Lightweight Post-Quantum Cryptography for Critical Embedded Systems in Defense, Health, and Finance

Abstract

The advent of cryptographically relevant quantum computers poses an existential threat to the security foundations of our digital world. Public-key cryptosystems such as RSA and ECC, which underpin secure communication and data protection, will be rendered obsolete by quantum attacks like Shor's algorithm. This impending reality creates a particularly acute challenge for embedded systems, the specialized, resource-constrained computers that are ubiquitous in critical infrastructure. These devices—ranging from military communication systems and implantable medical devices to financial payment terminals—are characterized by limited processing power, minimal memory, strict energy budgets, and long operational lifecycles, making the direct implementation of computationally intensive standard post-quantum cryptography (PQC) algorithms often infeasible. This research review addresses the development of lightweight PQC algorithms tailored for these constrained environments. It begins by establishing the dual imperatives of quantum resistance and lightweight design, analyzing the families of PQC algorithms and the pivotal standardization process led by the U.S. National Institute of Standards and Technology (NIST). We conduct a comprehensive performance analysis of the finalized NIST standards, including ML-KEM (Kyber) and ML-DSA (Dilithium), on relevant embedded architectures such as ARM Cortex-M4 and RISC-V, presenting concrete benchmark data on execution speed and memory consumption. This technical foundation is then applied to a detailed, sector-specific analysis of Defense, Health, and Finance. We examine the unique security requirements, regulatory landscapes, and migration strategies for each sector, revealing distinct challenges and recommending tailored solutions, from hardware-software co-design and crypto-agility in defense to cloud-offloading architectures and novel ultra-lightweight signature schemes for medical implants. The review concludes that while the path to a quantum-resilient embedded ecosystem is complex, it is viable through a combination of algorithmic optimization, architectural innovation, and strategic planning. Key future research must focus on developing efficient side-channel attack countermeasures and standardizing lightweight PQC optimizations to ensure a secure and interoperable future.

1. Introduction: The Twin Imperatives of Quantum Resistance and Lightweight Implementation

The digital infrastructure that underpins modern society is built upon a foundation of cryptographic trust. This foundation is now facing a seismic shift, driven by the concurrent advancement of two powerful technological forces: the rise of quantum computing and the ever-expanding proliferation of resource-constrained embedded systems into every facet of critical infrastructure. The former threatens to shatter our current security paradigms, while the

latter imposes severe limitations on our ability to deploy the next generation of defenses. This confluence creates a complex and urgent problem space, demanding solutions that are not only mathematically robust against future threats but also computationally efficient enough to operate within the stringent confines of miniature, low-power devices. This review examines the frontier of this challenge: the development and deployment of lightweight post-quantum cryptography.

1.1 The Inevitable Quantum Threat: Shor's Algorithm and the "Harvest Now, Decrypt Later" Crisis

Post-Quantum Cryptography (PQC), also known as quantum-resistant or quantum-safe cryptography, refers to the development of cryptographic algorithms that are secure against cryptanalytic attacks by both conventional and quantum computers. The need for PQC stems from the profound vulnerability of our current public-key infrastructure (PKI). Widely deployed public-key algorithms, most notably Rivest-Shamir-Adleman (RSA) and Elliptic Curve Cryptography (ECC), derive their security from the presumed intractability of certain mathematical problems for classical computers. Specifically, the security of RSA relies on the difficulty of integer factorization, while ECC's security is based on the difficulty of solving the elliptic curve discrete logarithm problem (ECDLP). For decades, these mathematical foundations have remained solid, as the resources required for a classical computer to solve them for sufficiently large key sizes are astronomical, estimated to take billions of years.
However, the theoretical landscape of computation was irrevocably altered in 1994 when mathematician Peter Shor developed a quantum algorithm capable of solving both the integer factorization and discrete logarithm problems in polynomial time. A sufficiently powerful quantum computer, leveraging the principles of quantum mechanics like superposition and entanglement to process a vast number of potential solutions simultaneously, could execute Shor's algorithm to break RSA and ECC encryption with ease, rendering much of our secure digital communication infrastructure obsolete. While large-scale, cryptographically relevant quantum computers (CRQCs) capable of executing this attack do not yet exist, the steady progress in quantum hardware development has transformed this threat from a distant academic concern into an urgent national security and economic issue.
The urgency is massively amplified by the "Harvest Now, Decrypt Later" (HNDL) attack strategy, also referred to as "store now, decrypt later". In this scenario, adversaries are currently capturing and storing vast quantities of encrypted data transmitted today. While they cannot decrypt this information with classical computers, they are stockpiling it with the expectation of decrypting it retroactively once a CRQC becomes operational. This paradigm creates an immediate and critical threat to any sensitive data with a long "shelf-life"—such as national security secrets, intellectual property, financial records, and personal health information—that must remain confidential for years or decades. The HNDL threat means that the migration to PQC cannot wait for the physical arrival of a CRQC; it must begin now to protect the long-term confidentiality of today's data.

1.2 The Ubiquity of Embedded Systems in Critical Infrastructure

Parallel to the rise of the quantum threat, the technological landscape has been reshaped by the silent proliferation of embedded systems. An embedded system is a specialized computing system, comprising both hardware and software, that is designed to perform a dedicated function within a larger mechanical or electrical system. Unlike general-purpose computers, they

are optimized for specific tasks, often with real-time computing constraints. These devices form the invisible backbone of modern critical infrastructure, and their presence in the sectors targeted by this review is pervasive and indispensable.

In the **Defense** sector, embedded systems are the core of modern military platforms. They are found in the avionics and fly-by-wire flight control systems of advanced aircraft, the mission control and guidance systems of missiles and unmanned aerial vehicles (UAVs), and the secure communication links of software-defined radios (SDRs). In the **Health** sector, the Internet of Medical Things (IoMT) relies on embedded systems for patient care and monitoring. This includes implantable medical devices (IMDs) such as pacemakers, cardiac defibrillators, and insulin pumps, as well as a vast array of wearable health monitors that transmit sensitive patient data wirelessly. In the **Finance** sector, embedded systems are at the heart of the payment ecosystem, powering millions of Point-of-Sale (POS) terminals and Automated Teller Machines (ATMs) that process financial transactions daily.

These systems share a set of defining characteristics that make securing them uniquely challenging. They often have extremely long service lifecycles, with devices in defense or industrial control expected to remain in the field for 10, 20, or even more years. This longevity means that a device deployed today must be secure against the threats of tomorrow. Furthermore, many of these systems are physically deployed in remote or accessible locations, making them difficult to update or service. A failed software update on a smartphone is an inconvenience; a failed update on a pacemaker or a missile guidance system can have catastrophic consequences. Their direct interaction with the physical world means that a security compromise can lead not just to data loss, but to significant asset damage, personal injury, or even death.

1.3 The Confluence of Challenges: Why Standard PQC is Unsuitable for Constrained Devices

The intersection of the quantum threat and the proliferation of embedded systems creates a formidable security challenge. While PQC offers a solution to the former, the algorithms themselves introduce new problems for the latter. In general, PQC algorithms exhibit different performance characteristics than their classical RSA and ECC counterparts. They often involve more complex mathematical operations and demand significantly more resources, resulting in larger public keys, ciphertexts, and signatures, as well as potentially higher computational and memory overhead.

This increase in resource requirements clashes directly with the fundamental nature of the embedded systems used in critical sectors. These devices are, by design, severely resource-constrained. Their hardware is optimized for low cost and low power consumption, not high-performance computation. They operate with limited processing power from low-frequency microcontrollers (MCUs), minimal Random Access Memory (RAM) for runtime operations, and restricted Read-Only Memory (ROM) or Flash for code storage, often measured in mere kilobytes (KiB). For battery-powered devices, such as IMDs or remote sensors, energy consumption is the paramount constraint, as every CPU cycle and memory access drains a finite power source, directly impacting the device's operational lifespan.

