Deep Reinforcement Learning for Optimal Pulse Shaping in Single-Photon Quantum Emitters

Syed Muhammad Abuzar Rizvi
Department of Electronics and Information Convergence Engineering
Kyung Hee University,
Yongin-si, Korea
smabuzarrizvi@khu.ac.kr

Jae Uk Roh
Department of Electronics and Information Convergence Engineering
Kyung Hee University,
Yongin-si, Korea
loadlumin@khu.ac.kr

Hyundong Shin
Department of Electronics and Information Convergence Engineering
Kyung Hee University,
Yongin-si, Korea
hshin@khu.ac.kr

***Abstract*—Single-photon sources are fundamental building blocks of quantum communication, quantum computing, and quantum sensing. Cascade quantum systems driven by short resonant pulses can, in principle, produce high-quality single photons in either emission channel. However, practical limitations such as finite lifetimes, spectral broadening, and re-excitation significantly reduce the single-photon probability. In this paper, we explore a deep reinforcement learning (DRL) framework to optimize the pulse width for maximizing the probability of single-photon emission in a selected channel while suppressing multi-photon events. We model a three-level cascade system under a time-dependent Hamiltonian with decay and define the DRL agent's action space as adjustments to the pulse width within physically realistic constraints. The reward function balances the single-photon probability against multi-photon probabilities. Numerical simulations demonstrate that the DRL-based controller discovers the optimal pulse width faster than conventional optimization methods. This work highlights the potential of DRL to enable robust and adaptive design of single-photon sources for scalable quantum networks.**

***Index Terms*—Cascade quantum systems, pulse shaping, quantum communication, quantum control, reinforcement learning, single-photon sources.**

I. Introduction

Single-photon sources are fundamental building blocks of quantum technologies, enabling a wide range of protocols in quantum communication, quantum computing, and quantum metrology. Their importance lies in the ability to generate deterministic, indistinguishable single photons that can be used for secure key distribution and photonic networking. Depending on the emission mechanism and control techniques, such sources can realize different classes of quantum states. The most direct outputs are Fock states [1], [2], particularly the single-photon state, which provides a fixed photon number with suppressed multiphoton contributions. Attenuated laser pulses, although not true single-photon states, generate coherent states [3]. More advanced engineered systems can produce squeezed states [4], supporting enhanced sensitivity in communication and sensing. Furthermore, tailored emission and sequential photon release can create multi-photon cluster and graph states, providing photonic resources for measurement-based quantum computing [5], [6].

Single-photon sources find wide-ranging applications across the spectrum of quantum technologies, making them indispensable for both near-term and long-term systems [7]–[9]. In quantum communication, they underpin quantum key distribution (QKD) protocols by enabling secure exchange of information through indistinguishable single photons, thereby mitigating vulnerabilities such as photon-number-splitting attacks [10], [11]. Beyond QKD, entangled photon pairs generated from cascade emitters or nonlinear crystals are critical for entanglement distribution and quantum teleportation, forming the foundation for the quantum internet and repeater-based long-distance networks [12]. In the domain of quantum metrology, engineered nonclassical states such as squeezed and multi-photon states enhance measurement sensitivity and precision [4]. Moreover, in photonic quantum computing, sequential emission of photons from quantum dots and trapped atoms enables the generation of photonic cluster and graph states, which act as universal resources for measurement-based computation [6]. These diverse applications highlight the versatility of single-photon sources as the enabling hardware for secure communication, distributed entanglement, precision sensing, and scalable quantum information processing.

Multi-level emitters, including biexciton–exciton cascade systems in quantum dots, are especially attractive because they can operate either as single-photon sources per monitored decay channel or as entangled photon-pair sources [13], [14]. In practice, however, achieving near-deterministic single-photon emission hinges on precise control of the excitation pulse. If the pulse duration is too long, re-excitation during decay leads to multi-photon contamination. If the pulse is too short, spectral broadening can excite unwanted transitions and degrade photon indistinguishability. Realistic devices further contend with finite lifetimes, dephasing, spectral diffusion, and instrumentation constraints that shift the optimal operating point away from simple analytic prescriptions. These considerations make pulse design a nontrivial optimization problem where competing effects must be balanced in a device- and setup-specific manner.

