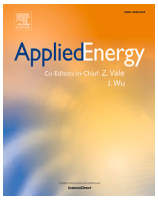
Applied Energy 408 (2026) 127418

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apen



Valuing operational flexibility in hybrid energy systems: A neural solution to the Hamilton–jacobi–bellman equation

Dávid Zoltán Szabó

Corvinus University of Budapest, Fővám tér 8., Budapest, H-1093, Hungary

HIGHLIGHTS

- Integrated stochastic control of hybrid gas-wind–storage assets under correlated price uncertainty.
- Neural HJB solver enables scalable solutions to high-dimensional control problems with non-linear operating frictions.
- Empirical calibration and backtesting using historical 2019–2024 European electricity and gas price data.
- Diagnostic benchmarking against analytical linear-quadratic models and classical finite-difference schemes.
- Learned policies reveal sparse, threshold-driven storage operations with extended no-trade regions.

ARTICLE INFO

Keywords:
Energy storage
Stochastic optimal control
Hamilton–jacobi–bellman equations
Physics-informed neural networks
Real options
Renewable energy systems

ABSTRACT

This paper develops a continuous-time stochastic control framework for the joint operation of gas-fired generation, wind power, and energy storage under correlated electricity and gas price uncertainty. The operator's decision problem is formulated as a finite-horizon Hamilton–Jacobi–Bellman (HJB) equation, capturing the trade-off between immediate revenues and the continuation value of storage under physical and operational constraints.

To address the resulting high-dimensional control problem, we employ a mesh-free neural approximation based on physics-informed and Deep Galerkin methods. An analytical linear–quadratic (LQ) formulation is derived as a benchmark, providing structural insight and a reference point under simplified assumptions.

Numerical experiments demonstrate stable convergence of the neural HJB solver and recovery of economically interpretable policy structures. When calibrated to historical electricity and gas price data and evaluated under realistic transaction costs, the learned policy exhibits sparse, threshold-driven storage operation with extended no-trade regions. In these regimes, optimal behavior leaves the storage inactive despite ongoing generation, reflecting the option-like and highly state-dependent value of operational flexibility.

Overall, the results show that neural HJB solvers provide an economically consistent and transparent framework for analyzing hybrid energy systems. By linking stochastic price dynamics, operational constraints, and realized storage decisions, the approach clarifies when flexibility is actively exercised and when it remains economically dormant in low-carbon power systems.

1. Introduction

The transition toward low-carbon power systems has fundamentally altered the operational environment of electricity producers. Rising penetration of intermittent renewable generation, tighter coupling between electricity and fuel markets, and the growing role of energy storage as a flexibility asset have transformed short-term operational decisions into a stochastic and multi-dimensional control problem.

In many liberalized power systems, gas-fired generation continues to play a central role in balancing renewable output and setting marginal prices. Consequently, electricity and natural gas prices exhibit strong empirical co-movement driven by merit-order effects and fuel-price pass-through [1]. Energy storage introduces an additional layer of optionality by allowing operators to shift energy across time, but its economic value depends critically on the joint evolution of prices, physical constraints, and market frictions.

Email address: davidzoltan.szabo@uni-corvinus.hu.

https://doi.org/10.1016/j.apenergy.2026.127418
Received 23 October 2025; Received in revised form 29 December 2025; Accepted 13 January 2026
Available online 19 January 2026

Stochastic optimal control provides a natural framework for modeling such problems. Continuous-time formulations based on Hamilton–Jacobi–Bellman (HJB) equations have a long tradition in energy economics, including applications to hydro scheduling [2], natural gas storage valuation [3,4], and optimal operation of storage systems [5]. While these models offer a clear economic interpretation, their practical use is often limited by the computational burden of solving high-dimensional HJB equations with classical numerical methods [6].

Recent advances in scientific machine learning have opened new avenues for addressing this challenge. Neural approaches such as the Deep Galerkin Method (DGM) [7] and Physics-Informed Neural Networks (PINNs) [8] approximate value functions by minimizing the residual of the governing partial differential equation without relying on spatial discretization. These methods have been shown to approximate solutions of high-dimensional PDEs under mild regularity conditions [9,10], and have found applications in finance, control, and engineering.

Despite this progress, the application of neural HJB solvers to empirically grounded energy-system problems remains limited. Existing studies often focus on low-dimensional settings or simplified market structures, and frequently abstract from correlated fuel and power prices, explicit physical constraints, or transaction costs. As a result, it remains unclear how neural HJB solvers translate into economically interpretable operational policies under realistic market conditions.

The objective of this paper is to address this gap. We develop a continuous-time stochastic control model for the joint operation of gas-fired generation, wind power, and energy storage under correlated electricity and gas price uncertainty. The operator's problem is formulated as a finite-horizon HJB equation with explicit physical constraints and transaction costs. To solve the resulting high-dimensional control problem, we employ a mesh-free neural approximation based on DGM/PINN principles.

An analytical linear–quadratic (LQ) formulation is derived as a benchmark to provide structural insight under simplified assumptions. The neural solver is then applied to the full nonlinear model and evaluated both under synthetic calibration and under parameters estimated from historical European electricity and gas price data. Rather than emphasizing pointwise numerical accuracy, the analysis focuses on the economic structure of the learned policies, including threshold behavior, shadow prices, no-trade regions, and the sensitivity of optimal operation to market frictions and temporal aggregation.

By combining classical stochastic control with modern neural PDE solvers, this paper contributes to the applied energy economics literature by clarifying when operational flexibility is economically exercised and when it remains dormant, with implications for market design, investment decisions, and the valuation of flexibility in low-carbon power systems.

2. Related literature

The present study connects three strands of literature: (i) stochastic control and HJB theory, (ii) optimal operation and storage valuation in energy markets, and (iii) neural-network-based methods for solving HJB equations.

2.1. Stochastic control and HJB theory

The foundations of continuous-time stochastic control are well established. Classical references such as [11,12] provide rigorous treatments of controlled diffusions, dynamic programming, and viscosity solutions of HJB equations. Linear–quadratic (LQ) control occupies a special role due to its analytical tractability and clear economic interpretation [13,14]. Extensions to singular and impulse control, which are particularly relevant for inventory and storage problems, are discussed in Ref. [15,16].

2.2. Energy economics and storage valuation

Stochastic control models have long been used to analyze operational decisions in energy systems. Early work includes hydroelectric planning [2] and real options approaches to resource extraction [17], with further extensions to decremental regulation and flexible capacity valuation [18]. Natural gas storage valuation has been extensively studied using HJB and dynamic programming formulations [3,4,19], highlighting the option-like nature of storage and the role of operational constraints. Related methods have been applied to electricity storage and switching problems [5] and more recently to the optimal management of virtual power plants under market uncertainty [20].

Most existing models, however, simplify fuel-price dynamics or treat electricity and gas prices independently. Empirical evidence suggests strong dependence between these markets, especially in systems where gas-fired generation frequently sets marginal prices [1]. Accounting for this dependence significantly increases model dimensionality and limits the applicability of classical numerical solvers.

2.3. Neural methods for HJB equations

Neural-network-based methods for solving PDEs have developed rapidly in recent years. The Deep Galerkin Method of [7] and the Physics-Informed Neural Network framework of [8] approximate PDE solutions by minimizing residuals at randomly sampled collocation points, thereby avoiding spatial grids. Subsequent work demonstrates that deep learning can approximate solutions of high-dimensional HJB equations under suitable conditions [9,10].

These methods have been applied to optimal control problems in finance and engineering, including hedging and stopping problems [21,22]. However, applications to energy systems typically remain stylized or focus on simplified PDE structures [23].

2.4. Positioning of the present study

The present paper does not claim methodological novelty in the neural architecture itself. Instead, its contribution lies in the integration of neural HJB solvers into a realistic energy-economics setting characterized by correlated electricity and gas prices, storage dynamics, and operational constraints. The resulting four-dimensional HJB equation already exceeds the stable reach of standard grid-based methods, motivating the use of a mesh-free neural approach.

By combining an analytical LQ benchmark with a PINN-based solver and empirical calibration, the paper emphasizes structural consistency, economic interpretability, and numerical stability rather than purely pointwise accuracy. This positioning aligns the contribution with applied energy research while leveraging recent advances in neural PDE methods.

3. Model formulation

This section describes the decision environment of an integrated operator who manages a hybrid portfolio composed of gas and wind generation units supported by an energy storage device. The aim is to provide a clear and flexible framework linking physical operations, market dynamics, and stochastic control. All assumptions are stated explicitly so that the model can be replicated or extended in both analytical and computational settings.

3.1. State variables and stochastic price dynamics

The system evolves over a finite horizon $[0, T]$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ endowed with the natural filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$. The state vector is

$$X(t) = (E(t), G(t), S(t)),$$

where

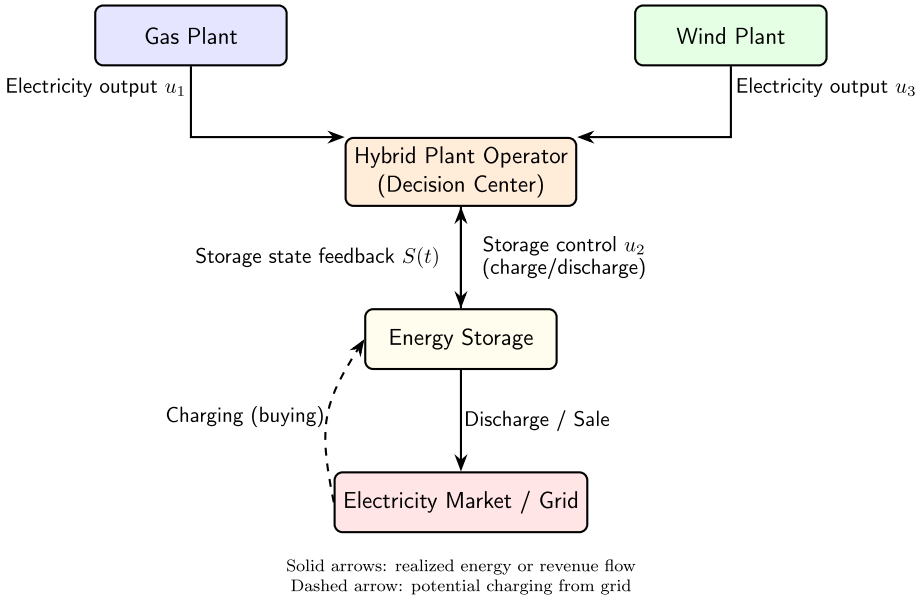


Fig. 1. Structure of the joint control problem linking gas, wind, storage, and market interactions. All controls (u_1, u_2, u_3) are chosen by the operator, who allocates generation and storage to maximize expected value under stochastic prices.

- $E(t)$: wholesale electricity price (EUR/MWh);
- $G(t)$: natural gas price (EUR/$\text{MWh}_{\text{thermal}}$);
- $S(t)$: stored energy (MWh).

Electricity and natural gas prices are modeled as correlated mean-reverting Ornstein–Uhlenbeck processes:

$$dE(t) = \kappa_E\big(\bar{E} - E(t)\big)\,dt + \sigma_E\,dW_E(t), \tag{1}$$

$$dG(t) = \kappa_G\big(\bar{G} - G(t)\big)\,dt + \sigma_G\,dW_G(t), \tag{2}$$

with instantaneous correlation

$$dW_E(t)\,dW_G(t) = \rho\,dt, \qquad |\rho| \le 1.$$

This specification captures the key empirical features of wholesale electricity and natural gas prices observed in European markets. Mean reversion reflects the tendency of prices to revert toward long-run levels driven by fuel costs, capacity constraints, and regulatory interventions, while the diffusion terms account for short-term uncertainty and market shocks.