This fundamental mismatch means that a naive migration strategy—simply replacing classical algorithms like ECC with standard implementations of PQC algorithms like CRYSTALS-Kyber—is infeasible for a vast and critical class of embedded devices. The computational load could overwhelm the processor, the memory footprint could exceed the

available RAM or ROM, and the energy drain could render a battery-powered device useless in a fraction of its intended lifespan. This reality necessitates a dedicated field of research focused on **lightweight post-quantum cryptography**: the design, optimization, and implementation of quantum-resistant algorithms specifically tailored to the severe constraints of embedded environments.

1.4 Review Objectives and Structure

This research review aims to provide a comprehensive and exhaustive analysis of the state-of-the-art in developing and implementing lightweight PQC for resource-constrained embedded systems, with a specific focus on the critical sectors of Defense, Health, and Finance. The objective is to synthesize the disparate fields of quantum-resistant algorithm design, embedded systems engineering, and sector-specific regulatory analysis into a unified, foundational reference for researchers, engineers, and policymakers.
To achieve this, the paper is structured as follows. Section 2 provides the necessary cryptographic background, detailing the families of PQC algorithms and analyzing the outcomes of the pivotal multi-year standardization process led by the U.S. National Institute of Standards and Technology (NIST). Section 3 formally defines the "lightweight" context, outlining the engineering constraints of embedded systems and the key performance metrics used to evaluate cryptographic solutions. Section 4 presents the empirical core of the review, analyzing real-world benchmark data for leading PQC algorithms on relevant embedded processor architectures, namely ARM Cortex-M4 and RISC-V, to assess their practical feasibility. Section 5 applies these technical findings in a deep-dive analysis of the Defense, Health, and Finance sectors, examining their unique operational landscapes, regulatory requirements, and strategic priorities for PQC migration. Section 6 explores advanced implementation strategies, such as crypto-agility and hybrid cryptography, and surveys the research frontier of novel lightweight PQC proposals. Finally, Section 7 synthesizes the key findings of the entire review, charting a course for a secure, quantum-resilient embedded future and identifying the most pressing challenges for future work.

2. The Post-Quantum Cryptographic Landscape: Algorithms and Standardization

The global effort to transition to a quantum-resistant cryptographic infrastructure is built upon two pillars: the development of new mathematical approaches to public-key cryptography and a rigorous, public process to standardize the most promising candidates. Understanding these two facets is essential to appreciating the options and challenges facing embedded system designers. The new algorithms offer a diverse set of trade-offs between security, performance, and size, while the standardization process provides the trust and interoperability necessary for global deployment.

2.1 A Taxonomy of PQC Families

PQC research has explored several distinct families of algorithms, each based on a different underlying mathematical problem believed to be hard for both classical and quantum computers to solve. This diversity is a strategic asset, as it hedges against the risk of a single mathematical breakthrough compromising the entire PQC ecosystem. The primary families include:

- **Lattice-Based Cryptography (LBC):** This is currently the most prominent and promising family of PQC algorithms. Its security is based on the geometric difficulty of problems on high-dimensional lattices, such as the Shortest Vector Problem (SVP) or the Learning with Errors (LWE) problem and its algebraic variants, Ring-LWE and Module-LWE. LBC schemes have emerged as leaders in the NIST standardization process due to their strong security proofs and, most importantly, their excellent balance of performance, key sizes, and signature sizes, making them highly suitable for a wide range of applications. The primary NIST standards, ML-KEM and ML-DSA, are both derived from this family.
- **Hash-Based Cryptography:** This family builds digital signatures using only the security of a cryptographic hash function (like SHA-256), which is a very well-understood and trusted primitive. The security assumption is minimal: as long as the hash function is collision-resistant and one-way, the signature is secure. This makes hash-based signatures, such as the Lamport signature scheme and its advanced derivatives like SPHINCS+, highly conservative and trustworthy choices. Their primary drawbacks are that they often produce very large signatures and can be "stateful," meaning a private key can only be used to sign a limited number of messages. However, stateless variants like SPHINCS+ overcome this limitation at the cost of performance and even larger signatures.
- **Code-Based Cryptography:** This is one of the oldest PQC approaches, with the McEliece cryptosystem being proposed in 1978. Its security relies on the difficulty of decoding a general linear error-correcting code. The original McEliece scheme, based on Goppa codes, has withstood four decades of scrutiny without a practical attack being found. This long history provides a high degree of confidence in its security. The primary disadvantage of code-based schemes is their extremely large public keys, which can be on the order of megabytes, posing a significant challenge for many applications, especially in constrained environments.
- **Multivariate Cryptography:** This approach bases its security on the difficulty of solving systems of multivariate polynomial equations over a finite field. While attempts to build encryption schemes from this family have largely failed, multivariate signature schemes like Rainbow have shown promise and were finalists in the NIST process.
- **Isogeny-Based Cryptography:** This is a relatively newer family that uses the complex map (an isogeny) between different elliptic curves as its hard problem. Schemes like the Supersingular Isogeny Diffie-Hellman (SIDH) were initially very attractive because they offered some of the smallest key sizes among all PQC candidates. However, the field suffered a major setback when significant cryptanalytic attacks were discovered against SIDH and its NIST finalist variant, SIKE, raising serious concerns about the maturity and security of the underlying assumptions.
- **Symmetric Key Quantum Resistance:** It is important to note that not all cryptography is broken by quantum computers. Symmetric key algorithms, such as the Advanced Encryption Standard (AES), are considered largely resistant to quantum attacks. The most effective known quantum attack against symmetric ciphers, Grover's algorithm, provides only a quadratic speedup, which effectively halves the key's security strength. This threat can be readily mitigated by simply doubling the key length—for example, by migrating from AES-128 to AES-256. This ensures that symmetric cryptography remains a robust and essential tool in the post-quantum era, particularly for bulk data encryption after a shared key has been established using a PQC Key Encapsulation Mechanism (KEM).

2.2 The NIST Standardization Process: A Multi-Year Global Effort

Recognizing the impending quantum threat, the U.S. National Institute of Standards and Technology (NIST) initiated a public, competition-like process in 2016 to solicit, evaluate, and standardize one or more quantum-resistant public-key cryptographic algorithms. The process began with an open call for proposals, which yielded 82 submissions from around the world, 69 of which were accepted as complete and proper candidates for the first round of evaluation in December 2017.

The NIST PQC standardization process was structured as a multi-round tournament designed to subject the candidate algorithms to intense public scrutiny from the global cryptographic community. Over several years, researchers and cryptanalysts analyzed the candidates for security flaws, implementation vulnerabilities, and performance characteristics. After each round, NIST, considering the public feedback and its own internal analysis, would select a smaller group of algorithms to advance to the next round.

- **Round 1 (2017-2019):** Whittled 69 candidates down to 26.
- **Round 2 (2019-2020):** Reduced the field to 7 finalists and 8 alternate candidates.
- **Round 3 (2020-2022):** Led to the selection of the first algorithms for standardization.

This open and transparent process was critical for building international consensus and trust in the final selected algorithms. It ensured that the standards would not only be mathematically sound but also practical to implement and interoperable on a global scale, a prerequisite for securing commercial hardware, software, and the broader internet.

2.3 Analysis of the Finalized NIST Standards (Published August 2024)

In August 2024, after years of rigorous evaluation, NIST published the first three finalized PQC standards. These algorithms form the initial bedrock of the global migration to quantum-resistant cryptography and are ready for immediate use.