Quantum optimal control (QOC) methods have been widely investigated to address these challenges. QOC refers to the systematic design and application of external fields, such as

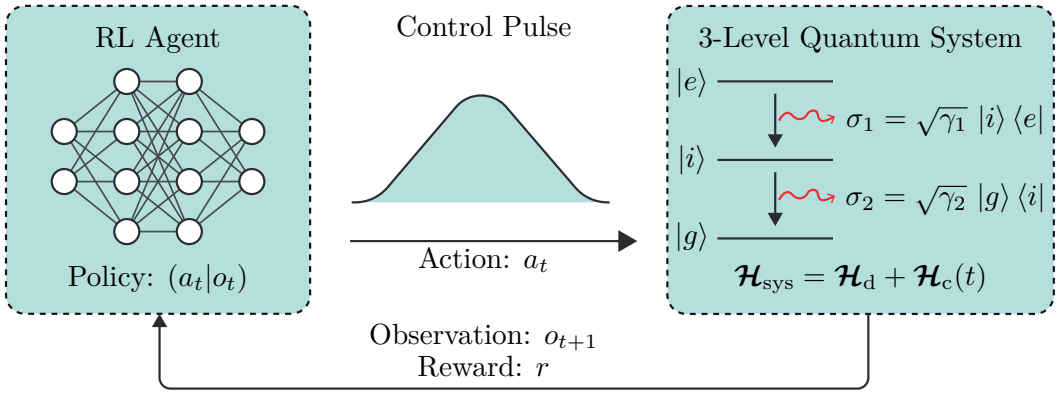

Fig. 1. DRL based control of a three-level quantum system. The agent generates a control pulse a_t to drive the system, which decays via $\sigma_1 = \sqrt{\gamma_1}\,|i\rangle\langle e|$ and $\sigma_2 = \sqrt{\gamma_2}\,|g\rangle\langle i|$. Observations and rewards are fed back to update the policy.

tailored laser pulses, microwave drives, or magnetic fields, to manipulate the dynamics of quantum systems in a precise and robust manner. The primary objective is to steer quantum states or operations toward a desired target despite the presence of decoherence, noise, and device imperfections, thereby enabling reliable quantum technologies. Gradient-based techniques such as gradient ascent pulse engineering (GRAPE) [15] and basis-expansion methods such as chopped random-basis (CRAB) [16] can deliver high-quality solutions when accurate models and gradients are available, but their performance may degrade under model mismatch, nonstationary noise, or limited diagnostic access. In contrast, DRL [17], [18] offers a model-free, data-driven approach that can optimize control policies directly from observed outcomes. By formulating pulse parameters as an episodic decision process, an reinforcement learning (RL) agent can explore pulse configurations, adapt to device idiosyncrasies and drift, and optimize a task-specific objective that directly reflects single-photon quality metrics. Crucially, the objective can prioritize single-photon probability while penalizing multi-photon events and incorporating auxiliary quality measures such as temporal-mode purity or source brightness.

This work investigates DRL for pulse-width optimization in three-level quantum system operated as single-photon sources. We build an RL environment that interfaces with an open-quantum-system to evaluate single-photon probability. The agent's action space adjusts pulse width within experimentally realistic bounds and the reward balances single-photon yield against multi-photon penalties. Furthermore, we compare the DRL based control strategy with results obtained from a conventional optimization method.

The remainder of this paper is organized as follows. Section II introduces the system model, while Section III formulates the DRL framework. Section IV presents the simulation setup along with baseline methods. Section V discusses the experimental results, and Section VI concludes the paper with potential directions for future research.

