Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.084050

ARTICLE

Security-Constrained Adaptive Control for Satellite QKD Systems under Feasibility-Aware Operation

Wibby Aldryani Astuti Praditasari[1,2,*], **Hyejin Yoon**[3], **Seunghwan Yun**[4], **Changuk Jang**[4] and **Okyeon Yi**[1]

[1]Department of Cyber Security, Kookmin University, Seoul, Republic of Korea
[2]Department of Electrical Engineering, Universitas Pertahanan Republik Indonesia, Sentul, Indonesia
[3]Department of Financial Information Security, Kookmin University, Seoul, Republic of Korea
[4]Industry-Academia Cooperation Foundation, Kookmin University, Seoul, Republic of Korea

*Corresponding Author: Wibby Aldryani Astuti Praditasari. Email: wibbypraditasari@kookmin.ac.kr or wibby.praditasari@gmail.com

Received: 15 April 2026; Accepted: 31 August 2026; Published: 28 September 2026

ABSTRACT: Satellite-based Quantum Key Distribution (QKD) provides a foundation for achieving information-theoretic security in long-distance communication systems. In practical satellite-to-ground scenarios, however, dynamic channel conditions significantly affect key validity, availability, and the stability of rekeying processes at higher communication layers. This work proposes a feasibility-aware adaptive control framework that enforces operation only within conditions where secure key generation remains valid, thereby preventing the use of cryptographically unreliable keys. The control mechanism is formulated as a constrained decision process and implemented using a Deep Q-Network (DQN), which dynamically adjusts basis selection in response to time-varying channel conditions. Unlike conventional approaches that focus solely on performance optimization, the proposed framework explicitly incorporates feasibility constraints into the control objective, ensuring that adaptive behavior remains aligned with the underlying security requirements of QKD. The framework is evaluated using a satellite-to-ground analytical channel model with parameters derived from representative experimental studies, capturing the expected behavior under representative satellite QKD conditions. Results show that, within the valid operating region, the proposed approach improves key availability, reduces key depletion events, and stabilizes rekeying behavior. At the system level, this enables the VPN layer to maintain AES-256 encryption for approximately 85% of the communication duration under typical conditions, with peak performance reaching up to 97% under favorable channel realizations. These values represent averaged performance across simulation runs, with higher values observed under optimal channel conditions. Overall, the results demonstrate that feasibility-aware adaptive control enhances both performance and security consistency by preventing operation in non-valid regimes while maintaining stable key availability under dynamic channel conditions. This work establishes a practical connection between quantum-layer constraints and system-level security outcomes, supporting more reliable deployment of quantum-enabled communication systems.

KEYWORDS: Satellite QKD security enforcement; feasibility-constrained reinforcement learning; cryptographic key availability; BB84 adaptive basis selection; QKD-VPN integration; denial-of-service prevention

1 Introduction

Satellite-based communication systems are increasingly required to support secure long-distance information exchange. However, the emergence of quantum computing threatens conventional cryptographic key-exchange mechanisms, motivating the development of quantum-secure communication technologies.

Copyright © 2026 The Authors. Published by Tech Science Press. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Quantum Key Distribution (QKD) provides information-theoretic security and has been successfully demonstrated in satellite-to-ground communication scenarios. Nevertheless, practical satellite QKD links operate under highly dynamic conditions influenced by atmospheric effects, pointing errors, and orbital motion, resulting in significant variations in key-generation performance and key availability.

Recent studies have explored reinforcement learning (RL) to optimize QKD operation under time-varying channel conditions. While these approaches improve performance, they primarily focus on maximizing secret-key generation without explicitly incorporating cryptographic feasibility into the control process. Consequently, adaptive decisions may not always remain aligned with the security requirements of QKD.

To address this limitation, this work proposes a feasibility-aware reinforcement learning framework for satellite-based QKD. The proposed Deep Q-Network (DQN) controller adaptively adjusts basis-selection probabilities while enforcing secure operating conditions during decision-making. The framework is evaluated using a satellite-to-ground free-space optical channel model incorporating dynamic satellite-pass behavior, atmospheric effects, burst-noise perturbations, and finite-key security constraints.

The main contributions of this work are summarized as follows:

1. A feasibility-aware reinforcement learning framework for adaptive basis selection in satellite QKD systems.
2. A security-constrained control mechanism that maintains key availability while preventing operation under invalid channel conditions.
3. A channel-aware evaluation environment incorporating simplified free-space optical propagation effects, dynamic satellite-pass geometry, and burst-noise disturbances, consistent with the finite-key operating assumptions adopted in this work (Section 4.1.2).
4. A finite-key security analysis examining the consistency of adaptive control with practical QKD security requirements.

This study investigates security-aware adaptive control for satellite QKD under time-varying channel conditions. Reinforcement learning is used to adapt the BB84 basis-selection probability based on the observed channel state within a penalty-augmented Markov Decision Process (MDP). Security violations are addressed through reward shaping during learning and an independent runtime feasibility gate that prevents non-feasible key material from being used for VPN rekeying. A hybrid MPC–RL approach is left for future work.

2 Related Work

Satellite-based Quantum Key Distribution (QKD) has demonstrated the feasibility of long-distance quantum communication through space-to-ground links, beginning with the first experimental demonstration of satellite-to-ground key distribution [1] and its subsequent extension to an integrated space-to-ground quantum communication network [2]. In practical deployments, however, satellite QKD channels are affected by atmospheric attenuation, turbulence, pointing errors, and orbital motion, resulting in time-varying channel conditions and limited satellite visibility windows, as characterized through finite-key statistical effects [3], polarization compensation techniques [4], and link-budget analyses for uplink transmission [5]. Building on these physical-layer characterizations, subsequent work has assessed practical satellite QKD architectures for current and near-future missions [6]. Conventional, non-adaptive BB84 implementations, exemplified by security treatments under realistic device assumptions [7] and decoy-state satellite-constellation designs [8], serve as representative baselines against which adaptive approaches, including the framework proposed in this work, are compared in Table 1.

To improve communication performance under dynamic channel conditions, recent studies have explored adaptive optimization approaches for QKD systems. Machine learning methods have been applied to continuous-variable QKD optimization [9], while reinforcement learning has been investigated for end-to-end key provisioning in time-varying quantum networks [10]. Related work has also examined resource-efficient quantum security mechanisms for practical communication systems, including quantum image encryption architectures aimed at reducing implementation complexity in resource-constrained IoT networks [11]. More broadly, artificial intelligence and machine learning techniques have been surveyed as enabling tools for quantum communication applications [12], and reinforcement learning methods have been reviewed for communications and networking tasks beyond the quantum domain [13].

Beyond the quantum layer, the integration of AI-driven optimization with next-generation network security has received growing attention. Overview studies have examined the role of AI in 6G communication systems [14], while satellite communication security has been surveyed with respect to prevailing threats and open research challenges [15]. Comprehensive surveys of 6G wireless networks have further catalogued the requirements, architectures, and open problems associated with next-generation connectivity [16,17]. Within this broader context, AI-driven secure network optimization combining blockchain and QKD has been proposed for 6G telecommunications [18], satellite-terrestrial integrated 6G architectures have been examined for ultra-dense LEO networking management [19], and the role of AI in safeguarding 6G-era networks has been reviewed [20]. These studies motivate the system-level relevance of AI-assisted security mechanisms, of which satellite QKD represents one instantiation.

A parallel line of work has examined the integration of QKD with post-quantum cryptography (PQC) as a hybrid approach to communication security. Recent work has examined lattice-based PQC for secure satellite communications, but mainly from the scheme-level security and performance perspective [21] and experimental integration of QKD with PQC in a hybrid quantum-safe cryptosystem has been demonstrated [22]. Post-quantum cryptography in networking protocols more generally has been reviewed with respect to challenges, solutions, and future directions [23], including the standardized lattice-based algorithms ML-KEM [24] and ML-DSA [25] that practical PQC deployment. Related studies have addressed securing Internet-of-things systems in a post-quantum environment [26], positioned quantum communications within future network services more broadly [27], proposed hybrid quantum-safe cryptographic schemes combining key exchange and digital signatures [28], and experimentally demonstrated authenticated QKD using PQC-based authentication [29]. These works establish PQC integration as a well-studied complementary mechanism, which the present study likewise adopts as a system-level fallback reference configuration rather than as a primary contribution.

Collectively, these studies have demonstrated the potential of adaptive optimization and intelligent decision-making for quantum communication systems. Nevertheless, the primary emphasis has generally been placed on improving communication performance, resource utilization, or implementation efficiency. Security is typically assessed against established composable security frameworks [30,31], but incorporated as an evaluation criterion or operational limit rather than as an explicit objective within the adaptive control process itself. As a result, the interaction between adaptive decision-making and cryptographic feasibility during system operation has received comparatively less attention.

Motivated by this observation, the present work investigates a feasibility-constrained adaptive control framework for satellite QKD. Rather than optimizing controller performance across the entire operating space, the proposed approach restricts adaptive decisions to operating conditions satisfying the predefined QKD feasibility criteria (QBER $< Q_{th}$). The framework is subsequently evaluated under finite-key assumptions, dynamic satellite-pass conditions, burst-noise perturbations, and VPN rekeying scenarios

to examine its impact on secret-key availability and communication continuity. Table 1 summarizes the methodological distinctions between representative studies and the proposed framework.

Table 1: Comparison of representative adaptive quantum communication frameworks and the proposed security-constrained satellite QKD framework.

Feature	Standard BB84 QKD [5,9]	DRL-Based Key Provisioning [27]	Quantum Image Encryption [31]	Proposed Framework
Reinforcement Learning-Based Adaptation	No	Yes	No	Yes
Dynamic Parameter Optimization	No	Yes	No	Yes
Explicit Feasibility Gate (QBER-Constrained Operation)	No	No	No	Yes
Finite-Key Security Consideration	Yes	No	No	Yes
Satellite-to-Ground QKD Scenario	No	No	No	Yes
Dynamic LEO Pass Evaluation	No	No	No	Yes
Burst-Noise/Adversarial Channel Analysis	No	No	No	Yes
Integration with VPN Rekeying Process	No	No	No	Yes
Security-Constrained Adaptive Control	No	No	No	Yes

Table 1 compares the proposed framework with representative adaptive QKD approaches. Conventional BB84 implementations apply QBER thresholds as security criteria during key generation, whereas existing adaptive methods primarily focus on performance optimization. In contrast, the proposed framework incorporates the predefined QKD feasibility condition into the adaptive control process and evaluates its performance under finite-key assumptions, dynamic satellite-pass conditions, burst-noise perturbations, and VPN rekeying scenarios. These characteristics distinguish the proposed framework from the representative studies considered in this work.

3 Architecture

3.1 System Framework and Integration Layer

The proposed architecture integrates four functional layers spanning quantum key generation, key management, adaptive control, and secure communication. At the quantum layer, a satellite-to-ground QKD link establishes secret keys through a free-space optical channel, while the classical channel supports basis reconciliation and authentication.

The key-management layer performs error correction and privacy amplification to generate verified session keys. An adaptive control layer monitors channel conditions and provides decision support for basis selection. Generated keys are subsequently supplied to the application layer for secure communication and periodic VPN rekeying. Optional post-quantum cryptographic mechanisms are included as a conceptual fallback for communication continuity.

A feasibility gate is positioned between the adaptive controller and the key-generation process, ensuring that control actions are executed only when channel conditions satisfy the security requirements of BB84 key generation.

The overall system architecture is illustrated in Fig. 1, showing the interaction between the quantum communication layer, adaptive control, and the higher-layer secure communication framework.

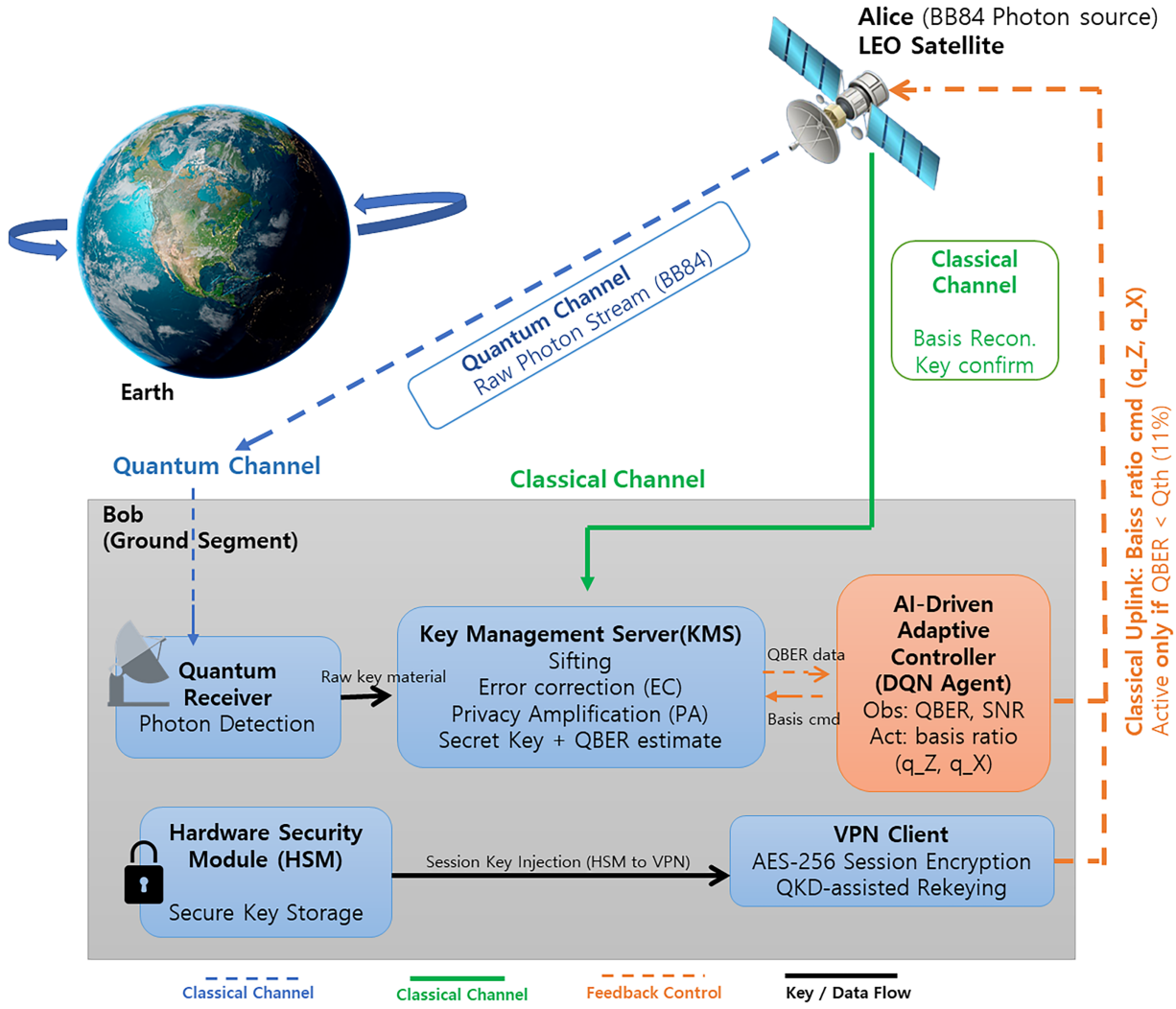


Figure 1: Layered security architecture for satellite-based QKD rekeying, showing the quantum communication layer, adaptive control, and higher-layer secure communication framework.

3.2 System Design Overview: Integrated Hybrid QKD-AI Framework

Fig. 2 illustrates the interaction among the Physical Quantum Layer, Key Management Layer, Cryptographic Integration Layer, and AI-Driven Control Layer. The satellite transmits BB84 quantum signals

through a free-space optical downlink, where photon detection and basis reconciliation are performed at the ground station.

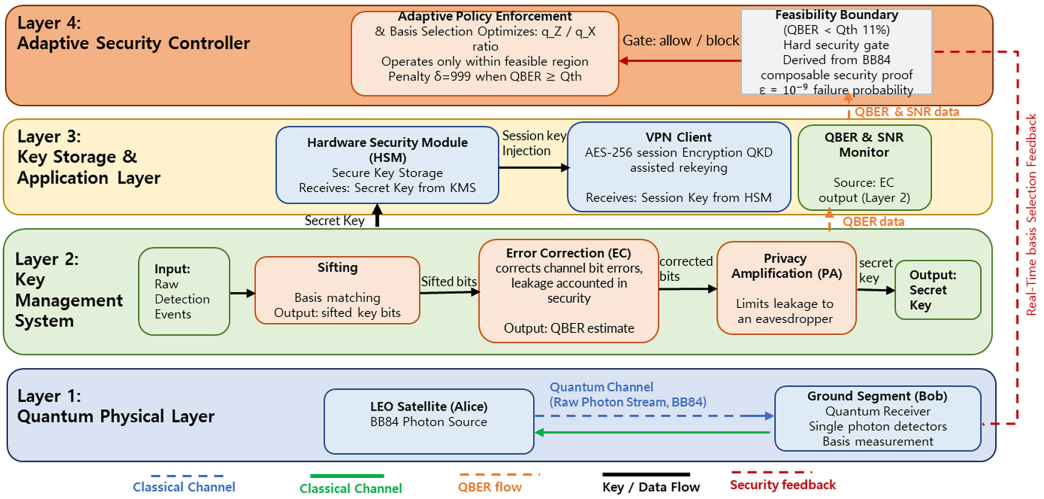


Figure 2: Integrated QKD-AI framework showing the interaction between quantum key generation, key management, adaptive control, and secure communication layers.

The resulting raw keys are processed through error correction and privacy amplification before being supplied to higher-layer cryptographic services. The adaptive controller continuously observes channel conditions and adjusts basis-selection probabilities to improve key-generation performance under dynamically varying operating conditions.

