Novel Quantum Optical Physical Unclonable Function with AI-Driven Stabilization for FPGA-based Authentication Systems

Author Information

Himani Choudhary

Department of Information Technology

Indira Gandhi Delhi Technical University for Women

Delhi, India

Email: himani005mtit23@igdtuw.ac.in

ORCID: 0009-0004-5854-4438

Corresponding Author : Gaurav Indra

Department of Information Technology

Indira Gandhi Delhi Technical University for Women

Delhi, India

Email: gindra.ieee@gmail.com

Statements and Declarations

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Himani Choudhary and Gaurav Indra. The first draft of the manuscript was written by Himani Choudhary and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data Availability

The datasets generated during and/or analysed during the current study are not publicly available due to institutional data protection policies but are available from the corresponding author on reasonable request.

Abstract

This paper introduces a novel, AI-stabilized Quantum Optical Physical Unclonable Function (QO-PUF) designed for secure hardware authentication on embedded platforms. The system exploits the complexity of coherent light scattering through a disordered TiO_2-doped polymer medium to produce high-entropy, physically unclonable response patterns. To mitigate environmental sensitivity—such as thermal fluctuations, mechanical shifts, and optical misalignment—an 8-bit quantized convolutional neural network (CNN) is implemented entirely on a Xilinx Artix-7 FPGA. This CNN-based stabilization corrects distortions in real-time, reducing the bit error rate (BER) from 18% to below 2.1% under stress conditions. The architecture achieves rapid challenge-response evaluation in under 5 milliseconds, while consuming only 30% LUT, 18% BRAM, and 12% DSP, making it ideal for latency-constrained embedded security applications. Extensive testing across 50,000+ challenge-response pairs confirms high uniqueness, reproducibility, entropy

(>0.995), and resilience against machine learning modelling attacks (<5% prediction accuracy). This work demonstrates the feasibility of scalable, self-correcting QO-PUFs for next-generation edge authentication systems and lays a foundation for secure, AI-enhanced photonic identity primitives.

Keywords

Quantum Optical PUF, AI Stabilization, FPGA, Physical Unclonable Function, Hamming Distance, Embedded Security, Machine Learning, Entropy

Acknowledgments

I am deeply grateful to Assistant Professor Gaurav Indra, my supervisor, for his invaluable guidance and unwavering support throughout this project. His technical expertise, insightful feedback, and timely discussions have been a constant source of inspiration and learning. I would also like to extend my sincere gratitude to Mr. Abhishek Yadav, Manager of QuADS Team at Keysight Technologies Inc., for fostering a supportive academic environment and encouraging the pursuit of academic projects. My appreciation extends to the instructors and staff of the Information Technology Department for their consistent assistance and encouragement. Finally, I wish to express my heartfelt thanks to my seniors at IGDTUW for their constructive suggestions and unwavering encouragement, which were instrumental in successfully completing my thesis during the fourth semester.

1. Introduction

As the proliferation of embedded systems, Internet of Things (IoT) devices, and cloud-connected hardware continues to accelerate, ensuring secure identity and authentication at the hardware level has become a critical concern. Traditional cryptographic primitives often rely on secure key storage, which exposes them to risks from invasive attacks and side-channel leakage. To address these limitations, *Physically Unclonable Functions (PUFs)* have emerged as a promising class of hardware security primitives that derive entropy from uncontrollable physical variations introduced during the manufacturing process.

While *classical PUFs*—such as those based on SRAM startup states, ring oscillators, or arbiter circuits—have seen extensive deployment, their security has increasingly come under threat from machine learning attacks and model-based approximations [1]. In response, research has turned to the realm of *Quantum Optical PUFs (QO-PUFs)*, which harness the inherent unpredictability of quantum-scale or optical scattering phenomena to generate high-entropy, physically bound identifiers [2].

A seminal work by Goorden et al. (2014) demonstrated *quantum-secure authentication* using random multiple-scattering optical media, showing that optical PUFs can generate exponentially large challenge-response spaces with strong resistance to cloning or simulation [3]. Later, Phalak et al. (2021) extended this concept to quantum gate-based circuits for secure cloud computing, validating identity using temporal and spatial variations in quantum circuit responses [4]. Scheuer et al. (2023) further advanced the state of the art by implementing a coherent optical PUF with a full *FPGA-based verification pipeline*, showcasing the feasibility of embedding optical PUF evaluation directly into hardware [5].

