

Revolutionizing Post-Quantum Cryptography Hardware: A Journey from Ultra-High Performance to Ultra-Lightweight NTT Accelerators

Bardia Taghavi, Prof. Reza Azarderakhsh

Florida Atlantic University

Introduction & Motivation

Quantum computing threatens traditional cryptography, driving the need for Post-Quantum Cryptography (PQC). NIST has standardized Module-Lattice-Based Cryptography (MLBC) schemes (ML-KEM, ML-DSA) that rely heavily on polynomial multiplication via Number Theoretic Transform (NTT). NTT consumes the most of the computation time in PQC, and the hardware acceleration of NTT is crucial.

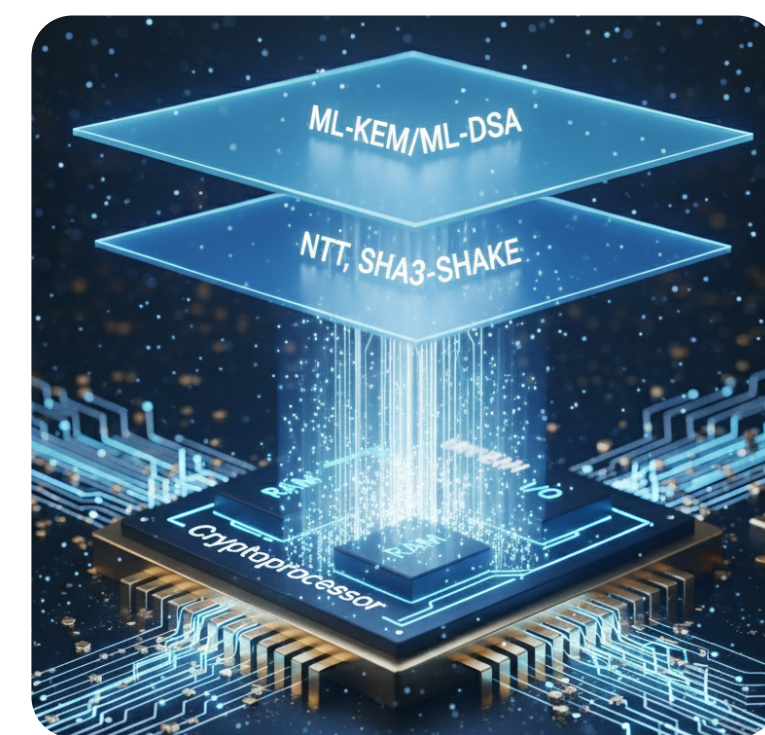


Figure 1. A Hierarchical View of a PQC Crypto-Processor.

Challenge: Hardware acceleration must address two extremes:

- ❖ **Ultra-Fast:** Maximum speed for servers/cloud.
- ❖ **Ultra-Lightweight:** Minimal area/power for IoT/smart cards.

Solution: Three NTT accelerators spanning the performance spectrum:

ParallelINTT, **LightNTT**, and **KNightCore**.

High-Performance: ParallelINTT

Revolutionary $\mathcal{O}(\log_2 n)$ Latency for both ML-KEM and ML-DSA.

Key Innovation: Fully parallelized architecture eliminates RAM bottlenecks by feeding all coefficients simultaneously to NTT stages.

Architecture:

- ❖ Constant-geometry structure for simplified routing
- ❖ K^2 -RED modular reduction with precomputed factors
- ❖ Zero-RAM: all coefficients in flip-flops

Results:

- ❖ **Latency:** $0.015\mu s$ (ML-KEM), $0.017\mu s$ (ML-DSA) at 400 MHz
- ❖ **Speedup:** **24× faster** than SOTA (ML-DSA), **21× faster** than SOTA (ML-KEM)

