

Research Article

Light as a Quantum Agent: Bridging Path Integrals and Reinforcement Learning

Bhushan Poojary^{*} 

Physics BSE, NIMS University, Jaipur, India

Abstract

We propose a novel and interdisciplinary conceptual framework that bridges Feynman's path integral formulation of quantum mechanics with reinforcement learning (RL) in artificial intelligence. In the path integral approach, a quantum system does not follow a single predetermined trajectory but instead explores all possible paths simultaneously, assigning each a complex amplitude weighted by $e^{iS/\hbar}$, where S represents the classical action. Constructive and destructive interference across these paths acts as a natural filter, amplifying trajectories of stationary action and suppressing suboptimal ones, thereby leaving paths of least action as the observable outcomes. We argue this process is strongly analogous to a quantum agent evaluating an entire policy space in superposition, where interference effectively encodes a reward mechanism that eliminates non-optimal policies. This perspective not only deepens our understanding of light's propagation in complex refractive media but also inspires the design of quantum-inspired Reinforcement learning architectures capable of leveraging the intrinsic parallelism of quantum mechanics. Furthermore, the advent of quantum computing, with its inherent properties of superposition, entanglement, and quantum interference, provides a tangible pathway for implementing such algorithms in practice. To illustrate this paradigm, we propose a toy model wherein policies are encoded as quantum states, rewards are mapped to phase shifts, and measurement collapses the superposed state into an optimal policy. The implications of this framework extend beyond algorithmic innovation, offering insights into the possibility that nature itself operates as a quantum learning system, with physical laws emerging from a process akin to reinforcement learning.

Keywords

Path Integral Formulation, Reinforcement Learning, Quantum Computing, Quantum Interference, Superposition, Quantum Agent

1. Introduction

Optimization is a universal principle spanning domains as diverse as physics, biology, and artificial intelligence. In the physical world, nature appears to optimize paths, energies, and configurations, manifesting in phenomena such as the principle of least action in classical mechanics and Fermat's principle of least time in optics [1]. These principles suggest

that physical systems inherently seek the most efficient or probable outcomes, whether in the propagation of light through varying media or the motion of celestial bodies in gravitational fields. In artificial intelligence, reinforcement learning (RL) similarly empowers agents to optimize their behavior through trial and error, iteratively refining policies to

^{*}Corresponding author: bhushanpoojary@gmail.com (Bhushan Poojary)

Received: 21 July 2025; **Accepted:** 1 August 2025; **Published:** 9 September 2025



Copyright: © The Author(s), 2025. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

maximize expected cumulative rewards within dynamic environments [2].

Richard Feynman's path integral formulation of quantum mechanics offers a profound perspective on how physical systems evolve. Rather than following a single classical trajectory, a quantum particle explores all conceivable paths between two points, each path contributing a probability amplitude weighted by $e^{\frac{iS}{\hbar}}$, where S is the classical action [3]. The observed trajectory emerges from the complex interplay of constructive interference among near-stationary action paths and destructive cancellation of less probable ones, revealing nature's inherent capacity for parallel evaluation of possibilities.

This paper draws an analogy between Feynman's path integrals and reinforcement learning, proposing that light's behavior in complex media closely resembles a quantum agent simultaneously exploring an entire policy space. While classical RL systems rely on stepwise sequential evaluation and updates, quantum systems intrinsically achieve fundamental parallelism through superposition and entanglement. The advent of quantum computing, with its ability to hold and manipulate superpositions of multiple states, presents a compelling opportunity to translate this analogy into a viable computational paradigm.

We aim to (1) establish a conceptual mapping between quantum pathfinding and reinforcement learning, (2) explore implications for developing quantum-inspired RL algorithms, and (3) propose a toy model to illustrate these ideas concretely. The broader significance lies in suggesting that nature itself may be viewed as performing a form of reinforcement learning at the quantum level, with physical laws emerging as optimal policies in a cosmic learning process.

2. Background

2.1. Feynman's Path Integral Formulation

In classical mechanics, a system's trajectory is determined by minimizing the action $S = \int L dt$, where L is the Lagrangian function describing the difference between kinetic and potential energy. The principle of least action dictates that among all possible paths, the one for which S is stationary corresponds to the classical trajectory.