2.3.1 FIPS 203: ML-KEM (Module-Lattice-Based Key-Encapsulation Mechanism)

- **Source and Purpose:** FIPS 203 specifies the ML-KEM algorithm, which is derived from the CRYSTALS-Kyber submission. It is designated as the primary standard for general-purpose encryption and key establishment. ML-KEM is a Key Encapsulation Mechanism (KEM), a primitive used to establish a secure shared secret between two parties over an insecure channel. This shared secret is then typically used with a highly efficient symmetric algorithm like AES to encrypt bulk data. It is the designated replacement for classical key exchange mechanisms like RSA encryption and Elliptic Curve Diffie-Hellman (ECDH).
- **Security and Parameters:** The security of ML-KEM is based on the hardness of the Module Learning with Errors (MLWE) problem in algebraic lattices. The standard specifies three parameter sets offering increasing levels of security: ML-KEM-512 (NIST Level 1, comparable to AES-128), ML-KEM-768 (NIST Level 3, comparable to AES-192), and ML-KEM-1024 (NIST Level 5, comparable to AES-256).
- **Key Characteristics:** ML-KEM was selected as the primary KEM due to its excellent all-around performance, strong security foundations, and relatively small key and ciphertext sizes, making it suitable for a wide variety of applications.

2.3.2 FIPS 204: ML-DSA (Module-Lattice-Based Digital Signature Algorithm)

- **Source and Purpose:** FIPS 204 specifies the ML-DSA algorithm, derived from the CRYSTALS-Dilithium submission. It is the primary standard for digital signatures, which are used to provide data integrity and authenticate the identity of a signer. ML-DSA is the designated replacement for classical signature schemes like RSA and the Elliptic Curve Digital Signature Algorithm (ECDSA).
- **Security and Parameters:** Like ML-KEM, its security is based on the hardness of problems over module lattices. The standard specifies three parameter sets: ML-DSA-44, ML-DSA-65, and ML-DSA-87, which correspond to NIST security strength categories 2, 3, and 5, respectively.
- **Key Characteristics:** ML-DSA was chosen for its strong performance across key generation, signing, and verification, and for offering a good balance between signature size and computational speed.

2.3.3 FIPS 205: SLH-DSA (Stateless Hash-Based Digital Signature Algorithm)

- **Source and Purpose:** FIPS 205 specifies the SLH-DSA algorithm, derived from the SPHINCS+ submission. It is also a standard for digital signatures but is intended to serve as a backup to ML-DSA.
- **Security and Parameters:** SLH-DSA's security is based entirely on the properties of its underlying cryptographic hash functions. This is considered a very conservative security assumption, as hash functions are among the most studied primitives in modern cryptography. Its security does not depend on any structured mathematical problems like those in lattices or number theory.
- **Key Characteristics:** The main rationale for standardizing SLH-DSA is to provide cryptographic diversity. Should an unexpected vulnerability be discovered in the mathematical foundations of lattice-based cryptography, SLH-DSA provides a secure alternative based on completely different principles. This resilience comes at a cost: SLH-DSA is generally slower and produces significantly larger signatures than ML-DSA.

The selection of these three initial standards reveals a clear and deliberate risk management strategy on the part of NIST. By choosing high-performance lattice-based algorithms (ML-KEM and ML-DSA) as the primary workhorses, the standards address the practical needs of most applications. Simultaneously, by standardizing a hash-based signature (SLH-DSA) and pursuing a code-based KEM, NIST is actively building a diverse portfolio. This "portfolio approach" is designed to mitigate the systemic risk of relying on a single family of mathematical problems. If a new quantum or classical algorithm were to emerge that weakens the assumptions behind lattice-based cryptography, the entire digital ecosystem would not be compromised, as alternative standards based on different hard problems would be available. This foresight, however, places a greater burden on system designers, particularly in the embedded space. The goal is no longer to simply pick the "best" algorithm, but to build systems with sufficient flexibility—or "crypto-agility"—to potentially support multiple algorithms from different families, each with its own unique performance and resource footprint. This makes the engineering challenge of the PQC transition significantly more complex.

2.4 The Role of Alternate and Future Candidates

NIST's standardization effort is a continuous process aimed at expanding and strengthening the PQC portfolio over time. Several other algorithms play a significant role in this ongoing strategy.

- **FALCON:** Another lattice-based signature scheme, FALCON was also selected for standardization, with a draft standard (FIPS 206) expected in late 2024. FALCON is notable for offering extremely compact signatures, among the smallest of any PQC candidate. However, its internal operations rely on floating-point arithmetic, which is more complex to implement correctly and securely on constrained devices compared to the integer arithmetic of Dilithium.
- **Fourth-Round KEMs and HQC:** To further diversify the KEM portfolio, NIST advanced several non-lattice-based candidates to a fourth round of evaluation, including the code-based schemes BIKE, Classic McEliece, and HQC. In March 2025, NIST announced its selection of **HQC (Hamming Quasi-Cyclic)** for standardization as a code-based alternative to the lattice-based ML-KEM. This is a critical development, as it provides a second, mathematically distinct foundation for quantum-resistant key establishment. A draft standard for HQC is anticipated in 2026, with final publication expected in 2027.
- **Additional Signature "On-Ramp":** Recognizing the need for even more signature options, particularly those with different trade-offs, NIST issued a new call for proposals in 2022. The call specifically sought signature schemes that are not based on structured lattices and that feature short signatures and fast verification times. This "on-ramp" process has accepted 40 new candidates into a first round of evaluation, signaling a vibrant and continuous effort to improve and diversify the suite of standardized digital signature algorithms.

3. The Lightweight Imperative: Constraints and Metrics for Embedded Systems

While the NIST standards provide a robust foundation for quantum-resistant security, their practical deployment hinges on their ability to function within the target environment. For a vast and growing number of critical applications, this environment is not a powerful server or desktop computer, but a highly constrained embedded system. The field of Lightweight Cryptography (LWC) has emerged specifically to address the unique challenges of securing these devices, focusing on a delicate balance between security, cost, and performance.

3.1 Defining the Resource-Constrained Environment

Lightweight Cryptography refers to the subfield of cryptography that designs algorithms and protocols to be extremely efficient in terms of computational resources, energy consumption, and implementation size. The goal is to provide robust security with minimal overhead, enabling secure communication and computation in environments where traditional cryptographic methods, including standard PQC implementations, would be too resource-intensive to be practical. The development of LWC is driven by the specific, severe constraints inherent to embedded systems. These defining constraints include:

- **Limited Processing Power:** Many embedded devices, particularly low-cost Internet of Things (IoT) nodes, are built around low-power microcontrollers (MCUs) with simple architectures (e.g., 8-bit or 32-bit cores like the ARM Cortex-M series) and low clock frequencies. These processors are incapable of executing computationally heavy algorithms without suffering from unacceptable latency or performance degradation that would impact their primary function.
- **Restricted Memory:** Memory is a premium resource in embedded design. Devices often have very limited non-volatile memory (ROM or Flash) for storing the program code and volatile memory (RAM) for runtime data storage, including the stack and heap. It is common for these devices to have only a few kilobytes (KiB) of RAM and tens or hundreds of KiB of Flash. This severely restricts the use of large cryptographic libraries, complex protocol stacks, or algorithms that require large intermediate state or key storage.
- **Energy Constraints:** For any battery-powered or energy-harvesting device, energy consumption is a critical design driver. This includes a vast range of applications from military field sensors and wearable health monitors to passive RFID tags. Every cryptographic operation consumes a portion of a finite energy budget, and high-energy algorithms can drastically reduce the operational lifespan of a device, necessitating frequent battery replacements (which may be impractical or, in the case of medical implants, require surgery) or rendering the device non-functional.
- **Physical Exposure and Side-Channel Vulnerabilities:** Unlike servers secured in data centers, embedded devices are often deployed in physically unsecured or even hostile environments. This makes them susceptible to physical tampering and, more subtly, side-channel attacks (SCA). SCAs exploit unintentional information leakage from the physical implementation of a cryptographic algorithm. By precisely measuring physical properties like power consumption, electromagnetic (EM) emissions, or execution timing during a cryptographic operation, an attacker can deduce secret key information without breaking the underlying mathematics of the algorithm. Designing LWC for embedded systems must therefore also consider resistance to these physical attack vectors.