ParallelINTT Hardware Design

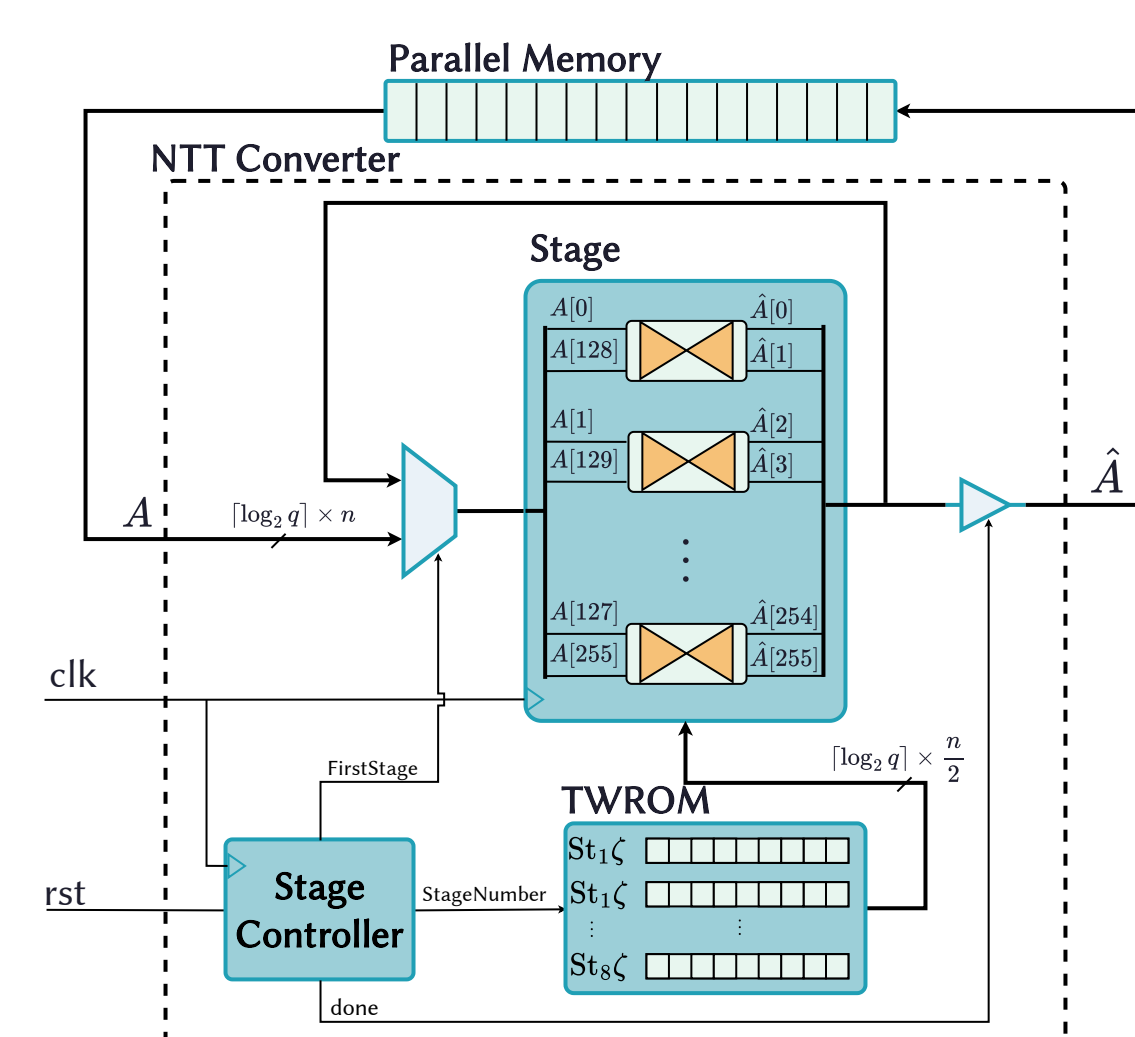


Figure 2. ParallelINTT Hardware

Lightweight Solutions

Ultra-Compact Implementations

LightNTT (ML-DSA):