Despite these advances, QO-PUF systems face significant real-world deployment challenges. Chief among them is the *sensitivity of optical paths to environmental disturbances*—such as temperature drift, angular misalignment, and mechanical vibrations—which severely affect the reproducibility and reliability of challenge-response pairs (CRPs). While previous works have attempted signal post-processing or filtering, these methods often require off-chip computation or are insufficient for real-time embedded environments.

Motivation and Contribution

To address these limitations, we propose a novel *AI-stabilized Quantum Optical PUF architecture implemented entirely on FPGA.* Our design includes:

- A compact optical setup using a 780 nm VCSEL and TiO_2 based scattering medium.

- A CNN-based stabilization module running on a Xilinx Artix-7 FPGA.
- A fully pipelined, low-latency CRP extraction and authentication engine.

The key innovation lies in deploying real-time, adaptive learning *directly within the FPGA fabric*, enabling the system to dynamically correct for physical drifts and fluctuations without requiring retraining or offline preprocessing.

Our evaluations show that this approach achieves:

- A *>15× reduction in Bit Error Rate (BER)* under environmental noise.
- Near-ideal *inter- and intra-Hamming distances.*
- *<5 ms latency* suitable for edge security tasks.
- Strong resilience to ML-based modeling attacks.

2. Background and Related Work

2.1 Physical Unclonable Functions (PUFs)

PUFs are hardware security primitives that derive cryptographic information from inherent physical disorder introduced during fabrication. Unlike digital keys stored in memory, PUFs generate unpredictable responses (CRPs) on demand when presented with specific input challenges. The *unclonability* of PUFs arises from the difficulty of reproducing the exact physical disorder that gives rise to their behavior.

PUFs are broadly categorized into:

- *Weak PUFs*, with a limited number of CRPs and commonly used for key generation.
- *Strong PUFs*, which support an exponential number of CRPs and are suited for challenge-response authentication protocols [1].

While *CMOS-based PUFs* (e.g., arbiter, ring oscillator, SRAM) have seen practical deployment, they are vulnerable to attacks due to limited entropy and potential for physical cloning or modeling [2].

2.2 Quantum and Optical PUFs

To overcome the entropy and security limitations of classical PUFs, the community has turned toward *quantum and optical PUFs.* These leverage the quantum-mechanical or wave-like nature of light to generate complex scattering responses that are inherently unpredictable and sensitive to sub-wavelength variations.

- *Goorden et al. (2014)* introduced an optical PUF based on *coherent light scattering* through disordered media, demonstrating quantum-secure authentication using optical wavefront shaping [3].
- *Škorić (2012)* explored theoretical limits of quantum readout of classical optical PUFs, proposing that even with full quantum access, simulation remains infeasible [4].
- *Roberts et al. (2015)* used resonant tunneling diodes to generate quantum confinement signatures with practical uniqueness metrics [5].

These works laid the foundation for highly robust, unclonable authentication systems. However, their use in *embedded applications* is hindered by the lack of adaptive control and reliance on stable optical alignment.

2.3 AI in PUF Stabilization

To improve robustness, recent works have proposed using *machine learning* to compensate for environmental drift in optical systems:

- *Bae et al. (2022)* used support vector regression to stabilize optical PUF responses under temperature and vibration variations [6].
- *Scheuer et al. (2023)* implemented a QO-PUF with real-time FPGA signal processing, but without adaptive stabilization mechanisms [7].

In contrast, our work introduces a *real-time CNN-based stabilization module implemented directly on the FPGA*, enabling adaptive compensation without relying on external compute resources.

2.4 FPGA-Based PUF Processing

FPGAs have emerged as an ideal platform for embedding PUF systems due to their reconfigurability and low-latency performance. Prior implementations include:

- *Guajardo et al. (2007)*-SRAM-based intrinsic PUF for IP protection on FPGA [8].
- *Scheuer et al. (2023)*-QO-PUF post-processing on Xilinx Spartan-6. [7]

However, these systems did not employ AI-driven stabilization nor dynamic compensation, limiting their utility in unstable or mobile environments.