3.2 Key Performance and Efficiency Metrics

To quantitatively evaluate and compare LWC and lightweight PQC algorithms, the research community has established a set of key performance indicators (KPIs) that directly map to the constraints of the target environment. A thorough analysis requires measuring performance across several dimensions:

- **Execution Speed (Latency & Throughput):** Speed is a primary concern, especially for real-time systems.
    - **Latency:** This is the time delay introduced by a cryptographic operation, typically measured in the number of CPU clock cycles required for completion. Lower latency is critical for time-sensitive applications like industrial control or vehicle-to-vehicle communication.
    - **Throughput:** This measures the rate at which data can be processed, typically expressed in bits or bytes per second. High throughput is important for applications that handle streams of data.

- **Memory Footprint (RAM & ROM):** This quantifies the memory resources required by the algorithm.
    - **RAM Usage:** This measures the amount of volatile memory (stack and heap) required during the execution of a cryptographic primitive. It is a critical metric, as exceeding the available RAM will cause the device to fail.
    - **ROM/Flash Usage (Code Size):** This measures the amount of non-volatile memory needed to store the algorithm's executable code and any constant data. A smaller code size is desirable as it leaves more space for other application features on a memory-constrained device.
- **Hardware Implementation Cost (Gate Equivalents - GE):** For implementations targeting Application-Specific Integrated Circuits (ASICs) or Field-Programmable Gate Arrays (FPGAs), the primary metric for cost is the circuit area. This is often expressed in Gate Equivalents (GE), a normalized unit representing the area of a basic logic gate. A smaller GE count indicates a smaller, cheaper, and typically lower-power hardware implementation.
- **Energy Consumption:** For battery-powered devices, this is arguably the most important metric. It is typically measured as the total energy consumed per cryptographic operation (e.g., in microjoules, μJ). Energy is the integral of power over time, so it is directly affected by both the algorithm's computational complexity (which influences execution time) and its hardware implementation efficiency (which influences power draw).

3.3 The Security-Efficiency Trade-off

A fundamental principle that governs the design of all lightweight cryptography is the trade-off between three competing goals: **security**, **cost** (area, energy), and **performance** (speed). It is generally straightforward to optimize for any two of these goals at the expense of the third. For example, one can achieve very high security and high performance with a large, power-hungry hardware accelerator, or one can achieve very low cost and high security with a very slow algorithm. The central challenge of LWC design is to find an optimal balance point that optimizes all three simultaneously for a given application context.

This trade-off often manifests in the choice of security level. While mainstream applications might default to the highest security levels (e.g., 256-bit security), this may be overkill and prohibitively expensive for a constrained device. LWC design often targets "adequate" or "sufficient" security levels, such as 80-bit or 128-bit security, which are deemed strong enough to protect the data for its required lifetime given the value of the asset and the capabilities of a potential attacker. For example, 64-bit to 80-bit security may suffice for simple one-way authentication on an RFID tag with a short operational life, whereas 128-bit security is a more typical target for mainstream IoT applications. This pragmatic approach to security is essential for making the deployment of cryptography feasible in the most constrained corners of the embedded world.

4. Performance Analysis of PQC on Embedded Architectures

The theoretical suitability of a PQC algorithm for embedded systems can only be confirmed through empirical analysis. Rigorous benchmarking on relevant hardware platforms is essential to quantify the real-world performance overheads in terms of speed, memory, and energy consumption. This section delves into the performance of leading PQC candidates on two key embedded architectures: the ubiquitous ARM Cortex-M series, representing the current state of commercial-off-the-shelf (COTS) microcontrollers, and the burgeoning RISC-V architecture, which represents the future of customizable, open-standard embedded processing.

4.1 Benchmarking on ARM Cortex-M4: Insights from the pqm4 Framework

The ARM Cortex-M family, particularly the Cortex-M4, is a dominant 32-bit microcontroller architecture found in a vast array of embedded devices. Its balance of performance, power efficiency, and feature set, including Digital Signal Processing (DSP) instructions, has made it a popular choice for moderately constrained applications. Recognizing its importance, NIST officially designated the ARM Cortex-M4 as a primary optimization target for PQC candidates, making performance on this platform a key evaluation criterion.

The de-facto standard for evaluating PQC performance on this architecture is the **pqm4** project. pqm4 is an open-source testing and benchmarking framework that provides a unified, fair, and reproducible environment for comparing PQC implementations. Its methodology is rigorous:

- **Physical Hardware Testing:** Benchmarks are run on widely available, low-cost development boards (such as the STM32F4Discovery or the newer Nucleo-L4R5ZI) to capture real-world performance, rather than relying on potentially inaccurate simulations.
- **Accurate Cycle Counting:** It uses on-chip hardware timers, like the 24-bit SysTick counter with an overflow interrupt handler, to precisely measure the number of CPU clock cycles for each cryptographic operation (key generation, encapsulation/signing, decapsulation/verification). Benchmarks are run at a reduced clock frequency (e.g., 24 MHz) to eliminate memory wait states, ensuring results are independent of memory controller speed and directly reflect the computational cost of the algorithm.
- **Stack Usage Measurement:** pqm4 measures the dynamic stack memory required by each primitive using a technique called "stack spraying," where a canary pattern is written to the stack space before execution and checked afterward to see how much was overwritten.
- **Unified Primitives:** To ensure fair comparisons between different PQC schemes, pqm4 provides a common, highly optimized library for underlying symmetric primitives like AES and SHA-3 (Keccak), so that performance differences reflect the PQC algorithms themselves, not their underlying hash function implementations.

The performance of the NIST-standardized algorithms on the ARM Cortex-M4 provides a crucial baseline for their feasibility in a vast number of existing and future embedded products. The following tables summarize the performance of optimized implementations (m4f for Kyber, m4f for Dilithium, and m4-ct for Falcon), which typically use hand-tuned assembly to leverage the Cortex-M4's specific instruction set for maximum efficiency.

Table 1: PQC KEM Performance on ARM Cortex-M4 (m4f/speed Implementations)

Scheme	Operation	Mean Cycle Count	Stack Usage (Bytes)
ML-KEM-512	KeyGen	392,423	4,372
	Encaps	390,881	5,436
	Decaps	428,167	5,412
ML-KEM-768	KeyGen	642,096	5,396
	Encaps	658,754	6,468
	Decaps	707,827	6,452
ML-KEM-1024	KeyGen	1,018,976	6,436
	Encaps	1,031,565	7,500
	Decaps	1,094,008	7,484

Data sourced and synthesized from pqm4 project benchmarks. Cycle counts are the mean of 10 executions. Stack usage is for the cryptographic primitive only.

The results for ML-KEM (Kyber) in Table 1 demonstrate its remarkable efficiency on a constrained platform. For the lowest security level (ML-KEM-512), all operations complete in under 430,000 cycles, with a stack usage of around 5.4 KiB. This level of performance makes it entirely feasible for a wide range of embedded applications that require secure key exchange. As expected, the resource requirements scale with the security level, with ML-KEM-1024 requiring roughly 2.5 times the cycles and a moderately larger memory footprint. This data clearly illustrates the security-performance trade-off that designers must navigate.