- ❖ Constant-geometry pipelined design, single butterfly unit
- ❖ Hardware-efficient OptimizedBarrett reduction for 23-bit modulus
- ❖ **Area:** 590 LUTs, 158 FFs, 2.5 BRAMs, 2 DSPs, **1.3× smaller** than SOTA
- ❖ **Performance:** $3.49\mu s$ at 300 MHz

KNightCore (ML-KEM):

- ❖ Unified datapath: NTT, iNTT, PWM in single pipeline
- ❖ Novel PWM: reuses NTT twiddle factors, reduced ROM usage
- ❖ **Area:** 650 LUTs, 130 FFs, 1.5 BRAMs, **0 DSPs**, **3.3× smaller** than SOTA
- ❖ **Performance:** $3.37\mu s$ (NTT/iNTT) + $2.35\mu s$ (PWM) at 274 MHz

LightNTT Hardware Design

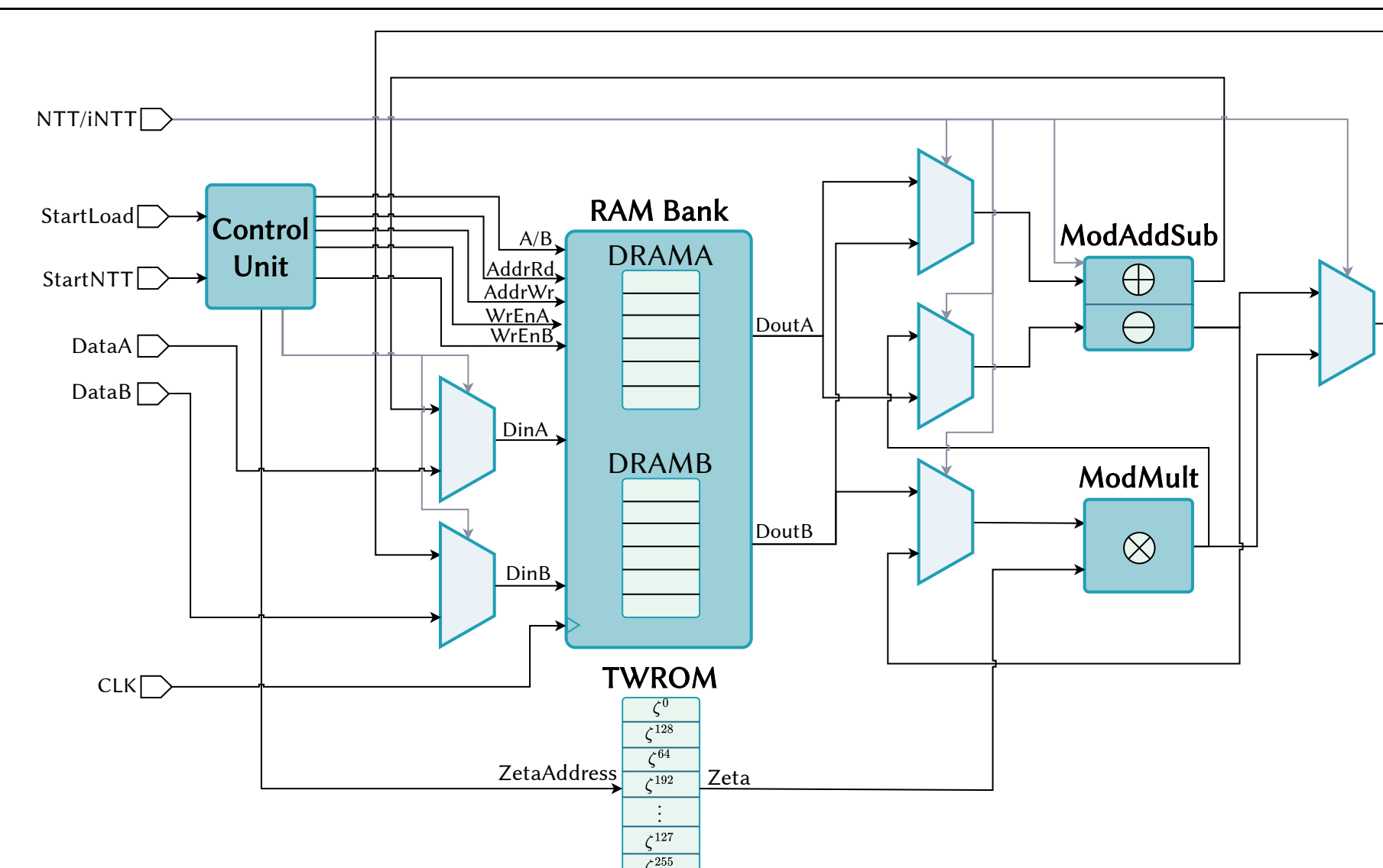


Figure 3. LightNTT Hardware

Key Results

Performance Summary:

