CECS

CENTER FOR EMBEDDED & CYBER-PHYSICAL SYSTEMS

Software and Hardware Implementation of Lattice-based Cryptography Schemes

Hamid Nejatollahi, Nikil Dutt, Sandip Ray,

Francesco Regazzoni, Indranil Banerjee and Rosario Cammarota

Center for Embedded and Cyber-Physical Systems

University of California, Irvine

Irvine, CA 92697-2620, USA

{hnejatol}@uci.edu

CECS Technical Report <17-04>

<11> <09>, <2017>

UNIVERSITY OF CALIFORNIA, IRVINE

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

Software and Hardware Implementation of Lattice-based Cryptography Schemes

HAMID NEJATOLLAHI, University of California Irvine
NIKIL DUTT, University of California Irvine
SANDIP RAY, University of Florida
FRANCESCO REGAZZONI, ALaRi
INDRANIL BANERJEE, Qualcomm Technologies Inc.
ROSARIO CAMMAROTA, Qualcomm Technologies Inc.

The advent of Quantum computing threatens to break many classical cryptographic schemes, leading to innovations in public key cryptography that focus on post-quantum cryptography primitives and protocols resistant to quantum computing threats. Lattice-based cryptography is a promising post-quantum cryptography family, both in terms of foundational properties as well as its application to both traditional and emerging security problems such as encryption , digital signature, key exchange, homomorphic encryption etc. While such techniques provide guarantees in theory, their realization on contemporary computing platforms requires careful design choices and trade-offs to manage both the diversity of computing platforms (e.g., high-performance to resource constrained), as well as the agility for deployment in the face of emerging and changing standards. In this work, we survey trends in lattice-based cryptographic schemes, some fundamental recent proposals for the use of lattices in computer security, challenges for their implementation in software and hardware, and emerging needs for their adoption. The survey means to be informative about the math, to allow the reader to focus on the mechanics of the computation, ultimately needed for mapping schemes on existing hardware, or synthesizing, part or all of a scheme, on special purpose hardware.

Additional Key Words and Phrases: Post-quantum cryptography; Public-key encryption; Lattice based cryptography; Ideal lattices; Ring-LWE

ACM Reference Format:
Hamid Nejatollahi, Nikil Dutt, Sandip Ray, Francesco Regazzoni, Indranil Banerjee, and Rosario Cammarota. 2017. Software and Hardware Implementation of Lattice-based Cryptography Schemes. *ACM Comput. Surv.* 1, 1, Article 0 (2017), 43 pages. https://doi.org/0000001.0000001

1 INTRODUCTION

Advances in computing efficiency steadily erode computer security at its foundation. These enable prospective attackers to use ever more powerful computing systems and the best cryptanalysis algorithms to increase the attack speed. One

This work was supported in part with a gift from Qualcomm Research.
Authors' addresses: Hamid Nejatollahi, University of California Irvine, Irvine, California, 92697-3435, hnejatol@uci.edu; Nikil Dutt, University of California Irvine, Irvine, California, 92697-3435, dutt@ics.uci.edu; Sandip Ray, University of Florida, sandip@ece.ufl.edu; Francesco Regazzoni, ALaRi, regazzoni@alari.ch; Indranil Banerjee, Qualcomm Technologies Inc. San Diego, CA, 92121-1714, ibanerje@qti.qualcomm.com; Rosario Cammarota, Qualcomm Technologies Inc. San Diego, CA, 92121-1714, ro.c@qti.qualcomm.com.

© 2017
University of California Irvine, CECS TR 17-04

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

aspect of this arises from Moore's Law[1], that enables traditional computing systems to become more powerful than ever before, allowing increasingly larger brute-force attacks.

Another aspect of concern is the rise of alternative computing paradigms, such as Quantum computing and its algorithms [120, 170] - that seem to be closer than ever reality,[2] [3], which in turn promises to further weaken the strength of current, standardized cryptography and its applications. Against quantum computers traditional public key cryptography is ineffective for any key length algorithm. The Shor's algorithm for quantum computers is designed to solve prime factorization of large primes and the discrete logarithm problem in polynomial time.

The emergence of new computing platforms, such as cloud Computing, software defined networks and Internet of Everything, demands the adoption of an increasing number of security standards, which in turn requires the implementation of a diverse set of cryptographic primitives, but this is only part of the story. At the computing platform level, we are seeing a diversity of computing capability, ranging from high-performance (real-time) virtualized environments, such as cloud computing resources and software defined networks, to highly resource-constrained IoT platforms to realize the vision of Internet of Everything. This poses tremendous challenges in the design and implementation of emerging standards for cryptography in a single embodiment, since the computing platforms exact diverging goals and constraints. On one end of the spectrum, in the cloud computing and software defined network space, applications demand high-performance, and energy efficiency of cryptographic implementations. This calls for the development of programmable hardware capable of running not only individual cryptographic algorithms, but full protocols efficiently, with the resulting challenge of designing for agility, e.g., designing computing engines that achieve the efficiency of Application-Specific Integrated Circuits (ASICs), while retaining some level of programmability. On the other end of the spectrum, in the IoT space, implementations of standardized cryptography to handle increased key sizes become too expensive in terms of cost, speed, and energy, but are necessary, e.g., in the case of long lived systems such as medical implants. In part, this demands the development of new and strong lightweight alternatives to symmetric key cryptography as well. Furthermore, given the variety of applications and their interplay with the cloud, even in this case agility in the implementation becomes a key requirement.

As a result of the trends in technology, the need to strengthen current practices in computer security, including strengthening and adding more variety in the cryptographic primitives in use, has become a widely accepted fact. The examples above, to name a few, form a compelling argument to call for innovation in the computer security space, including and beyond the foundations, i.e., down to the actual implementation and deployment of primitives and protocols to satisfy the emerging business models and their design constraints - latency, compactness, energy efficiency, tamper resistance and, more importantly, agility.

Innovation in public key cryptography focuses on the standardization of the so called post-quantum cryptography primitives and their protocols. Among the post-quantum cryptography families, the family of lattice-based cryptography (LBC) appears to be gaining acceptance. Its applications are proliferating for both traditional security problems (e.g., key exchange and digital signature), as well as emerging security problems (e.g., homomorphic schemes, identity-based encryption and even symmetric encryption). Lattice-based cryptographic primitives and protocols provides a rich set of primitives which can be used to tackle the challenges posed by deployment across diverse computing platforms, e.g., Cloud vs. Internet-of-Things (IoT) ecosystem, as well as for diverse use cases, including the ability to perform computation on encrypted data, providing strong (much better understood than before) foundations for protocols

[1] http://www.telegraph.co.uk/technology/2017/01/05/ces-2017-moores-law-not-dead-says-intel-boss
[2] http://spectrum.ieee.org/tech-talk/computing/software/rigetti-launches-fullstack-quantum-computing-service-and-quantum-ic-fab
[3] https://newatlas.com/ibm-next-quantum-processors/49590/

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

based on asymmetric key cryptography against powerful attackers (using Quantum computers and algorithms), and to offer protection beyond the span of traditional cryptography. Indeed, lattice-based cryptography promises to enhance security for long-lived systems, e.g., critical infrastructures, as well as for safety-critical devices such as smart medical implants [86].

In this paper, we review the foundations of lattice-based cryptography, some of the more adopted instances of lattices in security, their implementations in software and hardware, and their applications to authentication, key exchange and digital signatures. The purpose of this survey is to focus on the essential ideas and mechanics of the cryptosystems based on lattices and the corresponding implementation aspects. We believe that this survey is not only unique, but also important for guidance in the selection of standards, as well as in the analysis and evaluation of candidate implementations that represent state-of-the-art.

The rest of the paper is organized as follows. In Section 2 we provide the relevant background to make the paper self-contained. In particular, we briefly describe different Lattice constructions and Lattice-based schemes in Section 2.5. Section 3 discusses various implementations of lattice-based schemes. In particular, Section 3.1 discusses related art on improvements for arithmetic computation, i.e., polynomial/matrix multiplication or/and noise sampler. Sections 3.2 and 3.3 present software and hardware implementations of different lattice-based schemes respectively. Sections 4 concludes with an outlook.

2 BACKGROUND

2.1 Standard lattices

Lattice $\mathcal{L}$ is the infinite set (discrete) of points in n-dimensional Euclidean space with a periodic structure [157]. A basis (B) of the lattice is defined as $\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_n \in \mathbb{R}^{d\times n}$ to be n-linearly independent vectors in $\mathbb{R}^d$. B is a $d \times n$ matrix in which i^{th} column is $\mathbf{b}_i$ vector such that $B = [\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_n]$. Integer n and d are rank and dimension of the lattice $\mathcal{L}(B)$. If $n = d$, $\mathcal{L}(B)$ is a full rank (or dimension) lattice in $\mathbb{R}^d$. All integer combinations generated by the basis matrix B (with integer or rational entries) are forming the lattice $\mathcal{L}$.

$$\mathcal{L}(B) = \{\mathbf{Bx} : \mathbf{x} \in \mathbb{Z}^n\} = \mathcal{L}(\mathbf{b}_1, ..., \mathbf{b}_n) = \left\{ \sum_{i=1}^{n} x_i \mathbf{b}_i : x_i \in \mathbb{Z}, 1 \leq i \leq n \right\} \tag{1}$$

where $\mathbf{Bx}$ is matrix-vector multiplication. It should be mentioned that B is not unique and multiple basis vectors exist for a specific lattice. Mathematically, basis matrices B and B' produce the same lattice, $\mathcal{L}(B) = \mathcal{L}(B')$, if and only if there is a an invertible matrix U such that $B = B'U$. It is easy to prove that any invertible matrix U is unimodular; which means the absolute value of its determinant is one ($|det(U)| = 1$, for $U \in \mathbb{Z}^{n\times n}$).

Let det($\mathcal{L}_B$) be the determinant of lattice $\mathcal{L}(B)$ which is equal to absolute value of the determinant of the basis matrix B such that $det(\mathcal{L}_B) = \{|det(B)| : B \in \mathbb{Z}^{n\times n}\}$. The determinant of a lattice corresponds to the inverse of the lattice density; a lattice which consists of denser set of points has the higher determinant.

B and B' generate the same lattice if and only if $|det(B)| = |det(B')|$. In other words, det($\mathcal{L}_B$) is independent of the basis and multiplication of a basis B with a unimodular matrix U (with determinant of +1 or -1) generates new basis ($B' = BU$) that produces the same lattice.

$\mathcal{L}^*$ is defined as the dual of the lattice $\mathcal{L}$ in $\mathbb{R}^n$. $\mathcal{L}^*$ is the set of all vectors $y \in \mathbb{R}^n$ whose inner products with the vectors $x \in \mathcal{L}$ results in $z \in \mathbb{Z}$. It is proven that for each $B \in \mathbb{R}^{n\times n}$, $\mathcal{L}(\mathcal{B})^* = \mathcal{L}((B^{-1})^T)$, where T transposes a matrix.

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

In other words, $\mathcal{L}(B)$ is generated by columns of nonsingular matrix B, while $\mathcal{L}(\mathcal{B})^*$ is generated by the rows of matrix B^{-1}.

If A is generated by the columns of the nonsingular matrix B, then A' is the lattice generated by the rows of B.

$$\mathcal{L}^* = \{y : \langle x, y\rangle \in \mathbb{Z}, x \in \mathcal{L}\} \tag{2}$$

A `q-ary` lattice is a lattice $\mathcal{L}$ that satisfies $q\mathbb{Z}^n \subseteq \mathcal{L} \subseteq \mathbb{Z}^n$ which means membership of a vector x in lattice $\mathcal{L}$ is determined by x mod q with integer q much smaller than det($\mathcal{L}$). Most Lattice-based cryptographic schemes employ `q-ary` lattices as the hard on average problem.

Let n (rank), m (dimension) and q (preferably a prime integer) be positive integers and A is a matrix in $\mathbb{Z}_q^{n\times m}$. Prime p is a small integer with $O(\log n)$ bits and m is a small multiple of n such that $m = O(n)$. Two forms of `q-ary` lattices are common in practical lattice-based cryptographic systems [132].

$\Lambda_q(A)$ is lattice that is generated by the rows of matrix A (mod q) and is closely related to Learning With Error (`LWE`) problem.

$$\Lambda_q(A) = \{y \in \mathbb{Z}^m : y = A^T s(\mathrm{mod} q), \mathbf{s} \in \mathbb{Z}^n\} \tag{3}$$

$\Lambda_q^{\perp}(A)$ (orthogonal Λ_q form) is the lattice whose vectors are orthogonal modulo q to the rows of matrix A which corresponds to the code whose parity check matrix is A mod q. In other words, for vectors x,$y \in \Lambda_q^{\perp}$, $x + y$ is a member of lattice $\Lambda_q^{\perp}$. It should be mentioned that $\Lambda_q^{\perp}(A)$ is used to solve the Short Integer Solution (`SIS`) problem.

$$\Lambda_q^{\perp}(A) = \{y \in \mathbb{Z}^m : Ay = 0(\mathrm{mod} q), \mathbf{s} \in \mathbb{Z}^n\} \tag{4}$$

For short integer solution (`SIS`) lattice problems $\Lambda_q^{\perp}$ and for learning with error (`LWE`) lattice problems Λ_q is used.

From definition, lattice Λ_q is dual of lattice $\Lambda_q^{\perp}$ such that $\Lambda_q^{\perp}(A) = q.\Lambda_q(A)^*$ and $\Lambda_q(A) = q.\Lambda_q^{\perp}(A)^*$.

2.2 Ideal lattices

An ideal lattice is defined over a ring $R = \mathbb{Z}_q[x]/(f(x))$, where $f(x) = x^n + f_n x^{n-1} + \ldots + f_1 \in Z[x]$ (cyclotomic polynomial) and R contains all the polynomials with modulo q integer coefficients. In the case where $f(x)$ is monic (leading coefficient is 1) irreducible polynomial degree-n, $R = \mathbb{Z}_q[x]/(f(x))$ includes polynomials of degree less that equal $n - 1$. For instance, $R = \mathbb{Z}_q[x]/(x^n + 1)$ is an ideal lattice when n is a power of 2; however, $R = \mathbb{Z}_q[x]/(x^n - 1)$ is not an ideal lattice since $(x^n - 1)$ is a reducible polynomial.

2.3 Lattice problems and their applications to cryptography

The breakthrough work of Ajtai [6] provides confidence for adopting lattice based schemes in cryptography. Specifically, Ajtai proves that solving some `NP-hard` lattice problems, for instance Shortest Vector Problem (`SVP`), in average-case is as hard as to solve them in the worst case assumption. In other words, in order to solve `SVP` problem, an algorithm should solve the problem for any input basis B (all instances of `SVP` problem should be solved).

It is conjectured that there does not exist a probabilistic polynomial time algorithm that can approximate certain computational problems on lattices to within polynomial (in the dimension) factors [132]. This is the basis for security of lattice-based cryptography schemes. The fastest algorithm to solve the `SVP` problem has the time and memory complexity of $2^{O(n)}$ [7] [133] [4]. In a n-dimensional lattice $\mathcal{L}$, **successive minima** is defined as sequence of $\lambda = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ where the ith success minima, λ_i, is the radius of smallest ball that contains i linearly independent lattice vectors. We take the below definitions from [129]:

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

2.3.1 Shortest Vector Problem (SVP). Three variants of the SVP [129] exist in the literature, all of which can be reduced to each other. The first is to find the shortest nonzero vector; the second is to find length of the shortest nonzero vector; and the third determines if the shortest nonzero vector is shorter than a given real number.

Given B as the lattice basis, $v \in \mathcal{L}(B)\backslash\{0\}\}$ is defined as the shortest nonzero vector in lattice $\mathcal{L}(B)$ such that $\|v\| = \lambda_1(\mathcal{L}(B))$. Output of the SVP problem is **a**, the shortest nonzero vector in the lattice which is unique. It should be mentioned that SVP can be defined to an arbitrary norm (we use norm 2 here as the Euclidean norm). In γ-Approximate Shortest Vector Problem (SVP_γ), for $\gamma \geq 1$, the goal is to find the shortest nonzero vector $v \in \mathcal{L}(B)\backslash\{0\}$ where $\|v\| \leq \lambda_1(\mathcal{L}(B))$. The special case of $\gamma = 1$ is equivalent to the exact SVP. The decision variant of the Shortest Vector Problem ($G_{AP}SVP_\gamma$) is defined as determining if $d < \lambda_1(\mathcal{L}(B)) \leq \gamma.d$ where d is a positive real number.

To date, there is no known polynomial time solution to solve SVP and its approximate version. Ajtai proved that SVP is a NP-hard problem for randomized reductions and solving SVP in average case is as hard as solving SVP in the worst case assumption.

2.3.2 Closest Vector Problem (CVP). Let B and t be the lattice basis and the target point (might not be a member of the lattice), respectively; CVP is defined as finding vector $v \in \mathcal{L}$ where its distance ($\|v - t\|$) to a target point is minimized [129]. In γ-Approximate Closest Vector Problem (CVP_γ), for $\gamma \geq 1$, the goal is to find vector $v \in \mathcal{L}(B)$ such that $\|v - t\| \leq \gamma.dist(t, \mathcal{L}(B))$ where $dist(t, \mathcal{L}(\mathcal{B})) = inf\{\|v - t\| : v \in \mathcal{L}\}$ is the distance of target point t to the lattice $\mathcal{L}$.

2.3.3 Shortest Independent Vectors Problem (SIVP). Given lattice basis B and prime integer q, the shortest independent vector problem (SIVP) is defined as finding n linearly independent lattice vectors $\{v = v_1, ..., v_n : v_i \in \mathcal{L}(B) for 1 \leq i \leq n\}$ that minimizes $\|v\|$=$\max_i\|v_i\|$ [129]. Given an approximate factor $\gamma \geq 1$, the $\gamma - approximate$ Shortest Independent Vector Problem ($SIVP_\gamma$) is defined as finding n-linearly independent vectors lattice vectors $\{v = v_1, ..., v_n : v_i \in \mathcal{L}(B)$ such that $\max_i\|v_i\| \leq \lambda_n(\mathcal{L}(B)$ where λ_n denotes the nth success minima. For a n-dimensional lattice, λ_i, i^{th} success minima, is the radius of the smallest ball that contains i linearly independent lattice vectors. The decision version of SIVP is called ($G_{AP}SIVP_\gamma$) and is defined as determined if $d < \lambda_n(\mathcal{L}(B)) \leq \gamma.d$ where d is a positive real number.

In problems described above, the approximate factor is a function of the lattice dimension. By increasing the lattice dimension, solving computationally hard lattice problems become harder.

There is a close relationship between lattice hard problems (like CVP and SVP) and two common average-case lattice-based problems, Learning With Error (LWE) [157] and Shortest Integer Solution (SIS) [6]. In the following we introduce SIS and LWE problems with their complexity.

2.3.4 Shortest Integer Solution (SIS). Let a_1, a_2,..., $a_n \in \mathbb{Z}^{m\times n}$ be an arbitrary vector and q is a integer prime number. The SIS problem is defined as finding the vector $x \in \mathbb{Z}^{m\times n}$ such that

$$x_1.a_1 + x_2.a_2 + ... + x_n.a_n = 0 \pmod q \tag{5}$$

Short is usually translates as $z_i \in \{-1, 0, +1\}$. Considering q-ary lattices, let A=(a_1, a_2,..., a_n) be a vector in $\mathbb{Z}^{m\times n}$ and $\Lambda_q^\perp(A) = \{z \in \mathbb{Z}^m : Az = 0 \pmod q$; SIS problem is to find the shortest vector problem for the lattice $\Lambda_q^\perp$. Ajtai proved that if a lattice problem like SVP is hard to solve in the worst case, an average case one-way function exists [6]. Based on the Ajtai's theorem, if polynomial time algorithm A solves the SIS problem, an efficient algorithm B exists that can solve SVP (or SVIP) problem for any lattice of dimension n in polynomial time.

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

Ring-SIS. Let A=(a_1, a_2,..., a_n) be a vector with $a_i \in \mathbb{Z}_q[x]/(x^n+1)$ and $\Lambda_q^{\perp}(A) = \{z \in \mathbb{Z}_q[x]/(x^n+1) : Az = 0 \pmod q\}$; Ring-SIS problem is to find the shortest vector problem for the lattice $\Lambda_q^{\perp}$ [176].

2.3.5 Learning With Error (LWE). Let a be the polynomial with coefficients sampled uniformly at random in $\mathbb{Z}_q^n$, where n and q are degrees of lattice and modulus (prime integer), respectively. Recovering the (**unique**) random secret s (uniformly at random in $\mathbb{Z}_q^n$) from m ($m \geq n$) samples of the form ($a, a.s + e \bmod q$) is known as the Learning With Error (`LWE`) problem where e is the error that is sampled from error distribution χ. Worst case hardness assumption of the LWE problem is proven under the quantum [157] and classical [42] reductions. Besides, If secret s is sampled from the same error distribution as e, hardness assumption of the LWE problem is still valid [15].

Ring-LWE. let $R_q = \mathbb{Z}_q[x]/(x^n+1)$ to be the ring of polynomials where n is power of 2 and q is an integer. Recovering random secret s with uniform coefficients in R from $m \geq 1$ samples ($a_i, a_i.s + e_i \bmod q$) is known as ring learning with error problem (`Ring-LWE`) where $e \in R$ is the error with coefficients sampled from error distribution χ [127].

Learning With Rounding (LWR). Complexity and inefficiency of the decisional (`LWE`) problem prohibit it to be used for PRGs. The learning with rounding (`LWR`) problem [22] is the "derandomized" variant of the `LWE` with improved speedup (by eliminating sampling small errors from a Gaussian-like distribution with deterministic errors) and bandwidth (by rounding Z_q to the sparse subset R_p). `LWE` hides the lower order bits by adding a small error; however, `LWR` concealed the lower order bits with rounding. Let a_i be sampled uniformly at random in Z_q^n and $\lfloor \cdot \rceil : Z_q \rightarrow Z_p$ for $p < q$ be the modular "rounding function" where $\lfloor x \rceil_p = \lfloor (p/q) \cdot x \rceil \bmod p$. Similar to `LWE`, `LWR` is defined as recovering the (**unique**) secret s (uniformly at random in $\mathbb{Z}_q^n$) from $m \geq 1$ samples $(a_i, \lfloor \langle a_i, s \rangle \rceil_p) \in Z_q^n \times Z_p$. In other words, there is no probabilistic polynomial time algorithm that can distinguish pairs of $(a_i \leftarrow Z_q, \lfloor \langle a_i, s \rangle \rceil_p) \in Z_q^n \times Z_p$ with $(a_i \leftarrow Z_q, \lfloor u \rceil_p)$ where u is uniformly random in Z_q. `LWR` is assumed to be hard under the hardness assumption of `LWE` when the number of samples is bounded.