Table 2: PQC Signature Performance on ARM Cortex-M4 (m4f/m4-ct Implementations)

Scheme	Operation	Mean Cycle Count	Stack Usage (Bytes)
ML-DSA-44	KeyGen	1,425,492	38,296
	Sign	3,822,701	49,424
	Verify	1,421,600	8,912
ML-DSA-65	KeyGen	2,516,006	60,824
	Sign	6,193,171	68,872
	Verify	2,415,944	9,888
ML-DSA-87	KeyGen	4,274,513	97,688
	Sign	8,204,023	116,084
	Verify	4,193,228	12,060
Falcon-512	KeyGen	146,357,328	1,148
	Sign	40,191,597	2,428
	Verify	482,280	376
Falcon-1024	KeyGen	408,725,773	1,156
	Sign	87,706,019	2,508
	Verify	990,541	376

Data sourced and synthesized from pqm4 project benchmarks. Cycle counts are the mean of 1000 executions for Dilithium and 10 for Falcon. Stack usage is for the cryptographic primitive only.

Table 2 highlights the starkly different performance profiles of the two main signature standards. ML-DSA (Dilithium) offers a balanced performance, with key generation, signing, and verification times all in the low millions of cycles for the base security level. However, its stack usage is

substantial, with the signing operation for ML-DSA-87 requiring over 116 KiB of RAM, which would preclude its use on many smaller microcontrollers.
Falcon, in contrast, presents a highly asymmetric profile. Its key generation is extremely slow, taking hundreds of millions of cycles, making it unsuitable for on-device key generation in most scenarios. Signing is also computationally intensive. However, its verification is exceptionally fast and requires a minuscule amount of stack memory (under 1 KiB). This profile makes Falcon an excellent candidate for ecosystems where signatures are generated on a powerful server and verified on many constrained devices, such as in content delivery or broadcast authentication scenarios.

4.2 Benchmarking on RISC-V: The Frontier of Custom Acceleration

While ARM Cortex-M represents the established incumbent, the RISC-V architecture is rapidly gaining traction in the embedded space, offering a compelling alternative. RISC-V is a free and open-source Instruction Set Architecture (ISA), which means that anyone can design, manufacture, and sell RISC-V chips and software without paying licensing fees. Its most powerful feature for high-performance computing is its inherent modularity and extensibility. The base ISA can be extended with standardized or custom instruction set extensions (ISEs) to accelerate specific workloads. This capability for hardware-software co-design makes RISC-V a particularly promising platform for tackling the computational demands of PQC.
Software-only implementations of PQC algorithms on RISC-V processors reveal performance bottlenecks similar to those on ARM. The core lattice-based computations in Kyber and Dilithium are dominated by two main tasks: Keccak hashing (used for pseudorandom seed expansion and other functions) and the Number Theoretic Transform (NTT), which is an efficient algorithm for performing polynomial multiplication in a ring. On a 64-bit RISC-V core, Keccak can consume between 24% and 67% of the total cycles for Kyber, with NTT and other modular arithmetic operations accounting for most of the remainder. These computationally intensive routines are prime candidates for hardware acceleration.
Two primary strategies for acceleration have emerged in the RISC-V ecosystem:

1. **Loosely-Coupled Accelerators:** These are dedicated hardware blocks that function as co-processors, sitting alongside the main CPU core and communicating over a system bus. The CPU offloads an entire complex task, such as a full Keccak permutation or an NTT, to the accelerator, often using Direct Memory Access (DMA) to transfer data efficiently without CPU intervention. This approach offers great flexibility and can yield dramatic performance gains. Studies have shown speedups of 13x for Keccak and up to 20x for NTT/INVNTT compared to software-only implementations.
2. **Tightly-Coupled Accelerators (Custom Instructions):** This more integrated approach involves adding new, specialized instructions directly into the processor's pipeline. The RISC-V PQC Task Group is actively investigating ISEs for this purpose. Proposed instructions include a multi-round Keccak instruction (vkeccakp.wi) and a vector modular multiplication instruction (vmulq) specifically designed for the moduli used in Kyber and Dilithium. Projections show that such instructions could more than triple the performance of the NTT, significantly speeding up the most common PQC operations.

The performance impact of this architectural approach is profound, as demonstrated by benchmarking results comparing software-only execution with hardware-accelerated execution on RISC-V platforms.

Table 3: PQC Performance on RISC-V (Software vs. Hardware Acceleration for Kyber)

Scheme	Operation	Implementation	Mean Cycle Count	Achieved Speedup
Kyber-512	KeyGen	Software Only	1,052,145	1.0x
		HW Accelerated	292,660	**3.59x**
	Encaps	Software Only	1,106,228	1.0x
		HW Accelerated	365,167	**3.03x**
	Decaps	Software Only	1,231,155	1.0x
		HW Accelerated	460,374	**2.67x**
Kyber-768	KeyGen	Software Only	1,674,185	1.0x
		HW Accelerated	315,149	**5.31x**
	Encaps	Software Only	1,789,912	1.0x
		HW Accelerated	418,661	**4.29x**
	Decaps	Software Only	1,968,664	1.0x
		HW Accelerated	533,930	**3.69x**
Kyber-1024	KeyGen	Software Only	2,612,887	1.0x
		HW Accelerated	609,948	**4.29x**
	Encaps	Software Only	2,757,344	1.0x
		HW Accelerated	563,515	**4.89x**
	Decaps	Software Only	2,983,573	1.0x
		HW Accelerated	741,062	**4.02x**

Data sourced and synthesized from hardware acceleration studies on RISC-V. "Software Only" represents execution on a base RISC-V core. "HW Accelerated" represents the same core with loosely-coupled accelerators for Keccak and NTT.

The data in Table 3 provides compelling evidence for the power of hardware-software co-design. By offloading the most computationally intensive parts of the Kyber algorithm to dedicated hardware, overall performance is improved by a factor of 3-5x. This transforms the PQC implementation from merely feasible to highly efficient, enabling its use in applications with much stricter latency or throughput requirements.

This analysis reveals a fundamental divergence in the optimization pathways for the two leading embedded architectures. For a fixed-ISA platform like ARM Cortex-M, performance gains are achieved through meticulous software engineering and hand-tuned assembly code. While effective, these improvements are ultimately incremental. For an extensible ISA like RISC-V, the frontier of optimization lies in architectural innovation. The ability to design custom hardware accelerators or new instructions tailored to the specific bottlenecks of PQC algorithms offers a pathway to order-of-magnitude performance improvements. This suggests that the long-term trajectory for high-performance, lightweight PQC will likely favor these extensible architectures. Consequently, an organization's choice of PQC algorithm and its performance expectations may become deeply intertwined with its choice of processor architecture and hardware vendor, adding a new strategic layer to system design for critical applications.

5. Sector-Specific Analysis and Recommendations

While the technical benchmarks provide a universal measure of performance, the true test of a lightweight PQC algorithm is its suitability for a specific application context. The operational realities, risk tolerances, and regulatory frameworks of the Defense, Health, and Finance sectors are vastly different, leading to distinct challenges and priorities for their PQC migration. This section applies the preceding technical analysis to these three critical domains.

5.1 The Defense Sector: Securing Mission-Critical Systems with Long Lifecycles

5.1.1 The Landscape: High-Stakes, Long-Term Security

The defense sector is characterized by its reliance on a wide array of sophisticated embedded systems that are integral to mission success and national security. These include avionics and flight control systems in manned and unmanned aircraft (UAVs), software-defined radios (SDRs) for secure battlefield communications, guidance and fire-control systems for smart munitions and naval platforms, and satellite communication systems. The operational environment for these systems is demanding, requiring exceptional reliability, real-time performance, and resilience in physically contested or hostile settings. A defining feature of defense systems is their extremely long service lifecycle, often spanning 20 years or more. This means that cryptographic solutions deployed today must remain secure against the threats of the 2040s and beyond.