The instantaneous correlation coefficient ρ captures the structural dependence between fuel and power prices arising from merit-order effects and fuel-price pass-through. Through this channel, gas price fluctuations affect not only contemporaneous electricity prices but also the intertemporal value of storage via expectations of future price movements.

At this stage, the parameters $(\kappa_E, \kappa_G, \bar{E}, \bar{G}, \sigma_E, \sigma_G, \rho)$ define the stochastic structure of the control problem and are treated as exogenous. Their empirical estimation and calibration using historical market data are described in Section 6.2.

3.2. Control variables and operational decisions

At each time t, the hybrid plant operator allocates production and storage by choosing control actions adapted to the available information $\{\mathcal{F}_t\}$. In the general formulation, the control vector is

$$u(t) = (u_1(t), u_2(t), u_3(t)),$$

where

- $u_1(t)$ denotes electricity dispatched from the gas plant (MW),
- $u_2(t)$ denotes the net storage action, positive when charging and negative when discharging (MW),
- $u_3(t)$ denotes wind generation effectively sold to the market (MW).

In the numerical experiments reported in this paper, gas and wind generation are treated as exogenous and must-take processes. Accordingly, the only actively optimized control variable is the storage action $u_2(t)$, while $u_1(t)$ and $u_3(t)$ enter the model implicitly through market prices and the effective electricity–gas price spread. This reduced control specification preserves the essential economic trade-offs associated with storage flexibility while substantially simplifying the numerical solution of the stochastic control problem.

The admissible storage control satisfies power and energy constraints,

$$u_2^{\min} \le u_2(t) \le u_2^{\max}, \qquad 0 \le S(t) \le S_{\max},$$

which reflect charging and discharging limits as well as finite storage capacity. All decisions are non-anticipative and depend only on information available up to time t.

Fig. 1 illustrates the physical and economic structure of the integrated system. The operator acts as a decision center, routing energy between generation, storage, and the market. When prices are low, the operator typically charges the storage device by purchasing electricity from the grid or diverting available generation. When prices are high, stored energy is discharged and sold to the market. Through this mechanism, storage serves as a source of temporal flexibility that allows the operator to arbitrage stochastic price fluctuations.

3.3. Storage dynamics and boundary behavior

The storage level evolves according to the deterministic balance equation

$$\dot{S}(t) = \eta_c\,u_2^+(t) - \frac{1}{\eta_d}\,u_2^-(t),$$

where $u_2^+(t) = \max(u_2(t), 0)$ and $u_2^-(t) = -\min(u_2(t), 0)$ denote charging and discharging rates, and $\eta_c, \eta_d \in (0, 1]$ are the corresponding

efficiencies. The state of charge is constrained to the compact interval

$$0 \le S(t) \le S_{\max}.$$

In the numerical implementation, these bounds are enforced through reflecting (Neumann-type) boundary conditions in the Hamilton–Jacobi–Bellman equation, implying that the marginal value of storage vanishes at the boundaries. Economically, this captures the idea that additional charging when the storage is full, or additional discharging when it is empty, has no incremental value. This treatment avoids artificial impulses at the boundaries and yields smooth optimal policies near capacity limits.

3.4. Instantaneous profit and objective function

At each instant, the operator earns an instantaneous profit that reflects revenues from electricity sales net of operating and adjustment costs. In the general formulation, this profit is given by

$$\begin{aligned} L(X(t), u(t)) = {} & E(t)\,[u_1(t) + u_3(t) - u_2(t)] \\ & - C_G(u_1(t), G(t)) - C_W(u_3(t)) - C_S(u_2(t)), \end{aligned} \tag{3}$$

where the first term represents revenue from net electricity sales (generation minus storage charging), and the remaining terms capture operating, fuel, and adjustment costs associated with each asset.

Gas-related costs may include fuel expenditures and ramping penalties,

$$C_G(u_1, G) = G\,\kappa\,u_1 + \tfrac{1}{2}\alpha_G u_1^2,$$

where κ denotes the heat-rate conversion factor and $\alpha_G > 0$ penalizes rapid changes in gas-fired output. Wind generation costs can similarly reflect convex adjustment or curtailment penalties,

$$C_W(u_3) = \tfrac{1}{2}\alpha_W u_3^2,$$

with $\alpha_W > 0$ capturing operational frictions associated with ramping or curtailment decisions.

In the numerical experiments reported in this paper, gas and wind generation are treated as exogenous and must-take. Accordingly, explicit adjustment costs for u_1 and u_3 are set to zero, i.e. $C_G \equiv 0$ and $C_W \equiv 0$. All economic interaction between electricity and gas markets therefore enters through the effective price spread,

$$\text{spread}(E, G) = E - \eta_{\text{spark}}\,G,$$

which summarizes fuel-price pass-through and joint market effects in reduced form.

Under this specification, the instantaneous profit relevant for the numerical solution reduces to

$$L(X(t), u_2(t)) = -\,\text{spread}(E(t), G(t))\,u_2(t) - \tfrac{1}{2}R_{22}\,u_2(t)^2, \tag{4}$$

where $R_{22} > 0$ penalizes aggressive charging and discharging actions and captures short-run degradation and cycling costs of the storage device. This reduced form preserves concavity of the profit function and isolates the economic trade-off between stochastic prices and storage flexibility.

The operator maximizes expected discounted profit over the finite horizon $[0, T]$,

$$\begin{aligned} V(t, X) = \sup_{u_2(\cdot)\in\mathcal{A}(t,X)} \mathbb{E}\Bigg[& \int_t^T e^{-\beta(s-t)} L(X(s), u_2(s))\,ds \\ & + e^{-\beta(T-t)}\Phi\big(S(T), E(T), G(T)\big) \,\Bigg|\, X(t) = X \Bigg], \end{aligned} \tag{5}$$

where $\beta > 0$ denotes the continuous discount rate.

Consistent with the reduced profit specification, the terminal value of stored energy is taken to be

$$\Phi\big(S(T), E(T), G(T)\big) = \text{spread}(E(T), G(T))\,S(T),$$

corresponding to liquidation of the storage position at the prevailing effective electricity–gas price spread. This terminal condition is enforced explicitly in the numerical implementation and ensures internal consistency between the running payoff, the boundary conditions, and the learned value function.

3.5. Regularity and well-posedness

Under standard assumptions—Lipschitz-continuous drift and diffusion coefficients, compact control domains, and polynomial growth of the profit function—the associated Hamilton–Jacobi–Bellman equation admits a unique continuous viscosity solution [11]. The compactness of the storage domain simplifies boundary treatment, while bounded controls guarantee that the Hamiltonian remains finite for all admissible states.

In the present setting, reflecting (Neumann-type) boundary conditions at the storage limits ensure well-posedness of the HJB equation and exclude spurious boundary effects. These properties guarantee the existence of optimal Markov controls and provide a rigorous foundation for both the analytical formulation and the physics-informed neural network approximation employed in the subsequent sections.

3.6. Aggregate output constraint

Hybrid energy portfolios are often subject to limits on the total electrical power that can be delivered to the grid due to interconnection capacity or transformer constraints. Let $U_{\max}$ denote the maximum admissible aggregate output. At the modeling level, this leads to the portfolio-level constraint

$$u_1(t) + u_2(t) + u_3(t) \le U_{\max}, \qquad t \in [0, T], \tag{6}$$

which couples gas generation, wind dispatch, and storage operation.

Constraint (6) defines the admissible control domain of the stochastic control problem and reflects a common physical limitation in integrated generation–storage systems. It plays an important conceptual role by highlighting the competition between generation and storage for limited grid access and by clarifying how storage can be used to shift output across time rather than increase instantaneous delivery.

In the numerical experiments reported in Section 6, this aggregate output constraint is not enforced explicitly. Instead, the analysis focuses on the economic interaction between price uncertainty, storage dynamics, and bounded charging and discharging rates. This reduced specification allows us to isolate and study the structure of the optimal policy and the associated shadow prices without introducing additional active constraints.

The modeling framework itself readily accommodates constraint (6), and its numerical enforcement within the neural HJB solver follows standard penalty or projection techniques. Incorporating explicit grid-capacity constraints is therefore a natural extension of the present framework, but is left for future work in order to maintain clarity of exposition and focus on the core economic mechanisms.

4. HJB and analytical framework

The dynamic programming principle provides the foundation for deriving the HJB equation that characterizes the optimal operation of the hybrid gas–wind–storage portfolio. This section develops the continuous-time control formulation, interprets its components in economic and physical terms, and introduces the analytical benchmark used to validate the subsequent neural approximation.

4.1. Dynamic programming principle

For any admissible control trajectory $u(\cdot)$ and the associated state process $X_t = (E_t, G_t, S_t)$, the expected discounted profit is defined as

$$V(t,x) = \sup_{u(\cdot)\in\mathcal{A}(t,x)} \mathbb{E}\left[\int_t^T e^{-\beta(s-t)} L(X_s, u_s)\, ds + e^{-\beta(T-t)}\Phi(S_T) \;\middle|\; X_t = x\right].$$

This represents the maximal attainable value starting from state x at time t, given all feasible operational decisions. Applying the principle of optimality over a small time increment $\Delta t > 0$ yields

$$V(t,x) = \sup_{u\in\mathcal{U}(x)} \mathbb{E}\left[e^{-\beta\Delta t} V(t+\Delta t, X_{t+\Delta t}) + \int_t^{t+\Delta t} e^{-\beta(s-t)} L(X_s, u_s)\, ds\right],$$

which, after applying Ito's lemma and letting $\Delta t \to 0$, leads to the continuous-time HJB equation.

4.2. HJB formulation

Let $\mathcal{L}^u$ denote the infinitesimal generator of the controlled diffusion process under control u:

$$\mathcal{L}^u V = \mu_E \frac{\partial V}{\partial E} + \mu_G \frac{\partial V}{\partial G} + \dot{S}(u) \frac{\partial V}{\partial S} + \tfrac{1}{2}\sigma_E^2 \frac{\partial^2 V}{\partial E^2} + \tfrac{1}{2}\sigma_G^2 \frac{\partial^2 V}{\partial G^2} + \rho\,\sigma_E\sigma_G \frac{\partial^2 V}{\partial E\,\partial G}.$$

The value function $V(t,x)$ then satisfies

$$V_t + \sup_{u_2\in\mathcal{U}_2(S)} \left\{ L(x,u_2) + \mathcal{L}^{u_2} V - \beta V \right\} = 0, \qquad V(T,x) = \Phi(S_T),$$

where the supremum is taken over admissible storage actions u_2. In the numerical implementation, gas and wind generation decisions are treated as exogenous, so that the Hamiltonian is effectively one-dimensional in the control variable.

The expression in braces defines the *Hamiltonian* $\mathcal{H}(x, V, \nabla V, \nabla^2 V)$, which aggregates instantaneous profit $L(x,u)$ and the expected rate of change in continuation value. The maximizing control $u^*(t,x)$ defines the optimal operational policy, ensuring that every decision balances current profit and the discounted value of future flexibility.