II. Three-Level Cascade Quantum System

A. System model

We consider a quantum system composed of three energy levels, which serves as a model for a biexciton cascade in a quantum dot [13]. The states are the ground state $|g\rangle$, an intermediate exciton state $|i\rangle$, and an excited biexciton state $|e\rangle$. In the computational basis, these states are represented as:

$$|g\rangle = \begin{bmatrix}1\\0\\0\end{bmatrix}, \quad |i\rangle = \begin{bmatrix}0\\1\\0\end{bmatrix}, \quad |e\rangle = \begin{bmatrix}0\\0\\1\end{bmatrix}. \tag{1}$$

The total Hamiltonian of the system, $\mathcal{H}_{\rm sys}$, is composed of a static component also called drift Hamiltonian $\mathcal{H}_{\rm d}$, and a time-dependent interaction component called the control Hamiltonian $\mathcal{H}_{\rm c}(t)$, describing the coupling to a classical electric field. This can be written as:

$$\mathcal{H}_{\rm sys} = \mathcal{H}_{\rm d} + \mathcal{H}_{\rm c}(t). \tag{2}$$

The drift Hamiltonian defines the energy of the intermediate state relative to the biexciton binding energy, E_b:

$$\mathcal{H}_{\rm d} = \frac{E_b}{2}\,|i\rangle\langle i|\,. \tag{3}$$

The control Hamiltonian models a two-photon resonant excitation from the ground state to the excited state, driven by a laser pulse with a time-dependent electric field $E(t)$:

$$\mathcal{H}_{\rm c}(t) = \Omega(t)(|g\rangle\langle e| + |e\rangle\langle g|). \tag{4}$$

Here, the effective two-photon Rabi frequency, $\Omega(t)$, is given by:

$$\Omega(t) = \frac{(\mu E(t))^2}{E_b}, \tag{5}$$

where μ is the dipole coupling strength.

The system's interaction with the environment leads to spontaneous decay, which is modeled by two collapse operators representing the two stages of the photon cascade:

$$\sigma_1 = \sqrt{\gamma_1}\,|i\rangle\langle e|\,, \qquad \text{decay from } |e\rangle \text{ to } |i\rangle, \tag{6}$$

$$\sigma_2 = \sqrt{\gamma_2}\,|g\rangle\langle i|\,, \qquad \text{decay from } |i\rangle \text{ to } |g\rangle, \tag{7}$$

where γ_1 and γ_2 are the respective decay rates.

The full dynamics of the system, including both coherent evolution and incoherent decay, are described by the Lindblad master equation for the density matrix $\rho(t)$:

$$\frac{d\rho(t)}{dt} = \mathcal{L}\rho = -i[\boldsymbol{\mathcal{H}}_{\mathrm{c}}(t), \rho(t)] + \sum_{k=1,2} \mathcal{D}[\sigma_k]\rho(t). \quad (8)$$

The first term, the commutator, describes the coherent evolution driven by the laser pulse. The second term describes the dissipation, with the Lindblad dissipator $\mathcal{D}$ for a collapse operator c defined as:

$$\mathcal{D}[c]\rho = c\rho c^\dagger - \frac{1}{2}\left(c^\dagger c\rho + \rho c^\dagger c\right). \quad (9)$$

B. N-Photon Emission Probabilities

The probability of emitting exactly N photons is obtained using the conditioned evolution formalism [19]. It can be obtained by alternating between conditioned evolution and photon emission events. Starting from $\rho(0) = |g\rangle\langle g|$, the system evolves under the no-jump operator $\mathcal{K}$ until an emission time, where the collapse operator $\mathcal{S}$ is applied. This sequence is repeated N times, after which the system undergoes a final no-jump evolution. Integrating the resulting joint probability density over all emission times yields $P(N)$. The N-photon probability is expressed as

$$P(N) = \int_0^\infty dt_1 \int_0^\infty dt_2 \cdots \int_0^\infty dt_N \, p(t_1, t_2, \ldots, t_N), \quad (10)$$

where $p(t_1, t_2, \ldots, t_N)$ denotes the joint probability density of photon emissions occurring only at times $\{t_1, t_2, \ldots, t_N\}$.

