An 1.3μJ/Op 0.4mm^2 Agile Post-Quantum Crypto-Processor Towards IoT Applications

Tianze Huang, Jiahao Lu*, Dongsheng Liu, Siyuan Chen, Mingbo Wang, Xianqi Mei, Huanyu Wang, Aobo Li

School of Integrated Circuits, Huazhong University of Science and Technology, Wuhan 430074, China

***Abstract*—The looming threat of quantum computer attack raised a worldwide research on novel cryptographic algorithms. ML-KEM, ML-DSA, SLH-DSA are recognized as NIST-standardized post-quantum schemes due to their robust security and good performance. This paper presents a lightweight agile crypto-processor compatible with the 3 standards. Multi-mode data generator introduces align structure with resource/energy-saving features. Dual-path modular arithmetic elements enable semi-separation/complete-width operations for different algorithms. Efficient parallel execution condenses memory footprint through on-the-fly signature transmission. Fabricated on 28nm process, our processor occupies 0.4 mm^2, 38kB SRAM with 1.3 μJ/Op energy efficiency. Compared with similar SOTA works, we have achieved 1.4×/3.4× energy/area efficiency improvement with 81% area reduction.**

***Keywords*—post-quantum cryptography, lightweight hardware architecture, energy-efficiency, agile crypto-processor**

I. Introduction

The "harvest now, decrypt later" attack raised a worldwide concern on current security systems. Post-quantum cryptography (PQC) underlies novel mathematical problems and provides promised security against the adversary. In 2024, NIST officially issued 3 Federal Information Processing Standards (FIPS), which confirmed ML-KEM (Kyber), ML-DSA (Dilithium), SLH-DSA (SPHINCS+) as new standards for key encapsulation mechanism or digital signature. However, the migration of these quantum-resistant schemes still faces 3 major challenges, as shown in Fig. 1. Firstly, general terminal devices exhibit diverse cryptographic requirements across different application scenarios. Then, larger key size and more complex operations lead to significant area overhead, making it difficult to meet the stringent resource limitation in edge computing. Finally, higher power consumption also hinders their deployment in energy-constrained applications.

There are many integrated circuit (ASIC) implementations for post-quantum cryptography. Earlier research [1][2][3] achieves orders of magnitude speed up compared with software method. However, the unoptimized hardware architecture involves redundant operations and still results in high energy consumption. Some designs target on high-speed centralized platforms and support multiple PQC algorithms [4][5][6]. These implementations exhibit a weakness of heavy resource and energy burden. Although existing lightweight designs [7][8] demonstrate favorable area and power metrics, they focus on a specific PQC scheme and are insufficient for diverse IoT applications. Considering the practical requirements of edge computing scenarios, it's necessary to begin the research on compact area, low energy and agile PQC ASIC.

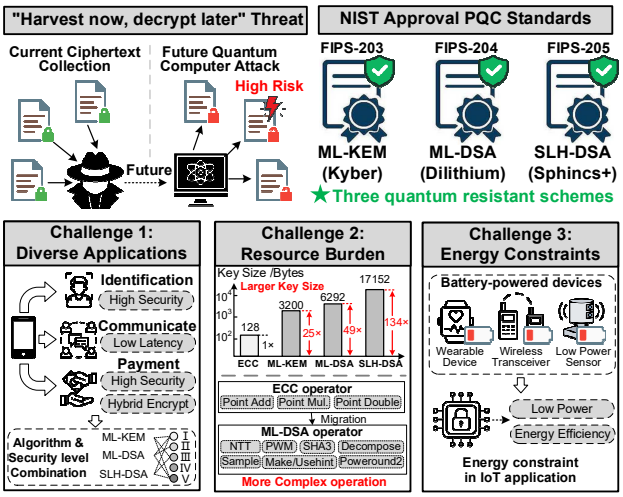


Fig. 1. Challenges for Post-Quantum Cryptography.