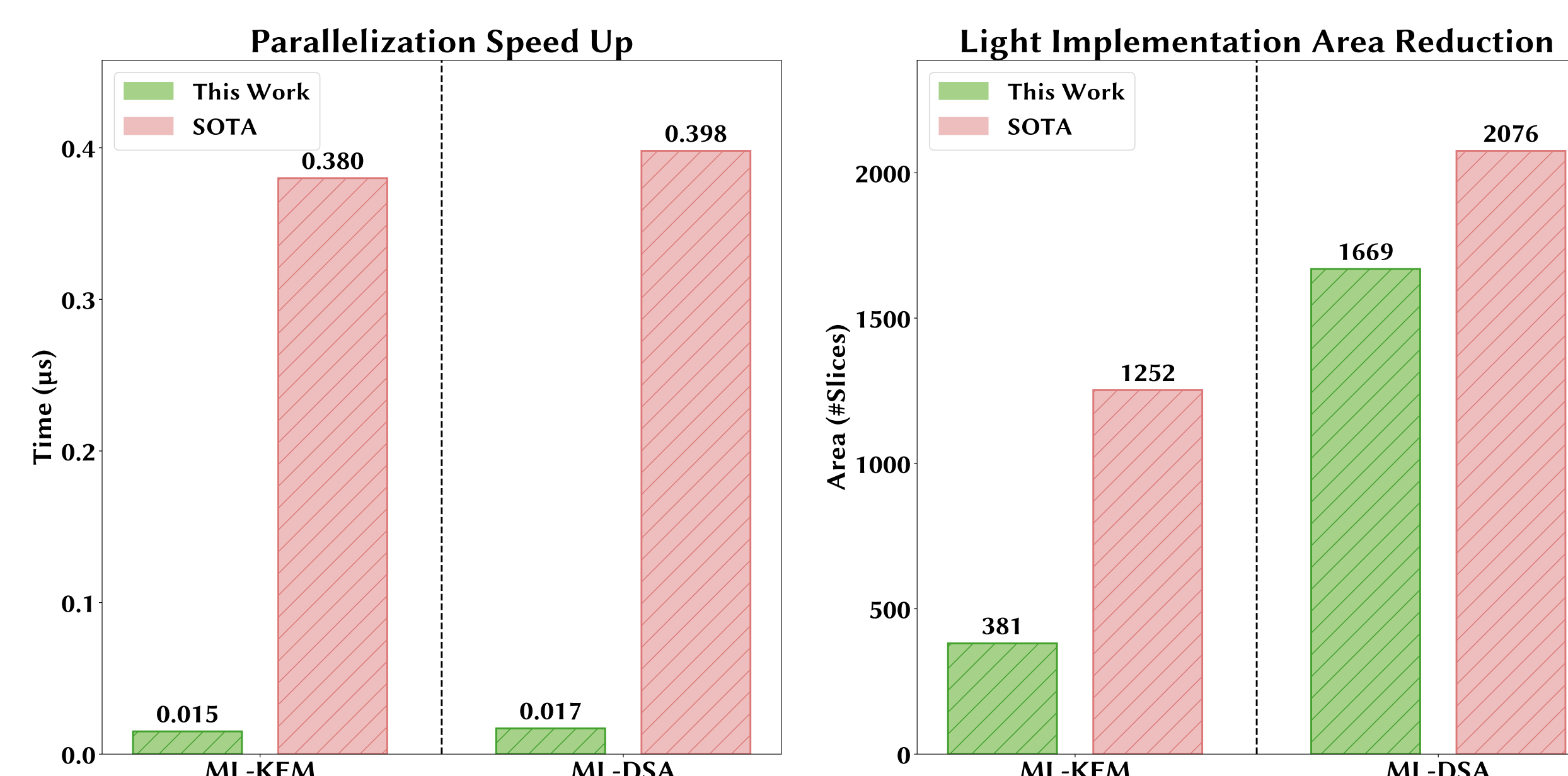


Figure 4. Comparison of the proposed NTT accelerators with the SOTA [1], [2], and [3].

Contributions & Impact

Unified PQC Hardware Solution

Revolutionary Breakthrough:

- ❖ **ParallelINTT:** First $\mathcal{O}(\log_2 n)$ latency NTT, up to 24× speedup
- ❖ **Industry Impact:** Enables real-time PQC encryption and digital signature in high-throughput servers

Ultra-Lightweight Solutions:

- ❖ **LightNTT:** Best ATP (5.82) for ML-DSA lightweight implementations, small and fast enough, 29% improvement in ATP
- ❖ **KNightCore:** First DSP-free design for ML-KEM with unified core for NTT/iNTT/PWM, with the smallest area usage
- ❖ **Deployment:** Enables PQC on IoT, smart cards, embedded systems

Research Impact:

- ❖ Demonstrated constant-geometry scalability across performance spectrum, for both schemes, with different NTT properties
- ❖ Advanced modular arithmetic optimization techniques
- ❖ Illustrated the performance of the proposed Barrett reduction in NTT among all the other reduction techniques
- ❖ Novel resource unification approaches for multiple operations

Future Work and Next Steps

- ❖ **Multi-Core Design:** Parallel lightweight cores for high-throughput applications