Due to the cascade structure of the system, the number of photons emitted into channels 1 and 2 is always equal. Without loss of generality, we condition on the emissions from channel 1 while tracing over channel 2. The corresponding collapse superoperator is defined as

$$\mathcal{S}\rho = \sigma_1 \rho \sigma_1^\dagger, \quad (11)$$

and the nonunitary (no-emission) evolution generator is given by

$$\mathcal{K} = \mathcal{L} - \mathcal{S}, \quad (12)$$

where $\mathcal{L}$ is the full Lindblad generator. The conditional probability density is then

$$p(t_1, t_2, \ldots, t_N) = \mathrm{Tr}\Big[\mathcal{K}(\infty, t_N)\mathcal{S}\mathcal{K}(t_N, t_{N-1}) \cdots \cdots \mathcal{S}\mathcal{K}(t_1, 0)\rho(0)\Big]. \quad (13)$$

Explicitly, the superoperator $\mathcal{K}$ takes the form

$$\mathcal{K}\rho = -i[H_I(t), \rho] + \mathcal{D}[\sqrt{\gamma_2}\sigma_2]\rho - \tfrac{1}{2}\left(\sigma_1^\dagger\sigma_1\rho + \rho\sigma_1^\dagger\sigma_1\right). \quad (14)$$

In practice, the recursive relation for evaluating $P(N)$ can be implemented numerically. At each recursion level, the system is first propagated with $\mathcal{K}$ up to a candidate emission time t_i, after which the collapse operator $\mathcal{S}$ is applied. The recursion continues until all N photon emissions are placed, at which point the final evolution with $\mathcal{K}$ yields the probability of ending in $|g\rangle$.

C. Problem Formulation for Pulse Optimization

The central problem is to determine the optimal laser pulse shape that maximizes the probability of emitting a single photon into a specific channel, while simultaneously suppressing the probability of emitting multiple photons. We model the laser pulse with a Gaussian envelope, which is characterized by its amplitude E_0, width (standard deviation) σ, and peak time t_0.

$$E(t) = E_0 \exp\left(-\frac{(t - t_0)^2}{2\sigma^2}\right) \quad (15)$$

The key parameters to be optimized are the pulse width σ and amplitude E_0. In our case, we will fix amplitude E_0 and focus on optimizing pulse width σ. We define an objective function that maximizes the single photon probability while suppressing both vacuum events and multiphoton contributions. Let $P_0(\sigma), P_1(\sigma), P_2(\sigma)$ denote the probabilities of emitting zero, one, and two photons, respectively. The objective is expressed as:

$$J(\sigma) = P_1(\sigma) - \lambda_0 P_0(\sigma) - \lambda_2 P_2(\sigma), \quad (16)$$

where $\lambda_2 \gg \lambda_0 > 0$ are penalty weights. This formulation ensures that the optimized pulse maximizes single photon generation efficiency while minimizing undesired outcomes.

III. DRL Model for Pulse Optimization

In this work, the optimization of the pulse width is formulated as a RL problem. The environment exposes a discrete action space of $N = 21$ candidate widths $\{w_k\}_{k=1}^N$ uniformly distributed over $[W_{\min}, W_{\max}] = [0.01, 0.20]$. At each episode, the agent samples an action $a \in \{1, \ldots, N\}$, corresponding to a width parameter $\sigma = w_a$. The environment evaluates this choice and returns a scalar reward

$$r = J(\sigma), \quad (17)$$

where $J(\sigma)$ denotes the objective function quantifying the system performance under the chosen pulse width.

The policy $\pi_\theta(a)$ is parameterized by a feed-forward neural network with two hidden layers of 32 ReLU-activated units each, followed by a softmax output over the 21 actions. Sampling from this categorical distribution ensures stochastic exploration. The training employs the REINFORCE algorithm, where the policy gradient update seeks to maximize $\mathbb{E}[J(\sigma)]$. A moving baseline b is used to reduce the variance of the gradient estimator, and the advantage is defined as $A = r - b$. The overall loss function is expressed as

$$L(\theta) = -A \, \log \pi_\theta(a) - \beta \, \mathcal{H}[\pi_\theta], \quad (18)$$

where $\mathcal{H}[\pi_\theta]$ denotes the entropy of the policy and $\beta = 10^{-3}$ is the entropy regularization weight. The baseline b is updated by exponential smoothing with momentum factor $\alpha = 0.9$. Policy parameters are optimized using Adam with learning rate 10^{-2}, over 40 training episodes.