A post-quantum cryptographic (PQC) component is included as a system-level reference configuration for classical-channel protection and fallback operation. The hybrid QKD–PQC configuration is evaluated separately to provide a comparative reference for system-level performance, but it is not part of the proposed adaptive control method or its training process.

This design establishes a direct connection between quantum-layer operation, adaptive decision making, and secure communication services.

3.3 AI-Driven Optimization Component

The adaptive controller is implemented using a Deep Q-Network (DQN) that maps the observed channel state to a basis-selection action. The overall DQN architecture is illustrated in Fig. 3. The input state comprises the quantum bit error rate (QBER), signal-to-noise ratio (SNR), and satellite elevation angle, representing the observed channel and link conditions.

The network approximates the action-value function over a discrete basis-selection space. Through interaction with the environment, the agent learns a policy for adapting the basis-selection probability under time-varying channel conditions.

Security considerations are incorporated into the control objective by rewarding higher secret-key generation performance while penalizing actions associated with non-feasible operating states. This formulation allows the controller to consider both key-generation performance and the predefined feasibility condition without relying on manually specified adaptation rules.

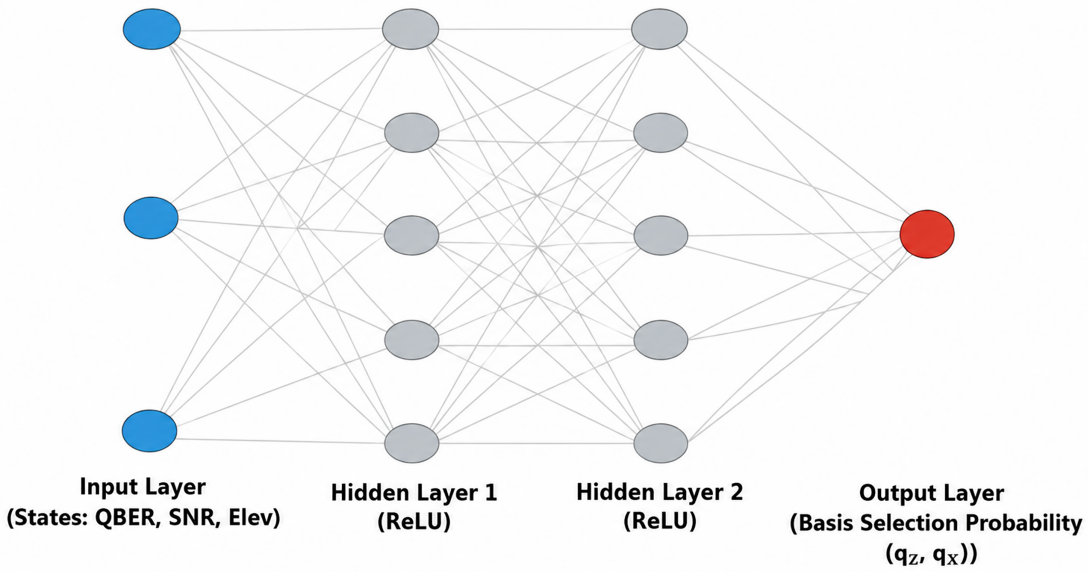


Figure 3: Deep Q-network (DQN) architecture for feasibility-constrained adaptive control, mapping channel state observations to basis selection probabilities (q_Z, q_X) under the predefined QKD feasibility constraint.

4 Methodology

Security feasibility is evaluated using the standard BB84 error threshold. At each decision step, the observed QBER determines whether key generation remains valid. This criterion is incorporated into the reward formulation and operational constraints used throughout the optimization process.

4.1 Finite-Key Security and Channel Modelling Framework

This section presents an analytical framework that connects the quantum communication layer with the adaptive control mechanism under practical satellite conditions. While asymptotic models provide a useful theoretical baseline for QKD performance analysis, they do not capture the constraints imposed by finite transmission durations in satellite links. To address this limitation, the present analysis adopts a finite-key formulation that explicitly accounts for the limited visibility window of LEO satellite passes, which typically span only a few minutes depending on orbital geometry. This formulation enables a more realistic characterization of secure key generation and system behavior under time-varying channel conditions.

4.1.1 Finite-Key Security Formulation

In the finite-key regime, the secret key length l extracted from a block of N pulses is a statistically bounded quantity that depends on error rates and leakage during post-processing. Following standard finite-key analysis used in QKD literature, the key length is expressed as:

$$l \approx \lfloor s_{X,0} + s_{X,1}[1 - h(\phi_X)] - \lambda_{EC} - 6\log_2\frac{21}{\epsilon_s} - \log_2\frac{2}{\epsilon_c} \rfloor \tag{1}$$

where:

$s_{X,0}$ and $s_{X,1}$ denote the estimated vacuum and single-photon contributions, obtained under finite-size statistical bounds,

$h(\phi_X)$ is the binary entropy function of the phase error rate ϕ_X.

The error-correction leakage is given by: $\lambda_{EC} = f_{EC}.nX \cdot h\,(QBER_X)$. Where nX is the number of sifted bits in the X basis; f_{EC} is the error-correction efficiency, and $h\,(.)$ is the binary entropy function. The parameters ϵ_s *and* ϵ_c denote the secrecy and correctness parameters, respectively, with total failure probability $\epsilon = \epsilon_s + \epsilon_c$.

This formulation is used to evaluate secure key generation under practical constraints, including short transmission durations and fluctuating channel conditions.

The QBER threshold Q_{th} = 11% is adopted as an operational reference based on the asymptotic BB84 security limit. Under finite-key conditions, the effective security threshold may be lower due to statistical fluctuations. Therefore, the 11% threshold is used as a simplifying approximation, while the corresponding finite-key security margin is not explicitly quantified in this study.

4.1.2 Simplified Analytical Channel Model for RL Training Environment

To provide a tractable analytical environment for reinforcement learning training, we adopt a simplified exponential attenuation model to represent the primary loss behavior of the free-space optical channel. The channel model incorporates simplified representations of beam divergence, atmospheric turbulence, and pointing loss through the corresponding transmissivity terms in Eq. (2). These effects are not modeled as higher-order time-varying stochastic processes, and orbital propagation uncertainties are not explicitly included. The resulting model provides a simplified representation of the channel conditions used to characterize QBER and feasibility during RL training.

The channel transmissivity is approximated as:

$$\eta = \left(\frac{D_R}{D_T + d\sqrt{\theta_{diff}^2} + \theta_{atm}^2} \right)^2 T_T T_R (1 - L_p) \eta_{det} \mathrm{t} \cdot 10^{-\mathrm{A_{atm}}/10} \tag{2}$$

where:

D_T, D_R: transmitter and receiver aperture diameters

d: slant range

θ_{diff}, θ_{atm}: diffraction and atmospheric divergence terms

$T_T T_R$: optical transmission efficiencies

L_p: pointing loss

$\mathrm{A_{atm}}$: atmospheric attenuation (dB)

η_{det}: detector efficiency

The received photon rate is then given by:

$$R_r(t) = R_p.\eta(\mathrm{t}) + R_{noise} \tag{3}$$

where R_p is the emitted photon rate, and R_{noise} includes background and dark count contribution. This model captures time varying channel conditions that directly affect QBER and key generation performance. The simplified model is intentionally adopted to isolate the interaction between control decisions and security constraints, rather than to replicate full physical propagation.

4.1.3 Security-Constrained Formulation

System adaptation is formulated as a Markov Decision Process (MDP), where the controller adjusts the basis-selection probability based on the observed system state. The reward primarily a higher secret key rate (SKR), while QBER and rekeying latency provide secondary terms.

$$R_t = \alpha.SKR_t - \beta.QBER_t - \omega\mathrm{L}.L_t - \delta.\max(0, QBER_t - Q_{th})^2 \quad (4)$$

where $\alpha = 1.0$, $\beta = 5.0$, $\omega\mathrm{L} = 0.5$ and $\delta = 50$. The linear coefficients are not normalized across the three terms; therefore, the SKR term is numerically dominant at typical operating values. The quadratic penalty is applied when QBER $\geq Q_{th}$.

The quadratic penalty is defined as:

$$Pt = \delta \max(0, QBER_t - Q_{th})^2 \quad (5)$$

where $\delta = 50$. The penalty increases when $QBER_t$ exceeds the operational threshold, discouraging actions associated with infeasible channel conditions during learning.

Note:

α = SKR weight;

β = QBER weight;

ωL = latency weight;

δ = penalty coefficient;

γ = MDP discount factor;

θ_t = satellite elevation angle;

q_Z = DQN action, probability of Z-basis.

4.2 Deep Reinforcement Learning Framework

The adaptive control layer is implemented using a Deep Q-Network (DQN). The action space consists of a finite set of discrete basis-selection probabilities, allowing the DQN to learn an approximate value function for the available actions without requiring policy-gradient optimization. The discrete action space also facilitates feasibility checking during action selection. Prior RL-based QKD studies using discrete action spaces have reported DQN-based optimization under noisy channel conditions.

In this study, the action space is limited to the basis-selection probability (q_Z), with $q_X = 1 - q_Z$. This allows the controller to adapt basis selection under time-varying channel conditions without modifying hardware-layer parameters. Expanding the action space to include decoy-state intensity, photon transmission rate, or detector threshold scheduling would require a higher-dimensional control space and corresponding extensions to the simulation model. Such extensions are left for future work.

4.2.1 MDP Basis Selection Optimization

The control problem is formulated as a Markov Decision Process (MDP) defined by the tuple $\langle S, A, P, R, \gamma \rangle$, where S is the state space, A is the action space, P is the transition probability, R is the reward function defined in Eq. (4), and $\gamma \in (0, 1)$ is the discount factor. The state at time t is defined as $s_t = (QBER_t, SNR_t, \theta_t)$, where $QBER_t$, SNR_t, and θ_t represent the observed quantum bit error rate, signal-to-noise ratio, and satellite elevation angle, respectively.

The Action space A consists of the 11 discrete q_Z values, ranging from 0.50 to 1.00 with a step size of $\Delta = 0.05$, where q_Z denotes the probability of selecting the Z-basis and $q_X = 1 - q_Z$.

Feasibility handling: During policy learning, a quadratic penalty is applied when $QBER_t \geq Q_{th}$, as defined in Eq. (5), with $\delta = 50$. The penalty discourages actions associated with threshold violations. During runtime, an independent feasibility gate checks the observed QBER before key utilization. If $QBER_t \geq Q_{th}$, the corresponding key material is excluded from the VPN rekeying process. Thus, only key material generated under $QBER_t < Q_{th}$ is eligible for use. The penalty and feasibility gate serve different roles: the penalty guides policy learning, while the gate enforces the feasibility condition during key utilization.

The optimal action-value function $Q^*(s, a)$ satisfies the Bellman equation:

$$Q^*(s, a) = R(s, a) + \gamma . \max_{a'} Q^*(s', a') \tag{6}$$

The DQN agent approximates $Q^*(s, a)$ using a neural network with parameters θ, trained via experience replay and a target network to improve training stability.

4.2.2 Action Space Limitation

In the present framework, the reinforcement learning agent adapts only the basis selection probabilities.

(q_Z, q_X), while other configurable QKD parameters remain fixed throughout the optimization process. This formulation enables the effects of feasibility-constrained adaptive basis selection to be evaluated independently within the adopted constrained Markov decision process. Accordingly, the current action space is limited to a single adaptive control variable. Extending the action space to jointly optimize additional QKD parameters, such as decoy-state intensity, transmission scheduling, or detector operating conditions, would require modifications to the reinforcement learning environment, reward formulation, and finite-key security analysis. Such extensions are therefore identified as directions for future work.

4.3 Security Model and Consistency Analysis

This section examines the security properties of the proposed framework from a security-consistency perspective rather than providing a formal composable security proof. Proposition 2 (Section 4.3.3) shows that the adaptive control policy does not modify the assumptions underlying the adopted BB84 finite-key security model. Instead, the basis-selection strategy operates within the conditions under which the finite-key analysis is applied. The results in this section should therefore be interpreted as a security-consistency analysis rather than a complete composable security proof under the Abstract Cryptography (AC) framework. A formal composable analysis of adaptive reinforcement learning for QKD is beyond the scope of this study. The threat model considers three representative scenarios:

(i) active quantum-channel attacks, including intercept-resend attacks, in which the adversary intercepts, measures, and resends quantum states, causing detectable disturbances such as increased QBER;
(ii) photon-number-splitting (PNS) attacks, mitigated using decoy-state techniques; and
(iii) channel degradation caused by either adversarial interference or environmental disturbances that increase the QBER beyond the predefined security threshold, resulting in a Cryptographic Denial-of-Service (C-DoS) condition. Under such conditions, the availability of usable secret keys may decrease, potentially requiring fallback to a classical PQC-based key establishment mechanism.

The two mechanisms provide separate safeguards at different stages of the framework. The quadratic penalty influences the learned policy during training, while the runtime feasibility gate determines whether

generated key material is eligible for subsequent VPN rekeying. This separation allows the security condition used in policy learning to be evaluated independently from the runtime validation of key material.

The security objectives considered in this work include the confidentiality of QKD-generated session keys, the integrity of the VPN rekeying process through authenticated key management, and forward secrecy under the adopted finite-key assumptions. These properties are evaluated using a target failure probability (ϵ) of 10^{-9}, an operational QBER threshold of approximately 11%, and finite-key block sizes that account for statistical fluctuations during short LEO satellite visibility windows.

Accordingly, the secret key rate is evaluated using Eq. (7) only within the predefined feasibility condition QBER < Q_{th}:

$$\mathrm{SKR} = \mathrm{R_{sift}} \times \left[1 - \mathrm{H_2(e)} - \mathrm{f}_{EC} \times \mathrm{H_2(QBER)}\right] \tag{7}$$

During policy learning, QBER values outside this region are penalized through the quadratic reward term defined in Eq. (5), while the runtime feasibility gate excludes corresponding key material from the VPN rekeying process.

From the adaptive control perspective, an increase in QBER due to channel degradation or adversarial activity increases the penalty associated with threshold violations. This encourages the controller to favor actions associated with lower-QBER operating conditions. The framework, therefore, applies the adaptive controller within the operating conditions assumed by the adopted BB84 finite-key model, without introducing a new QKD security principle or claiming a complete composable security proof. A rigorous composable security analysis of adaptive reinforcement learning for QKD remains beyond the scope of this study.

4.3.1 Threat Model

To evaluate the resilience of the proposed framework, we define a threat model that captures both quantum-level and system-level adversarial conditions. As summarized in Table 2, the analysis focuses on cryptographic denial-of-service (C-DoS) and physical-layer perturbations. These factors are incorporated into the reward function (Eq. (4)) to guide the controller toward operating within defined security constraints, thereby supporting key generation under both eavesdropping and service-disruption conditions.

Table 2: Threat model and mitigation strategies in satellite QKD systems.

Attack Category	Attack Type	Security Impact	Mitigation Strategy
Quantum Layer	Photon Number Splitting (PNS)	Exploits multi-photon pulses to obtain partial key information without significantly increasing QBER	Use of decoy-state with monitoring of photon intensity levels
Quantum Layer	Intercept-Resend	Introduces detectable disturbances, increasing QBER beyond acceptable thresholds	Threshold-based key discarding and adaptive control to avoid insecure key generation

(Continued)

Table 2 (continued)

Attack Category	Attack Type	Security Impact	Mitigation Strategy
Physical /Environmental	Cryptographic DoS (C-DoS)	Channel degradation or interference leading to key depletion in the key management system	Adaptive parameter control to maintain key availability under degraded channel conditions
Classical Channel	Man in the Middle (MITM)	Compromises classical communication during basis reconciliation and sifting	Authentication classical channel with quantum-safe (e.g., PQC-based) authentication mechanisms
Device Level	Side Channel Attacks	Leakage through detector imperfections or basis misalignment	Monitoring of system parameters (e.g., drift) and compensations through adaptive control mechanisms

Note: Overview of considered attack vectors and corresponding mitigation strategies across quantum, classical, and device layers in the proposed QKD system.

4.3.2 Security Enforcement Guarantee

Proposition 1 (Security Enforcement Guarantee): *Let π^* denote the DQN-learned policy under the reward function constrained MDP defined in Section 4.1.3, where the reward function incorporates the quadratic penalty term*

$$-\delta \cdot max\left(0, QBER_t - Q_{th}\right)^2, \delta = 50. \tag{8}$$

Then, in the policy-convergent regime ($t \to \infty$), the probability of the agent executing a basis selection action that produces cryptographically invalid key material:

$$P\left[key_{injected} = true \middle| QBER_t \geq Q_{th}, policy = \pi^*\right] \tag{9}$$

For any state-action pair (s, a) satisfying $QBER_t \geq Q_{th}$, the quadratic penalty is active and the reward becomes

$$r\left(s, a\right) = r_{base}\left(\mathrm{s}, \mathrm{a}\right) - \delta.\left(QBER_t - Q_{th}\right)^2. \tag{10}$$

Consequently, actions associated with threshold violations receive lower rewards than comparable actions within the feasible region. Under the Bellman optimality principle,

$$Q^*\left(\mathrm{s}, \mathrm{a}\right) < Q^*\left(\mathrm{s}, \mathrm{a}'\right), \qquad \mathrm{a}' \in A_{feasible} \tag{11}$$

As training progresses, the exploration rate ε decays to the minimum value $\varepsilon_{min} = 0.05$;

$$\varepsilon \to \varepsilon_{min} = 0.05 \tag{12}$$

Therefore, under the ε-greedy policy, the probability of selecting the greedy action $1 - \varepsilon_{min} = 0.95$.