Feynman's path integral formulation generalizes this concept in quantum mechanics by postulating that a particle does not traverse a single trajectory but instead explores all conceivable paths $x(t)$ between initial and final states. The total probability amplitude for a particle to move from point (x_a, t_a) to (x_b, t_b) is given by:

$$\langle x_b, t_b | x_a, t_a \rangle = \int D[x(t)] e^{\frac{iS(x(t))}{\hbar}}$$

Here, $D[x(t)]$ denotes a functional integral over the infi-

nite-dimensional space of trajectories, and $e^{\frac{iS}{\hbar}}$ is a phase factor for each path determined by its classical action $S(x(t))$ [4].

This framework elegantly explains quintessential quantum phenomena such as interference and tunneling. In the famous double-slit experiment, for instance, the probability amplitude at a point on the screen arises from summing contributions of all possible photon paths, including those looping or reflecting multiple times, through both slits [5].

In optics, Fermat's principle emerges as a classical approximation of this process: paths of stationary action (least time) constructively interfere, while those far from stationary action cancel out due to rapid oscillations in phase. This provides an intuitive bridge between classical determinism and quantum probabilistic evolution, emphasizing the fundamental role of path summation in shaping observable reality.

2.2. Reinforcement Learning

Reinforcement learning (RL) is a subfield of machine learning where an agent interacts dynamically with an environment, which is often modeled as a Markov Decision Process (MDP). An MDP is defined by a set of states S , a set of possible actions A , a transition probability function $P(s'|s, a)$ describing the likelihood of moving from state s to state s' given action a , and a reward function $R(s, a)$ that quantifies the immediate benefit of taking action a in state s . The agent's objective is to learn a policy $\pi: S \rightarrow A$ that maximizes the expected cumulative discounted reward:

$$V(s) = E [\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) | s_0 = s]$$

where γ ($0 < \gamma \leq 1$) is the discount factor controlling the weight of future rewards.

Classical RL algorithms, such as Q-learning and SARSA, update value estimates iteratively to converge toward an optimal action-value function $Q(s, a)$

$$Q(s, a) \leftarrow Q(s, a) + \alpha [R(s, a) + \gamma \max_{a'} Q(s', a') - Q(s, a)]$$

where α is the learning rate controlling how quickly new experiences override old knowledge.

A key challenge in RL is balancing exploration (sampling unfamiliar actions to gather information) and exploitation (leveraging known rewarding actions) to achieve efficient and optimal learning outcomes [6, 7]. Advanced methods, such as deep Q-networks (DQN) and policy gradient approaches, extend these principles to high-dimensional and continuous action spaces.

2.3. Quantum Reinforcement Learning

Quantum reinforcement learning (QRL) seeks to harness the unique features of quantum mechanics—superposition, entanglement, and interference—to accelerate classical rein-

forcement learning tasks and explore exponentially large state-action spaces more efficiently [8]. In this paradigm, quantum systems can encode entire policy spaces in a superposed state, allowing parallel evaluation of multiple trajectories simultaneously. Variational quantum algorithms (VQAs), such as the Quantum Approximate Optimization Algorithm (QAOA) and the Variational Quantum Eigensolver (VQE), demonstrate promising approaches for tackling combinatorial optimization problems within quantum processors [9]. Additionally, quantum walks, as generalizations of classical random walks, have been proposed as powerful models for quantum agents navigating complex environments and solving search problems with potential polynomial or exponential speedups [10, 11].

Despite these theoretical advantages, most current QRL implementations are constrained by sequential update structures and hybrid classical-quantum processing. This limitation prevents them from fully exploiting quantum mechanics' intrinsic parallelism, leaving significant room for the development of novel algorithms that are natively quantum in nature.

3. Light as a Learning Agent

3.1. Mapping Path Integral to RL Concepts

The behavior of light in a complex medium can be reinterpreted through RL terminology (Table 1):

Table 1. Mapping between quantum mechanical phenomena and reinforcement learning concepts.

Quantum Mechanics	Reinforcement Learning
All paths explored (superposition)	Policy space explored simultaneously
Amplitudes $e^{\frac{iS}{\hbar}}$	Reward-weighted policy evaluation
Destructive interference	Penalization of suboptimal policies
Least action path observed	Optimal policy selected

This analogy reframes light's propagation as an optimization process where interference acts as a natural reinforcement mechanism.

3.2. Complex Media as Reward Landscapes

A complex medium with spatially varying refractive indices can be reinterpreted as a rugged "reward landscape" through the lens of reinforcement learning. In such a medium, regions of higher refractive index cause light to slow down,

introducing phase delays that are analogous to negative rewards or penalties in an RL framework. Conversely, regions with lower refractive indices allow light to traverse more quickly, resulting in phase advances similar to positive rewards. This dynamic interplay creates a highly non-trivial environment where multiple competing paths contribute to the total probability amplitude of reaching a destination.

