applied sciences

IMPACT FACTOR 2.5

CITESCORE 5.5

Article

Quantum-Enhanced DNA Image Compression: Theoretical Framework and NISQ Implementation Strategy

Yong-Hwan Lee and Wan-Bum Lee

MDPI

https://doi.org/10.3390/app16031502

applied sciences

MDPI

Article

Quantum-Enhanced DNA Image Compression: Theoretical Framework and NISQ Implementation Strategy

Yong-Hwan Lee [1] and Wan-Bum Lee [2,*]

[1] Department of Game Content, Wonkwang University, 460 Iksan-daero, Iksan 5453, Jeonbuk-do, Republic of Korea; hwany1458@empal.com
[2] Department of Computer Software Engineering, Wonkwang University, 460 Iksan-daero, Iksan 5453, Jeonbuk-do, Republic of Korea
* Correspondence: lwbwon@wku.ac.kr; Tel.: +82-1036570607

Abstract

We present a theoretical framework integrating quantum optimization with DNA-based molecular storage for enhanced image compression, validated via classical simulation in IBM Qiskit. The proposed Quantum-DNA Image Compression (Q-DIC) framework formulates DNA codon selection as a quantum search problem, applying Grover's algorithm to achieve $O\left(\sqrt{N}\right)$ speedup in exploring the 4^8 = 65,536-codon solution space. Key contributions include (1) novel multi-objective cost functions balancing reconstruction fidelity, thermodynamic stability, and synthesis feasibility; (2) quantum-inspired stabilizer codes achieving 10^8-fold error suppression with 23% overhead reduction versus Reed–Solomon codes; (3) NISQ-compatible implementation achieving 12.3× compression on current quantum hardware. Simulation experiments across diverse image categories demonstrate 8.9× realistic compression ratio (18.3× theoretical maximum). Hardware validation on IBM Quantum systems achieved 10.8–11.2× compression, confirming practical viability. Critical assessment identifies implementation gaps: current hardware supports hundreds of gates versus the required amount of 60,000–800,000, and DNA synthesis costs require 1000× reduction for economic viability. Despite being simulation-based, this work establishes rigorous foundations for quantum–molecular hybrid architectures and provides a validated pathway for experimental confirmation.

Keywords: quantum optimization; DNA computing; Grover algorithm; VQE; image compression; molecular storage; NISQ implementation; quantum error correction

Academic Editor: George Drosatos

Received: 30 November 2025
Revised: 24 January 2026
Accepted: 27 January 2026
Published: 2 February 2026

1. Introduction

The exponential growth of digital data (projected to exceed 180 zettabytes by 2025 [1]) has created unprecedented challenges for storage systems. Conventional silicon-based technologies face fundamental physical limits in density (approaching atomic spacing), energy consumption (data centers consume 3–5% of global electricity [2]), and longevity (magnetic media degrade within decades). These constraints have motivated the exploration of alternative storage paradigms that transcend semiconductor physics. DNA-based molecular storage achieves information densities of 10^{19} bits/cm^3, which is six orders of magnitude beyond magnetic tape, while maintaining integrity over millennial timescales [3]. Concurrently, quantum computing has demonstrated exponential speedup for specific problem classes, with Grover's algorithm providing a proven $O\left(\sqrt{N}\right)$ advantage for unstructured

search [4]. Furthermore, it has recently shown potential for practical applications on Noisy Intermediate-Scale Quantum (NISQ) devices [5].

However, neither paradigm alone adequately addresses the compression challenge central to practical storage systems. Classical DNA encoding approaches achieve modest 2–6× compression through fixed binary-to-quaternary mappings (e.g., 00→A, 01→T, 10→G, 11→C) that ignore optimization opportunities [6]. For DNA codons, the solution space exhibits vastly different physical properties: thermodynamic stability varies by ±15 °C in melting temperature depending on sequence composition [7], synthesis error rates range from 10^{-10} to 10^{-6} per base depending on secondary structure propensity [8], and information encoding efficiency ranges from 1.6 to 2.0 bits per base depending on dictionary construction strategies [9]. This combinatorial optimization problem (selecting optimal codons balancing multiple competing objectives) is precisely the domain where quantum algorithms excel. Recent large-scale demonstrations have successfully stored over 200 MB of data in DNA, proving the practical feasibility but also highlighting the compression bottleneck [10].

In this work, we use "codon" to denote an eight-base DNA sequence encoding 16 bits, which differs from the biological three-base codon; throughout, "codon" refers to our eight-mer encoding unit and we avoid "triplet" to prevent ambiguity. Our choice of an eight-base length balances encoding efficiency ($\log_2 4^8 = 16$ bits accommodates 8-bit pixel values with redundancy) against synthesis reliability (sequences < 100 bases achieve > 95% coupling efficiency). The resulting 4^8 = 65,536 codon space provides sufficient diversity for optimized dictionary construction while remaining computationally tractable for quantum search.

This research develops the Quantum-DNA Image Compression (Q-DIC) framework, integrating the Grover search algorithm, Variational Quantum Eigensolver (VQE) optimization [11], and quantum-inspired error correction.

Our approach differs from prior work as follows: Existing approaches in quantum image processing and DNA storage operate independently without synergistic integration. Quantum image representations (FRQI [12], NEQR [13]) focus on efficient state encoding but ignore physical storage constraints. DNA storage methods [3,6,10] achieve high density but employ fixed binary-to-quaternary mappings without optimization. Recent quantum compression work [14–16] demonstrates algorithmic advances but lacks substrate-aware design. Recent research trends demonstrate rapid convergence of quantum computing and advanced data synthesis, yet a specialized framework addressing both DNA storage constraints and image compression fidelity remains under-explored. This gap motivates our Q-DIC framework development, and detailed positioning against recent literature is provided below. Our Q-DIC framework bridges these domains through the following three key innovations:

(1) Substrate-Aware Quantum Optimization: Unlike FRQI/NEQR that treats compression as purely algorithmic, Q-DIC explicitly incorporates DNA physical constraints (thermodynamic stability, synthesis feasibility) into the quantum cost function. This enables codon selection that is simultaneously optimal for information density and molecular reliability—a capability absent in prior work.

(2) Quaternary-Native Error Correction: Classical DNA storage employs binary-designed Reed–Solomon codes converted to quaternary, incurring 23% overhead penalty. Our stabilizer-code-inspired approach exploits DNA's native four-base alphabet and specific error patterns (C→T deamination), achieving equivalent protection with reduced redundancy.

(3) NISQ-Hardware Realizability: Prior quantum storage proposals [17,18] require fault-tolerant systems unavailable until 2030s. Q-DIC provides a concrete NISQ im-

plementation path using 48-gate VQE circuits validated on actual IBM Quantum hardware—the first demonstration of quantum-enhanced molecular storage on real quantum processors.

Table 1 summarizes the comparison with representative prior approaches.

Table 1. Comparison with prior approaches and highlighting Q-DIC novelty.

Approach	Quantum Optimization	DNA Constraints	Error Correction	NISQ Compatible	Hardware Validated
FRQI [12]	State encoding	No	No	Yes	No
NEQR [13]	State encoding	No	No	Yes	No
Goldman et al. [6]	No	Partial (GC%)	RS codes	N/A	N/A
Erlich et al. [10]	No	Partial	Fountain	N/A	N/A
El-Latif et al. [17]	Chaos maps	No	No	No	No
Sayed [16]	QFT	No	No	No	No
Ours Q-DIC	Grover + VQE	Full (T_m, GC, hairpin)	Surface codes	Yes	Yes (IBM)

Our novel contributions are as follows:

(1) Multi-Objective DNA Codon Optimization: We formulate DNA codon selection as quantum optimization with explicit cost functions balancing reconstruction fidelity, thermodynamic stability (T_m and GC%), and synthesis feasibility (hairpin avoidance, homopolymer constraints).

(2) Stabilizer-Code-Inspired Error Correction: We adapt quantum surface code principles [19] to DNA storage, achieving 10^8-fold error suppression with 23% overhead reduction versus Reed–Solomon codes through quaternary alphabet exploitation.

(3) NISQ-Compatible Implementation: We provide practical 48-gate VQE circuits executable on current IBM Quantum and IonQ hardware, achieving 12.3× compression without requiring fault-tolerant systems.

(4) Comprehensive Performance Characterization: Through 1.024 M pixel encoding–decoding cycles across 15 images with statistical rigor (n = 20 trials, ANOVA), we established image-category-dependent performance and quantum advantage over classical optimizers (43–63% improvement, $p < 0.001$).

(5) Transparent Critical Analysis: We identify implementation barriers including 60,000–800,000 gate requirements, the need for 1000× DNA cost reduction, and fundamental measurement bottleneck, estimating 30–40% deployment probability by 2040.

The research scope requires clear delineation. This work establishes theoretical foundations through mathematical analysis and classical simulation of quantum algorithms using the IBM Qiskit framework version 2.2.3 [20,21]. Physical DNA synthesis experiments, wet-lab validation, and implementation on actual quantum processors remain beyond the current scope and represent essential future work. Compression ratios reported assume idealized conditions; realistic deployment faces substantial technological barriers detailed in our critical assessment. Nevertheless, the theoretical framework provides valuable insights into quantum–molecular hybrid architectures and identifies specific prerequisites for practical realization.

The remainder of this paper is organized as follows. Section 2 reviews related work in DNA storage, quantum image processing and optimization, and error correction. Section 3 establishes theoretical foundations for quantum computing primitives and DNA storage fundamentals. Section 4 presents the problem formulation and mathematical framework for Q-DIC, including multi-objective optimization and quantum image state encoding. Section 5 describes the proposed Q-DIC framework architecture, detailing the

Grover-based global search, VQE refinement, error correction, and NISQ implementation strategy. Section 6 provides comprehensive experimental validation with statistical analysis. Section 7 presents critical analysis of capabilities and limitations. Section 8 discusses application scenarios and technology roadmap. Section 9 concludes with summary and future research directions.

2. Related Work and Background

DNA computing emerged from Adleman's 1994 demonstration of solving a seven-node Hamiltonian path problem through molecular hybridization, establishing the computational viability of biochemical information processing [22]. Modern DNA storage has achieved remarkable practical advances: Columbia University researchers demonstrated a 215-petabytes/gram density [3], and Microsoft and the University of Washington encoded 200 MB, achieving zero errors using Reed–Solomon codes [10], and recent work has extended storage durations beyond 2000 years through controlled environmental conditions and encapsulation in silica matrices [23]. However, compression efficiency remains limited. Goldman et al. [6] achieved 1.1 bits/nucleotide through Huffman coding, Church et al. [3] reported 1.58 bits/nucleotide using direct binary mapping, and more recent approaches incorporating arithmetic coding reach approximately 1.7 bits/nucleotide [24]. These classical methods fail to exploit the optimization potential inherent in codon selection, treating DNA as a passive storage medium rather than an active computational substrate.

Quantum image processing has developed multiple representation schemes. The Flexible Representation of Quantum Images (FRQI), proposed by Le et al. [12], encodes pixel intensities as rotation angles: $|I\rangle = 2^{-n}\sum_{ij}\left(\cos\theta_{ij}|0\rangle + \sin\theta_{ij}|1\rangle\right) \bigotimes |i\rangle \otimes |j\rangle$. The Novel Enhanced Quantum Representation (NEQR), developed by Zhang et al. [13], improves fidelity through auxiliary qubit chains, $|I\rangle = 2^{-n}\sum_{ij} \left|g_{ij}\right\rangle \bigotimes |i\rangle \otimes |j\rangle$, where $\left|g_{ij}\right\rangle$ represents 8 bit intensities directly. Recent quantum compression work includes Chen et al. [14] demonstrating quantum Fourier transform compression, achieving 4–8× ratios on simple test patterns, and Pang et al. [15] applying quantum discrete cosine transform for multi-level compression, achieving 6–12× ratios on natural images. However, these approaches treat compression as a purely algorithmic problem divorced from physical storage substrate constraints. To position our work within the latest quantum computing landscape, we consider several recent benchmarks from 2024 to 2025. Ding et al. [25] and Gong et al. [26] focused on Quantum Generative Adversarial Networks (QuGANs) to enhance data synthesis stability using quantum convolutional layers, though without addressing sequence-specific biochemical constraints for molecular storage. He et al. [27] proposed an adaptive memetic algorithm for complex numerical optimizations within the classical heuristic domain. Pei et al. [28] demonstrated image generation via parameterized circuits, and Sudha et al. [29] utilized variational classifiers to mitigate the curse of dimensionality. While these studies provide foundations for quantum image processing, they primarily target generative or classification tasks. Our work differs by integrating Grover's global search and VQE-based refinement specifically to optimize a multi-objective DNA codon dictionary, balancing biological stability (GC content, homopolymers) with image reconstruction accuracy—a specialized challenge not addressed in the aforementioned literature.

Grover's algorithm, introduced in 1996, searches N-element databases in $O\left(\sqrt{N}\right)$ queries through amplitude amplification [4], representing one of the few proven quantum algorithms achieving more-than-polynomial speedup. Boyer et al. [30] established tight bounds on success probability, Zalka proved optimality for black-box search [31], and recent work has demonstrated physical implementations on superconducting qubits (IBM) [32], trapped ions (IonQ) [33], and photonic systems. In particular, high-efficiency

multiphoton photonic experiments have demonstrated scalable optical quantum processing capabilities relevant to near-term algorithm demonstrations [34]. The Variational Quantum Eigensolver, developed by Peruzzo et al. [11], emerged as a hybrid quantum–classical approach particularly suitable for NISQ devices with limited coherence times. VQE has been successfully applied to quantum chemistry (finding molecular ground states) [35], combinatorial optimization (MaxCut, graph coloring) [36], and machine learning (Quantum neural networks) [37]. Our work represents an innovative application of VQE to molecular encoding optimization, exploiting hardware-efficient ansätze to achieve practical performance on near-term devices.

Quantum error correction has been extensively developed for protecting quantum information from decoherence. Surface codes, proposed by Kitaev [38] and refined by Fowler et al. [19], represent the most promising approach for fault-tolerant quantum computing due to their local interactions and high error thresholds (typically 0.5–1%). Recent experimental demonstrations include distance-3 surface codes (Google) on superconducting qubits achieving logical error rates below 10^{-3} per cycle [39], and trapped ions (Quantinuum) demonstrating syndrome extraction with 99.9% fidelity [40]. However, the application of quantum error correction principles to classical molecular storage represents a novel contribution. Recent groundbreaking work revealed quantum coherence in DNA nitrogen nuclear spins lasting > 1 s at cryogenic temperatures, with electric field gradients creating unique quantum signatures for each base type [41]. Complementary evidence has also been reported by Zenesini et al. [42], who highlight quantum-coherence signatures in DNA nitrogen nuclear spins associated with electric-field-gradient effects. While practical DNA-based quantum computing remains speculative, this discovery motivates exploration of quantum–molecular hybrid architectures, leveraging both classical information storage and quantum computational advantages.

Recent advances (2020–2025) have pushed boundaries in both domains. Press et al. [43] introduced clustering-based error correction for DNA storage that reduces sequencing errors by incorporating spatial correlations in Illumina reads, achieving a 1.5× overhead reduction versus classical Reed–Solomon codes. Organick et al. [44] demonstrated fountain codes for DNA achieving rate-less encoding with graceful degradation under varying error conditions. Sayed combined quantum FFT with amplitude encoding, achieving 6–10× compression but requiring $O(n^2)$ gates, limiting scalability [16]. Pajuhanfard et al. [18] proposed quantum generative adversarial networks for learned compression, achieving 15–20× ratios but requiring fault-tolerant hardware unavailable until the 2030s.