5.1.2 The Mandates: A Government-Driven Transition

Unlike commercial sectors, the PQC transition in defense is driven by explicit government mandates and a comprehensive cryptographic modernization strategy. Key directives and frameworks include:

- **The "Harvest Now, Decrypt Later" Imperative:** For the Department of Defense (DoD), the HNDL threat is not a theoretical risk but a clear and present danger to national security. Adversaries are presumed to be actively collecting classified and sensitive military communications today for future decryption. This makes the transition to PQC a non-negotiable, near-term priority to protect the long-term confidentiality of state secrets.
- **The DoD PQC Migration Roadmap:** The U.S. government, through agencies like the NSA, CISA, and OMB, has established a clear roadmap for migrating federal information systems to PQC. The core tenets of this strategy are to:
    1. **Create a comprehensive cryptographic inventory** to identify all systems using quantum-vulnerable algorithms.
    2. **Prioritize systems for migration** based on risk and the sensitivity of the data they protect.
    3. **Engage with vendors** to ensure the supply chain is prepared for the transition.
    4. **Build cryptographic agility** into systems to accommodate future changes.
- **DoD Instruction 8500.01 "Cybersecurity":** This foundational directive establishes a multi-tiered, risk-based approach to securing all DoD IT throughout its entire lifecycle, from acquisition and design to operation and disposal. PQC requirements must be identified and integrated into this lifecycle management process.

- **Commercial National Security Algorithm (CNSA) Suite 2.0:** The NSA defines the specific cryptographic algorithms approved for use in National Security Systems (NSS). CNSA 2.0 mandates the use of the new NIST PQC standards for protecting sensitive government information.

5.1.3 Analysis and Recommendations

The defense sector's requirements place a premium on long-term security assurance and risk mitigation. The long lifecycles of military hardware mean that crypto-agility is not just a best practice but a mission-critical necessity. Systems deployed today must be capable of being updated to new cryptographic standards that may emerge a decade from now.
For specific applications, the performance benchmarks from Section 4 provide clear guidance.

- **Authentication and Secure Updates:** For tasks like secure boot and over-the-air firmware updates in systems like UAVs or SDRs, a robust and efficient digital signature scheme is paramount. The benchmark data shows that **ML-DSA (Dilithium)** offers a strong, balanced performance profile suitable for these tasks.
- **Cryptographic Diversity:** The principle of cryptographic diversity is especially acute in national security contexts. Relying solely on lattice-based cryptography, even with its strong security case, introduces a potential single point of failure. Therefore, it is a prudent risk mitigation strategy for critical systems to also have the capability to use algorithms from different mathematical families. The standardization of the hash-based **SLH-DSA (SPHINCS+)** and the code-based **HQC** provides these alternatives. While SLH-DSA has larger signatures and is slower, its conservative security basis makes it an excellent choice for high-assurance functions like signing root-of-trust firmware. HQC provides a vital non-lattice alternative for key exchange.
- **Hardware Acceleration:** Given the need for high performance in real-time systems (e.g., missile guidance, electronic warfare), the defense sector is a prime candidate for leveraging hardware-accelerated PQC on platforms like RISC-V. Custom ISEs can provide the necessary performance while maintaining a low size, weight, and power (SWaP) profile.

5.2 The Health Sector: Protecting Patient Safety in Connected Medical Devices

5.2.1 The Landscape: Ultra-Constrained, Life-Critical Devices

The health sector's embedded systems landscape is dominated by the rapid growth of the Internet of Medical Things (IoMT). While this includes hospital equipment and bedside monitors, the most challenging devices from a security perspective are **implantable medical devices (IMDs)** and wearables. Devices like pacemakers, implantable cardioverter-defibrillators (ICDs), insulin pumps, and neurostimulators are life-sustaining and life-critical. Compromised authentication could lead to a failure to deliver a life-saving therapy, while a breach of confidentiality could expose highly sensitive patient data.
These devices operate under the most extreme resource constraints of any critical sector. The primary design drivers are **minimal physical size** and **extreme low-power operation**. Battery life is paramount, as replacement often requires an invasive surgical procedure, posing a direct risk to the patient's well-being. Consequently, the computational and energy overhead of any

security mechanism must be minimized to an absolute degree.

5.2.2 The Regulatory Framework: The FDA and Crypto-Agility

The key regulatory body for medical devices in the United States is the Food and Drug Administration (FDA). The FDA's approach to cybersecurity is guided by the principle of ensuring a "reasonable assurance of safety and effectiveness" throughout the device's entire lifecycle. The latest **FDA guidance, "Cybersecurity in Medical Devices" (June 2025)**, solidifies this approach.

- **Legal Authority:** The guidance implements Section 524B of the Federal Food, Drug, and Cosmetic (FD&C) Act, which grants the FDA explicit authority to require and enforce cybersecurity measures for "cyber devices" (devices with software and connectivity).
- **Lifecycle Management:** The FDA mandates a Secure Product Development Framework (SPDF) that integrates cybersecurity into all phases of the device lifecycle, from design to postmarket surveillance.
- **Crypto-Agility Mandate:** Crucially, the FDA guidance does **not** yet explicitly mandate the use of PQC algorithms. However, it strongly emphasizes the need for **crypto-agility**. Manufacturers must design devices to be **"secure and timely updatable and patchability"** and to be capable of mitigating emerging cybersecurity risks over their entire lifecycle. The guidance explicitly states that a change to a device's authentication or encryption algorithms may require a new premarket submission to the FDA. This effectively requires manufacturers to plan for a future PQC transition, especially for devices with long field lives.

5.2.3 Analysis and Recommendations

The extreme resource constraints of IMDs make the direct implementation of even optimized NIST PQC standards like Kyber and Dilithium highly impractical. The computational and energy overhead would unacceptably shorten battery life. This reality forces the health sector to pursue alternative strategies:

1. **Architectural Solutions (Cloud Offloading):** A highly promising approach is to offload the heavy cryptographic work from the IMD to a more powerful, external device like a patient's smartphone or a secure cloud service. In this model, the IMD would perform a computationally expensive PQC key exchange (e.g., using ML-KEM) only once—at initial pairing or on a very infrequent basis (e.g., daily). This one-time operation establishes a shared symmetric key. All subsequent communication between the IMD and the external system would then be protected using a highly efficient, low-power, hardware-accelerated symmetric cipher like AES. This architecture provides full quantum-resistant security for the channel while minimizing the cryptographic burden on the IMD itself.
2. **Novel Ultra-Lightweight Algorithm Design:** The research community is actively developing new PQC signature schemes specifically for this asymmetric use case, where the signer is severely constrained but the verifier is powerful. Schemes like **LiteQSign** and **INF-HORS** are hash-based signatures designed for near-optimal efficiency on the signer side. They require only a small, constant number of hash operations to generate a signature, dramatically reducing the computational and energy cost on the IMD. The verification process is more computationally intensive, but this workload is pushed to the hospital's servers or physician's programmer, which have ample resources. Performance results on an 8-bit MCU show these schemes can be up to 20 times faster at signing than

other PQC alternatives, making them a viable path for securing IMDs directly.

5.3 The Finance Sector: Fortifying the Transaction Ecosystem

5.3.1 The Landscape: A Vast, Interconnected Ecosystem

The financial sector's embedded systems landscape is dominated by the payment processing ecosystem, particularly the millions of Point-of-Sale (POS) terminals and Automated Teller Machines (ATMs) deployed globally. These devices are the frontline for capturing and transmitting sensitive cardholder data. Their security is paramount for maintaining the trust that underpins the entire financial system. The primary challenge in this sector is not necessarily the resource constraints of an individual device—a modern POS terminal is significantly more powerful than a pacemaker—but rather the sheer scale and interconnectedness of the ecosystem, which includes a complex web of merchants, acquiring banks, payment processors, and third-party service providers, many of whom rely on legacy systems.