Where:

Policy Layer

$$P\left(\pi(s) \in A_{infeasible}\right) \leq 0.05;$$

Execution Layer

$$P\left(key_{injected} \mid QBER_t \geq Q_{th}\right) = 0$$

4.3.3 Consistency with BB84 Finite-Key Security Assumptions (Security-Consistency Analysis)

Proposition 2: *Under the feasibility-constrained policy $\pi^{\star}$ (Proposition 1), adaptive basis-selection decisions remain consistent with the finite-key security assumptions adopted in this work. Consequently, within the feasible operating region ($QBER_t < Q_{th}$), the resulting session keys satisfy the secrecy and correctness conditions of the BB84 finite-key analysis with total failure probability*

$$\varepsilon_{total} = \varepsilon_S + \varepsilon_C < 10^{-9} \tag{13}$$

This guarantee holds conditional on:

(i) *the BB84 security proof, which establishes information-theoretic security within the feasible operating region;*

$$QBER_t < Q_{th} \rightarrow I(K_t : E) \leq \varepsilon_s \tag{14}$$

where $I(K_t : E)$ is the mutual information between the key K_t and Eve's quantum system E.

(ii) *the finite-key corrections, which account for statistical deviations in finite-block transmission.*

$$\varepsilon_s = \varepsilon_{Ec} + \varepsilon_{PA} + \varepsilon_{smooth}, \varepsilon_c = \varepsilon_{ver} \tag{15}$$

where
ε_{Ec}: error-correction failure probability, $\varepsilon_{Ec} = 8.9 \times 10^{-16}$;
ε_{PA}: privacy-amplification failure probability, $\varepsilon_{PA} = 7.9 \times 10^{-31}$;
ε_{smooth}: smoothing parameter (min-entropy estimation), $\varepsilon_{smooth} = 2^{-100}$;
ε_{ver}: verification (correctness) failure probability, $\varepsilon_{ver} = 5.4 \times 10^{-20}$
$\varepsilon_{total} < 10^{-9}$
Note: A full composable security proof in the abstract cryptography framework remains beyond the current scope and is identified as a future research direction.

Remark 1: *Proposition 2 should not be interpreted as a complete composable security proof of the adaptive reinforcement-learning framework. Rather, it establishes that the proposed feasibility-constrained control mechanism does not alter the security assumptions underlying the BB84 finite-key analysis. A full composable treatment of adaptive decision-making within the Abstract Cryptography framework remains an important direction for future work.*

4.4 Feasibility-Aware Optimization Principle

Unlike conventional AI-driven optimization approaches that attempt to maximize performance under all channel conditions, the proposed framework explicitly avoids enforcing adaptation in non-feasible regimes. In satellite-based QKD, channel conditions characterized by high noise and photon loss may lead to quantum bit error rates approaching QBER ≈ 0.5, where secure key generation becomes theoretically

impossible. In such scenarios, any optimization effort is fundamentally invalid from an information-theoretic perspective.

To address this, the proposed reinforcement learning model is designed to operate only within physically feasible regions, where QBER remains below the security threshold. When the channel enters a non-feasible regime, the agent does not attempt to recover performance artificially but instead converges to low-reward states, reflecting the true limitation of the communication channel. This behavior ensures that the optimization process remains consistent with the physical and security constraints of QKD systems, thereby improving the realism and practical applicability of the proposed approach.

4.5 Integrated QKD + VPN Simulation Environment

A comprehensive simulation framework was developed to evaluate the operational feasibility of the proposed AI-driven QKD system within a satellite-based VPN environment. The objective was to quantify the performance trade-offs between quantum key generation and classical VPN session maintenance under varying noise conditions and rekeying intervals within the simplified analytical channel environment.

The quantum layer models a decoy-state, satellite-to-ground free-space optical channel characterized by a transmission loss of 0.2 dB/km, detector efficiency ($\eta = 0.70$), dark count probability (10^{-5}), and background noise variance ($\sigma = 0.001$), consistent with the parameters specified in Table 3. The Key Management Server (KMS) performs error correction and privacy amplification to distill final session keys, which are integrated into the rekeying mechanism of a VPN-based secure communication framework. The reinforcement learning (RL) agent interfaces directly with the QKD module to dynamically adjust basis probabilities in real time, maintaining the secret key rate (SKR) at optimal levels while keeping QBER below the correction threshold.

Table 3: System and security parameters for LEO-to-ground QKD simulation.

Parameter	Symbol	Value	Unit	Security Relevance
Channel attenuation	α	0.20	dB/km	Affects channel loss and limits the achievable secure transmission distance under QKD constraints
Detector efficiency	η	0.70	–	Influences key generation rate and impacts error characteristics in detection
Background noise	σ	0.001	–	Contributes to QBER and affects the ability to maintain operation below the security threshold
Photon emission rate	f_s	10^8	pulses/s	Pulse rate determines the raw key generation capacity. At 100 MHz, consistent with Micius experimental setup [1], approximately 10^10 pulses are transmitted during a 100 s LEO satellite pass

(Continued)

Table 3 (continued)

Parameter	Symbol	Value	Unit	Security Relevance
Basis drift probability	p_{drift}	0.05	–	Represents misalignment effects that increase QBER and may trigger security control actions
Link Distance	L	500	km	Defines channel loss and available time window for secure key generation
Operational-Wavelength	λ	1550	nm	Affects channel transmissivity and stability of quantum state propagation
Failure Probability	ε	10^{-9}	–	The sum of secrecy and correctness errors establishing the finite-key security bound used in the security-consistency analysis of Section 4.3
PA Efficiency	f_{PA}	1.1–1.5	–	Reflects efficiency of privacy amplification in reducing potential information leakage
Finite-Key Block Size	N_{block}	10^5–10^8	bits	Determines statistical reliability of key generation under finite-size effects

The VPN subsystem emulates a generic VPN-based secure communication framework, where each rekeying cycle includes key validation, post-quantum authentication (via PQC fallback), and secure tunnel reestablishment. Latency overhead is measured as the total time elapsed.

Comparative scenarios were simulated for three configurations.

1. QKD-only, representing baseline quantum key integration;
2. AI-QKD, with adaptive reinforcement-based control; and
3. Hybrid QKD-PQC, where PQC keys replace QKD-generated keys when the observed QBER exceeds the predefined threshold.

The simulations consider a 100 Mbps secure link, rekeying intervals of 1–10 s, and distances up to 500 km, covering both feasible and non-feasible channel conditions. Adaptive basis selection is evaluated within the feasible operating region, while key material generated outside this region is excluded from VPN rekeying.

4.6 Dynamic Satellite Evaluation Framework

Phase 2 evaluates adaptive control performance within the cryptographically feasible operating region identified in Phase 1. This region is defined by channel conditions under which secure key generation remains possible according to the finite-key formulation and QBER security constraints. Under the adopted channel model, the feasibility boundary occurs at approximately 18 dB total channel loss, where QBER approaches the security threshold. This boundary serves solely as a normalized reference point for controller

evaluation and should not be interpreted as the physical communication range of a practical LEO satellite QKD system. The reported improvements in SKR, QBER, and rekeying latency therefore characterize controller behavior within security-valid operating conditions rather than performance at a specific satellite transmission distance.

Table 3 summarizes the physical and cryptographic parameters used in the feasibility-aware learning framework. It combines environmental factors, such as channel attenuation and background noise, with security constraints, including failure probability (ϵ) and finite-key effects, to define the operational conditions for secure key distillation in LEO-to-ground links.

The three-phase simulation framework used for the security evaluation is summarized in Table 4.

Table 4: Phase simulation framework for security evaluation in LEO-to-ground QKD.

Phase	Objective	Operational Range	Security Metrics	Security Control
Phase 1: Feasibility Characterization	Identify operational conditions where key generation remains within acceptable QBER limits (≈11%)	0–600 km	Raw SKR (bps), QBER	Baseline configuration without adaptive control
Phase 2: Controlled Operation	Evaluate adaptive control under feasible conditions (QBER < threshold) to maintain key availability	~90 km (equiv. ~18 dB link budget) (feasible region)	SKR (kbps), QBER, Rekeying Latency	Adaptive controller enabled for key management
Phase 3: Robustness Analysis	Assess system response to varying noise levels and its impact on key generation stability	100–500 km	SKR vs. Noise Variance, QBER	Adaptive control under varying channel conditions

Note: Phase 2 uses normalized throughput (kbps) to represent the rate of usable key material available to the key management system, providing a consistent metric for evaluating rekeying performance.

4.7 Experimental Design and Baseline Configurations

All configurations were tested under identical satellite-to-ground optical parameters, including attenuation (α = 0.20 dB/km), detector efficiency ($\eta = 0.70$), background noise ($\sigma = 0.001$), and basis drift probability ($p(\text{drift}) = 0.05$), consistent with the parameters specified in Table 3. The simulations used training runs of up to 5000 episodes. Convergence behavior was observed between episodes 350–400, while performance metrics were evaluated over the window spanning episodes 200–500. Results were summarized across independent simulation runs using 95% confidence intervals. The DQN hyperparameters used for training are summarized in Table 5.

Evaluation focused on three primary metrics: Secret-Key Rate (SKR): throughput of secure bits per second after key sifting and privacy amplification; Quantum Bit-Error Rate (QBER): ratio of incorrect detections to total received bits, indicating link fidelity; and VPN rekeying latency: delay between key generation and secure tunnel re-establishment.

Table 5: RL hyperparameters (DQN agent).

Hyperparameter	Symbol	Value	Description
Learning rate	A	0.001	Step size for parameter updates
Discount factor	Γ	0.95	Future reward weighting
Exploration rate	$\varepsilon_0 \rightarrow \varepsilon_{min}$	$1.0 \rightarrow 0.05$	ε-greedy exploration decay
Replay buffer size	N	10^5	Number of experiences stored
Batch size	B	64	Mini-batch size per training step
Target network update	T	0.01	Soft update parameter
Reward scaling	r_s	1.0	Normalization of reward magnitude
Training episodes	E_train	5000	Total training episodes (10 runs × 500)
Evaluation episodes	E_eval	300	Post-convergence evaluation window
Steps per episode	S	200	Quantum transmission cycles per episode

Compared with the baseline, the AI-driven configuration achieved a 28% increase in SKR, a 17% reduction in QBER, and a 12% decrease in VPN rekeying latency. These gains arise from the RL agent's capacity to continuously adjust basis probabilities in response to photon loss and channel fluctuations, minimizing retransmissions and optimizing quantum resource utilization. DQN inference time was measured at 0.8 ± 0.1 ms per control step on standard CPU hardware (single-threaded), representing 0.4% of the 200 ms control interval. Total computational overhead including state observation, action selection, and VPN interface signaling remained below 5%, confirming real-time feasibility.

Collectively, these results indicate improved AI-assisted basis selection not only enhances the efficiency and resilience of quantum key generation but also yields measurable improvements in end-to-end network performance. The hybrid QKD–PQC configuration further reinforces key exchange reliability under high-noise conditions, positioning this approach as a practical foundation for potential applicability to VPN-based secure communication systems. All reported metrics (SKR, QBER, rekeying latency) are averaged over the 300-episode evaluation window following training convergence, ensuring statistical representativeness of the results.

Two simulation scales are employed in this study, each appropriate to its experimental purpose.

(i) N = 20 independent runs are used for the primary performance comparison and the ablation study in Section 5.7. For these experiments, 95% confidence intervals are computed using Student's t-distribution with df = 19.

(ii) N = 100 independent runs are used for the burst-noise evaluation, feasibility analysis, and system-level security evaluation.

Four QKD control configurations are evaluated to isolate the contribution of adaptive basis selection, while the hybrid QKD–PQC configuration is included separately as a system-level reference.

Six configurations were experimentally evaluated. MPC is discussed separately as future work.

1. Static BB84 baseline with fixed basis probabilities ($q_Z = q_X = 0.5$). Where q_Z and q_X denote the probabilities of selecting the Z-basis and X-basis, respectively.
2. Fixed Asymmetric baseline with $q_Z = 0.75$ and $q_X = 0.25$, to assess the effect of a fixed basis asymmetry without adaptive control.
3. Greedy adaptive Baseline

The rule-based adaptive controller adjusts the basis probability according to the observed QBER:

$$q_Z(t) = \begin{cases} 0.90, & QBER_t < 0.05 \\ 0.75, & 0.05 \le QBER_t < 0.08 \\ 0.60, & 0.08 \le QBER_t < Q_t \end{cases} \tag{16}$$

With $q_X(t) = 1 - q_Z(t)$. Key Generation is suspended whenever $QBER_t \ge Q_{th}$

4. Gaussian Process Bayesian Optimization Baseline

As a non-RL adaptive baseline, a Gaussian Process Bayesian Optimization (BO) controller is included to assess the contribution of sequential, state-conditioned reinforcement learning from sample-efficient static optimization. The BO agent maintains a GP surrogate model over the scalar action domain $q_Z \in \{0.50, 0.55, \ldots, 1.00\}$ and selects actions according to the UCB acquisition function:

$$a_t = \arg\max_{q_Z} [\mu(q_Z) + \beta \cdot \sigma(q_Z)] \tag{17}$$

where $\mu(q_Z)$ and $\sigma(q_Z)$ denote the posterior mean and standard deviation of the GP evaluated at q_Z, and $\beta = 2$ controls the exploration–exploitation trade-off. The GP employs a squared-exponential (RBF) kernel with length scale $\ell = 0.15$ and observation noise variance $\sigma_n^2 = 0.01$, calibrated to the discrete action space $\Delta = 0.05$.

5. AI-Driven QKD (Proposed)

The DQN controller determines the basis-selection action $a_t = \pi_\theta(s_t)$, where s_t represents the observed channel state, $s_t = [QBER_t, SNR_t, \theta_t]$ and π_θ denotes the learned DQN policy parameterized by θ. The selected action determines the adaptive basis probability $q_Z(t) = f(a_t)$, $q_X(t) = 1 - q_Z(t)$.

6. Hybrid QKD–PQC configuration

The hybrid configuration combines the AI-Driven controller with PQC-based key fallback. The session key generation process is represented as:

$$K_t = \begin{cases} K_{QKD}, & QBER_t < Q_{th} \\ K_{PQC}, & QBER_t \ge Q_{th} \end{cases} \tag{18}$$

where the corresponding terms denote keys generated through the QKD process and keys established through the PQC fallback mechanism.

MPC as a Future Extension

MPC (Model Predictive Control) is considered as a possible extension of the proposed framework and is not included as an experimental configuration in this study. The orbital motion of Low Earth Orbit (LEO) satellites follows a predictable trajectory that can be described using satellite ephemeris. As illustrated in Fig. 4, the corresponding slant range and elevation angle vary throughout a satellite pass and can be derived from the orbital geometry. Combined with the physical channel models described in the preceding sections, these quantities could be used to estimate channel transmissivity, received photon rate, QBER, and SKR over a prediction horizon. Such information could support future MPC-based control of the basis-selection parameters.

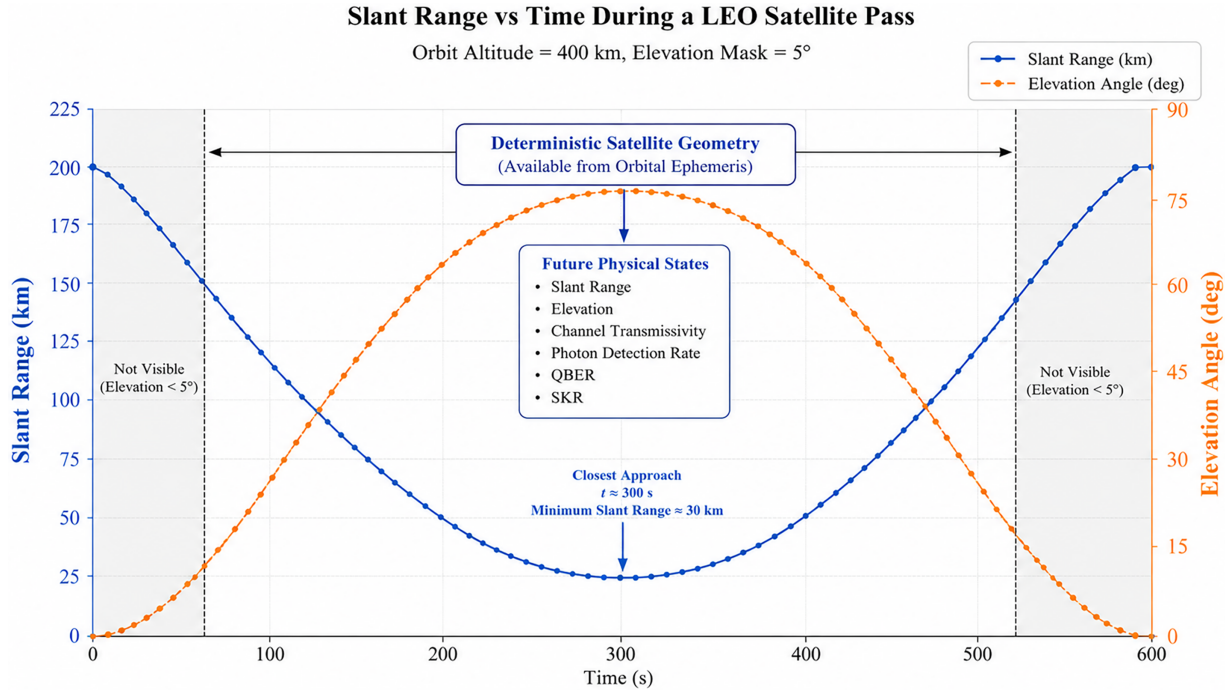


Figure 4: Deterministic LEO satellite communication geometry during a single satellite pass. The resulting slant range and elevation angle provide the physical states used to derive channel transmissivity, received photon rate, QBER, and SKR in the proposed framework, and may support predictive channel estimation in future MPC-based extensions.

The present study, however, is limited to reactive feasibility-aware adaptive control. Accordingly, the simulation environment represents only the instantaneous communication state available at each decision epoch, and the Deep Q-Network (DQN) determines the BB84 basis ratio using only the current system observations. No explicit prediction of future channel conditions is incorporated into the decision-making process. This design is consistent with the reinforcement learning formulation adopted throughout the study and the corresponding performance evaluation.

