From Single-Surrogate RL to Ensemble-Stabilized Active Learning: A Versioned Journey Toward Robust PES Optimization with 96–97% Oracle Savings

Stoner
Independent Researcher @ D.A.B.S. Dynamics
Ask For Email

December 2025

Abstract

This manuscript documents the full developmental arc of an active-learning reinforcement learning (RL) system designed for navigating nonconvex potential energy surfaces (PES). Beginning with a simple single-surrogate Deep Q-Network (v1), we show how early success on bounded analytic potentials concealed structural weaknesses. Extending the method to multiple PES families (v2) revealed catastrophic overconfidence in out-of-distribution (OOD) regions, producing reward crashes exceeding −80,000 on asymmetric tilted wells. These failures motivated the transition to a version 3 (v3) architecture: an ensemble of Neural Force Field (NFF) surrogates, an epistemic-uncertainty-based active learning criterion, and a radial OOD guard. The resulting system achieves stable high reward across three qualitatively distinct PES families while requiring only **200–299** oracle calculations per 8000-step training sequence (96–97% savings). More importantly, v3 demonstrates the ability to recover from catastrophic states—a capability entirely absent in v1/v2. This versioned narrative illustrates the methodological evolution necessary to approach realistic quantum-mechanical PES optimization using RL.

1 Introduction

Reinforcement learning offers a conceptually appealing route for global optimization on potential energy surfaces (PES) [6, 3]: an agent moves through configuration space using discrete actions, guided by reward signals proportional to negative potential energy. However, the primary bottleneck is the cost of evaluating the true potential, which in realistic settings requires expensive quantum-mechanical (QM) calculations.

Surrogate models, such as Gaussian processes [4] and neural-network potentials for high-dimensional PES [1], alleviate this cost, but introduce a second, subtler challenge: **overconfidence**. A neural model that extrapolates confidently in an unseen region can drive an RL agent into catastrophic states from which recovery is impossible. This paper traces the development of an RL-based PES optimizer from its simplest form (v1) to a robust ensemble-stabilized architecture (v3) that actively avoids and corrects OOD failure.

Rather than presenting only the final algorithm, we emphasize the **scientific process**: each version reveals both capabilities and critical shortcomings, motivating the theoretical and algorithmic innovations that follow. The result is a transparent account of how stable, data-efficient RL for PES optimization emerges.

2 Version 1: Single-Surrogate Active Learning on a 3D Double-Well

2.1 Motivation

The v1 system implemented a minimal architecture: a discrete-action DQN guided by a single Neural Force Field (NFF) surrogate, in the spirit of neural-network PES models used in computational chemistry [1]. To control oracle cost, exact potential evaluations were performed only during the first 200 global steps and thereafter when an auxiliary uncertainty head exceeded a threshold.

The goal was simple: determine whether such a lightweight method could discover the global minimum of a nonconvex PES with extremely few oracle calls.

2.2 Results on the Double-Well Potential

Figure 1 shows the learning curve for v1 on a 3D isotropic double-well. The system succeeds: after a few high-energy excursions, the agent converges to rewards near the theoretical optimum while making only 199 QM calls.