4.3. Economic interpretation and structure

The HJB equation has a clear operational interpretation. The term V_t measures the temporal change in system value, while the drift and diffusion components of $\mathcal{L}^u V$ describe how price movements affect expected revenues. The gradient $\partial_E V$ and $\partial_G V$ represent the sensitivities of portfolio value to electricity and gas price shocks—analogous to "delta" exposures in option theory—while $\partial_S V$ is the shadow value of stored energy, i.e., the marginal worth of one additional MWh in the battery. Second-order derivatives capture the curvature of the value function and thus quantify risk exposure to volatility and cross-market correlation. When σ_E, σ_G, or ρ increase, the portfolio's optionality rises, raising V even if mean prices remain constant.

The discount term $-\beta V$ reflects time preference and risk adjustment, ensuring that future profits are valued less and improving numerical stability of the PDE. At the storage boundaries $S = 0$ and $S = S_{\max}$, reflecting (Neumann-type) conditions $\partial_S V = 0$ apply, implying that the marginal value of charging or discharging vanishes when the storage is empty or full. This property reproduces the real-world behavior of batteries operating near capacity limits, where technical and economic constraints prevent abrupt actions.

4.4. Linear–quadratic analytical benchmark

To provide an interpretable reference solution and a validation anchor for the neural approximation, we consider a linear–quadratic (LQ) specialization of the stochastic control problem. LQ formulations are classical in optimal control and yield closed-form feedback policies that clarify the qualitative structure of optimal decisions under simplifying assumptions [6,13].

Electricity and gas prices are approximated locally by correlated mean-reverting Ornstein–Uhlenbeck dynamics,

$$dE_t = \kappa_E(\bar{E} - E_t)\,dt + \sigma_E\, dW_{E,t},$$
$$dG_t = \kappa_G(\bar{G} - G_t)\,dt + \sigma_G\, dW_{G,t},$$
$$dW_{E,t}\, dW_{G,t} = \rho\, dt,$$

which captures the dominant short- to medium-term behavior observed in wholesale energy markets. Linearizing around the long-run means $\bar{E}$ and $\bar{G}$ yields affine price dynamics that are compatible with the LQ framework.

Under quadratic adjustment costs and unconstrained dynamics, the associated Hamilton–Jacobi–Bellman equation admits a quadratic value function of the form

$$V(x,t) = \tfrac{1}{2} x^\top P(t) x,$$

where the symmetric matrix $P(t)$ solves a Riccati differential equation. The resulting optimal control is a linear state-feedback rule,

$$u^*(t,x) = -K(t)\, x,$$

with gain matrix $K(t)$ determined by the model parameters governing mean reversion, adjustment costs, and storage responsiveness.

This analytical benchmark provides two key insights. First, it highlights the threshold-driven and approximately planar structure of optimal storage policies in regions where constraints are inactive. Second, it serves as a structural validation tool: when the full nonlinear model is simplified to the LQ setting, the neural HJB solver should recover a policy consistent with the Riccati-based feedback.

Importantly, the linear–quadratic formulation is not used as a training target for the neural solver. Instead, it serves as a structural and qualitative reference. The PINN is trained on the full nonlinear problem with bounded controls and state constraints, and comparison with the LQ benchmark is performed at the level of policy shape and economic interpretation, rather than pointwise accuracy.

4.5. Existence, regularity, and neural approximation link

Under standard conditions—Lipschitz coefficients, compact control sets, and polynomial growth of L—the HJB admits a unique continuous viscosity solution $V \in C([0,T]\times\mathbb{R}^3)$ [11]. The LQ benchmark corresponds to a special case where the HJB has a closed-form quadratic solution. For general nonlinear specifications, however, no analytical solution exists, motivating numerical approximation.

The neural method introduced below does not require the problem to be linear–quadratic. PINNs approximate the value function $V_\theta(t,x)$ by minimizing the residual of the HJB equation directly, without assuming any structural form for dynamics or profit. The LQ system instead provides a *validation anchor*: in the limit where dynamics and profits satisfy the linear–quadratic conditions above, the neural solver should recover the analytical Riccati-based feedback. This benchmark thus serves as both a calibration test and a sanity check for the high-dimensional, nonlinear extension.

5. Conceptual neural solution (DGM/PINN)

While the analytical form of the HJB equation provides theoretical clarity, its direct numerical solution becomes infeasible once the number of state variables increases. Classical discretization-based methods—finite difference, finite element, or tree-based dynamic programming—suffer from the exponential growth of computational cost with each additional dimension, a manifestation of the "curse of dimensionality." Neural approaches overcome this limitation by representing the value function as a differentiable function approximator whose parameters

are optimized to satisfy the HJB residual and boundary conditions in expectation. This section outlines the conceptual structure of such a solver, linking recent developments in machine learning with classical stochastic control in energy systems.

5.1. Approximation of the value function

Let $V_\theta(t,x)$ denote a neural approximation of the true value function, parameterized by weights θ. The input vector $(t,E,G,S) \in [0,T]\times\mathcal{X}$ captures the temporal and state dimensions, and the output is a scalar estimate of portfolio value. A fully connected feedforward architecture with L layers and smooth activation functions (e.g., $\tanh$ or $\texttt{softplus}$) is typically employed to ensure differentiability of all derivatives required by the HJB operator. The universal approximation theorem [24,25] guarantees that sufficiently wide neural networks with nonpolynomial activation functions can approximate any continuous function on compact domains to arbitrary accuracy. Hence, V_θ is, in principle, capable of representing complex and nonlinear value surfaces induced by correlated stochastic market dynamics.

Automatic differentiation provides exact gradients and Hessians of V_θ with respect to its inputs, allowing the HJB operator to be evaluated continuously without spatial grids or finite-difference stencils. This capability is particularly advantageous in stochastic energy systems, where control decisions depend on smooth yet high-dimensional functions of correlated prices and storage levels.

5.2. Residual loss and training objective

The key to neural PDE solvers lies in enforcing the HJB equation in a weak sense by minimizing its residual. Define the interior loss as

$$\mathcal{L}_{\mathrm{PDE}}(\theta) = \mathbb{E}_{(t,x)\sim\mathcal{D}_{\mathrm{int}}}\left[\left|V_{t,\theta} + \sup_{u\in\mathcal{U}(x)}\{L(x,u) + \mathcal{L}^u V_\theta - \beta V_\theta\}\right|^2\right],$$

where $\mathcal{D}_{\mathrm{int}}$ denotes the sampling distribution of interior collocation points. Because the control supremum is not analytically available in general, a differentiable relaxation is applied: the maximizing control $u^*(x,\nabla V_\theta)$ is approximated using a smooth softmax or quadratic penalty, ensuring differentiability and stability of gradient-based optimization.

Boundary and terminal conditions contribute additional loss terms:

$$\mathcal{L}_{\mathrm{BC}}(\theta) = \mathbb{E}_{x\in\partial\mathcal{X}}\left[|B(V_\theta,x)|^2\right], \qquad \mathcal{L}_{\mathrm{T}}(\theta) = \mathbb{E}_{x\in\mathcal{X}}\left[|V_\theta(T,x)-\Phi(S)|^2\right],$$

where $B(\cdot)$ encodes reflecting storage constraints ($\partial_S V = 0$ at $S = 0, S_{\max}$).

In addition, the total output constraint $u_1+u_2+u_3 \le U_{\max}$ is enforced softly through an auxiliary penalty term,

$$\mathcal{L}_{\mathrm{grid}}(\theta) = \lambda_{\mathrm{grid}}\,\mathbb{E}_{(t,x)}\left[\max\{0, u_1^*+u_2^*+u_3^*-U_{\max}\}^2\right],$$

[1] which keeps the learned control within feasible operational limits without altering the underlying HJB structure. The overall objective function becomes

$$\mathcal{L}(\theta) = \lambda_1\mathcal{L}_{\mathrm{PDE}} + \lambda_2\mathcal{L}_{\mathrm{BC}} + \lambda_3\mathcal{L}_{\mathrm{T}} + \mathcal{L}_{\mathrm{grid}},$$

[2] with coefficients λ_i calibrated to balance gradient magnitudes across components. Minimizing $\mathcal{L}(\theta)$ ensures that the learned V_θ satisfies both the HJB equation and its economic and physical constraints in an average sense.

5.3. Sampling and optimization strategy

Collocation points (t,E,G,S) are sampled from low-discrepancy Sobol or Latin hypercube sequences to achieve near-uniform coverage of the domain. Mini-batches of such samples are used to estimate expectations in $\mathcal{L}(\theta)$, and the parameters θ are updated via stochastic gradient descent or Adam optimization. Adaptive learning rates and early stopping help prevent overfitting or divergence. In time-dependent problems, a curriculum approach can be adopted, starting from short horizons and gradually expanding to the full $[0,T]$, which stabilizes learning for backward-evolving PDEs.

Crucially, all derivatives required for the PDE residual are obtained by automatic differentiation, eliminating the need for any discretization grid. Hence, the computational cost grows linearly with the number of collocation points rather than exponentially with the number of state variables, allowing the method to scale to problems with up to five or more stochastic factors. In the present application, this property enables the joint treatment of electricity and gas prices, storage dynamics, and operational constraints within a unified neural framework.

Importantly, the use of random uniform sampling does not alter the qualitative structure of the learned policy or the location of switching regions. Differences are primarily numerical, manifesting as slower convergence and noisier shadow-price estimates, which motivates the use of quasi-random sampling in all reported experiments.

5.4. Control recovery and interpretation

Once the network has been trained, approximate optimal controls can be extracted from the first-order conditions embedded in the Hamiltonian. For continuously differentiable Hamiltonians, the optimal feedback control satisfies

$$\nabla_u[L(x,u) + (\nabla V_\theta)\cdot f(x,u)] = 0,$$

which can be solved analytically or via one-step gradient ascent at each (t,x). This produces explicit control maps $u_\theta^*(t,x)$, yielding interpretable policy surfaces. Visualizing u_θ^* as a function of electricity price and storage level reveals intuitive operational patterns: charging at low prices, discharging at high prices, and throttling gas generation when fuel costs spike. Such mappings provide economic interpretability and practical decision support, bridging machine-learning predictions with physical system operations.

5.5. Relation to the DGM and PINN literature

The numerical approach adopted in this paper builds directly on the Deep Galerkin Method (DGM) introduced by Ref. [7] and the broader class of Physics-Informed Neural Networks (PINNs) formalized by Ref. [8]. These methods approximate solutions of partial differential equations by minimizing residuals at randomly or quasi-randomly sampled collocation points, thereby avoiding spatial discretization and mitigating the curse of dimensionality inherent in grid-based schemes.

From a numerical perspective, residual minimization can be interpreted as a Galerkin-type projection of the unknown solution onto a function space spanned by neural network basis functions. Recent theoretical results establish convergence and stability of such neural PDE solvers under mild regularity assumptions [10], showing that they inherit key consistency properties of classical finite-element methods while remaining fully mesh-free.

The contribution of the present work is not the development of a novel neural architecture or training paradigm. Rather, it lies in the disciplined integration of existing DGM/PINN solvers into a constrained stochastic control setting that is economically interpretable and directly motivated by energy system operation. The neural network is used as a surrogate for the value function of a finite-horizon Hamilton–Jacobi–Bellman equation with hard state and control constraints, transaction costs, and correlated market dynamics.

[1] In all numerical experiments reported in the manuscript, the aggregate output constraint is not explicitly enforced, and the corresponding grid-penalty term is therefore set to zero. It is included here to emphasize that the proposed framework readily accommodates such constraints without altering the structure of the HJB equation or the neural training procedure.

[2] In all numerical experiments, the loss weights were set to $\lambda_1 = \lambda_2 = \lambda_3 = 1$. We verified that moderate rescaling of these coefficients does not affect the qualitative structure of the learned policy or the reported diagnostics.