Module-LWE. Let n and d be dimensions of R (degree of ring R_q) and rank of module $M \in R^d$. The hardness of a scheme with $R_q = \mathbb{Z}_q[x]/(x^n+1)$ and $d = 1$ is based on Ring-LWE and Ring-SIS problems; however $R_q = \mathbb{Z}_q$ and $d > 1$ end up with LWE and SIS problems. Module-LWE/SIS is a generalization of LWE/SIS and Ring-LWE/SIS in which the parameters are $R = \mathbb{Z}_q[x]/(x^n+1)$ and $d \geq 1$. Security reduction of lattice problems in module M depends on $N = n \times d$ (dimension of the corresponding module lattice) [117]. Suppose A is a $d \times d$ random matrix; security reduction of LWE/SIS problem for $d = 1$ (Ring-SIS/LWE) and ring of dimension n is the same as $d = i$ and a ring of dimension n/i. In the former case, matrix A contains n elements in Z_q; however, in the latter case, A contains $i^2 \times n$ elements in Z_q.

Let $R_q = \mathbb{Z}_q[x]/(x^n+1)$ and $R = \mathbb{Z}[x]/(x^n+1)$ be rings of polynomials where n is power of 2 and $q \in Z$. Vector a is uniformly sampled in R_q^d. Recovering the random secret s with coefficients sampled from the error distribution χ^d in R from $m \geq 1$ samples $(a_i, a_i.s + e_i \bmod q) \leftarrow R_q^d \times R_q$ is known as module learning with error (`Module-LWE`) problem where $e_i \in R$ is the error with coefficients sampled from error distribution χ [117].

`Module-LWE`/`Module-LWR` is a middle ground problem for `LWE`/`LWR` and their ring variant `RLWE`/`RLWR` which reduces computational pressure and bandwidth of the standard lattices and improve security of the ideal lattices (with the ring structure). With the same arithmetic foundation, *Module-LWE*/*Module-LWR* provides a trade-off between the security and cost (computation and bandwidth).

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

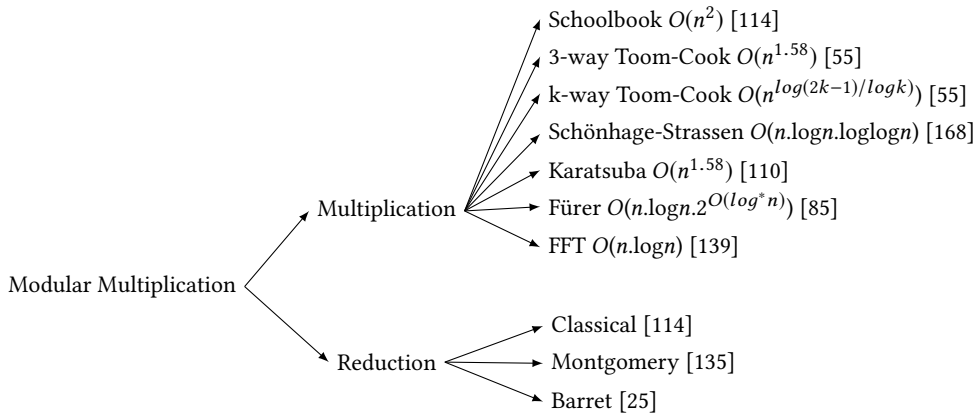

Fig. 1. Modular Multiplication

2.4 Arithmetic and Components of Lattices

In this section we provide an evaluation of the components in a lattice-based cryptosystem that guides the actual implementation. To components are critical: (a) the Polynomial multiplication for ideal lattices, and matrix multiplication for standard lattice are the main speedup bottlenecks; (b) The discrete Gaussian sampling is used to sample noise in order to hide the secret information. There are various algorithms for the sampler and multiplier in the literature, with providing the designer a specific goal [138]. We briefly review different algorithms and outline their practical implementations in Section 3.1

There exist two main classes of lattice-based algorithms used in cryptography, namely NTRU and Learning with Error (LWE). The security of NTRU is based on hardness not-provably reducible to solving the Closest Vector Problem (CVP) in a lattice, whereas the security of LWE relies on provably reducible solving the Shortest-Vector Problem (SVP) in a lattice. Consequently, NTRU suffers from security guarantees, but in practice provides more flexibility and efficiency in the implementation. On the contrary, LWE problems are resistant to quantum attacks, while their relatively inefficient nature led researchers to devise more efficient formulations, e.g., over rings - Ring Learning with Errors (Ring-LWE).

Implementations are broadly classified in pure software, pure hardware, and hardware/software co-design cryptographic engines [138].

The implementation of the modulo arithmetic (multiplication and addition of big numbers) is important, as the arithmetic clearly is one of the most time consuming parts in lattice-based schemes. For Standard LWE schemes, matrix multiplication algorithms are adopted, whereas number theoretic transform (NTT) is a better choice for polynomial multiplication in Ring-LWE. A summary of modular arithmetic is presented in Figure 1.

In addition to the arithmetic portion, a bottleneck in lattice-based schemes is the extraction of the random term, which usually is implemented with a discrete noise sampler (from a discrete Gaussian distribution) and can be done with rejection, inversion, Ziggurat, or Knuth-Yao sampling with moderate standard deviation for key exchange and public key encryption, and small standard deviation for digital signature to achieve a compact and secure signature.

When implemented, Standard LWE-based schemes exhibit a relatively large memory footprint due to large key size - hundreds of kilobyte for public key - which render a straightforward implementation of standard LWE-based schemes impractical. The adoption of specific ring structures, e.g., Ring-LWE, offers key size reduction by a factor of n compared

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

to Standard LWE [127], making Ring-LWE an excellent candidate for resource constrained devices, such as Wi-Fi capable smart devices, including medical implants. Another avenue to address resource constrained devices is that memory footprint can be traded off with security assurance, which improves both efficiency and memory consumption.

High-performance Intel/AMD processors, which are notoriously equipped with AVX vector instructions, and ARM/AVR micro-controllers are popular platforms for software implementations, as we will see in more detail in the coming sections. Practical software implementations of standard lattices, encryption scheme [103] and key exchange [37], have been published recently. For hardware implementations, FPGAs provide flexibility and customization but not agility. Hardware implementations of lattice-based schemes (e.g., BLISS-I [71] with higher security level) are about an order of magnitude faster than hardware implementation of RSA-2048 [104]. Furthermore, another candidate platform to implement lattice-based schemes are application specific integrated circuits (ASICs) of which there appear to be no such implementations in the literature at this time. However, the main advantages and challenges for ASIC design of lattice-based schemes are presented in [140].

2.4.1 The multiplier component. Standard lattice arithmetic operations involve calculations over matrices; however, ideal lattices benefit from their specific structure and operate on polynomials. Matrix multiplication is used for the standard lattices, while polynomial multiplication is employed for ideal lattices.

Arithmetic operations for a Ring-LWE based scheme are performed over a ring of polynomials. The most time and memory consuming part is the polynomial multiplication. The easiest way to multiply two polynomials is to use the Schoolbook algorithm with the time complexity of $O(n^2)$ [114]. Let n and p be degree of the lattice (n is a power of 2) and a prime number ($p = 1 \bmod 2n$), respectively. Z_p denotes the ring of integers modulo p and x^{n+1} is an irreducible degree n polynomial. The quotient ring R_p, contains all polynomials with the degree less than n in Z_p, that defines as $R_p = Z_p/[x^{n+1}]$ in which coefficients of polynomials are in the range $[0,p)$.

The number theoretic transform (NTT) is a generalization of Fast Fourier Transform (FFT), which is carried out in a finite field instead of complex numbers. The latter could achieve time complexity of $O(n\log n)$. In other words, exp(-2πj/N) with n^{th} primitive root of unity ω_n which is defined as the smallest element in the ring that $\omega_n^n = 1 \bmod p$ and ${\omega_n}^i \neq 1 \bmod p$ for i ≠n. The main idea behind this is to use the point value representation instead of the coefficient representation by applying NTT in $O(n\log n)$; thereafter performing point-wise multiplication in $O(n)$ and finally converting the result to coefficient representation by applying Inverse NTT (INTT) in $O(n\log n)$.

$$a(x) \times b(x) = NTT^{-1}(NTT(a) \odot NTT(b)) \tag{6}$$

Where $\odot$ is the point-wise multiplication of the coefficients. If NTT is applied to $a(x)$ with $(a_0, \ldots, a_{n-1})$ as coefficients, we would have:

$$(\hat{a_0}, \ldots, \hat{a_{n-1}}) = NTT(a_0, \ldots, a_{n-1}) \tag{7}$$

$$\hat{a_i} = \sum_{j=0}^{n-1} a_i \omega^{ij} mod(p), i = 0, 1, \ldots, n-1 \tag{8}$$

In order to retrieve the answer from point value representation using NTT^{-1}, it is sufficient to apply NTT function with a slight change as indicated below:

$$a_i = \sum_{j=0}^{n-1} \hat{a}_i \omega^{-ij} mod(p), i = 0, 1, \ldots, n-1 \tag{9}$$

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

As can be seen, if instead of ω we use $-\omega$ and divide all the coefficients by n, NTT^{-1} is computed.

In order to compute NTT(a), we pad the vector of coefficients with n zeros which culminates in doubling the input size. By employing negative wrapped convolution technique [182], there is no need to double the input size.

To improve efficiency of polynomial multiplications with NTT, combining multiplications of powers of ω with powers of ψ and ψ^{-1} ($\psi^2 = \omega$) can be beneficial which requires storage memory for precomputed powers of ω and ψ^{-1} in bit-reversed order. The combining idea is used in [121, 153, 161]. It should be mentioned that NTT can be used only for the $p = 1 \pmod{2n}$ case where p is a prime integer. Suppose $a' = (a_0, \psi a_1, ..., \psi^{n-1} a_{n-1})$, $b' = (b_0, \psi b_1, ..., \psi^{n-1} b_{n-1})$, $c' = (c_0, \psi c_1, ..., \psi^{n-1} c_{n-1})$ to be coefficient vectors of the a, b, c that are multiplied component-wise by $(1, \psi^1, ..., \psi^{n-1})$. Based on the negative wrapped convolution theorem, modulo $(x^n + 1)$ is eliminated and the degree of NTT and NTT^{-1} is reduced from $2n$ to n.

$$c' = NTT^{-1}(NTT(a') \odot NTT(b')) \tag{10}$$

where $c(x)$ is the polynomial multiplication of $a(x)$ and $b(x)$.

$$c = (\psi^0 c'_0, \psi^{-1} c'_1, ..., \psi^{-n+1} c'_{n-1}) \tag{11}$$

Other popular multiplication algorithms in literature are the Karatsuba algorithm (with time complexity of $O(n^{log3/log2})$ [110]), and subsequent variants of it [59]. Schönhage-Strassen with time complexity of $O(n.\log n.\log\log n)$ [168] outperforms the Karatsuba algorithm [48].

An extensive analysis and comparison for hardware complexity of various modular multiplications, including Schoolbook (classical), Karatsuba, and FFT, with different operand size are presented in [59]. Authors calculate hardware complexity of each multiplier by decomposing it into smaller units such as full adder, half adder, multiplexer, and gate. Rafferty et al. [156] adopt the same approach to analyze large integer multiplications of both combined and individual multipliers. Karatsuba multiplier outperforms for operands greater of equal to 32 bits. Schoolbook imposes large memory footprint in order to store partial products which negatively impact performance that is mitigated by Comba [54] with the same time complexity but relaxing memory addressing by optimizing the sequence of partial products. Rafferty et al. compare hardware complexity, in terms of $+$, $-$, and $*$ units, of different combinations of classical (Comba), Karatsuba, and FFT (NTT for integers) for up to multipliers with 65536-bit operands. However, they evaluate the latency and clock frequency by implementing in hardware (Xilinx Virtex-7 FPGA) for up to 256-bits for combination of NTT+Comba and 4096-bit for Karatsuba+Comba which are not in a good range for lattice-based cryptography (e.g. $1024 \times 14 = 14336$ bits are used for Newhope key exchange). Based on their (analytical) hardware complexity analysis, combination of Karatsuba-Schoolbook is the best choice for operands under 64 bits. Karatsuba-Comba is preferable for operands for 64 bits to 256 bits. For larger operands, the lowest hardware complexity is achieved by combined multiplier NTT-Karatsuba-Schoolbook. It should be mentioned that results are for a single multiplication. Authors make some assumption for the sake of simplicity such as the shift operation is free and inputs are ready.

2.4.2 The Sampler Component. The quality of a discrete Gaussian sampler is determined by a tuple of three parameters: (σ,λ,τ). In such a tuple σ is the standard deviation (adjusts dispersal of data from the mean), λ is the precision parameter (controls statistical difference between a perfect and implemented discrete Gaussian sampler), and τ is the distribution tail-cut (determines amount of the distribution that we would like to ignore). Each of these parameters affects the security and efficiency of the sampler. For instance, smaller standard deviation decreases the memory footprint required to store precomputed tables. For encryption/decryption schemes σ=3.33 [118] is suggested. Digital

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

signature sampling from a Gaussian sampler involves large σ=215 [71]. However, by employing Peikert's convolution lemma [142], standard deviation can be reduced by an order of magnitude which is a remarkable improvement on the precomputed table size. Speed and memory footprint is λ dependent, i.e. higher λ results in more secure but a slower and bigger sampler. The tail of the Gaussian distribution touches the x-axis at x=+∞ (considering only the positive side due to the symmetry) with negligible probability. Tail-cut parameter (τ) defines the amount of the distribution that we would like to ignore, hence random number e is sampled in $|e| \in \{0, \sigma \times \tau\}$ instead $|e| \in \{0, \infty\}$. Instead of the statistical distance, Rényi divergence technique [21] can be employed to measure the distance between two probability distributions (e.g., [12]). More precisely, in lattice based cryptography, Rényi divergence is used to generalized security reductions. Sampling the discrete Gaussian distribution is one the most time and memory hungry parts of lattice-based cryptosystems, due to the demands of high precision, numerous random bits, and huge lookup tables. To be more specific, obligatory negligible statistical distance between the implementation of a Gaussian sampler (approximated) and the theoretical (perfect) discrete Gaussian distribution imposes expensive precise floating point arithmetic (to calculate the exponential function) or large memory footprint (to store precomputed probabilities). To keep statistical distance less than $2^{-\lambda}$, floating point precision with more than standard double-precision is obligatory which is not natively supported by the underlying platform; thus software libraries should be used to perform higher floating-point arithmetic. It is impractical to sample from a perfect Gaussian sampler; hence an λ-bit uniform random integer is used to approximate the discrete sampler. Fortunately, it is proven that the Gaussian sampler could be used to achieve $\lambda/2$ security level (approximated sampler) instead of λ level (perfect sampler), since (to date) there is no algorithm that can distinguish between a perfect sampler (λ bits) and an approximate sampler ($\lambda/2$ bits) [163]. In other words, we can cut half of the bits in the sampler which results in a smaller and faster sampler [24, 103, 164]. Reduction in precision parameter (from λ to $\lambda/2$) changes tail-cut parameter (τ) as $\tau = \sqrt{\lambda \times 2.ln(2)}$ [102]. Sampling from a Gaussian distribution may lead to a timing side channel attack, which can be avoided by using a generic constant time Gaussian Sampling over integers [134]. Folláth provides a survey on different Gaussian samplers in lattice based cryptography schemes with a more mathematical outlook [82]. Gaussian samples are classified into six categories and guidelines provided for choosing the best candidate on different platforms for specific parameter ranges. However, we organize Gaussian samplers into the following types discussed below. A Summary of advantage and disadvantages of each sampler are listed in Table 1.

Rejection Sampler. Firstly, x is sampled in (-$\tau\sigma$,$\tau\sigma$) uniformly at random where τ and σ are tail-cut and standard deviation of Gaussian distribution. Thereafter, x is rejected with the probability proportional to 1-exp(-$x^2/2\sigma^2$). High rejection rate of samples (on average 8 trials to reach acceptance) along with expensive calculation of `exp()` are the main reasons for the inefficiency [181].

The rejection sampler was employed for the first time within lattice-based cryptosystems in [89]. Remarkable speed and area improvement could be achieved by accomplishing rejection operation using lazy floating-point arithmetic [75].

Bernoulli Sampler. Bernoulli is an optimized version of rejection sampling in order to reduce average required attempts for a successful sampler from 10 to around 1.47 with no need to calculate `exp()` function or precomputed tables [147]. Bernoulli is introduced in [71] for lattice-based cryptography and used widely in the research community [103, 149, 152].

The main idea behind Bernoulli sampling is to approximate sampling from $D_{\mathbb{Z}, k\sigma_2}$ using the distribution $k \cdot D_{\mathbb{Z}^+, \sigma_2} + \mathbb{U}(\{0, ..., k-1\})$ where $\mathbb{U}$ is the uniform distribution. The procedure for Bernoulli sampler is shown below in 5 steps.

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

(1) sample $x \in \mathbb{Z}$ according to $D_{\mathbb{Z}^+,\sigma_2}$ with probability density of $\rho_{\sigma_2} = e^{-x^2/(2\sigma_2^2)}$
(2) sample $y \in \mathbb{Z}$ uniformly at random in $\{0, ..., k-1\}$ and calculate $z \leftarrow y + kx$, $j \leftarrow y(y + 2kx)$
(3) sample $b \leftarrow \mathcal{B}_{-j/2\sigma^2}$ where $\sigma = k\sigma_2$ and $\mathcal{B}$ is Bernoulli distribution. To sample from $\mathcal{B}_c$ where c is a precomputed constant value, a uniform number $u \in [0, 1)$ with λ-bit precision is sampled; 1 is returned if $u < c$, otherwise 0 should be returned.
(4) if $b = 0$ goto to step (1)
(5) if $z = 0$ go to step (1), otherwise generate $b \leftarrow \mathcal{B}_{1/2}$ and return $(-1)^b z$ as the output

The standard deviation of target Gaussian sampler $D_{\mathbb{Z},k\sigma_2}$, equals $k\sigma_2$ where $\sigma_2 = \sqrt{\frac{1}{2ln2}} \approx 0.849$ is standard deviation of the binary Gaussian sampler $D_{\mathbb{Z},\sigma_2}$ and $k \in \mathbb{Z}^+$ is the uniform distribution parameter.For schemes with small standard deviation (e.g. public key encryption) sampling from binary Gaussian distribution can be eliminated [152], while for digital signatures with large standard deviation, using Gaussian distribution is mandatory [149]. It should be mentioned that Gaussian distribution can be replaced with other distributions (e.g., uniform distribution [91]). The Bernoulli approach eliminates long integer calculation; instead, single bit operations are the most frequent operation, hence it is a good candidate for hardware implementation. However, the time dependency of Bernoulli makes it vulnerable to timing attacks which is resolved in hardware implementation of BLISS in [149]. Precomputed tables in Bernoulli sampling are small, besides binary Gaussian distribution (easy to sample intermediate sampler) is independent of σ hence the Peikert convolution lemma (smaller σ) does not have a considerable effect on the area. However, convolution lemma reduces area of precomputed tables in CDT sampler by a factor of 23× for `BLISS` (reduces σ from 215 to 19.47).

Binomial Sampler. Centered Binomial distribution (ψ_k) is a close approximation of rounded Gaussian sampler (ξ_σ) which eliminate the need both computing $exp()$ and precomputed large tables. Let $\sigma = \sqrt{8}$ be standard deviation of ξ_σ and a binomial distribution is parameterized with $k = 2\sigma^2$; choosing ψ_k as the sampling distribution has negligible statistical difference with rounded Gaussian sampler with $\sigma = \sqrt{8}$ [12]. Centered Binomial distribution(ψ_k) for integer $k \geq 0$ is defined as sampling $2 \cdot k$ random numbers uniformly from $\{0, 1\}$ as $(a_1, ..., a_k, b1, ..., b_k)$ and output $\sum_{i=1}^{k}(a_i, b_i)$ as the random sample [12]. Since k scales with power 2 of σ, it is not practical to use binomial sampling for digital signatures with large standard deviation.

Binomial sampler has been employed inside software implementation of NewHope [13, 14, 148, 178], HILA5 [165], LAC [123], LIMA [172], Kyber [16, 38] and Titanium [177]. It also is used in hardware implementation of NewHope [180].

Ziggurat Sampler. Ziggurat sampler is a variation of rejection sampler introduced in [128] for a continuous Gaussian sampler [4]. Discrete version of Ziggurat sampler is proposed in [43] which is suitable for schemes with large standard deviation. The area under the probability density function is divided into n rectangles with the same area whose size is proportional to the probability of sampling a point in each rectangle. The left and right corner of each rectangle is on the y-axis and Gaussian distribution curve, respectively. Each rectangle is stored using its lower right coordinates **??**. Firstly, rectangle R_i and point x_i inside the rectangle is chosen uniformly at random. Since we are considering positive x_i, a random sign bit, s, is also required. If $x_i \leq x_{i-1}$, x_i resides below the curve and would be

[4]Another method to sample from a continuous Gaussian is **Box-Muller** [40]. Box-Muller method transforms two independent uniforms into two independent discrete Gaussian distributions. pqNTRUSign [46] and NTRUEncrypt [184] use Box-Muller based Gaussian sampler.

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

accepted. Otherwise, a uniformly random y_i is sampled and if $y_i \leq exp(x_i)$, the random point (x_i, y_i) is accepted; otherwise new random x should be sampled and the process is repeated.