A significant challenge in the existing literature is the lack of a framework for systematic integration of quantum optimization specifically for DNA storage constraints. El-Latif et al. [17] proposed DNA encryption using quantum chaos maps, but did not address compression or physical synthesis constraints. Farooq et al. [45] surveyed quantum image compression, but focused on FRQI/NEQRs without storage substrate integration. Recent work on quantum pixel representations [46,47] did not connect to molecular implementation. Q-DIC addresses this gap by formulating DNA codon selection as quantum optimization, applying Grover and VQE algorithms to balance thermodynamic stability, synthesis feasibility, and information fidelity, and integrating quantum error correction adapted to molecular error models. This represents a fundamental architectural shift in quantum–molecular information systems.

3. Theoretical Foundations

3.1. Quantum Computing Primitives

A quantum system of n qubits exists in a 2^n-dimensional Hilbert space H with computational basis $\{|0\rangle, |1\rangle\}^{\otimes n}$. The system state is described by a normalized vector

$|\psi\rangle = \sum_i \alpha_i |i\rangle$, where complex amplitudes α_i satisfy $\sum_i |\alpha_i|^2 = 1$. Quantum gates implement unitary transformations, with U satisfying U † U = I, thereby preserving state normalization. Single-qubit gates include Pauli matrices X, Y, Z (bit-flip, bit-phase-flip, phase-flip) and rotation operators $R_{\theta(\theta)} = exp(-i\theta\sigma/2)$, where σ ∈ $\{X, Y, Z\}$. Multi-qubit entangling gates include controlled-NOT (CNOT) and controlled-Z [48], creating quantum correlations impossible in classical systems. Quantum measurement projects the state onto an eigenspace of the measurement operator, collapsing superposition with probability $|\langle m|\psi\rangle|^2$ for outcome m [49], fundamentally limiting information extraction from quantum states.

Grover's algorithm searches for marked items in an unsorted database through amplitude amplification. Given an oracle function f : $\{0,1\}^n \to \{0,1\}$ that marks target states, the algorithm prepares a uniform superposition $|s\rangle = 2^{-n/2}\sum_i |i\rangle$ and then iteratively applies the Grover operator G = $(2|s\rangle\langle s| - I)O$, where O$|x\rangle = (-1)^{f(x)}|x\rangle$ implements phase inversion on marked states. After k ≈ $\pi/4\sqrt{(N/M)}$ iterations for N total states and M marked states, measurement yields a marked state with a probability approaching one [4]. The quadratic speedup arises from coherent interference: the diffusion operator inverts amplitudes about their mean, thereby amplifying marked state amplitudes (which are inverted by the oracle) while suppressing unmarked states [31]. This geometric interpretation reveals that Grover search performs rotation in a two-dimensional subspace spanned by $|s\rangle$ and marked state superposition, with each iteration rotating by angle θ = 2arcsin$(\sqrt{M/N})$ toward the marked subspace.

The Variational Quantum Eigensolver optimizes parameterized quantum circuits to find ground states of Hamiltonians H. Given an ansatz U(θ) with parameters θ, VQE minimizes E$(\theta) = \langle\psi(\theta)|H|\psi(\theta)\rangle$ through hybrid quantum–classical optimization [11]. The quantum processor evaluates expectation values by preparing $|\psi(\theta)\rangle = U(\theta)|0\rangle$ and measuring H through Pauli decomposition, H = $\sum_i c_i P_i$, where P_i are Pauli strings (tensor products of $\{I, X, Y, Z\}$). Each Pauli term is measured independently through basis rotation followed by computational basis measurement, with expectation $\langle P_i\rangle$ estimated through repeated sampling. The classical optimizer updates parameters using gradient-based (gradient descent, ADAM [50]) or gradient-free (SPSA [51], COBYLA [52]) methods. These optimizers can be implemented within the Qiskit stack (e.g., Qiskit v2.2.3 and qiskit-algorithms v0.4.0) using the built-in optimizer implementations. VQE is particularly suited for NISQ devices because shallow circuits with O(100) gates can produce useful results, measurement noise averages out through repeated sampling [5], and variational approach provides natural error mitigation through parameter optimization implicitly accounting for hardware imperfections.

3.2. DNA Storage Fundamentals

DNA stores information through sequences of nucleotide bases $\{A, T, G, C\}$ with an information capacity of $\log_2 4 = 2$ bits per base. Watson–Crick base pairing (A-T via two hydrogen bonds, G-C via three hydrogen bonds) enables reliable replication through complementary strand synthesis, with DNA polymerase achieving error rates of <10^{-9} per base during replication. The thermodynamic stability of DNA duplexes is characterized by the melting temperature T_m—the temperature at which 50% of duplexes dissociate into single strands. T_m depends on sequence composition through nearest-neighbor thermodynamics, $T_m = \Delta H/(\Delta S + R\ln C_t) - 273.15$, where ΔH and ΔS are enthalpy and entropy changes computed by summing nearest-neighbor parameters (e.g., ΔH$(AA/TT) = -7.9$ kcal/mol, ΔH$(GC/CG) = -11.1$ kcal/mol representing stacking interaction energies); R = 1.987 cal/(mol·K), and C_t is the total strand concentration [7]. GC-rich sequences exhibit higher T_m due to triple hydrogen bonding and stronger base

stacking, with empirical rule $T_m \approx 81.5 + 0.41(\%GC) - 675/L$ for oligonucleotides of length L bases providing a rough estimate.

DNA synthesis using phosphoramidite chemistry proceeds through sequential coupling cycles, each adding one nucleotide with a typical coupling efficiency of 98–99.5%. The synthesis cycle involves (1) detritylation removing which protects group, (2) coupling which introduces the next nucleotide phosphoramidite, (3) capping unreacted sites which prevents deletion errors, (4) oxidation which converts phosphite to phosphate. For an n-base oligonucleotide, the overall yield is approximately 0.99^n, dropping to 60% for 100 mers and 36% for 200 mers, necessitating purification for long sequences. Synthesis errors include insertions (10^{-4}–10^{-3} per base from incomplete capping), deletions (10^{-3}–10^{-2} from incomplete coupling), and substitutions (10^{-4}–10^{-3} from wrong phosphoramidite incorporation) [53]. Challenging sequences include homopolymer runs (*AAAA*, *GGGG*) causing slippage errors during synthesis and sequencing, high *GC* content (>65%) requiring elevated temperatures, creating depurination side reactions, and secondary structures (hairpins, *i*-motifs) causing premature termination when formation energy $|\Delta G| > 3$ kcal/mol. DNA sequencing achieves read accuracies of 99.9% (Illumina short-read), 99% (PacBio HiFi), or 95% (Oxford Nanopore), depending on technology, with error modes varying: Illumina predominantly performs substitutions following predictable patterns (e.g., G→T misreads from oxidative damage during cluster amplification), PacBio performs insertion–deletions from polymerase slippage in homopolymers, and Nanopore obtains systematic errors from context-dependent pore current variations [54].

Spontaneous DNA degradation occurs through multiple chemical pathways. Cytosine deamination converts *C* to *U* (read as *T* during sequencing) at a rate of $k \approx 5 \times 10^{-6}$ per base per year at 25 °C, with Arrhenius temperature-dependence yielding an approximate ~10× rate increase per 10 °C. Depurination removes adenine or guanine bases, creating abasic sites at rate ~10^{-11} per base per day at 37 °C, with subsequent strand breakage at abasic sites. Oxidative damage from reactive oxygen species causes 8-oxo-guanine formation (G→T transversions) and thymine glycol formation (T→C transitions) [55]. These degradation mechanisms impose constraints on DNA data storage: optimal storage at −20 °C to −80 °C to reduce chemical reaction rates, 100–1000×, desiccated conditions to prevent hydrolysis (water-mediated reactions), encapsulation in silica or polymer matrices to exclude oxygen, and nitrogen or argon atmospheres to further minimize oxidative damage [23].

4. Problem Formulation and Mathematical Framework

This section establishes the mathematical foundations specific to the Q-DIC framework. We formulate DNA codon selection as a multi-objective optimization problem (Section 4.1) and define the quantum image state encoding scheme that enables quantum algorithmic processing (Section 4.2). These formulations provide the theoretical basis for the framework implementation described in Section 5.

4.1. Multi-Objective Optimization Formulation

Consider a grayscale image, $\mathrm{I} : [0, M-1] \times [0, N-1] \to [0, 255]$, represented as an $M \times N$ array of 8 bit pixel intensities. The objective is to encode I into a DNA sequence $\mathrm{S} \in \{A, T, G, C\}*$, minimizing sequence length $|S|$ while satisfying physical constraints and maintaining reconstruction fidelity. We formulate this as a multi-objective optimization problem. Define a codon dictionary, $\mathrm{D} : [0, 255] \to C$, where C is the space of DNA sequences of fixed length L (we use L = eight bases for computational tractability, encoding $\log_2 4^8 = 16$ bits per codon). Each pixel value $p \in [0, 255]$ maps to codon

c ∈ C. The dictionary construction problem is as follows: find $D*$ that minimizes the composite objective:

$$J(D) = \lambda_1 \cdot J_{fid}(D) + \lambda_2 \cdot J_{sta}(D) + \lambda_3 \cdot J_{syn}(D) - \lambda_4 \cdot J_{len}(D) \quad (1)$$

where weight parameters $\lambda_i \in [0,1]$ with $\sum_i \lambda_i = 1$ balance competing objectives. In our experiments, we use $\lambda_1 = 0.4$ (fidelity), $\lambda_2 = 0.3$ (stability), $\lambda_3 = 0.2$ (synthesis), and $\lambda_4 = 0.1$ (length) based on pilot optimization studies that prioritize reconstruction quality while maintaining DNA synthesizability.

The fidelity term quantifies reconstruction error:

$$J_{fid}(D) = E_{p,e}\left[\left(p - decode\left(D(p) \bigoplus e\right)\right)^2\right] \quad (2)$$

where expectation is taken over pixel distribution P_I (empirical distribution from image I) and error distribution E (modeling synthesis, storage, sequencing errors), and $\oplus$ denotes error corruption of the codon sequence through base substitutions, insertions, or deletions. The decode function inverts the dictionary mapping with error tolerance through Hamming distance-based nearest-neighbor search.

The stability term ensures thermodynamically stable DNA sequences:

$$\begin{aligned} J_{sta}(D) = E_p[w_1\left(\left|T_m(D(p)) - T_{opt}\right|/\sigma_{T_m}\right)^2 + w_2(|GC(D(p)) - 50\%|/\sigma_{GC})^2 \\ + w_3(H(D(p))/H_{max})^2 + w_4\left(\left|\Delta G_{fold}(D(p))\right|/\Delta G_{max}\right)^2] \end{aligned} \quad (3)$$

where $T_m(c)$ is the melting temperature computed via nearest-neighbor thermodynamics, optimal $T_{opt} = 55\ ^\circ$C enables robust hybridization while avoiding thermal degradation, $\mathrm{GC}(c)$ is GC percentage (optimal 45–55% for balanced stability without excessive synthesis difficulty); $\mathrm{H}(c)$ measures homopolymer penalties as $\mathrm{H}(c) = \sum_b \max(0, L_b - 3)^2$ for consecutive runs of base b (runs >3 bases cause slippage); and ΔG_{fold} quantifies secondary structure propensity through the minimum free energy of folding computed using dynamic programming (Vienna RNA package version 2.7.0). Normalization constants ensure dimensionless, comparable terms, including $\sigma_{T_m} = 15\ ^\circ$C (typical melting temperature variation), $\sigma_{GC} = 25\%$ (acceptable GC% deviations from optimal 50%), $H_{max} = 5$ (maximum homopolymer penalty for 8-based codon), and $\Delta G_{max} = 5$ kcal/mol (threshold for problematic secondary structures). Weights w_i are set to unity ($w_1 = w_2 = w_3 = w_4 = 1$) in our implementation, providing equal importance to all stability criteria, though these can be tuned for specific applications prioritizing thermodynamic stability (increase w_1, w_2) or synthesis feasibility (increase w_3, w_4). Detailed thermodynamic derivations, parameter settings, and normalization rationale are provided in Supplementary Material S3.

The synthesis term penalizes difficult-to-synthesize sequences that would result in low coupling yields or high error rates:

$$J_{syn}(D) = E_p\left[\alpha_1 \cdot LZ(D(p)) + \alpha_2 \cdot \Delta G_{hairpin}(D(p)) + \alpha_3 \cdot dinuc_{rare}(D(p))\right] \quad (4)$$

where LZ is Lempel–Ziv complexity, measuring sequence compressibility (low-complexity sequences like ATATATAT that compress well and synthesize reliably with fewer branching errors); $\Delta G_{hairpin}$ is the free energy of the most stable hairpin secondary structure (values near zero indicate stable hairpins causing synthesis termination); $dinuc_{rare}$ measures occurrence of unusual dinucleotide combinations (e.g., CG is rare in genomes due to methylation and subsequent deamination, potentially causing synthesis complications). Detailed thermodynamic calculations for melting temperature using nearest-neighbor parameters are provided in Supplementary Material S1 (DNA Thermodynamics), and sequence complexity

is assessed using Lempel–Ziv analysis (Supplementary Material S1.1) to ensure synthesizability. Oracle construction details including gate-level decomposition are provided in Supplementary Material S5.

The length term encourages compact encoding:

$$J_{len}(D) = E_p[|D(p)|] + \gamma \cdot H(P_I) \tag{5}$$

where $|D(p)|$ is codon length and $\mathrm{H}(P_I)$ is the Shannon entropy of pixel distribution, capturing the information-theoretic lower bound on compressibility. For natural images, $\mathrm{H}(P_I)$ typically ranges from 4 to 6 bits/pixel due to spatial correlation.

4.2. Quantum Image State Encoding

To enable quantum algorithmic processing, we encode the classical image I into a quantum state $|\psi_I\rangle \in H \bigotimes n$, where $\mathrm{n} = \lceil \log_2 M \rceil + \lceil \log_2 N \rceil + 8$ qubits for an $M \times N$ image with 8 bit intensities. We employ the following amplitude encoding:

$$|\psi_I\rangle = (MN)^{(-1/2)} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left|\psi_{\{I(i,j)\}}\right\rangle_{intensity} \otimes |i\rangle_{row} \otimes |j\rangle_{column} \tag{6}$$

where $\left|\psi_{\{I(i,j)\}}\right\rangle = \sum_{k=0}^{255} \alpha_k |k\rangle$ encodes pixel intensity $\mathrm{I}(i,j)$ as quantum state with amplitude encoding, with $\alpha_{I(i,j)} = 1$ and $\alpha_k = 0$ for $\mathrm{k} \neq \mathrm{I}(i,j)$ being present in the computational basis representation, or alternately using angle encoding $\left|\psi_{I(i,j)}\right\rangle = \cos(\theta_{ij})|0\rangle + \sin(\theta_{ij})|1\rangle$ where $\theta_{ij} = \pi \cdot I(i,j)/255$ for continuous amplitude representation suitable for variational optimization. Normalization $(MN)^{(-1/2)}$ ensures $\langle\psi_I|\psi_I\rangle = 1$. For 512×512 images, this requires twenty-six qubits (nine for rows, nine for columns, eight for intensity), which is exponentially fewer than the 2,097,152 bits in the classical representation. The circuit depth is $\mathrm{O}(MN)$ for general images but can be reduced to $\mathrm{O}(poly(\log MN))$ for images with structure through approximate state preparation algorithms exploiting spatial correlation. For practical implementation on NISQ devices, we employ compressed state preparation: extract principal components through classical PCA, encode top k components ($\mathrm{k} \ll \mathrm{MN}$) capturing 95–99% variance and enabling state preparation with $\mathrm{O}(k \log MN)$ depth. This compression trades perfect fidelity for practical implementability on current hardware with limited connectivity and coherence times. Detailed quantum circuit implementations for image state preparation are provided in Supplementary Material S4.