Algorithm 1 PINN-based HJB solver for hybrid gas–wind–storage operation.

Require: Time horizon T, discount rate β, price dynamics (E_t, G_t), storage dynamics S_t, control constraints $\mathcal{U}(x)$, neural network architecture $V_\theta(t, E, G, S)$

Ensure: Approximate value function V_θ and feedback control u^*_θ

Initialize neural network parameters θ
Define sampling distributions for interior, boundary, and terminal points
for training epoch $k = 1, \dots, K$ **do**
  Sample collocation points (t_i, E_i, G_i, S_i) from interior domain
  Sample boundary points at $S = 0$ and $S = S_{\max}$
  Sample terminal points at $t = T$
  Evaluate V_θ and its derivatives using automatic differentiation
  Compute approximate optimal control

$$u^*(x) = \arg\max_{u \in \mathcal{U}(x)} \left\{ L(x,u) + \nabla V_\theta(x) \cdot f(x,u) \right\}$$

  Compute HJB residual loss $\mathcal{L}_{\text{PDE}}$
  Compute boundary loss $\mathcal{L}_{\text{BC}}$ and terminal loss $\mathcal{L}_{\text{T}}$
  Combine losses into total objective

$$\mathcal{L}(\theta) = \lambda_1 \mathcal{L}_{\text{PDE}} + \lambda_2 \mathcal{L}_{\text{BC}} + \lambda_3 \mathcal{L}_{\text{T}}$$

  Update parameters θ using Adam optimizer
end for
Recover feedback policy $u^*_\theta(t, x)$ from trained V_θ
return V_θ, u^*_θ

Within this framework, the emphasis is placed on structural and economic validation rather than pointwise numerical accuracy. By combining shadow-price diagnostics, saturation maps, and comparison with an analytical linear–quadratic benchmark, the neural solver is assessed in terms of its ability to recover economically meaningful switching behavior and marginal value signals. This focus distinguishes the present application from many existing studies that primarily benchmark neural PDE solvers against known analytical solutions or synthetic test equations.

In the context of energy systems, this approach offers a scalable and flexible alternative to classical finite-difference or dynamic programming methods, which become numerically fragile even in low-dimensional constrained settings. Once trained, the learned value function provides a differentiable and interpretable representation of operational flexibility that can be readily adapted to alternative price dynamics, regulatory environments, or cost structures without redesigning the numerical scheme.

5.6. Algorithmic summary and practical implications

The preceding sections introduced the HJB formulation and its neural approximation in a continuous-time setting. Algorithm 1 summarizes the complete numerical procedure used to approximate the value function and recover the associated feedback control. Rather than relying on grid-based discretization, the solver operates directly on randomly sampled collocation points drawn from the interior, boundary, and terminal domains.

Each training iteration enforces three complementary conditions: (i) satisfaction of the HJB equation in the interior of the state space, (ii) consistency with physical storage constraints at the boundaries, and (iii) alignment with the terminal valuation. The optimal control is recovered implicitly through the Hamiltonian maximization using gradients of the learned value function, ensuring economic consistency of the resulting policy.

Table 1
Neural architecture and training hyperparameters used in all numerical experiments.

Component	Value
Hidden layers	4
Neurons per layer	256
Activation function	tanh
Optimizer	Adam
Learning rate	1×10^{-3}
Training epochs (HJB phase)	6500
Warm-start epochs	1500
Interior collocation points / epoch	2048
Boundary points / epoch	512
Terminal points / epoch	1024

From a practical perspective, the algorithm highlights the modularity of the proposed framework. Alternative price dynamics, cost specifications, or policy constraints can be incorporated by modifying the drift, diffusion, or instantaneous payoff terms without altering the training structure. Once trained, the neural value function serves as a differentiable surrogate model that can be evaluated in real time, enabling scenario analysis, operational backtesting, and policy stress testing without re-solving the HJB equation.

For reference, a concise methodological comparison between the analytical linear–quadratic benchmark, the finite-difference diagnostic, and the proposed HJB–PINN solver is provided in Appendix A (Table A.4). This overview helps contextualize the numerical results discussed below by summarizing the role, dimensionality, constraint handling, and numerical behavior of the different solution approaches.

6. Empirical results, policy structure, and operational implications

This section presents numerical results for the proposed four-dimensional HJB–PINN framework under empirically calibrated electricity and gas price dynamics. The focus is on three complementary aspects: (i) to assess numerical stability and convergence of the neural HJB solver in a constrained multi-factor setting driven by real-world market data; (ii) to interpret the learned control policies using economically meaningful diagnostic quantities such as shadow prices and saturation regions; and (iii) to evaluate the resulting operational behavior through historically grounded backtesting experiments under realistic market frictions.

All numerical results reported in this section are obtained using the training procedure outlined in Algorithm 1. Unless stated otherwise, the neural architecture, loss components, and training hyperparameters are kept fixed across all experiments to ensure comparability between calibration regimes and operational scenarios. Table 1 summarizes the baseline network configuration used throughout this section.

6.1. Model setup and calibration

This section describes the structural components of the storage control problem that are kept fixed across all numerical experiments. Empirical calibration of the stochastic price dynamics is introduced separately in the subsequent sections.

Electricity and gas prices follow correlated Ornstein–Uhlenbeck dynamics,

$$dE_t = \kappa_E(\bar{E} - E_t)\,dt + \sigma_E\,dW_{E,t},$$
$$dG_t = \kappa_G(\bar{G} - G_t)\,dt + \sigma_G\,dW_{G,t},$$
$$dW_{E,t}\,dW_{G,t} = \rho\,dt,$$

as specified in Section 3.1. The storage state S_t evolves deterministically subject to round-trip efficiencies (η_c, η_d) and hard physical bounds $S_t \in [0, S_{\max}]$, with $S_{\max} = 100$. The planning horizon is $T = 1$, and future cash flows are discounted at rate $\beta = 0.05$.

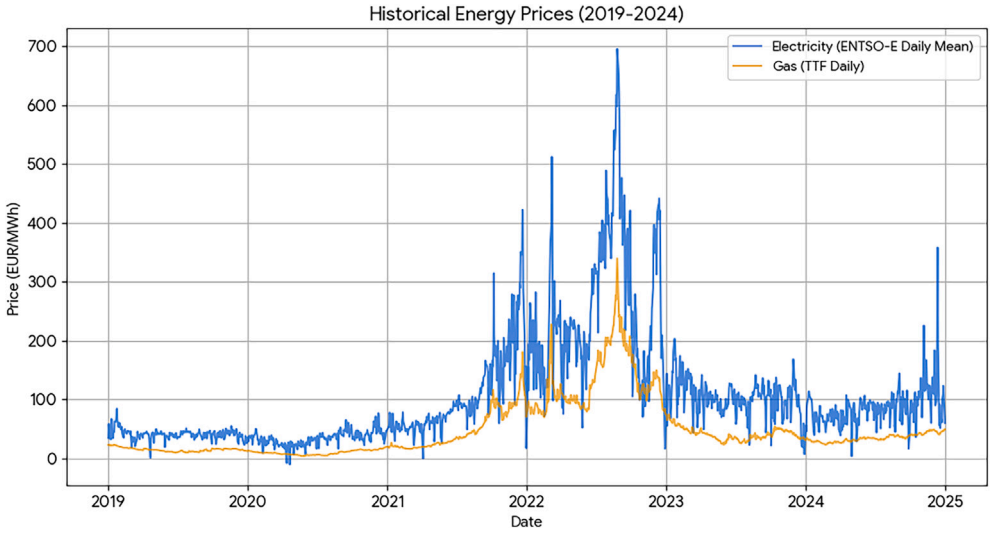


Fig. 2. Historical daily mean electricity prices (ENTSO-E, DE–LU) and natural gas prices (TTF), January 2019 to December 2024.

Unless stated otherwise, charging and discharging efficiencies are fixed at $\eta_c = 0.95$ and $\eta_d = 0.90$, consistent with values reported for utility-scale battery storage systems. The admissible storage control satisfies $|u| \le 40$, and storage adjustment costs are modeled by a quadratic penalty

$$C_S(u) = \tfrac{1}{2} R_{22} u^2, \qquad R_{22} = 0.25.$$

No explicit costs are imposed on electricity generation or wind dispatch in the numerical solver, i.e. $C_G \equiv 0$ and $C_W \equiv 0$, in order to isolate the interaction between stochastic prices and storage flexibility.

The dependence on the gas price enters the control problem through the effective electricity–gas price spread,

$$\text{spread}(E, G) = E - \eta_{\text{spark}}\, G,$$

which represents the implicit marginal cost of gas-fired generation. In all empirical experiments, the spark pass-through coefficient is fixed at $\eta_{\text{spark}} = 0.6$, reflecting a stylized but economically reasonable conversion rate between gas prices and marginal electricity production costs. This parameter is imposed exogenously and held constant across all numerical experiments to ensure internal consistency between model calibration, policy interpretation, and operational backtesting.

While the terminal valuation of stored energy is linear in this spread, the gas price influences the value function more broadly through its own stochastic dynamics and its correlation with electricity prices. As a result, variations in G affect both the immediate profitability of charging and discharging actions and the expected continuation value via correlated future price movements. Despite the apparent linearity of the terminal payoff, the gas price cannot be eliminated from the state vector, and the resulting HJB equation remains genuinely four-dimensional. This structure allows the learned policy to adapt endogenously across different gas-price regimes without imposing exogenous switching rules.

6.2. Historical data and descriptive price dynamics

The empirical analysis is based on daily electricity and natural gas price data covering the period from January 2019 to December 2024. This time span encompasses several distinct market regimes, including a phase of relative pre-crisis stability, the extreme price dislocations observed during the European energy crisis in 2022, and a subsequent period of high-price stabilization in 2023–2024. The presence of these markedly different regimes provides a rich empirical environment for evaluating storage operation under realistic market conditions.

Table 2
Empirical calibration parameters (2019–2024).

Process	Long-run Mean (μ)	Mean Reversion (κ)	Volatility (σ)
Electricity (E)	105.45	0.0507	30.83
Natural Gas (G)	49.37	0.0068	5.59

The dataset consists of two synchronized price series. Electricity prices correspond to day-ahead wholesale prices for the DE–LU bidding zone, covering Germany and Luxembourg, obtained from the ENTSO–E Transparency Platform. Natural gas prices are represented by daily settlement prices of Dutch TTF (Title Transfer Facility) futures, which serve as the dominant benchmark for marginal gas-fired generation costs in European electricity markets. Both series are aggregated to daily mean values and aligned by calendar date, resulting in a final dataset of 1510 common trading days.

Fig. 2 illustrates the historical evolution of electricity and gas prices over the sample period. The figure highlights the strong co-movement between the two markets as well as the pronounced volatility observed during the 2022 crisis. Electricity prices frequently exceeded 400 EUR/MWh during peak stress events, while gas prices temporarily surged above 200 EUR/MWh, reflecting the tight coupling between fuel markets and power prices under the European merit-order system.

Descriptive statistics confirm substantial heterogeneity in price behavior across the sample period. The long-run mean electricity price is approximately $\bar{E} = 105.45$ EUR/MWh, while the distribution exhibits pronounced right skewness and elevated volatility, with a standard deviation of $\sigma_E = 30.83$ EUR/MWh. The gas market displays lower volatility in absolute terms ($\sigma_G = 5.59$ EUR/MWh), but markedly higher persistence, reflecting slower mean reversion and trend-following behavior. These empirical characteristics motivate the use of correlated stochastic price dynamics and play a central role in shaping the optimal storage policy examined in the subsequent sections.