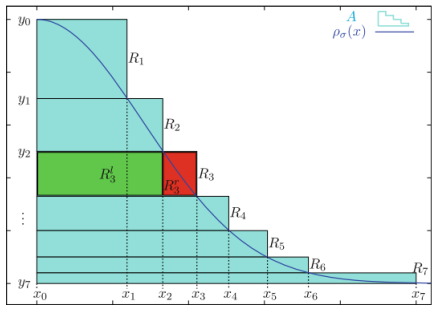

Fig. 2. A partition of Ziggurat [43].

More rectangles reduce the number of rejections and hence results in higher performance and precision. Due to its flexibility offers a trade-off between memory consumption and performance, Ziggurat sampler is a suitable candidate for resource constrained embedded devices. More precision and performance require more precomputed rectangles which impose remarkable memory overhead; however the increase in number of precomputed rectangles could drop the performance if the cache is fully occupied. Consequently, software implementation is preferred to any hardware implementation. There is only one hardware implementation of Ziggurat in the literature which hints at the impracticality of realizing the Zigguart sampler in hardware [102].

Cumulative Distribution Table (CDT) Sampling. CDT is also known as the inversion sampling method which introduced in [142]. CDT is faster than rejection and Ziggurat samplers due to elimination of expensive floating-point arithmetic [179]. Since in cumulative distribution all the numbers are less than 1, it is sufficient to use binary expansion of the fraction. CDT requires a large table to store values of the cumulative distribution function (CDF) of the discrete Gaussian (highest memory footprint [82]) for which their size is a function of distribution tail-cut (τ) and Gaussian parameter (σ). Variable r is sampled uniformly at random in the range [0,1) with λ bits of precision. The goal is to find an x whose probability is $\rho(x) = S[x+1] - S[x]$ where $S[x]$ equals the value of CDF at x. CDF sampler performs a search, usually binary search, on the CDF, which is precomputed and stored in the memory, to find the corresponded x. A standard CDT needs table of size at least $\sigma\tau\lambda$ bits; for instance `BLISS-IV` ($\sigma = 13.4, \tau = 215, \lambda = 128$) and `BLISS-I` ($\sigma = 13.4, \tau = 19.54, \lambda = 128$) need at least pre-computed tables of size 630 kbits and 370 kbits for 192 and 128 bit post quantum security, which is not practical for resource constrained embedded devices [71]. By employing Peikert's convolution lemma [142], the size of precomputed tables are dropped by the factor of 11 by sampling twice from $\tau' = \frac{\tau}{\sqrt{1+k^2}} = 19.47$ instead of sampling from a $D_{Z,\tau}$ with $\tau = 215$. Further improvements on table size are presented in [147] by employing adaptive mantissa size which halves the table size.

A hardware implementation of CDT sampler is proposed in [150]. Further improvements of hardware implementations are accomplished in [65, 149, 151]. An efficient software implementation of CDT is proposed in [64]; however it is vulnerable to timing attack which is resolved in [112] by suggesting a time independent CDT sampler. Authors of [152] combine Cumulative Distribution Function (CDF) and rejection sampling to achieve a compact and reasonably fast Gaussian sampler.

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

Table 1. Comparison of different Gaussian samplers; partially extracted from [71].

Sampler	Speed	FP exp()	Table Size	Table Lookup	Entropy	Features
Rejection	slow	10	0	0	$45+10log_2\sigma$	suitable for constrained devices
Ziggurat	flexible	flexible	flexible	flexible	flexible	suitable for encryption requires high precision FP arithmetic not suitable for HW implementation
CDT	fast	0	$\sigma\tau\lambda$	$log_2(\tau\sigma)$	$2.1+log_2\sigma$	suitable for digital signature easy to implement
Knuth-Yao	fastest	0	$1/2\sigma\tau\lambda$	$log_2(\sqrt{2\pi e}\sigma)$	$2.1+log_2\sigma$	not suitable for digital signature
Bernoulli	fast	0	$\lambda log_2(2.4\tau\sigma^2)$	$\approx log_2\sigma$	$\approx 6+3log_2\sigma$	suitable for all schemes
Binomial	fast	0	0	0	$4\sigma^2$	not suitable for digital signature

Knuth-Yao sampler. The Knuth-Yao sampler [115] provides a near optimal sampling (suitable for the high precision sampling) thanks to its near entropy consumption of the random bits required by the sampling algorithm. Assume n to be number of possible values for each random variable r with the probability of p_r. The probability matrix is constructed based on the binary expansion of each variable whose r^{th} row denotes the binary expansion of p_r. According to probability matrix, a discrete distribution generating binary tree (DDG) is built whose i^{th} level corresponds to the ith column of the probability matrix. Sampling is the procedure of walking through the DDG tree until a leaf is reached and its value returned as the sampling value. At each level, a uniformly random bit indicates whether the left child or right child of the current node should be visited in the future. The Knuth-Yao sampler is suitable for schemes with small standard deviations; thus Knuth-Yao is not suitable for digital signature because of its slow sampling caused by high number of random bits. In order to minimize the statistical distance between the approximated distribution and the true Gaussian distribution, Knuth-Yao sampler needs large memory to store probability of the sample points with high precision, which is a an issue on resource constrained platforms. Combination of Knuth-Yao and CDT results in about halving the table sizes, which is still prohibitively large; however, the Bernoulli sampler offers the best precomputed table size [71].

De Clercq et al. [60] introduce an efficient software implementation of the Ring-LWE based cryptosystem by using Knuth-Yao as a Gaussian sampler. Using a column-wise method for sampling, Roy et al. [162] propose the first hardware implementation of the Knuth-Yao sampling with small standard deviation which results in faster sampling. Same authors improve their implementation in [160].

Based on the presented results on [43], with the same memory budget, CDT beats rejection sampling and discrete Ziggurat. The Ziggurat sampler outperforms CDT and rejection sampling for larger values of the standard deviation. Ziggurat sampler bears almost the same speedup as Knuth-Yao, while it improves the memory footprint by a factor of 400. As stated earlier, due to the large standard deviation necessary for digital signature, Knuth-Yao sampler is not suitable for digital signatures. Inside an encryption scheme with $\sigma_{LP} = 3.3$ [118], with the same security level ($\lambda = 64$), Knuth-Yao sampler beats CDT in terms of number of operations performed in one second per slice of FPGA (Op/s/S) for time independent implementation [112]. However, for time dependent Gaussian sampler with the same security level ($\lambda = 94$), the CDT sampler proposed by Du and Bai [65] outperforms Knuth-Yao implementation of [160] in term of Op/s/S.

The Size of precomputed tables in a Bernoulli sampler is two orders of magnitude smaller than that of CDT and Knuth-Yao sampler [71], however CDT has three times more throughput than Bernoulli for hardware implantation of `BLISS` in [149].

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

2.5 Lattice-based schemes

Security of the lattice-based cryptography schemes are based on hardness of solving two average-case problems, a.k.a Short Integer Solution (SIS) [131] and the Learning With Errors (LWE) problem [157].

Regev [157] proposed the Learning With Errors problem which can be reduced to a worst-case lattice problem like the Shortest Independent Vectors Problem (SIVP). In the proposed scheme, the ciphertext is $O(n\log n)$ times bigger than plaintext; however, in [144] ciphertext has the same length order compared to the plaintext. A smaller key size for LWE-based encryption is introduced as the LP scheme in [118]. Another difference between Regev's encryption scheme and LP scheme is that the former uses discrete Gaussian sampler in the key generation step, while LP employs Gaussian Sampler in encryption in addition to the key generation step.

2.5.1 Public Key Encryption. Public key encryption (PKE) is employed to encrypt and decrypt messages between two parties. Additionally, it is a basis for other cryptography schemes such as digital signatures. Generally, it consists of three steps including key generation, encryption, and decryption. The first party (Alice) generates two set of keys and keeps one of them private (sk_{Alice}) and distributes the other key (pk_{Alice}) to other party (Bob). In order to send the message M to Alice, Bob encrypts the message as $S = Enc(M, pk_{Alice})$ where pk_{Alice} is Alice's public and Enc is encryption cipher. Thereafter, Alice decrypts the message as $M = Dec(S, sk_{Alice})$ where Dec is the decryption cipher and sk_{Alice} is Alice's private key. Similarly, Alice should encrypt her message with Bob's public key (pk_{Bob}) if she wants to send message to Bob. Encryption and decryption ciphers are one-way functions and are known publicly; hence the only secret data are private keys of the recipients which should be impossible to uncover using their corresponding public keys. Practical lattice-based PKE schemes are either based on NTRU related [99, 174] or LWE [157] (and its variants including RLWE [127], MLWE [117], ILWE [95] and MPLWE [159]) assumptions.

2.5.2 Digital Signature. Digitally signing a document involves sending the signature and document separately. In order to verify the authenticity of the message, the recipient should perform verification on both the signature and the document. Digital signature consists of three steps including key generation, Sign_{sk}, and Verify_{pk}. In the first step, secret key (sk) and public key (pk) are generated; signer keeps the secret key and all verifier parties have the public key of signer. During the sign step, signer applies the encryption algorithm on input message M with its private key and produces output S as $S = \text{Sign}_{sk}(M, sk_{signer})$. Signer sends the tuple (M,S) to the Verifier who applies Verify_{pk} (M,S) and outputs 1 if M and S are a valid message and signature pair; otherwise, S is rejected as the signature of message M. As an example of hash and sign procedure: in the sign step, message M is hashed as $D = h(M)$ where D is the digest. Signer applies the encryption algorithm on D with its private key with the output of $S = Enc(D, sk_{signer})$. Afterwards, signer sends the pair of (M,S) to the verifier. Subsequently, verifier uses public key to decrypt S as $D' = Dec(S, pk_{signer})$. Then verifier compares $D = h(M)$ (same hash function is shared between signer and verifier) and D'; signature is verified if $D' = D$; otherwise, the signature is rejected. The steps outlined for sign and verify are just an example (used in RSA); hence sign and verify steps might be slightly different for various signature schemes, but follow the same idea.

Lattice-based signature schemes belong to one of two classes including hash-and-sign (e.g. GPV [88]) and Fiat-Shamir signatures (e.g. BG [20], GLP [91], and BLISS [71]). GPV is a provably secure framework to obtain "hash-and-sign lattice-based signature schemes"; GGH [1] and NTRUSign [96] were the two first works that propose lattice-based signatures which are not provably secure (due to the deterministic signing). Because of the high standard deviation of Gaussian sampler, implementation of lattice-based digital signatures that use Gaussian sampler is challenging. Besides, the need for hash function components and rejection step makes digital signature more complex.

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

2.5.3 Key exchange mechanism. Key exchange is the process of exchanging keys between two parties in the presence of adversaries. If parties use symmetric keys, the same key is shared between them; otherwise, public key of parties should be exchanged. There are numerous public key exchange methods reported in the literature. For classical cryptography, Diffie–Hellman is a practical public key exchange that has been used widely. Establishing a shared key in a lattice-based key encapsulation (KEM) or key exchange (KEX) scheme can be done by either a *reconciliation-based* or *encryption-based* method [13]. Ding [63] propose the first lattice-based key agreement using the reconciliation-based method. There are plenty of reconciliation-based key agreement schemes based on the LWE (e.g., Frodo [37]), RLWE (e.g., BCNS [39] and NewHope [12]) and MLWE (e.g., Kyer[38]) that require less bandwidth compare to the simper encryption-based ones (e.g., NewHope-Simple [13] and NewHope-512/1024 [148]).

Fujisaki-Okamoto transform. In the random oracle model, an IND-CPA [5] public key encryption (PKE) can be transformed into a IND-CCA [6] PKE using the $Fujisaki - Okamoto$ (FO) transform [84]. Hofheinz et al. [101] introduce a variant of the FO transform that performs transformation from CPA-security into CCA-security in the quantum random oracle model. By applying this variant of FO on a CPA-secure PKE, an IND-CCA key encapsulation mechanism (KEM) is achieved. This transformation is widely used in submitted proposals to "NIST call for post-quantum algorithms" [3] where authors first make a IND-CPA PKE and then build the CCA-KEM, with the same parameter space, by applying the KEM version of the transform [101]. **Note**. There are plenty of submitted proposals to the NIST PQC standardization call that a single proposal (e.g., LOTUS [145]) supports both public key encryption (e.g., LOTUS-PKE) and key agreement (e.g., LOTUS-KEM); in Table 2 and Table 3, we list those proposals as two separate schemes. To avoid the redundancy, in section 3, we describe each scheme under only one of the categories of public key encryption (section 3.2.1) or key exchange mechanism (section 3.2.2).

Similarly, to avoid the redundancy, an article with both software and hardware implementations is mentioned only once (software implementation in section 3.2 or hardware implementation in section 3.3).

The contemporary lattice-based schemes existed in the literature are listed in Table 2. Details on the security level, public key, secret key and ciphertext size of lattice-based public key encryption (PKE), key establishment (KEX/KEM) schemes can be seen in Table 3. Furthermore, details on the security level, secret key, public key, and signature size of the lattice-based digital signature schemes are listed in Table 4.

3 IMPLEMENTATION CHALLENGES

In this section, we consider various implementations of lattice-based cryptographic schemes using software, hardware, software/hardware codesign and DSP techniques. First we focus on the implementation of key arithmetic modules such as Gaussian sampler and matrix/polynomial multiplication in Section 3.1. Software and hardware implementations of lattice-based cryptographic schemes are described in Section 3.2 and Section 3.3, respectively. Implementations of lattice-based schemes using hardware/software codesign techniques are described in Section 3.4. The only implementation of a lattice-based scheme using DPS implementation is described in Section 3.5.

Table 5 presents a birds-eye view of popular implementation schemes, and can be helpful as a visual/organizational reference during the discussion of various implementation schemes.

[5]Indistinguishability under chosen plaintext attack
[6]Indistinguishability under chosen ciphertext attack

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

Table 2. Contemporary lattice-based schemes.

Lattice Type	Schemes		
	Public Key Encryption	**Digital Signature**	**Key Exchange**
Standard Lattices	LP [118] Lizard [49, 50] NTRUEnrypt [98] EMBLEM.CPA [169] FrodoPKE [137] LOTUS-PKE [145] Odd-Manhattan [146] uRound2.PKE [87]	GGH [1] NTRUSign[1][96] GPV [88] Lyubashevsky [124] BG [20] TESLA [11]	Frodo [37] EMBLEM [169] FrodoKEM [137] spKEX[31] OKCN/AKCN-LWE/LWR[4][109, 185] Lizard [50] LOTUS-KEM [145] Odd-Manhattan [146] uRound2.PKE [87]
Ideal Lattices	NTRU [99] NTRU Prime[28] Ring-Lizard (RLizard) [49, 50] trunc8 [167] HILA5 [165, 166] R.EMBLEM.CPA [169] NewHope-CPA-PKE [148] ntru-pke, ss-ntru-pke [184] u/nRound2.PKE [87]	Lyubashevsky [125] GLP [91] GPV [79] BLISS [71] BLISS-B[70] Ring-TESLA [51] TESLA# [24] BLZZRD [164] GLYPH[2][52] FALCON [83] qTESLA [32]	JARJAR, NewHope [12] NewHope-Simple [13] BCNS [39] HILA5 [165, 166] NTRU KEM [107] Ding Key Exchange [62] R.EMBLEM [169] OKCN/AKCN-RLWE[4][109, 185] AKCN/OKCN-SEC[4][185] LIMA-sp/2p [172] RLizard [50] NewHope-CPA/CCA-KEM [148] ntru-kem, ss-ntru-kem [184] NTRU-HRSS-KEM [108] Streamlined-NTRU-Prime, NTRU-LPRime [29] u/nRound2.KEM [87]
Module Lattices	Kyber PKE [16, 38] AKCN-MLWE-CCA[4][185] KINDI$_{CPA}$ [23] SABER [58]	Dilithium [72, 73] pqNTRUSign [46]	Kyber KEM[16, 38] CNKE, OKCN/AKCN-MLWE[4][185] KINDI$_{CCA-KEM}$ [23] SABER [58] THREEBEARS[3][95]
Middle Product Lattices	Titanium-CPA [177]	-	Titanium-CCA [177]

[1] Adapted from GGH digital signature scheme;
[2] Adapted from GLP digital signature scheme;
[3] Based on the integer version of the MLWE problem;
[4] From the KCL [185] family;

3.1 Implementation of Arithmetic Modules

In this section, practical implementations of Gaussian sampler and polynomial multiplication on both hardware and software platforms are presented. There is only one hardware implementation of matrix multiplication (for standard lattices) available in the literature which we detail in Section 3.1.2.

3.1.1 Gaussian Sampler. Dwarakanath and Galbraith [77] provide a survey on different algorithms of efficiently computing the exponential function (Gaussian distribution curve) on resource constrained devices in terms of memory capacity. In order to decrease memory footprint, pipelining the sampling algorithm is offered. To be more specific, authors divide distribution curve into rectangles with the same probability and choose the rectangles according to the Knuth-Yao method which means the Knuth-Yao method is employed another round for rectangle itself. Tail probabilities in discrete distribution are relatively small which provide the chance of approximating them with lower precision arithmetic. Consequently, on the fly tail construction using standard double-precision floating-point precision is suggested. Although the offered idea could significantly reduce memory (lookup table) footprint since lookup tables just store non-tail probabilities; however, considerable floating point arithmetic overhead is imposed on the system.

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

Table 3. Comparison of contemporary lattice-based public key encryption and key exchange schemes.

Scheme	PQ security		Failure	Size (bytes)		
	SVP	CCA	Probablity[1]	Secret Key	Public Key	Ciphertext
BCNS (KEX) [39]	78	-	?	4096	4096	4224
JarJar (KEX) [12]	118	-	55	896	928	1024
NewHope (KEX) [12]	255	-	61	1792	1824	2048
Frodo rec. (KEX) [37] [2]	130	-	36	1280	11296	11288
Kyber light (KEX) [38]	102	169	169	832	736	832
Kyber rec. (KEX)[38] [2]	161	142	142	1248	1088	1184
Kyber paranoid (KEX) [38]	218	145	145	1664	1440	1536
NTRU KEM [107]	123	∞	∞	1422	1140	1281
NTRU Prime (KEM) [28]	129	∞	∞	1417	1232	1141
HILA5 (KEM/PKE) [166]	255	135	135	1792	1824	2012
trunc8 [167] [§]	131	-	45	128	1024	1024
NTRU ees743ep1 (PKE) [99]	159	-	112	1120	1027	980
Ding Key Exchange [62] [4]	AES-256 [*]	-	60	3072	2064	2176
EMBLEM (KEM)[169]	AES-128	-	140	2039180	2036736	78368
R.EMBLEM (KEM) [169]	AES-128	-	140	6144	4096	3104
FrodoKEM (Frodo-976) [137]		AES-192	199	31272	15632	15762
FrodoKEM (Frodo-640) [137]		AES-128	148	19872	9616	9736
KCL (e.g., AKCN-RLWE) (KEM) [185] [4]		AES-256	40	1664	1,696	2083
KINDI (e.g., KINDI-512-3-2-1) (PKE/KEM) [23] [4+]		AES-256	276	2752	2368	3392
LAC (e.g., LAC256) [123] [4+]		AES-256	115	2080	1056	2048
LIMA (e.g., CCA.LIMA-2p2048) [172] [4+]		SHA-512	314	18433	12289	7299
LIMA (e.g., CCA.LIMA-sp2062)[172] [4+]		SHA-512	244	24745	16497	9787
Lizard.CCA (PKE) [50] [4]		AES-256	381	557056	6553600	3328
RLizard.CCA (PKE) [50] [4]		AES-256	305	513	8192	8512
Lizard.KEM [50] [4]		AES-256	381	34880	4587520	35904
RLizard.KEM [50] [4]		AES-256	305	769	8192	8256
LOTUS (PKE/KEM) [145] [4+]		AES-256	256	1630720	1470976	1768
NewHope1024 (KEM) [148] [5]		AES-256	216	3680	1824	2208
NewHope512 (KEM) [148] [5]		AES-128	213	1888	928	1220
Streamline-NTRU-Prime (KEM) [29]		AES-256	∞	1600	1218	1047
NTRU-LPRime (KEM) [29]		AES-256	∞	1238	1047	1175
NTRU-HRSS-KEM [108]		AES-128	∞	1418	1138	1278
NTRUEncrypt (e.g., ntru-kem-743) [184] [4]		AES-256	112	1173	1023	1023
Odd-Manhattan (KEM) [146] [4]		AES-256	?	4456650	4454241	616704
uRound2 (uround2_kem_nd_l5) [87] [4]		AES-256	65	169	709	868
nRound2 (nround2_kem_nd_l5) [87] [4]		AES-256	45	165	691	818
uRound2 (uround2_pke_nd_l5) [87] [4]		AES-256	137	1039	830	953
nRound2 (nround2_pke_nd_l5) [87] [4]		AES-256	164	1039	830	1017
SABER (PKE/KEM) [58] [4]		AES-256	165	3040	1312	1472
THREEBEARS (KEM) [95] [4]		AES-256	188	40	1584	1697
Titanium-CPA (PKE) [177] [4]		AES-256	30	32	23552	8320
Titanium-CCA (KEM) [177] [4]		AES-256	85	26944	26912	8352
DH-3072	-	-	?	416	384	384
ECDH-256	-	-	?	32	32	64

[1] Failure is $-log_2$ of the failure probability;
[2] For recommended parameters;
[3] AES-128 and AES-192 are also available;
[4] Scheme with the highest security is selected;
[5] INP-CPA KEM is also available;
[§] Ring-LWE encryption and authentication system;
[+] INP-CPA PKE is available;
[*] The scheme is at least as hard as AES-256 as a requirement by NIST (same is applied to schemes with security of AES-128, AES-192 and SHA-512); CECS TR 17-04

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

Table 4. Comparison of popular lattice-based key digital signature schemes

Scheme	Security		Size		
	PreQ	PostQ	Secret Key	Public Key	Signature
GPV [88]	100	?	256 B	1.5 kB	1.186 kB
BG [20]	128	?	0.87 MB	1.54 MB	1.46 kB
TESLA–128 [11]	128	?	1.01 MB	1.33 MB	1.280 kB
TESLA–256 [11]	256	128	1.057 MB	2.2 MB	1.688 kB
GLP [91]	100	<80	256 B	1.5 kB	1.186 kB
BLISS-I [71] [2] BLISS-BI [70][1]	128	<66	256 B	896 B	700 B
TESLA#–I [24]	128	64	2.112 kB	3.328 kB	1.616 kB
TESLA#–II [24]	256	128	4.608 kB	7.168 kB	3.488 kB
Ring-TESLA-II [51]	118	64	1.92 kB	3.328 kB	1.568 kB
Dilithium rec. [73] [3]	138	125	3.504 kB	1.472 kB	2.701 kB
Dilithium high. [73][4]	176	160	3.856 KB	1.760 kB	3.366 kB
FALCON (falcon1024) [83]	AES256	128	8.193 kB	1.793 kB	1.233 kB
FALCON (falcon768) [83]	AES192	96	6.145 kB	1.441 kB	1.077 kB
FALCON (falcon512) [83]	AES128	64	4.097 kB	897 B	690B
pqNTRUsign [46] [5]	AES256	128	2.604 kB	2.065 kB	2.065 kB
qTESLA (qTesla_256) [32] [5]	AES256	128	8.256 kB	8.224 kB	6.176 kB
qTESLA (qTesla_192) [32] [5]	AES192	96	8.256 kB	8.224 kB	6.176 kB
qTESLA (qTesla_128) [32] [5]	AES128	64	2.112 kB	4.128 kB	3.104 kB
DSA-3072	128	0	416 B	384 B	384 B
ECDSA-256	128	0	32 B	32 B	64 B

[1] BLISS-BI speeds up BLISS-I by factor of 1.2×;
[2] Speed optimized;
[3] For recommended parameters;
[4] The highest security level;
[5] For both Gaussian-1024 and Unifrom-1024 variants;

Table 5. Popular implementation of lattice-based schemes.