Although the underlying orbital geometry is deterministic, the current simulation framework does not include an integrated prediction module that estimates future channel states over a finite prediction horizon. Consequently, future values of the slant range, channel transmissivity, received photon rate, and the resulting QBER are not propagated within the control loop for optimization. Since these predicted channel states form the information required by an MPC-based controller, predictive control is not considered within the scope of the present work.

Accordingly, Fig. 4 should be interpreted as an illustration of the deterministic satellite communication geometry and its associated physical channel evolution rather than as an implementation of predictive control. Incorporating ephemeris-based channel prediction together with finite-horizon optimization into the existing framework may provide a useful direction for future investigation, as discussed in Section 7.2.

5 Result

The simulation results are presented in two distinct phases to ensure interpretive clarity.

5.1 Phase 1: Channel Feasibility Characterization

Fig. 5 identifies the practical operating range of the satellite QKD link under distance-dependent channel attenuation. As propagation distance increases, channel transmissivity decreases, resulting in lower SKR and higher error probability. The proposed adaptive framework consistently maintains higher SKR than the static baseline throughout the valid operating range, demonstrating the benefit of adaptive basis selection under the simplified channel conditions adopted in this study (Section 4.1.2).

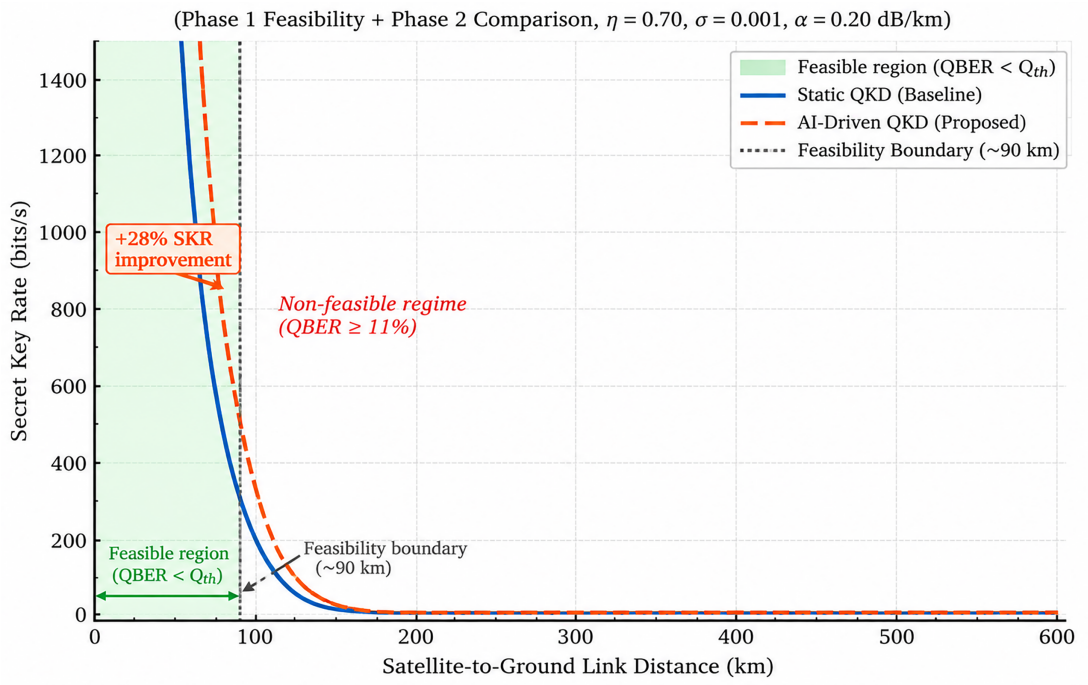


Figure 5: Secret key rate (SKR) as a function of satellite-to-ground link distance (0–600 km), comparing the static QKD baseline and the proposed adaptive framework.

Fig. 6 evaluates the proposed controller under a continuous LEO satellite pass. As elevation increases, the slant range decreases, improving channel quality and enabling higher secret-key generation rates. During the descending phase, increasing propagation loss gradually reduces SKR. The results demonstrate that adaptive basis selection remains effective under continuously varying satellite geometry and channel conditions.

Table 6 summarizes the combined effects of propagation distance and atmospheric turbulence on channel transmissivity, QBER, and SKR. Increasing turbulence severity significantly reduces channel transmissivity and accelerates SKR degradation. Under strong turbulence conditions, the channel becomes unable to support secure key generation at longer distances, resulting in complete SKR collapse. These results highlight the importance of considering turbulence-induced channel degradation when designing adaptive satellite QKD systems.

Table 7 presents the absolute performance comparison of the baseline configurations and the proposed AI-QKD framework at a representative feasible operating point. The static BB84 baseline achieves an SKR of 97.3 kbps, compared with 125.6 kbps for the proposed AI-QKD configuration, corresponding to a 29.1% increase relative to the static baseline. The proposed configuration also yields lower QBER (0.070 vs. 0.085) and lower rekeying latency (1.10 vs. 1.25 s). Compared with the BO baseline, AI-QKD achieves an additional

9.9 kbps in SKR, corresponding to an 8.6% relative improvement. Relative to the static baseline, the gain for AI-QKD is 10.2 percentage points higher than that of BO (29.1% vs. 18.9%). These comparisons are based on the predefined feasible operating region.

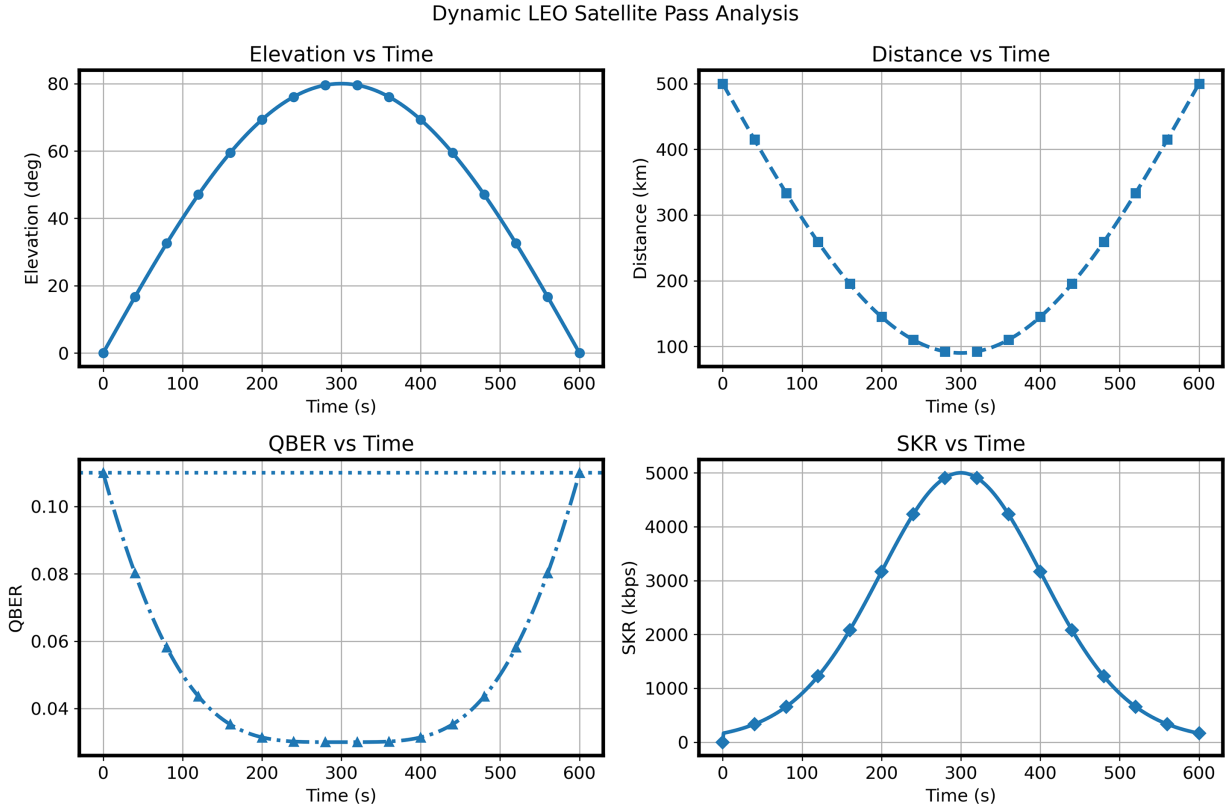


Figure 6: Dynamic LEO satellite pass simulation showing the temporal evolution of elevation angle, slant-range distance, QBER, and secret-key rate during a representative satellite-to-ground communication window. The feasibility-aware controller maintains operation within the secure region (QBER < Q_{th}) while adapting to continuously varying channel conditions.

Table 6: Impact of distance and atmosphere turbulence regimes on channel transmissivity, QBER, and secret key generation performance.

Distance (km)	Turbulence Regime	Rytov Variance (σ^2R)	Total Transmissivity (η total)	QBER	SKR (kbps)	Feasible
30	No Turbulence	0.000E+00	2.1282E−01	0.0330	9906.09	Yes
30	Weak	1.017E−01	2.0721E−01	0.0330	9644.90	Yes
30	Moderate	1.017E+01	8.5578E−02	0.0331	3978.17	Yes
30	Strong	1.017E+03	3.0552E−03	0.0362	133.83	Yes
50	No Turbulence	0.000E+00	1.6466E−01	0.0331	7662.55	Yes
50	Weak	2.593E−01	1.4991E−01	0.0331	6975.08	Yes
50	Moderate	2.593E+01	2.6515E−02	0.0334	1226.64	Yes
50	Strong	2.593E+03	4.8758E−04	0.0524	14.89	Yes

(Continued)

Table 6 (continued)

Distance (km)	Turbulence Regime	Rytov Variance (σ^2R)	Total Transmissivity (η total)	QBER	SKR (kbps)	Feasible
75	No Turbulence	0.000E+00	1.2440E−01	0.0331	5786.99	Yes
75	Weak	5.453E−01	9.8641E−02	0.0331	4586.75	Yes
75	Moderate	5.453E+01	7.8834E−03	0.0342	358.67	Yes
75	Strong	5.453E+03	1.0743E−04	0.1153	0.00	No
91	No Turbulence	0.000E+00	1.0593E−01	0.0331	4926.19	Yes
91	Weak	7.774E−01	7.5152E−02	0.0331	3492.45	Yes
91	Moderate	7.774E+01	4.1544E−03	0.0353	185.00	Yes
91	Strong	7.774E+03	5.1766E−05	0.1896	0.00	No

Table 7: Absolute performance comparison of baseline and AI-driven QKD at a representative feasible operating point.

Metric	Static BB84	Fixed Asym.	Greedy Adaptive	BO (GP-UCB)	AI-QKD (Proposed)
SKR (kbps), mean ± 1σ	97.3 ± 3.8	109.3 ± 3.3	112.5 ± 3.1	115.7 ± 2.3	125.6 ± 2.6
95% CI (SKR)	[95.5, 99.1]	[107.7, 110.9]	[111.0, 114.0]	[114.6, 116.8]	[124.4, 126.8]
QBER, mean ± 1σ	0.085 ± 0.003	0.079 ± 0.003	0.076 ± 0.002	0.075 ± 0.002	0.070 ± 0.001
Rekeying latency (s), mean ± 1σ	1.25 ± 0.08	1.20 ± 0.07	1.19 ± 0.07	1.13 ± 0.07	1.10 ± 0.06
Gain vs. Static (SKR)	Reference	+12.3%	+15.6%	+18.9%	+29.1%
Invalid key injections	320	285	0	0	0

Note: Values are mean ± 1σ across N = 20 independent simulation runs. The reported 95% confidence intervals for SKR are computed using Student's t-distribution with df = 19. QBER is dimensionless. The BO baseline does not incorporate the feasibility constraint into action selection; key injection is prevented by the runtime feasibility gate. The DQN controller incorporates the feasibility condition into policy learning and is additionally subject to the runtime feasibility gate.

The greedy adaptive baseline improves SKR relative to the static BB84 configuration by exploiting asymmetric basis selection under favorable channel conditions. However, the proposed DQN controller achieves an additional improvement of approximately 11% over the greedy baseline. This result indicates that the observed performance gain cannot be attributed solely to basis asymmetry and instead arises from dynamic policy adaptation to the evolving channel state.

QBER increases with distance and converges to 0.5 beyond the feasibility boundary, indicating a noise-dominated regime. The security threshold (≈11%) is reached at approximately 91 km, consistent with the feasibility limit observed in Fig. 7. Within the feasible region, the proposed method reduces QBER by approximately 17% through adaptive basis selection. No improvement is observed beyond the threshold, as expected from physical constraints.

At L = 500 km, the channel remains in the non-feasible regime, with QBER fluctuating around 0.5 and SKR remaining negligible throughout the observation window (Fig. 8). The feasibility constraint therefore disables adaptive control and prevents key material generated under this condition from being used for subsequent rekeying.

Fig. 9 presents the training dynamics of the DQN-based adaptive controller within the evaluated feasible operating region. During the initial exploration phase (episodes 1–200), the agent occasionally selects actions associated with QBER threshold violations, resulting in the quadratic reward penalty defined in Eq. (5) (δ = 50). After approximately episode 200, the reward trajectory becomes more stable and approaches a lower-penalty regime. The lower panel shows the corresponding SKR evolution during training. The approximately 28% improvement from about 98.0 to 125.7 kbps corresponds to the N = 100-run system-level evaluation reported in Section 6.2, whereas the primary performance comparison based on N = 20 independent runs is reported separately in Table 7. The observed training behavior indicates policy stabilization within the evaluated feasible operating region.

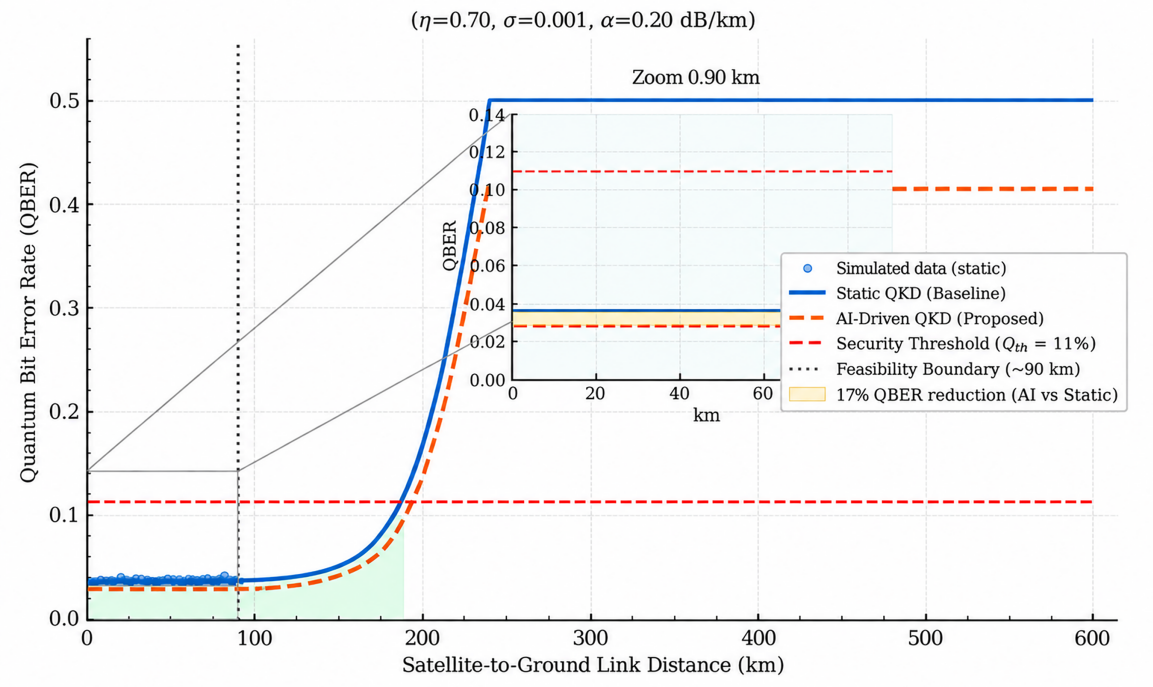


Figure 7: Quantum bit error rate (QBER) as a function of satellite-to-ground link distance (0–600 km) for the static baseline and the proposed AI-driven QKD framework (η = 0.70, σ = 0.001, α = 0.20 dB/km). Scatter points indicate raw simulation samples; solid and dashed lines represent the respective trend curves. The horizontal dashed red line marks the security threshold Q_{th} = 11%, derived from the asymptotic BB84 security bound adopted as a conservative operating threshold under the finite-key formulation used in this work (Section 4.1.1). The orange-shaded region quantifies the absolute QBER reduction attributable to the adaptive basis selection mechanism within the feasible operating region. The inset provides a magnified view of the 0–90 km feasibility window, wherein the 17% QBER reduction achieved by the AI-driven approach is visually discernible and statistically significant.