6.3. Stochastic parameter estimation

The parameters for the coupled Ornstein–Uhlenbeck processes were estimated using a discrete-time Maximum Likelihood (ML) approach. This methodology captures the mean-reverting behavior and the intrinsic volatility of the coupled markets. The estimated values, which provide the foundation for the operational stress-testing of the model, are reported in Table 2.

Table 3
PINN training diagnostics and stability metrics under empirical calibration (averaged over the final 2000 training epochs).

Metric	Mean value	Final state
Shadow price mean $\mathbb{E}[\partial_S V]$	15.29	16.16
Shadow price volatility std$(\partial_S V)$	25.43	26.63
Saturation region share (SRS)	0.357	0.462
Control variability std(u)	17.40	18.26

The analysis identifies a high instantaneous correlation ($\rho = 0.9230$), reflecting the structural dependency of electricity prices on marginal gas-fired generation in the European merit-order system. The electricity market exhibits significantly higher daily volatility and faster mean reversion compared to the gas market ($\kappa_G \approx 0.007$). These empirically grounded distributions form the basis for the empirical policy evaluation and operational backtesting presented in the subsequent sections.

6.4. Solver stability and diagnostic metrics

The numerical reliability of the HJB–PINN solver under empirically calibrated electricity and gas price dynamics is assessed using diagnostic quantities recorded during the terminal phase of training. Following a warm-start phase, the full HJB system is trained for 6500 epochs, with stability metrics aggregated over the final 2000 epochs. Table 3 reports the resulting diagnostics for the empirically calibrated model.

The shadow price

$$V_S(t, E, G, S) := \partial_S V(t, E, G, S)$$

represents the marginal economic value of stored energy and constitutes the central signal governing optimal charge–discharge decisions. Its mean level and dispersion therefore provide a direct diagnostic of whether the learned value function exhibits economically meaningful state dependence.

Due to the hard control constraint $|u| \le u_{\max}$, the optimal policy may become saturated in parts of the state space. We define the *saturation region share (SRS)* as the fraction of evaluated states for which $|u^*| \ge 0.95\, u_{\max}$. The complement of this region corresponds to the interior control domain, where storage decisions are governed by marginal value signals rather than binding constraints.

The shadow price exhibits a clearly positive mean level together with substantial dispersion, indicating that the learned value function remains strongly state-dependent under historical calibration. The magnitude of std$(\partial_S V)$ reflects the pronounced volatility and persistence of observed electricity and gas prices over the 2019–2024 period.

The reported saturation region share shows that a non-negligible fraction of the state space is constraint-dominated under empirical dynamics, particularly during extreme price realizations. At the same time, the remaining interior region displays meaningful variation in both the shadow price and the control signal, implying that the learned policy is not purely bang–bang but combines interior optimization with constraint-induced saturation.

Overall, these diagnostics confirm that the HJB–PINN solver remains numerically stable under historical calibration and converges to an economically interpretable value function. Differences relative to the synthetic experiments reported in Appendix B reflect structural properties of real market data rather than numerical artifacts, highlighting the role of volatility, persistence, and hard constraints in empirically grounded storage operation.

6.5. Policy interpretation and structural mapping

The learned operational logic is interpreted through a structural comparison with a simple linear–quadratic (LQ-like) reference policy defined on the electricity–gas price spread. Throughout this section, the *immediate spread* refers to the contemporaneous price signal

$$\text{spread}(E, G) = E - \eta_{\text{spark}}\, G,$$

which governs the instantaneous profitability of charging or discharging actions. In contrast, optimal decisions are determined by comparing this immediate spread to the *continuation value* of stored energy, captured by the storage shadow price $V_S = \partial_S V$. Together, these components determine the *effective economic incentive* to adjust inventory.

Fig. 3 visualizes both feedback laws at $t = 0$ on the two-dimensional slice (E, S) with the gas price fixed at its empirical mean $G_0 = \bar{G}$. While the PINN is trained on the full four-dimensional state space (t, E, G, S), this slice provides a transparent representation of the switching geometry under a representative gas-price regime.

The LQ-like reference is used as a structural baseline rather than a numerical target. Accordingly, the comparison focuses on monotonicity, switching-region orientation, and the implied charge–discharge logic across inventory levels. In both panels, the dominant feature is a diagonal switching structure in the (E, S) plane, consistent with storage arbitrage. Charging occurs only when the continuation value of inventory, as reflected by the shadow price V_S, sufficiently exceeds the immediate spread and associated adjustment costs. Conversely, when the immediate spread dominates the marginal value of holding energy in storage, the optimal policy prescribes discharging.

The empirically calibrated PINN policy exhibits a relatively narrow charging region. This reflects the presence of transaction costs, quadratic adjustment penalties, and efficiency losses, which together create a wide no-trade band. As a result, under historical price dynamics, inventory

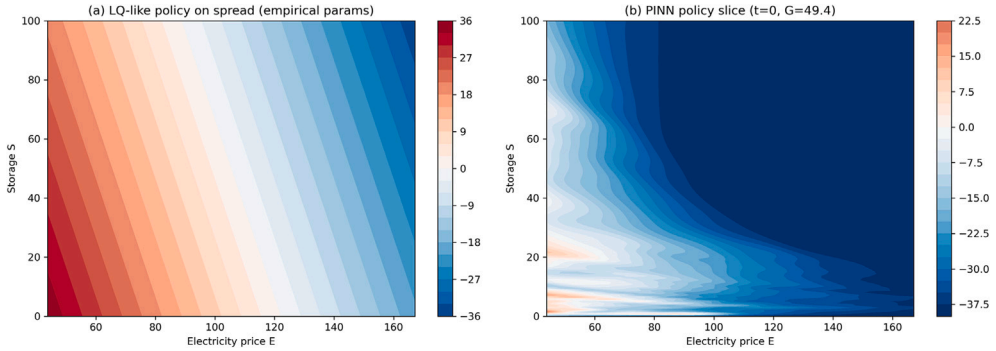


Fig. 3. Optimal storage policy comparison $u^*(E, S)$ under empirical calibration, evaluated at $t = 0$ and $G_0 = \bar{G}$. (a) LQ-like reference policy with planar feedback structure. (b) PINN policy slice at $t = 0$ and $G = \bar{G}$. The horizontal axis denotes the electricity price E (EUR/MWh), and the vertical axis denotes the storage level S (MWh).

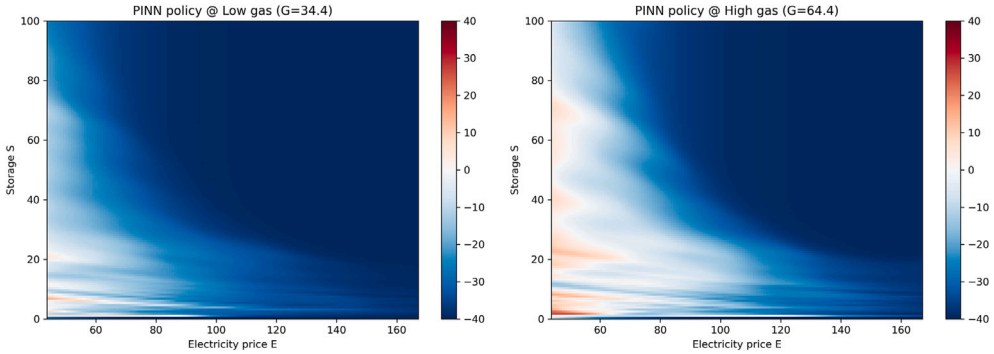


Fig. 4. Gas sensitivity of the empirically calibrated PINN policy at $t = 0$. Left: policy slice at $G_{\text{low}} = \bar{G} - 15$. Right: policy slice at $G_{\text{high}} = \bar{G} + 15$. Axes correspond to electricity price E (EUR/MWh) and storage level S (MWh).

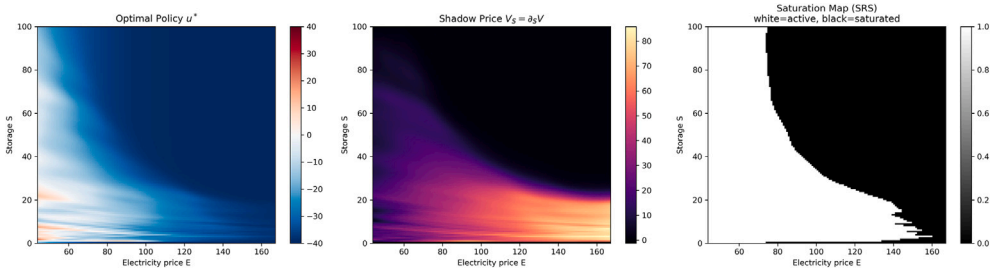


Fig. 5. Diagnostic visualization of the learned HJB–PINN solution at $t = 0$ and fixed gas price $G = \bar{G}$. Left: optimal storage control $u^*(E, S)$, where positive values correspond to charging and negative values to discharging. Center: shadow price $V_S(E, S) = \partial_S V$, representing the marginal economic value of stored energy. Right: saturation map, where white indicates the active interior control region ($|u^*| < 0.95\, u_{\max}$) and black denotes constraint-dominated regimes. In all panels, the horizontal axis corresponds to the electricity price E (EUR/MWh), and the vertical axis denotes the storage level S (MWh).

accumulation is optimal only in a limited subset of states characterized by unusually low electricity prices or elevated continuation values.

To illustrate how gas-price conditions remap the operating regions, Fig. 4 reports the same $t = 0$ policy slice under a lower and higher gas regime,

$$G_{\text{low}} = \bar{G} - 15, \qquad G_{\text{high}} = \bar{G} + 15.$$

Changes in G shift the immediate spread and therefore translate the switching boundary in the (E, S) plane, while preserving the overall directional structure of the learned control.

6.6. Shadow price structure and saturation effects

Building on the shadow price diagnostic introduced in Section 6.4, we now examine the structure of the storage shadow price and its interaction with physical control constraints in order to assess the internal consistency and economic interpretability of the learned HJB–PINN solution. The shadow price

$$V_S(t, E, G, S) := \partial_S V(t, E, G, S),$$

represents the marginal economic value of stored energy and captures the expected incremental benefit of holding one additional unit of energy in storage.

In the present formulation, no explicit running cost is imposed directly on the storage *state*. In particular, there is no inventory holding cost or penalty that depends explicitly on the level of stored energy $S(t)$. Instead, the shape of the shadow price emerges endogenously from three structural elements of the control problem: (i) the linear terminal valuation $V(T, E, G, S) = E\, S$, (ii) quadratic *control* costs penalizing rapid charge–discharge actions, and (iii) hard physical bounds on both storage capacity and admissible control rates. Together, these features induce a non-trivial, state-dependent marginal value landscape even in the absence of an explicit inventory penalty.

Fig. 5 illustrates the interaction between marginal economic signals and physical constraints for the trained four-dimensional HJB–PINN model, evaluated at initial time $t = 0$ and fixed gas price $G = \bar{G}$. The left panel reports the optimal feedback control $u^*(E, S)$, the center panel shows the learned shadow price $V_S(E, S)$, and the right panel displays the saturation map identifying constraint-dominated regions.

The shadow price surface is smooth and monotonic over a wide region of the state space, indicating stable convergence of the neural approximation and the absence of spurious oscillations. Higher marginal values are concentrated at lower inventory levels, reflecting the asymmetric option value of scarce storage capacity: when inventory is low, future arbitrage opportunities become more valuable due to the irreversibility imposed by the lower storage bound.