Lattice Type	Schemes		
	Software	Hardware	Hardware/Software
Standard Lattices	PKE: [49, 50] [98] [169] [137] [145] [146] [87] DS: [20] [79][57] [11] KEX: [31] [37] [169] [137] [31] [109, 185] [50] [145] [146] [87]	PKE: [103] [169]	-
Ideal Lattices	PKE: [60] [119] [153] [158] [45] [183] [28] [49, 50] [167] [165, 166] [169] [148] [184] [87] DS: [91] [93] [79] [141] [149] [36] [153, 154] [9] [79] [71] [70] [51] [24] [164] [52] [83] [32] KEX: [63] [109] [12] [13] [14] [90] [107] [13] [39] [165, 166] [62] [169] [109, 185] [185] [172] [50] [148] [184] [108] [29] [87]	PKE: [89] [151] [161] [152] [169] [158] DS: [91] [149] [92] [105] KEX: [169] [180] [116]	DS: [19] [18]
Module Lattices	PKE: [16, 38] [58] [23] [185] DS: [72, 73] [46, 100] KEM: [16, 38] [95] [58] [23] [185]	-	-
Middle Product Lattices	PKE:[177] KEM:[177]	-	-

Software Implementation. Buchmann et al. [43] design and implement (C++) a flexible Gaussian sampler named Ziggurat which sets a trade-off between memory footprint, precision, and execution time of noise sampling from discrete Gaussian distribution. The area under the probability density function (PDF) is divided into rectangles with the same area which are employed to minimize calculation of expensive exponential function. More rectangles (more memory footprint) results in higher precision and better performance. The Ziggurat sampler is attractive because of its

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

potential flexibility that makes it a good candidate to use in cryptoengines of either high performance servers (allocate more memory to reach better performance and precision) or low speed resource constraint embedded devices (use few number of rectangles to minimize memory consumption). As a reminder, two points are chosen and crossed line from those points is used as an approximation of the PDF curve. In order to increase performance of Ziggurat sampler, more rectangles are required which imposes significant memory overhead since it needs to re-compute all the rectangles and save them in the memory. A better way to improve Ziggurat sampler is to increase number of approximation lines (by adding two extra points) which culminates in minimizing number of calculating the exponential function [136]. In other words, more approximation lines decreases probability of rejection by providing a more precise approximation of the curve. Consequently, performance and memory occupation are improved by employing more approximate lines.

Hardware Implementation. Roy et al. [162] implement the first high precision and low area hardware implementation of Knuth-Yao sampler with small standard deviation on a Xilinx Virtex 5 FPGA. Knuth-Yao involves a random walk tree traversal which imposes expensive sequential bit scanning and wide ROM footprint. To improve performance, authors traverse the discrete distribution generating (DDG) tree by employing relative distance of intermediate nodes. Column-wised, instead of row-wised, storing of the samples probability in ROM remarkably improves performance of the sampler. Numerous zeros in probability matrix is well compressed by applying one-step compression which culminates in near-optimal number of random bits to generate a sample point. Presented Knuth-Yao sampler suffers from vulnerability to timing and power attacks due to the non-constant time random walk which is solved by a random shuffle approach in order to eliminate leaking the timing information [160]. It should be mentioned that authors just offer the solution (random shuffle) and do not evaluate hardware implementation of the shuffler. In the new implementation, efficiency of the Knuth-Yao is enhanced by employing small LUTs with the input of the random bits and output of the sample point with high probability or an intermediate node positioned in the discrete distribution generation tree. Employing a lookup table with 8-bit input results in hitting a sample point (eliminate need of expensive bit-scanning procedure) with probability of 97% by eight random bits. Additionally, a more compact sampler is achieved by reducing the width of ROM and random bit generator.

Du and Bai [65] implement a highly precise (large tail bound) and area efficient cumulative distribution function (CDF) inversion variant of discrete Gaussian sampler on Xilinx Spartan-6 FPGA. Authors reduce area occupation by employing piecewise comparison (results in 90% saving of the random bits) and avoiding comparison of large numbers which is an improvement on [151]. Further improvement is achieved by employing small lookup tables with high hit rate which culminates in performance improvement. Performance of the proposed Gaussian sampler is improved twofold by the same authors with a software implementation (Intel Core i7-4771) [64]. The main challenge to improve performance on a general purpose processor is the large number of random bits that are consumed by the sampler to generate a random number. This obstacle is alleviated (around 30%) by employing multi-level fast lookup table (using 8 smaller lookup table instead of one). Further speed improvement of the sampler is gained by applying multithreading technique. The main security drawback of both hardware and software implementation of discrete Gaussian sampler by Du and Bai is its vulnerability to timing attacks due to the fact that traversal of binary tree is not accomplished in constant time [102].

Howe et al. [102] propose a comprehensive evaluation of various practical hardware implementation of time-independent discrete Gaussian samplers, including Bernoulli, discrete Ziggurat (First hardware design on FPGA), CDT, and Knuth-Yao. They present each sampler's weaknesses/strengths and perform comparison with state-of-the-are designs in terms of memory footprint, performance, and FPGA resource consumption is presented. Authors analyze

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

different quantum secure parameters for public key encryption scheme and digital signature. CDT Gaussian sampler provides higher throughput with lower when it is used for digital signature. Due to the disappointing performance of hardware Ziggurat implementation, authors prohibit use of Ziggurat sampler for digital signature schemes. Similarly, CDT sampler achieves better balanced area and throughput for public key encryption. However, allowing use of BRAMs, makes Knuth-Yao variant a much superior design in terms of area and throughput. Authors use BRAMs in order to decrease occupied slices in FPGA and to save precomputed values which significantly improves performance.

Pöppelmann and Güneysu [152] propose an area optimized hardware implementation of Bernoulli sampler that employs Bernoulli evaluation instead of evaluation of `exp()`. Rejection probability is high due to the absence of binary Gaussian distribution (easy to sample intermediate sampler) which results in increasing the entropy consumption and runtime. Although proposed Gaussian sampler is suitable for encryption schemes, it would be challenging to be employed inside digital signatures as the sampling component. Hardware implementation of Bernoulli sampler with binary Gaussian distribution is presented in [149] for `BLISS` scheme on Xilinx Spartan-6 FPGA.

Göttert et al. [89] propose the first hardware implementation of discrete Gaussian sampler. Authors use rejection sampling in their software implementation; however, because of the obligatory floating point arithmetic, authors prefer to employ lookup tables in their hardware implementation of implement Gaussian which the Gaussian distributed values in the stored array are indexed using a pseudo random bit generator. It should be mentioned that proposed sampler has unsatisfactory precision which has far distance from golden discrete Gaussian distribution due to the small tail bound. Authors employ a fully parallel architecture to design a polynomial multiplier which provides high throughput but makes the design extremely big which cannot be fitted into the largest Virtex-7 FPGA family.

Roy et al. [161] propose a compact Ring-LWE cryptoprocessor with the main goal of optimizing NTT multiplication which is accomplished by the reduction in fixed computation (4 NTT instead of 5 NTT) and in reducing the pre-scaling overhead. Besides, NTT Memory access is minimized by storing two coefficients in a single word, processing two pairs of coefficients together, and eliminating idle cycles. Authors avoid using ROM to save the twiddle factors and instead compute the twiddle factors on the demand. Besides, they reduce security by limiting coefficients of secret key to be binary, instead of Gaussian distributed, which gives the opportunity to replace multiply with addition operations. Small lookup tables are used in the Knuth-Yao discrete Gaussian sampler to avoid expensive bit scanning [162] to improve speedup; additionally, ROM widths are reduced which results in a more compact and faster sampler than Bernoulli [152].

3.1.2 Multiplier.

Software Implementation. Emeliyanenko [80] proposes an efficient 24-bit modular multiplication to achieve high throughput polynomial multiplications using NTT algorithm (on Nvidia GeForce GTX 280 GPU). Employing CUDA FFT kernel and Chinese Remainder Theorem (CRT), the proposed method provides better speedup compared to the GMP and NTL libraries for moderate coefficient bit-length.

Akleylek et al. [10] propose sparse polynomial multiplication for lattice-based schemes for the first time which improves the performance of digital signature proposed in [92] by 34%. Authors implement the Schönhage-Strassen polynomial multiplication on NVIDIA Geforce GT 555M GPU and compare its performance with existed multiplication schemes, including iterative NTT, parallel NTT and CUDA-based FFT (cuFFT) for different integer size.

Akleylek et al. [8] propose a software implementation (Intel Core i5-3210M processor) of the sparse polynomial multiplication using sliding window that bears around 80% of speed improvement compared to NTT [93]. Authors assume to have polynomials with coefficients with three possible values including -1,0 and +1. Multiplications by zero

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

are avoided and for +1 and -1 cases addition and subtraction are used, respectively. It should be mentioned that their idea can be used polynomials with arbitrary coefficients by performing substituting a multiplication with a loop of additions.Performance of the proposed method depends on high number zeros and numerous identical patterns which make the system prone to timing attacks.

`FNLlib` [5] is a scalable, efficient, and open source C++ library contains optimized arithmetic operations on the polynomials for the ideal lattice-based cryptography schemes. Compared to the generic libraries for polynomial arithmetic, several orders of magnitude improvement in the speed is achieved by employing algorithm optimizations, including fixed sized Chinese Remainder Theorem (CRT), scalar modular multiplication and NTT algorithm, and programming level optimizations, such as SSE and AVX2 SIMD. Authors use NFLlib for the RLWE encryption scheme and homomorphic encryption and compare their efficiency with classical cryptographic schemes (like RSA) and libraries (like NTL).

Longa and Naehrig [122] propose an efficient modular reduction with the main idea of limiting the coefficient length to 32 bits. Consequently, by employing the new technique in NTT, reduction is only required after multiplication. Combined with lazy reduction in NTT, speed improvement of factor 1.9 for C implementation (on 64 bit platform) and 1.25 for AVX2 vector implementation compared to NewHope (tolerant against timing attacks) is achieved. However, due to lack of 64-bit register, proposed reduction technique does not provide any speed up on 32-bit microcontrollers [14]. Additionally, authors use signed integer arithmetic which optimizes number of add operations in both sampling and polynomial multiplication.

Hardware Implementation. Howe et al. [103] propose the only hardware implementation of standard lattice-based encryption scheme based on LWE problem. Authors perform the multiply-accumulate (MAC) operations of matrices in the encryption scheme by utilizing a dedicated DSP48A1 unit of the Spartan-6 FPGA to achieve an area optimized hardware implementation of standard LWE based encryption engine.

Pöppelmann et al. [150] propose the first hardware optimization of polynomial multiplication (NTT) for the ideal lattice-based encryption schemes on a Xilinx Spartan-6 FPGA with the main goal of minimizing the area. Authors design a sequential NTT (one butterfly operator) that stores twiddle factors in a dedicated ROM which imposes memory overhead but achieves decent performance. Twiddle factors refer to different powers of the multiplicative operands. An optimized version of presented NTT polynomial multiplier is employed in [151] to design in a Ring-LWE encryption engine. With acceptable runtime, authors of [17, 161] present an optimized hardware implementation of NTT introduced in [150] to compute polynomial multiplication in a Ring-LWE on the smallest Spartan-3 Xilinx FPGA. The main idea is to compute twiddle factors on the demand instead of storing them in the ROM. By replacing the modulus with a Fermat number, modular reduction is improved and shift operation could be used instead to polynomial exponentiation. However, these proposed optimization could not take advantage of inherent parallelism in NTT. It should be mentioned that authors of [17] do not provide implementation of the whole cryptoengine.

Chen et al. [47] present a high performance polynomial multiplication for Ring-LWE encryption cryptosystems is implemented in hardware on a Spartan-6 FPGA by exploiting the parallel property of the NTT. Authors provide different secure set of parameters by which efficient Ring-LWE encryption and SHE could be achieved. They prove that polynomial multiplication can be done by computing the negative wrapped convolution by which there is no need to compute the modular reduction. Besides, size of FFT/inverse-FFT and point-wise multiplication is halved compared to the zero padding method. To be more specific, the proposed architecture for polynomial multiplication consists of two

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

butterflies and two point-wise modulo p multiplier (p has flexible length) which produce outputs (equivalent to two parallel NTT) that could be used to perform the inverse-FFT.

Du and Bai [67] propose a scalable and efficient polynomial multiplier architecture that take advantage of NTT's parallelism (implemented on Xilinx Spartan-6 FPGA) which provides a speed and area trade-off. In [150] and [17, 161] one and two butterfly operators are employed, respectively; however, [67] use b (power of 2) butterfly operators to improve speed of the polynomial multiplier which perform multiplication of two n-degree polynomials in $(1.5n + 1.5n\log n)/b$ cycles. To improve area (minimize required area), authors employ the cancellation lemma to minimized number of constant factors. Butterfly operation takes two coefficients (x,y) and one constant factor (ω) and makes two new coefficients ($[x+\omega y]$ mod p, $[x-\omega y]$ mod p). Butterfly can be used as a modulo p multiplier (by setting $x = 0$). Besides, in the first stage of inverse NTT the constant factor (ω) is 1, hence new coefficients are $[x-y] \bmod p$ and $[x+y] \bmod p$. Employing presented facts about butterfly operation, necessary clock cycles to calculate a sequential polynomial multiplication is $(1.5n + 1.5n\log n)$ which reduces $3.5n$ of required cycles.

Györfi et al. [94] perform a thorough evaluation of various candidates to implement modular FFT is perform on Xilinx Kintex-7 FPGA. Authors study three architectures in the diminished-one number system (computations over Z_{2k+1}) for different parameters in order to meet various factors such as run-time execution, throughput, occupation, and scalability. The first architecture yields the best performance, using pipelined modular butterfly-based FFT in which FFT core is throughput optimized. The other two architectures are serial distributed arithmetic-based and nested multiplication which occupy less area than butterfly-based FFT.

Du and Bai [68] demonstrate 30% savings in time and space (compared to [150]) on a Spartan-6 FPGA by performing on-the-fly performing bit-reversal step along with a new memory access scheme, (to load/store coefficients in calculating NTT). The idea is to load/store at address bit-reverse(i) instead of load/store at address i in memory, which means ith coefficient of NTT's output is located in the $bit - reverse(i)$th memory location. Consequently,the bit-reversal step in the inverse-NTT is eliminated. Authors employ two Block RAMs on FPGA which provide interleaving, hence two parallel NTT can be interleaved. Authors apply their optimization to the NTT of a RLWE base public key cryptosystem [66]. In addition, it is assumed uniformly random polynomial in the public key scheme to be fixed, hence precomputing the NTT of it offline improves the performance. In the above two articles, only one butterfly operator is used; however, by adapting bit-reversal and memory saving methods of above articles, authors propose a high-speed polynomial multiplication architecture with four butterfly operators that achieve on average 2.2 speedup improvement [69]. Two butterfly operators are used to calculate the ith level and other two perform $(i + 1)$th level calculations in pipeline by using results of the ith stage.

3.2 Software Implementations

3.2.1 Public Key Encryption. de Clercq et al. [60] propose an efficient software implementation of Ring-LWE based encryption for ARM Cortex-M4F micro-controller. The main goal was maximizing the speed (using assembly level optimization) and minimizing memory footprint (by storing two coefficients in one word). They employ the Knuth-Yao algorithm to achieve fast noise sampling, and use the platform's True Random Number Generator (`TRNG`) to generate random numbers. Authors employ optimization of the paper [161] including instruction-level parallelization. Additionally, polynomial multiplication is optimized by integrating multiple coefficients into one large word allowing load/store operations to be performed with a single instruction.

In [119], the authors use a byte-wise scanning method to improve the performance of the Gaussian sampler based on the Knuth-Yao algorithm. This allows them to implement a Ring-LWE based public key encryption scheme on a

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

resource-constrained 8-bit ATxmega128 AVR processor. By applying sophisticated memory alignments for storing coefficients, about 20 percent decrease in RAM usage is achieved. For NTT computation a couple of optimization techniques are employed including approximation based reduction and negative wrapped convolution.

By replacing the Gaussian noise distribution with a uniform binary error distribution, a high performance and lightweight public key encryption scheme implemented on small and 8-bit ATXmega128 and 32-bit Cortex-M0 micro-controllers [45]. By evaluating hardness of binary LWE against hybrid attack, security of scheme is determined. The main advantage of this scheme over Lindner-Peikert's proposal (`LP`) [118] is its smaller key and chipertext size. In terms of the speed, it is beaten by the scheme in [119] with slightly higher memory footprint. Similarly, proposed design in [153], uses NTT with precomputed twiddle factors and eliminating the bit reversal step which results in twofold performance improvement.

Yuan et al. [183] provide a portable JavaScript implementation of lattice-base cryptography schemes on PC web browsers, Tessel (an embedded system for IoT applications), and Android devices. To compute polynomial multiplication in Ring-LWE schemes NTT is used, while Karatsuba algorithms [110] is employed for NTRU schemes. In order to reduce the execution time, inverse transform sampling is employed in which possible values are precomputed and stored in a lookup table.

Reparaz et al. [158] implement a masked Ring-LWE scheme on a Virtex-II FPGA and 32-bit ARM Cortex-M4F which is Differential Power Analysis (DPA) resistant. In order to be resilient to first-order side-channel attacks, a constant time masked decoder with high success probability is implemented. Entire computation is done in the masked domain by employing a dedicated masked decoder which imposes considerable time and area overhead compared with unprotected design .

Cheon et al. [49] exploits learning with rounding (`LWR`) problem [22] and present `Lizard` and its ring variant (`Ring-Lizard`). Discrete Gaussian error distribution is replaced with an efficient rounding process with smaller modulus. Based on the results, Lizard beats NTRU and RSA encryption schemes by factors of 3 and 5, respectively. The main idea behind the Lizard is to eliminate the least significant bits of the ciphertext rather than integrating the message with some error. Cheon et al. [50] submitted `Lizard` (IND-CPA/CCA PKE and IND-CCA2 KEM) and its ring variant, `RLizard`, to the NIST PQC standardization call (NIST security category of 1, 3 and 5). Sparse and small secrets version of LWE and LWR (RLWE and RLWR) are the security basis of the Lizard (RLizard) IND-CPA PKE.

Besides Intel Xeon E5-2620 CPU, authors provide performance evaluation on smart phone (Samsung Galaxy S7) for their recommended parameter of Lizard.CPA (128-bit quantum security). Authors claim that Lizard is suitable for smart phones (memory usage of 20 megabytes). Authors provide datapath and finite state machine for hardware implementation of the Lizard PKE using Lizard.CPA and RLizard.CPA.

Chen et al. propose `NTRUEncrypt` [184] (at the NIST standardization call), a family of IND-CCA2 (resistant to subfield attacks) PKE and KEM schemes at 85, 159 and 198-bit post quantum security (NIST security category 1, 5 and 5). Based on the original NTRU scheme [99] and using parameters set of [98], `ntru-pke` and `ntru-kem` are achieved by applying NAEP transformation [106]. Using the same transformation, base on the provably secure NTRU encryption scheme [174] (based on RLWE problem), `ss-ntru-pke` and `ss-ntru-pke` are derived. Modulus is chosen to be a power of 2 (2^{11}) to enhance efficiency of the modulo arithmetic and integer multiplications. Authors adopt a PRNG from Salsa20 [27] to expand the seed. Box-Muller [40] is employed (only in ss-ntru-pke and ss-ntru-kem) to sample from the discrete Gaussian distribution.

The performance results are reported only for Intel i7-6600U processor (AVX2 optimization of NTT is not performed).

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

Bernstein et al. [29] introduce two ideal-lattice-based KEMs named Streamlined-NTRU-Prime and NTRU-LPRime with 248-bit and 225-bit security (NIST security category 5), respectively, with ciphertext and key size of around 1kB that are designed to reduce attacker's success probability by eliminating the ring homomorphisms. Schemes are IND-CCA2 where a a key can be used multiple times, hence large key generation latency is tolerable. Authors implement the reference code on a Intel Xeon E3-1275 v3. Streamlined-NTRU-Prime is faster than NTRU-LPRime in terms of the encapsulation and decapsulation time with slower key generation.