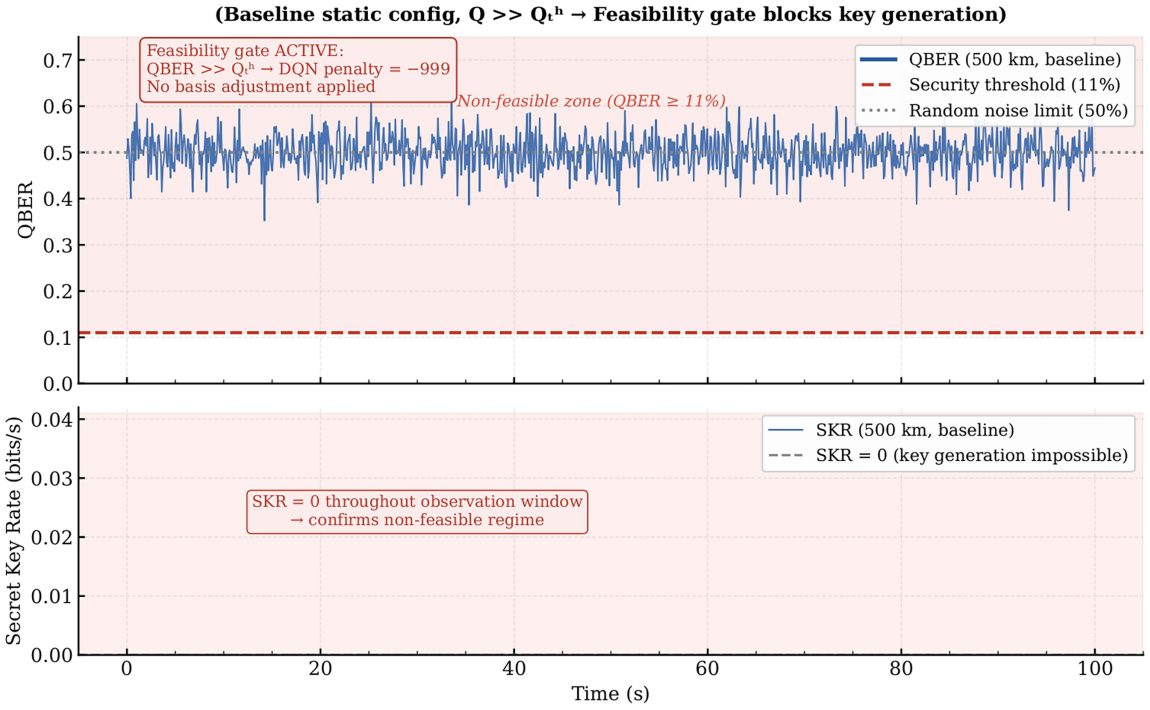


Figure 8: Temporal evolution of (upper panel) the quantum bit error rate (QBER) and (lower panel) the secret key rate (SKR, bits/s) at a fixed satellite-to-ground link distance of 500 km under the static QKD baseline configuration (η = 0.70, σ = 0.001, α = 0.20 dB/km, observation window: 100 s). At this distance, the channel operates entirely within the non-feasible regime (QBER >> Q_{th} = 11%). The upper panel illustrates the stochastic QBER fluctuations around the random noise limit (QBER ≈ 0.50), while the lower panel confirms that the resulting SKR is effectively indistinguishable from zero throughout the observation window. The feasibility gate, a hard non-parametric execution-layer block (Section 4.2.1), is active at all time steps, excluding all key material generated in this regime from the VPN rekeying process. The reward-shaping penalty (δ = 50, Eq. (5)) discourages the policy from favoring this regime during training, but the exclusion shown here is enforced independently of that penalty value.

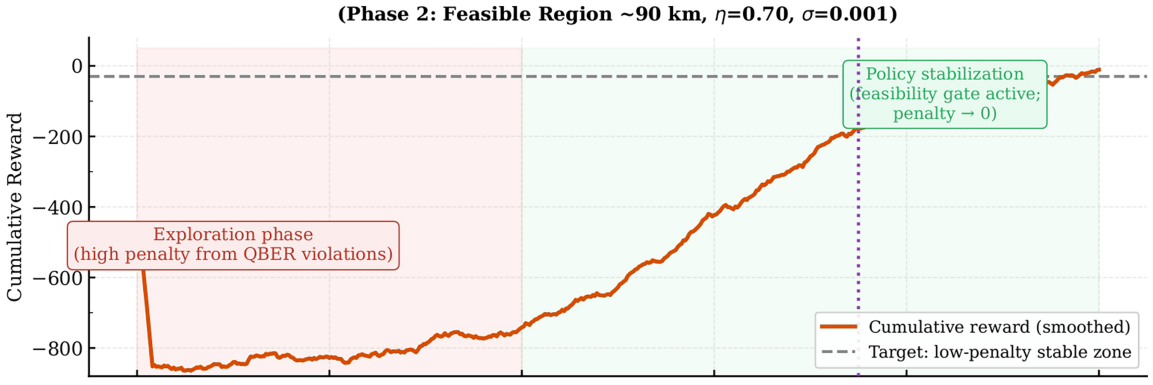


Figure 9: (Continued)

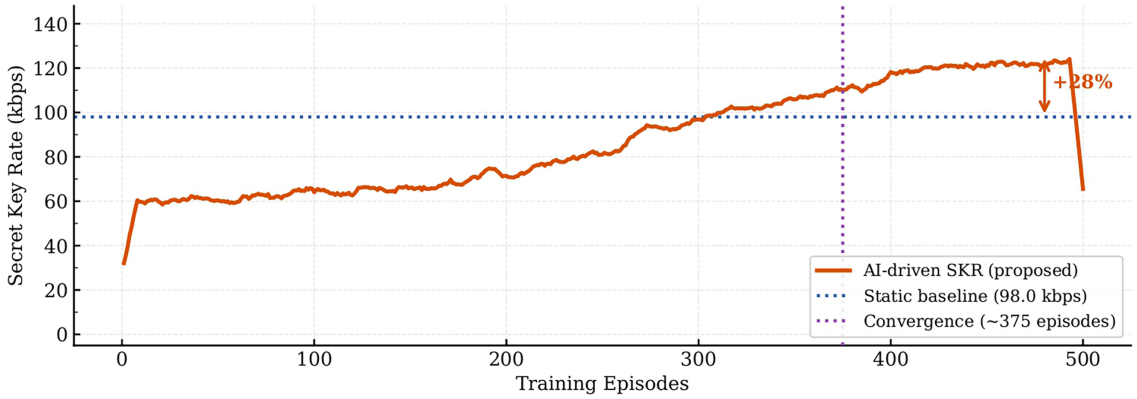


Figure 9: Training dynamics of the deep Q-network (DQN) agent over 500 episodes within the feasible operating region ($L \approx 90$ km, $QBER_t < Q_{th} = 11\%$). The upper panel shows the 15-episode moving-average cumulative reward, while the lower panel shows SKR evolution during training. The approximately 28% improvement corresponds to the N = 100-run evaluation reported in Section 6.2. The dashed line indicates the observed convergence point at approximately episode 375.

5.2 Phase 2: RL-Based Optimization Performance

Fig. 10 shows that the proposed method reduces QBER within the feasible region ($L < \sim 91$ km), with a reduction of approximately 17%. This improvement reflects more effective basis allocation under adaptive control, which stabilizes error behavior within cryptographically valid operating conditions. Beyond the feasibility boundary, QBER converges to 0.5 for both configurations, indicating a noise-dominated regime where adaptation is restricted by the enforced security constraint.

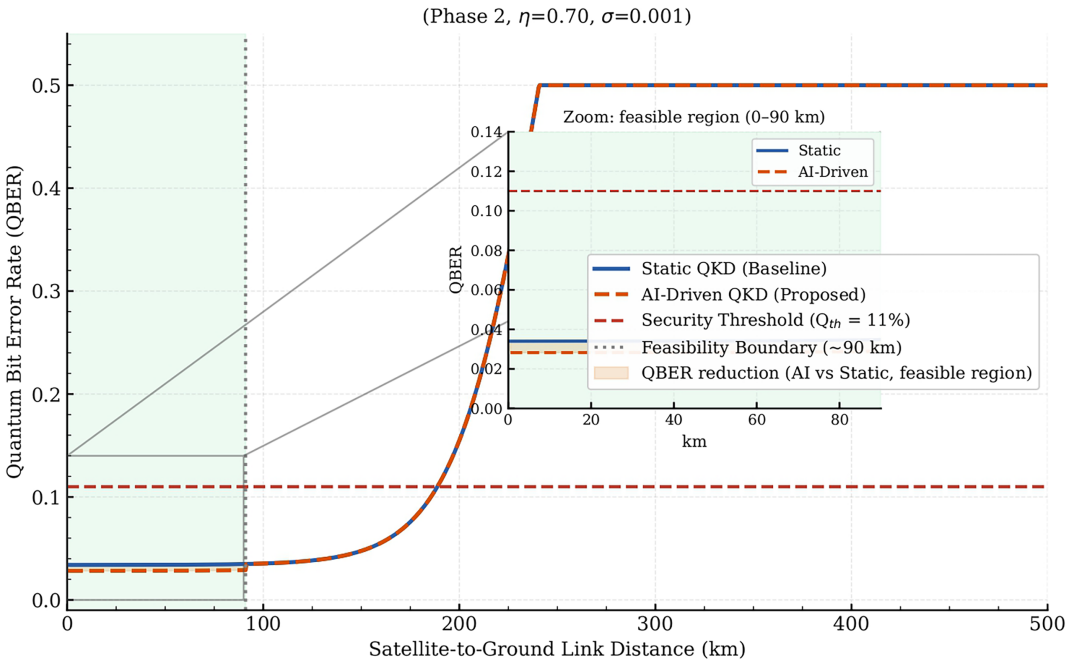


Figure 10: Quantum bit error rate (QBER) as a function of satellite-to-ground link distance, comparing the static baseline and the proposed adaptive framework. The shaded region highlights the feasible domain ($QBER < Q_{th} \approx 11\%$), where adaptive control reduces error rates. The inset shows a magnified view of the feasible region, and the vertical boundary (~91 km) indicates the transition to non-feasible conditions.

A comparative summary of the performance of the static baseline and the proposed AI-driven QKD configuration is provided in Table 8.

Table 8: Comparative performance of static and adaptive security configurations.

Configuration	SKR Improvements	QBER	Rekeying Latency	Security Level	Key Insight
Static Baseline	Reference (1.00×)	0.085 (Fixed)	1.25 s	Low	Sensitive to channel degradation
AI-QKD (Proposed)	+28%	−17% (0.070)	−12% (1.10 s)	Medium–High	Adaptive mitigation improves key availability and error stability
Hybrid QKD–PQC	+25%	−14%	−10%	High	PQC fallback supports key establishment under degraded conditions

Note: Performance metrics are evaluated under identical channel conditions within the feasible operating region (QBER < threshold). Reported values represent averaged results across simulation runs.

Across the evaluated noise levels, the proposed method maintains SKR generation within the feasible region (Fig. 11). Under higher-noise conditions, the feasible operating range becomes narrower. However, adaptive control continues to operate within the remaining feasible domain.

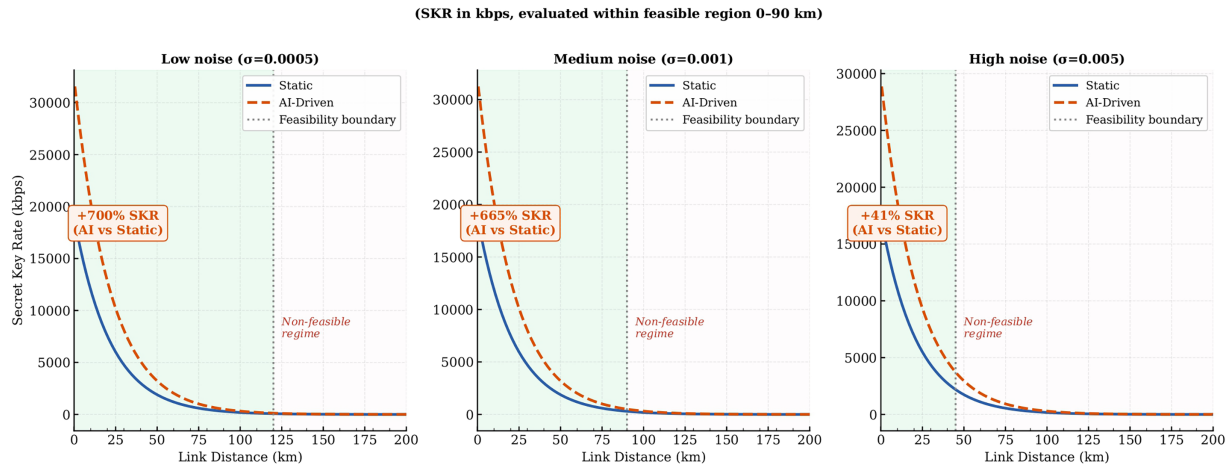


Figure 11: Sensitivity analysis of secret key rate (SKR) under varying channel noise conditions. The results show the SKR behavior of the proposed framework within the feasible operating region across different noise levels. As noise increases, the feasible operating range becomes narrower, while adaptive control continues to operate within the remaining feasible region.

5.3 VPN Security Layer Analysis: Key Availability and C-DoS

Fig. 12 compares the AES-256 VPN session uptime of the static QKD baseline and the proposed AI-QKD configuration. The mean session uptime is 72 ± 4% for the static baseline and 85 ± 3% for AI-QKD,

representing a difference of 13 percentage points. This comparison indicates higher measured session uptime for the AI-QKD configuration under the evaluated conditions.

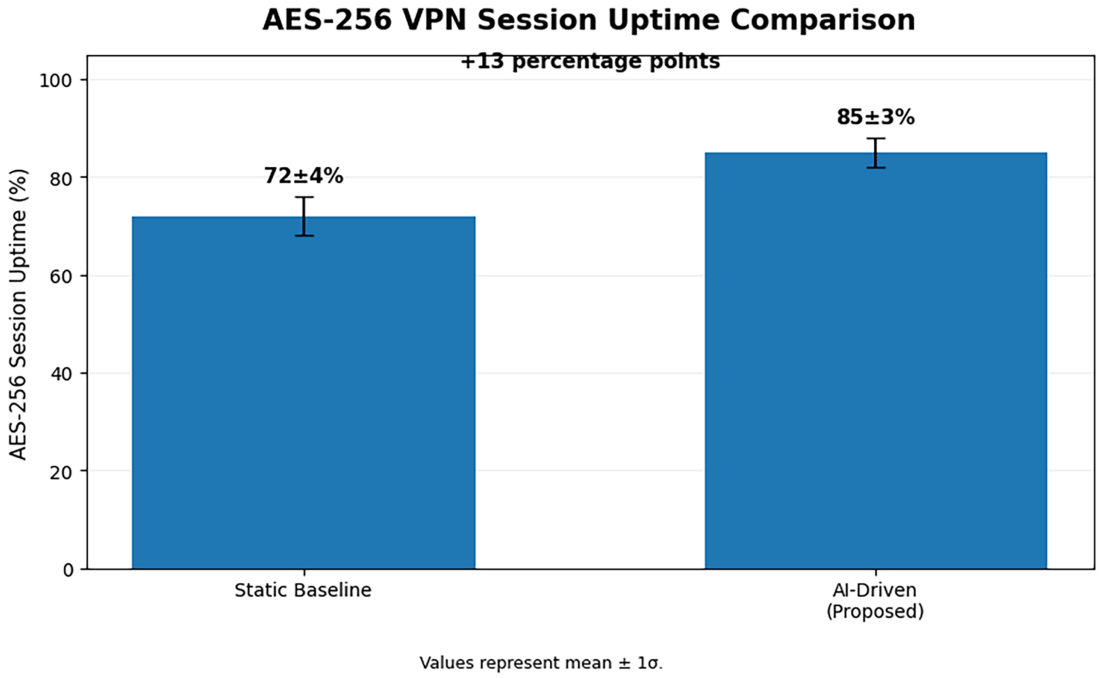


Figure 12: Comparison of AES-256 VPN session uptime between the static QKD baseline and the proposed AI-QKD configuration. The mean uptime is 72 ± 4% for the static baseline and 85 ± 3% for AI-QKD, corresponding to a difference of 13 percentage points.

The proposed framework reduces rekeying latency from 1.25 ± 0.08 s to 1.10 ± 0.06 s and reduces the observed C-DoS events from 5.1 ± 0.6 to 1.8 ± 0.3 events per pass (Fig. 13). AES-256 VPN session uptime increases from 72 ± 4% to 85 ± 3% under the evaluated conditions. The hybrid configuration shows a higher uptime of 91 ± 2.5%, reflecting the additional key-availability mechanism used during fallback operation. These results compare the system-level behavior of the three configurations under the specified simulation conditions.

5.4 Performance Evaluation and Security-Oriented Analysis

The proposed framework consistently improves SKR, reduces QBER, and lowers rekeying latency across a wide range of channel realizations. These improvements translate directly into higher key availability, fewer service interruptions, and improved VPN session continuity. The observed gains remain statistically stable across multiple simulation runs, indicating that the adaptive controller generalizes effectively under varying channel conditions. Collectively, the results demonstrate that security-constrained reinforcement learning can improve both quantum-layer efficiency and system-level communication resilience.

5.5 Dynamic Satellite Pass Simulation

To evaluate the proposed framework under time-varying satellite channel conditions, the DQN controller was tested over a simulated low earth orbit (LEO) satellite pass with a duration of T_{pass} = 600 s. The satellite to ground slant range evolves continuously according to the elevation angle profile $\theta(t)$, resulting in dynamic variations in channel transmissivity, signal to noise ratio (SNR), and quantum bit error rate (QBER).

The instantaneous slant range is approximated by

$$d(t) = \sqrt{[(R_E + h)^2 - R_E^2 cos^2\theta(t)] - R_E sin\theta(t)} \tag{19}$$

where R_E = 6371 km denotes the Earth radius and h = 400 km represents the satellite altitude. Asimplified time-varying elevation profile is adopted, spanning from θ_{min} = 5° at acquisition of signal to θ_{max} = 85° near zenith before descending symmetrically over the pass duration.