The saturation map highlights regions where the optimal control reaches its admissible limits. In these states, decisions are governed primarily by physical constraints rather than marginal economic trade-offs. Conversely, the interior control region corresponds to states where the policy responds smoothly to the balance between the current electricity price E and the shadow price V_S.

The transition between interior and constraint-dominated regimes induces visible kinks in both the policy and shadow-price surfaces at intermediate storage levels. These features arise naturally from switching between unconstrained and constraint-binding dynamics and do not indicate numerical instability or approximation error. Their coherent alignment across the control, shadow price, and saturation panels confirms that the learned policy is driven by economically meaningful marginal signals wherever operational flexibility is available.

Overall, the shadow price and saturation diagnostics demonstrate that the HJB–PINN solver captures the correct economic structure of the storage problem: marginal value governs decisions in flexible regions

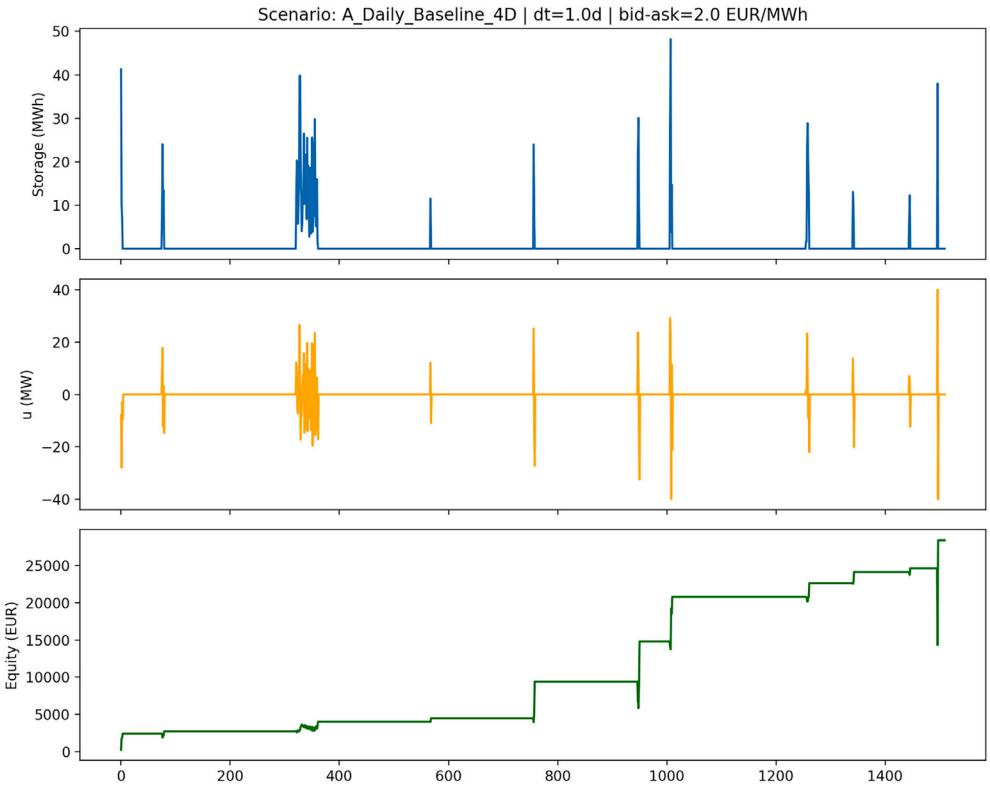


Fig. 6. Daily operation under baseline transaction costs. The learned policy exhibits intermittent, event-driven storage actions triggered by pronounced price movements, with cumulative equity increasing in discrete jumps rather than smoothly over time.

of the state space, while hard constraints appropriately dominate behavior near capacity limits. This separation yields a transparent and economically interpretable validation of the learned solution at the level of internal structure, independent of out-of-sample performance considerations.

6.7. Operational backtesting and friction sensitivity

This section concludes the numerical analysis by evaluating the operational behavior of the PINN-based storage controller under historical market calibration. Electricity and gas price time series are used directly to assess how the learned value function translates into realized storage actions in the presence of realistic market frictions. The analysis focuses on the emergence of storage activity and its robustness with respect to transaction cost assumptions.

We first consider daily operation under a baseline transaction cost calibration. Fig. 6 reports the resulting storage trajectory, control actions, and cumulative equity for a bid–ask spread of approximately ± 1 EUR/MWh around the mid-price, reflecting typical short-term wholesale market conditions. Under this calibration, the learned policy produces intermittent but economically meaningful storage actions. Charging and discharging are concentrated around pronounced price movements, while extended periods of inactivity persist when price signals are insufficient to overcome transaction costs and quadratic control penalties. The resulting equity curve grows in discrete steps, reflecting the episodic realization of arbitrage gains rather than continuous trading activity.

Fig. 7 reports the same daily backtest under a reduced bid–ask spread of approximately ± 0.25 EUR/MWh. At this lower transaction cost level, the qualitative structure of the optimal control remains unchanged. Storage actions occur at the same market events and retain their impulse-like character, indicating that the learned policy operates in the same economic regime. The primary effect of reduced transaction costs is a modest increase in terminal equity, while the timing and structure of storage utilization remain essentially identical.

Additional diagnostics at coarser decision frequencies are reported in Appendix C. In particular, the weekly backtest illustrates how temporal aggregation alters the effective arbitrage landscape and induces a distinct, event-driven operating regime.

At first glance, the sparse trading activity observed in the historical backtests may appear at odds with the large saturation regions identified in the policy maps. This is not a contradiction. Saturation regions characterize conditional optimal actions given a particular state (t, E, G, S), whereas realized backtesting reflects the actual state trajectory induced by historical prices.

In the full four-dimensional problem, time t is an explicit state variable. At each trading date, the controller optimizes based not only on current prices and inventory, but also on the remaining time to maturity. As the horizon shortens, the continuation value of storage decreases, and the policy becomes increasingly selective. As a result, the historical price path spends most of the time in interior no-trade regions, while saturation regions are entered only during rare and pronounced price dislocations.

Consequently, storage is exercised infrequently but aggressively. When the state enters a saturation region, the optimal response is a near-maximal charge or discharge action. Outside these regions, the optimal policy prescribes inaction, reflecting the option-like nature of storage under transaction costs and efficiency losses. This separation between large conditional action regions and sparse realized trading is a central implication of the stochastic control framework and is consistent with the interpretation of storage as a real option rather than a continuously cycled asset.

Taken together, these backtests demonstrate that the proposed PINN framework yields operationally realistic and economically interpretable control behavior under historical calibration. Daily storage operation is governed by sparse, event-driven decisions that are robust to moderate

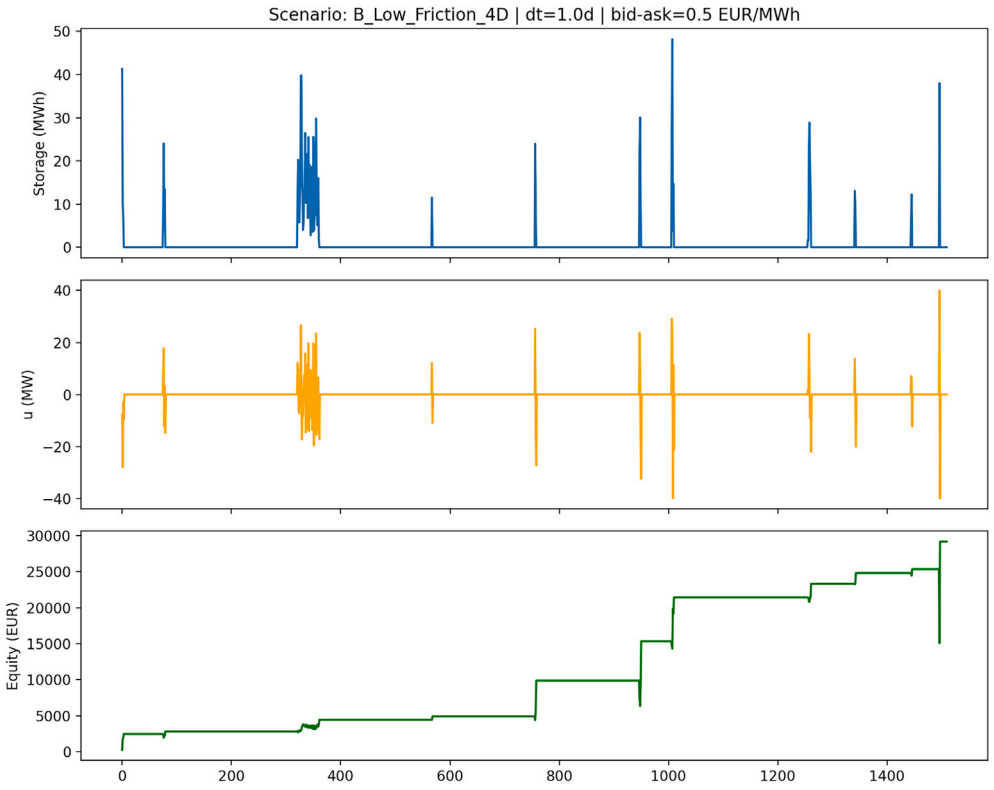


Fig. 7. Daily operation under reduced transaction costs. The optimal control exhibits the same event-driven structure as in the baseline case, with a slightly higher terminal equity due to lower execution costs.

changes in transaction costs, with differences primarily reflected in terminal profitability rather than in the qualitative structure of the learned policy. Across the daily scenarios considered, the terminal equity remains positive and of comparable magnitude under both transaction cost calibrations, indicating that reduced execution costs primarily affect net profitability rather than the timing or structure of storage decisions.

7. Discussion and policy implications

This section synthesizes the numerical findings and places them into a broader economic and policy context. Rather than focusing on absolute profitability, the discussion emphasizes what the learned policies reveal about the operational value of flexibility under realistic market conditions and existing market design constraints. The results indicate that storage value is highly state-dependent and threshold-driven, and that it is strongly shaped by market frictions, with important implications for both system design and regulatory frameworks. The discussion below builds on policies obtained through the unified training procedure outlined in Algorithm 1.

7.1. Flexibility as a non-linear economic option

A central insight emerging from the empirical analysis is that the value of flexibility is inherently non-linear. Across all historically calibrated scenarios, storage does not behave as a continuously active arbitrage device. Instead, it operates as a real option that remains dormant for extended periods and is exercised only when price movements exceed a critical threshold. Below this threshold, optimal behavior collapses into a no-trade region, even in the presence of ongoing generation.

This behavior reflects the combined effects of transaction costs, efficiency losses, and control penalties. Small or moderate price fluctuations are insufficient to justify inventory adjustments once these frictions are accounted for. As a result, flexibility value does not scale smoothly with volatility, but instead exhibits a discontinuous response: either storage is economically inactive, or it becomes sharply valuable during short-lived price dislocations. This option-like structure explains why empirically learned policies appear sparse and impulse-driven rather than cyclic.

7.2. Transaction costs and the suppression of storage activity

The numerical diagnostics highlight the decisive role played by transaction costs, in particular bid–ask spreads. Even moderate spreads introduce a fundamental asymmetry between immediate sale and deferred sale via storage. While direct generation-to-market transactions incur trading costs only once, storage-mediated arbitrage implicitly bears these costs twice: first through the opportunity cost of foregone immediate revenue, and later through the execution cost at liquidation.