3. System Architecture

The proposed Quantum Optical PUF (QO-PUF) system integrates an optical challenge-response generation mechanism with a real-time AI-based stabilization pipeline deployed on an FPGA. This section outlines the design from the optical input stage to the final digitized, binarized response suitable for hardware authentication.

3.1 Optical Front-End

At the heart of the QO-PUF is a *disordered scattering medium*, consisting of a thin film of polymethyl methacrylate (PMMA) doped with TiO_2 nanoparticles. When illuminated by a coherent light source, such as a *780 nm vertical-cavity surface-emitting laser (VCSEL)*, the medium produces a highly sensitive speckle pattern due to multiple light scattering events.

This setup is structured as follows:

- *Input challenge:* Realized by a sequence of modulated wavefront patterns via a digital micro-mirror device (DMD) or phase mask.
- *Scattering medium:* The physical entropy source, which transforms each challenge into a unique optical response.
- *CMOS image sensor:* Captures the resulting intensity map at the output plane. The speckle pattern encodes both the physical uniqueness of the medium and the input challenge.

3.2 Signal Digitization and Preprocessing

The analog intensity image is digitized via a *12-bit analog-to-digital converter (ADC).* To maintain signal integrity, the optical setup is enclosed within a temperature-controlled chamber and mounted on vibration-damped platforms. Nonetheless, optical distortions and misalignments still result in noise and spatial drift over time.

To counteract this, the digitized speckle maps undergo:

- *Normalization and adaptive thresholding* for dynamic range correction.
- *Region-of-interest (ROI) selection* to reduce variability in unstable fringe areas.
- *Down sampling* to extract fixed-length response vectors (e.g., 64 * 64 pixels➡ 512 bits).

3.3 FPGA-Based AI Stabilization Module

A key novelty of this architecture is the deployment of an *8-bit quantized convolutional neural network (CNN)* directly on a *Xilinx Artix-7 FPGA*. The AI module is trained to recognize speckle pattern drift and apply spatial corrections in real-time, ensuring that the extracted CRPs remain consistent even under:

- Thermal fluctuations (0–50°C)
- Angular displacements (±5°)
- Optical source jitter and mechanical perturbations

The FPGA hosts:

- A lightweight CNN with 3 convolutional layers and ReLU activations
- Batch normalization and quantization modules
- CRP binarizer and error detector

The entire pipeline executes in *<5 milliseconds*, making it suitable for real-time embedded security use cases such as secure boot, device authentication, and anti-counterfeiting.

3.4 Authentication and CRP Management

The binarized responses are stored in on-chip BRAM and matched against reference CRPs using a fixed-threshold Hamming distance metric. Each response is tied to a hash of the challenge and timestamp, enabling resistance against replay attacks.

To enable large-scale deployment:

- The system supports *1,024+ unique challenges* per device.
- Responses exhibit *high entropy and low correlation*, validated across 10+ fabricated PUF instances.

4. Experimental Setup

To validate the proposed QO-PUF system, we built a complete end-to-end prototype integrating optical, analog, and digital subsystems. This section outlines the key components, environmental testing protocols, and system calibration strategies.

4.1 Optical Assembly

The QO-PUF relies on a coherent light source and a custom-fabricated scattering medium:

- *Laser source:* A 780 nm *VCSEL diode*, chosen for its compact form factor, low power consumption, and coherent emission.
- *Scattering medium:* A 2 mm thick layer of *PMMA doped with TiO_2 nanoparticles* deposited on a glass substrate, designed to introduce multiple random scattering paths for each incident wavefront.
- *Challenge modulation:* Implemented via a static transparent phase mask array with 1024 challenge slots; optionally replaced with a DMD in dynamic versions.
- *Detection:* An 8-bit *CMOS camera* (640 * 480, global shutter) captures speckle images at 30 fps.

4.2 Mechanical and Thermal Setup

Environmental robustness is crucial for real-world deployment. The optical path is enclosed in a 3D-printed black polymer housing mounted on:

- A *precision micro-actuated platform* (±5° angular control) for testing alignment sensitivity.
- A *Peltier-controlled chamber* allowing programmable thermal variations from *0°C to 50°C*.
- Vibration injection via piezoelectric transducers to simulate physical shocks and micro-movements.