Since the optimization is single-step, no temporal discounting is applied. The framework additionally records (i) the entropy $\mathcal{H}[\pi_\theta]$ to monitor exploration, (ii) the KL divergence $D_{\mathrm{KL}}(\pi_{\theta_t} \| \pi_{\theta_{t-1}})$ to quantify policy shift, and (iii) gradient norms to assess training stability.

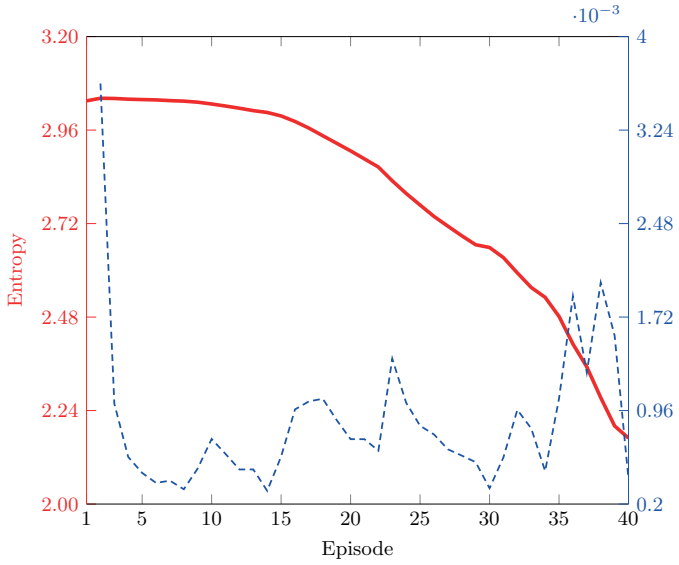

Fig. 2. Evolution of policy diagnostics during DRL, where entropy decreases steadily with episodes, while KL divergence remains bounded, indicating stable policy updates

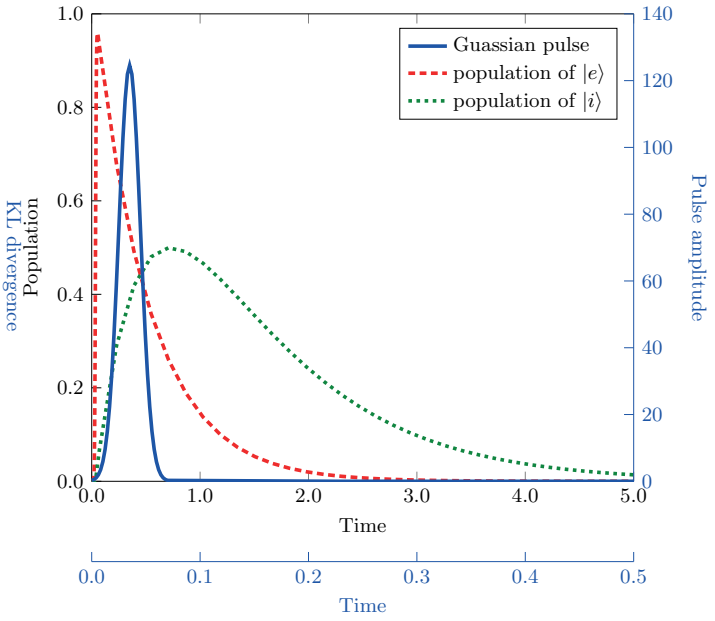

Fig. 3. System dynamics under optimized Gaussian excitation pulse to generate single photon source.