Hülsing ea al. propose NTRU-HRSS [108], a One-way CPA secure (OW-CPA) PKE and NTRU-HRSS-KEM, a CCA2-secure KEM derived from NTRU [99] with 123-bit post quantum security (NIST security category 1). In contrast to NTRUEncrypt [184] and standard NTRU [2], KEM is derived directly from NTRU-HRSS without using padding techniques (e.g., [106]). Contrary to Streamlined NTRUPrime [28] and standard NTRU, correctness of NTRU-HRSS does not rely on the fixed weight distinctions. NTRU-HRSS is designed based on the worrisome algebraic structure of cyclotomic rings (in contrast to Streamlined NTRUPrime). Additionally, NTRU-HRSS has probabilistic encryption, while Streamlined NTRUPrime has deterministic encryption. NTRU-HRSS use power of 2 modulus rather than the prime modulus (used in Streamlined NTRUPrime) which leads to faster arithmetic computation. NTRU-HRSS employs trinary secret key and messages and large modulus in order to avoid decryption failure (in contrast to LWE-based schemes with non-zero probability of failure) which leads to lower security and higher communication cost. Authors report the performance results of the reference and AVX2 implementation on Intel Core i7-4770K CPU.

Bansarkhani proposes KINDI [23] (at NIST PQC standardization call), a trapdoor-based encryption scheme based on LARA [155] in which data is concealed into the error without changing the target distribution. As a result, more data can be encrypted per ciphertext bit which reduces the message expansion factor (beneficial in "sign-then-encrypt"). $\mathtt{KINDI}_{\mathtt{CPA}}$, (Module-LWE based IND-CPA PKE) has been proposed with 5 different parameter sets ranging from 164 to 330-bit security (NIST security category of 2, 4 and 5). By applying a variant of Fujisaki-Okamoto (FO) transformation [101] on the $\mathtt{KINDI}_{\mathtt{CPA}}$, $\mathtt{KINDI}_{\mathtt{CCA-KEM}}$ with the same parameter space, can be built.

3.2.2 Key Exchange. Ding et al. [63] propose a provably secure Ring-LWE key exchange mechanism which is not passively secure since it produces biased keys. Peikert improves the protocol by using a new reconciliation method which generates unbiased keys [143]. A practical constant-time software implementation of the Peikert's Ring-LWE key exchange protocol is proposed in [39], namely BCNS, which can be added as the key exchange protocol to the transport layer security (TLS) protocol in OpenSSL along with RSA as the authentication and key SHA-256 as the hashing method. The most time consuming part of the protocol is the Gaussian sampler, which is done by employing a constant-time search on the Cumulative Distribution Table (CDT). For polynomial arithmetic, authors adapt the FFT from Nussbaumer's method [139], in cyclotomic rings whose degree is power of two which provides efficient modular reduction. BCNS employs a fixed polynomial as the system parameter which could be a potential weak link of the protocol. In addition, selection of large modulus results in lower efficiency and security level, 78-bit quantum security, than expected from a Ring-LWE scheme. In contrast to the digital signature and encryption schemes, key exchange scheme does not need a high quality Gaussian sampler [12], which BCNS uses; consequently, a simpler noise distribution is used in NewHope instead of Gaussian sampler. BCNS caches keys, which can be very dangerous to the security of the protocol because of the shared-key reused attacks [81], which is solved in the NewHope.

Alkim et al. [12] introduce NewHope, a portable C and highly optimized SIMD implementation (AVX2) of unauthenticated key exchange scheme, that solves the inefficiency (10 times better performance) and security drawbacks (increase quantum security level from 78-bit to 128-bit) of BCNS by optimizing the key exchange algorithm and better

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

parameter selection. A better analysis of failure probability which results in smaller modulus, on the fly generation of the polynomial system parameter, efficient polynomial arithmetic (combining Montgomery and Barret reduction and employing polynomial encoding), and using the centered binomial instead of the of discrete Gaussian distribution are the main improvements of NewHope over BCNS. NewHope has attracted attention of research and industry communities such that Google released the Chrome Canary which uses NewHope as the key exchange protocol along with elliptic curve Diffie–Hellman as the authentication protocol [41].

Alkim et al. [13] propose `NewHope-Simple`, a simpler variant of the Newhope with the same performance and security level. Simplicity is achieved by eliminating the error-reconciliation mechanism [63] with 6% message size overhead. Authors discard the least significant bits of each coefficient due to their negligible impact on the successful plaintext recovery. Additionally, authors encode a single key bit into 4 coefficients that results in reduction of the ciphertext length. In `NewHope-Simple`, polynomial a can be fixed, while the original `NewHope` generates a on the fly for every single run of the scheme. Software implementation of NewHope on ARM Cortex-M family, low power Cortex-M0 and high performance Cortex-M4, is presented in [14] which is the first key exchange scheme with quantum security level of 128-bit on constrained embedded devices. Authors optimize all hot regions of protocol in assembly, including error reconciliation, the uniform noise generation by ChaCha20 stream cipher [26], and NTT/NTT^{-1}. For NTT, authors set a memory-time trade-off for precomputing powers of constants (design parameters) by which only a subset of the powers of constants are precomputed and stored in the table. Gueron and Schlieker [90] further optimize the NewHope by optimizing the pseudorandom generation part which results in 1.5× better performance on the Intel Skylake processors. Authors improve the sampling step by lowering the rejection rate (from 25% to 6%) and exploit the parallelism in pseudorandom generation (replace SHAKE-128 with the parallelized SHA-256 or AES block cipher) and rejection sampling (employing AVX vector instructions). Longa and Naehrig [122], employ a reduction technique (during the NTT calculation) that eliminates the modular reduction after additions of two polynomials which results in the speed improvement of 1.9 and 1.25 for C and AVX implementations (compared to the reference NewHope [12]), respectively.

Adopted from the NewHope-Simple, Alkim et al. [148] propose NewHope as a family of KEMs, at NIST PQC standardization call. Submitted proposal includes NewHope512-CPA-KEM and NewHope512-CCA-KEM ($n = 1024, q = 12289$) which target 101-bit security (NIST security category level 1) and NewHope1024-CPA-KEM and NewHope1024-CCA-KEM with 233-bit security (NIST category security category 5) with comparable performance as the elliptic curve based cryptosystems. Four mentioned KEMs are derived from NewHope-CPA-PKE (which does not support arbitrary length messages, hence can not be used as a standalone encryption scheme) by applying a variant of Fujisaki-Okamoto transform [101]. In order to generate the random number and shared secret, hash function `SHAKE256` [78] is used as a pseudorandom function; generation of the shared polynomial a is done by expanding a 32-byte seed using `SHAKE128` [78]. Besides the reference and vectorized (using AVX instructions) implementations on the Intel Core i7-4770K (Haswell) processor, authors provide implementation and optimization of KEMs on a 64-bit MIPS architecture (MIPS64).

Ding et al. [62] propose (at NIST PQC standardization call) *Ding Key Exchange*, an ephemeral IND-CPA secure error reconciliation-based key exchange protocol from RLWE problem. At the same security level, Ding Key Exchange reduces communication cost (due to its rounding technique) compare to the similar schemes (NewHope, NewHope-Simple and Kyber). It provides equivalent security to AES-128, AES-192 and AES-256 (NIST security category 1,3 and 5) with flexible parameter choices and key size of n-bit where n can be 512 and 1024; as a result, it is more resistant to Grover algorithm compare to NewHope, NewHope-Simple and Kyber with the key size of 256-bit. NTL library [171] and CDT sampler are used for polynomial multiplication and sampling from the Gaussian distribution.

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

HILA5 [166], a side channel resistant Ring-LWE-based key exchange scheme (and also public key encryption) with the same security parameters ($n = 1024, q = 12289$) and sampler (binomial sampler ψ_{16}) as NewHope and has been tested on Intel Core i7-6700 CPU; also it has been integrated into OQS and OpenSSL. HILA5 uses SafeBits, improved version of the Peikert reconciliation mechanism [143], to reach slightly smaller messages than NewHope (36 bytes which is 0.9%) at the same security level by generating unbiased secret bits and hence less randomness in secret bits. HILA5 employs an efficient constant time error correction block to correct 5 bits of error which results in decryption failure of 2^{-128} compared to NewHope's failure rate of 2^{-64} (Frodo [37] and Kyber [38] with failure rate of $2^{-38.9}$ and $2^{-71.9}$) by sacrificing less than 4% of performance. Considering the higher reliability of HILA5, it can be employed as a PKE scheme. HILA5 [165] is submitted to the NIST PQC standardization call as a family of PKE and KEM schemes that provide equivalent security to AES-256 (NIST security category 5). Optimized polynomial multiplication and error sampling are performed by employing Cooley-Tukey [56] method and binomial distribution. Besides, SHAKE-256 is used to sample from uniform distribution.

Frodo (FrodoCCS) [37], the first practical implementation of public key exchange scheme based on standard lattices, original LWE problem [157], is secure against cache-timing attacks. Like BCNS, Frodo could be integrated into OpenSSL such that Google has announced that Frodo is used in 1% of Chrome web browsers. Matrix arithmetic compared to polynomial arithmetic imposes considerable overheads on the bandwidth (4.7 times more than NewHope), throughput (1.2 less throughput than NewHope), and performance (8 times slower than NewHope). Massive memory overhead is imposed if matrix variant should be saved in the memory. By generating and afterwards discarding the matrix variant (on-the-fly), memory overhead is alleviated. Besides, authors use an efficient Gaussian sampler, inversion sampling method, which employs precomputed tables. Based on the authors's claim, integrating Frodo (LWE-based post-quantum key exchange protocol) into TLS halves the server throughput. To tackle the ring related attack, NTRUPrime is proposed as a more secure scheme by using a combination Karatsuba, schoolbook, and Toom's multiplier [28]. Consequently, NTT-friendly prime and polynomial are not crucial which results in negligible drop in performance compared to NewHope [12].

FrodoKEM [137], is a family of IND-CCA secure KEMs based on the LWE problem with brute-fore security of at least AES-128 (FrodoKEM-640) and AES-192 (FrodoKEM-640). FrodoPKE is transformed by a variant of
tt FO transformation [101] to build the FrodoKEM. Authors generate public matrix A from a small seed using PRNG (AES128 or cSHAKE128) which results in more balanced ciphertext and key sizes, but remarkable computational overhead. Timing and cache attacks are prevented by prohibiting use of secret address accesses and branches. A portable C code reference and its optimized implementation (of generating the public matrix A and matrix arithmetic) are provided. Besides, authors report the results of the implementing on a 64-bit ARM Cortex-A72 (with the best performance achieved by using OpenSSL AES implementation, that benefits from the NEON engine) and an Intel Core i7-6700 (x64 implementation using AVX2 and AES-NI instructions). Employing modular arithmetic ($q \leq 2^{16}$) results in using efficient and easy to implement single-precision arithmetic. Sampling of the error term (16 bits per sample) is done by inversion sampling using small lookup table corresponds to the discrete cumulative density functions (CDT sampling).

Open Quantum Safe (OQS) [173] software platform is designed to evaluate proposed quantum-resistant schemes which has an open-source library (contains C implementation of BCNS, NewHope, and Frodo) of post-quantum cryptographic schemes. More importantly, OQS offers the chance to integrate the quantum resistant schemes into classical applications and protocols with the goal of minimizing the software change; besides, OQS provide opportunity to compare post-quantum schemes with each other or with classical cryptographic algorithms.

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

Jin and Zhao [109] present symmetric (OKCN) and asymmetric (AKCN) LWE and Ring-LWE based key exchange mechanisms (NewHope and Frodo are Ring-LWE and LWE-based symmetric key exchange schemes). OKCN, optimally-balanced key consensus with noise, can be used for key transport and encryption, while AKCN, asymmetric key consensus with noise, only can be employed for the key transport. In the proposed scheme, the server sets the session key before starting the key exchange mechanism. Consequently, it provides opportunity to encrypt the message offline which provides higher security and better workload balance. Compares with Frodo, OKCN-LWE produces a much smaller matrix by eliminating the least significant bits of each LWE sample which results in less computation for matrix arithmetic; smaller matrix also results in faster generation and sampling of the matrix. With the same set of parameters (same security level), OKCN-LWE consumes more bandwidth (30%) than Frodo, while its failure probability is remarkably lower. Employing the same optimization techniques, Ring-LWE based version of OKCN, which adopts the same noise distribution and parameters of NewHope, provides a more computationally efficient scheme than NewHope. It should be mentioned that authors integrate the OKCN-LWE scheme into the open safe project platform [173].

Zhao et al. [185] extend [109] and present a generic construction of authenticated key exchange, PKE and KEM schemes based on LWE/RLWE, LWR, and MLWE problems (submitted to NIST PQC standardization call as `KCL` (pka OKCN/AKCN/CNKE)). OKCN-LWE and OKCN-LWR key exchanges require less bandwidth (18% and 28%) compare to Frodo at the same security level (shared key size of 256-bit). The most efficient key exchange mechanism with share the key size of 512-bit is achieved by AKCN. Authors prove that the errors in different positions in the shared key are independent and propose single-error correction (SEC) code to correct at least one bit error; using the SEC, with the same security and error rate, OKCN/AKCN-RLWE based KEX schemes generate 765-bit shared key with less bandwidth than NewHope and NewHope-Simple (with 256-bit shared key). Authors claim that they provide the most efficient lattice-based key exchange scheme with share key size of 256-bit by applying OKCN/AKCN to MLWE-based key exchange mechanism. Additionally, authors provide a new authenticated key exchange scheme named concealed non-malleable key-exchange (CNKE).

Seo et al. [169] propose error-blocked multi-bit key encapsulation mechanism named EMBLEM and (R.EMBLEM) which is (secure against adaptive chosen ciphertext attack) based on the small secret LWE (RLWE) problem. During the decryption phase, the error does not affect the message by separating the message and error and concatenating each message block with the error-blocking bit. Secret key is sampled uniformly at random in $[-B, B]$ (where B is a positive integer smaller than σ) instead of Gaussian distribution. Consequently, key size is notably reduced since secret key can be generated by a 256-bit seed which eliminates the need for storing the whole matrix; however, it imposes computational overhead to generate the the secret key from the seed using pseudorandom functions. For polynomial multiplication in R.EMBLEM, Cooly-Tukey butterfly and Gentleman-Sande butterfly are used in NTT and inverse NTT, respectively. Besides the software implementation on Intel core-i7-7600, authors implement schemes on the Zynq 7 FPGA platform.

`Kyber` [38] is a highly optimized IND-CCA KEM with the post quantum security based on the hardness of solving the module learning with error problem (`Modulue-LWE`) [117]. Ideal lattices with their ring structure decrease public key and ciphertext size of standard lattices schemes by sacrificing security assumption. Module lattices proposed to fill the gap by believing that full ring structure is excessive [117]. Authors define IND-CPA PKE scheme under Module-LWE hardness assumption and apply a variant of Fujisaki-Okamoto transform [101] to build a IND-CCA KEM. Employing IND-CCA KEM, they design IND-CCA KEX and AKEX under hardness assumption in the classical and quantum random-oracle models. Kyber works over only one ring, $R_q = \mathbb{Z}_{7681}[x]/(x^{256} + 1)$, which provides flexibility (e.g. performing polynomial multiplication) to sacrifice security (from 128-bit to 102-bit) to improve performance and

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

communication size (33%) (by only changing k from 3 to 2). This flexibility is exclusive to Kyber (Module-LWE schemes); in Ring-LWE schemes, changing the security parameters results in building a new ring R_q and ring operations. Kyber has been submitted to the NIST PQC standardization call for as Kyber512, Kyber768, Kyber1024 at 102, 161 and 218-bit security (NIST security category 1, 3 and 5) [16].

Lu et al. [123] present `LAC` (**LA**ttice-based **C**ryptosystems) that includes an IND-CPA PKE (`LAC.CPA`), a passively secure KEX (`LAC.KE`), an IND-CCA KEM (`LAC.CCA`) and an AKEX (`LAC.AKE`) all of which are based on the RLWE problem. The main design concern is to enhance bandwidth efficiency (reduce key and ciphertext size) by setting modulus q to be small ($q = 251$) which prevents direct use of NTT in LAC. However, employing AVX2 vector instructions improves performance of the polynomial multiplication by a factor of 30, polynomial multiplication imposes remarkable computational pressure on the systems without the support of vector instructions. Sampling the secret and error term is done using the centered binomial distributions. LAC is proposed with three set of parameters that are much more expensive to break than AES128, AES192 and AES-256 (NIST security category of 1, 3 and 5).

Smart et al. [172] propose `LIMA` (**L**att**I**ce **MA**thematics), a family of IND-CCA and IND-CPA RLWE-based PKE (based on based on `LP` [118]) and KEM schemes, to the NIST PQC standardization call as 6 set of parameters with claimed post-quantum security from 143 to 274 (NIST security category of 1, 2, 3 and 5). Authors employ the Fujisaki-Okamoto [84] and Dent transform [61] to obtain IND-CCA PKE and IND-CCA KEM schemes. In addition to power-of-two cyclotomic rings (`LIMA-2p`), authors propose safe-prime cyclotomics (`LIMA-sp`) that reduces probability of subfield attacks by scarifying the efficiency compare to the power-of-two cyclotomics. In order to avoid decryption failure, authors perform rejection sampling (from a centered binomial distribution) at the encryption stage which makes the implementation to be non-constant time. In order to perform polynomial multiplication with FFT, large modulus should be selected for `LIMA-sp`.

Phong et al. [145] present `LOTUS` (**L**earning with err**O**rs based encryption with chosen ciphertex**T** for po**S**t quantum era) IND-CCA2 secure LWE-based PKE (`LOTUS-PKE`) and KEM (`LOTUS-KEM`) with 128-bit, 192-bit and 256-bit security (NIST security category of 1, 3 and 5). Knuth-Yao algorithm [115] is employed to sample the error term from discrete Gaussian distribution. In order to reduce sampling's overhead, DDG tree is built online and probability matrix is stored column-wised. Besides the reference and optimized implementations, vectorized implementation (employing AVX2 vector instructions) of LOTUS-PKE and LOTUS-KEM are provided.

`NTRU-KEM` [107], an IND-CCA2-secure KEM based on NTRU cryptosystem with 128-bit classical security, is the first timing attack resistant NTRU software thanks to its constant-time noise sampler. NTRU-based KEM has active security which allows parties to cache the ephemeral keys, however passive secure key exchange mechanisms like NewHope and Kyber must not use cached values. Compared to NewHope (255-bit PQ security), NTRU-KEM (123-bit PQ security) improves (secret key size, public key size,cipher text size) by (20%, 37%, 37%) and halves the required clock cycles for encryption/encapsulation step. However, it increases required clock cycles for key generation and decryption/encapsulation by a factor of 3.47.

Plantard [146] presents `Odd-Manhattan`, a IND-CCA KEM by using Dent transform [61] on IND-CPA PKE, at 126,192 and 256-bit security (NIST security category 1,3 and 5). `Odd-Manhattan` is based on the *α-Bounded Distance Parity Check (BDPCα)* [126], which impose considerable increase in time and size of key generation, encryption and decryption. To alleviate the time overhead, computational reuse (store the results of the k consecutive additions (constant time) in memory) with notable memory penalty has been employed.

Garcia-Morchon et al. [87] introduce `Round2`, a family of CCA-PKE (`Round2.PKE`) and CPA-KEM (`Round2.KEM`) based on the general learning with rounding (GLWR) problem. By having d (dimension), n (system parameter), q

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

(large modulus), p (rounding modulus) $\in \mathbb{Z}^+$ where $q \leq p$ and $n \in 1, d$, if $n = 1$, instantiated scheme is based on the LWR problem, while $n = d$ ($n + 1$ is prime) results in a RLWR based scheme. Round2 provides two set of parameters including Unified-Round2 (`uRound2`, q is a power of 2) and NTT-Round2 (`nRound2`, q is prime, $n = d$ ($n + 1$ is prime)). With uRound2, schemes can be seamlessly instantiated from LWR or RLWR [22] (for all NIST security levels), both $n = 1$ and $n = d$, with the same code which provides agility (i.e., switching from RLWR-based schemes to LWR-based schemes without recompilation). GLWR ,compare to LWE, results in decreasing random data generation due to avoiding sampling from the non-uniform noise distribution; besides, required bandwidth is reduced since fewer bits are needed per coefficient. Secret terms can be either a sparse-trinary (reduces probability of error in decryption) or uniformly sampled in $\mathbb{Z}_q^d$. In order to have a unique implementing for LWR and RLWR, a common multiplier that implements polynomial multiplication as the matrix multiplication is employed. Compare to NewHope [12] and Kyber [38], RLWR-based uRound2 requires smaller the public-key and ciphertext in total for NIST security category 5. Over the same ring as NTRU-KEM scheme, Round2 bears better speedup due to its faster key generation. Performance evaluation of the reference implementation is performed on Intel Core i7 2.6GHz.

D'Anvers et al. [58] propose `SABER`, a family of Module-LWR based IND-CPA PKE and IND-CCA KEM schemes including `LightSaber-KEM`, `Saber-KEM` and `FireSaber-KEM` with 115, 180 and 245-bit security (NIST security category 1, 3 and 5). Integers are chosen to be the power-of-two modulus which results in avoiding explicit modular reduction and relaxing complicated sampling methods (e.g., rejection) by efficiently, constant time, sampling from a (modulo power 2) uniform distribution; however, power-of-two modulus prevents using the NTT for polynomial multiplication (Karatsuba and Toom-Cook algorithms are used instead). Switching among SABER schemes is accomplished by choosing a modulus of higher rank in the fixed polynomial ring $\mathbb{Z}_{2^{13}}[x]/(x^{256} + 1)$. Failure rate is reduced by using reconciliation method introduced in [13].

Hamburg [95] introduces *ThreeBears*, a family of IND-CPA and IND-CCA KEM schemes adopted from Kyber [38] and based on integer Module-LWE (`ILWE`) [53] problem. *ThreeBears* includes *BabyBear*, *MamaBear* (recommended)and *PapaBear* with NIST security security category of 2, 4 and 5, respectively. For each scheme, deterministic CCA-secure (with `FO` transform) and ephemeral (without `FO` transform) implementations are presented. In order to reduce the memory footprint, private key is rapidly generated by expanding a seed; similarly, public key and large public matrix (uniformly at random sample each element) are generated modulus N where N is a large Mersenne prime. Sampling the noise is performed by expanding a seed to only 1B per digit. Although the Saarinen's error correction [166, 167] can notably improve *ThreeBears* security, author prefers to use Melas BCH code as the two-error-correcting code to maintain the code simplicity. To preserve simplicity, NTT is not used which results in slower integer arithmetic, on devices without vector unit support in particular. Performance analysis of the *ThreeBears* implementations are provided on Intel Skylake, ARM Cortex-A8 and ARM Cortex-A53. With 15% smaller ciphertext and public key size, *MamaBear* (resp. *PapaBear*) is stronger than *Kyber-Paranoid* (resp. Hila5 [166, 167] and NewHope [12]).