5.3.2 The Standards: PCI DSS and FS-ISAC Guidance

The primary standards body for the payment industry is the **PCI Security Standards Council (SSC)**, founded by the major card brands. Its flagship standard, the **PCI Data Security Standard (PCI DSS)**, provides the baseline technical and operational requirements for all entities that store, process, or transmit cardholder data. Related standards like **PCI PTS POI** (PIN Transaction Security Point of Interaction) and **PCI P2PE** (Point-to-Point Encryption) provide specific security requirements for payment terminals and end-to-end encryption solutions, respectively.
To date, the PCI SSC has not issued a formal PQC standard or migration roadmap. However, the **Financial Services Information Sharing and Analysis Center (FS-ISAC)** has taken a proactive leadership role. Its PQC Working Group, composed of experts from across the financial industry, has published a series of influential papers to guide the sector's transition. This guidance includes:

- An analysis of the impact of quantum computing on the payment card industry.
- Technical papers on creating a cryptographic inventory, modeling quantum risk, and achieving crypto-agility.
- Detailed use cases for migrating critical infrastructure, including ATM and POS systems.

5.3.3 Analysis and Recommendations

The financial industry's PQC transition is primarily driven by risk management, the need to protect against future fraud, and the imperative to maintain consumer trust in the payment system. Given the complexity and legacy nature of the ecosystem, a "flag day" transition where all systems switch to PQC simultaneously is impossible.

- **Hybrid Cryptography:** This makes **hybrid cryptography** (discussed in detail in Section 6.2) a critical enabling technology for the financial sector. A hybrid approach, which combines a classical algorithm with a PQC algorithm, allows new PQC-enabled terminals to remain backward-compatible and interoperable with older parts of the network that have not yet been upgraded. This provides a gradual and manageable migration path.
- **Algorithm Performance:** The performance of digital signatures is crucial for the speed of transaction authorization. The benchmark data shows that **ML-DSA** offers a strong,

balanced performance that is well-suited for this purpose. The extremely fast verification times of **Falcon** could also be attractive, especially in online transaction scenarios where a merchant's server verifies a signature from a consumer's device.

- **Crypto-Agility:** As with the other sectors, achieving crypto-agility is a key strategic goal identified by FS-ISAC. The ability to update cryptographic protocols and algorithms across the vast network of payment terminals and backend systems is essential for long-term security management.

5.4 Synthesis of Critical Sector Requirements

The distinct operational contexts and regulatory pressures of the Defense, Health, and Finance sectors result in different strategic priorities for the adoption of lightweight PQC. Table 4 provides a synthesized comparison of these requirements.

Table 4: Synthesis of Critical Sector Requirements for Lightweight PQC

Feature	Defense Sector	Health Sector (IoMT/IMDs)	Finance Sector
Key Embedded Systems	UAVs, SDRs, missile guidance, tactical sensors, command & control systems	Implantable devices (pacemakers, ICDs, insulin pumps), wearable monitors	POS terminals, ATMs, hardware security modules (HSMs)
Primary Constraints	Long lifecycle (20+ years), high reliability, real-time operation, physical security	Extreme low power, minimal size, battery life, patient safety, difficult to update	High transaction volume, ecosystem interoperability, legacy system support, trust
Key Regulatory Body	DoD, NSA, NIST (via CNSA Suite)	FDA (Food and Drug Administration)	PCI SSC (Payment Card Industry Security Standards Council), FS-ISAC
Primary Migration Driver	National security (protecting long-lived state secrets from HNDL attacks)	Patient safety and data privacy (ensuring device integrity and functionality)	Risk management and maintaining consumer/ecosystem trust
Recommended PQC Strategies	Crypto-agility, cryptographic diversity (Lattice + Hash/Code), HW/SW co-design (RISC-V)	Cloud/device offloading of PQC operations, novel ultra-lightweight algorithms (e.g., LiteQSign)	Hybrid cryptography (for backward compatibility), crypto-agility, standardized protocols

6. Advanced Implementation Strategies and Future Research Directions

Successfully deploying lightweight PQC in critical embedded systems requires more than just selecting and optimizing an algorithm. It demands broader strategic thinking about system architecture, long-term maintainability, and the evolving nature of cryptographic threats. This

section explores advanced implementation strategies that are crucial for a successful migration and outlines the key research frontiers that will shape the next generation of secure embedded systems.

6.1 Crypto-Agility: Designing for an Evolving Threat Landscape

The single most important strategic principle to emerge from the global PQC transition effort is the need for **crypto-agility**. Formally, crypto-agility is the capacity of a computing system to be updated to new cryptographic algorithms, primitives, and protocols with minimal disruption. For decades, cryptographic standards were relatively static, and algorithms were often hardcoded deep within systems. The PQC migration represents the first large-scale, ecosystem-wide cryptographic transition in the modern internet era, and it has exposed the profound brittleness of this static approach.
The requirement for crypto-agility is now a central theme in the migration roadmaps for all three critical sectors.

- The **DoD** views agility as essential for managing the security of systems with multi-decade lifecycles and for responding to future unforeseen threats.
- The **FDA**'s guidance effectively mandates agility by requiring that medical devices be designed for secure updates and by noting that changing an encryption algorithm may trigger a new regulatory review, thus incentivizing manufacturers to build flexible architectures from the start.
- The **FS-ISAC** has identified crypto-agility as the core prerequisite for the financial sector to manage the transition while maintaining interoperability across a complex and heterogeneous network.

The PQC transition is therefore the primary catalyst forcing the entire technology industry to elevate crypto-agility from a theoretical best practice to a mandatory design requirement. The long-term benefit of this shift will extend far beyond quantum resistance; it will prepare the entire digital infrastructure for any future cryptographic transition, whether necessitated by new mathematical breakthroughs, implementation vulnerabilities, or evolving standards. In practice, achieving crypto-agility in embedded systems involves a commitment to modular software design, such as using cryptographic libraries that can be updated or "hot-swapped," and implementing robust, authenticated, and secure over-the-air (OTA) update mechanisms.

6.2 Hybrid Cryptography: Bridging the Classical and Quantum-Resistant Worlds

As sectors begin their migration, they will face a long transitional period where new, PQC-enabled systems must coexist and interoperate with legacy systems that only support classical cryptography. The **hybrid approach** has emerged as the key tactical solution to this challenge. Hybrid cryptography combines a classical algorithm (e.g., ECDH) with a PQC algorithm (e.g., ML-KEM) in parallel. In a key exchange, for instance, both algorithms are run, and the two resulting shared secrets are cryptographically combined (e.g., by concatenation followed by a key derivation function) to produce the final session key.
This approach offers two powerful benefits:

1. **Backward Compatibility and Interoperability:** A hybrid system can communicate with a PQC-only system, a classical-only system, or another hybrid system. Protocols like TLS can be designed to negotiate the strongest mutually supported algorithm set, falling back to classical-only modes when necessary to maintain connectivity with legacy infrastructure. This provides a smooth, gradual migration path without a disruptive "flag day."
2. **Risk Mitigation:** The security of a hybrid key exchange rests on the assumption that *at least one* of the constituent algorithms is secure. This provides a valuable hedge against the possibility that a new, yet-to-be-discovered flaw exists in the first generation of PQC standards. If ML-KEM were to be broken, the connection would remain secure thanks to the classical ECDH component. Conversely, if a CRQC comes online, the connection is protected by ML-KEM.