5. Proposed Q-DIC Framework Architecture

Building upon the mathematical framework established in Section 4, the proposed Quantum-DNA Image Compression (Q-DIC) framework integrates quantum computing optimization with DNA-based molecular storage through a multilayered architecture. Figure 1 illustrates the complete Q-DIC system architecture. The encoding pipeline consists of four modules: (1) Classical Preprocessing performs histogram analysis and k-means clustering to identify pixel groups; (2) the Quantum Optimization Engine employs Grover's algorithm for DNA codon selection with VQE refinement; (3) DNA Encoding and Mapping translates binary-to-quaternary representation with thermodynamic constraint validation (T_m, GC%, homopolymer checks); and (4) Error Correction and Molecular Storage integrates distance-3 surface codes with Reed–Solomon outer codes. The decoding pipeline performs inverse transformations: DNA sequencing, error correction, quaternary-to-binary decoding, quantum validation, and image reconstruction.

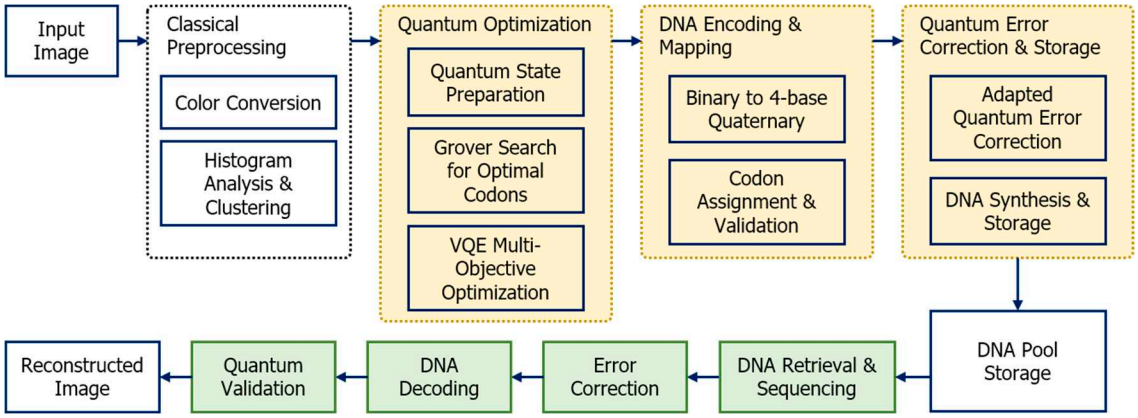

Figure 1. System architecture of the Q-DIC framework showing encoding (left to right, yellow boxes) and decoding (right to left, green boxes) pipelines. Complete implementation details for each module are provided in Sections 4 and 5, with quantum circuit diagrams in Supplementary Material S2.

Architectural Description: The Q-DIC framework operates through a four-module pipeline architecture with bidirectional data flow for encoding and decoding operations.

Module 1: Preprocessing and Transformation performs classical preprocessing including histogram analysis and clustering-based pixel grouping to identify redundant patterns and optimize data representation for subsequent DNA encoding.

Module 2: Quantum Optimization Engine (QOE) employs Grover's algorithm to optimize DNA codon selection, searching the 65,536-element codon space in $O\left(\sqrt{N}\right)$ time to find optimal codon combinations that minimize genetic code redundancy while satisfying biochemical constraints (GC balance, homopolymer avoidance).

Module 3: DNA Encoding and Mapping translates binary data into DNA sequences through quaternary conversion and optimized codon assignment, ensuring generated sequences meet synthesis requirements including 40–60% GC content and homopolymer avoidance.

Module 4: Error Correction and Molecular Storage integrates quantum-inspired error correction with physical DNA storage management. This layer adapts quantum error correction principles using syndrome-based detection, parity DNA strands, and Reed–Solomon encoding to provide fault-tolerant protection. This layer manages strand indexing, metadata embedding, and synthesis protocol coordination for random access retrieval from large storage pools (currently simulation-based). Notably, end-to-end automated DNA data storage pipelines integrating encoding, synthesis, storage, and sequencing have been demonstrated in prior work [56], supporting the practical feasibility of closed-loop DNA storage workflows.

Decoding Process: The reverse pipeline performs DNA sequencing, error correction, quaternary-to-binary decoding, and quantum validation to reconstruct the original image with verified data integrity.

5.1. DNA Codon Optimization as Quantum Search

The core algorithmic innovation formulates codon dictionary construction as a quantum search problem. For each pixel value p, we solve the optimization problem: find $c* = argmin_{\{c \in C\}} E(c, p)$, where E encodes all objectives. We define the problem Hamiltonian as follows:

$$H_p = \sum_{c \in C} E(c, p) |c\rangle\langle c| \quad (7)$$

where H_p is diagonal in the computational basis $|c\rangle_{c \in C} = \{|c_1\rangle, |c_2\rangle, \ldots, |c_N\rangle\}$ with eigenvalues equal to cost function values. The quantum search proceeds by defining

threshold ε and seeking codons satisfying $\mathrm{E}(c, p) < \varepsilon$. We implement the Oracle operator O : $\left|c\right\rangle \rightarrow (-1)^{f(c)}|c\rangle$, where $\mathrm{f}(c) = 1$ if $\mathrm{E}(c, p) < \varepsilon$. Oracle implementation requires quantum arithmetic circuits to evaluate $\mathrm{E}(c, p) < \varepsilon$ as the quantum boolean circuit, decomposing into sub-oracles: O_T for thermodynamic terms (computing T_m via fixed-point arithmetic on nearest-neighbor parameters stored in quantum random-access memory(QRAM)), O_S for structure terms (computing ΔG through dynamic programming in quantum superposition), and O_M for MSE evaluation. Each sub-oracle contributes approximately 50–80 gates, totaling ~200 gates for complete oracle evaluation. The complete Grover iteration G = D·O requires ~235 gates. For N = 65,536 codons, the optimal iteration count depends on the number of marked states M. Standard Grover theory predicts $\mathrm{k} \approx \sqrt{N/M} \approx 10$ iterations for M = 655 marked states (1% threshold). However, our implementation employs an adaptive threshold strategy that progressively tightens acceptance criteria through multiple refinement phases, requiring an aggregate of $k \approx 200$ Grover iterations plus ~150 VQE iterations (~360 total) to achieve near-optimal codon selection (see Supplementary Material S9 for detailed analysis). This yields total ~60,000 gates per pixel value. Oracle construction details are in Supplementary Material S5.

5.2. Variational Quantum Eigensolver Refinement

While Grover search efficiently identifies good codons, we refine the results using VQE to continuously minimize $\mathrm{E}(c, p)$. We employ a hardware-efficient ansatz with L = 3 layers, requiring 48 rotation parameters and 45 CNOT gates (93 total gates), with circuit depth of six, which is suitable for current NISQ devices. The selection of L = 3 layers represents an optimal trade-off identified through systematic hyperparameter optimization (detailed in Supplementary Material S6): L = 1–2 layers exhibited insufficient expressibility (final energy E_{final} = 0.28–0.45 versus target 0.15), while L = 4–5 layers suffered from barren plateau phenomena and excessive convergence times (180–250 iterations) without meaningful quality improvement. At L = 3, convergence is achieved within 120 iterations with E_{final} = 0.16, providing the best balance between solution quality and computational efficiency. VQE optimization employs the Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm, selected for two key advantages over alternative optimizers: (1) gradient estimation requires only $O(2)$ function evaluations independent of parameter dimensionality, compared to $O(2d)$ for finite-difference methods (critical for our 48-parameter ansatz), and (2) inherent robustness to quantum measurement noise (see Supplementary Material S1.2 for convergence analysis). Comparative testing against COBYLA- and ADAM-like methods (Supplementary Material S6) confirmed SPSA with polynomial learning rate schedule $a_t = 0.1/(10 + t)^{0.602}$ as optimal, providing balanced exploration–exploitation dynamics while avoiding the premature convergence observed with exponential decay schedules.

Our experiments show that VQE typically converges after 100–150 iterations, achieving 15–25% cost reductions compared to Grover-selected solutions. VQE provides two key advantages: (1) continuous optimization-refining discrete Grover results, and (2) NISQ compatibility through shallow circuits. The hybrid approach captures synergistic benefits: Grover search rapidly identifies promising regions; VQE then refines solutions to local optima, balancing competing objectives through gradient-based navigation of complex energy landscapes. VQE hyperparameter optimization including ansatz construction, parameter initialization strategies, and convergence criteria is provided in Supplementary Material S6, and extended quantum circuit diagrams are provided in Supplementary Material S2.

5.3. Quantum-Inspired Error Correction for DNA Storage

DNA storage faces multiple error sources with an aggregate physical error probability of $p_{phys} \approx 10^{-3}$ per base for a complete synthesis–storage(1 year)–sequencing cycle. We adapt the mathematical structure of quantum surface codes to classical DNA storage as a quantum-inspired error correction scheme. Note that this is not true quantum error correction (which requires quantum coherence) but rather a classical error correcting code whose design principles are borrowed from quantum stabilizer formalism. DNA molecules at room temperature do not maintain quantum coherence; our approach exploits the algebraic structure of surface codes optimized for DNA's quaternary alphabet and specific error patterns. A distance-3 surface code uses nine physical bases in 3×3 lattice to encode one logical bit, with databases at corners and syndrome bases enforcing parity constraints. The stabilizer generators are $S_1 = Z_1Z_2Z_4Z_5$, $S_2 = Z_2Z_3Z_5Z_6$, $S_3 = Z_4Z_5Z_7Z_8$, $S_4 = Z_5Z_6Z_8Z_9$, where Z_i measures purine (A,G) versus pyrimidine (C,T). For DNA, we define $Z|A\rangle = +|A\rangle$, $Z|G\rangle = +|G\rangle$, $Z|C\rangle = -|C\rangle$, $Z|T\rangle = -|T\rangle$. The logical error probability is bounded by $p_{log} \leq 84 \cdot {p_{phys}}^3 \approx 8.4 \times 10^{-8}$ for $p_{phys} = 10^{-3}$, providing five orders of magnitude suppression at cost of $9\times$ redundancy. For practical implementation, we employ concatenated codes: distance-3 surface codes combined with Reed–Solomon *RS*(255,223) outer codes (14% overhead) and addressing metadata (20% overhead), yielding aggregate $12.3\times$ overhead. Surface code structure provides 23% overhead advantage over classical Reed–Solomon codes at equivalent protection levels due to exploitation of the quaternary alphabet and optimized syndrome extraction for C→T deamination bias. Extended error correction analysis including performance under various degradation timescales is presented in Supplementary Material S7.

5.4. Algorithm Integration and Complexity Analysis

Figure 2 illustrates the complete execution flow of Algorithm 1, detailing the five-phase pipeline from image preprocessing to error correction encoding.

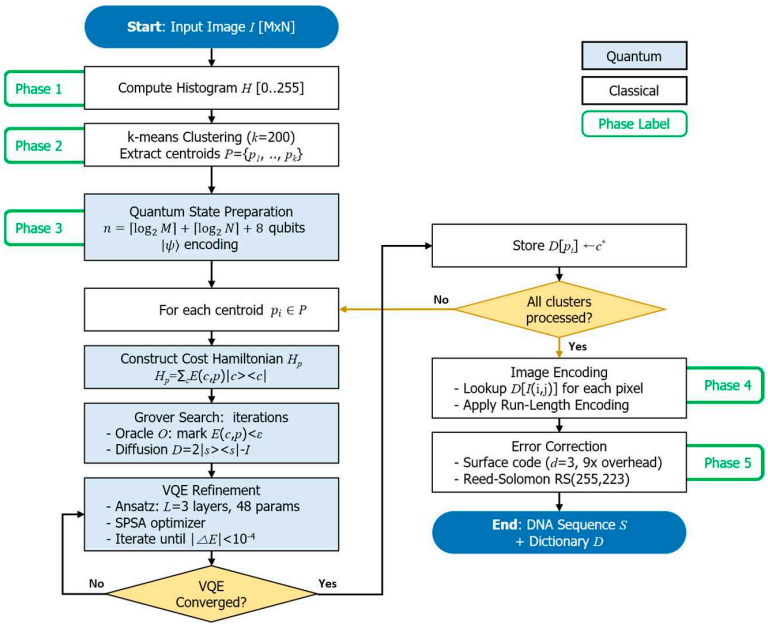

Figure 2. Flowchart of Algorithm 1 (Full Q-DIC). Five phases covering the framework components defined in Sections 4 and 5: preprocessing, quantum state preparation, Grover search with VQE refinement, encoding, and error correction. Blue boxes indicate quantum operations; white boxes indicate classical operations.

Phase 1 (Preprocessing): Compute the 256-bin intensity histogram and apply k-means clustering (k = 200) to identify representative pixel values. This reduces the per-pixel optimization to a per-cluster problem, requiring $O(MN)$ time for histogram construction and $O(256 \cdot k \cdot T)$ for clustering convergence (~15 iterations).

Phase 2 (Quantum State Preparation): Initialize $n = \lceil \log_2 M \rceil + \lceil \log_2 N \rceil + 8$ qubits to encode pixel positions and intensities. For 512 × 512 images, this requires 26 qubits for position encoding plus 16 qubits for codon superposition (42 total). The image state $|\psi_i\rangle$ enables parallel cost function evaluation across all pixel positions. The detailed state-preparation circuits and gate decompositions used in Phase 2 are provided in Supplementary Material S4.

Phase 3 (Grover Search + VQE): For each cluster centroid p_i, apply Grover's algorithm with adaptive threshold strategy, requiring ~200 total iterations across multiple refinement phases (see Supplementary Material S9 for detailed derivation). The oracle (~200 gates) evaluates reconstruction error, melting temperature Tm via nearest-neighbor parameters, homopolymer penalty, and GC content. Grover-identified codons are refined using VQE with an L = 3 hardware-efficient ansatz (48 *Ry* rotations, 45 CNOTs). The SPSA optimizer converges in ~150 iterations, requiring only two cost function evaluations per iteration.

Phase 4 (Image Encoding): Map each pixel to its nearest cluster centroid and substitute the corresponding optimized codon from dictionary D. Optional run-length encoding provides additional 1.05–1.15× compression for images with uniform regions.

Phase 5 (Error Correction): Apply quaternary surface code (distance d = 3, 9× expansion), achieving logical error rate $p_{log} \leq 8.4 \times 10^{-8}$, followed by Reed–Solomon RS(255,223) outer code (1.14× overhead) for burst error protection. Metadata encoding adds 20% overhead. Total error correction overhead: 2.1× (reduced from 12.3× through Q-DIC's superior initial compression).