Steinfeld et al. [177] present `Titanium`, a family of IND-CPA PKE (`Titanium-CPA`) and IND-CCA KEM (`Titanium-CCA`) schemes based on the middle product LWE (`MPLWE`) problem [159]. Schemes are tightly and provably secure based on the hardness of Polynomial-LWE problem over polynomial ring $\mathbb{Z}[x]/f(x)$ where f is the member of a large group of ring polynomials. Titanium is a middle ground scheme that achieves a trade-off between security and efficiency such that, in terms of ciphertext size and performance, it is superior to *Frodo* [37] but inferior to *Kyber* [38]. Among 6 suggested parameter sets, $Std128$, $Med160$, $Hi192$ and $Super256$ satisfy minimum security specified by NIST (1,1,3 and 5, respectively). However, security analysis of Titanium assumes the classical random oracle model. NTT and binomial difference error distribution are used for polynomial multiplication and error sampling; secret key coordinates are

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

sampled uniformly at random over $\mathbb{Z}_q$. It should be mentioned that error correction or reconciliation techniques are not employed in Titanium. In addition to the reference and optimized implementation, authors provide performance analysis of vectorized implementation (AVX2) on Intel i7-7700K.

3.2.3 Digital Signature. Lyubashevsky [124] presents Short Integer Solution (SIS) problem based digital signature schemes adopted from Fiat-Shamir transformation. Lyubashevsky improves this scheme by establishing it in an efficient manner for Ring-SIS and Ring-LWE which culminates in smaller signature and key size [125]. BLISS signature [71] is an optimized version of Lyubashevsky's signature where a binomial Gaussian sampler is used in the rejection sampler component, resulting in a remarkable reduction in the standard deviation of the Gaussian distribution. Bai and Galbraith [20] propose a provably secure small signature (BG signature) based on the standard LWE and SIS problems that can be implemented using uniform distributions. Standard worst-case computational assumptions on lattices is the basis of security for the BG signature scheme.

Pöppelmann et al. [153, 154] evaluate implementations of various lattice based schemes, including RLWE-based public key encryption schemes and BLISS [71] on the 8-bit AVR micro-controllers. They review various NTT algorithms (written in C) and optimize (using assembly language and ignoring zero coefficients) polynomial multiplication (column-wise) for ideal lattice-based cryptography schemes. However, using precomputed twiddle factors in NTT computations requires more memory footprint. Official release code of Keccak [30] for AVR is used for the random oracle which is needed for signing and verification. Other optimization offered by the authors are removing bit reversal step during polynomial multiplication and applying the Gaussian sampling in a lazy manner. Compared to RLWE encryption, BLISS needs larger standard deviation for sampling from Gaussian distribution; candidates for Gaussian sampling are CDT sampling [35] with binary search , which results in large tables, and Bernoulli [36], that impose remarkable performance overhead. Consequently, for Gaussian sampler, KL-convolution sampler [149] is used which consume less flash memory compared with the CDT and Bernoulli samplers. The BLISS implementation consumes the least flash footprint on AVR and has the lowest published runtime through 2015.

BLISS-B [70] (with the same security level) improves performance of original BLISS 2.8 times by employing the ternary representation of polynomials in order to shorten length of the random numbers. During the key generation, keys will not be rejected which leads to 5-10 times enhancement of key generation step runtime. Generated signatures by BLISS and BLISS-B are compatible with each other, allowing signatures generated by one to be valid for the other. Although generated keys of BLISS-B could not be used in BLISS, BLISS generated keys are compatible with BLISS-B.

Oder et al. [141] present an efficient software implementation of BLISS on ARM Cortex-M4F to optimize the throughput along with minimizing memory footprint. Authors evaluate efficiency of variety of Gaussian samplers including the Bernoulli, Knuth-Yao, and Ziggurat [43]. In order to improve NTT computation, assembly level optimization along with precomputed coefficients are employed. They conclude that Knuth-Yao sampler is the best candidate for large devices, while for constrained devices Bernoulli is more favorable.

Güneysu et al. [93] present a highly optimized SIMD implementation of GLP signature [91] and implement it on Intel's Sandy and Ivy Bridge processors. In the proposed scheme, Gaussian sampler is replaced with the uniform sampling from {-1,0,+1}; To exploit SIMD support of Advanced Vector Extensions (AVX), each 512 double-precision floating-point array of coefficients is 32-byte aligned. besides, modular reduction of coefficient is done in a lazy manner. However, signature size and security level of implemented scheme is inferior to BLISS.

El Bansarkhani and Buchmann [79] implement the first software implementation (space and speed optimized) of GPV signature scheme [88] by employing the Micciancio and Peikert (MP) trapdoors [130] on the Sun XFire 4400 server

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

equipped with 16 Quad-Core AMD Opteron. Besides the matrix version, a much faster Ring-LWE based variant of the scheme is provided which has around 3-6 and 3-9 times better speed than matrix version for sign and verification steps, respectively. Due to the small number of stored entries, instead of rejection sampling, the inversion transform method is used for discrete Gaussian sampling during integer key generation; however, rejection sampling [88] is used in the randomize rounding. It worth mentioning that for random oracle SHA256 and a pseudo random number generator from [44] are used.

Dagdelen et al. [57] propose a high-speed software implementation of the BG signature on an Intel processor with AVX and an ARMv7 with Neon vector instructions support. Compare to the BG signature, authors proposed an optimized rejection sampling. Small performance degradation is observed by employing standard lattices instead of ideal lattices which is a great achievement because there is no quasi-logarithmic arithmetic scheme like NTT for standard lattices.

Boorghany and Jalili [35, 36] propose efficient software implantation of lattice-based GLP and BLISS authentication protocols for resource constrained smart cards and micro-controllers (ARM and AVR). Authors perform a design space exploration by choosing different parameter sets for FFT and Gaussian Sampler (Knuth-Yao and Bernoulli) along with the various PKE schemes. They conclude that lattice-based schemes are efficient enough to be implemented on the constrained devices.

Alkim et al. [11] introduce `TESLA` (a tightly secure signature in random oracle model resulted) by a tight reduction to LWE-based problems on the standard lattices and implement it on Intel Core-i7 4770K (Haswell). Their proposed design is adopted from BG signature [20] which is faster and smaller than the same scheme in [57] due to employing parallel matrix vector multiplication and lazy reduction. Authors propose two variants TESLA-128 and TESLA-256; the former one, TELSA-I, is not quantum resistant, while the latter, TELSA-II, provides the first lattice based digital signature with 128 bit security against quantum computers. Large public key size (about 1 MB) makes it impractical to implement. A fast, small, and provably secure Ring-LWE based software implementation of TESLA is presented in [9] by the same authors on the same platform which reduce the key size about 3 order of magnitude. To generate signatures, uniform sampling is employed instead of Gaussian sampler which BLISS uses. The propose Ring-TESLA benefits from the AVX2 instructions, which has a one cycle throughput for eight doubles integers. To instantiate from the Ring-TESLA there is a problem in parameter selection, which leads to rejection of valid signatures in verification stage. The problem is solved in [51] by new parameter selection method and in [24] by altering the algorithm. Based on the claims in [24], TESLA and Ring-TESLA are using global parameters which results in employing a fixed lattice for all the signatures that could weaken the signature scheme. A recent version of TESLA [11] fixes the problem by adding a new condition to the signing step which results in dropping the speedup, creating a more complex signature scheme, with less success in signing. Barreto et al. [24] introduce TESLA#, a high performance version of Ring-TESLA, on Intel Core i7-4770 Haswell processor, which resolves the security problems of TESLA. Further improvement is achieved by designing a more efficient Gaussian which accelerates the key generation step along with avoiding to store all 32 bits of coefficients of polynomial. Bindel et al. [32] propose qTESLA a familiy of (provably existentially unforgeability under chosen-message attack (`EUF-CMA`) secure in the quantum random oracle model) Ring-LWE based digital signature schemes, including qTESLA-128, qTESLA-256 and qTESLA-192 with NIST's security categories of 1, 3 and 5. qTESLA adopts a simpler version [24] of the bimodal Gaussian sampler [71] that is only employed in key generation. Although qTESLA performs polynomial multiplication by NTT, it is compatible with schoolbook algorithm. qTESLA uses cSHAKE [111] to deterministically generate the random bits for driving the seeds in the key generation and generation of a fresh polynomial for every key pairs. In the signing step, qTESLA employs SHA-3 as the hash function and cSHAKE as the pseudo-random function. Contrary to the Ring-TESLA, qTESLA is secure against cache

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

side channel attacks by applying countermeasures introduced in [34]; however, qTESLA is vulnerable to fault attacks [33] similar to Ring-TESLA.

BLZZRD [164], a side-channel resistant lattice-based signature based on BLISS-B [70], is implemented on an Intel Core-i7 Haswell processor with small signature size but with costly signing step. Authors achieve optimal compression for discrete Gaussian distribution by using Binary Arithmetic Coding (BAC) which leads to a more compact signature compared to advanced Huffman-based signature compressors. Further security improvement is gained by prohibiting leak of information about the execution time and power consumption of the arithmetic operation by applying randomization which makes the signature resistant to timing and power attacks. Additionally, masking properties of Gaussian samples is achieved by randomizing and combining multiple number the of sample vectors.

Dilithium [73], a simple and efficient digital signature scheme resistant to the lattice reduction attacks (with the same conservative security parameter in [12]) is adapted from the designs in [91] and [20] that uses Fiat-Shamir Abort Framework [124] which is secure in random oracle model (no security proof in quantum random oracle model is presented). Authors implement Dilithium and its variant Dilithium-G on Intel Core-i7 4770k with comparable efficiency to BLISS. In [91], *hints* are generated (by the signer to help verifier to verify the signature) to make the signature smaller; Dilithium improves hint-generation and halves the public key size with less than 5% increase in the signature size. In fact, authors set *total size = signature size + public key size* as their size parameter. Dilithium over ring of $Z_q[x]/[x^n + 1]$ ($n = 256, q = 2^{23} - 2^{13} + 1 = 8380417$) has slightly bigger *total size* than BLISS (over ring of $Z_q[x]/[x^{1024} + 1]$) with the same security level. Dilithium samples polynomial noises from the uniform distribution S_η in $[-\eta, +\eta]$ where η is in the range of 3 to 7 for very high secure to weakly secure scheme, respectively. However, Dilithium-G extract noises from Gaussian sampler which results in better security but vulnerable to timing attacks. Rejection sampling is the same for both schemes as if individual coefficients of a signature is not within a certain range, signing procedure must be restarted. Dilithium employs the standard NTT-based polynomial multiplication, however in vectorized version, Dilithium uses integer instructions instead of floating point vector instructions [12].

Dilithium has been submitted to the NIST PQC standardization call at three security levels (all use the same ring) including *medium*, *recommended* and *very high* (NIST security category 1,2 and 3) [72]. Dilithium is tightly secure in the quantum random oracle model based on the *Module-LWE*, *Module-SIS* and *SelfTargetMSIS* [113] problem (adopted from combined security of *MSIS* problem and hash function H). Signature size of Dilithium is about 2× bigger than that of BLISS [71] and [74] (smallest schemes among lattice-based digital signatures) that use discrete Gaussian sampler which Dilithium avoids. However, compare to the most efficient lattice based digital signature schemes that avoids Gaussian sampler, Dilithium achieves 2.5× smaller public key size. Coole-Tukey and Gentleman-Sande butterflies are used in NTT and inverse NTT, respectively, to perform the polynomial multiplication in which Montgomery reduction is used after multiplication (avoid reduction after addition and subtraction). Vectorized (AVX2 instruction set on Intel Core i7-4770K) version of NTT gives 4.5× speed improvement over the (reference) integer NTT implementation which is 2× faster than floating point NTT [12]. Dilithium uses SHAKE128 and SHAKE256 to drive the matrix ($A \in R_q^{k \times l}$ in NTT domain) and vectors. Vectorization improves speedup of the matrix and vector expansion by sampling four coefficients in parallel.

Fouque et al. [83] propose FALCON, a familiy of compact lattice-based hash-and-sign digital signature schemes with quantum security of 103, 172 and 230 bits (NIST security category 1,3 and 5) with the main goal of minimizing the *total size = signature size + public key size*. FALCON is the result of combining GPV framework [88], NTRU lattices [99] and Fast Fourier sampling [76]; provably secure NTRUSign is built by combining the GVP framework and NTRU lattices [175]. instantiation of the GPV IBE over NTRU lattices is presented in [74] which can be transformed to FALCON by

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

employing the Fast Fourier sampling in private key operations. NTRU lattices along with the capability of the message recovery (completely from the signature) result in compactness of the FALCON. Verification step in FALCON is fairly fast and simple and can be performed by a hash function followed by NTT operations. FALCON uses double precision floating-point arithmetic in signing which can be challenging to implement on the devices without floating point units. Another downside of the FALCON is the extensive use of the discrete Gaussian sampling over the integers which is hard to protect against the timing and side-channel attacks. FFT over the complex numbers is used for private key operations, while public key operations and key generation are perform using NTT over $\mathbb{Z}_q$. FALCON uses bimodal Gaussian in the reject in sampler, ChaCha20 as the PRNG and SHAKE-256 as XOF for all security levels. FALCON can be easily transformed into an IBE scheme [74]. Authors only provide performance evaluation of the reference implementation on an Intel Core i7-6567U CPU with 15% of error margin due to not disabling the boosting feature of the processor. Falcon has the smallest *total size* among all the post quantum digital signature schemes at the same security level.

Chen et al. [46] present `pqNTRUSign`, a modular lattice based hash-then-sign digital signature scheme (introduced in [97]) at post quantum security of 149-bit (NIST security category 5). pqNTRUSign provides sampler agility as the user can choose the sampler based on the design goal; constant time uniform sampling for the security and (bimodal) Gaussian sampling for the performance goals, respectively. The key generation is the same for both set of parameters (Gaussian-1024 and Uniform-1024); other steps have different implementations for various parameter sets. The public key, forgery and transcript security are provided by NTRU assumption, LWE problem over NTRU lattices and rejection sampler, respectively. Like `NTRUEncrypt` [184], Box-Muller [40] is employed to sample from the discrete Gaussian distribution. AVX optimization for polynomial multiplication is not included in pqNTRUSign; a naive NTT with the time comlexity of $O(N^2/2)$ is used.

3.3 Hardware Implementations

3.3.1 Public Key Encryption. Göttert et al. [89] propose the first hardware implementation of Ring-LWE based public key encryption on the Xilinx Virtex-7 FPGA. Due to the large area occupation of the full Ring-LWE public key encryptoin scheme, only LWE-polynomial variants are chosen to be implemented. Proposed implementations are based on LP lattice-based encryption scheme [118] which achieve 316 times higher throughput compared to software implementation. Using a fully parallel architecture (which makes their design extremely big) high throughput is achieved by minimizing required clock cycles in computing the NTT. The main optimization metric is performance for which they show speedups for encryption and decryption schemes by factors of 200 and 70, respectively, in comparison to the software implementation with the same level of security.

Pöppelmann and Güneysu [151] provide a flexible Ring-LWE encryption engine in which one core is used to perform key generation, encryption, and decryption steps with the primary goal of optimizing throughput per unit of area. In addition, by applying optimizations including different encoding technique and removing some LSBs of the ciphertext coefficnents, enryption engine, which is 60 times smaller than the design in [89], it could be fit on the Xilinx Spartan-6 FPGA with 3 times slower encryption step. They employ a Gaussian sampler with relatively low precision, using the CDT sampling method that compares random probabilities with a cumulative distribution table. The proposed Gaussian Sampler is fast (one sampler per cycle at 60 MHz) with the cost of numerous random bits (85) to produce a single random number. The Gaussian sampler is time independent with the cost of an array of parallel comparators, one per each word of the table.

Roy et al. [161] implement a compact Ring-LWE crypto processor on Virtex 6 FPGA where they optimize NTT multiplication by reducing the fixed computation and pre-scaling overhead. Authors suggest to combine pre-computation

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

stage and NTT computation. Besides, NTT Memory access is minimized by storing two coefficients in a single word, processing two pairs of coefficients together, and eliminating idle cycles. Small lookup tables are used in the Knuth-Yao discrete Gaussian sampler [162] which leads to more compact and faster sampler than [152].

Pöppelmann and Güneysu [152] implement the smallest lattice-based encryption engine on Spartan-6 and Virtex-5. Compared with the high speed implementation of [151], this is one order of magnitude slower due to non-applicability of using NTT (using DSP-enabled schoolbook polynomial multiplier). Besides, considerable area is saved by using special modulus, power of 2, by which modular reduction is almost cost free. Further area saving is achieved by using a Bernoulli distribution [71] with small precomputed tables in order to optimized simple rejection sampling by eliminating computing of $exp()$ function.

Howe et al. [103] present the first, and the only, hardware implementation of lattice-based encryption engine based on learning with error problem over standard lattices on the lightweight Spartan-6 FPGA. The main concern is optimizing the area, while the scheme maintains balance between area and performance by using a larger Gaussian sampler. The proposed encryption engine is smaller in comparison to the design of [89]; besides, it can closely compete with the encryption scheme of [151]. To maximize the performance, authors use a larger Gaussian sampler, Bernoulli sampler, which generates samples in parallel with no adverse effect on the critical path.

Reparaz et al. [158] implement a masked Ring-LWE scheme on a 32-bit ARM Cortex-M4F and Virtex-II FPGA which is Differential Power Analysis (DPA) resistant. In order to be first-order side-channel attack resilient, a constant time masked decoder with high success probability is implemented. The entire computation is done in the masked domain by employing a dedicated masked decoder which imposes considerable time and area overhead compared with an unprotected design.

3.3.2 Digital Signature. Howe et al. [104] provide evaluation and summary of practical instantiations of digital signature schemes based on lattice problems on different platforms. Evaluation metrics are secret key, public key and signature size. Additionally, they give a survey on various implementations of basic blocks including NTT and sampling. Authors evaluate Bernoulli, Ziggurat, Knuth-Yao, and cumulative distribution table (CDT) variants of Gaussian sampler.

Güneysu et al. [91] implement an efficient lattice-based signature scheme (GLP signature) on a Xilinx Virtex 6 FPGA which is the first practical lattice-based signature scheme that could resist transcript collision attacks. Author removes the need for Gaussian noise sampling by using the rejection sampling which leads to hide the secret key contained in each signature. Security of the proposed scheme is lower than standard lattice-based signatures due to building the hardness assumption based on the Decisional Compact Knapsack problem. Because of the regular structure of the schoolbook algorithm, authors achieve high speed and small size for implementation of the polynomial multiplier. Compared with BLISS, the proposed signature is sub-optimal in terms of signature size and security level.

Pöppelmann et al. [149] implement a high throughput hardware implementation of BLISS [71] on Xilinx Spartan-6 FPGA. Authors improve and parallelize the column-wise schoolbook multiplier presented in [91]. Authors employ an efficient CDT based Gaussian sampler with large tables. To improve performance of the CDT sampler, author deploy an decrease number of comparisons by improving the binary search and reducing size of precomputed large tables by using an optimized floating-point representation (adaptive mantissa size) with negligible effect on the performance. Authors provides enhanced CDT which uses two smaller samples (Peikert convolution theorem [142]). A standard CDT needs table of the size at least $\eta \times \tau \times \lambda = 215.73 \times 13.4 \times 128 = 370kb$ while enhanced CDT needs around 23× smaller table. Authors evaluate performance and resource consumption of BLISS-I ($n = 512, q = 12289$) by employing the CDT and two parallel Bernoulli samplers and conclude that CDT consumes less FPGA resources than Bernoulli; besides,

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

enhanced CDT achieves 17.4 Million Operation per Seconds (MOPS) which is 2.3× more than that of Bernoulli sampler. Based on the results, performance of enhanced CDT is almost the same for BLISS-I ($\sigma = 215$), BLISS-III ($\sigma = 250$) and BLISS-IV ($\sigma = 271$).

Güneysu et al. [92] propose an optimized and flexible implementation of a lattice-based digital signature on Xilinx Spartan-6 and Virtex-6 FPGAs which has the same theoretical basis as [91] but with major improvements. Compared with [91], instead of Schoolbook multiplier, authors employ a parallelized NTT for polynomial multiplications (the most time consuming part of the digital signature), which leads to smaller and faster signing/verification engines. Authors develop a flexible processing core with VHDL that could be configured as either signing or/and verification engine which does not impose any overhead to provide flexibility. The signing step is broken into three separate blocks, including lattice processing engine, random oracle, and, sparse multiplication along with compression unit, which are running in parallel. The digital signature processor is adapted from the lattice processor for the public key encryption scheme in [151].

3.3.3 Key Exchange. Oder et al. [180] propose an area optimized constant time implementation of NewHope-Simple [13], on Xilinx Artix-7 FPGA, with decent performance level. With the same post-quantum security level as NewHope (128-bit), server and client work with clock frequency of 125 and 117 MHz, respectively. Authors design two separate modules for the client and server sides which forces an embedded system to be only either a server or a client, hence results in the lack of re-usability as a huge disadvantage. For the sake of the area optimization, 512 butterfly operations are performed serially, while they can be performed in parallel.

Kou et al. [116] provide a high performance pipelined implementation of NewHope [12] on Xilinx Artix-7 FPGA which is 19.1× (4 ×) faster (bigger) than hardware implementation of the NewHope-Simple [180]. In order to improve the performance, NTT operations are computed using four butterfly units; besides, Longa-Naehrig modular reduction [121] is used instead of Barrett reduction.