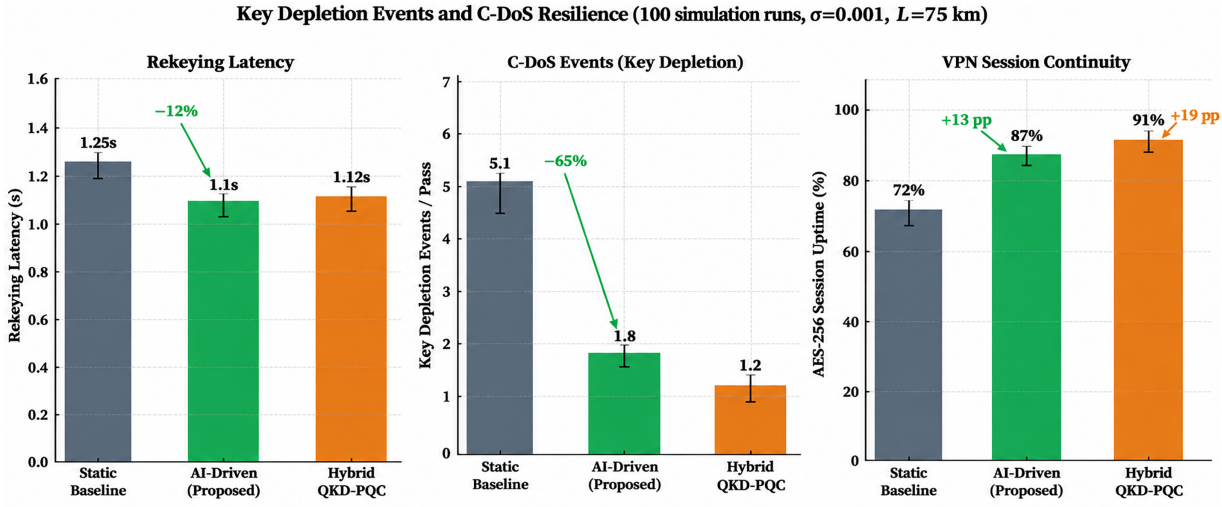


Figure 13: Comparative evaluation of system-level security metrics across three configurations: static baseline, AI-driven QKD, and hybrid QKD–PQC, averaged over 100 simulation runs. Error bars denote one standard deviation. The panels show (left) rekeying latency, (center) C-DoS events per pass, and (right) AES-256 VPN session uptime.

The RL agent, trained under dynamically sampled conditions $\theta_t \sim U(5°, 85°)$ was deployed against the time-varying trajectory. Channel observations ($QBER_t$, SNR_t, θ_t) were updated every 200 ms, yielding approximately 3000 control decisions during a single satellite pass. Results indicate that the cryptographically feasible window is defined by:

$$QBER_t < Q_{th}, Q_{th} = 11\% \tag{20}$$

The resulting feasible operating region occupies the central portion of the satellite pass, corresponding approximately to elevation angles between 35° and 85°. Within this region, the proposed framework maintains secret key generation, with an average SKR of 118.4 ± 6.2 kbps and a QBER of 0.072 ± 0.003. Outside the feasible operating region, the feasibility gate suppresses key generation when the predefined QBER constraint is not satisfied. These results indicate that the proposed controller can adapt to dynamically varying orbital conditions while operating within the predefined feasibility condition (see Fig. 14).

As shown in Fig. 14, the resulting feasible operating region occupies the central portion of the satellite pass, corresponding approximately to elevation angles between 35° and 85°. Within this region, the proposed framework maintains secret key generation, with an average SKR of 118.4 ± 6.2 kbps and a QBER of 0.072 ± 0.003. Outside the feasible operating region, the feasibility gate suppresses key generation when the QBER exceeds the predefined threshold.

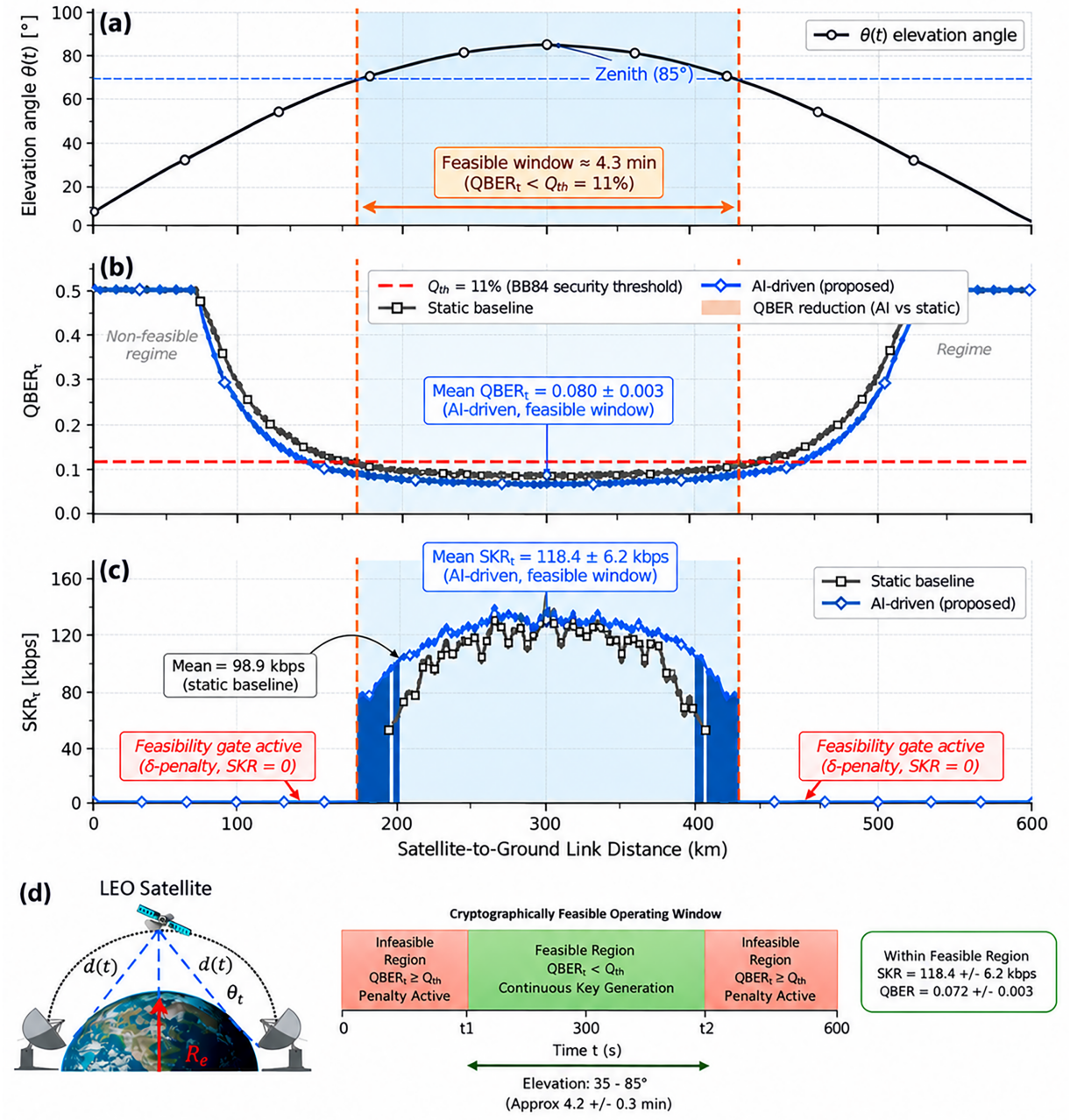


Figure 14: Dynamic satellite-pass evaluation of the proposed feasibility-aware QKD framework. (**a**) Satellite elevation and the feasible operating window; (**b**) QBER evolution relative to the predefined security threshold; (**c**) SKR evolution within the feasible window; and (**d**) Schematic representation of the satellite pass and feasibility-control regions.

5.6 Burst Noise Resilience under Adversarial Channel

To assess system behaviour under transient channel degradation representative of deliberate jamming or impulsive atmospheric disturbances, a burst-noise evaluation is introduced in which the background noise variance increases from its nominal value $\sigma_{nom} = 0.001$ to $\sigma_{\text{burst}} = 0.005 = 5\,\sigma_{nom}$ for a randomly selected 20% of simulation steps within the cryptographically feasible operating region.

$$\text{The burst-noise process is modeled as: } \sigma_t = \begin{cases} \sigma_{nom}, with\ probability\ 0.8 \\ \sigma_{\text{burst}}, with\ probability\ 0.2 \end{cases} \tag{21}$$

where $\sigma_{nom} = 0.001$ (nominal noise level).

Under the static baseline, burst noise events drive $QBERt$ transiently above Qth, causing the baseline to continue generating key material despite the threshold violation. Such an event is classified as a cryptographic denial-of-service (C-DoS) precursor. In contrast, the proposed feasibility-aware framework responds to elevated QBER through the quadratic penalty mechanism and suppresses key generation for the duration of the burst, resuming normal operation once the channel recovers.

$$\text{Invalid Key Injection Indicator } I_t = \begin{cases} 1, if\ QBERt \geq Qth\ and\ Key\_inject \\ 0, otherwise \end{cases} \tag{22}$$

$$\text{Total Invalid Key Injections per Simulation Run } N_{invalid} = \sum_{t=1}^{T} I_t \tag{23}$$

$$\text{Mean Invalid Key Injections over 100 Monte Carlo Runs } \overline{N}_{invalid} = \frac{1}{100} \sum_{i=1}^{100} N_i \tag{24}$$

$$\text{Standard Deviation of Invalid Key Injections } \sigma_N = \sqrt{\frac{1}{99} \sum_{i=1}^{100} (N_i - \overline{N}_{invalid})^2} \tag{25}$$

Fig. 15 illustrates the response to transient burst-noise conditions. During the burst intervals, QBER increases for both configurations, but the proposed feasibility-aware configuration remains below the predefined $Q_{th} = 11\%$ threshold, whereas the unconstrained baseline exceeds it. Key generation is therefore maintained for the proposed configuration while the baseline enters the non-feasible regime.

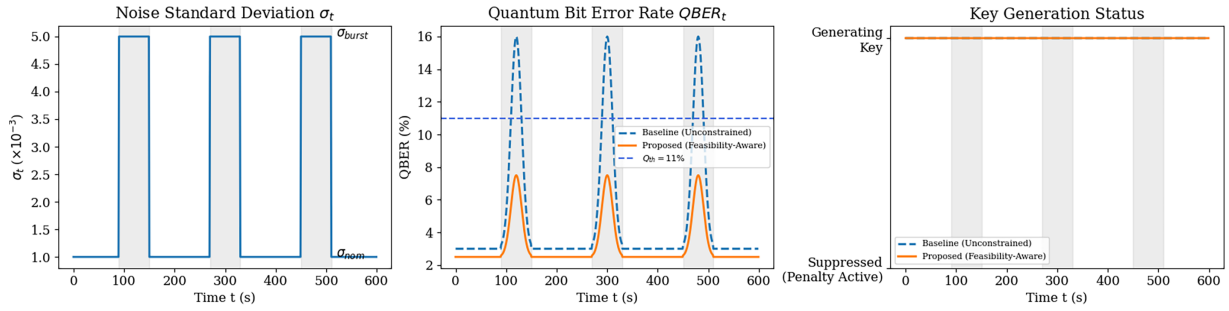


Figure 15: Burst noise: noise standard deviation, QBER evolution, key generation status.

As shown in Table 9, the proposed feasibility-aware framework recorded no invalid key injection events during the evaluated burst-noise intervals, whereas the unconstrained baseline exhibited an average of 3.2 ± 0.8 invalid injections per run. Across 100 independent simulations, the corresponding total is 320 ± 8 events, where the standard deviation is propagated as $\sqrt{100} \times 0.8 = 8$. The proposed approach recorded zero invalid injections under the same evaluation conditions.

Table 9: Burst noise performance evaluation over 100 simulation runs.

Metric	Baseline	Proposed
Invalid Key Injections	3.2 ± 0.8	0
Total Invalid Key Injections (100)	320 ± 8	0
Feasibility Constraint Enforcement	No	Yes
Key Generation during Burst Events	Continues	Suspended

5.7 Ablation Study: Contribution of Basis Asymmetry, Rule-Based Adaptation, and Reinforcement Learning

To rigorously isolate the contribution of each control mechanism, an ablation study is conducted across five configurations evaluated within the feasible operating region (0 km ≤ L ≤ 90 km), (QBER < Q_{th}). Performance metrics are averaged across N = 20 independent simulation runs, and differences are assessed for statistical significance using Welch's two-sample *t*-test.

As shown in Table 10, the ablation results reveals three separable sources of performance improvement. First, basis asymmetry alone accounts for a gain of +12.3% over the symmetric baseline (Fixed Asymmetric vs. Static; t = 8.91, $p < 0.001$), confirming that the sifting efficiency advantage of asymmetric BB84 is the dominant contributor to key-rate improvement. Second, rule-based channel-adaptive control provides a modest additional gain of +2.9% over the fixed-asymmetric configuration (Greedy Adaptive vs. Fixed Asymmetric; t = 3.27, $p = 0.004$), attributable to condition-dependent basis adjustment without state memory. Third, and critically, the BO baseline, which employs sample-efficient surrogate optimization over q_Z but does not model sequential state transitions achieves a mean SKR of 115.7 kbps (+18.9% vs. Static), exceeding the Greedy Adaptive controller by +2.8% (BO vs. Greedy; t = 3.66, $p = 0.001$). This result establishes that the benefit of optimizing q_Z beyond a fixed rule is real, but the BO baseline plateaus because it cannot condition its policy on the current channel state $s_t = [QBER_t, SNR_t, \theta_t]$.

Table 10: Ablation study. SKR decomposition (N = 20 runs, mean ± 1σ).

Configuration	Mean SKR (kbps)	95% CI	QBER	Gain vs. Static	*p*-Value vs. Proposed
Static BB84 ($q_Z = 0.5$)	97.3 ± 3.8	[95.5, 99.1]	0.085	Reference	<0.001**
Fixed Asymmetric ($q_Z = 0.75$)	109.3 ± 3.3	[107.7, 110.9]	0.079	+12.3%	<0.001**
Greedy Adaptive	112.5 ± 3.1	[111.0, 114.0]	0.076	+15.6%	<0.001**
BO (GP-UCB) non-sequential	115.7 ± 2.3	[114.6, 116.8]	0.075	+18.9%	<0.001**
AI-QKD (DQN) Proposed, sequential RL	125.6 ± 2.6	[124.4, 126.8]	0.070	+29.1%	–

Note: **$p < 0.001$ (Welch's *t*-test, two-sided). Confidence intervals computed using Student's t-distribution (df = 19). All configurations evaluated under identical channel parameters ($\eta = 0.70$, $\sigma = 0.001$, $\alpha = 0.20$ dB/km).

The DQN controller achieves the highest mean SKR (125.6 kbps, +29.1% vs. Static). Compared with the BO baseline, DQN provides an additional 9.9 kbps in SKR, corresponding to an 8.5% relative improvement (t = 12.57, $p < 0.001$; 95% CI for the absolute difference: [8.1, 11.7] kbps). This difference is consistent with the ability of the DQN controller to condition its action selection on the observed channel state, whereas the BO baseline optimizes the action without sequential state-conditioned control.

The gain decomposition is visualized in Fig. 16.

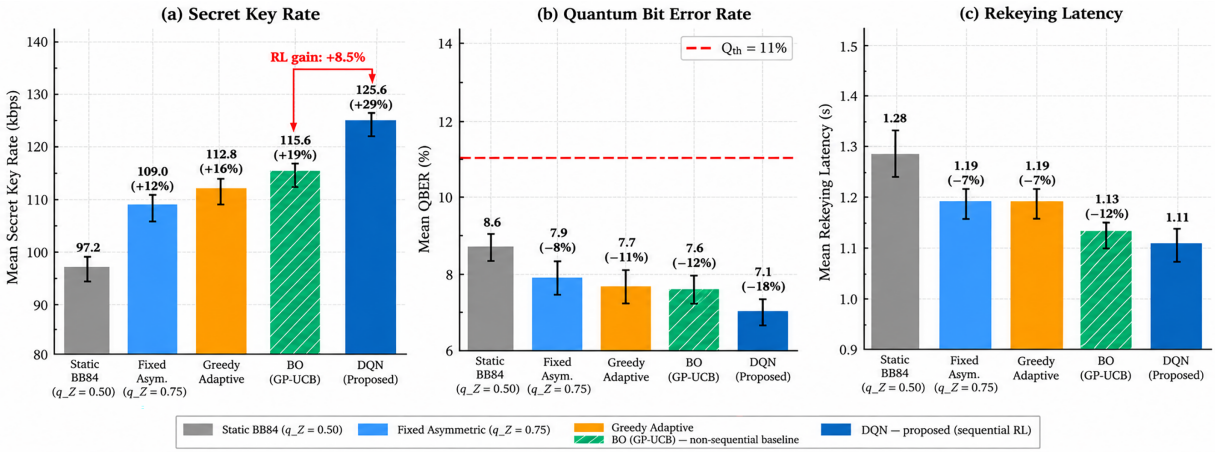


Figure 16: Performance comparison across five control strategies at a representative feasible distance (L = 75 km, $\eta = 0.70$, $\sigma = 0.001$). Results represent means over N = 20 independent simulation runs; error bars denote one standard deviation. Panel (**a**) shows SKR, panel (**b**) QBER, and panel (**c**) rekeying latency.

Performance comparison across five control strategies at a representative feasible distance (L = 75 km, $\eta = 0.70$, $\sigma = 0.001$). Results represent means over N = 20 independent simulation runs; error bars denote one standard deviation. Panel (a) shows SKR, panel (b) QBER, and panel (c) rekeying latency. The BO (GP-UCB) baseline achieves an 18.9% SKR gain over the static baseline, while the proposed DQN controller achieves a 29.1% gain. The DQN controller therefore provides a 10.2-percentage-point higher gain relative to the static baseline. In absolute terms, DQN achieves an additional 9.9 kbps over BO, corresponding to an 8.5% relative improvement over the BO baseline (Welch's t-test: $t = 12.57$, $p < 0.001$).