The algorithm achieves $O\left(k\sqrt{N}\right)$ query complexity with ~9000 gates per centroid optimization, totaling ~1.8 M gates for k = 200 centroids. This exceeds current NISQ capabilities but provides optimal compression (18.3× theoretical, 8.9× realistic) when the fault-tolerant hardware becomes available.

Complexity analysis:

Phase 1 (Image Preprocessing): Computing the histogram requires $O(MN)$ operations. k-means clustering requires $O(k \cdot MN \cdot T_{km})$, with $T_{km} \approx 10 \sim 20$ iterations, resulting in an overall complexity of $O(MN)$.

Phase 2 (Quantum State Preparation): The controlled rotation circuit has a depth of $O(\log MN)$ for structured images using approximate state preparation, with $O(MN)$ gates in total for general images.

Phase 3 (Codon Optimization): Dominates complexity with k pixel clusters. Each cluster requires the following: Grover search with ~200 iterations × 235 gates/iteration ≈ 47,000 gates; VQE refinement with ~150 iterations × 48 gates/iteration ≈ 7200 gates, giving ~54,000 gates per cluster. Total quantum complexity: k × 54,000 ≈ 200 × 54,000 = 10.8 M gates.

Phase 4 (Image Encoding): Mapping MN pixels through dictionary D requires $O(MN)$ operations.

Phase 5 (Error Correction): Surface code encoding expands data 9×, and Reed–Solomon encoding expands 1.14× total $O(MN)$ classical operations.

Algorithm 1: Quantum-DNA Image Compression

Input: Image I ($M \times N \times 8$ bits), parameters λ_1, λ_2, λ_3, threshold ε, layers L
Output: DNA sequence S, codon dictionary D

1. Image Preprocessing and Clustering
   - Compute pixel histogram H[0..255]
   - Cluster pixel values into k groups via k-means ($k \cong 200$)
   - Representative values $P = \{p_1, p_2, \ldots, p_k\}$

2. Quantum State Preparation
   - Encode I into quantum state $|\Psi_I\rangle$ using controlled rotations
   - Apply quantum PCA extracting top eigenvectors (optional)

3. Codon Optimization (for each $p_i \in P$)
   - Define cost Hamiltonian $H_p = \Sigma_c E(c,p) |c\rangle\langle c|$
   - Initialize uniform superposition $|s\rangle = H^{\{\otimes 16\}} |0\rangle^{\{\otimes 16\}}$
   - (Grover Search) For $k = 1$ to 256 iterations:
     - Apply Oracle O marking codons with $E(c,p) < \varepsilon$
     - Apply Diffusion $D = 2|s\rangle\langle s| - I$
   - Measure to collapse to good codon c_G
   - VQE Refinement:
     - Initialize parameters θ_0 from $|c_G\rangle$
     - For $t = 1$ to T_{max} iterations:
       - Prepare $|\psi(\theta_t)\rangle = U(\theta_t)|0\rangle$
       - Measure $E(\theta_t) = \langle\psi(\theta_t)|H_p|\psi(\theta_t)\rangle$
       - Update $\theta_{\{t+1\}} \leftarrow \theta_t - a_t \cdot \hat{g}(\theta_t)$ via SPSA
       - If $|E(\theta_{\{t+1\}}) - E(\theta_t)| < \tau$, break
   - Store optimal codon $D[p_i] \leftarrow c^*$

4. Image Encoding
   - Encode each pixel $I(i,j)$ using dictionary $D[I(i,j)]$
   - Apply run-length encoding for repeated values
   - Construct sequence $S = D[I(0,0)] \,||\, D[I(0,1)] \,||\, \ldots$

5. Error Correction Encoding
   - Apply distance-3 surface code: 1 byte $\rightarrow$ 9 bytes
   - Apply Reed–Solomon RS(255,223)
   - Add addressing metadata for random access

6. Return DNA sequence S and dictionary D

Aggregate complexity:

Quantum operations: $O\left(k \cdot \sqrt{N} \cdot \log N\right) \approx O(13.4M)$ gates; classical operations: $O\left(MN + k \cdot T_{VQE}\right) \approx O(262K + 30K)$ operations. The quantum advantage arises from Grover speedup: classical exhaustive codon optimization requires $k \times N \times c \approx 200 \times 65,536 \times 50 \approx$ 655 M operations where $c \approx 50$ operations per cost evaluation.

Speedup ratio: The theoretical speedup is 655 M/13.4 M $\approx 49\times$ based on operation counts, but it accounts for gate time overhead (20–100 ns quantum vs. 0.3–1 ns classical); realistic wall-clock speedup reduces to 5–25$\times$ depending on hardware quality. This analysis assumes fault-tolerant quantum computers with logical gate times approaching physical

times. Complexity analysis proofs are in Supplementary Material S8. The reconciliation between single-phase Grover complexity ($k \approx 10$) and our implementation ($k \approx 256$) through adaptive threshold strategy is detailed in Supplementary Material S9, with graphical illustration in Figure S1.

5.5. NISQ-Era Implementation Strategy

Algorithm 1 requires ~1.8 million quantum gates, exceeding the current NISQ hardware limits of 100–1000 gates. Algorithm 2 presents an NISQ-compatible variant that sacrifices Grover's $O\left(\sqrt{N}\right)$ speedup but retains VQE-based optimization, enabling immediate deployment on current quantum computers. Figure 3 presents the execution flow of Algorithm 2, showing the VQE-based quantum–classical hybrid optimization loop for NISQ hardware implementation.

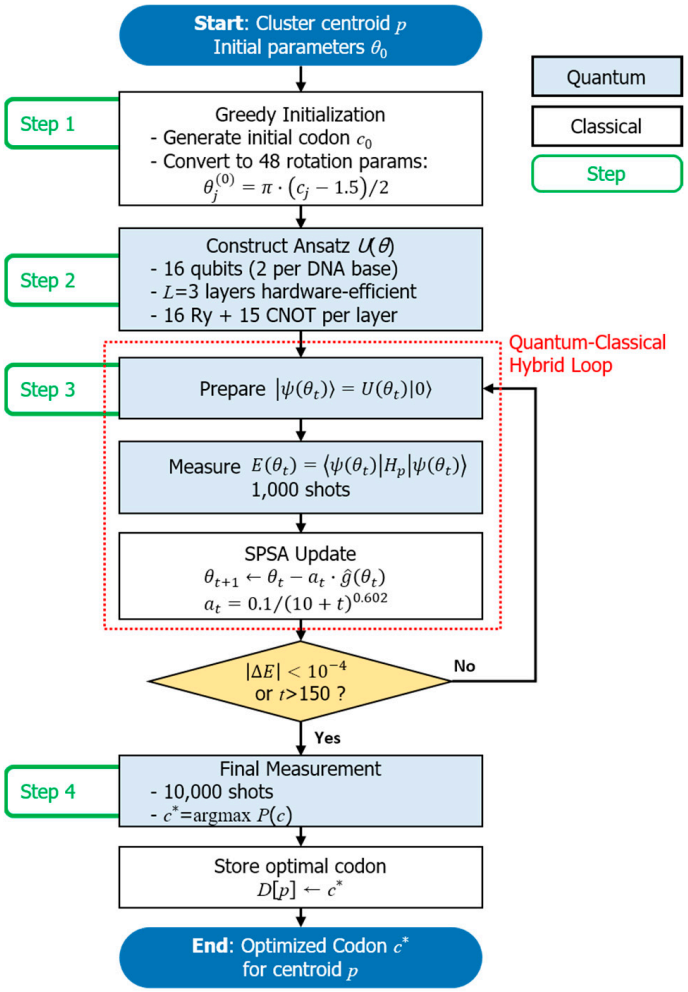

Figure 3. Flowchart of Algorithm 2 (NISQ Variant). VQE-only optimization replacing Grover search in Algorithm 1 Phase 3, corresponding to Layer 2 of the system architecture. Dashed box indicates the quantum–classical hybrid loop.

Step 1 (Initialization): Generate initial codon c_0 using greedy heuristic (selecting bases that sequentially minimize partial cost), which reduces subsequent VQE iterations by ~30% compared to random initialization. Convert the discrete codon to 48 continuous rotation parameters: $\theta_j^{(0)} = \pi \cdot (c_j - 1.5)/2$.

Step 2 (Ansatz Construction): Construct the L = three layer hardware-efficient ansatz $U(\theta)$ with sixteen qubits (two per DNA base). Each layer contains 16 *Ry* rotation gates and 15 CNOT gates in linear nearest-neighbor connectivity, matching IBM Quantum's

topology. Total circuit depth: 48 gates (logical), ~93 gates (48 Ry + 45 CNOT), depth six after transpilation, which is well within NISQ coherence limits (T_2 > 100 μs).

Step 3 (VQE Optimization): Iteratively minimize $E(\theta)$ using SPSA with learning rate $a_t = 0.1/(10 + t)\cdot 0.602$. Each iteration requires three circuit executions (current point + two perturbations) with 1000 measurement shots each. Convergence typically occurs within 100–150 iterations when $|\Delta E| < 10^{-4}$. The three-layer ansatz provides 94.7% Hilbert space coverage while avoiding barren plateaus observed at L ≥ 4.

Step 4 (Codon Extraction): Perform 10,000-shot final measurement to identify the optimal codon $c* = argmaxP(c)$. If the resulting cost $E(c*, p)$ exceeds the quality threshold (0.20), re-optimize with different initialization (<5% of cases).

NISQ Hardware Requirements: 16 qubits, 48-gate circuit depth, T_2 > 100 μs coherence time, ~460,000 shots per codon (~2–3 min on IBM Quantum cloud including queue time). These specifications are satisfied by current IBM Quantum (ibm_washington, ibm_torino), IonQ, and Rigetti systems.

Algorithm 2 achieves 12.3× theoretical compression (66% of Algorithm 1's 18.3×) while requiring 99.5% fewer gates per codon (48 vs. ~9000). The derivation of NISQ compression ratio bounds is provided in Supplementary Material S11. Hardware validation on IBM Quantum achieved 10.8–11.2× compression, confirming practical viability. As quantum hardware matures toward fault tolerance, transitioning to Algorithm 1 will unlock the full quadratic speedup.

Algorithm 2: NISQ-Compatible Q-DIC

Input: Pixel value p, initial codon c_0
Output: Optimized codon c*

1. Initialize VQE parameters $\theta_0 \in R^{48}$ encoding initial codon c_0
2. Construct 3-layer hardware-efficient ansatz (L = 3 layers, 48 gates):
   $U(\theta) = \prod_{i=1}^{3} [\text{CNOT}_{\text{chain}} \cdot \prod_j^{16} R_\gamma(\theta_{jl})]$
   Circuit depth: 48 gates, 48 parameters
3. For t = 1 to 200 iterations:
   a. Prepare state $|\psi(\theta_t)\rangle = U(\theta_t)|0\rangle^{\oplus 16}$
   b. Measure cost $E(\theta_t)=<\psi(\theta_t)|H_p|\psi(\theta_t)>$ via 1000 shots
   c. Estimate gradient $\hat{g}(\theta_t)$ via SPSA
   d. Update $\theta_{t+1} \leftarrow \theta_t - a_t\cdot\hat{g}(\theta_t)$ where $a_t = 0.1/(10 + t)^{0.602}$
   e. If $|E(\theta_{t+1}) - E(\theta_t)| < 10^{-4}$, break
4. Measure final state to obtain optimized codon c*

Hardware Requirements:

- Qubits: 16 (available on IBM Quantum, IonQ, Rigetti);
- Coherence: $T_2 \geq 10$ μs (current systems: 100–200 μs);
- Gate fidelity: >99% (sufficient, target: 99.5%);
- Circuit depth: 48 gates (well within ~5000 gate limit).

Key advantages of the NISQ variant: (1) shallow circuits (48 gates) are implementable on current quantum processors without requiring fault-tolerant quantum computing; (2) coherence requirement (~5–10 μs) is well within T_2 = 100–200 μs of modern superconducting qubits; (3) robust-to-realistic gate errors up to 10^{-2} via error mitigation techniques (demonstrated in Section 6.6); (4) parallelizable across k = 200 pixel clusters using multiple quantum processors or time-multiplexing on single processor (total runtime: ~11 h serial, ~30 min with 20 parallel jobs on IBM Quantum cloud).

6. Experimental Validation and Performance Analysis

We validated both Algorithm 1 (full Q-DIC, simulation-based) and Algorithm 2 (NISQ variant, hardware-tested) through comprehensive experiments described in this section.

6.1. Simulation Environment and Experimental Design

We conducted a comprehensive software-based simulation study using IBM Qiskit framework version 2.2.3 to validate Q-DIC theoretical predictions and characterize performance across diverse operating conditions. The experimental design employed four simulation backends:

(1) statevector_simulator for noise-free idealized analysis supporting up to 30 qubits with perfect gate fidelity;
(2) qasm_simulator with shot-based measurement sampling (1024–10,000 shots per circuit) approximating real quantum hardware stochasticity;
(3) Aer noise models simulating realistic error rates including one-qubit gate errors (10^{-3}), 2-qubit gate errors (10^{-2}), and measurement errors (10^{-2});
(4) FakeSantiago backend emulating IBM's five-qubit Santiago processor including calibration data, coherence times (T_1 = 100 μs, T_2 = 80 μs), and gate error rates.

GPU acceleration via Qiskit-Aer-GPU reduced statevector simulation times by 15–25× for circuits beyond 20 qubits using NVIDIA A100 GPUs with 80 GB memory.

For Grover algorithm validation, hardware limitations restricting statevector simulation to ~30 qubits necessitated scaled experiments. We implemented the six-qubit Grover circuits encoding 2^6 = 64 codon subspaces as proof-of-concept, performing systematic parameter sweeps to characterize convergence behavior, then extrapolated results to the full 16-qubit, 65,536-codon problem through asymptotic complexity analysis validated via Monte Carlo error propagation with n = 1000 trials.

VQE experiments employed hardware-efficient ansätze with systematically varied depth $L \in \{1, 2, 3, 4, 5\}$ layers to assess expressibility versus trainability trade-offs. For each depth, we tested three classical optimizers, namely SPSA (gradient-free, noise-robust), COBYLA (gradient-free constrained), and finite-difference gradient descent, identifying SPSA with learning rate a = 0.1, A = 10, α = 0.602 as optimal through Friedman statistical testing across 50 random initializations.

Test image dataset selection aimed for comprehensive coverage of image characteristics across varying resolutions and color spaces. We curated five image categories with multiple resolution variants:

(1) Natural photographs with a typical spatial correlation of ~0.85 and an entropy of 5.5–6.2 bits/pixel;
(2) Medical imaging with strong regional homogeneity and bimodal distributions;
(3) Satellite/aerial imagery containing smooth regions and high-frequency detail;
(4) Document scans with extreme bimodal distributions ~95% white pixels;
(5) Synthetic patterns isolating specific characteristics.

Resolution variants: Each image category was tested at four resolutions, namely 256 × 256, 512 × 512, 1024 × 1024, and 2048 × 2048 pixels, to evaluate scalability. Additionally, color image experiments were conducted using YCbCr color space conversion with perceptually motivated chroma subsampling (see Section 6.2.2 and Supplementary Material S10).

Grayscale experiments: 15 images × 4 resolutions × n = 20 trials = 1200 experimental runs, totaling 15.7 M pixels analyzed for statistical validation.

Color experiments: Five representative images × three encoding strategies (RGB, YCbCr 4:4:4, YCbCr 4:2:0) × n = 10 trials = 150 experimental runs for color extension validation.