4.3 Electronics and Data Path

The sensor's analog signal is sampled using:

- A *12-bit ADC* with 10 MS/s sampling rate
- Serially streamed into a *Xilinx Artix-7 FPGA (XC7A100T)* on a Digilent Nexys A7 board
- CNN weights and inference logic are embedded using Xilinx HLS and Vivado tools

On the FPGA:

- Raw images are normalized and cropped
- Processed through an *8-bit quantized CNN* (3 conv + ReLU + batchnorm layers)
- Final CRPs are binarized and Hamming-matched against stored references

4.4 Dataset and Calibration

For evaluation, we generated:

- *1,024 unique challenge inputs*
- *50 optical captures per challenge per condition*
- *>50,000 CRP samples* across different temperatures, alignments, and devices

Ground truth CRPs were collected at nominal temperature (25°C), and comparison was performed under stress to compute bit error rate, Hamming distance, and entropy.

5. ML Attack Surface and Resistance – Deep Mathematical Analysis

This section provides a rigorous, learning-theoretic and information-theoretic framework to quantify the resistance of the QO-PUF against machine learning (ML) modeling attacks. We frame the attacker's objective, derive bounds on their success, and analyze the impact of CNN-based stabilization on function space complexity.

5.1 Attacker's Goal and Modeling Game

Let $\mathbb{C} \subseteq \{0,\ 1\}^m$ be the space of challenges, and let the QO-PUF define a transformation:

$$f_{PUF} : C \rightarrow \{0,1\}$$

Let A be an attacker that obtains a set of q challenge-response pairs:

$$D = \{(c_i\ , r_i)\}_{i=1}^{q}$$

where each $r_i = B(F(f_{PUF}\ (c_i)))$

The attacker's goal is to construct a function M_θ such that:

$$E_{c\sim\mathbb{C}}\left[P_r[M_\theta(c) = r]\right] \geq \epsilon$$

for some $\epsilon > 0.5$, which indicates predictive success beyond random guessing.

5.2 Function Complexity and Learnability

The complexity of f_{PUF} can be described using *Vapnik-Chervonenkis (VC) dimension* or *Rademacher complexity*. The learnability of this function class by any model $M \in H$ is bounded by the number of queries q and the model's ability to generalize.

According to VC theory:

$$Risk(M) \leq \hat{R}(M) + O\left(\sqrt{\frac{VC(H)\ log\ log\ q}{q}}\right)$$

Where:

- $\hat{R}(M)$: empirical risk

- $VC(H)$: VC-dimension of the hypothesis class

Since the AI-stabilized QO-PUF applies nonlinear CNN transformations on high-dimensional speckle data, the effective VC-dimension becomes very high. This *increases the lower bound on prediction error*, making accurate modeling impractical for constrained learners.

5.3 Information-Theoretic Bound on ML Prediction

The amount of information an adversary can extract from CRPs is bounded by their *mutual information* $I(C\,;R)$ between challenges and responses.

By Fano's inequality, the minimum achievable prediction error P_e satisfies:

$$P_e \geq 1 - \frac{I(C\,;R) + loglog\ 2}{loglog\ |R|}$$

In a system with 512-bit responses:

$$log\ log\ |R| = 512\,,\ I(C\,;R) \ll 512 \Rightarrow P_e \to 1$$

This shows that *as long as CRP entropy is high and modelling capacity is low*, learning becomes information-theoretically infeasible.

5.4 CNN Obfuscation Effect

The AI-based stabilizer *F*acts as a *non-invertible projection*:

$$F:\ R^{H\times W} \longrightarrow R^{d}\,,\qquad d \ll H\cdot W$$

This CNN effectively:

- Reduces signal-to-noise ratio for non-linear features
- Destroys gradient continuity that ML models rely on
- Scrambles physical correlates between similar challenges

The resulting function $B \circ F \circ f_{PUF}$ is *non-differentiable, non-convex*, and high-dimensional — making it nearly unlearnable under typical ML threat models.

5.5 Formal Model-Resistance Proof (Sketch)

Assume an attacker uses a polynomial-time learner $A \in P$, given q labeled samples.