3.4 Hardware/Software Codesign

Because of the probabilistic inherent feature of rejection sampling in the lattice-based schemes, the probability of generation of an invalid signature exists. Leveraging precompution, failure possibility of signature generation could be minimized. Aysu et al. [18] divide the signature scheme in hash-based cryptographic signatures into two separate phases in order to minimize the energy and latency of signature generation. During the offline phase, input (message) independent computations, for instance key and random number generation, are performed and results are stored in a memory buffer as coupons. Subsequently, output is generated using the precomputed coupons and the input (message) during the online phase. Employing the same idea, Aysu et al. [19] implement a latency optimized lattice based signature with hardware/software co-design technique. The main objective is to optimize latency which is achieved by focusing on the signature generation step that is done on the embedded device. On the other hand, verification step is performed on high performance platform servers. Signature generation scheme is divided into two separate phases including offline and online phases which are performed on NIOS soft-core as software, and Altera Cyclone-IV FPGA as hardware, respectively. Hardware is responsible for low latency hash function and polynomial multiplication; however, the software part computes and stores polynomials.

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

3.5 DSP Implementation

In order to perform the multiply-accumulate (MAC) operations of matrices in the encryption scheme, Howe et al. [103] utilize a dedicated DSP48A1 unit of the Spartan-6 FPGA to achieve an area optimized hardware implementation of standard LWE based encryption engine. The main goal is optimizing area, while the scheme maintains balance between area and performance by using a larger Gaussian sampler.

4 CONCLUSION

Lattice-based cryptographic algorithms and protocols promise to tackle the challenges posed by deployment across diverse computing platforms, as well as for diverse use cases within reasonable security, performance and energy efficiency guarantees.

Numerous schemes and implementations tackle different trade-offs, such as memory footprint, security, performance, and energy, are mapped on a variety of platforms and are applicable to specific use cases. However, current designs are still deficient in addressing the need for agility, which is paramount to tackle the needs of emerging business models at the computing platform level. In addition, securing such platforms against physical attacks is a topic that needs to be researched.

In this manuscript, we provided a review of lattice-based cryptography, some of the proposals for lattices in computer security, their implementations in software and hardware, and their applications to key exchange and digital signatures.

5 ACKNOWLEDGEMENT

This work was supported in part with a gift from Qualcomm Research.

REFERENCES

[1] 1997. Public-Key Cryptosystems from Lattice Reduction Problems. In *Proceedings of the Annual International Cryptology Conference on Advances in Cryptology (CRYPTO '97)*.

[2] 2009. IEEE Standard Specification for Public Key Cryptographic Techniques Based on Hard Problems over Lattices. *IEEE Std 1363.1-2008* (2009).

[3] 2017. NIST: National institute for standards and technology. Postquantum Crypto Project. (2017). available at http://csrc.nist.gov/groups/ST/post-quantum-crypto/.

[4] Divesh Aggarwal, Daniel Dadush, Oded Regev, and Noah Stephens-Davidowitz. 2015. Solving the Shortest Vector Problem in 2N Time Using Discrete Gaussian Sampling: Extended Abstract. In *Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing (STOC '15)*. ACM, New York, NY, USA, 733–742. https://doi.org/10.1145/2746539.2746606

[5] Carlos Aguilar-Melchor, Joris Barrier, Serge Guelton, Adrien Guinet, Marc-Olivier Killijian, and Tancrède Lepoint. 2016. NFLlib: NTT-based Fast Lattice Library. In *Proceedings of the RSA Conference on The Cryptographers' Track (CT-RSA '16)*.

[6] Miklós Ajtai. 1996. Generating Hard Instances of Lattice Problems (Extended Abstract). In *Proceedings of the Annual ACM Symposium on Theory of Computing (STOC '96)*.

[7] Miklós Ajtai, Ravi Kumar, and Dandapani Sivakumar. 2001. A Sieve Algorithm for the Shortest Lattice Vector Problem. In *Proceedings of the Annual ACM Symposium on Theory of Computing (STOC '01)*.

[8] Sedat Akleylek, Erdem Alkım, and Zaliha Yüce Tok. 2016. Sparse Polynomial Multiplication for Lattice-Based Cryptography with Small Complexity. *The Journal of Supercomputing* (2016).

[9] Sedat Akleylek, Nina Bindel, Johannes Buchmann, Juliane Krämer, and Giorgia Azzurra Marson. 2016. An Efficient Lattice-Based Signature Scheme with Provably Secure Instantiation. In *Proceedings of the International Conference on Cryptology in Africa (AFRICACRYPT '16)*.

[10] Sedat Akleylek, Özgur Dağdelen, and Zaliha Yüce Tok. 2015. On the Efficiency of Polynomial Multiplication for Lattice-Based Cryptography on GPUs Using CUDA. In *Proceedings of the International Conference on Cryptography and Information Security in the Balkans*.

[11] Erdem Alkim, Nina Bindel, Johannes Buchmann, Özgür Dagdelen, Edward Eaton, Gus Gutoski, Juliane Krämer, and Filip Pawlega. 2015. Revisiting TESLA in The Quantum Random Oracle Model. Cryptology ePrint Archive, Report 2015/755. (2015).

[12] Erdem Alkim, Léo Ducas, Thomas Pöppelmann, and Peter Schwabe. 2015. Post-Quantum Key Exchange - a New Hope. Cryptology ePrint Archive, Report 2015/1092. (2015).

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

[13] Erdem Alkim, Léo Ducas, Thomas Pöppelmann, and Peter Schwabe. 2016. NewHope Without Reconciliation. Cryptology ePrint Archive, Report 2016/1157. (2016).

[14] Erdem Alkim, Philipp Jakubeit, and Peter Schwabe. 2016. NewHope on ARM Cortex-M. In *Proceedings of the International Conference on Security, Privacy, and Applied Cryptography Engineering (SPACE '16)*.

[15] Benny Applebaum, David Cash, Chris Peikert, and Amit Sahai. 2009. Fast Cryptographic Primitives and Circular-Secure Encryption Based on Hard Learning Problems. In *Proceedings of the Annual International Cryptology Conference on Advances in Cryptology (CRYPTO '09)*.

[16] Roberto Avanzi, Joppe Bos, Léo Ducas, Eike Kiltz, Tancrede Lepoint, Vadim Lyubashevsky, John M. Schanck, Peter Schwabe, Gregor Seiler, and Damien Stehlé. 2017. *CRYSTALS-KYBER*. Technical Report. National Institute of Standards and Technology.

[17] Aydin Aysu, Cameron Patterson, and Patrick Schaumont. 2013. Low-Cost and Area-Efficient FPGA Implementations of Lattice-Based Cryptography. In *Proceedings of the IEEE International Symposium on Hardware-Oriented Security and Trust (HOST '13)*.

[18] A. Aysu and P. Schaumont. 2016. Precomputation Methods for Hash-Based Signatures on Energy-Harvesting Platforms. *IEEE Trans. Comput.* (2016).

[19] Aydin Aysu, Bilgiday Yuce, and Patrick Schaumont. 2015. The Future of Real-Time Security: Latency-Optimized Lattice-Based Digital Signatures. *ACM Transactions on Embedded Computing Systems* (2015).

[20] Shi Bai and Steven D Galbraith. 2014. An Improved Compression Technique for Signatures Based on Learning with Errors. In *Proceedings of topics in Cryptology (CT-RSA '14)*.

[21] Shi Bai, Adeline Langlois, Tancrède Lepoint, Damien Stehlé, and Ron Steinfeld. 2015. Improved Security Proofs in Lattice-Based Cryptography: Using the Rényi Divergence Rather Than the Statistical Distance. In *Proceedings, Part I, of the 21st International Conference on Advances in Cryptology – ASIACRYPT 2015*.

[22] Abhishek Banerjee, Chris Peikert, and Alon Rosen. 2012. Pseudorandom Functions and Lattices. In *Proceedings of the Annual International Conference on Theory and Applications of Cryptographic Techniques (EUROCRYPT '12)*.

[23] Rachid El Bansarkhani. 2017. *KINDI*. Technical Report. National Institute of Standards and Technology.

[24] Paulo S. L. M. Barreto, Patrick Longa, Michael Naehrig, Jefferson E. Ricardini, and Gustavo Zanon. 2016. Sharper Ring-LWE Signatures. Cryptology ePrint Archive, Report 2016/1026. (2016).

[25] Paul Barrett. 1986. Implementing the Rivest Shamir and Adleman Public Key Encryption Algorithm on a Standard Digital Signal Processor. In *Proceedings of the Annual International Cryptology Conference on Advances in Cryptology (CRYPTO '86)*.

[26] Daniel J Bernstein. 2008. ChaCha, a variant of Salsa20. In *Workshop Record of SASC 2008: The State of the Art of Stream Ciphers*.

[27] Daniel J. Bernstein. 2008. New Stream Cipher Designs. Chapter The Salsa20 Family of Stream Ciphers.

[28] Daniel J. Bernstein, Chitchanok Chuengsatiansup, Tanja Lange, and Christine van Vredendaal. 2016. NTRU Prime: Reducing Attack Surface at Low Cost. Cryptology ePrint Archive, Report 2016/461. (2016).

[29] Daniel J. Bernstein, Chitchanok Chuengsatiansup, Tanja Lange, and Christine van Vredendaal. 2017. *NTRU Prime*. Technical Report. National Institute of Standards and Technology.

[30] Guido Bertoni, Joan Daemen, Michaël Peeters, and Gilles Van Assche. 2013. (2013).

[31] Sauvik Bhattacharya, Oscar Garcia-Morchon, Ronald Rietman, and Ludo Tolhuizen. 2017. spKEX: An Optimized Lattice-Based Key Exchange. Cryptology ePrint Archive, Report 2017/709. (2017).

[32] Nina Bindel, Sedat Akleylek, Erdem Alkim, Paulo S. L. M. Barreto, Johannes Buchmann, Edward Eaton, Gus Gutoski, Juliane Kramer, Patrick Longa, Harun Polat, Jefferson E. Ricardini, and Gustavo Zanon. 2017. *qTESLA*. Technical Report. National Institute of Standards and Technology. available at https://csrc.nist.gov/projects/post-quantum-cryptography/round-1-submissions.

[33] Nina Bindel, Johannes Buchmann, and Juliane Krämer. 2016. Lattice-Based Signature Schemes and their Sensitivity to Fault Attacks. Cryptology ePrint Archive, Report 2016/415. (2016). https://eprint.iacr.org/2016/415.

[34] Nina Bindel, Johannes Buchmann, Juliane Krämer, Heiko Mantel, Johannes Schickel, and Alexandra Weber. 2017. Bounding the cache-side-channel leakage of lattice-based signature schemes using program semantics. Cryptology ePrint Archive, Report 2017/951. (2017). https://eprint.iacr.org/2017/951.

[35] Ahmad Boorghany and Rasool Jalili. 2014. Implementation and Comparison of Lattice-based Identification Protocols on Smart Cards and Microcontrollers. Cryptology ePrint Archive, Report 2014/078. (2014).

[36] Ahmad Boorghany, Siavash Bayat Sarmadi, and Rasool Jalili. 2015. On Constrained Implementation of Lattice-Based Cryptographic Primitives and Schemes on Smart Cards. *ACM Transactions on Embedded Computing Systems* (2015).

[37] Joppe Bos, Craig Costello, Léo Ducas, Ilya Mironov, Michael Naehrig, Valeria Nikolaenko, Ananth Raghunathan, and Douglas Stebila. 2016. Frodo: Take off the Ring! Practical, Quantum-Secure Key Exchange from LWE. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS '16)*.

[38] Joppe Bos, Léo Ducas, Eike Kiltz, Tancrède Lepoint, Vadim Lyubashevsky, John M. Schanck, Peter Schwabe, and Damien Stehlé. 2017. CRYSTALS – Kyber: a CCA-secure Module-Lattice-Based KEM. Cryptology ePrint Archive, Report 2017/634. (2017).

[39] Joppe W. Bos, Craig Costello, Michael Naehrig, and Douglas Stebila. 2015. Post-Quantum Key Exchange for the TLS Protocol from the Ring Learning with Errors Problem. In *Proceedings of the IEEE Symposium on Security and Privacy (SP '15)*.

[40] George EP Box, Mervin E Muller, et al. 1958. A note on the generation of random normal deviates. *The annals of mathematical statistics* (1958).

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

[41] Matt Braithwaite. 2016. Experimenting with post-quantum cryptography. (2016). https://security.googleblog.com/2016/07/experimenting-with-post-quantum.html

[42] Zvika Brakerski, Adeline Langlois, Chris Peikert, Oded Regev, and Damien Stehlé. [n. d.]. Classical Hardness of Learning with Errors. In *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing (STOC '13)*.

[43] Johannes Buchmann, Daniel Cabarcas, Florian Göpfert, Andreas Hülsing, and Patrick Weiden. 2013. Discrete Ziggurat: A Time-Memory Trade-Off for Sampling from a Gaussian Distribution over the Integers. In *Proceedings of the International Conference on Selected Areas in Cryptography (SAC '13)*.

[44] Johannes Buchmann, Erik Dahmen, Elena Klintsevich, Katsuyuki Okeya, and Camille Vuillaume. 2007. Merkle Signatures with Virtually Unlimited Signature Capacity. In *Proceedings of the International Conference on Applied Cryptography and Network Security (ACNS '07)*.

[45] Johannes Buchmann, Florian Göpfert, Tim Güneysu, Tobias Oder, and Thomas Pöppelmann. 2016. High-Performance and Lightweight Lattice-Based Public-Key Encryption. In *Proceedings of the ACM International Workshop on IoT Privacy, Trust, and Security (IoTPTS '16)*.

[46] Cong Chen, Jeffrey Hoffstein, William Whyte, and Zhenfei Zhang. 2017. *pqNTRUSign: A modular lattice signature scheme.* Technical Report. National Institute of Standards and Technology.

[47] Donald Donglong Chen, Nele Mentens, Frederik Vercauteren, Sujoy Sinha Roy, Ray CC Cheung, Derek Pao, and Ingrid Verbauwhede. 2015. High-Speed Polynomial Multiplication Architecture for Ring-LWE and SHE Cryptosystems. *IEEE Transactions on Circuits and Systems I: Regular Papers* (2015).

[48] Donald Donglong Chen, Gavin Xiaoxu Yao, Ray CC Cheung, Derek Pao, and Cetin Kaya Koç. 2016. Parameter Space for the Architecture of FFT-Based Montgomery Modular Multiplication. *IEEE Trans. Comput.* (2016).

[49] Jung Hee Cheon, Duhyeong Kim, Joohee Lee, and Yongsoo Song. 2016. Lizard: Cut off the Tail! Practical Post-Quantum Public-Key Encryption from LWE and LWR. Cryptology ePrint Archive, Report 2016/1126. (2016).

[50] Jung Hee Cheon, Sangjoon Park, Joohee Lee, Duhyeong Kim, Yongsoo Song, Seungwan Hong, Dongwoo Kim, Jinsu Kim, Seong-Min Hong, Aaram Yun, Jeongsu Kim, Haeryong Park, Eunyoung Choi, Kimoon kim, Jun-Sub Kim, and Jieun Lee. 2017. *Lizard.* Technical Report. National Institute of Standards and Technology.

[51] Arjun Chopra. 2016. Improved Parameters for the Ring-TESLA Digital Signature Scheme. Cryptology ePrint Archive, Report 2016/1099. (2016).

[52] Arjun Chopra. 2017. GLYPH: A New Insantiation of the GLP Digital Signature Scheme. Cryptology ePrint Archive, Report 2017/766. (2017).

[53] Gu Chunsheng. 2017. Integer Version of Ring-LWE and its Applications. Cryptology ePrint Archive, Report 2017/641. (2017). https://eprint.iacr.org/2017/641.

[54] Paul G. Comba. 1990. Exponentiation cryptosystems on the IBM PC. *IBM systems journal* (1990).

[55] Stephen A Cook and Stål O Aanderaa. 1969. On the minimum computation time of functions. *Ph.D. dissertation, Harvard University* (1969).

[56] James W. Cooley and John W. Tukey. 1965. An Algorithm for the Machine Calculation of Complex journal = Mathematics of Computation, Fourier Series. (1965).

[57] Özgür Dagdelen, Rachid El Bansarkhani, Florian Göpfert, Tim Güneysu, Tobias Oder, Thomas Pöppelmann, Ana Helena Sánchez, and Peter Schwabe. 2014. High-Speed Signatures from Standard Lattices. In *Proceedings of the International Conference on Cryptology and Information Security in Latin America (LATINCRYPT '14)*.

[58] Jan-Pieter D'Anvers, Angshuman Karmakar, Sujoy Sinha Roy, and Frederik Vercauteren. 2017. *SABER: Mod-LWR based KEM.* Technical Report. National Institute of Standards and Technology.

[59] Jean Pierre David, Kassem Kalach, and Nicolas Tittley. 2007. Hardware Complexity of Modular Multiplication and Exponentiation. *IEEE Trans. Comput.* (2007).

[60] Ruan de Clercq, Sujoy Sinha Roy, Frederik Vercauteren, and Ingrid Verbauwhede. 2015. Efficient Software Implementation of Ring-LWE Encryption. In *Proceedings of the Design, Automation & Test in Europe Conference & Exhibition (DATE '15)*.

[61] Alexander W. Dent. 2003. A Designer's Guide to KEMs. In *Cryptography and Coding*, Kenneth G. Paterson (Ed.).

[62] Jintai Ding, Tsuyoshi Takagi, Xinwei Gao, and Yuntao Wang. 2017. *Ding Key Exchange.* Technical Report. National Institute of Standards and Technology.

[63] Jintai Ding, Xiang Xie, and Xiaodong Lin. 2012. A Simple Provably Secure Key Exchange Scheme Based on the Learning with Errors Problem. Cryptology ePrint Archive, Report 2012/688. (2012).

[64] C. Du and G. Ba. 2016. High-Performance Software Implementation of Discrete Gaussian Sampling for Lattice-Based Cryptography. In *Proceedings of the IEEE Information Technology, Networking, Electronic and Automation Control Conference*.

[65] Chaohui Du and Guoqiang Bai. 2015. Towards Efficient Discrete Gaussian Sampling for Lattice-Based Cryptography. In *Proceeding of the International Conference on Field Programmable Logic and Applications (FPL '15)*.

[66] Chaohui Du and Guoqiang Bai. 2016. Efficient polynomial multiplier architecture for Ring-LWE based public key cryptosystems. In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS' 16)*.

[67] Chaohui Du and Guoqiang Bai. 2016. A Family of Scalable Polynomial Multiplier Architectures for Ring-LWE Based Cryptosystems. (2016).

[68] Chaohui Du and Guoqiang Bai. 2016. Towards Efficient Polynomial Multiplication for Battice-Based Cryptography. In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS' 16)*.

[69] Chaohui Du, Guoqiang Bai, and Xingjun Wu. 2016. High-Speed Polynomial Multiplier Architecture for Ring-LWE Based Public Key Cryptosystems. In *Proceedings of the International Great Lakes Symposium on VLSI (GLSVLSI '16)*.

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

[70] Léo Ducas. 2014. Accelerating Bliss: The Geometry of Ternary Polynomials. Cryptology ePrint Archive, Report 2014/874. (2014).

[71] Léo Ducas, Alain Durmus, Tancrède Lepoint, and Vadim Lyubashevsky. 2013. Lattice Signatures and Bimodal Gaussians. In *Proceedings of the Annual International Cryptology Conference on Advances in Cryptology.*

[72] Léo Ducas, Eike Kiltz, Tancrède Lepoint, Vadim Lyubashevsky, Peter Schwabe, Gregor Seiler, and Damien Stehlé. 2017. *CRYSTALS-Dilithium.* Technical Report. National Institute of Standards and Technology.

[73] Léo Ducas, Tancrède Lepoint, Vadim Lyubashevsky, Peter Schwabe, Gregor Seiler, and Damien Stehlé. 2017. CRYSTALS – Dilithium: Digital Signatures from Module Lattices. Cryptology ePrint Archive, Report 2017/633. (2017).

[74] Léo Ducas, Vadim Lyubashevsky, and Thomas Prest. 2014. Efficient Identity-Based Encryption over NTRU Lattices. In *Proceedings of the International Conference on The Theory and Application of Cryptology and Information Security (ASIACRYPT '14).*

[75] Léo Ducas and Phong Q. Nguyen. 2012. Faster Gaussian Lattice Sampling Using Lazy Floating-point Arithmetic. In *Proceedings of the International Conference on The Theory and Application of Cryptology and Information Security (ASIACRYPT '12).*

[76] Léo Ducas and Thomas Prest. [n. d.]. Fast Fourier Orthogonalization. In *Proceedings of the ACM on International Symposium on Symbolic and Algebraic Computation (ISSAC '16).*

[77] Nagarjun C. Dwarakanath and Steven D. Galbraith. 2014. Sampling from Discrete Gaussians for Lattice-based Cryptography on a Constrained Device. *Applicable Algebra in Engineering, Communication and Computing* (2014).

[78] Morris J Dworkin. 2015. *SHA-3 standard: Permutation-based hash and extendable-output functions.* Technical Report.

[79] Rachid El Bansarkhani and Johannes Buchmann. 2013. Improvement and Efficient Implementation of a Lattice-Based Signature Scheme. In *Proceedings of the International Conference on Selected Areas in Cryptography (SAC '13).*

[80] Pavel Emeliyanenko. 2009. Efficient Multiplication of Polynomials on Graphics Hardware. In *Proceedings of the International Symposium on Advanced Parallel Processing Technologies (APPT '09).*

[81] Scott Fluhrer. 2016. Cryptanalysis of Ring-LWE Based Key Exchange with Key Share Reuse. Cryptology ePrint Archive, Report 2016/085. (2016).

[82] János Folláth. 2014. Gaussian Sampling in Lattice Based Cryptography. *Tatra Mountains Mathematical Publications* (2014).

[83] Pierre-Alain Fouque, Jeffrey Hoffstein, Paul Kirchner, Vadim Lyubashevsky, Thomas Pornin, Thomas Prest, Thomas Ricosset, Gregor Seiler, William Whyte, and Zhenfei Zhang. 2017. *FALCON: Fast-Fourier Lattice-based Compact Signatures over NTRU.* Technical Report. National Institute of Standards and Technology.

[84] Eiichiro Fujisaki and Tatsuaki Okamoto. 1999. How to Enhance the Security of Public-Key Encryption at Minimum Cost. In *Proceedings of the Second International Workshop on Practice and Theory in Public Key Cryptography (PKC '99).*

[85] Martin Fürer. 2009. Faster integer multiplication. *SIAM J. Comput.* (2009).