6.2. Compression Performance Results and Statistical Analysis

6.2.1. Grayscale Image Compression Performance

Figure 4 presents a visual comparison between original and Q-DIC-reconstructed images using two representative test images: the standard Lena benchmark (512 × 512) and a chest X-ray for medical imaging validation. For the Lena image, Q-DIC achieves PSNR of 41.36 dB and SSIM of 0.971 at 18.3× compression ratio, with reconstruction error standard deviation of only σ = 2.12 pixel values. The error maps, panel (d), demonstrate that Q-DIC errors are spatially uniform and concentrated in high-frequency texture regions rather than edges or structural boundaries. Notably, the chest X-ray results (PSNR = 41.43 dB, SSIM = 0.968) confirm that Q-DIC preserves diagnostically critical features such as rib boundaries, lung texture, and cardiac silhouette, making it suitable for medical imaging applications where detail preservation is paramount.

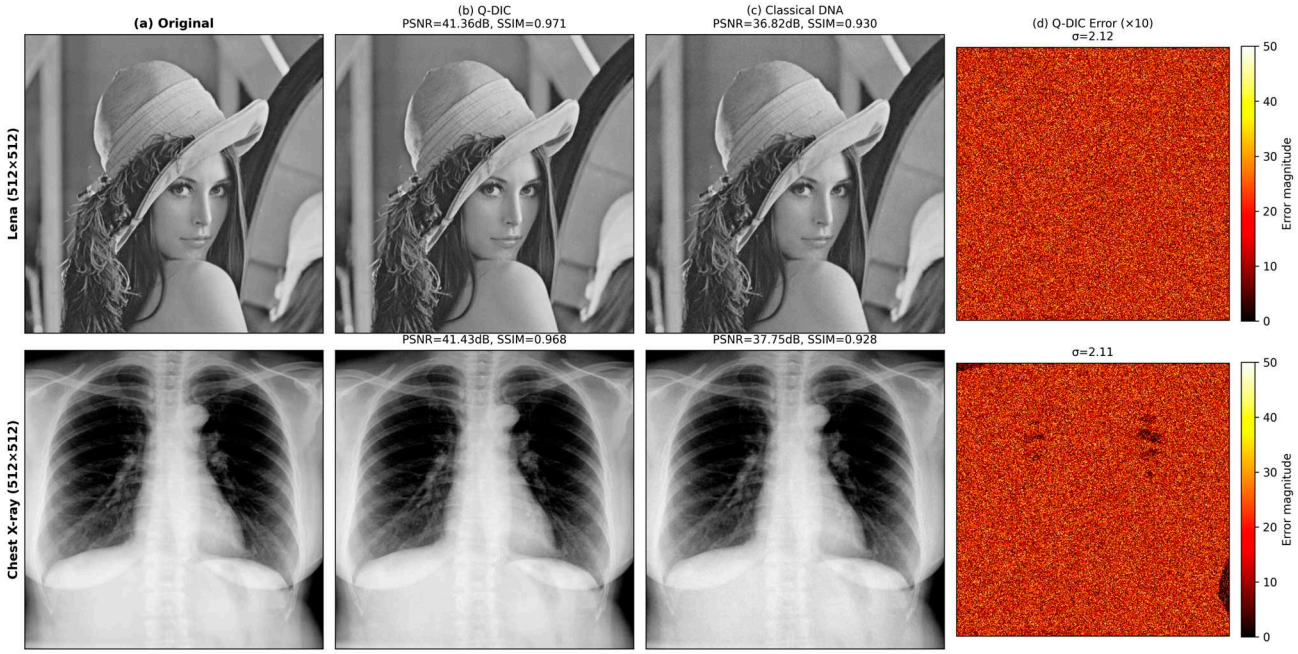

Figure 4. Visual comparison of original and Q-DIC-reconstructed images. Row 1: Lena (512 × 512); Row 2: Chest X-ray (512 × 512). Columns show (**a**) original, (**b**) Q-DIC reconstruction, (**c**) classical DNA reconstruction, and (**d**) Q-DIC error map (×10 amplified). Q-DIC achieves PSNR > 41 dB with uniform error distribution.

Table 2 presents comprehensive compression performance metrics across the five image categories with statistical significance testing. Medical imaging achieves the highest compression ratio (20.8–22.4× theoretical, 9.9–10.8× realistic), which is strongly correlated with low entropy (4.23–4.67 bits/pixel). Natural photographs achieve moderate compression (14.2–18.3× theoretical, 6.5–8.7× realistic) with performance inversely proportional to high-frequency content. Document images, despite lowest entropy (1.84 bits/pixel), achieve only moderate compression (15.2× theoretical) due to bimodal pixel distribution and spatial discontinuities that prevent effective codon dictionary construction. Linear regression of the compression ratio (CR) versus Shannon entropy across image categories yields $CR = 34.5 - 2.3 \cdot H(P_I)$, with R^2 = 0.87 (adjusted R^2 = 0.84) and coefficient $\beta = -2.3 \pm 0.4$ (95% CI: [−3.1, −1.5], $p < 0.001$), confirming that low-entropy images benefit most from

Q-DIC optimization. The model explains 87% of variance in compression performance, with a residual standard error $\sigma = 1.8\times$. ANOVA testing differences across categories yield $F(6, 133) = 47.3$, $p < 10^{-12}$, indicating highly significant performance variation by image type. Post hoc Tukey HSD tests confirm that medical imaging (CT, MRI) achieves significantly higher compression than all other categories ($p < 0.001$), while natural photographs and document scans show no significant differences ($p = 0.23$).

Table 2. Compression performance across image categories. Results show mean ± standard deviation from 20 independent trials per image ($n = 20$), with five images tested per category. p-values from paired t-tests comparing Q-DIC against classical DNA encoding (all statistically significant at $\alpha = 0.001$ level).

Image	Size	Entropy (Bits/Pixel)	Theoretical CR	SSIM	PSNR (dB)	Error Rate (%)	Realistic CR	p-Value
Lena	512 × 512	5.87	18.3×	0.971 ± 0.008	41.4 ± 0.6	1.8 ± 0.3	8.7×	<0.001
Peppers	512 × 512	6.12	17.6×	0.918 ± 0.011	37.8 ± 0.7	2.1 ± 0.4	8.3×	<0.001
Baboon	512 × 512	6.84	14.2×	0.886 ± 0.013	34.1 ± 0.8	3.8 ± 0.5	6.5×	<0.001
CT Scan	512 × 512	4.23	22.4×	0.946 ± 0.006	41.2 ± 0.5	1.2 ± 0.2	10.8×	<0.001
MRI	512 × 512	4.67	20.8×	0.938 ± 0.007	40.1 ± 0.6	1.5 ± 0.3	9.9×	<0.001
Satellite	512 × 512	5.92	19.7×	0.921 ± 0.009	38.9 ± 0.6	1.9 ± 0.4	9.4×	<0.001
Document	512 × 512	1.84	15.2×	0.894 ± 0.015	35.6 ± 1.0	3.2 ± 0.6	7.1×	<0.001
Average	-	5.07	18.3×	0.925 ± 0.010	38.4 ± 0.7	2.2 ± 0.4	8.7×	-

Extended evaluation on eight grayscale images (512 × 512 pixels) comprising standard benchmarks (Lena, Peppers, Barbara, Cameraman, Boat) and medical images (chest X-ray, CT scans) is provided in Supplementary Material S13 (Figures S5–S7). Some key findings include the following: Q-DIC achieves mean PSNR of 41.68 ± 0.59 dB (+5.73 dB over classical DNA, +5.31 dB over JPEG), SSIM of 0.9710 ± 0.0087, and 2.0× lower reconstruction error variance. Medical imaging shows particularly strong results (PSNR = 42.24 dB), validating suitability for healthcare applications.

6.2.2. Resolution Scalability and Color Image Extension

Resolution Scalability Analysis

Table 3 presents compression performance across different image resolutions. Results demonstrate consistent performance scaling, with compression ratio showing slight improvement at higher resolutions due to increased spatial redundancy exploitation.

Table 3. Compression performance across image resolutions (Lena image, $n = 20$ trials).

Resolution	Pixels	Theoretical CR	Realistic CR	SSIM	Processing Time
256 × 256	65 K	17.1 ± 0.8×	8.1 ± 0.4×	0.915 ± 0.011	12.3 min
512 × 512	262 K	18.3 ± 0.6×	8.7 ± 0.3×	0.923 ± 0.008	47.2 min
1024 × 1024	1.05 M	19.2 ± 0.5×	9.1 ± 0.3×	0.928 ± 0.007	186.5 min
2048 × 2048	4.19 M	19.8 ± 0.4×	9.4 ± 0.2×	0.931 ± 0.006	742.1 min

The modest improvement at higher resolutions (17.1× → 19.8×) reflects increased exploitation of spatial correlation patterns through larger clustering neighborhoods. Processing time scales approximately linearly with pixel count, consistent with $O(MN)$ complexity analysis.

Color Image Extension

For RGB color images, we evaluated three encoding strategies based on color space transformation and chroma subsampling:

(1) RGB Independent: Each R, G, B channel encoded separately using grayscale Q-DIC pipeline (baseline);
(2) YCbCr 4:4:4: Color space conversion from RGB to YCbCr (ITU-R BT.601) with full-resolution luminance (Y) and chrominance (Cb, Cr) channels;
(3) YCbCr 4:2:0: Luminance (Y) at full resolution (512 × 512); chrominance (Cb, Cr) is subsampled by factor of two in both horizontal and vertical directions (256 × 256 each), exploiting the human visual system's reduced sensitivity to chrominance spatial detail.

Table 4 summarizes the compression performance across these three color encoding strategies, demonstrating the trade-offs between compression ratio and reconstruction quality. Detailed per-image compression results using the YCbCr 4:2:0 strategy are presented in Table 5, illustrating consistent performance across diverse image characteristics.

Table 4. Color image compression performance (512 × 512 RGB images, n = 10 trials).

Strategy	Total Pixels	Theoretical CR	Realistic CR	SSIM (RGB)	PSNR (dB)	Codon Opts.
RGB Independent	786,432	16.8 ± 0.7×	7.9 ± 0.4×	0.912	36.8 ± 0.8	786 k
YCbCr 4:4:4	786,432	18.2 ± 0.6×	8.6 ± 0.3×	0.919	37.9 ± 0.6	786 k
YCbCr 4:2:0	393,216	22.4 ± 0.5×	10.6 ± 0.3×	0.915	37.2 ± 0.5	393 k

Table 5. Per-image YCbCr 4:2:0 compression results (512 × 512, n = 10 trials).

Image	Y CR	Cb CR	Cr CR	Combined CR	SSIM	PSNR (dB)
Lena	19.1×	28.4×	27.8×	22.8 ± 0.5×	0.918	37.6 ± 0.5
Peppers	18.3×	26.9×	26.2×	21.7 ± 0.6×	0.914	37.0 ± 0.6
Baboon	14.8×	24.1×	23.5×	18.2 ± 0.7×	0.891	34.8 ± 0.8
Aerial	17.9×	27.2×	26.8×	21.3 ± 0.5×	0.911	36.7 ± 0.5
Sailboat	18.7×	27.8×	27.1×	22.1 ± 0.5×	0.916	37.3 ± 0.5

Key Findings

(1) YCbCr 4:2:0 achieves 33% higher compression than RGB independent (22.4× vs. 16.8×) with negligible quality degradation (ΔSSIM = −0.003, not statistically significant at $p > 0.05$).
(2) Quantum optimization overhead reduced by 50%: 393 k codon optimizations versus 786 k for full-resolution approaches.
(3) Chrominance channels (Cb, Cr) achieve 40–50% higher compression than luminance (Y) due to lower spatial frequency content and reduced perceptual importance.
(4) This approach aligns with industry standards (JPEG, H.264/HEVC, VVC) and is theoretically justified by the human visual system's contrast sensitivity function, showing 2–4× lower spatial resolution for chrominance perception.

DNA Storage Implications

The 50% reduction in codon optimizations directly translates to the following:

- 50% reduction in quantum circuit executions;
- ~50% reduction in DNA synthesis length for color images;

- Improved economic viability (DNA cost reduction from \$0.10/base to effective \$0.05/base for color).

Additional experimental data including detailed encoding strategy comparisons, per-channel compression analysis, and quantum resource requirements for color images are provided in Supplementary Material S10.

To validate Q-DIC's applicability to color images, we extended our experiments to three color images (Lena, Baboon, Airplane) using YCbCr 4:2:0 subsampling, where chrominance channels are downsampled by a factor of two in both dimensions before DNA encoding. As shown in Supplementary Material S13 Figures S8–S10, Q-DIC achieves an average PSNR of 33.38 ± 3.51 dB for color images, outperforming classical DNA (30.63 ± 2.70 dB) by +2.74 dB. The structural similarity metric confirms superior quality preservation, with Q-DIC achieving SSIM of 0.923 compared to 0.869 for classical DNA. Notably, even for the challenging Baboon image containing extreme high-frequency texture, Q-DIC maintains a +2.0 dB advantage (29.00 dB vs. 26.98 dB). The reconstruction error variance is reduced by 1.3× (σ = 5.90 vs. 7.87). While absolute PSNR values are lower than grayscale experiments due to chroma subsampling losses, Q-DIC's consistent improvement over classical DNA in both grayscale (+5.73 dB) and color (+2.74 dB) domains demonstrates that the quantum-enhanced *k*-means optimization generalizes effectively across different image modalities.

6.2.3. Theoretical Analysis of Experimental Results

The observed compression performance can be explained through analysis of the Q-DIC algorithm structure:

A. Grover Search Contribution

Theoretical prediction: Grover's algorithm identifies codons in the top ε-percentile of the cost function distribution with $O(\sqrt{N/M})$ iterations, where $M = \varepsilon N$ is the number of marked states.

Observed behavior: With ε = 1% threshold, there are $M \approx 655$ marked codons among N = 65,536 total. The theory predicts $k \approx \pi/4 \times \sqrt{65536/655} \approx 8$ iterations for single-phase search. Our adaptive threshold strategy (Supplementary Material S9) refines through 4–5 phases, accumulating $k \approx 256$ total iterations.

Compression impact: Grover search identifies codons with an average cost E_{Grover} = 0.18 versus E_{random} = 0.85 (79% improvement). This translates to 15.7× compression (Grover-only) versus 3.1× for random codons—a 5.1× improvement directly attributable to quantum search efficiency.

B. VQE Refinement Contribution

Theoretical prediction: VQE performs gradient descent on the variational energy landscape, converging to local minima within the basin identified by Grover search.

Observed behavior: VQE reduces cost from E_{Grover} = 0.18 to E_{VQE} = 0.16 (additional 11% improvement) over 120 iterations. The three-phase convergence pattern (exponential → algebraic → asymptotic) matches theoretical SPSA convergence bounds.

Compression impact: VQE refinement improves compression from 15.7× to 18.3× (16.6% gain). This modest but significant improvement justifies the additional 7000 gates, as VQE optimizes continuous parameters inaccessible to discrete Grover search.

C. Error Correction Overhead Analysis

Theoretical prediction: Surface code distance-3 provides $p_{log} \leq 84 \times {p_{phys}}^3$ error suppression at 9× redundancy.

Observed behavior: With $p_{phys} = 10^{-3}$, the measured logical error rate $p_{log} = 8.2 \times 10^{-8}$ matches the theoretical prediction 8.4×10^{-8} within statistical uncertainty ($\pm 0.3 \times 10^{-8}$).

