CryptoLight: An Electro-Optical Accelerator for Fully Homomorphic Encryption

Mengxin Zheng
Indiana University
Bloomington, USA
zhengme@iu.edu

Qian Lou
UCF
Orlando, USA
qian.lou@ucf.edu

Fan Chen
Indiana University
Bloomington, USA
fc7@iu.edu

Lei Jiang
Indiana University
Bloomington, USA
jiang60@iu.edu

Yongxin Zhu
CAS
Shanghai, China
zhuoyongxin@sari.ac.cn

ABSTRACT

Fully homomorphic encryption (FHE) protects data privacy in cloud computing by enabling computations to directly occur on ciphertexts. To improve the time-consuming FHE operations, we present an electro-optical (EO) FHE accelerator, CryptoLight. Compared to prior FHE accelerators, on average, CryptoLight reduces the latency of various FHE applications by >94.4% and the energy by >95%.

CCS CONCEPTS

• Security and privacy → Privacy protections; • Hardware → Emerging optical and photonic technologies.

KEYWORDS

Fully homomorphic encryption, Accelerator.

ACM Reference Format:

Mengxin Zheng, Qian Lou, Fan Chen, Lei Jiang, and Yongxin Zhu. 2022. CryptoLight: An Electro-Optical Accelerator for Fully Homomorphic Encryption. In Proceedings of 17th IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH '22). ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3565478.3572544

1 INTRODUCTION

Data privacy is gaining tremendous importance around the world. This leads a surge in demand for privacy-preserving computing solutions protecting data confidentiality while in transit, rest, and in-use. Fully homomorphic encryption (FHE) [3] emerges as one of the most promising solutions to guaranteeing data privacy by allowing computations to directly happen on ciphertexts. FHE operations are extremely time-consuming, i.e., one FHE bootstrapping costs several seconds on a CPU. Prior work proposes GPU [3], -based accelerators to process FHE operations. Among all, the ACS-based FHE accelerators, CraterLake [5] and BTS [2], obtain the state-of-the-art performance.

However, their performance is seriously limited by their narrow datapaths and intensive matrix transpositions. An FHE ciphertext consists of two polynomials of large degrees (e.g., several thousand) with large integer coefficients (e.g., several hundred bits). To efficiently compute with polynomials, FHE schemes (e.g., CKKS [5]) adopt Residue Number System (RNS) [5], and Number Theoretic Transform (NTT). First, to compute with large integer coefficients, RNS divides each coefficient into multiple smaller bit-width (e.g., 60-bit) residues, each of which can be processed by the datapath of prior FHE accelerators. The latency of an FHE multiplication, rotation, or bootstrapping is dominated by expensive key-switching (KS) primitives [3] that makes output ciphertexts to be encrypted by the same secret key as the input ciphertext(s).

The computational overhead of KS greatly increases with an enlarging number of residues. Second, for two large degree-N polynomials, NTT and inverse NTT (INTT) reduces the time complexity of their multiplications to $O(N \log N)$. But it is difficult to perform an (i)NTT, i.e., NTT or INTT, on a large degree-N polynomial directly. Prior FHE accelerators [2, 5] place it as an $n \times n$ matrix, where $N = n^2$, perform an (i)NTT on each row, multiply the matrix with some constants, transpose the matrix, and perform an (i)NTT on each row again. As a result, frequent matrix transpositions greatly prolong the latency of KS in various FHE operations by introducing huge volumes of on-chip memory traffic.

We propose an electro-optical (EO) FHE accelerator, CryptoLight, to support a large bit-width datapath and free its computing units from matrix transpositions. Our contribution is summarized as:

• A 512-bit EO CU. We propose a 512-bit EO Computing Units (CUs) built upon ultra-fast EO integer adders and multipliers to process polynomials with large coefficients.

• An in-SPM TU. We build a low-power eDRAM-based on-chip scratchpad (SPM) system.

2 CRYPTOLIGHT

2.1 A 512-bit EO CU

We build a 512-bit EO CU featured by an NTT unit, a modular add/mult unit, an automorphism unit, and a true random number generation (TRNG) unit. Its most important component is the 512-bit EO NTT unit, which has an arithmetic and inversion unit, an address generation unit, and two butterfly units. A 512-bit EO NTT unit also supports the kernel of iNTT working in a different data flow. We also use EO adders and multipliers to construct the other CU components.

An EO NTT Unit. We present an EO NTT unit for the CU. Matrix transpositions are done by TUs in SPM banks, but the other three steps of the NTT on a large polynomial are computed by an NTT unit. CryptoLight aims to support 64K-element NTTS, so a NTT unit supports 256-element NTT operations. The details are summarized as follows.

• A butterfly unit: We propose an EO butterfly unit (BU) to accelerate radix-2 NTT butterflies, as shown in Figure 1(a). A BU consists of an EO pipelined integer array multiplier, an EO Montgomery modular reduction unit, and two EO modular adders. The
Matrix transpositions are heavily used. One bank has a level-2 TU to enable data movements inside the bank without sending it to BTS. Figure 1(e) shows, we implement the recursive algorithm. We modeled CryptoLight by a cycle-accurate FHE accelerator simulator, Sapphire-Sim [1], which is validated against several crypto-processor chips. We compared CryptoLight against the state-of-the-art ASIC-based FHE hardware accelerators, CraterLake (Lake) [5] and BTS [2]. We studied the performance and energy of CKKS FMUL, FROT, and FBOT operations on ciphertexts.

The latency comparison between CryptoLight and various accelerator baselines is shown in Figure 2(a). Compared to Lake, BTS decreases the latency of FMUL, FROT and FBOT by 60% on average, due to its larger bit-width datapath and larger SPM system. Because of the 512-bit EO datapath and the TUs in the SPM, CryptoLight reduces the latency of FMUL, FROT and FBOT by 96% on average over BTS.

The energy comparison between CryptoLight and various accelerator baselines is shown in Figure 2(b). BTS consumes slightly larger power than Lake. Compared to BTS, CryptoLight reduces the energy consumption of FMUL, FROT and FBOT by 98.8% on average.

REFERENCES

[1] Utsav Banerjee et al. 2019. Sapphire: A Configurable Crypto-Processor for Post-Quantum Lattice-based Protocols. IACR Transactions on Cryptographic Hardware and Embedded Systems (2019).

[2] SangyouKim et al. 2022. BTS: An Accelerator for Bootstrappable Fully Homomorphic Encryption. In ACM International Symposium on Computer Architecture.

[3] Palisade. 2020. PALISADE Operations using CUDA.

[4] M. SadeghRiazi, Kim Laine, Blake Polton, and Wei Dai. 2020. HEAX: An Architecture for Computing on Encrypted Data. In Architectural Support for Programming Languages and Operating Systems.

[5] Nikola Samardzic et al. 2022. CraterLake: A Hardware Accelerator for Efficient Unbounded Computation on Encrypted Data. In IEEE/ACM International Symposium on Computer Architecture.

[6] Zhoufeng Ying et al. 2018. Electro-Optic Ripple-Carry Adder in Integrated Silicon Photonics for Optical Computing. Journal of Selected Topics in Quantum Electronics (2018).