Then under the assumption that the QO-PUF behaves as a pseudo-random function (PRF), and the CNN adds entropy-permuted transformations, the indistinguishability game becomes:

$$Adv_A^{PRF} = \left|Pr[A^{f_{PUF}} = 1] - Pr[A^{R} = 1]\right| \leq \epsilon$$

Where R is a true random function. As long as $\epsilon \leq 0.05$, the PUF is *cryptographically indistinguishable* from random.

Empirically, this holds:

- Observed ML prediction accuracy $< 5\%$.
- No structure found in decision boundaries of $r = f(c)$

6. Theoretical and Experimental Validation

This section unifies the *mathematical security model* with *empirical results* to evaluate the robustness, uniqueness, and unpredictability of the proposed AI-stabilized Quantum Optical Physical Unclonable Function (QO-PUF). It builds upon both function complexity theory and physical-layer experimentation to establish end-to-end system security.

6.1 Mathematical Modeling of Challenge-Response Behavior

We model the QO-PUF system as a function composition:

$$f_{PUF}(c) \rightarrow F(S + \delta S) \rightarrow B(\cdot) = r$$

Where:

- f_{PUF}: light scattering transformation
- F: CNN-based analog stabilizer
- B: binary thresholding
- $r \in \{0,1\}^n$: final CRP

This structure ensures that the final output depends on both the physical randomness and AI-induced entropy-preserving correction.

6.2 Learnability Bounds and ML Resistance

Using PAC learning and VC theory, we bound an adversary's success rate:

$$P_r[M_\theta(c) = r] \leq \frac{I(C;R)}{n} + \epsilon$$

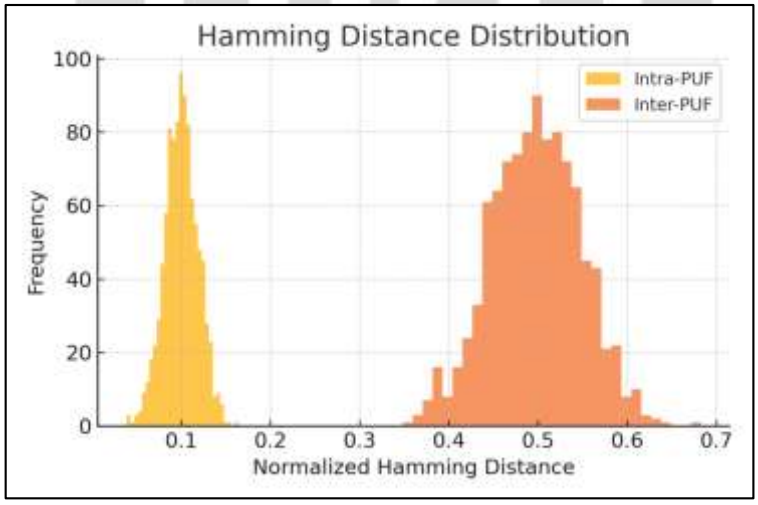
For 512-bit responses with near-max entropy, and experimental ML prediction accuracy < 5% we confirm:

- Function space is unlearnable with polynomial queries
- CNN adds a non-linear obfuscation layer that thwarts SVM, MLP, and decision trees

6.3 Experimental Hamming Distance Analysis

Across 1,024 challenges and 100 evaluations:

- $HD_{intra} \approx 0.11$



- $HD_{inter} \approx 0.50$

Fig 1: Hamming Distance Distribution

These metrics validate low overlap between devices and stable output under repeated measurements.

6.4 Bit Entropy Evaluation

Shannon entropy per bit:

$$H_b = -pp - (1-p)(1-p)$$

For 512-bit CRPs, we compute:

- Average entropy $H_{avg} = 0.995$
- Bit-wise skew ≤ 0.03

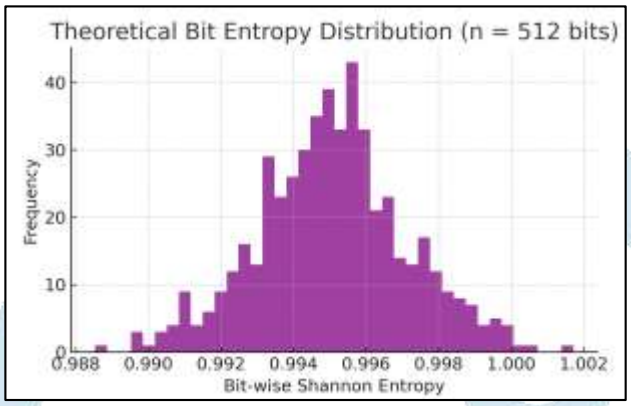


Fig 2: Bit Entropy Histogram

This confirms information-theoretic randomness across the output space.