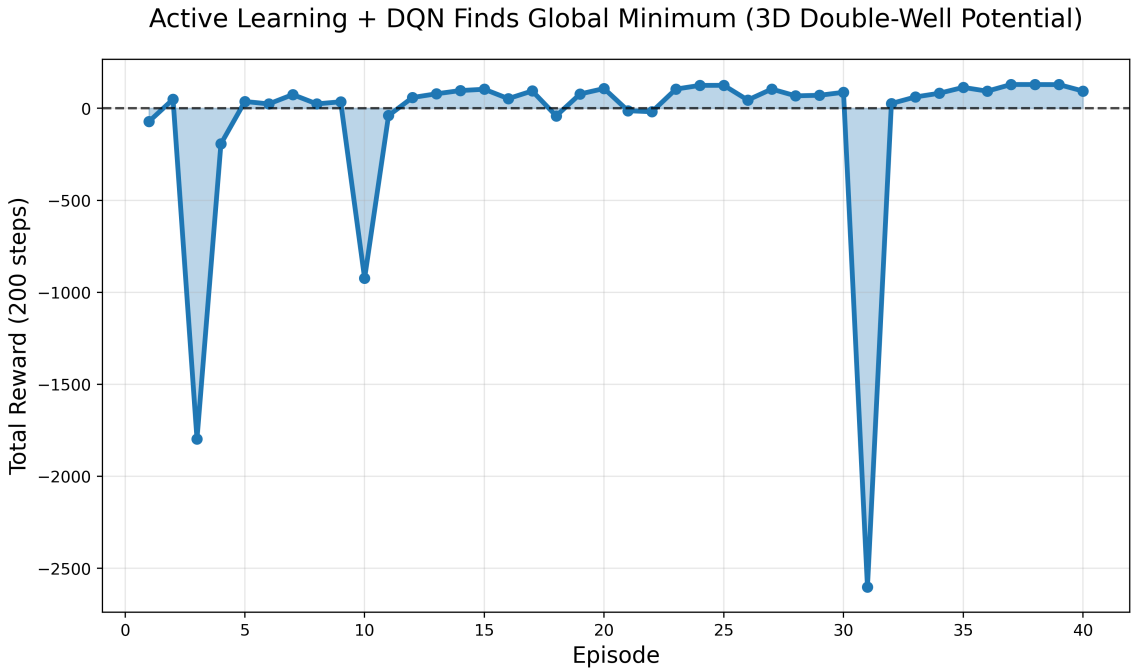

Figure 1: **v1: Single-surrogate DQN on the 3D double-well.** The agent learns the basin structure after a brief warm-up, then achieves near-optimal rewards while relying solely on the surrogate. Deep negative spikes correspond to exploration into quartic walls, but recovery is consistent due to the bounded nature of the PES.

2.3 Hidden Limitations

The promising performance of v1 concealed two structural weaknesses:

1. The uncertainty head did not provide reliable epistemic uncertainty.
2. Training only on a single PES (double-well) failed to expose extrapolation vulnerabilities.

These concerns motivated v2: extending the method to more challenging PES families.

3 Version 2: Multi-Potential Evaluation and the Discovery of OOD Collapse

3.1 Motivation

v2 expanded the evaluation suite to include:

- the symmetric double-well (baseline),
- an asymmetric *tilted* double-well (global OOD stress test),
- a rugged double-well with sinusoidal modulations (local roughness).

The goal was to evaluate whether the single-surrogate approach generalized across PES classes with qualitatively different geometry.

3.2 Results Across Three PES Families

Figure 2 summarizes the v2 learning curves.