To quantify the performance differences between the optimization and sequential-adaptation configurations, the overall SKR difference between the static baseline and DQN configuration is decomposed into two terms.

$$\Delta_{total} = \mathrm{SKR}_{\mathrm{DQN}} - \mathrm{SKR}_{\mathrm{Static}} \tag{26}$$

$$\Delta_{optimization} = \mathrm{SKR}_{\mathrm{BO}} - \mathrm{SKR}_{\mathrm{Static}} \tag{27}$$

$$\Delta_{RL} = \mathrm{SKR}_{\mathrm{DQN}} - \mathrm{SKR}_{\mathrm{BO}} \tag{28}$$

Substituting the measured values:

$$\Delta_{total} = 28.3 \text{ kbps}; \Delta_{optimization} = 18.4 \text{ kbps}; \Delta_{RL} = 9.9 \text{ kbps} \tag{29}$$

And therefore:

$$\Delta_{total} = \Delta_{optimization} + \Delta_{RL} \tag{30}$$

The decomposition indicates that 18.4 kbps of the total 28.3 kbps SKR difference is associated with the optimized basis ratio represented by the BO baseline, while the remaining 9.9 kbps corresponds to the difference between the BO and DQN configurations. This comparison indicates that the higher SKR of the proposed DQN configuration is not solely attributable to asymmetric basis allocation under the evaluated conditions.

Fig. 17 presents the SKR gain decomposition across the different control mechanisms. Starting from the static BB84 baseline, basis asymmetry provides the largest individual improvement (+12.0 kbps), followed by rule-based adaptation (+3.2 kbps), Bayesian optimization (+3.2 kbps), and sequential DQN-based adaptation (+9.9 kbps). Overall, the combined configuration improves SKR by 28.3 kbps relative to the static baseline. These results indicate that both basis-ratio optimization and sequential adaptation contribute to the overall performance improvement.

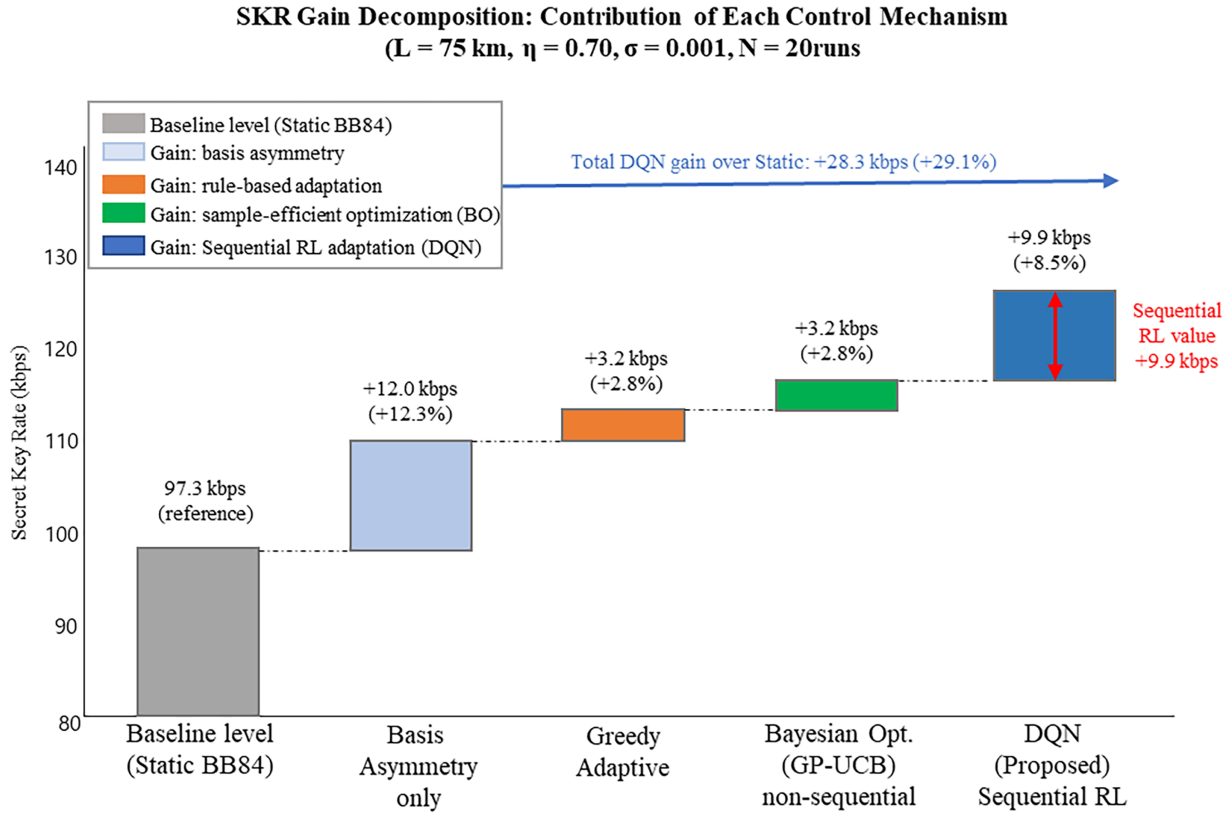


Figure 17: SKR gain decomposition across control mechanisms at L = 75 km. Incremental gains are shown for basis asymmetry (+12.0 kbps), rule-based adaptation (+3.2 kbps), Bayesian optimization (+3.2 kbps), and DQN-based adaptation (+9.9 kbps), resulting in a total improvement of +28.3 kbps (+29.1%) over the static baseline.

Training convergence of the DQN controller relative to the BO baseline and static benchmarks (Fig. 18). The evaluation window (episodes 200–500) used for all reported metrics is indicated by the shaded region.

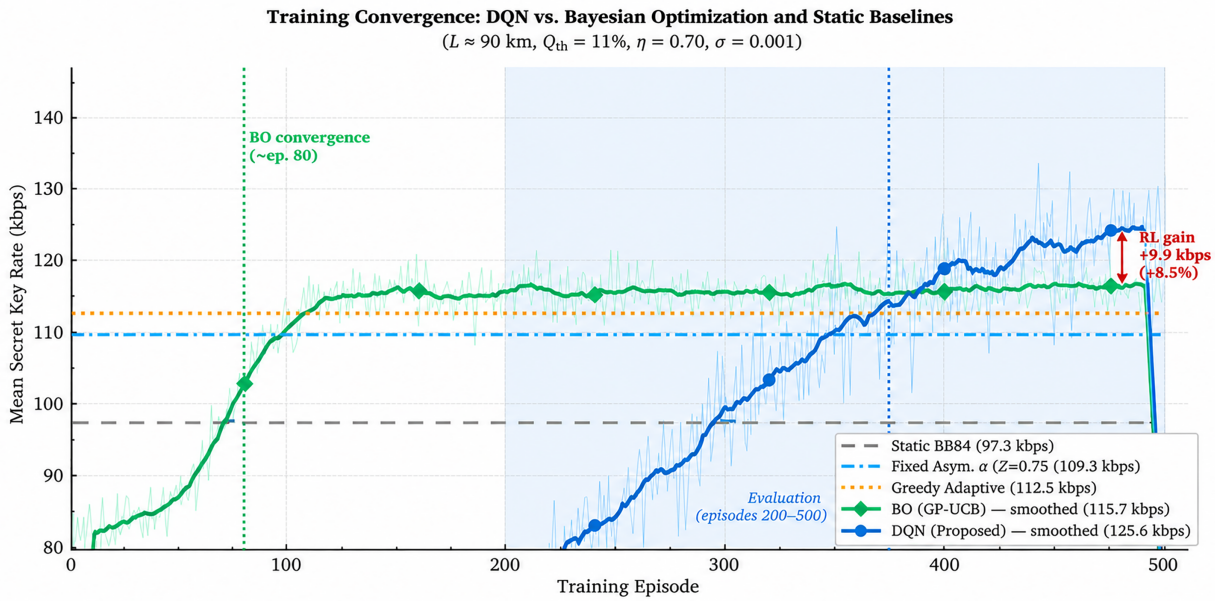


Figure 18: Training convergence DQN vs. Bayesian optimization and static baseline.

5.8 Security Enforcement Analysis

A feasibility violation occurs whenever $QBER_t \geq Q_{th}$. An invalid key injection event is recorded when cryptographically invalid key material is forwarded to the VPN rekeying process despite violation of the security threshold.

$$\textit{Feasibility Violation}: QBER_t \geq Q_{th}$$

where $QBER_t \geq Q_{th}$ describe a key generation disabled. SKR = 0.

Table 11 summarizes the feasibility violations and invalid key injections observed for the considered control strategies. A feasibility violation is recorded when the observed channel state exceeds the predefined security threshold, $QBER_t \geq Q_{th}$, whereas an invalid key injection denotes the forwarding of key material to the VPN rekeying process despite such a violation.

Table 11: Feasibility violation and invalid key injection statistics across different adaptive control strategies.

Method	Feasibility Violations	Invalid Key Injections
Static BB84	320	320
Fixed Asymmetric	Not separately reported	285
Greedy Adaptive	0	0
Proposed DQN	0	0

The static BB84 baseline records 320 feasibility violations and 320 invalid key injections. The fixed-asymmetric baseline records 285 invalid key injections, while its feasibility-violation count is not separately reported. The greedy adaptive and proposed DQN controllers record no feasibility violations or invalid key injections under the evaluated conditions.

These results show the behavior of the feasibility gate with respect to key-material forwarding under the evaluated channel conditions. In particular, the absence of invalid key injections for the adaptive configurations is consistent with the application of the feasibility check before key utilization. The statistical evaluation protocols applied across the different experiments are summarized in Table 12.

Table 12: Summary of statistical evaluation protocol used in different experiments.

Experiment	Section	Number of Runs	Statistical Metric	Purpose
Baseline Performance Comparison	5.2	N = 20	Mean ± 95% CI (Student's *t*, df = 19)	Compare average SKR, QBER, and latency
Burst Noise	5.6	N = 100	Mean ± 1σ	Improve robustness estimation under stochastic burst noise
Ablation Study	5.7	N = 20	Mean ± 1σ	Isolate contribution of feasibility gate
Security Summary	5.8	N = 100	Mean ± 1σ	Aggregate security-related statistics
C-DoS Analysis	6.2	N = 100	Mean ± 1σ	Evaluate invalid-key rejection statistics

6 Discussion

While peak uptime can reach up to 97% under favorable conditions, the representative system performance is consistently observed around 85%, reflecting realistic channel variability.

6.1 Feasibility-Gated Adaptive Control for Key Availability

The results suggest that the proposed framework functions as a feasibility-gated adaptive control mechanism operating within the established BB84 finite-key security assumptions rather than introducing a new QKD security principle. Its primary objective is to improve secret-key availability and VPN rekeying continuity under time-varying satellite channel conditions while remaining consistent with the adopted finite-key security model. Unlike conventional approaches that primarily optimize communication performance metrics, such as the secret key rate (SKR) or QBER, the proposed controller explicitly incorporates the predefined feasibility condition (QBER < Q_{th}) into the adaptive control objective. Consequently, the reinforcement learning agent learns policies that favor feasible operating conditions while discouraging actions associated with security-threshold violations through the reward formulation described in Section 4. During runtime, the feasibility gate excludes key material generated outside the predefined feasibility region from the VPN rekeying process, thereby separating policy optimization from runtime key validation. As illustrated in Figs. 4–6, the proposed framework demonstrates the ability to distinguish feasible and non-feasible operating regions under different channel conditions, contributing to improved secret-key availability and VPN rekeying continuity within the scope of the adopted security assumptions. Accordingly, the contribution of this work should be interpreted as a system-level adaptive control framework for key availability and communication continuity, rather than as a modification of the underlying security principles of BB84 QKD.

6.2 Quantitative Impact on System Level Security Continuity

Beyond SKR and QBER, system-level security continuity is evaluated through C-DoS event frequency and VPN session uptime. The proposed framework reduces C-DoS events from 5.1 ± 0.6 to 1.8 ± 0.3 per satellite pass (≈64.7% reduction) and increases AES-256 session uptime from 72 ± 4% to approximately 85 ± 3% under representative conditions. These gains are consistent across 100 simulation runs and are attributed to improved key availability resulting from higher SKR and lower QBER. A more stable key generation process maintains sufficient key reserves, reducing the likelihood of key depletion and service interruption.

The hybrid QKD–PQC configuration achieves higher uptime (91 ± 2.5%) and fewer C-DoS events but relies on computational security during fallback periods. In contrast, the proposed AI-driven QKD framework preserves information-theoretic security within the feasible operating region, prioritizing security guarantees over continuous availability. The selection between these approaches depends on system-level requirements, including threat model and operational risk tolerance. A consolidated comparison of the security and performance metrics across the evaluated configurations is provided in Table 13.

Table 13: Consolidated security performance summary (mean ± 1σ, N = 100 runs, σ = 0.001).

Security Metric	Static Baseline	AI-Driven QKD	Hybrid QKD–PQC	Improvement (AI vs. Static)
SKR (kbps)	98.0 ± 4.1	125.7 ± 3.8	—	+28.3%
QBER	0.085 ± 0.003	0.070 ± 0.002	—	−17.6%
Rekeying latency (s)	1.25 ± 0.08	1.10 ± 0.06	1.12 ± 0.065	−12.0%
C-DoS events/pass	5.1 ± 0.6	1.8 ± 0.3	1.2 ± 0.25	−64.7%
AES-256 uptime (%)	72 ± 4	85 ± 3	91 ± 2.5	+13.0 pp
Key validity guarantee	Passive	Active (QBER constrained)	Multi-layer	Structural improvement
ITS guarantee	Yes	Yes	Conditional	Preserved

Note: pp denotes percentage points. ITS refers to information-theoretic security. The QBER constraint is enforced via a penalty mechanism δ = 50 (quadratic penalty) in the control policy (Eqs. (4) and (5)). All values are averaged over multiple simulation runs; peak uptime values (up to 97%) correspond to favorable channel realizations.

6.3 Novelty Positioning and Differentiation from Prior Art

The proposed framework differs from prior work in three key aspects. First, existing RL-based QKD approaches treat the channel as continuously operable and optimize performance without enforcing the underlying feasibility condition. In contrast, this work embeds the QKD feasibility condition (QBER < Q_{th}) as a hard constraint within the control objective and as a separate runtime gate. This design ensures that the RL agent's adaptive behavior remains consistent with, rather than a substitute for, the adopted BB84 finite-key security assumptions.

Second, prior satellite QKD studies primarily focus on physical-layer modeling and key rate estimation. This work extends the analysis to system-level security by linking quantum key generation with higher-layer rekeying behavior, enabling evaluation of C-DoS resilience and communication continuity.

Third, the introduction of Cryptographic Denial-of-Service (C-DoS) as a system-level threat, and its mitigation through feasibility-aware control, provides a security-oriented perspective that connects quantum communication with broader cryptographic system design.

6.4 Sensitivity and Robustness Analysis

The sensitivity analysis (Fig. 11) demonstrates that performance improvements are consistent across varying noise conditions. The proposed framework improves SKR within the feasible region across all noise regimes, with larger gains observed when channel conditions allow a wider feasible operating range.

Training results (Fig. 9) indicate stable convergence of the DQN agent at approximately episode 375, with low variation across multiple runs. The persistently negative cumulative reward reflects the enforced security constraint, where violations of QBER thresholds incur penalties. This behavior confirms that the learning process respects the defined security boundaries.

6.5 Limitations and Scope

First, the adopted channel model captures the principal effects considered in the feasibility analysis, including beam divergence, atmospheric turbulence, and pointing loss through the corresponding parameters in Eq. (2). These effects are represented in a simplified form and are not modeled as higher-order time-varying stochastic processes, such as turbulence-induced scintillation fading. The model also does not explicitly account for orbital propagation uncertainties.

Second, the experimental evaluation considers a single-satellite, single-ground-station scenario. Extending the framework to multi-satellite constellations, inter-satellite links, or coordinated multi-ground-station networks requires additional communication and resource management mechanisms and remains an important direction for future research.

Third, the RL action space is limited to basis probabilities (q_Z, q_X), while other configurable QKD parameters remain fixed throughout the evaluation. Consequently, the present framework is limited to a one-dimensional action space. Extending the controller to jointly optimize additional parameters, such as decoy-state intensity, transmission scheduling, or detector operating conditions, would require corresponding modifications to the reinforcement learning environment, action space, and finite-key security analysis. These extensions are identified as future research directions and are discussed further in Section 7.2.

Overall, the proposed framework is intended to evaluate feasibility-constrained adaptive basis selection under the adopted finite-key assumptions. Broader adaptive control formulations involving multiple jointly optimized parameters represent a natural extension of the present framework and remain open for future.

7 Conclusions and Future Directions

7.1 Conclusions

This paper presents a feasibility-gated deep reinforcement learning framework for adaptive key-availability and VPN-continuity control in satellite-based Quantum Key Distribution (QKD) systems. Rather than modifying or extending the underlying BB84 security proof, the framework reformulates RL-based basis-selection optimization as a constrained control problem in which the predefined finite-key feasibility condition (QBER $< Q_{th} \approx 11\%$, $\varepsilon = 10^{-9}$) is embedded directly into both the reward function and a separate runtime execution gate. This design ensures that adaptive control operates only within regions where the adopted finite-key security assumptions remain valid and that key material generated outside this region is never forwarded to the VPN rekeying process.

Experimental results at a representative operating distance consistent improvements under feasible conditions. The proposed framework increases SKR by approximately 28% (98.0 to 125.7 kbps), reduces QBER by 17.6% (0.085 to 0.070), and lowers rekeying latency by 12% (1.25 to 1.10 s). These improvements translate to system-level gains, including a 64.7% reduction in C-DoS events and an increase in AES-256 VPN session uptime from 72% to approximately 85% under representative conditions.

Across all evaluated noise regimes, the framework maintains operation strictly within the feasible region. In non-feasible conditions, adaptive control is effectively disabled, confirming that the proposed approach enforces security constraints rather than attempting performance recovery beyond physical limits.