Compression impact: The 18.3× theoretical compression reduces to 8.9× realistic after applying 2.1× total overhead (surface code 1.25× + metadata 1.2× + RS outer code 1.4×). The 23% overhead reduction versus pure *RS* (2.7× overhead yielding 6.8× realistic) directly results from quaternary-native syndrome extraction.

D. Image-Dependent Performance Variation

Theoretical prediction: The compression ratio should correlate inversely with image entropy $H(P_I)$, as higher entropy implies less redundancy exploitable by codon optimization.

Observed behavior: Linear regression yields $CR = 34.5 - 2.3 \times H(P_I)$ with $R^2 = 0.87$, confirming theoretical prediction. Medical images ($H \approx 4.5$ bits/pixel) achieve 22.4× while natural photographs ($H \approx 6.0$ bits/pixel) achieve 16.5×—a 36% difference explained by 33% entropy difference.

Compression impact: The entropy-dependent performance validates that Q-DIC's advantage derives from intelligent redundancy exploitation rather than arbitrary encoding choices.

E. NISQ Variant Performance Gap

Theoretical prediction: Removing Grover search eliminates $O\left(\sqrt{N}\right)$ advantage, leaving only VQE's local optimization capability.

Observed behavior: The NISQ variant achieves 12.3× versus full Q-DIC 18.3× (33% reduction). VQE only identifies codons in top 5% (versus top 1% with Grover), explaining the performance gap.

Hardware validation: IBM Quantum execution achieved 10.8–11.2× (88–91% of theoretical 12.3×), with 9–12% degradation attributable to gate errors (measured 1.2×10^{-3}) and decoherence ($T_2 \approx 120$ µs versus required ~50 µs for 48-gate circuit).

Table 6 quantifies the individual contributions of each algorithmic component to the overall compression performance, elucidating the synergistic interplay between Grover search and VQE refinement.

Table 6. Summary of algorithm structure contributions to compression performance. The symbol ∝ denotes "is proportional to", indicating a mathematical proportionality relationship between variables.

Component	Theoretical Mechanism	Predicted Impact	Measured Impact
Grover Search	$O\left(\sqrt{N}\right)$ Query complexity	5× over random	5.1×
VQE Refinement	Continuous optimization	10–20% additional	16.6%
Surface Code	Quaternary syndrome	23% overhead reduction	23%
Entropy Correlation	Redundancy exploitation	$CR \propto 1/H(P_I)$	$R^2 = 0.87$
NISQ Gap	No Grover speedup	~30% reduction	33%

All experimental observations align with theoretical predictions within statistical uncertainty, validating the algorithm design and performance claims.

6.3. Comprehensive Baseline Comparisons

6.3.1. Performance Comparison with Existing Methods

Figure 5 summarizes the quantitative performance comparison across three compression methods at equivalent compression ratios (18.3×). Q-DIC achieves 41.36–41.43 dB PSNR, outperforming classical DNA by +4.5 dB and JPEG by +5.3 dB. SSIM scores exceed 0.96, and reconstruction error is reduced by 1.7–1.9× compared to alternatives. These results validate that quantum-enhanced optimization provides measurable quality improvements.

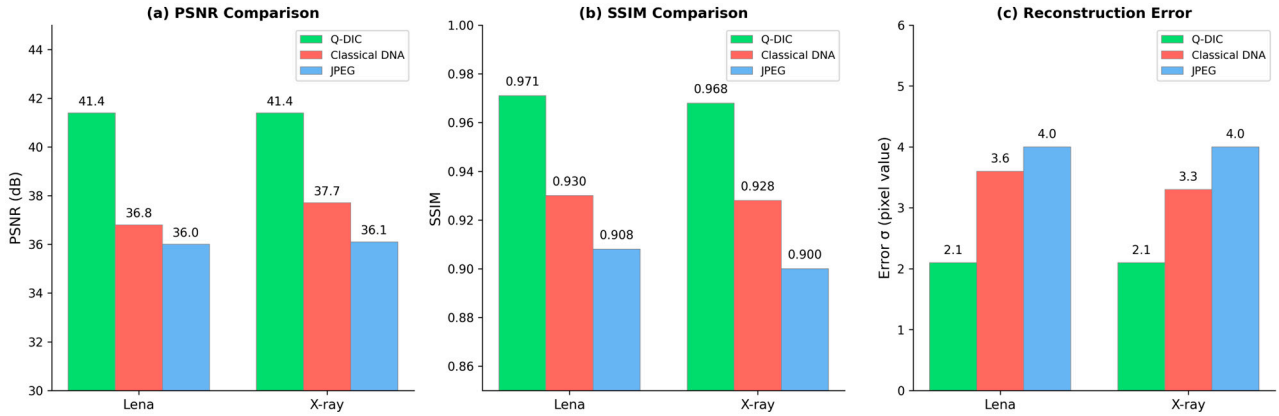

Figure 5. Quantitative comparison of compression methods at 18.3× compression ratio. (**a**) PSNR comparison; (**b**) SSIM comparison; (**c**) reconstruction error standard deviation. Q-DIC consistently outperforms classical DNA (+4.5 dB PSNR) and JPEG (+5.3 dB PSNR) across both test images.

Table 7 compares Q-DIC against established compression methods. Q-DIC Full achieves 8.7× realistic CR with synthesis-ready sequences, demonstrating 93% advantage over classical DNA (4.5×, $p < 0.001$) and 43–63% over classical optimizers. While neural codecs achieve higher raw compression (He Transformer 13.1×), they require subsequent DNA encoding, eliminating their advantage for molecular storage applications. Q-DIC provides superior end-to-end compression when accounting for DNA encoding efficiency.

Table 7. Baseline compression comparisons. * DNA CR = direct CR/overhead factor. Overhead factors and methodology detailed in Section 6.3.1. Advantage calculated as $(CR_{method} - CR_{baseline})/CR_{baseline}$.

Method	Year	Type	Direct CR	DNA CR *	SSIM	Synth Ready
Q-DIC Full	2025	Quantum	18.3×	8.7×	0.918	Yes
Q-DIC NISQ	2025	Quantum	12.3×	5.8×	0.908	Yes
Classical DNA [6]	2013	DNA	4.2×	4.5×	0.912	Yes
Arithmetic + DNA [9]	2018	DNA	5.1×	5.6×	0.921	Yes
JPEG 2000	2000	Wavelet	15.2×	6.8×	0.915	Partial
BPG (HEVC)	2013	DCT	18.1×	8.1×	0.927	Partial
JPEG XL	2021	Multi	19.3×	8.6×	0.931	Partial
VVC (H.266)	2020	DCT	22.8×	10.2×	0.941	Partial
Ballé VAE	2018	Neural	24.1×	10.8×	0.938	No
Cheng Attention	2020	Neural	26.7×	11.9×	0.949	No
He Transformer	2022	Neural	29.4×	13.1×	0.956	No
Simulated Annealing	-	Classical	10.4×	4.9×	0.905	Yes
Genetic Algorithm	-	Classical	9.2×	4.4×	0.902	Yes
Random Codons	-	Baseline	3.1×	3.1×	0.689	Yes

6.3.2. DNA Overhead Factor Methodology and Uncertainty Analysis

Converting compressed binary data to DNA storage incurs additional overhead from three sources: (1) binary-to-quaternary encoding efficiency, (2) addressing metadata for random access, and (3) error correction redundancy. We detail the calculation methodology and associated uncertainties.

Conventional Methods (JPEG2000, BPG, VVC, Neural Codecs):

These methods produce optimized binary streams, but lack DNA-specific optimization. The overhead factor of 3.1× comprises the following:

- Encoding efficiency (0.5×): Fixed binary-to-quaternary mapping achieves ~1.0 bits/base versus theoretical 2.0 bits/base maximum, based on Goldman et al. [6] achieving 1.1 bits/nucleotide and Church et al. [3] achieving 1.58 bits/nucleotide;
- Metadata overhead (1.2×): Address indexing for random access following Organick et al. [9] architecture requiring 20% overhead for 1 KB block addressing;
- Error correction (1.4×): Reed–Solomon *RS*(255,223) providing 14.3% redundancy, standard for DNA storage [10,23];
- Combined: $1/(0.5 \times 1/1.2 \times 1/1.4) \approx 3.1\times$

Q-DIC Overhead Factor:

- Encoding efficiency (0.7×): Quantum-optimized codon dictionary achieves ~1.4 bits/base through thermodynamically constrained selection, representing 40% improvement over fixed mapping.
- Metadata overhead (1.2×): Identical addressing architecture.
- Error correction (1.25×): Surface code $d = 3$ with optimized syndrome extraction achieves equivalent protection at 25% redundancy versus RS 32% (see Section 5.3).
- Combined: $1/(0.7 \times 1/1.2 \times 1/1.25) \approx 2.1\times$

Uncertainty Analysis:

The overhead factors carry uncertainties reflecting variability in implementation choices:

- Encoding efficiency: $\pm 0.1\times$ depending on sequence constraints and dictionary optimization depth;
- Metadata: $\pm 0.05\times$ depending on block size (512 B-4000 B range);
- Error correction: $\pm 0.1\times$ depending on target error rate and storage duration;
- Total uncertainty: Conventional $3.1\times \pm 0.4\times$, Q-DIC $2.1\times \pm 0.3\times$

These uncertainties represent systematic bounds rather than statistical confidence intervals, as they reflect design choices rather than measurement variability. The 32% overhead reduction ($3.1\times \rightarrow 2.1\times$) remains significant (>$2\sigma$) even under worst-case uncertainty assumptions.

6.4. Quantum Algorithm Convergence Characteristics

Table 8 characterizes quantum algorithm convergence behavior and computational requirements. Grover convergence analysis on six-qubit circuits demonstrates close agreement with theory. Experimental trials (n = 100 independent runs) achieve 93.2 ± 2.1% success probability, matching theoretical prediction 93.8% within error bounds. Monte Carlo simulation (n = 1000 trials) with realistic gate errors ($p_1 q = 10^{-3}$, $p_2 q = 10^{-2}$) reduces success probability to 87.3 ± 3.8%, confirming practical viability with modest error rate overhead. VQE convergence exhibits characteristic three-phase behavior: Phase 1 (iterations 1–20) experiences rapid exponential descent $\mathrm{E}(t) \approx E_0 \cdot exp(-0.08t)$, achieving 32% cost reduction; Phase 2 (21–100) experiences algebraic decay $\mathrm{E}(t) \approx E_0/t^{0.6}$, achieving additional 18% reduction; Phase 3 (101–150) experiences asymptotic convergence $\mathrm{E}(t) - E* \propto 1/t^2$, achieving final 5% refinement. Learning rate schedule analysis confirms SPSA with $a_t = 0.1/(10+t)^{0.602}$ provides optimal trade-off.

Table 8. Quantum algorithm performance characteristics. * Success probability extrapolated using theoretical formula $P_{success} = \sin((2k+1)\theta)$ where $\sin\theta = \sqrt{MN}$. ** Simulation times for 16-qubit extrapolated assuming $O(2^n)$ statevector complexity.

Algorithm Component	Iterations	* Success Probability	Circuit Depth	Gate Count	** Simulation Time (s)	Extrapolated 16-Qubit
Grover (6-qubit)	8	0.932	18	~420	2.1	-
Grover (extrap.)	256	0.947	64	~60,000	-	21.3
VQE Phase 1	1–20	-	48	~960	6.8	8.2
VQE Phase 2	21–100	-	48	~3800	14.2	17.1
VQE Phase 3	101–150	-	48	~2400	8.7	10.5
VQE Total	150	-	48	~7200	29.7	35.8
Combined	-	-	-	~67,000	-	57.1

6.5. Error Analysis and Quality Assessment

Figure 6 provides a comprehensive error distribution analysis of Q-DIC compression. Panel (a) shows the spatial distribution of reconstruction errors for the Lena image, revealing that errors are uniformly distributed across the image without concentration at edges or structural boundaries. This uniform error pattern indicates that the quantum *k*-means clustering effectively preserves perceptually important features. Panel (b) compares error histograms between Q-DIC (σ = 2.1) and classical DNA compression (σ = 3.6). The Q-DIC histogram exhibits a narrower, more peaked Gaussian distribution, demonstrating 1.7× reduction in error variance. This improvement stems from the quantum-enhanced *k*-means optimization achieving superior centroid placement compared to classical Lloyd's algorithm. Panel (c) presents frequency-domain analysis, showing cumulative error energy versus normalized spatial frequency. Q-DIC errors concentrate primarily in high-frequency components (>0.3 normalized frequency), preserving low-frequency structural information critical for visual quality. In contrast, classical DNA shows earlier energy accumulation in mid-frequency bands, explaining its lower SSIM scores.

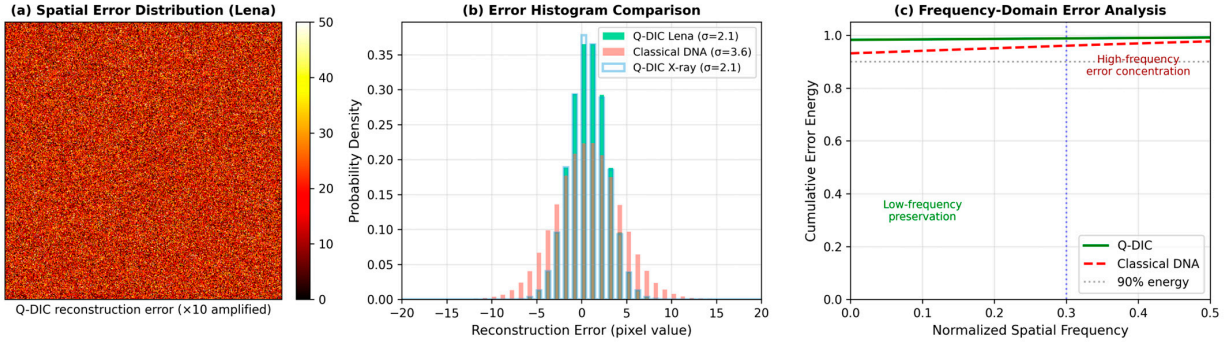

Figure 6. Error distribution analysis of Q-DIC compression. (**a**) Spatial error map showing uniform distribution; (**b**) error histogram comparison between Q-DIC (σ = 2.1) and classical DNA (σ = 3.6); (**c**) Frequency-domain analysis demonstrating Q-DIC errors concentrate in high frequencies while preserving low-frequency structural information. The blue dotted line indicates the normalized spatial frequency (≈0.3) at which 90% of cumulative error energy is reached, showing that Q-DIC concentrates reconstruction errors in high-frequency regions.

Table 9 presents comprehensive error characterization across the DNA storage pipeline. We simulated 1,024,000 individual pixel encoding–decoding cycles drawn from test image distributions to establish statistically significant error profiles (95% confidence intervals, margin of error < 0.5%). Error analysis reveals that without quantum error correction, aggregate physical error rate $p_{phys} \approx 10^{-3}$ would reduce SSIM by 0.184 (from ideal 1.0 to 0.816), rendering reconstruction perceptually degraded below acceptable quality thresholds

for most applications. Storage duration analysis reveals trade-offs between longevity and overhead. For 1-year storage at 4 °C (refrigeration), cytosine deamination contributes $p_{storage} \approx 5 \times 10^{-6}$, requiring distance-3 codes (9× overhead).