Of course, this dual-algorithm approach incurs a performance overhead. The system must bear the computational cost of both the classical and the PQC operations, and the public keys and/or ciphertexts transmitted will be larger. However, empirical analysis of Hybrid Public Key Encryption (HPKE) shows that this overhead can be manageable. One study found that a hybrid of X25519 (an ECC curve) and Kyber incurred a 52% performance overhead for encrypting a small 1 KB plaintext compared to classical-only HPKE. However, because the asymmetric operations are only performed once for the initial key exchange, the cost is amortized over the size of the data being transferred. For a 1 MB plaintext, the overhead of the hybrid scheme dropped to just 17%. This demonstrates that for many applications, particularly those involving bulk data transfer, the performance penalty of a hybrid approach is modest and a worthwhile price to pay for the benefits of interoperability and risk mitigation.

6.3 Novel Lightweight PQC Proposals: A Look at the Research Frontier

While the NIST standards provide a stable and trusted baseline for deployment, they represent a snapshot of the state-of-the-art from a few years ago. The cryptographic research community continues to innovate, developing new algorithms and techniques specifically optimized for the most constrained environments. These proposals, while not yet standardized, point toward the future of lightweight PQC and may offer even better performance trade-offs.

- **Rudraksh:** This is a lattice-based KEM that was designed from the ground up for lightweight hardware implementation. It makes several design choices to reduce resource usage, most notably using the ASCON lightweight hash function (the winner of NIST's separate LWC standardization project) instead of the more resource-intensive Keccak (SHA-3) used in Kyber. Its designers claim it can achieve a 3x improvement in hardware area compared to an area-optimized Kyber implementation while providing an equivalent NIST Level 1 security guarantee.
- **LiteQSign and INF-HORS:** These are two closely related hash-based digital signature schemes designed for the extreme asymmetric use case of devices like IMDs, where the signer is severely constrained but the verifier is not. They achieve near-optimal signature generation efficiency by requiring only a small, constant number of hash operations on the signing device. This pushes the bulk of the computational work to the verifier. Experiments on an 8-bit MCU showed that LiteQSign can generate signatures up to 20 times faster and with significantly higher energy efficiency than other PQC signature schemes, while also producing smaller keys and signatures.

- **LWPQC:** This proposal explores a tweakey-based block cipher architecture, a design paradigm that integrates a "tweak" along with the key into the cryptographic operation. This approach aims to provide enhanced flexibility and security while maintaining a lightweight implementation suitable for resource-constrained environments.

The existence of this active research frontier demonstrates that the field is not static. Future PQC standards and solutions may well incorporate these or other novel techniques to push the boundaries of efficiency even further.

6.4 Open Challenges and Unanswered Questions

Despite the significant progress in standardizing and benchmarking PQC, several major challenges remain that will require sustained research and development efforts.

- **Efficient Side-Channel Attack (SCA) Countermeasures:** While PQC algorithms are mathematically secure against quantum computers, their physical implementations are not inherently secure against physical attacks. Protecting implementations against SCAs and fault injection attacks is a critical and difficult problem. Standard countermeasures, such as masking, often introduce significant overhead in performance and memory, potentially doubling or tripling the cost of an operation. Developing efficient, low-overhead countermeasures specifically for the new mathematical structures found in PQC algorithms (e.g., lattice operations) is a major area of ongoing research.
- **Standardization of Lightweight Optimizations:** As shown in Section 4, the most significant performance gains often come from platform-specific assembly code or custom hardware accelerators. While effective, this creates a risk of a fragmented ecosystem where interoperability is compromised. There is a pressing need to standardize these optimizations, for example by defining official PQC instruction set extensions for architectures like RISC-V, to ensure that high-performance implementations remain interoperable and widely accessible.
- **Supply Chain Complexity and Coordination:** The PQC migration is not a task a single organization can accomplish in isolation. It requires deep coordination across a complex global supply chain, involving silicon vendors, hardware security module (HSM) manufacturers, device manufacturers, operating system developers, and application software providers. Ensuring that all components in the supply chain support the new PQC standards and can interoperate securely is a monumental logistical challenge that will require years of coordinated effort.

7. Conclusion: Charting a Course for a Quantum-Resilient Embedded Future

The journey toward a quantum-resilient digital world is one of the most significant cryptographic transitions in history. It presents a particularly formidable challenge for the vast and critical ecosystem of embedded systems, where the demand for robust, long-term security collides with the reality of severe resource constraints. This review has sought to comprehensively map this complex landscape, from the foundational mathematics of PQC algorithms to the granular performance details on embedded hardware and the unique strategic imperatives of the Defense, Health, and Finance sectors.

7.1 Synthesis of Key Findings

The analysis conducted throughout this review leads to several key conclusions that can guide the path forward.

First, the quantum threat is real, and the "Harvest Now, Decrypt Later" strategy makes the migration to PQC an immediate necessity, not a future problem. The successful multi-year, global effort led by NIST has provided a trusted and robust set of initial standards—ML-KEM, ML-DSA, and SLH-DSA—that form a solid foundation for this transition.

Second, the feasibility of implementing these standards on embedded systems is highly dependent on the specific device's capabilities and the application's requirements. Our analysis of performance benchmarks shows that for higher-end 32-bit microcontrollers, such as the ARM Cortex-M4, optimized implementations of the primary lattice-based standards, **ML-KEM (Kyber)** and **ML-DSA (Dilithium)**, are demonstrably feasible. They offer strong performance at a resource cost that, while higher than their classical predecessors, is manageable for a wide range of applications.

Third, for the most severely resource-constrained devices, particularly the ultra-low-power implantable and wearable devices in the health sector, direct implementation of the current NIST standards is often impractical. For these use cases, the most promising solutions are not purely algorithmic but architectural. Strategies that **offload** the heavy PQC computations to a more powerful host device or a cloud service, using PQC only to establish a symmetric session key, provide a viable path to quantum-resistant security without draining the device's precious battery life. Concurrently, novel **asymmetric-workload algorithms** like LiteQSign, which are explicitly designed to minimize the computational burden on the signing device, represent a vital research direction.

Finally, the future of high-performance PQC in the embedded space is inextricably linked to **hardware-software co-design**. Extensible architectures like **RISC-V**, with their support for custom accelerators and instruction set extensions, offer a clear path to overcoming the performance bottlenecks of PQC. This approach transforms PQC from merely "possible" to "highly efficient," enabling its use in the most demanding real-time and high-throughput applications.

7.2 A Final Perspective on Viability and Adoption

The migration to lightweight PQC is not a monolithic technical problem but a multifaceted strategic endeavor that will unfold differently and at a different pace in each critical sector. The

transition will be shaped by the interplay of technology, regulation, and risk management. The pace will be set by the stringent, security-first mandates of the **Defense** sector; moderated by the cautious, safety-focused regulatory environment of the **Health** sector; and driven by the complex, trust-based, interoperability-dependent ecosystem of the **Finance** sector.

Perhaps the most significant and lasting outcome of this global transition will be the widespread adoption of **crypto-agility** as a core design principle. The immense challenge of migrating the world's digital infrastructure to PQC is forcing all industries to abandon the brittle, hardcoded cryptographic designs of the past and build systems that are inherently more flexible, adaptable, and resilient to change. This will not only secure our systems against the quantum threat but will better prepare us for any cryptographic challenge the future may hold.

7.3 Concluding Remarks on the Collaborative Effort Required

Securing the future of our critical embedded infrastructure is a task too large for any single entity. It demands a sustained and collaborative effort across the entire technological and societal spectrum. Cryptographers must continue to innovate, pushing the boundaries of efficiency in quantum-resistant algorithm design. Electrical engineers and computer architects must develop novel hardware and software optimizations to bring these algorithms to life on constrained platforms. Standards bodies like NIST and industry consortia like FS-ISAC must continue to provide clear, trusted, and practical guidance. Regulators like the FDA must skillfully balance the adoption of new security technologies with the paramount goals of patient safety and device effectiveness. Only through this dedicated, multi-stakeholder collaboration can we successfully navigate the transition and ensure that the embedded systems that underpin our modern world remain secure and trustworthy in the quantum era and beyond.