This paper presents a lightweight agile crypto-processor that supports the latest 3 post-quantum cryptography standards. It introduces dual-path modular arithmetic elements for different algorithms and lowers memory consumption with on-the-fly signature transmission. The align structure in multi-mode data generator achieves resource and energy reduction. Fabricated in 28nm process, it consumes 0.4 mm^2 with an energy efficiency of 1.3μJ/Op.

II. Hardware Implementation

A. Overall Architecture

Fig. 2 shows the overall architecture of our crypto-processor. It consists of memory, controller, polynomial arithmetic unit (PAU) and multi-mode data generator (MDG). The controller provides control signals and distributes memory ports to the other modules. PAU executes various modular operations with 4×/2× parallel in ML-KEM/ML-DSA. Hash digests and polynomial coefficients are provided by MDG. Memory contains three secret RAMs (12kB/12kB/6kB) and one public RAM (8kB). The former saves private data and requires restricted access, while the latter can be accessed normally. Finally, we attach the independent PQC crypto-core to RISC-V processor as general accelerator, which facilitates system-level verification.

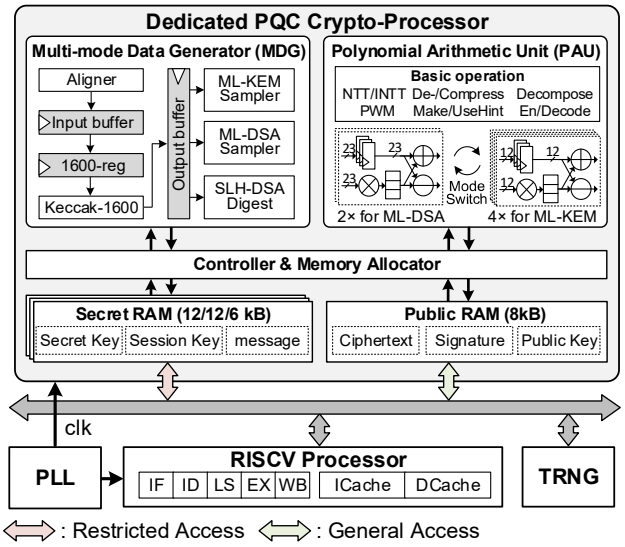


Fig. 2. Architecture of the Proposed Dedicated PQC Processor.

B. *Polynomial Arithmetic Unit*

There are various polynomial operations in ML-KEM/ML-DSA. Number-theoretic transform (NTT) based multiplication occupies the majority of execution time and we mainly discuss it. The operation contains basic modular addition, subtraction and multiplication. Existing designs utilize 24-bit data-path for adapting the 23-bit coefficients in ML-DSA [1][3]. However, they directly applied the circuit to ML-KEM (12-bit coefficient). The computation in upper redundant bits lead to more energy consumption and lower area efficiency.

Fig. 3 (a)(b) shows the proposed dual-path modular adder/subtractor. They both instantiates two 12-bit adders and subtractors. Through controlling the carry and borrow signals, a single full-width ML-DSA operation or two parallel half-width ML-KEM operations are supported.

The structure of dual-path modular multiplier is shown in Fig. 3(c). In ML-KEM mode, two basic multipliers work independently and each 24-bit result is divided into three parts. The highest 4-bit x only contains 11 possible values because two operands are in range [0, 3328]. We pre-store 11 × 12-bit constants in look-up table and perform related modular reduction. Then middle 8-bit y is reduced with formula $2^{12} \equiv 2^9+2^8\text{-}1 \bmod q_k$. Simplified operation is shown in the red dashed box, which involves several small-width additions. Finally, lowest 12 bits z directly participates in subsequent computations. In ML-DSA mode, two multipliers perform $a \times b[11:0]$ and $a \times b[22:12]$, respectively. Since formula $2^{23} \equiv 2^{13}\text{-}1 \bmod q_d$ significantly reduces the data width and only incurs one subtraction, following modular reduction all adopts this method. Compared with the method that consumes 48 additions (24-bit) [3], our design consumes about 50% less hardware resource.