This double exposure to market frictions creates a substantial economic barrier to storage utilization. Unless expected future prices exceed current prices by a margin sufficient to offset the full spread, efficiency losses, and ramping penalties, it is rational for the controller to leave the storage empty. The prevalence of near-zero inventory levels in several scenarios therefore does not indicate a modeling failure, but rather a consistent economic response to historically observed price dynamics and trading frictions. This mechanism is consistent with the backtesting results, where reductions in transaction costs affect terminal profitability only marginally, without altering the qualitative structure of optimal storage operation.

7.3. Temporal granularity and the role of short-term price signals

The comparison between daily and weekly decision frequencies further clarifies the conditions under which storage flexibility becomes valuable. At daily resolution, short-term price dispersion occasionally generates arbitrage opportunities large enough to overcome transaction costs, leading to sparse but economically meaningful storage actions. At weekly resolution, however, these short-lived price signals are largely averaged out, while trading frictions remain unchanged.

As a consequence, the learned policy at weekly frequency converges almost entirely to a no-trade regime, as illustrated by the weekly backtesting results reported in Fig. C.11 in Appendix C. This finding underscores the importance of temporal granularity in market design: flexibility assets derive much of their value from access to high-frequency price signals. Coarser settlement intervals or aggregated pricing mechanisms can unintentionally suppress the very arbitrage incentives that justify investment in storage capacity.

7.4. Relation to linear–quadratic benchmark models

The observed behavior contrasts sharply with the smooth and persistent cycling predicted by analytical linear–quadratic (LQ) benchmark models examined earlier in the paper. In the LQ framework, continuous control arises naturally from quadratic cost structures and frictionless trading assumptions. Under these conditions, storage responds gradually to marginal price incentives and remains active across a wide range of states.

The empirically calibrated PINN solution departs from this paradigm once realistic transaction costs and historical price dynamics are introduced. Smooth LQ-type control gives way to threshold-based, option-like behavior characterized by extended inactivity punctuated by discrete interventions. This divergence is not a numerical artifact, but a direct consequence of relaxing the linearity and frictionless assumptions embedded in classical models. The comparison highlights the risk of overestimating practical storage utilization when relying exclusively on stylized analytical benchmarks.

7.5. Implications for market design and investment decisions

From a policy perspective, these results suggest that the economic value of storage flexibility is highly sensitive to market rules governing transaction costs and temporal resolution. Market designs that rely on coarse settlement intervals or impose excessive trading frictions may inadvertently neutralize the incentives for active storage operation, even when physical flexibility is available.

For investors, the results emphasize that storage profitability cannot be inferred from volatility alone. Instead, value depends on the frequency and magnitude of price dislocations relative to trading frictions. The shadow value of inventory, as reflected in the marginal value function, provides a natural diagnostic for identifying periods of flexibility scarcity and for assessing the marginal benefit of additional storage capacity under specific market conditions.

Overall, the findings position storage not as a continuously exploited arbitrage mechanism, but as a conditional asset whose value materializes only under sufficiently favorable market configurations. By revealing where and why flexibility becomes economically inactive, the proposed PINN framework offers a transparent tool for evaluating both operational strategies and policy interventions in low-carbon power systems.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the author used **ChatGPT (OpenAI, GPT-5 model)** and **Gemini (Google AI)** to assist with language editing, code refinement, and technical phrasing of the manuscript. After using these tools, the author carefully reviewed and edited the content and takes full responsibility for the final version of the text.

Declaration of competing interest

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.

Acknowledgements

The author gratefully acknowledges Yerkin Kitapbayev and Paul V Johnson for their valuable comments and insightful suggestions, which greatly improved the clarity and rigor of this work.

Appendix A. Additional methodological diagnostics

A.1. Finite-difference benchmark in two dimensions

For completeness, we consider a two-dimensional finite-difference (FD) discretization of the HJB equation restricted to the (E, S) subspace, with the gas price fixed at its long-run mean $G = \bar{G}$. The purpose of this exercise is not to construct a competitive alternative to the proposed PINN solver, but to document the numerical behavior of classical grid-based policy iteration schemes in a constrained, advection-dominated setting.

The HJB equation is discretized on a uniform grid using an upwind treatment for the mean-reverting drift in the electricity price, a central discretization for the diffusion term, and reflecting (Neumann-type) boundary conditions at the storage limits. Backward time integration is performed using an implicit Euler scheme. The resulting nonlinear discrete problem is solved via Howard-type policy iteration. Several stabilization strategies are tested, including under-relaxation of the policy update, temperature-based smoothing of the control saturation, and semi-Lagrangian treatment of the storage advection term.

Across all tested variants, robust convergence of the policy iteration is not observed. In particular, the maximum policy update $\max |u^{k+1} - u^k|$ fails to decay monotonically across outer iterations and may increase, indicating a lack of contractivity in the discrete policy map. This behavior is consistent with a bang–bang switching mechanism commonly encountered in constrained HJB problems: small perturbations in the discrete storage gradient can trigger alternating activation of opposite control bounds, thereby preventing stable convergence of Howard iterations even under implicit time stepping.

Representative outcomes for several FD variants are summarized in Table A.5. While saturation of the control near physical limits is

Table A.4
Methodological comparison of solution approaches considered in this study. “Dim.” denotes the effective dimensionality of the state space. “Constr.” indicates whether hard physical or control constraints are explicitly enforced. “Stability” summarizes the observed numerical behavior of the solver. “Interpr.” reflects the economic interpretability of the resulting policy. “Role” describes the conceptual function of each method within the paper. Abbreviations: LQ = linear–quadratic benchmark, FD = finite difference, PINN = physics-informed neural network.

Method	Dim.	Constr.	Stability	Interpr.	Role
LQ	3	No	Closed-form	High	Reference
FD	2	Yes	Unstable	Medium	Diagnostic
PINN (ours)	4	Yes	Stable	High	Main solver

Table A.5
Finite-difference benchmark in two dimensions: qualitative behavior of policy iteration under bounded controls. FD–E–PI denotes an explicit finite-difference scheme combined with policy iteration. FD–I–PI corresponds to an implicit (backward Euler) discretization with Howard-type policy iteration. FD–I–PI–S additionally incorporates stabilization mechanisms, including under-relaxation, control smoothing, and semi-Lagrangian treatment of storage advection.

Scheme	Time stepping	Observed behavior
FD–E–PI	Explicit	Divergent
FD–I–PI	Implicit	Flip–flop
FD–I–PI–S	Implicit	Oscillatory

economically meaningful, a non-decaying policy update norm reflects numerical instability of the discrete solver rather than convergence to a stable fixed point.

These observations highlight a practical limitation of classical grid-based policy iteration methods for constrained, advection-dominated HJB problems. Even in two dimensions, robust convergence is difficult to guarantee without highly specialized monotone discretizations. This motivates the mesh-free HJB–PINN framework adopted in this paper, which remains stable in the full four-dimensional state space and yields economically interpretable policies without grid-induced oscillations.

Appendix B. Additional numerical experiments under synthetic calibration

This appendix reports supplementary numerical results obtained under a synthetic calibration of the joint electricity–gas price dynamics. These experiments are not intended as a realistic market representation, but serve as controlled diagnostic tests that complement the empirically calibrated results presented in the main text. In particular, they are used to (i) verify solver stability, (ii) illustrate the structural relation between the PINN solution and analytical benchmarks, and (iii) provide a clean reference setting for sensitivity analysis.

B.1. Synthetic model specification

The synthetic experiments are based on mean-reverting Ornstein–Uhlenbeck dynamics for electricity and gas prices with moderate persistence and volatility. The parameters are summarized below:

$$\bar{E} = 85, \quad \kappa_E = 0.45, \quad \sigma_E = 24, \qquad \bar{G} = 35, \quad \kappa_G = 0.38, \quad \sigma_G = 9,$$

with instantaneous correlation $\rho = 0.55$ and discount rate $\beta = 0.05$. The storage model, control bounds, efficiencies, and cost parameters are identical to those used in the empirical calibration. The PINN architecture and loss construction remain unchanged.

Training proceeds via a two-stage procedure consisting of a warm-start phase (1500 epochs) followed by full HJB training (6500 epochs). Control saturation is defined consistently with the main text as

$$\mathrm{SRS} = \mathbb{P}\big(|u^*| > 0.95\, u_{\max}\big),$$

and diagnostic statistics are averaged over the final 2000 epochs.

B.2. Policy structure and LQ comparison

Fig. B.8 compares the learned PINN policy with a linear–quadratic (LQ) reference constructed on the spark spread, evaluated at $t = 0$ and fixed gas price $G = \bar{G}$. As expected in this stylized setting, the LQ policy exhibits a strictly planar feedback structure, while the PINN solution recovers a similar global switching geometry with additional curvature induced by state constraints and nonlinearities.

This experiment confirms that the PINN solver reproduces the expected LQ structure in the interior of the state space when dynamics are well behaved, while retaining the ability to handle constraints without imposing linearity.

B.3. Shadow price structure and saturation diagnostics

Fig. B.9 reports the diagnostic panel for the synthetic case, showing the optimal policy $u^*(E, S)$, the shadow price $V_S = \partial_S V$, and the saturation map at $t = 0$ and $G = \bar{G}$. Compared to the empirical calibration, the shadow price surface is smoother and less persistent over time, reflecting the more regular and controlled nature of the synthetic price dynamics.

Quantitative diagnostics summarizing shadow price statistics, control variability, and saturation prevalence are reported in Table B.6. As expected, the saturation region share is lower than under historical calibration, reflecting the reduced role of binding constraints in the synthetic setting. At the same time, shadow price variability remains pronounced within the interior control region, providing a clean and well-controlled baseline for structural comparison.

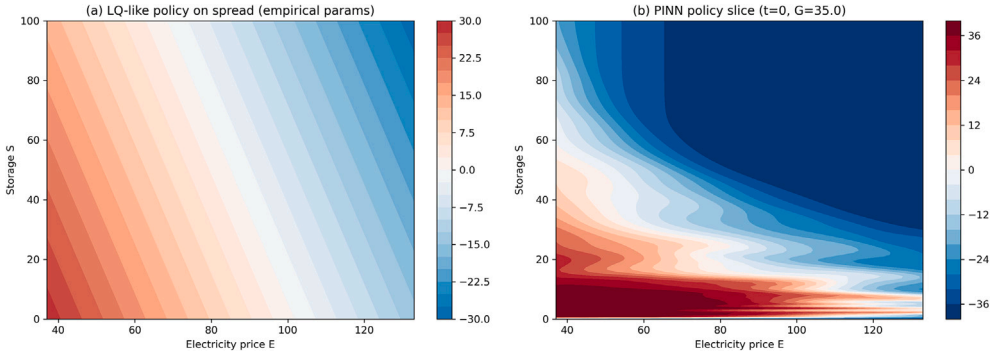


Fig. B.8. Policy comparison under synthetic calibration at $t = 0$ and $G = \bar{G}$. Left: LQ benchmark with planar feedback structure. Right: PINN policy slice learned from the full four-dimensional HJB. Axes correspond to electricity price E and storage level S.