As light propagates, every possible trajectory experiences unique cumulative phase shifts determined by the medium's local properties. Constructive interference between trajectories reinforces paths that align well with the medium's optimal "reward gradient," while destructive interference suppresses those that do not. The resulting observed light path thus emerges as the optimal trajectory selected by this natural reinforcement process—analogous to an agent discovering a globally optimal policy in a complex reward landscape filled with local optima. This perspective highlights how physical systems can perform optimization implicitly and in parallel.

4. Simultaneity vs Sequential Iteration

4.1. Simultaneity in Quantum Systems

Quantum superposition enables the simultaneous evaluation of all possible paths in a system's configuration space. This parallelism is exemplified in quantum search algorithms such as Grover's, which achieves quadratic speedup by encoding and exploring the entire search space in a superposed quantum state, leveraging interference to amplify correct solutions while suppressing incorrect ones [12].

4.2. Sequential Iteration in Classical RL

In contrast, classical RL relies on iterative policy updates where each policy or trajectory is evaluated step by step. This process scales poorly in environments with large or continuous state spaces due to the combinatorial explosion of possibilities. Even when parallelism is introduced via distributed systems or GPU acceleration, the resulting speedups are limited and cannot match the intrinsic simultaneity of quantum systems.

4.3. Implications for Algorithm Design

This contrast suggests that quantum-inspired RL algorithms could emulate interference-based filtering mechanisms, where complex probability amplitudes encode reward signals, and destructive interference naturally eliminates suboptimal or low-reward policies. Such designs would enable faster convergence toward globally optimal strategies by exploiting quantum parallelism.

5. Quantum Computing: Unlocking True Parallelism

Quantum computing architectures, including superconducting qubits (IBM, Google) and photonic processors (Xanadu), provide platforms for implementing superposition-based RL. Variational quantum circuits can encode policy states, with parameter optimization performed via quantum gradient descent [13, 14].

Photonic systems naturally simulate path integrals, offering experimental platforms for testing these concepts [15, 16].

6. Proposed Toy Model

We propose a model where:

- 1) Policies are represented as qubit superpositions $|\psi\rangle = \sum_i c_i |\pi_i\rangle$
- 2) Reward signals modulate phases $e^{iR(\pi_i)}$
- 3) Quantum interference filters trajectories, and measurement collapses the state to an optimal policy $|\pi^*\rangle$.

This model can be simulated on quantum devices using frameworks like Qiskit or Cirq [17].

7. Discussion

This paradigm suggests nature may optimize via quantum reinforcement. While practical implementation faces challenges (decoherence, error correction), hybrid classical-quantum architectures could bridge the gap [18].

Philosophically, viewing physical laws as emergent from a learning process challenges traditional notions of determinism [19, 20]. Future work should focus on formal mathematical mappings and experimental verification.

8. Conclusion

We propose a novel analogy between Feynman's path integral formulation and reinforcement learning. Light's behavior in complex media inspires a vision of quantum agents exploring policy spaces simultaneously, with quantum computing offering tools to realize this paradigm. This framework may revolutionize AI and provide insights into the optimization principles underlying physical laws.

Abbreviations

DQN	Deep Q-Network
MDP	Markov Decision Process
QAOA	Quantum Approximate Optimization Algorithm
QRL	Quantum Reinforcement Learning
RL	Reinforcement Learning
VQE	Variational Quantum Eigensolver
VQA	Variational Quantum Algorithm

Author Contributions

Bhushan Poojary is the sole author. The author read and approved the final manuscript.

Conflicts of Interest

The author declares no conflicts of interest.