We instantiate 2 reconfigurable butterfly unit as shown in Fig. 4. They instantiate above modular arithmetic elements and perform 4×/2× parallel operations for ML-KEM/ML-DSA . The different data paths support Cooley-Tukey (CT), Gentleman-Sande (GS) and point-wise-multiplication (PWM) operations

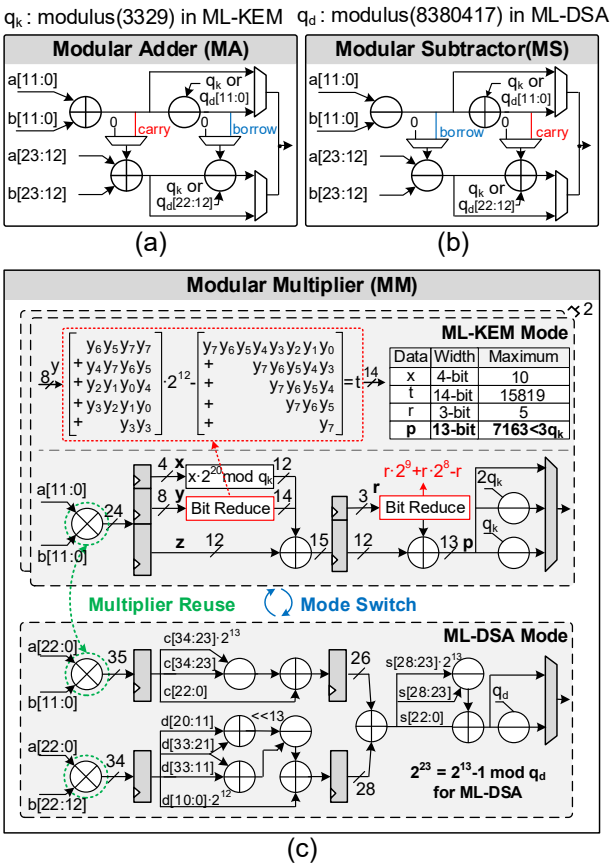


Fig. 3 Dual-path Modular Arithmetic Elements (a) Modular Adder (b) Modular Subtractor (c) Modular Multiplier

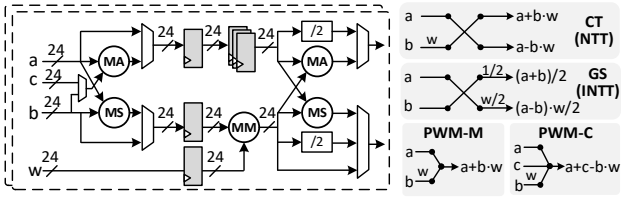


Fig. 4. Reconfigurable Butterfly Unit

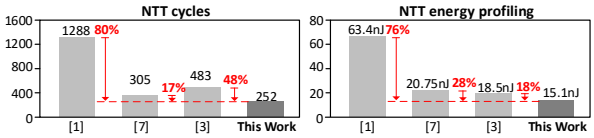


Fig. 5. NTT Cycles and Energy Profiling.

for polynomial multiplication. Fig. 5 shows the result for NTT cycles and energy. Compared with similar lightweight design [3], our design achieved 48%/18% reduction in cycles/energy.

C. *Multi-mode Data Generator*

Multi-mode data generator provides digest and polynomial coefficients for ML-KEM/ML-DSA/SLH-DSA. However, as

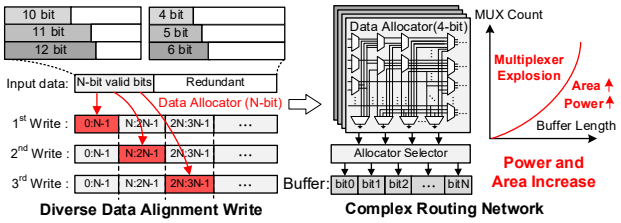


Fig. 6. Challenge in Multi-mode Data Generator Due to Diverse Data Widths.