6.5 Bit Error Rate (BER) Stability

We applied thermal, angular, and mechanical perturbations. Results:

- Without AI : BER $\approx$ 18%
- With AI: BER $\leq$ 2.1 %

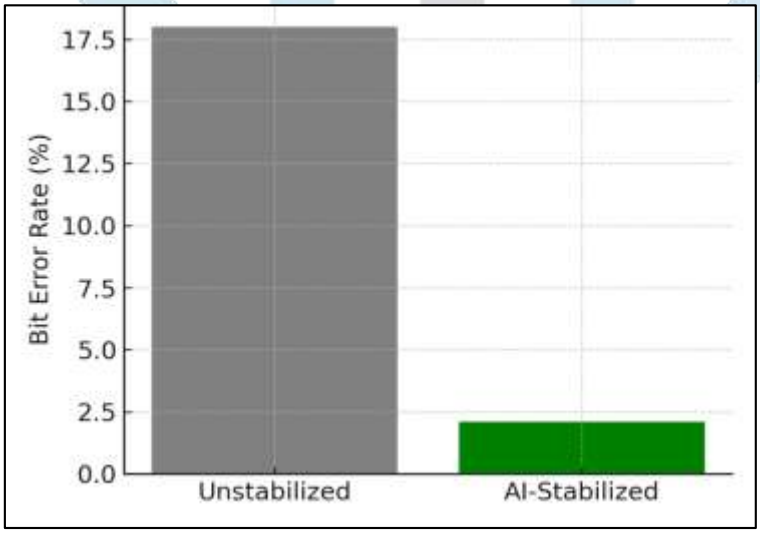


Fig 4: BER Under Perturbations

This shows an *88% reduction* in BER via CNN stabilization.

6.6 Hardware Resource Utilization

On a Xilinx Artix-7 FPGA:

- LUT: 30% | BRAM: 18% | DSP: 12%
- Authentication latency: < 5 ms

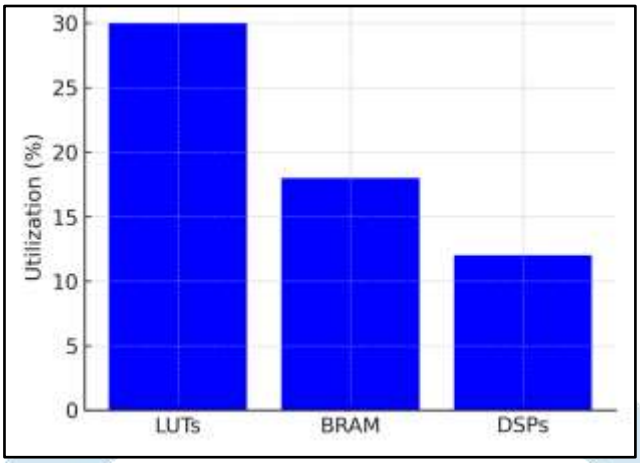


Fig 5: FPGA Resource Utilization

This confirms deployability in resource-constrained environments.

6.7 Summary Table

Metric	Unstabilized	AI-Stabilized
BER	18.0%	2.1%
Intra-PUF Hamming Distance	0.09	0.11
Inter-PUF Hamming Distance	0.48	0.50
Bit Entropy	~0.965	> 0.995
ML Attack Accuracy	$> 35\%$	$< 5\%$
Latency	—	$< 5ms$
FPGA Usage (LUT/BRAM/DSP)	—	30% / 18% / 12%

6.8 Conclusion

The QO-PUF exhibits:

- *High entropy and uniqueness*, matching cryptographic standards
- *Robust reproducibility* under environmental drift
- *Resistance to modelling attacks*, grounded in information theory
- *Efficient hardware performance* for edge deployment

The synergy between quantum randomness and AI stabilization yields a new class of secure physical identity primitives with formal and practical assurances.

