后量子密码算法与芯片设计研究进展

刘冬生,李奥博,胡昂,陆家昊,黄天泽,杨朔,李翔,张嘉明

华中科技大学集成电路学院,武汉 430074

摘要 后量子密码是用于抵御量子计算机攻击的新一代密码技术,被视为传统密码系统的可靠替代,国际上相关标准也正在逐步形成。综述了后量子密码的发展历程,分析了当前算法研究的最新进展、数学原理及属性特点。从算法、硬件实现架构及具体电路实现3个层次展开分析,提出了未来研究所需要攻克的高效硬件实现、动态可重构、侧信道攻击防御及安全SoC集成等关键技术。通过对低功耗后量子密码芯片、高性能后量子密码芯片及芯片中的哈希散列、随机采样、运算加速和逻辑处理等核心模块进行了综述,总结了当前芯片实现在核心算子高效IP设计、多场景应用兼容、多元防御机制和信息基础设施融合等重点领域的应用现状与研究价值,以及产业化与多元化方面的未来发展趋势。

关键词 后量子密码;信息安全;算法原理;高效实现;芯片设计

现代信息安全以信息的保密性、完整性和可用性三大基本要素为核心,并具有可靠性、可控性及抗抵赖属性,作为信息化时代其他领域建设和发展的保障,已经上升到了国家战略层面。从无处不在的物联网终端设备到超大规模的云计算平台,各种应用信息的传输、交换与存储皆在人们认为足够安全的密码体制保护下进行。

当今信息安全领域广泛使用的公钥密码体制主要基于难以求解的数学难题所构造,例如,RSA(Rivest-Shamir-Adleman)是基于大整数分解难题,Diffie-Hellman和ElGamal是基于离散对数问题,椭圆曲线密码(Elliptic Curve Cryptography,ECC)体制则是基于椭圆曲线离散对数问题。在经典的计算机架构下,RSA、ECC加密算法及Diffie-Hellman密钥交换算法等所依赖的底层数学问题足够困难而无法在有效时间内求解。然而,随着量子技术的发展,面对拥有海量算力且计算方式特殊的量子计算机,传统公钥密码体制通过采用增加密钥长度和参数大小来抵御安全攻击的方式不再有效,将暴露出更大的安全隐患。Peter Shor于1994年提出采用量子傅里叶变换的Shor算法[1],这是第一个能在多项式时间内求解大整数分解问题的量子算法,并经一定改进后被证实能够有效求解椭圆曲线离散对数问题。

收稿日期:2022-11-17;修回日期:2023-01-18

作者简介:刘冬生,教授,研究方向为后量子密码算法及密码芯片等,电子信箱:dsliu@mail.hust.edu.cn

引用格式:刘冬生, 李奥博, 胡昂, 等. 后量子密码算法与芯片设计研究进展[J]. 科技导报, 2024, 42(2): 20-30; doi: 10.3981/j.issn.1000-7857.2024.02.003

据研究人员预估,未来10年商用量子计算机将会面世[2]。在量子计算机面前,构造传统公钥密码体制所基于的数学难题将毫无安全性可言,进而依赖密码体制构建的信息安全系统及各种应用将面临严峻的安全问题,甚至存在被完全破解的潜在威胁。因此,亟待研究能够抵御量子攻击的密码体制及相关技术,来应对量子计算时代所面临的各种信息安全问题,也就是后量子密码。

后量子密码不仅具备传统计算范式下的安全性,且具有独特的量子计算攻击抗性,因此也被称为抗量子密码[3]。2022年4月,美国众议院推动联邦政府信息技术系统向后量子密码算法过渡,中国也正在积极推动相关研究和产业应用。可以预见,,后量子密码将全面替代传统的公钥密码体制,届时与信息安全相关的行业都会产生巨大的变化与需求。考虑到后量子密码从云到端的部署与应用方面都离不开后量子密码芯片的支撑,因此,设计出实用性、灵活性、高效性与安全性有机统一的后量子密码芯片至关重要。

1　后量子密码算法研究

为了尽早部署能抵抗量子计算机攻击的密码算法,2012年美国国家标准技术研究院(National Institute of Standards and Technology,NIST)宣布,现有的公钥加密技术需要逐渐过渡到具有量子安全或者说后量子替代方案(post-quantum alternatives)上,并且正式启动了后量子密码(post-quantum cryptography,PQC)的研究工作。2016年2月,NIST的PQC Project宣布开展后量子密码标准征集工作[4],主要聚焦于2类后量子密码算法的征集:即公钥加密算法(包括密钥封装机制)和数字签名。