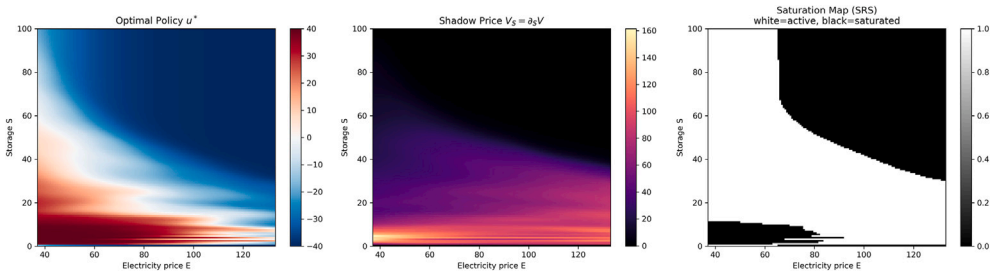


Fig. B.9. Diagnostic panel under synthetic calibration at $t = 0$ and $G = \bar{G}$. Left: optimal control $u^*(E, S)$. Center: shadow price $V_S(E, S)$. Right: saturation map, where white indicates interior control ($|u^*| < 0.95\, u_{\max}$) and black denotes constraint-dominated regions.

Table B.6
PINN diagnostics under synthetic calibration. Mean values are computed over the final 2000 training epochs.

Metric	Mean value	Final state
Shadow price mean $E[\partial_S V]$	26.74	28.62
Shadow price volatility std$(\partial_S V)$	34.82	37.17
Saturation region share (SRS)	0.27	0.40
Control variability std(u)	24.76	27.43

B.4. Gas price sensitivity

Finally, Fig. B.10 illustrates the sensitivity of the learned policy to shifts in the gas price, evaluated at $G = \bar{G} \pm 15$. As in the empirical case, changes in the gas price shift the effective spark spread and translate the switching boundary in the (E, S) plane. Higher gas prices move the charging–discharging threshold toward higher electricity prices, while lower gas prices shift it in the opposite direction, thereby altering the set of states in which storage charging becomes optimal.

Overall, the synthetic experiments confirm that the proposed HJB–PINN framework behaves consistently in a controlled environment, reproduces classical benchmark structures where appropriate, and provides a stable reference point against which the empirically calibrated results in the main text can be interpreted.

Appendix C. Weekly backtesting and temporal aggregation effects

This appendix reports an additional backtesting experiment designed to isolate the effect of temporal aggregation on the learned storage policy. In contrast to the daily backtests discussed in the main text, control decisions are restricted to a weekly frequency, while all other model components—price paths, transaction costs, physical constraints, and the learned value function—remain unchanged.

Fig. C.11 displays the resulting storage trajectory, control actions, and cumulative equity for the weekly decision scenario. The top panel shows the storage level, the middle panel reports the applied control u, and the bottom panel depicts the resulting equity process.

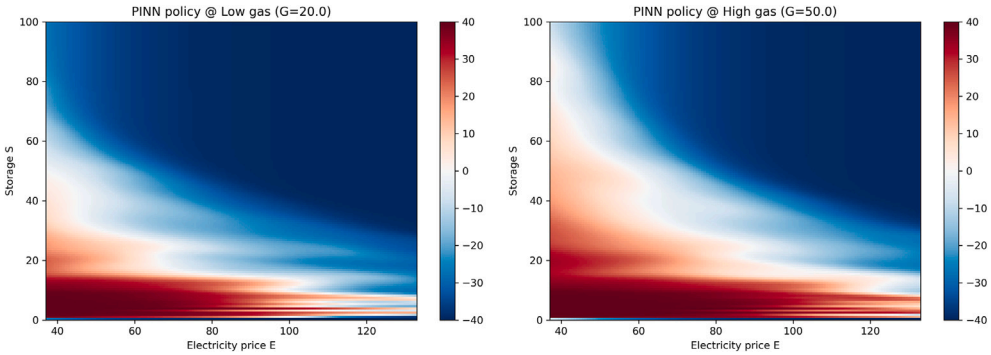


Fig. B.10. Gas price sensitivity of the PINN policy under synthetic calibration. Left: low gas price $G = \bar{G} - 15$. Right: high gas price $G = \bar{G} + 15$.

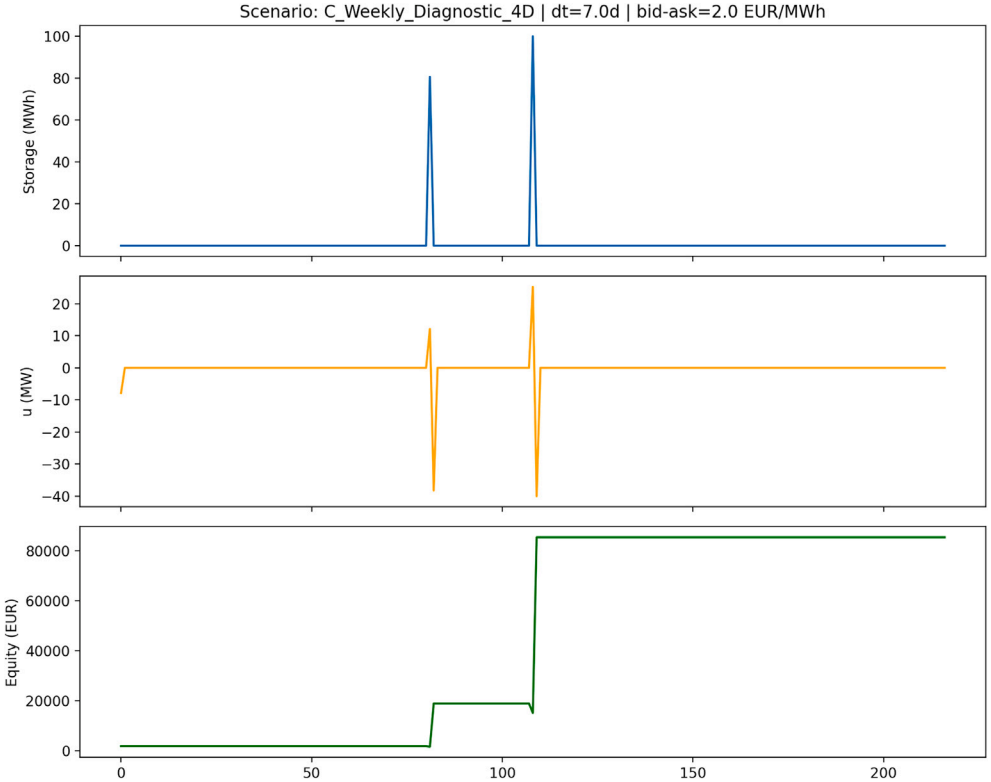


Fig. C.11. Weekly backtest of the 4D HJB–PINN policy under empirical price dynamics. Top: storage level S. Middle: applied control u. Bottom: cumulative equity. Control decisions are taken at weekly intervals with the same bid–ask spread as in the daily baseline scenario.

The weekly backtest exhibits an extreme form of event-driven behavior. Storage remains inactive for most of the sample, punctuated by a small number of large charging and discharging actions concentrated around pronounced price dislocations. As a result, both the storage level and the equity trajectory evolve in discrete jumps rather than through gradual adjustments.

This behavior is economically intuitive. At weekly resolution, short-lived price fluctuations are averaged out, while transaction costs and efficiency losses remain unchanged. Consequently, only exceptionally strong and persistent price signals are able to overcome the effective no-trade region implied by the learned value function. When such signals occur, the optimal response is to utilize storage capacity aggressively over a short horizon, leading to impulse-like control actions.

Importantly, the qualitative structure of the control policy remains consistent with the daily backtests. Storage actions are clustered around distinct market events rather than forming regular charge–discharge cycles. The sparsity of activity is therefore not an artifact of the chosen decision frequency, but a direct consequence of the threshold-based structure of the empirically learned value function.

Overall, the weekly backtest highlights the critical role of temporal resolution in the economic value of flexibility. While coarser decision intervals may amplify realized profits when rare but pronounced opportunities occur, they substantially reduce the frequency of storage utilization. This finding reinforces the interpretation of storage as a conditional, option-like asset whose value materializes only when market signals are both sufficiently strong and persistent.

Data availability

Data will be made available on request.

References

[1] Benth FE, Benth JS, Koekebakker S. Stochastic modelling of electricity and related markets, vol. 11. World Scientific; 2008.

[2] Pereira MVF, Pinto LMVG. Multi-stage stochastic optimization applied to energy planning. Math Program 1991;52(1):359–75.

[3] Thompson M, Davison M, Rasmussen H. Natural gas storage valuation and optimization: a real options application. Nav Res Logist 2009;56(3):226–38.

[4] Chen Z, Forsyth PA. Implications of a regime-switching model on natural gas storage valuation and optimal operation. Quantitative Finance 2010;10(2):159–76.

[5] Carmona R, Ludkovski M. Valuation of energy storage: an optimal switching approach. Quant. Finance 2010;10(4):359–74.

[6] Bertsekas DP. Dynamic programming and optimal control. 4th ed. vol. II. Belmont, MA: Athena Scientific; 2017.

[7] Sirignano J, Spiliopoulos K. DGM: a deep learning algorithm for solving partial differential equations. J Comput Phys 2018;375:1339–64.

[8] Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J Comput Phys 2019;378:686–707.

[9] Han J, Jentzen A, Weinan E. Solving high-dimensional partial differential equations using deep learning. Proc Natl Acad Sci 2018;115(34):8505–10.

[10] Beck C, Gonon L, Jentzen A. Overcoming the curse of dimensionality in the numerical approximation of high-dimensional semilinear elliptic partial differential equations. Part Differ Equ Appl 2024;5(6):31.

[11] Fleming WH, Soner HM. Controlled markov processes and viscosity solutions. Springer; 2006.

[12] Yong J, Zhou XY. Stochastic controls: Hamiltonian systems and HJB equations, vol. 43. Springer Science & Business Media; 1999.

[13] Anderson BDO, Moore JB. Optimal control: linear quadratic methods. Mineola, NY: Dover Publications; 2007, reprint of the 1989 Prentice Hall edition.

[14] Duncan TE, Pasik-Duncan B. A direct approach to linear-quadratic stochastic control. Opusc Math 2017;37(6):821–7.

[15] Beneš VE. Existence of optimal strategies based on specified information, for a class of stochastic decision problems. SIAM J Control 1970;8(2):179–88.

[16] Øksendal B, Sulem A. Applied stochastic control of jump diffusions, vol. 3. Springer; 2007.

[17] Smith JE, McCardle KF. Valuing oil properties: integrating option pricing and decision analysis approaches. Oper Res 1998;46(2):198–217.

[18] Szabó DZ, Martyr R. Real option valuation of a decremental regulation service provided by electricity storage. Philos Trans R Soc A Math Phys Eng Sci 2017;375(2100):20160300.

[19] Szabó DZ, Duck P, Johnson P. Optimal trading of imbalance options for power systems using an energy storage device. Eur J Oper Res 2020;285(1):3–22.

[20] Johnson P, Szabó DZ, Duck P. Optimal trading with regime switching: numerical and analytic techniques applied to valuing storage in an electricity balancing market. Eur J Oper Res 2024;319(2):611–24.

[21] Buehler H, Gonon L, Teichmann J, Wood B. Deep hedging. Quantitative Finance 2019;19(8):1271–91.

[22] Becker S, Cheridito P, Jentzen A. Deep optimal stopping. J Mach Learn Res 2019;20(74):1–25.

[23] Misyris GS, Venzke A, Chatzivasileiadis S. Physics-informed neural networks for power systems. In: 2020 IEEE power & Energy Society general Meeting (PESGM). IEEE; 2020. p. 1–5.

[24] Cybenko G. Approximation by superpositions of a sigmoidal function. Math Control Signals Syst 1989;2(4):303–14.

[25] Hornik K. Approximation capabilities of multilayer feedforward networks. Neural networks 1991;4(2):251–7.