IV. Methods

A. Environment Parameters

In the considered system, the decay rates are simplified by setting $2\gamma_2 = \gamma_1 = \gamma$, with γ normalized to unity. The dipole coupling strength μ and the biexciton binding energy E_b are also taken as unity. The excitation pulse is normalized such that its area is π/ζ^2, where $\zeta = 1$ represents the combined contribution of the dipole coupling strength and the binding energy.

B. Benchmarking Configuration

For comparison, a population-based black-box optimizer, the cross-entropy method (CEM), is employed to maximize $J(\sigma)$. Unlike REINFORCE, CEM does not rely on gradients, instead it iteratively refines a sampling distribution over the continuous search space $[W_{\min}, W_{\max}]$. At each iteration, a population of $M = 12$ candidate widths is drawn from a Gaussian distribution $\mathcal{N}(\mu_d, \tau^2)$, where μ_d and τ denote the mean and standard deviation of the search distribution. Each candidate σ_i is then evaluated to obtain the reward $J(\sigma_i)$. The top-performing candidates, known as the "elites" (in this case, the top 25%), are selected. The mean and standard deviation of these elite samples are then used to update the search distribution for the next iteration.

The wall-clock time of both the methods are measured. This allows a direct comparison of the convergence behavior of CEM and DRL with respect to runtime.

V. Results

The numerical results highlight the convergence behavior of the proposed DRL framework, as observed in Fig. 2, the policy entropy decreases gradually with training episodes, demonstrating a consistent reduction in exploration, while the KL

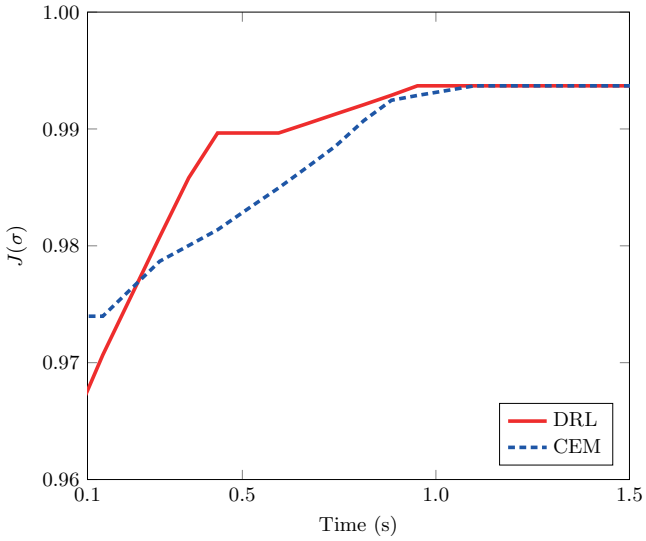

Fig. 4. Convergence time comparison between the DRL approach and the CEM method.

divergence remains bounded, confirming stable policy updates. Fig. 3 depicts the temporal evolution of the system driven by optimized Gaussian excitation pulse, where the applied field successfully transfers population between the intermediate and excited states, validating the pulse design. The comparative performance in Fig. 4 shows that both DRL and CEM achieve near-optimal objective values $J(\sigma)$. However, DRL converges more smoothly within the first few seconds of wall-clock time, whereas CEM exhibits some variability in later iterations. These results confirm that the DRL approach provides a robust and stable strategy for optimizing pulse widths in the

considered quantum system.

VI. Conclusion

In this paper, we presented a DRL framework for optimizing excitation pulse widths in three-level cascade quantum emitters to maximize single-photon generation probability while suppressing multi-photon contributions. By framing the optimization as a RL problem, the proposed method demonstrated stable convergence and faster adaptation compared to a conventional optimizer, confirming its robustness for model-free quantum control. The results highlight the promise of DRL in addressing challenges that arise from decoherence, re-excitation, and device-specific imperfections in realistic single-photon sources. In future work, the DRL paradigm can be extended beyond simple width optimization to model and explore more complex pulse shapes, multi-parameter controls, and adaptive protocols tailored to experimental constraints. Such advances will further enable the design of intelligent, high-performance single-photon sources, strengthening the foundations of scalable quantum communication, computation, and sensing networks.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) under RS-2025-00556064 and by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2025-2021-0-02046) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

References

[1] A. J. Shields, "Semiconductor quantum light sources," *Nat. Photonics*, vol. 1, no. 4, pp. 215–223, Apr. 2007.