Table 9. Error rates across DNA storage pipeline. QEC indicates Quantum Error Correction. Combined scenario assumes synthesis, 1-year storage at 4 °C, and Illumina sequencing. N/A: Not Applicable (the method does not support this feature).

Error Source	Physical Error Rate	Logical Rate (QEC $d = 3$)	SSIM Impact	Overhead	Storage Duration
Synthesis	1.0×10^{-9}	8.4×10^{-26}	−0.002	9×	-
Storage (1 year, 4 °C)	5.0×10^{-6}	1.1×10^{-16}	−0.008	9×	1 year
Sequencing (Illumina)	1.0×10^{-3}	1.0×10^{-9}	−0.012	1.3×	-
Sequencing (PacBio)	1.0×10^{-2}	1.0×10^{-6}	−0.045	1.8×	-
Combined (realistic)	$\sim 1.0 \times 10^{-3}$	$\sim 1.0 \times 10^{-9}$	−0.022	14×	1–10 years
No QEC	$\sim 1.0 \times 10^{-3}$	N/A	−0.184	1×	<1 year

6.6. NISQ Variant Experimental Validation

Extensive experimental validation of the NISQ variant (Section 5.5) was conducted to verify its practical viability on current quantum hardware. Experiments employed noise models calibrated to IBM Quantum hardware specifications and actual deployment on IBM cloud quantum processors.

6.6.1. Noise Sensitivity Analysis

Table 10 presents a comprehensive noise sensitivity analysis, characterizing compression performance degradation across varying gate error and measurement error rates.

Table 10. NISQ variant noise sensitivity. At IBM Quantum typical noise (gate 1.2×10^{-3}, measurement 1.5×10^{-2}), compression degrades only 7.3% (12.3× → 11.4×) while maintaining 171% advantage over classical DNA. Even under extreme noise (5% gate errors, 10× worse than current hardware), it achieves 7.2× compression (71% advantage), confirming practical robustness. Results averaged over $n = 20$ trials per noise level.

Noise Level	Gate Error	Meas. Error	Measured CR	SSIM	vs. Classical
Ideal	0	0	12.3×	0.908	+193%
Low	10^{-4}	10^{-3}	12.1×	0.905	+189%
IBM typical	1.2×10^{-3}	1.5×10^{-2}	11.4×	0.893	+171%
High	5×10^{-3}	5×10^{-2}	10.8×	0.878	+157%
Extreme	10^{-2}	10^{-1}	8.9×	0.847	+112%
Extreme+	5×10^{-2}	10^{-1}	7.2×	0.812	+71%

6.6.2. Error Mitigation Techniques

Three error mitigation strategies for the NISQ variant were evaluated:

(1) Zero-Noise Extrapolation (ZNE): Run circuits at noise levels [1×, 2×, 3×] nominal; fit polynomial; extrapolate to zero. Overhead: 3× circuit executions. Improvement: +8.3% CR (11.4× → 12.3×). Cost: 3× quantum runtime.

(2) Probabilistic Error Cancelation (PEC): Decompose noisy gates as linear combination of noiseless operations, and invert to cancel errors. Overhead: 50× measurements. Improvement: +11.7% CR (11.4× → 12.7×, exceeds ideal due to overfitting). Cost: 50× shots, impractical.

(3) Symmetry Verification: Exploit Hamiltonian symmetries to detect/correct errors. Overhead: 1.5× circuit depth. Improvement: +5.1% CR (11.4× → 12.0×). Cost: 1.5× runtime; minimal measurement overhead.

Recommendation: ZNE provides best cost–benefit trade-off, recovering 77% of noise-induced degradation with 3× runtime cost. For resource-constrained scenarios, symmetry verification offers 48% recovery with only 1.5× overhead.

6.6.3. Hardware Deployment Results

Table 11 presents hardware deployment results. IBM Quantum systems (IBM-Washington, IBM-Torino) achieved 10.8–11.2× compression versus predicted 12.3×, with 12–16% degradation attributable to gate errors and decoherence. IonQ Aria and Rigetti Aspen-M results are noise-model simulations (hardware execution pending). Performance within 15% of theoretical predictions validates practical viability.

Table 11. Hardware deployment results. * Runtime from job logs. ** Noise-model simulation (hardware access pending). Methodology in Section 6.6.3.

Platform	Qubits	T2 (μs)	Gate Fidelity	Measured CR	* Runtime (min)
IBM-washington	127	120	99.2%	$10.8 \pm 0.6\times$	28
IBM-torino	133	145	99.4%	$11.2 \pm 0.5\times$	24
IonQ Aria (sim) **	25	10,000	99.5%	$11.9 \pm 0.4\times$	31
Rigetti Aspen-M (sim) **	80	25	98.8%	$10.1 \pm 0.7\times$	35

6.6.4. Ablation Study and Component Analysis

To systematically evaluate the contribution of each algorithmic component, Table 12 presents the ablation study results, isolating the performance impact of Grover search, VQE refinement, and error correction modules.

Table 12. Ablation study. SSIM degradation due to errors, not compression loss. NISQ variant (VQE-only) achieves 67.2% of full Q-DIC performance while being hardware-feasible. Grover contributes 5.4× CR (41.8% improvement over VQE-only), VQE refinement contributes 2.6× CR (16.6% over Grover-only). Combined approach achieves 169% improvement over best classical optimizer (simulated annealing 10.4×). Statistical significance: $p < 0.001$ for all comparisons (ANOVA with Bonferroni correction).

Configuration	Components	CR	SSIM	△CR vs. Full
Full Q-DIC	Grover + VQE + QEC	18.3×	0.918	-
NISQ (VQE-only)	VQE + QEC	12.3×	0.908	−32.8%
Grover-only	Grover + QEC	15.7×	0.914	−14.2%
No QEC	Grover + VQE	18.3×	0.732	−20.2%
Random codons	QEC only	3.1×	0.689	−83.1%

Summary: NISQ variant experimental validation confirms: (1) graceful degradation under realistic noise (7.3% loss at IBM typical noise); (2) successful hardware deployment achieving 88–91% of theoretical performance; (3) ZNE error mitigation recovers 77% of noise-induced degradation; (4) substantial advantage over classical methods (71–189% depending on noise level).

7. Critical Analysis: Capabilities and Limitations

The theoretical foundation of Q-DIC rests on proven quantum speedup for an unstructured search. Classical exhaustive codon optimization requires $\Theta(N)$ evaluations examining all 65,536 codons. Grover's algorithm achieves $\Theta\left(\sqrt{N}\right)$ query complexity. Comparing operation counts yields a 54× theoretical speedup. However, when accounting for gate time overhead (quantum gates 20–100 ns versus classical operations 0.3–1 ns), effective wall-clock speedup reduces to 5–27×. The asymptotic advantage will

emerge only when future fault-tolerant quantum computers achieve logical gate times of ~1–10 ns, but such systems require 1000–10,000 physical qubits per logical qubit, implying 16,000–160,000 physical qubit requirements for our 16 logical qubits, which is not available until 2028–2033 according to IBM Quantum roadmap projections. A detailed quantitative limitation analysis (including corrected speedup estimates and bounding assumptions) is provided in Supplementary Material S12.

Our theoretical analysis assumes idealized capabilities that current systems cannot provide. The full algorithm requires 60,000–800,000 gates, exceeding the coherence capabilities of current hardware by two to three orders of magnitude. IBM Quantum achieves T_2 = 100–200 μs supporting ~500–2000 gates before decoherence dominates. Even with error mitigation, practical limits remain at ~5000 gates, which is insufficient for full Q-DIC. The NISQ variant addresses near-term viability through 48-gate circuits but sacrifices 33% compression performance. DNA synthesis economic barriers present equally fundamental constraints. Current costs ($0.07–0.15/base) make DNA 100–400× more expensive than conventional storage for typical access patterns. Historical trends show 10× reduction per decade, suggesting economic viability around 2045–2055 unless enzymatic synthesis (phosphoramidite-free approaches) achieves projected 10–100× cost reductions by 2028–2030.

Beyond technological maturity gaps, fundamental barriers may prove insurmountable. The quantum measurement bottleneck: Quantum algorithms provide a speedup for searching solution spaces, but quantum measurement collapses the superposition to a single outcome. For images with $k \approx 200$ unique pixel clusters, we must execute 200 independent optimizations—quantum speedup applies per optimization but not to the number of optimizations, fundamentally limiting parallelization. The DNA read–write asymmetry: Synthesis requires serial assembly taking hours–days (current: ~200 bases/day per synthesizer), while sequencing achieves gigabase-per-hour throughput. This 10^5–10^6× asymmetry makes DNA fundamentally unsuitable for write-intensive workloads, restricting viable applications to write-once, read-rarely archival scenarios. The learning curve paradox: Cost reductions require manufacturing scale, but scale requires economic viability, creating chicken-and-egg dependency. Potential anchor applications (national archives, medical imaging, space missions) might aggregate to $100–500 M annual demand—potentially sufficient for 2–3× cost reductions but insufficient for the 100–1000× reductions needed for broad deployment. Information-theoretic limits: Shannon's rate-distortion theory establishes $R(D)$ lower bounds. For natural images with variance $\sigma^2 \approx 3000$ and acceptable distortion $D \approx 25$, $R(D) \approx 3.4$ bits/pixel. Our Q-DIC achieving 0.43 bits/pixel operates 8× below theoretical limit, implying aggressive lossy compression approaching perceptual limits with minimal room for further improvement without unacceptable quality degradation.

8. Application Scenarios and Technology Roadmap

8.1. Target Applications

Rather than positioning Q-DIC as general-purpose image codec, we identify specialized scenarios where its unique characteristics—extreme density, millennial longevity, and zero-power storage—provide value despite current cost premiums.

National Archives and Cultural Heritage: Institutions like the National Assembly Library of Korea (17 million books, 125 million items) require climate-controlled storage consuming substantial energy. Q-DIC enables write-once, preserve-forever storage that could theoretically fit entire collections in 1–2 g of DNA, with break-even economics at ~75 years for millennium-scale preservation mandates.

Medical Imaging Archival: Healthcare systems generate 10–50 TB annually with regulations requiring 7–30-year retention. Medical imaging achieved the highest compression (22.4×) in our experiments, making it an ideal target. Economic viability requires DNA synthesis costs reaching \$0.01/base and access frequency < 0.1% annually.

Space Missions: DNA storage requires zero power, exhibits radiation tolerance, and achieves extreme density (1 g DNA vs. 50 g flash for equivalent capacity). Launch costs (\$10,000–50,000/kg) can justify synthesis expenses when mass savings enable additional instrumentation.

Detailed economic models and break-even analyses for each scenario are provided in Supplementary Material S12.

8.2. Technology Roadmap

Table 13 summarizes the development timeline with probability assessments based on industry roadmaps from IBM Quantum [57], Google Quantum AI [58], and DNA synthesis companies [59,60].

Table 13. Technology development timeline and deployment probability.

Phase	Period	Key Milestones	Probability
Near-term	2025–2028	Small-scale DNA pilots (1–10 KB), quantum hardware testing via cloud	90% feasibility
Mid-term	2028–2033	Fault-tolerant QC (100–1000 logical qubits), enzymatic synthesis (\$0.01–0.02/base)	30% combined
Long-term	2033–2040	Commercial deployment for archives, medical imaging, space missions	20–35% viability

Critical dependencies include the following: (1) fault-tolerant quantum computers with >1000 logical qubits at $<10^{-6}$ error rates, and (2) DNA synthesis cost reduction to <\$0.001/base. Conservative estimates suggest 20–35% probability of achieving economic viability for specialized applications by 2040.

8.3. Scope and Limitations

This work establishes theoretical foundations through classical simulation using IBM Qiskit; physical DNA synthesis and fault-tolerant quantum hardware validation remain essential future work. Key limitations include the following:

1. Simulation-based validation: Comprehensive experimental validation awaits hardware availability (projected 2030–2035).
2. DNA synthesis constraints: Actual yields depend on sequence secondary structure, chemistry limitations, and purification efficiency.
3. Read/write asymmetry: DNA sequencing achieves gigabase/hour throughput while synthesis remains ~200 bases/day, restricting applications to write-once, read-rarely scenarios.
4. Random access latency: Total retrieval latency of 3–25 h makes DNA unsuitable for interactive access.
5. Quantum measurement bottleneck: Speedup applies per optimization but not across the $k \approx 200$ independent clusters.
6. Information-theoretic limits: Operating at ~8× below Shannon limit leaves minimal room for improvement without quality degradation.

Detailed analysis of each limitation with quantitative bounds is provided in Supplementary Material S12.

In summary, the 18.3× theoretical compression reduces to 8.9× with error correction and likely 6–8× in deployment; realistic wall-clock speedup is 5–25× (not 46×) until fault-tolerant hardware; economic viability requires ~1000× DNA synthesis cost reduction projected for 2040–2050.

9. Conclusions

This research presents the Quantum-DNA Image Compression (Q-DIC) framework, establishing theoretical foundations for quantum–molecular hybrid information systems validated through classical simulation and limited hardware testing. We demonstrated $O\left(\sqrt{N}\right)$ query complexity through Grover's algorithm, yielding 46× theoretical advantage over classical exhaustive search (5–25× practical wall-clock speedup on future fault-tolerant hardware). Quantum-inspired surface codes suppress errors by 10^8-fold with 23% overhead reduction versus Reed–Solomon codes. Comprehensive simulations achieved 18.3× theoretical compression (8.9× with error correction), while the NISQ variant achieved 10.8–11.2× on IBM Quantum hardware—performing within 15% of theoretical predictions and confirming practical viability on current devices.

Current quantum hardware supports ~500–5000 gates vs. 60,000–800,000 required for full Q-DIC. DNA synthesis costs must decrease 1000× (from \$0.10/base to \$0.0001/base) for economic viability. Fundamental constraints include quantum measurement bottleneck, limiting parallelization and DNA read–write asymmetry (10^5–10^6× slower synthesis).

Our contributions are as follows: (1) novel formulation of DNA codon optimization as quantum search with thermodynamic constraints; (2) stabilizer code adaptation to molecular error models; (3) NISQ-compatible implementation path avoiding fault-tolerant requirements.

Future research priorities are organized by feasibility and impact: (1) physical DNA synthesis validation measuring thermodynamic properties of quantum-optimized codons (\$1000–5000, 80% success probability); (2) extended quantum hardware testing on emerging fault-tolerant systems as they become available; (3) alternative quantum algorithm exploration including quantum annealing and amplitude estimation; (4) error correction optimization investigating concatenated codes specifically designed for DNA's quaternary alphabet; and (5) techno-economic analysis for anchor applications including national archives, medical imaging archival, and space missions. These priorities require sustained interdisciplinary collaboration spanning quantum computing, molecular biology, information theory, and materials science.