此次算法标准征集面向全球范围展开,共25个国家和地区的密码学家参与。PQC标准的第一轮草案提交截止于2017年11月30日,NIST共收到82个后量子密码算法草案。在进行初步审查后,NIST最终公布了69个"完整且合适"的草案正式进入第一轮筛选。在这些候选草案中,主要包括以下数学方法构造的后量子密码算法:基于格(Lattice-based)的共计28项、基于编码(Code-based)的共计20项、基于多变量的(Multivariate-based)的共计10项、基于哈希(Hash-based)的共计3项,以及包含基于超奇异同源(Isogeny-based)在内的其他种类共计8项。随后,26个后量子密码方案进入了第二轮筛选,包括NewHope(密钥交换)、Rainbow(数字签名)等知名算法。2020年7月,NIST PQC计划宣布只有7个后量子密码方案入围了第三轮决赛筛选[5],其中包括CRYSTALS-KYBER、CRYSTALS-Dilithium(数字签名)、FALCON(数字签名)、NTRU和SABER这5个基于格的后量子密码方案以及基于编码的Classic McEliece与基于多变量的Rainbow方案。2022年7月,NIST公布了竞赛结果,宣布CRYSTALS-KYBER(以下简称KYBER/Kyber)、CRYSTALS-Dilithium(以下简称Dilithium)、FALCON和SPHINCS+这4种后量子密码算法将进入待标准化进程中。出于后量子密码算法多样化的目的,对部分算法将进行第四轮评选工作。

按照NIST PQC Project的规划,2022—2024年PQC的标准化工作将正式完成,这些草案中将有1个或多个算法会成为PQC标准。未来RSA、ECC等传统公钥加密方案的后量子密码替代标准将逐步融入到人们信息化的工作和生活中,为抵御量子计算机的攻击做好充足的准备。不同的后量子密码算法具有相异的原理与特征,并在后续的应用范围和场景上有所区别。表1为待标准化算法与第四轮评选算法的基本属性与比较。

目前,最新的后量子密码方案所基于的数学原理仅包括以下4种:基于格(Lattice-based)、基于编码(Code-based)、基于哈希(Hash-based)和基于超奇异同源(Isogeny-based)。不同的数学原理会给算法方案带来差异化的参数和性能。

基于格的后量子密码算法依靠其公私钥尺寸小、计算速度快、灵活性强等优点,成为近年来的研究热点。格理论不仅可以实现公钥加密和数字签名算法,也可用于全同态加密等不同密码构造。与格相关的基本计算性难题有最短向量问题(shortest vector problem,SVP)和最近向量(closest vector problem,CVP)问题。这2个问题的困难性已经被

表1 后量子密码待标准化算法及第四轮评选算法特点

进程	待标准化				第四轮			
种类	公钥加密	数字签名			公钥加密			
方案	Kyber	Dilithium	FALCON	SPHINCS+	BIKE	Classic McEliece	HQC	SIKE
数学原理	基于格			基于哈希	基于编码			基于超奇异同源
密钥长度	小	小	小	大	适中	大	适中	小
运算性能	快	快	一般	慢	一般	慢	一般	慢
功能多样性	很好	很好	好	有限	好	有限	好	有限

证明是NP(non-deterministic polynomial)难。目前不存在多项式时间内求解格上困难问题的量子算法,因此基于格的密码算法被认为是能够抵御量子计算机攻击的。在格密码的发展历程中,密码界的研究者们提出的大部分密码算法是基于平均情况下的格困难问题,例如误差学习(learning with erros,LWE)问题、环域上的误差学习(ring-LWE,RLWE)问题[6]和模误差学习问题(module-LWE,MLWE)等。由于MLWE方案在安全级别和运算性能方面取得了良好的均衡,在2022年NIST的后量子密码评审中,由MLWE问题构造的Kyber和Dilithium方案成功入选并即将被标准化。

基于编码的后量子密码算法[7]的安全性依赖于编码理论中的困难问题,例如校验子解码问题和LPN(learning parity with noise)问题。这些密码系统利用纠错码来构造一个单向函数,其安全性在于消息解码和编码结构恢复过程中存在困难性,以此来实现密码学上的加解密操作和密钥交换,并为公钥加密和密钥封装提供了一种保守的方法。而线性码是纠错码的一个子类,最初用于控制不可靠或有噪声的通信信道上的数据错误,Berlekamp等[8]的工作证明了线性编码的问题为NP难。后来这边促生了使用线性码解码作为主要困难问题的McEliece密码系统。目前,该类型密码方案的主要缺陷是密钥长度过大,一些减少密钥大小的尝试使这些算法易受攻击。部分研究人员有针对性地提出了一些在不影响安全的情况下减小密钥大小的技术。