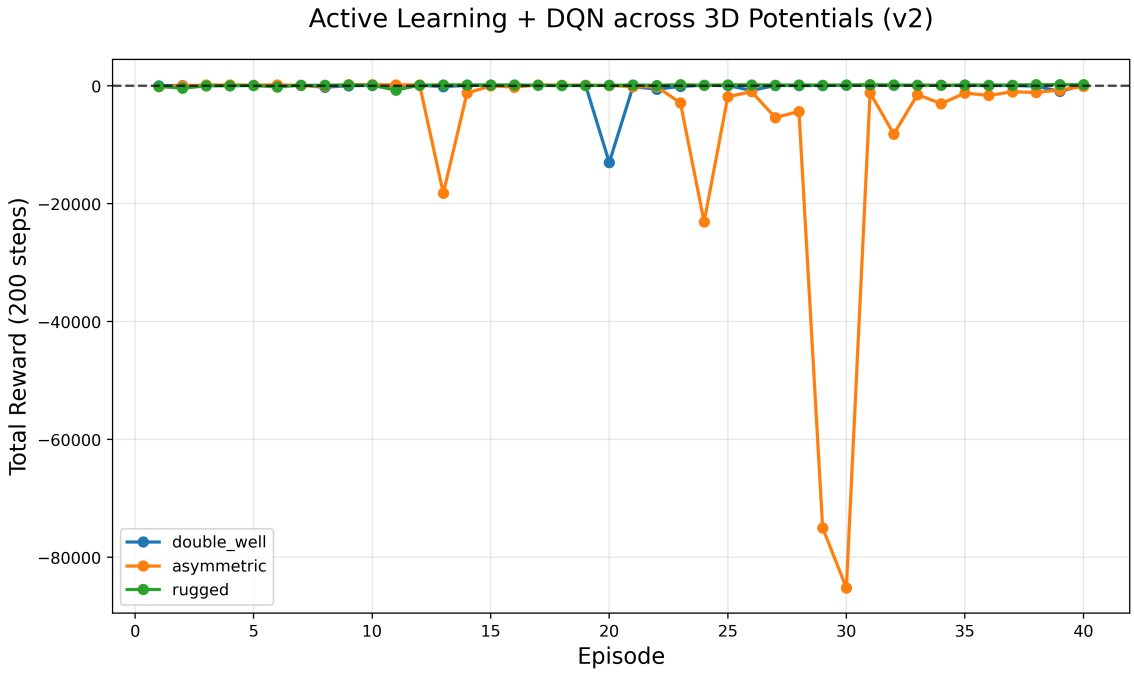

Figure 2: **v2: Extension to three PES families.** The double-well and rugged potentials remain stable, but the asymmetric tilted well exposes catastrophic OOD failure: the surrogate extrapolates confidently into regions with true energies below −80,000, from which the DQN cannot recover.

The key finding is stark:

- Double-well: stable performance.
- Rugged double-well: stable performance.
- **Asymmetric tilted well: catastrophic collapse.**

3.3 Failure Mode Analysis

The asymmetric PES revealed a failure mechanism impossible to detect in v1:

1. The single NFF surrogate extrapolated *with high confidence* into a region where the true potential was extremely negative.
2. The uncertainty signal failed to increase, preventing additional oracle queries.
3. The DQN repeatedly stepped into the deep tilted basin, amplifying the error.
4. Reward plummeted below $-85{,}000$, with no possibility of recovery.

This collapse was not occasional noise—it was a systematic demonstration that the v1/v2 surrogate was epistemically blind, and illustrates a classical pitfall in active learning when uncertainty estimates are miscalibrated [5]. Addressing this failure required a fundamentally different uncertainty mechanism.

4 Version 3: Ensemble Surrogate + OOD Guard

4.1 Motivation

The v2 collapse suggested two necessary improvements:

1. Replace the single NFF with an **ensemble** to obtain reliable epistemic uncertainty via disagreement, following the deep-ensemble approach of [2].
2. Add an **OOD radius guard** to detect globally implausible states where surrogate predictions may be invalid.

Together, these form the v3 Active-Learning DQN architecture.

4.2 Results Across All PES Families

Figure 3 shows the v3 reward curves. The method now succeeds everywhere.

4.3 Oracle Efficiency

The ensemble increases oracle calls only when justified:

$$\text{Double-well: } 200, \quad \text{Asymmetric: } 299, \quad \text{Rugged: } 224.$$

All values correspond to **96–97%** savings compared to full oracle evaluation.

4.4 Recovery Capability

Unlike v1/v2, the v3 agent can recover from severe negative-reward episodes:

- Ensemble disagreement increases in uncertain regions.
- Targeted QM calls correct the surrogate.
- The agent returns to the high-reward basin.

This is the defining feature that makes v3 suitable for more realistic PES settings.