[2] P. Michler, A. Kiraz, C. Becher, W. V. Schoenfeld, P. M. Petroff, L. Zhang, E. Hu, and A. Imamoglu, "A quantum dot single-photon turnstile device," *Science*, vol. 290, no. 5500, pp. 2282–2285, Dec. 2000.

[3] W.-Y. Hwang, "Quantum key distribution with high loss: Toward global secure communication," *Phys. Rev. Lett.*, vol. 91, no. 5, p. 057901, Aug. 2003.

[4] R. Schnabel, N. Mavalvala, D. E. McClelland, and P. K. Lam, "Quantum metrology for gravitational wave astronomy," *Nat. Commun.*, vol. 1, no. 121, pp. 1–10, Nov. 2010.

[5] N. H. Lindner and T. Rudolph, "Proposal for pulsed on-demand sources of photonic cluster state strings," *Phys. Rev. Lett.*, vol. 103, no. 11, p. 113602, Sep. 2009.

[6] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge, UK: Cambridge University Press, 2000.

[7] F. Grosshans and P. Grangier, "Continuous variable quantum cryptography using coherent states," *Phys. Rev. Lett.*, vol. 88, no. 5, p. 057902, Feb. 2002.

[8] N. J. Cerf, G. Leuchs, and E. S. Polzik, *Quantum Information with Continuous Variables of Atoms and Light*. London, U.K.: Imperial College Press, 2007.

[9] S. M. A. Rizvi, U. Khalid, S. Chatzinotas, T. Q. Duong, and H. Shin, "Controlled quantum semantic communication for industrial CPS networks," *IEEE Trans. Netw. Sci. Eng.*, Jul. 2025 (Early Access), Doi:10.1109/TNSE.2025.3589296.

[10] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, "Quantum cryptography," *Rev. Mod. Phys.*, vol. 74, no. 1, pp. 145–195, Mar. 2002.

[11] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev, "The security of practical quantum key distribution," *Rev. Mod. Phys.*, vol. 81, no. 3, pp. 1301–1350, Sep 2009.

[12] S. N. Paing, J. W. Setiawan, T. Q. Duong, D. Niyato, M. Z. Win, and H. Shin, "Quantum anonymous networking: A quantum leap in privacy," *IEEE Netw.*, vol. 38, no. 5, pp. 131–145, Sep. 2024.

[13] L. Hanschke, K. A. Fischer, S. Appel, and et al., "Quantum dot single-photon sources with ultra-low multi-photon probability," *npj Quantum Inform.*, vol. 4, p. 43, Sep. 2018.

[14] K. A. Fischer, R. Trivedi, and D. Lukin, "Particle emission from open quantum systems," *Phys. Rev. A*, vol. 98, p. 023853, Aug. 2018.

[15] P. de Fouquieres, S. G. Schirmer, S. J. Glaser, and I. Kuprov, "Second order gradient ascent pulse engineering," *J. Magn. Reson.*, vol. 212, no. 2, pp. 412–417, Aug. 2011.

[16] T. Caneva, T. Calarco, and S. Montangero, "Chopped random-basis quantum optimization," *Phys. Rev. A*, vol. 84, no. 2, p. 022326, Aug. 2011.

[17] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[18] X.-M. Zhang, Z. Wei, R. Asad, X.-C. Yang, and X. Wang, "When does reinforcement learning stand out in quantum control? A comparative study on state preparation," *npj Quantum Inform.*, vol. 5, no. 1, p. 85, Nov. 2019.

[19] H. Carmichael, *An Open Systems Approach to Quantum Optics*, 1st ed. Berlin, Heidelberg: Springer-Verlag, 1993.