applications*, 58(2): 1843–1857.

Mnih, V.; Badia, A. P.; Mirza, M.; Graves, A.; Harley, T.; Lillicrap, T. P.; Silver, D.; and Kavukcuoglu, K. 2016. Asynchronous methods for deep reinforcement learning. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML'16, 1928–1937. JMLR.org.

OpenDSS Project. 2017. IEEE 13-Bus Test Feeder. GitHub repository. [Online]. Available: https://github.com/tshort/OpenDSS/tree/master/Distrib/IEEETestCases/13Bus. Accessed: Jul. 21, 2026.

Qi, J.; Yang, C.-H. H.; Chen, S. Y.-C.; and Chen, P.-Y. 2025. Quantum Machine Learning: An Interplay Between Quantum Computing and Machine Learning. In *2025 IEEE International Symposium on Circuits and Systems (ISCAS)*, 1–5. London, United Kingdom: IEEE.

Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347*.

Sedzro, K. S.; Shi, X.; Lamadrid, A.; and Zuluaga, L. 2019. A Heuristic Approach to the Post-Disturbance and Stochastic Pre-Disturbance Microgrid Formation Problem. *IEEE Transactions on Smart Grid*, 10(5): 5574–5586.

Singh, R.; Mehfuz, S.; and Kumar, P. 2016. Intelligent Decision Support Algorithm for Distribution System Restoration. *SpringerPlus*, 5(1): 1175.

Ullah, I.; Butler-Purry, K. L.; Bent, R.; Nagarajan, H.; and Barnes, A. 2025. Studying a Sequential Black Start Restoration Method Using PowerModelsONM.jl. In *2025 57th North American Power Symposium (NAPS)*, 1–6. IEEE.

Wang, F.; Cetinay, H.; He, Z.; Liu, L.; Van Mieghem, P.; and Kooij, R. E. 2023. Recovering Power Grids Using Strategies Based on Network Metrics and Greedy Algorithms. *Entropy*, 25(10): 1455.

Xanadu. 2026. Quantum Circuits. *PennyLane Documentation*, version 0.45.1. [Online]. Available: https://docs.pennylane.ai/en/stable/introduction/circuits.html. Accessed: Jul. 21, 2026.

Yang, L.-J.; Zhao, Y.; Wang, C.; Gao, P.; and Hao, J.-H. 2019. Resilience-Oriented Hierarchical Service Restoration in Distribution System Considering Microgrids. *IEEE Access*, 7: 152729–152743.

Yao, S.; Gu, J.; Zhang, H.; Wang, P.; Liu, X.; and Zhao, T. 2020. Resilient Load Restoration in Microgrids Considering Mobile Energy Storage Fleets: A Deep Reinforcement Learning Approach. In *2020 IEEE Power & Energy Society General Meeting (PESGM)*, 1–5. Montreal, QC, Canada: IEEE.

Yao, T.; Chen, X.; Yao, Y.; Huang, W.; and Chen, Z. 2025. Offline Prompt Reinforcement Learning Method Based on Feature Extraction. *PeerJ Computer Science*, 11: e2490.

Zhao, H.; Wei-Kocsis, J.; Akhijahani, A. H.; and Butler-Purry, K. L. 2024. Exploring a Physics-Informed Decision Transformer for Distribution System Restoration: Methodology and Performance Analysis. *arXiv preprint arXiv:2407.00808*.

Zhao, H.; Wei-Kocsis, J.; Heidari Akhijahani, A.; and Butler-Purry, K. L. 2025. Dual-Head Physics-Informed Graph Decision Transformer for Distribution System Restoration. *arXiv preprint arXiv:2508.06634*.

Zhao, T.; and Wang, J. 2022. Learning Sequential Distribution System Restoration via Graph-Reinforcement Learning. *IEEE Transactions on Power Systems*, 37(2): 1601–1611.