[86] Oscar Garcia-Morchon, Ronald Rietman, Sahil Sharma, Ludo Tolhuizen, and Jose Luis Torre-Arce. 2015. DTLS-HIMMO: Achieving DTLS Certificate Security with Symmetric Key Overhead. In *Proceedings of the European Symposium on Computer Security (ESORICS '15).*

[87] Oscar Garcia-Morchon, Zhenfei Zhang, Sauvik Bhattacharya, Ronald Rietman, Ludo Tolhuizen, and Jose-Luis Torre-Arce. 2017. *Round2.* Technical Report. National Institute of Standards and Technology.

[88] Craig Gentry, Chris Peikert, and Vinod Vaikuntanathan. 2008. Trapdoors for Hard Lattices and New Cryptographic Constructions. In *Proceedings of the Annual ACM Symposium on Theory of Computing (STOC '08).*

[89] Norman Göttert, Thomas Feller, Michael Schneider, Johannes Buchmann, and Sorin Huss. 2012. On the Design of Hardware Building Blocks for Modern Lattice-based Encryption Schemes. In *Proceedings of the International Conference on Cryptographic Hardware and Embedded Systems (CHES '12).*

[90] Shay Gueron and Fabian Schlieker. 2016. Speeding up R-LWE Post-Quantum Key Exchange. Cryptology ePrint Archive, Report 2016/467. (2016).

[91] Tim Güneysu, Vadim Lyubashevsky, and Thomas Pöppelmann. 2012. Practical Lattice-based Cryptography: A Signature Scheme for Embedded Systems. In *Proceedings of the International Conference on Cryptographic Hardware and Embedded Systems (CHES '12).*

[92] Tim Güneysu, Vadim Lyubashevsky, and Thomas Pöppelmann. 2015. Lattice-Sased Signatures: Optimization and Implementation on Reconfigurable Hardware. *IEEE Trans. Comput.* (2015).

[93] Tim Güneysu, Tobias Oder, Thomas Pöppelmann, and Peter Schwabe. 2013. Software Speed Records for Lattice-Based Signatures. In *Proceedings of the International Workshop on Post-Quantum Cryptography (PQCrypto '13).*

[94] Tamás Györfi, Octavian Cret, and Zalán Borsos. 2013. Implementing Modular FFTs in FPGAs – A Basic Block for Lattice-Based Cryptography. In *Proceedings of the Euromicro Conference on Digital System Design (DSD '13).*

[95] Mike Hamburg. 2017. *Three Bears.* Technical Report. National Institute of Standards and Technology.

[96] Jeffrey Hoffstein, Nick Howgrave-Graham, Jill Pipher, Joseph H. Silverman, and William Whyte. 2003. NTRUSign: Digital Signatures Using the NTRU Lattice. In *Proceedings of the RSA Conference on The Cryptographers' Track (CT-RSA '03).*

[97] Jeff Hoffstein, Jill Pipher, John M. Schanck, Joseph H. Silverman, and William Whyte. 2014. Transcript Secure Signatures Based on Modular Lattices. In *Proceedings of the International Workshop on Post-Quantum Cryptography (PQCrypto '14).*

[98] Jeff Hoffstein, Jill Pipher, John M. Schanck, Joseph H. Silverman, William Whyte, and Zhenfei Zhang. 2017. Choosing Parameters for NTRUEncrypt. In *Proceedings of the RSA Conference on The Cryptographers' Track (CT-RSA '17).*

[99] Jeffrey Hoffstein, Jill Pipher, and Joseph H. Silverman. 1998. NTRU: A Ring-Based Public Key Cryptosystem. In *Proceedings of the International Symposium on Algorithmic Number Theory (ANTS-III).*

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

[100] Jeffrey Hoffstein, Jill Pipher, William Whyte, and Zhenfei Zhang. 2017. A signature scheme from Learning with Truncation. Cryptology ePrint Archive, Report 2017/995. (2017).

[101] Dennis Hofheinz, Kathrin Hövelmanns, and Eike Kiltz. 2017. A Modular Analysis of the Fujisaki-Okamoto Transformation. Cryptology ePrint Archive, Report 2017/604. (2017).

[102] J. Howe, A. Khalid, C. Rafferty, F. Regazzoni, and M. O'Neill. 2016. On Practical Discrete Gaussian Samplers For Lattice-Based Cryptography. *IEEE Trans. Comput.* (2016).

[103] J. Howe, C. Moore, M. O'Neill, F. Regazzoni, T. Güneysu, and K. Beeden. 2016. Lattice-based Encryption Over Standard Lattices In Hardware. In *Proceedings of the Annual Design Automation Conference (DAC '16).*

[104] James Howe, Thomas Pöppelmann, Máire O'neill, Elizabeth O'sullivan, and Tim Güneysu. 2015. Practical Lattice-Based Digital Signature Schemes. *ACM Transactions on Embedded Computing Systems* (2015).

[105] J. Howe, C. Rafferty, A. Khalid, and M. O'Neill. 2017. Compact and Provably Secure Lattice-Based Signatures in Hardware. (2017).

[106] Nick Howgrave-Graham, Joseph H. Silverman, Ari Singer, and William Whyte. 2003. NAEP: Provable Security in the Presence of Decryption Failures. Cryptology ePrint Archive, Report 2003/172. (2003). https://eprint.iacr.org/2003/172.

[107] Andreas Hülsing, Joost Rijneveld, John Schanck, and Peter Schwabe. 2017. High-Speed Key Encapsulation from NTRU. In *Proceedings of the International Conference on Cryptographic Hardware and Embedded Systems (CHES '17).*

[108] Andreas Hülsing, Joost Rijneveld, John M. Schanck, and Peter Schwabe. 2017. *NTRU-HRSS-KEM.* Technical Report. National Institute of Standards and Technology.

[109] Zhengzhong Jin and Yunlei Zhao. 2017. Optimal Key Consensus in Presence of Noise. Cryptology ePrint Archive, Report 2017/1058. (2017).

[110] Anatolii Karatsuba and Yu Ofman. 1963. Multiplication of Many-Digital Numbers by Automatic Computers. In *Proceedings of the USSR Academy of Sciences.*

[111] John Kelsey. 2016. SHA-3 Derived Functions: cSHAKE, KMAC, TupleHash, and ParallelHash. *NIST special publication* (2016).

[112] A Khalid, J Howe, C Rafferty, and M O'Neill. 2016. Time-Independent Discrete Gaussian Sampling For Post-Quantum Cryptography. In *Proceedings of the International Conference on Field-Programmable Technology (FPT '16).*

[113] Eike Kiltz, Vadim Lyubashevsky, and Christian Schaffner. 2017. A Concrete Treatment of Fiat-Shamir Signatures in the Quantum Random-Oracle Model. Cryptology ePrint Archive, Report 2017/916. (2017). https://eprint.iacr.org/2017/916.

[114] Donald E. Knuth. 1997. *The Art of Computer Programming, Volume 2 (3rd Ed.): Seminumerical Algorithms.*

[115] Donald E Knuth and Andrew C Yao. 1976. The Complexity of Nonuniform Random Number Generation. *Algorithms and complexity: new directions and recent results* (1976).

[116] Po-Chun Kuo, Wen-Ding Li, Yu-Wei Chen, Yuan-Che Hsu, Bo-Yuan Peng, Chen-Mou Cheng, and Bo-Yin Yang. 2017. High Performance Post-Quantum Key Exchange on FPGAs. Cryptology ePrint Archive, Report 2017/690. (2017). https://eprint.iacr.org/2017/690.

[117] Adeline Langlois and Damien Stehlé. 2012. Worst-Case to Average-Case Reductions for Module Lattices. Cryptology ePrint Archive, Report 2012/090. (2012).

[118] Richard Lindner and Chris Peikert. 2011. Better Key Sizes (and Attacks) for LWE-based Encryption. In *Proceedings of the International Conference on Topics in Cryptology (CT-RSA'11).*

[119] Zhe Liu, Hwajeong Seo, Sujoy Sinha Roy, Johann Großschädl, Howon Kim, and Ingrid Verbauwhede. 2015. Efficient Ring-LWE Encryption on 8-bit AVR Processors. (2015).

[120] Gui-Lu Long. 2001. Grover Algorithm with Zero Theoretical Failure Rate. *Physical Review A* (2001).

[121] Patrick Longa and Michael Naehrig. 2016. Speeding up the Number Theoretic Transform for Faster Ideal Lattice-Based Cryptography. Cryptology ePrint Archive, Report 2016/504. (2016).

[122] Patrick Longa and Michael Naehrig. 2016. Speeding up the Number Theoretic Transform for Faster Ideal Lattice-Based Cryptography. In *Proceedings of the International Conference on Cryptology and Network Security (CANS '16).*

[123] Xianhui Lu, Yamin Liu, Dingding Jia, Haiyang Xue, Jingnan He, and Zhenfei Zhang. 2017. *LAC.* Technical Report. National Institute of Standards and Technology.

[124] Vadim Lyubashevsky. 2009. Fiat-Shamir with Aborts: Applications to Lattice and Factoring-Based Signatures. In *Proceedings of the International Conference on the Theory and Application of Cryptology and Information Security: Advances in Cryptology (ASIACRYPT '09).*

[125] Vadim Lyubashevsky. 2012. Lattice Signatures Without Trapdoors. In *Proceedings of the Annual International Conference on Theory and Applications of Cryptographic Techniques (EUROCRYPT '12).*

[126] Vadim Lyubashevsky and Daniele Micciancio. 2009. On Bounded Distance Decoding, Unique Shortest Vectors, and the Minimum Distance Problem. In *Proceedings of the Annual International Cryptology Conference on Advances in Cryptology (CRYPTO '13).*

[127] Vadim Lyubashevsky, Chris Peikert, and Oded Regev. 2010. On Ideal Lattices and Learning with Errors over Rings. In *Proceedings of the Annual International Conference on Theory and Applications of Cryptographic Techniques (EUROCRYPT'10).*

[128] George Marsaglia, Wai Wan Tsang, et al. 2000. The Ziggurat Method for Generating Random Variables. *Journal of statistical software* (2000).

[129] Daniele Micciancio. 2010. *Cryptographic Functions from Worst-Case Complexity Assumptions.*

[130] Daniele Micciancio and Chris Peikert. 2012. Trapdoors for Lattices: Simpler, Tighter, Faster, Smaller. In *Proceedings of the Annual International Conference on Theory and Applications of Cryptographic Techniques (EUROCRYPT '12).*

[131] Daniele Micciancio and Oded Regev. 2007. Worst-Case to Average-Case Reductions Based on Gaussian Measures. *SIAM J. Comput.* (2007).

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

[132] Daniele Micciancio and Oded Regev. 2009. *Lattice-based cryptography.*

[133] Daniele Micciancio and Panagiotis Voulgaris. 2010. Faster Exponential Time Algorithms for the Shortest Vector Problem. In *Proceedings of the Twenty-first Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '10).* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1468–1480. http://dl.acm.org/citation.cfm?id=1873601.1873720

[134] Daniele Micciancio and Michael Walter. 2017. Gaussian Sampling over the Integers: Efficient, Generic, Constant-Time. Cryptology ePrint Archive, Report 2017/259. (2017).

[135] Peter L Montgomery. 1985. Modular Multiplication Without Trial Division. *Mathematics of computation* (1985).

[136] Shruti More and Raj Katti. 2015. Discrete Gaussian Sampling for Low-Power Devices. In *Proceedings of the IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM '15).*

[137] Michael Naehrig, Erdem Alkim, Joppe Bos, Léo Ducas, Karen Easterbrook, Brian LaMacchia, Patrick Longa, Ilya Mironov, Valeria Nikolaenko, Christopher Peikert, Ananth Raghunathan, and Douglas Stebila. 2017. *FrodoKEM.* Technical Report. National Institute of Standards and Technology.

[138] Hamid Nejatollahi, Nikil Dutt, and Rosario Cammarota. 2017. Trends, Challenges and Needs for Lattice-based Cryptography Implementations: Special Session. In *Proceedings of the IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis Companion (CODES '17).*

[139] Henri Nussbaumer. 1980. Fast Polynomial Transform Algorithms for Digital Convolution. *IEEE Transactions on Acoustics, Speech, and Signal Processing* (1980).

[140] Tobias Oder, Tim Güneysu, Felipe Valencia, Ayesha Khalid, Maire O'Neill, and Francesco Regazzoni. 2016. Lattice-Based Cryptography: From Reconfigurable Hardware to ASIC. In *Proceedings of the International Symposium on Integrated Circuits (ISIC '16).*

[141] Tobias Oder, Thomas Pöppelmann, and Tim Güneysu. 2014. Beyond ECDSA and RSA: Lattice-based digital signatures on constrained devices. In *Proceedings of the Annual Design Automation Conference (DAC '14).*

[142] Chris Peikert. 2010. An Efficient and Parallel Gaussian Sampler for Lattices. In *Proceedings of the Annual Conference on Advances in Cryptology (CRYPTO'10).*

[143] Chris Peikert. 2014. Lattice Cryptography for the Internet. In *Proceedings of the International Workshop on Post-Quantum Cryptography (PQCrypto '14).*

[144] Chris Peikert, Vinod Vaikuntanathan, and Brent Waters. 2008. A Framework for Efficient and Composable Oblivious Transfer. In *Proceedings of the Annual Conference on Cryptology: Advances in Cryptology (CRYPTO '08).*

[145] Le Trieu Phong, Takuya Hayashi, Yoshinori Aono, and Shiho Moriai. 2017. *LOTUS.* Technical Report. National Institute of Standards and Technology.

[146] Thomas Plantard. 2017. *Odd Manhattan.* Technical Report. National Institute of Standards and Technology.

[147] Thomas Pöppelmann. 2016. *Efficient Implementation of Ideal Lattice-Based Cryptography.* Ruhr-Universität Bochum.

[148] Thomas Poppelmann, Erdem Alkim, Roberto Avanzi, Joppe Bos, Léo Ducas, Antonio de la Piedra, Peter Schwabe, and Douglas Stebila. 2017. *NewHope.* Technical Report. National Institute of Standards and Technology.

[149] Thomas Pöppelmann, Léo Ducas, and Tim Güneysu. 2014. Enhanced Lattice-Based Signatures on Reconfigurable Hardware. In *Proceedings of the International Conference on Cryptographic Hardware and Embedded Systems (CHES '14).*

[150] Thomas Pöppelmann and Tim Güneysu. 2012. Towards Efficient Arithmetic for Lattice-Based Cryptography on Reconfigurable Hardware. In *Proceedings of the International Conference on Cryptology and Information Security in Latin America (LATINCRYPT '12).*

[151] Thomas Pöppelmann and Tim Güneysu. 2013. Towards Practical Lattice-Based Public-Key Encryption on Reconfigurable Hardware. In *Proceedings of the Revised Selected Papers on Selected Areas in Cryptography (SAC '13).*

[152] Thomas Pöppelmann and Tim Güneysu. 2014. Area Optimization of Lightweight Lattice-Based Encryption on Reconfigurable Hardware. In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '14).*

[153] Thomas Pöppelmann, Tobias Oder, and Tim Güneysu. 2015. High-Performance Ideal Lattice-Based Cryptography on 8-Bit ATxmega Microcontrollers. In *Proceedings of the International Conference on Cryptology and Information Security in Latin America (LATINCRYPT '15).*

[154] Thomas Pöppelmann, Tobias Oder, and Tim Güneysu. 2015. High-Performance Ideal Lattice-Based Cryptography on 8-bit ATxmega Microcontrollers. Cryptology ePrint Archive, Report 2015/382. (2015).

[155] El Bansarkhani Rachid. 2017. LARA - A Design Concept for Lattice-based Encryption. Cryptology ePrint Archive, Report 2017/049. (2017).

[156] Ciara Rafferty, Maire O'Neill, and Neil Hanley. 2017. Evaluation of Large Integer Multiplication Methods on Hardware. *IEEE Trans. Comput.* (2017).

[157] Oded Regev. 2005. On Lattices, Learning with Errors, Random Linear Codes, and Cryptography. (2005).

[158] Oscar Reparaz, Sujoy Sinha Roy, Ruan de Clercq, Frederik Vercauteren, and Ingrid Verbauwhede. 2016. Masking Ring-LWE. *Journal of Cryptographic Engineering* (2016).

[159] Miruna Rosca, Amin Sakzad, Ron Steinfeld, and Damien Stehlé. 2017. Middle-Product Learning With Errors. Cryptology ePrint Archive, Report 2017/628. (2017). https://eprint.iacr.org/2017/628.

[160] Sujoy Sinha Roy, Oscar Reparaz, Frederik Vercauteren, and Ingrid Verbauwhede. 2014. Compact and Side Channel Secure Discrete Gaussian Sampling. Cryptology ePrint Archive, Report 2014/591. (2014).

[161] Sujoy Sinha Roy, Frederik Vercauteren, Nele Mentens, Donald Donglong Chen, and Ingrid Verbauwhede. 2014. Compact Ring-LWE Cryptoprocessor. In *Proceedings of the International Conference on Cryptographic Hardware and Embedded Systems (CHES'14).*

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343

[162] Sujoy Sinha Roy, Frederik Vercauteren, and Ingrid Verbauwhede. 2013. High Precision Discrete Gaussian Sampling on FPGAs. In *Proceedings of the International Conference on Selected Areas in Cryptography (SAC '13)*.

[163] Markku-Juhani O. Saarinen. 2015. Gaussian Sampling Precision in Lattice Cryptography. Cryptology ePrint Archive, Report 2015/953. (2015).

[164] Markku-Juhani O Saarinen. 2017. Arithmetic Coding and Blinding Countermeasures for Lattice Signatures. *Journal of Cryptographic Engineering* (2017).

[165] Markku-Juhani O. Saarinen. 2017. *HILA5*. Technical Report. National Institute of Standards and Technology.

[166] Markku-Juhani O. Saarinen. 2017. HILA5: On Reliability, Reconciliation, and Error Correction for Ring-LWE Encryption. Cryptology ePrint Archive, Report 2017/424. (2017).

[167] Markku-Juhani Olavi Saarinen. 2017. Ring-LWE Ciphertext Compression and Error Correction: Tools for Lightweight Post-Quantum Cryptography. In *Proceedings of the ACM International Workshop on IoT Privacy, Trust, and Security (IoTPTS '17)*.

[168] Arnold Schönhage and Volker Strassen. 1971. Schnelle Multiplikation Grosser Zahlen. *Computing* (1971).

[169] Minhye Seo, Jong Hwan Park, Dong Hoon Lee, Suhri Kim, and Seung-Joon Lee. 2017. *EMBLEM and R.EMBLEM*. Technical Report. National Institute of Standards and Technology.

[170] Peter W. Shor. 1997. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. *SIAM J. Comput.* (1997).

[171] Victor Shoup. 2016. NTL: a Library for Doing Number Theory. *URL http://www.shoup.net/ntl* (2016).

[172] Nigel P. Smart, Martin R. Albrecht, Yehuda Lindell, Emmanuela Orsini, Valery Osheter, Kenny Paterson, and Guy Peer. 2017. *LIMA*. Technical Report. National Institute of Standards and Technology.

[173] Douglas Stebila and Michele Mosca. 2016. Post-Quantum Key Exchange for the Internet and the Open Quantum Safe Project. Cryptology ePrint Archive, Report 2016/1017. (2016).

[174] Damien Stehlé and Ron Steinfeld. 2011. Making NTRU As Secure As Worst-case Problems over Ideal Lattices. In *Proceedings of the 30th Annual International Conference on Theory and Applications of Cryptographic Techniques: Advances in Cryptology (EUROCRYPT '11)*.

[175] Damien Stehlé and Ron Steinfeld. 2011. Making NTRU as Secure as Worst-Case Problems over Ideal Lattices. In *Proceedings of the 30th Annual International Conference on Theory and Applications of Cryptographic Techniques: Advances in Cryptology (EUROCRYPT '11)*.

[176] Damien Stehlé, Ron Steinfeld, Keisuke Tanaka, and Keita Xagawa. 2009. Efficient Public Key Encryption Based on Ideal Lattices. In *Proceedings of the 15th International Conference on the Theory and Application of Cryptology and Information Security: Advances in Cryptology (ASIACRYPT '09)*.

[177] Ron Steinfeld, Amin Sakzad, and Raymond K. Zhao. 2017. *Titanium*. Technical Report. National Institute of Standards and Technology. available at https://csrc.nist.gov/projects/post-quantum-cryptography/round-1-submissions.

[178] Silvan Streit and Fabrizio De Santis. 2017. Post-Quantum Key Exchange on ARMv8-A – A New Hope for NEON made Simple. Cryptology ePrint Archive, Report 2017/388. (2017).

[179] David B Thomas, Wayne Luk, Philip HW Leong, and John D Villasenor. 2007. Gaussian Random Number Generators. *Comput. Surveys* (2007).

[180] Oder Tobias and Güneysu Tim. 2017. Implementing the NewHope-Simple Key Exchange on Low-Cost FPGAs. In *Proceedings of the International Conference on Cryptology and Information Security in Latin America (LATINCRYPT '17)*.

[181] John Von Neumann. 1951. Various Techniques Used in Connection With Random Digits. *National Bureau of Standards Applied Mathematics Series* (1951).

[182] Franz Winkler. 1996. Polynomial Algorithms in Computer Algebra. In *Texts & Monographs in Symbolic Computation*.

[183] Ye Yuan, Chen-Mou Cheng, Shinsaku Kiyomoto, Yutaka Miyake, and Tsuyoshi Takagi. 2016. Portable Implementation of Lattice-Based Cryptography Using JavaScript. *Proceedings of the International Symposium on Computing and Networking*.

[184] Zhenfei Zhang, Cong Chen, Jeffrey Hoffstein, and William Whyte. 2017. *NTRUEncrypt*. Technical Report. National Institute of Standards and Technology.

[185] Yunlei Zhao, Zhengzhong jin, Boru Gong, and Guangye Sui. 2017. *A Modular and Systematic Approach to Key Establishment and Public-Key Encryption Based on LWE and Its Variants*. Technical Report. National Institute of Standards and Technology. available at https://csrc.nist.gov/projects/post-quantum-cryptography/round-1-submissions.

Network System Technologies Ex. 2063
Qualcomm v. Network System Technologies - IPR 2024-00343