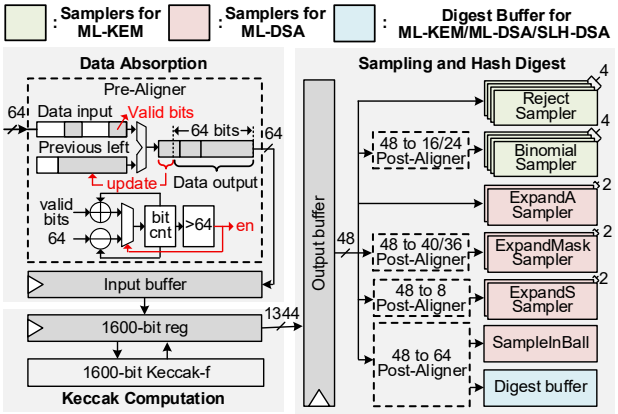


Fig. 7. Multi-mode Data Generator

shown in Fig. 6, diverse data widths alignment will lead to a multiplexers explosion for large register array, which increases area and power. We introduce aligner modules as shown in Fig. 7. The pre-aligner extracts valid bits and collects them serially. Once the bit counter is larger than 64, 64 bits valid data will be written to input buffer. After 24-round Keccak iterations, output buffer saves the result and delivers it to post-aligners in a fixed 48-bit width. With these aligner modules, we ensure that large register array (input/output buffer) operates with a fixed data width, which saves the extra area and power caused by multiplexers explosion.

Another advantage of our design lies on the orderly execution. Fig. 8(a) shows timing diagram in reject sampling, similar method is also extended to the other samplers. The process requires multiple 24-round Keccak computation. After completing the 1st one, next 24-round Keccak iterations and sampling can be executed simultaneously. Fig. 8 (b) shows the orderly execution specifically adopted in SLH-DSA, which only involves hash operations. In this scheme, data absorption and computation are performed simultaneously. This reduces the idle time for computational circuit. Better metrics for sampling operation is shown in Fig. 9. Compared with [3], our design achieved 51%/35% reduction in sampling cycles/energy.

D. Parallel Module Collaboration

Towards the two dominant operations in ML-KEM/ML-DSA, polynomial multiplication and sampling are executed synchronously , as shown in Fig. 10(a). Further, since signatures in SLH-DSA are generated uniformly over time, we adopt an

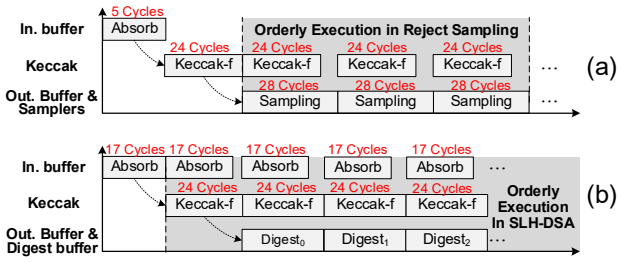


Fig. 8. (a) Orderly Execution in Reject Sampling
(b) Orderly Execution in SLH-DSA

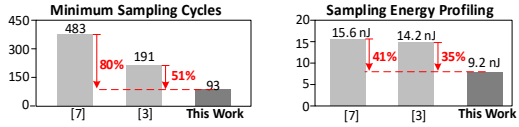


Fig. 9. Hash and Sampling Cycles, Energy Profiling.

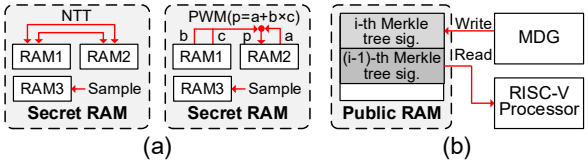
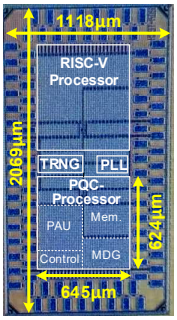


Fig. 10. (a) Parallel Collaboration for Polynomial Multiplication
(b) On-the-fly Signature Transmission in SLH-DSA