References

- [1] Feynman, R. P. *Space-Time Approach to Non-Relativistic Quantum Mechanics*. Reviews of Modern Physics. 1948, 20(2), 367-387. <https://doi.org/10.1103/RevModPhys.20.367>
- [2] Sutton, R. S.; Barto, A. G. *Reinforcement Learning: An Introduction*, 2nd ed.; The MIT Press: Cambridge, MA, 2018. Available online: <http://incompleteideas.net/book/the-book-2nd.html>
- [3] Nielsen, M. A.; Chuang, I. L. *Quantum Computation and Quantum Information*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2010. <https://doi.org/10.1017/CBO9780511976667>
- [4] Dirac, P. A. M. *The Lagrangian in Quantum Mechanics*. Physikalische Zeitschrift der Sowjetunion. 1933, 3, 64-72. <https://doi.org/10.4236/ns.2014.614103>
- [5] Born, M.; Wolf, E. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7th ed.; Cambridge University Press: Cambridge, UK, 1999. <https://doi.org/10.1017/CBO9781139644181>
- [6] Watkins, C. J. C. H.; Dayan, P. Q-Learning. *Machine Learning*. 1992, 8(3), 279-292. <https://doi.org/10.1007/BF00992698>
- [7] Silver, D.; Huang, A.; Maddison, C. J.; Guez, A.; Sifre, L.; van den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; Dieleman, S.; Grewe, D.; Nham, J.; Kalchbrenner, N.; Sutskever, I.; Lillicrap, T.; Leach, M.; Kavukcuoglu, K.; Graepel, T.; Hassabis, D. Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature*. 2016, 529(7587), 484-489. <https://doi.org/10.1038/nature16961>
- [8] Dunjko, V.; Briegel, H. J. Machine Learning & Artificial Intelligence in the Quantum Domain: A Review of Recent Progress. *Reports on Progress in Physics*. 2018, 81(7), 074001. <https://doi.org/10.1088/1361-6633/aab406>
- [9] Farhi, E.; Goldstone, J.; Gutmann, S. A Quantum Approximate Optimization Algorithm. *arXiv: 1411.4028* (2014). <https://doi.org/10.48550/arXiv.1411.4028>
- [10] Paparo, G. D.; Dunjko, V.; Makmal, A.; Mart ín-Delgado, M. A.; Briegel, H. J. Quantum Speedup for Active Learning Agents. *Physical Review X*. 2014, 4, 031002. <https://doi.org/10.1103/PhysRevX.4.031002>
- [11] Schuld, M.; Petruccione, F. *Supervised Learning with Quantum Computers*. Springer: Cham, 2018. <https://doi.org/10.1007/978-3-319-96424-9>

- [12] Grover, L. K. A Fast Quantum Mechanical Algorithm for Database Search. STOC (1996), pp. 212-219.
<https://doi.org/10.1145/237814.237866>
- [13] McClean, J. R.; Boixo, S.; Smelyanskiy, V. N.; Babbush, R.; Neven, H. Barren Plateaus in Quantum Neural Network Training Landscapes. *Nature Communications*. 2018, 9, 4812.
<https://doi.org/10.1038/s41467-018-07090-4>
- [14] Preskill, J. Quantum Computing in the NISQ Era and Beyond. *Quantum*. 2018, 2, 79.
<https://doi.org/10.22331/q-2018-08-06-79>
- [15] Flamini et al. Photonic Quantum Information Processing: A Review. *Reports on Progress in Physics*, 2018, 82(1), 016001.
<https://doi.org/10.1088/1361-6633/aad5b2>
- [16] Carolan, J.; Harrold, C.; Sparrow, C.; Martin-Lopez, E.; Russell, N. J.; Silverstone, J. W.; Shadbolt, P. J.; Matsuda, N.; Oguma, M.; Itoh, M.; Matthews, J.; Hashimoto, T.; O'Brien, J. L. Universal linear optics in a reprogrammable photonic chip. *Science*. 2015, 349(6249), 711-716.
<https://doi.org/10.1126/science.aab3642>
- [17] Abraham, H.; Akhalwaya, I. Y.; Aleksandrowicz, G.; Alexander, T.; Barkoutsos, P.; Bielawski, K.; Bucher, D.; Capelluto, L.; Carballo, C.; Chen, A.; Córcoles, A.; Costales, E.; Cross, A.; Deutsch, J.; Dhand, I.; et al. Qiskit: An open-source framework for quantum computing. *Zenodo*. 2019,
<https://doi.org/10.5281/zenodo.2562110>
- [18] Cerezo, M.; Arrasmith, A.; Babbush, R.; Benjamin, S. C.; Endo, S.; Fujii, K.; McClean, J. R.; Mitarai, K.; Yuan, X.; Cincio, L.; Coles, P. J. Variational Quantum Algorithms. *Nature Reviews Physics*. 2021, 3(9), 625-644.
<https://doi.org/10.1038/s42254-021-00348-9>
- [19] Tegmark, M. *Our Mathematical Universe: My Quest for the Ultimate Nature of Reality*; Alfred A. Knopf: New York, 2014. Print, 432 pp., ISBN 978-0307599803.
- [20] Lloyd, S. Ultimate Physical Limits to Computation. *Nature*. 2000, 406(6799), 1047-1054.
<https://doi.org/10.1038/35023282>