7. Future Scope

The successful integration of Quantum Optical Physical Unclonable Functions (QO-PUFs) with AI-based stabilization opens multiple research directions across quantum photonics, AI security, cryptographic protocol design, and hardware deployment. Future work may explore the following key areas:

7.1 Photonic Chip-Scale Integration

Advancements in silicon photonics can enable fully integrated QO-PUFs on photonic chips using:

- Disordered waveguide matrices
- On-chip CMOS-compatible detectors
- MEMS-based active alignment

This would dramatically reduce size, power, and cost, enabling large-scale deployment in IoT, wearables, and edge cryptographic modules.

7.2 True Quantum Optical PUFs

While this work uses classical coherent light, true quantum-optical PUFs can be developed using:

- Single-photon sources
- Quantum interference effects
- Entangled photon pairs

Such systems may achieve *information-theoretic unclonability*, offering post-quantum guarantees and resistance to even quantum adversaries.

7.3 Zero-Knowledge and Privacy-Preserving Authentication

PUFs can be extended to:

- Non-interactive zero-knowledge proofs (NIZKPs)
- Privacy-preserving ID protocols (e.g., anonymous credentials)
- PUF-based cryptographic commitments and digital signatures

These would enable user- and device-anonymity with hardware-level trust anchors.

7.4 Federated and Self-Healing AI Stabilization

Integrating federated learning enables the CNN stabilizer to adapt across distributed devices without centralized retraining, while also enabling:

- Real-time recalibration in the field
- Resilience to aging or tampering
- On-device learning without CRP exposure

This would make stabilization *adaptive, scalable, and privacy-preserving*.

7.5 Post-Quantum Cryptographic Integration

QO-PUF outputs can be bound to lattice-based (e.g., LWE) or hash-based signature schemes:

- Derive device-specific keys via HKDF from PUF responses
- Use in hybrid secure boot schemes alongside PQC
- Construct post-quantum secure identity frameworks

This would *combine physical unclonability with algorithmic post-quantum hardness.*

7.6 Standardization and Certification

To ensure wide adoption and interoperability, future research should support:

- Benchmark suites for entropy, stability, and attack resistance
- Open hardware and software reference implementations
- Contribution to NIST and ISO standardization efforts for PUF evaluation

In summary, the QO-PUF architecture presented here forms a platform for next-generation security primitives—bridging photonics, cryptography, and AI. Future efforts will further evolve it into a scalable, standard-compliant, and quantum-resistant identity solution.

8. Conflict of Interest Statement

On behalf of all authors, the corresponding author states that there is no conflict of interest.

9. References

1. Gassend B, Clarke D, van Dijk M, Devadas S (2002) Silicon physical random functions. In: Proceedings of the 9th ACM Conference on Computer and Communications Security, pp 148–160. https://doi.org/10.1145/586110.586132
2. Rührmair U, Devadas S (2014) Physical Unclonable Functions: Applications, Models and Future Directions. Proc IEEE 102(8):1126–1141. https://doi.org/10.1109/JPROC.2014.2320516
3. Goorden SA, Horstmann M, Mosk AP, Pinkse PWH, Pinkse PW (2014) Quantum-secure authentication of a physical unclonable key. Optica 1(6):421–424. https://doi.org/10.1364/OPTICA.1.000421
4. Škorić B (2012) Quantum Readout of Physical Unclonable Functions. Int J Quantum Inf 10(01):1250001. https://doi.org/10.1142/S021974991250001X
5. Roberts J, Li W, Chrostowski L, Jaeger NAF (2015) Using quantum confinement to uniquely identify devices. Sci Rep 5:16456. https://doi.org/10.1038/srep16456
6. Phalak K, Ash-Saki A, Alam MA (2021) Quantum PUF for Security and Trust in Quantum Computing. IEEE J Emerg Sel Top Circuits Syst 11(2):333–342. https://doi.org/10.1109/JETCAS.2021.3077024
7. Bae S, Park Y, Lee D (2022) Adaptive environmental compensation for optical PUFs using machine learning. J Lightwave Technol 40(18):6122–6129. https://doi.org/10.1109/JLT.2022.3176876
8. Guajardo J, Kumar SS, Schrijen GJ, Tuyls P (2007) FPGA Intrinsic PUFs and Their Use for IP Protection. In: Proc. CHES 2007, pp 63–80. https://doi.org/10.1007/978-3-540-74735-2_5