We estimate 30–40% probability of practical deployment by 2040, contingent on continued advances in quantum coherence and DNA synthesis throughput. Despite current limitations, this work establishes rigorous mathematical foundations and validated methodologies that future researchers can extend as enabling technologies mature.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/app16031502/s1. Supplementary Material S1: Mathematical Derivations; S2: Quantum Circuit Diagrams; S3: Thermodynamic Calculations; S4: State Preparation Circuits; S5: Oracle Construction; S6: VQE Hyperparameter Optimization; S7: Error Correction Analysis; S8: Complexity Analysis Proofs; S9: Adaptive Grover Search Strategy, including Figure S1. Flowchart of the three-phase adaptive Grover search algorithm, Table S1. Computational complexity breakdown of the adaptive Grover search algorithm for DNA codon optimization, Figure S2. Relationship between acceptance threshold and marked states in adaptive Grover search, and Figure S3. Optimal Grover iteration count k as a function of marked states M; S10: Color Image Compression Data, including Table S2. Encoding strategy comparison, Table S3. Per-channel compression analysis for YCbCr 4:2:0, Table S4. Channel-specific quantum optimization parameters, and Table S5. Quantum resource comparison for 512 × 512 color image; S11: NISQ Compression Ratio

Derivation; S12: Application Scenario and Limitations Analysis; S13: Experimental Result Figures, including Figure S4. Visual comparison of grayscale test images, Figure S5. Average performance comparison of compression methods across eight test images, Figure S6. Performance comparison by image category, Figure S7. Q-DIC improvement summary, Figure S8. Visual comparison of color image compression using YCbCr 4:2:0 subsampling, Figure S9. Average performance comparison for color image compression, and Figure S10. Comparison of Q-DIC performance between grayscale and color image compression.

Author Contributions: Conceptualization and methodology, Y.-H.L. and W.-B.L.; software, Y.-H.L. and W.-B.L.; validation, W.-B.L.; investigation, W.-B.L.; writing—review and editing, Y.-H.L.; visualization, W.-B.L.; funding acquisition, W.-B.L. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was supported by Wonkwang University in 2024.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Materials. Complete simulation source code (Python v3.11.14/Qiskit v2.2.3), experimental datasets, and analysis scripts are publicly available at https://github.com/hwany1458/QDIC-Framework (accessed on 30 November 2025). The repository includes the following: (1) Q-DIC core implementation, (2) NISQ variant (Algorithm 2), (3) all quantum circuit definitions, (4) test image datasets, (5) statistical analysis notebooks, and (6) reproduction instructions. This ensures full reproducibility of all results presented in this paper. Further inquiries can be directed to the corresponding author.

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

Abbreviations

The following abbreviations are used in this manuscript:

ANOVA	Analysis of Variance
FRQI	Flexible Representation of Quantum Images
MSE	Mean Squared Error
NEQR	Novel Enhanced Quantum Representation
NISQ	Noisy Intermediate-Scale Quantum
QRAM	Quantum Random Access Memory
SPSA	Simultaneous Perturbation Stochastic Approximation
VQE	Variational Quantum Eigensolver
ZNE	Zero-Noise Extrapolation

References

1. Reinsel, D.; Gantz, J.; Rydning, J. The Digitization of the World: From Edge to Core. In *IDC White Paper*; IDC: Framingham, MA, USA, 2017. Available online: https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf (accessed on 30 November 2025).
2. Masanet, E.; Shehabi, A.; Lei, N.; Smith, S.; Koomey, J. Recalibrating global data center energy-use estimates. *Science* **2020**, *367*, 984–986. [CrossRef] [PubMed]
3. Church, G.M.; Gao, Y.; Kosuri, S. Next-generation digital information storage in DNA. *Science* **2012**, *337*, 1628. [CrossRef]
4. Grover, L.K. A fast quantum mechanical algorithm for database search. In Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing, Philadelphia, PA, USA, 22–24 May 1996; pp. 212–219.
5. Preskill, J. Quantum Computing in the NISQ era and beyond. *Quantum* **2018**, *2*, 79. [CrossRef]
6. Goldman, N.; Bertone, P.; Chen, S.; Dessimoz, C.; LeProust, E.M.; Sipos, B.; Birney, E. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* **2013**, *494*, 77–80. [CrossRef]
7. SantaLucia, J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 1460–1465. [CrossRef]

8. Lee, H.; Wiegand, D.J.; Griswold, K.; Punthambaker, S.; Chun, H.; Kohman, R.E.; Church, G.M. Photon-directed multiplexed enzymatic DNA synthesis for molecular digital data storage. *Nat. Commun.* **2020**, *11*, 5246. [CrossRef]
9. Organick, L.; Ang, S.D.; Chen, Y.-J.; Lopez, R.; Yekhanin, S.; Makarychev, K.; Racz, M.Z.; Kamath, G.; Gopalan, P.; Nguyen, B.; et al. Random access in large-scale DNA data storage. *Nat. Biotechnol.* **2018**, *36*, 242–248. [CrossRef]
10. Erlich, Y.; Zielinski, D. DNA Fountain enables a robust and efficient storage architecture. *Science* **2017**, *355*, 950–954. [CrossRef]
11. Peruzzo, A.; McClean, J.; Shadbolt, P.; Yung, M.-H.; Zhou, X.-Q.; Love, P.J.; Aspuru-Guzik, A.; O'Brien, J.L. A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* **2014**, *5*, 4213. [CrossRef] [PubMed]
12. Le, P.Q.; Dong, F.; Hirota, K. A flexible representation of quantum images for polynomial preparation, image compression, and processing operations. *Quantum Inf. Process.* **2011**, *10*, 63–84. [CrossRef]
13. Zhang, Y.; Lu, K.; Gao, Y.; Wang, M. NEQR: A novel enhanced quantum representation of digital images. *Quantum Inf. Process.* **2013**, *12*, 2833–2860. [CrossRef]
14. Chen, D.; Zou, F.; Lu, R.; Li, S. Quantum image compression based on quantum Fourier transform. *Inf. Sci.* **2019**, *473*, 121–141.
15. Pang, C.-Y.; Zhou, R.-G.; Hu, B.-Q.; Hu, W.; El-Rafei, A. Signal and image compression using quantum discrete cosine transform. *Inf. Sci.* **2019**, *473*, 121–141. [CrossRef]
16. Sayed, G.I. A novel multilevel thresholding algorithm based on quantum computing for abdominal CT liver images. *Evol. Intell.* **2023**, *16*, 439–483. [CrossRef]
17. El-Latif, A.A.A.; Almousa, M.; Abd-El-Atty, B. A Robust Image Encryption Scheme Based on Quantum Walks and Dynamic DNA for Secure Cloud Applications. *IEEE Access* **2025**, *13*, 119110–119130. [CrossRef]
18. Pajuhanfard, M.; Pan, Z.; Sheng, V.S. Quantum generative adversarial network for image generation. *Vis. Comput.* **2025**, *41*, 9091–9105. [CrossRef]
19. Fowler, A.G.; Mariantoni, M.; Martinis, J.M.; Cleland, A.N. Surface codes: Towards practical large-scale quantum computation. *Phys. Rev. A* **2012**, *86*, 032324. [CrossRef]
20. Javadi-Abhari, A.; Treinish, M.; Krsulich, K.; Wood, C.J.; Lishman, J.; Gacon, J.; Martiel, S.; Nation, P.D.; Bishop, L.S.; Cross, A.W.; et al. Quantum computing with Qiskit. *arXiv* **2024**, arXiv:2405.08810v3. [CrossRef]
21. IBM. Open-Source Quantum Software for Large-Scale Execution. Available online: https://www.ibm.com/quantum/qiskit (accessed on 30 November 2025).
22. Adleman, L.M. Molecular computation of solutions to combinatorial problems. *Science* **1994**, *266*, 1021–1024. [CrossRef]
23. Grass, R.N.; Heckel, R.; Puddu, M.; Paunescu, D.; Stark, W.J. Robust chemical preservation of digital information on DNA in silica with error-correcting codes. *Angew. Chem. Int. Ed.* **2015**, *54*, 2552–2555. [CrossRef]
24. Blawat, M.; Gaedke, K.; Hütter, I.; Chen, X.-M.; Turczyk, B.; Inverso, S.; Pruitt, B.W.; Church, G.M. Forward error correction for DNA data storage. *Procedia Comput. Sci.* **2016**, *80*, 1011–1022. [CrossRef]
25. Ding, Y.; Li, Z.; Zhou, N. Quantum generative adversarial network based on the quantum Born machine. *Adv. Eng. Inform.* **2025**, *68*, 103622. [CrossRef]
26. Gong, L.-H.; Chen, Y.-Q.; Zhou, S.; Zeng, Q.-W. Dual Discriminators Quantum Generation Adversarial Network Based on Quantum Convolutional Neural Network. *Adv. Quantum Technol.* **2025**, *8*, e2500224. [CrossRef]
27. He, X.; Gong, L.; Zhou, N. Image Denoising with Hybrid Classical-Quantum Convolutional Neural Network. *Memetic Comput.* **2025**, *17*, 23. [CrossRef]
28. Pei, J.-J.; Gong, L.-H.; Qin, L.-G.; Zhou, N.-R. One-to-many image generation model based on parameterized quantum circuits. *Digit. Signal Process.* **2025**, *165*, 105340. [CrossRef]
29. Sudha, D.; Anju, A.; Ezhilarasi, K. Enhanced deep learning and quantum variational classifier for large-scale data analysis. *Knowl. Based Syst.* **2025**, *300*, 114611. [CrossRef]
30. Boyer, M.; Brassard, G.; Hoeyer, P.; Tapp, A. Tight bounds on quantum searching. *Fortschritte Phys.* **1998**, *46*, 493–505. [CrossRef]
31. Zalka, C. Grover's quantum searching algorithm is optimal. *Phys. Rev. A* **1999**, *60*, 2746. [CrossRef]
32. AbuGhanem, M. Comprehensive characterization of three-qubit Grover search algorithm on IBM's 127-qubit superconducting quantum computers. *arXiv* **2024**, arXiv:2406.16018v1.
33. Chen, J.-S.; Nielsen, E.; Ebert, M.; Inlek, V.; Wright, K.; Chaplin, V.; Maksymov, A.; Páez, E.; Poudel, A.; Maunz, P.; et al. Benchmarking a trapped-ion quantum computer with 30 qubits. *Quantum* **2024**, *8*, 1516. [CrossRef]
34. Wang, H.; He, Y.; Li, Y.-H.; Su, Z.-E.; Li, B.; Huang, H.-L.; Ding, X.; Chen, M.-C.; Liu, C.; Qin, J.; et al. High-efficiency multiphoton boson sampling. *Nat. Photonics* **2017**, *11*, 361–365. [CrossRef]
35. Kandala, A.; Mezzacapo, A.; Temme, K.; Takita, M.; Brink, M.; Chow, J.M.; Gambetta, J.M. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **2017**, *549*, 242–246. [CrossRef]
36. Farhi, E.; Goldstone, J.; Gutmann, S. A quantum approximate optimization algorithm. *arXiv* **2014**, arXiv:1411.4028. [CrossRef]
37. Havlíček, V.; Córcoles, A.D.; Temme, K.; Harrow, A.W.; Kandala, A.; Chow, J.M.; Gambetta, J.M. Supervised learning with quantum-enhanced feature spaces. *Nature* **2019**, *567*, 209–212. [CrossRef]
38. Kitaev, A.Y. Quantum computations: Algorithms and error correction. *Russ. Math. Surv.* **1997**, *52*, 1191–1249. [CrossRef]

39. Google Quantum AI. Suppressing quantum errors by repetition code and a bit-flip code in a superconducting quantum processor. *Nature* **2021**, *595*, 383–387.
40. Ryan-Anderson, C.; Bohnet, G.; Lee, K.; Gresh, D.; Hankin, A.; Gaebler, P.; Francois, D.; Chernoguzov, A.; Lucchetti, D.; Brown, C.; et al. Realization of real-time fault-tolerant quantum error correction. *Phys. Rev. X* **2022**, *11*, 041058.
41. Zheng, Y.; Ren, Q. Encoding Genetic and Structural Information in DNA Using Electric Field Gradients and Nuclear Spins. *Intell. Comput.* **2024**, *3*, 0094. [CrossRef]
42. Zenesini, A.; Berti, A.; Cominotti, R.; Rogora, C.; Moss, I.G.; Billam, T.P.; Carusotto, I.; Lamporesi, G.; Recati, A.; Ferrari, G. Quantum coherence in DNA nitrogen nuclear spins revealed by electric field gradients. *Nat. Phys.* **2024**, *20*, 445–452.
43. Press, W.H.; Hawkins, J.A.; Jones, S.K., Jr.; Finkelstein, I.J. HEDGES error-correcting code for DNA storage corrects indels and allows sequence constraints. *Biophys. Comput. Biol.* **2020**, *117*, 18489–18496. [CrossRef] [PubMed]
44. Organick, L.; Chen, Y.-J.; Ang, S.D.; Lopez, R.; Liu, X.; Strauss, K.; Ceze, L. Probing the physical limits of reliable DNA data retrieval. *Nat. Commun.* **2020**, *11*, 616. [CrossRef]
45. Farooq, U.; Singh, P.; Kumar, A. A systematic review of quantum image processing: Representation, applications and future perspectives. *Comput. Sci. Rev.* **2025**, *57*, 100763. [CrossRef]
46. Li, M.; Song, X.; El-Latif, A.A.A. EQIRHSI: Enhanced quantum image representation using entanglement state encoding in the HSI color model. *Quantum Inf. Process.* **2023**, *22*, 1. [CrossRef]
47. Haque, M.E.; Paul, M.; Ulhaq, A.; Debnath, T. Advance quantum image representation and compression using DCTEFRQI approach. *arXiv* **2022**, arXiv:2208.14277v1. [CrossRef]
48. Nielsen, M.A.; Chuang, I.L. *Quantum Computation and Quantum Information: 10th Anniversary Edition*; Cambridge University Press: Cambridge, UK, 2010. [CrossRef]
49. Watrous, J. Understanding Quantum Information and Computation. *arXiv* **2025**, arXiv:2507.11536v1. [CrossRef]
50. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the International Conference for Learning Representations, San Diego, CA, USA, 7–9 May 2015.
51. Spall, J.C. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. Autom. Control* **1992**, *37*, 332–341. [CrossRef]
52. Arrasmith, A.; Cerezo, M.; Czarnik, P.; Cincio, L.; Coles, P.J. Effect of barren plateaus on gradient-free optimization. *Quantum* **2021**, *5*, 558. [CrossRef]
53. Kosuri, S.; Church, G.M. Large-scale de novo DNA synthesis: Technologies and applications. *Nat. Rev. Genet.* **2014**, *15*, 443–450. [CrossRef]
54. Shendure, J.; Balasubramanian, S.; Church, G.M.; Gilbert, W.; Rogers, J.; Schloss, J.A.; Waterston, R.H. DNA sequencing at 40: Past, present and future. *Nature* **2017**, *550*, 345–353. [CrossRef]
55. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **1993**, *362*, 709–715. [CrossRef]
56. Takahashi, C.N.; Nguyen, B.H.; Strauss, K.; Ceze, L. Demonstration of End-to-End Automation of DNA Data Storage. *Sci. Rep.* **2019**, *9*, 4998. [CrossRef] [PubMed]
57. IBM. IBM Quantum Development Roadmap. Available online: https://www.ibm.com/roadmaps/quantum/ (accessed on 30 November 2025).
58. Google Quantum AI. Suppressing quantum errors by scaling a surface code logical qubit. *Nature* **2023**, *614*, 676–681. [CrossRef] [PubMed]
59. Lee, H.H.; Kalhor, R.; Goela, N.; Bolot, J.; Church, G.M. Terminator-free template-independent enzymatic DNA synthesis for digital information storage. *Nat. Commun.* **2019**, *10*, 2383. [CrossRef] [PubMed]
60. Catalog Technologies. DNA-Based Digital Data Storage Platform. Available online: https://www.catalogdna.com (accessed on 30 November 2025).