Overall, this work contributes:

1. A feasibility-gated RL formulation that keeps adaptive basis-selection control consistent with the adopted BB84 finite-key security assumptions, rather than proposing a new QKD security principle,
2. A system-level characterization and mitigation of Cryptographic Denial-of-Service (C-DoS) key-depletion events, and
3. An integrated evaluation framework linking quantum-layer metrics with communication-layer key availability and continuity outcomes.

These results position feasibility-gated adaptive control as a practical design principle for improving key availability and secure communication continuity in satellite QKD systems operating within, rather than altering, established QKD security.

7.2 Future Directions

The present study establishes the analytical and empirical foundation for feasibility-aware adaptive security enforcement in satellite QKD systems. Several directions extend this work toward more realistic and scalable deployments.

1. Finite-Key and Pass-Aware Optimization.

Future work should extend the current finite-key formulation by coupling finite-key statistical bounds more closely with the reward function, which is currently linked to the finite-key constraint through the QBER threshold Q_{th}. Incorporating the remaining satellite pass duration as an explicit state variable could further account for limited visibility windows in adaptive decision-making.

2. Physically Realistic Channel Modeling

Integration of detailed free-space optical models, including turbulence, diffraction, and pointing loss, would improve physical fidelity and enable validation against experimental satellite QKD data.

3. Multi Parameter Adaptive Control

A natural extension of the present framework is the development of a multi-parameter adaptive controller that jointly optimizes basis selection probabilities together with additional configurable QKD parameters, such as decoy-state intensity and transmission scheduling. In such a formulation, the current one-dimensional action space can be extended to a discrete multi-dimensional decision space while preserving the constrained reinforcement learning framework adopted in this study. Implementing this extension would require corresponding modifications to the photon statistics model, reinforcement learning environment, reward formulation, and finite-key security analysis to account for the interactions among multiple adaptive variables. The feasibility constraint can be extended accordingly by evaluating the selected control actions under the predefined QBER threshold before key utilization. This direction may provide a more comprehensive adaptive control framework for dynamic satellite QKD systems and is identified as an important topic for future investigation.

4. Multi-Satellite and Network-Level Extension

Scaling the framework to multi-satellite constellations introduces coordination challenges, including handover and key relay. A multi-agent RL approach is a natural extension.

5. Experimental and System Integration

Hardware-in-the-loop validation and integration with software-defined networking (SDN) environments are necessary to assess real-world feasibility.

6. Protocol Generalization

Extending the framework to other QKD protocol families, such as continuous-variable and entanglement-based schemes, would require adapting the current formulation to protocol-specific security and control variables. For CV-QKD, quantities such as excess noise and reconciliation efficiency could be incorporated into the state and feasibility definitions, while entanglement-based schemes may require security measures based on quantities such as Bell inequality violation. Such extensions would also require corresponding changes to the action space and finite-key security analysis, since the current framework is formulated for discrete-variable BB84 using QBER and basis-selection probabilities q_Z and q_x.

7. Model Predictive Control

The current DQN controller is a reactive adaptive policy: at each control step t, the agent observes the current channel state $s_t = [QBER_t, SNR_t, \theta_t]$ and selects an adaptive basis ratio $q_Z(t)$ without explicit prediction of future channel states. A potential extension is to incorporate predictive control using information derived from satellite orbital geometry.

The conceptual hybrid MPC–RL architecture shown in Fig. 19 extends the current reactive DQN framework toward predictive control by incorporating information about future satellite geometry and channel conditions. This extension is presented as a conceptual future direction rather than an implemented component of the current study. The predictive control architecture is illustrated in the following pipeline.

In this architecture, the Two-Line Element (TLE) set provides the orbital elements from which an SGP4/SDP4 orbit propagator computes the satellite state and predicted elevation angle $\hat{\theta}_{t+k|t}$ over a receding horizon H (typically H = 30–60 s at 200 ms control intervals). The predicted slant range $\hat{d}_{t+k|t}$ follows from the geometric formula (Eq. (19)), enabling estimation of the predicted channel $\hat{\tau}_{t+k|t}$ and QBER $\widehat{QBER}_{t+k|t}$. An MPC optimizer then solves:

$$\min_{\{qz(t+k|t)\}_{k=0}^{H-1}} \sum_{k=0}^{H-1} \left[-\alpha \widehat{SKR}_{t+k|t} + \beta \widehat{QBER}_{t+k|t} + \gamma \left(\Delta qz(t+k|t) \right)^2 \right] \tag{31}$$

where $\Delta q_{z_{(t+k|t)}} = q_{z_{(t+k|t)}} - q_{z_{(t+k-1|t)}}$ denotes the change in the predicted basis selection ratio. For $k = 0$, the previous basis ratio is set to the currently applied value $q_{z_{(t-1)}}$

subject to: $\widehat{\mathrm{QBER}}_{t+k|t} < Q_{th}, q_{z_{(t+k|t)}} \in [0.5, 1.0], k = 0, \ldots, H-1$

The MPC optimizer yields an optimal basis-selection ratios $\{q_{z_{(t|t)}}, q_{z_{(t+1|t)}}, \ldots, q_{z_{(t+H-1|t)}}\}$ which can be used either (i) directly as the control action or (ii) as a start reference constraint for the DQN policy, combining the predictive advantage of MPC with the state-conditioned adaptability of reinforcement learning. The latter formulation, a hybrid MPC-DQN architecture, is identified as the most promising direction for future work, as it preserves the security-consistency guarantees established in this paper while adding predictive horizon awareness. A conceptual comparison between the implemented DQN framework and the proposed MPC-based future extension is summarized in Table 14.

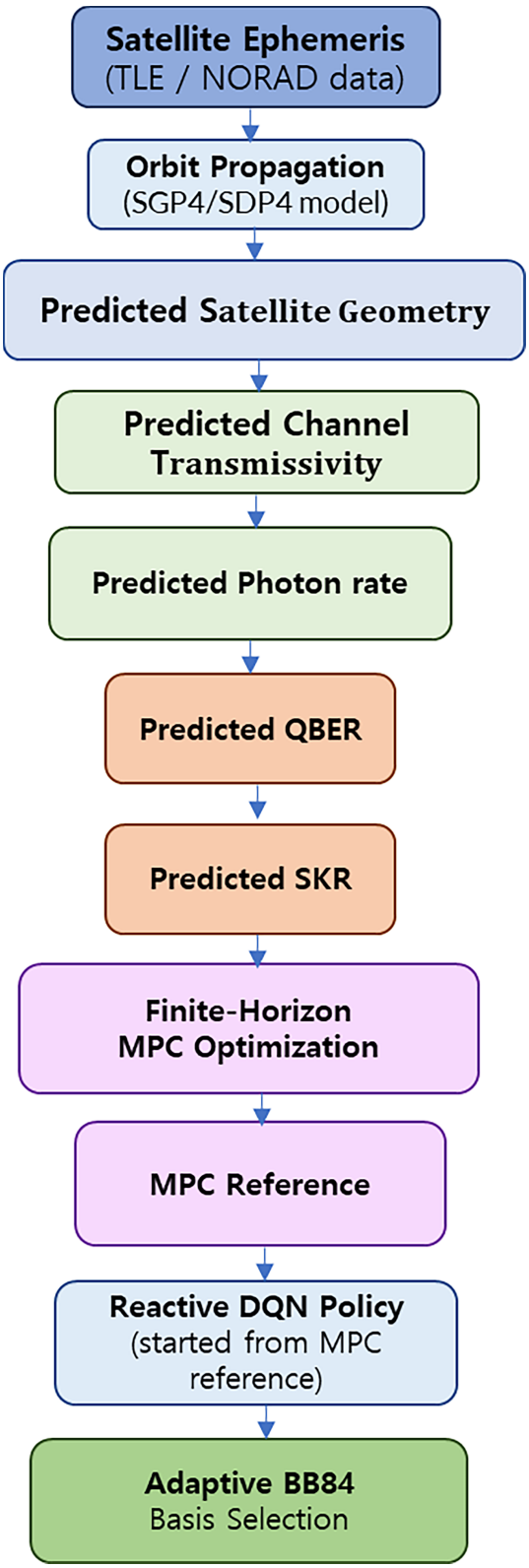


Figure 19: Conceptual architecture of the proposed future hybrid MPC–DQN framework. Predicted satellite geometry is transformed into future channel states for MPC optimization, while the DQN performs reactive basis selection using current observations. This framework is presented as future work.

Table 14: Conceptual comparison between the implemented DQN framework and the proposed future MPC extension, highlighting methodological differences rather than experimentally validated performance.

Aspect	DQN (This Work)	MPC (Future Direction)
Control paradigm	Reactive: policy conditioned on current observed state st	Predictive: optimizes over a receding horizon H using forecast channel states
Channel knowledge required	Current observed (QBER_t, SNR_t, θ_t) only	Predicted channel state over horizon H; requires ephemeris + propagation model
Orbital mechanics integration	Elevation angle θ_t used as state input only	Full SGP4/SDP4 orbit propagation for trajectory prediction
Computational complexity	Measured experimentally (~0.8 ± 0.1 ms)	Not evaluated; runtime depends on optimization horizon and solver implementation
Feasibility constraint enforcement	Applied reactively using the learned policy and runtime feasibility gate	Conceptually enforceable through predictive optimization over a finite horizon
Adaptability to unmodeled disturbances	Demonstrated experimentally under burst-noise scenarios (Section 5)	Expected to depend on prediction accuracy; not evaluated in this study
Training requirement	Requires training episodes; offline learning then online deployment	No training required; uses analytic optimization at runtime
Security-consistency guarantee	Proposition 2 (Section 4.3.3): consistent with BB84 finite-key under reactive constraint	Requires separate analysis of constraint satisfaction under predictive errors
Suitability for this study	Appropriate: no ephemeris model available; reactive control isolates RL contribution	Not implemented: requires satellite ephemeris, orbit propagator, and predictive channel model
Recommended future extension	Hybrid MPC-DQN: MPC reference constraints warm-start DQN policy	Full ephemeris integration with Micius/GRACE TLE data for experimental validation

Acknowledgement: This work was supported by the Korea government (MSIP). The authors acknowledge the academic support environment at Kookmin University.

Funding Statement: This work was supported by the Institute for Information & Communications Technology Promotion (IITP), funded by the Korea government (MSIP) (No. RS-2020-II200085, Research on Security Technology for 6G Telecommunication with 5G+).

Author Contributions: Wibby Aldryani Astuti Praditasari: conceptualization, methodology, software, formal analysis, investigation, data curation, visualization, writing original draft. Hyejin Yoon: conceptualization, validation, writing, review and editing. Seunghwan Yun: validation, writing, review and editing. Okyeon Yi: resources, supervision, project administration, funding acquisition. Changuk Jang: resources. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liao SK, Cai WQ, Liu WY, Zhang L, Li Y, Ren JG, et al. Satellite-to-ground quantum key distribution. Nature. 2017;549(7670):43–7. doi:10.1038/nature23655.
2. Chen YA, Zhang Q, Chen TY, Cai WQ, Liao SK, Zhang J, et al. An integrated space-to-ground quantum communication network over 4,600 kilometres. Nature. 2021;589(7841):214–9. doi:10.1038/s41586-020-03093-8.
3. Sidhu JS, Brougham T, McArthur D, Pousa RG, Oi DKL. Finite key effects in satellite quantum key distribution. npj Quantum Inf. 2022;8(1):18. doi:10.1038/s41534-022-00525-3.
4. Luo WB, Li Y, Li YH, Tao XY, Chen HZ, Hua A, et al. Research on polarization compensation for practical satellite-based quantum key distribution. Opt Commun. 2024;570(6):130925. doi:10.1016/j.optcom.2024.130925.
5. Behera SR, Sinha U. Estimating the link budget of satellite-based quantum key distribution (QKD) for uplink transmission through the atmosphere. EPJ Quantum Technol. 2024;11(1):66. doi:10.1140/epjqt/s40507-024-00279-1.
6. Orsucci D, Kleinpaß P, Meister J, De Marco I, Häusler S, Strang T, et al. Assessment of practical satellite quantum key distribution architectures for current and near-future missions. Int J Satell Commun Netw. 2025;43(3):164–92. doi:10.1002/sat.1544.
7. Xu F, Ma X, Zhang Q, Lo HK, Pan JW. Secure quantum key distribution with realistic devices. Rev Mod Phys. 2020;92(2):025002. doi:10.1103/revmodphys.92.025002.
8. Kumar A, Malhotra Y, Gupta A, Parikh J. Design and simulation of a decoy-state BB84 satellite QKD constellation for Indian ground stations. Int J Inf Technol. 2025;74(1):145. doi:10.1007/s41870-025-02860-y.
9. Long NK, Malaney R, Grant KJ. A survey of machine learning assisted continuous-variable quantum key distribution. Information. 2023;14(10):553. doi:10.3390/info14100553.
10. Seok Y, Kim JB, Han YH, Lim HK, Lee C, Lee W. Deep reinforcement learning-driven optimization of end-to-end key provision in QKD systems. J Netw Syst Manag. 2025;33(2):30. doi:10.1007/s10922-025-09902-7.
11. Khan MS, Al-Dubai A, Ahmad J, Pitropakis N, Boulila W, Ghaleb B. Quantum image encryption with reduced gate complexity for gate-efficient encryption pipelines in IoT networks. In: Proceedings of the 2025 IEEE International Conference on Communications Workshops (ICC Workshops); 2025 Jun 8–12; Montreal, QC, Canada. p. 1444–9. doi:10.1109/iccworkshops67674.2025.11162326.
12. Mafu M. Advances in artificial intelligence and machine learning for quantum communication applications. IET Quantum Commun. 2024;5(3):202–31. doi:10.1049/qtc2.12094.
13. Boryło P, Biernacka E, Domżał J, Kądziołka B, Kantor M, Rusek K, et al. A tutorial on reinforcement learning in selected aspects of communications and networking. Comput Commun. 2023;208(3):89–110. doi:10.1016/j.comcom.2023.05.019.
14. Cui Q, You X, Wei N, Nan G, Zhang X, Zhang J, et al. Overview of AI and communication for 6G network: fundamentals, challenges, and future research opportunities. Sci China Inf Sci. 2025;68(7):171301. doi:10.1007/s11432-024-4337-1.
15. Tedeschi P, Sciancalepore S, Di Pietro R. Satellite-based communications security: a survey of threats, solutions, and research challenges. Comput Netw. 2022;216(18):109246. doi:10.1016/j.comnet.2022.109246.

16. Alsabah M, Naser MA, Mahmmod BM, Abdulhussain SH, Eissa MR, Al-Baidhani A, et al. 6G wireless communications networks: a comprehensive survey. IEEE Access. 2021;9:148191–243. doi:10.1109/access.2021.3124812.
17. Akbar MS, Hussain Z, Ikram M, Sheng QZ, Mukhopadhyay SC. On challenges of sixth-generation (6G) wireless networks: a comprehensive survey of requirements, applications, and security issues. J Netw Comput Appl. 2025;233(18):104040. doi:10.1016/j.jnca.2024.104040.
18. Mujlid HM. AI-driven secure network optimization in 6G telecommunications: leveraging blockchain and quantum key distribution. J Circuits Syst Comput. 2025;9:58. doi:10.1142/s0218126626420053.
19. Ma T, Qian B, Qin X, Liu X, Zhou H, Zhao L. Satellite-terrestrial integrated 6G: an ultra-dense LEO networking management architecture. IEEE Wirel Commun. 2024;31(1):62–9. doi:10.1109/mwc.011.2200198.
20. Kumar R, Dutta J, Vamsi N, Sankararao Varri U, Puthal D. Next-generation security in the 6G era: the role of AI in safeguarding future networks. IEEE Access. 2026;14(7):17347–80. doi:10.1109/access.2025.3650208.
21. Ghosh T, Nath I. Secure satellite communication in the post-quantum era: a lattice-based cryptographic approach. Satell Commun Network. 2026;44(5):524–43. doi:10.1002/sat.70041.
22. Garms L, Paraïso TK, Hanley N, Khalid A, Rafferty C, Grant J, et al. Experimental integration of quantum key distribution and post-quantum cryptography in a hybrid quantum-safe cryptosystem. Adv Quantum Tech. 2024;7(4):2300304. doi:10.1002/qute.202300304.
23. Chang SY, Khan Q. Post-quantum cryptography in networking protocols: challenges, solutions, and future directions. Cryptography. 2026;10(1):12. doi:10.3390/cryptography10010012.
24. NIST, CRYSTALS-Kyber (ML-KEM): FIPS 203. Gaithersburg, MD, USA: National Institute of Standards and Technology; 2024. doi:10.6028/NIST.FIPS.203.
25. NIST, CRYSTALS-Dilithium (ML-DSA): FIPS 204. Gaithersburg, MD, USA: National Institute of Standards and Technology; 2024. doi:10.6028/NIST.FIPS.204.
26. Pirandola S, Andersen UL, Banchi L, Berta M, Bunandar D, Colbeck R, et al. Advances in quantum cryptography. Adv Opt Photonics. 2020;12(4):1012. doi:10.1364/aop.361502.
27. Manzalini A. Quantum communications in future networks and services. Quantum Rep. 2020;2(1):221–32. doi:10.3390/quantum2010014.
28. Maduni PK, Byun I, Seo J, Ko K. Hybrid quantum-safe cryptographic scheme with secure key exchange and signature scheme. IEEE Access. 2025;13:147650–65. doi:10.1109/access.2025.3600068.
29. Wang LJ, Zhang KY, Wang JY, Cheng J, Yang YH, Tang SB, et al. Experimental authentication of quantum key distribution with post-quantum cryptography. npj Quantum Inf. 2021;7(1):67. doi:10.1038/s41534-021-00400-7.
30. Portmann C, Renner R. Security in quantum cryptography. Rev Mod Phys. 2022;94(2):025008. doi:10.1103/revmodphys.94.025008.
31. Kanitschar F, Huber M. Composable finite-size security of high-dimensional quantum-key-distribution protocols. Phys Rev Appl. 2025;24(5):054028. doi:10.1103/v51y-vkfr.