- ❖ **Security:** Measured and mitigated side-channel attack countermeasures for constant-geometry architectures
- ❖ **Unified Core:** A unified NTT core supporting both ML-KEM and ML-DSA in a single core
- ❖ **Integration:** A complete ultra-fast or ultra-lightweight implementation of PQC system-on-chip

Acknowledgements

Generously supported by NSF RINGS 214796. Thank you to my advisor, Prof. Reza Azarderakhsh, for his guidance and support.

References

- [1] S. Mandal and D. B. Roy, "Kid: A hardware design framework targeting unified NTT multiplication for crystals-kyber and crystals-dilithium on FPGA," in *37th International Conference on VLSI Design and 23rd International Conference on Embedded Systems, VLSID 2024, Kolkata, India, January 6-10, 2024*, pp. 455-460, IEEE, 2024.
- [2] C. Zhang, D. Liu, X. Liu, X. Zou, G. Niu, B. Liu, and Q. Jiang, "Towards efficient hardware implementation of NTT for kyber on fpgas," in *IEEE International Symposium on Circuits and Systems, ISCAS 2021, Daegu, South Korea, May 22-28, 2021*, pp. 1-5, IEEE, 2021.
- [3] Y. Geng, X. Hu, M. Li, and Z. Wang, "Rethinking parallel memory access pattern in number theoretic transform design," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 70, no. 5, pp. 1689-1693, 2023.