Technology	TSMC 28nm
Die size	2.069×1.118 mm^2
Core Area	0.4 mm^2 (PQC Processor) / 1.0mm^2(with RISC-V Processor)
Voltage	0.68 - 1.1 V
SRAM	38 KB
Frequency	10 - 550 MHz
Hash function	SHA3-256/SHA3-512 SHAKE128/SHAKE256
Power	0.78mW @0.68 V ~ 93mW @1.1V

Fig. 11. Die Micrograph and Chip Characteristics

on-the-fly transmission strategy as shown in Fig. 10(b). Whenever one Merkle tree is completed, corresponding signature is transferred immediately to host device. Thus, we only requires the memory footprint for 2 Merkle tree signatures, which reduces about 90% storage from 49856 bytes to 4544 bytes and saves corresponding transmission time.

III. Measurement Results

The proposed processor is evaluated and implemented on TSMC 28nm CMOS process. The micrograph of chip die is shown in Fig. 11. Our independent PQC-processor consumes 0.4mm^2. The power supply range is 0.68V-1.1V, and frequency range is 10MHz-550MHz.

The throughput and energy efficiency for ML-KEM/ML-DSA/SLH-DSA are listed in Fig. 12. Our design reaches a peak throughput 54 KOPS with an energy efficiency of 1.3μJ/Op for

TABLE I COMPARISON OF PERFORMANCE AND RESOURCE

	ISSCC'19 [1]	ISSCC'22 [4]	CICC'24 [8]	ESSERC'22 [3]	ESSERC'24 [7]	VLSI'25 [5]	ISSCC'24 [6]	This Work
Technology	40nm	28nm	40nm	28nm	40 nm	28nm	28nm	28nm
Frequency(MHz)	12-72	500	10-180	35-190	10-270	50-800	275-750	10-550
Voltage(V)	0.68-1.1	0.9	0.65-1.2	0.65-1.35	0.62-1.2	0.65-1.1	0.7-1.1	0.68-1.1
Area(mm^2)	0.28[c]	3.6	0.43	0.18[c]	0.34	2.11	3.2	0.4[d]
Power(mW)	7-10	39-368	0.69-31.36	1.21	0.273-35.0	8.45-351.45	91-420	0.78-93
NIST FIPS	Kyber Dilithium	Kyber Dilithium	Kyber	Kyber Dilithium	Kyber	Kyber Dilithium SPHINCS+	Kyber Dilithium SPHINCS+	ML-KEM ML-DSA SLH-DSA
Kyber-512/ML-KEM-512 (Keygen+Encaps.+Decaps.)								
Throughput (OPS)	207	47925	18935	1942	12770	84951	69473	54201[a]
Energy eff. (μJ/Op)	26.6	3.40	1.26	5.22	0.72/1.50[e]	1.82	4.40	1.3[a]
Area eff.[b]	0.739	13.3	44.0	10.8	37.5	40.3	21.7	135.5[a]
Dilithium-II/ML-DSA-44 (Keygen+Sign+Verify)								
Throughput (OPS)	87	11527	-	-	-	23325	15832	12991[a]
Energy eff. (μJ/Op)	87.2	20.6	-	-	-	6.79	22.80	5.0[a]
Area eff.[b]	0.3	3.2	-	-	-	11.1	4.9	32.5[a]
SPHINCS+128/SLH-DSA-128 (Keygen+Sign+Verify)								
Throughput (OPS)	-	-	-	-	-	1283	219.6	192[a]
Energy eff. (μJ/Op)	-	-	-	-	-	172	807.5	230[a]
Area eff.[b]	-	-	-	-	-	0.6	0.068	0.48[a]

a: Measured at 0.9V and 500MHz.
b: Area efficiency (Area eff.) is calculated as throughput/area.
c: Not include on chip memory.
d: All parts in our independent crypto-processor are calculated, including polynomial arithmetic unit, multi-mode data generator, controller and SRAM.
e: 0.72μJ is measured at 0.62V, it is normalized to 1.5 μJ at 0.9V