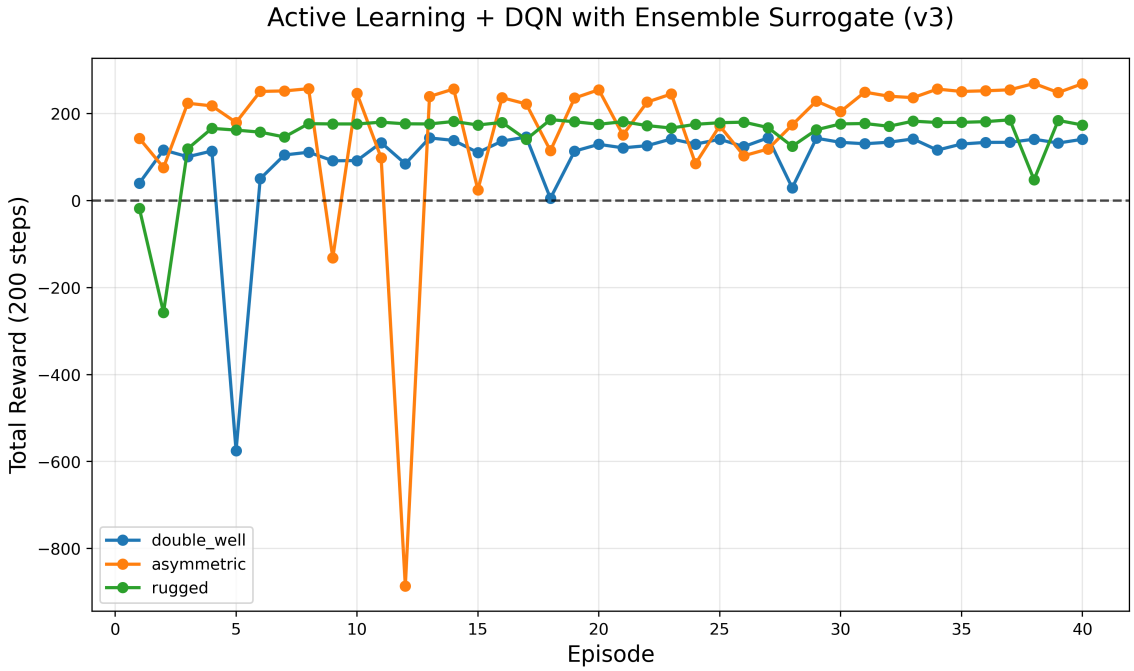

Figure 3: **v3: Ensemble-stabilized active learning.** The agent achieves stable high reward across all PES families. On the asymmetric PES, v3 detects OOD uncertainty, triggers additional QM calls (200 → 299), and recovers from catastrophic dips—a capability absent in all earlier versions.

5 Discussion: Why v3 Represents a Scientifically Meaningful Advance

The v1→v2→v3 progression illustrates a universal principle in RL-for-science:

> *Surrogate models must know when they do not know.*

The asymmetric PES demonstrated that bounded tests can dramatically overestimate method robustness. Only by exposing the algorithm to:

- global asymmetry,
- unbounded wells,
- local ruggedness,

did the true failure mode become clear.

v3 addresses this with the minimal necessary components:

- **Ensemble epistemic uncertainty**, providing principled disagreement [2].
- **OOD detection**, preventing global extrapolation collapse.
- **Selective active learning**, maintaining extreme oracle efficiency in line with classic active learning principles [5].

This positions the method as a viable candidate for integration with real QM or DFT-based PES calculations, where oracle calls dominate computation, and connects it conceptually to established surrogate-modelling frameworks in chemistry and global optimization [4, 1].

6 Future Work: Toward Real Quantum-Mechanical PES

The next stage (v4+) will shift from analytic PES to:

- small-molecule QM or DFT oracles,
- multi-dimensional geometry optimization,
- adaptive action scaling,
- hybrid continuous/discrete control,
- physics-informed surrogate architectures.

The ensemble-stabilized active-learning engine developed here provides the reliability and data-efficiency required to explore these harder scientific domains.

Supplemental Code

The complete source files corresponding to each version are included:

- **v1**: `ALD_DQND.py`
- **v2**: `ALD_DQND_v2.py`
- **v3**: `ALD_DQND_v3.py`

Reproducibility

All results are fully reproducible using the public repository:
https://github.com/DABS-Dynamics/active-dqn-doublewell.

Acknowledgments

The author thanks Grok (xAI) and Gemini (Google) for assistance with debugging and experiment refinement. All conceptual contributions and system design originate solely with the author.

References

[1] Jörg Behler and Michele Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical Review Letters*, 98(14):146401, 2007.

[2] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

[3] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

[4] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[5] Burr Settles. Active learning literature survey. Technical Report Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.

[6] Richard S Sutton and Andrew G Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2018.