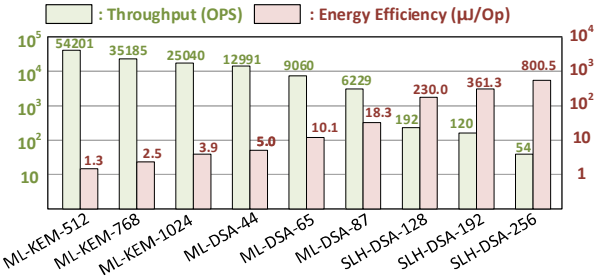


Fig. 12. Throughput and Energy Efficiency at 500MHz & 0.9V

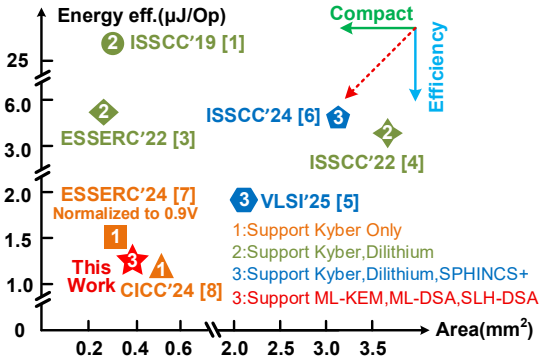


Fig. 13. Comparison with ASIC Works.

ML-KEM-512. A more detailed comparison with other advanced works is shown in Table I and Fig. 13. Compared with the same agile designs [5][6], we have achieved 81%/87% area reduction, 1.4×/3.4× energy efficiency and 3.4×/3.6× area efficiency improvement.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (No. 62134002), the Wuhan Key R&D Program (No. 2025050602030108), the Fund of National Key Laboratory of Multispectral Information Intelligent Processing Technology (Grant No. 202410487401), the Innovation Project of JinYinHu Laboratory (Grant No. 2024JYH011401), the National Key R&D Program of China (No.2024YFB4505400).

REFERENCES

[1] U. Banerjee, A. Pathak, and A. P. Chandrakasan, "2.3 An energy-efficient configurable lattice cryptography processor for the quantum-secure Internet of Things," ISSCC, pp. 46-48, Feb. 2019.

[2] A. Ghosh, J. M. B. Mera, A. Karmakar, D. Das *et al.*, "A 334uW 0.158 mm^2 saber learning with rounding based post-quantum crypto accelerator," CICC, pp. 1-2, Apr. 2022.

[3] B. Kim, J. Park, S. Moon, K. Kang *et al.*, "Configurable energy-efficient lattice-based post-quantum cryptography processor for iot devices," ESSCIRC, pp. 525-528, Sep. 2022.

[4] Y. Zhu, W. Zhu, M. Zhu, C. Li *et al.*, "A 28nm 48KOPS 3.4 μJ/Op Agile Crypto-Processor for Post-Quantum Cryptography on Multi-Mathematical Problems," ISSCC, pp. 514-516, Feb. 2022.

[5] J. Lu et al., "A 28nm 84.9KOPS 1.82μJ/op RISC-V Crypto-SoC with Primitive-Based Deep-Coupling Unified Post-Quantum Engine," VLSI, pp. 1-3, June 2025.

[6] Y. Zhu, W. Zhu, Y. Ouyang, J. Sun *et al.*, "16.2 A 28nm 69.4 kOPS 4.4 μJ/Op Versatile Post-Quantum Crypto-Processor Across Multiple Mathematical Problems," ISSCC, pp. 298-300, Feb. 2024.

[7] A. Li et al., "A 273μW 0.34mm^2 Efficient CRYSTALS-KYBER Processor for PQC Towards Edge Computing," ESSERC, pp. 472-475, Sep. 2024.

[8] A. Li, J. Lu, D. Liu and X. Li, "A 40nm 1.26μJ/Op Energy-Efficient CRYSTALS-KYBER Post-Quantum Crypto-Processor with Comprehensive Side Channel Security Analysis and Countermeasures," CICC, pp. 1-2, Apr. 2024.