基于哈希的后量子密码算法使用散列函数加密,用于数字签名方案,它们的安全性依赖于该哈希函数的抗碰撞性[9]。抗冲突散列函数的存在可以被视为该数字签名方案的最低要求,如果可以构造2个具有相同数字签名的信息,则签名方案不再被认为是安全的。由于这些方案基于底层散列函数的安全属性(抗碰撞性和抗第一原像性),没有有效的量子算法能快速找到哈希函数的碰撞。此外,每个新的加密散列函数都会产生一个新的基于散列的签名方案。因此,安全签名方案的构建独立于数论或代数中的算法难题。

基于超奇异同源的后量子密码算法是目前历史较短的一种新型算法。与密码学领域相关最著名的工作是由Bostan等[10]在2008年完成的。从那时起,David Jao在Bostan等开发的算法基础上建立了一个密码系统[11]。该密码系统经优化后成为SIKE算法并被NIST所接受。超奇异同源可以将一个给定多项式函数映射到另一个多项式函数,椭圆曲线密钥交换将点分布在给定的多项式函数曲线上,以建立一个临时的私钥。基于超奇异椭圆曲线同源的方案相对于其他后量子密码具有密钥尺寸短的优势,但其实现的效率相比于基于编码的算法和基于格的算法均不占优势,这是由于同源密码学建立在已经比较复杂的椭圆曲线密码之上,导致其构造非常复杂。

2 后量子密码关键技术

当前,后量子密码算法标准的角逐已渐渐步入尾声,待标准化算法与仅剩的数个参评方案也将在短时间内成为标准,并取代现有的公钥基础设施

(public key infrastructure, PKI)。面对即将到来的量子计算时代与新信息安全标准革新[12],如何行之有效地应用后量子密码,占据新标准主导下的有利地位是关键所在。

如今各类信息系统与设备都是在各类通用或领域专用的平台上进行运行,而构成这些平台的核心是各类电子系统及其所依托的芯片。芯片设计是目前后量子密码发展与应用中的重要内容,包括专用集成电路(application specific integrated circuit, ASIC)和可供集成的知识产权核(intellectual property core, IP)。对后量子密码芯片及其关键技术需开展深入研究,探索核心算子高效硬件实现技术、后量子密码可重构技术、侧信道攻击防御技术及安全SoC集成技术,推动后量子密码的研究发展与应用,使中国处于一定的战略高度,否则会影响中国电子信息及物联网产业的发展,甚至波及到国家安全。

2.1　高效硬件实现技术

通过对各种后量子密码算法进行抽取和分析可以发现,其主要涉及模运算、系数生成等核心算子。后量子密码的核心计算运算涵盖了多项式、矩阵和向量间的模加、模减和模乘等运算。核心算子的硬件实现结果直接影响着整个后量子密码系统的资源开销和运算性能[13]。针对信息安全领域应用多样性和开发通用性、安全级别的多元化需求问题,在对后量子密码方案中的核心算子进行硬件实现时,需要综合考虑功耗、性能与面积的折衷关系:在优化硬件资源开销的同时保证一定的运算性能,在追求高性能的同时避免不必要的资源开销,在电路设计的过程中加入低功耗技术,从而最终取得特定应用场景下最高的硬件实现效率。

2.2　动态可重构技术

对不同后量子密码算法方案中核心运算的计算方式与特点进行研究,针对算法在加解密过程中的关键运算特征,如基本数据位宽、向量基元长度、迭代运算次数和计算属性等方面,设计可配置数据位宽、计算规模与计算方式的可重构后量子密码核心计算单元[14]。研究支持多种算法协议的可重构后量子密码芯片架构,深入分析哈希运算、采样运算、多项式运算、数论变换运算及模乘运算等关键步骤,探究差异化的高效多项式乘法器实现技术与高速可配置哈希硬件实现方案,通过可重构技术来提高后量子密码安全芯片的灵活性与适配性。

2.3　侧信道攻击防御技术

后量子密码算法以某种具体的方式实现(如芯片、程序等)后,其采样、模运算等核心算子在运行时会通过计算时间、功耗和电磁等侧信道信息泄漏敏感信息,对密码系统带来巨大的安全隐患[15],针对算法漏洞的攻击也具有很大威胁[16]。应首先在算法层面通过对易遭受侧信道攻击的核心算子进行改进优化,减少信息泄露的可能性。在具体硬件实现上,分析各个功能模块的工作机理以及可能遭受的侧信道攻击类型,重点关注密钥参与计算的解密模块,在内部设计具备抗侧信道攻击的电路结构,以有效提升系统抵御侧信道攻击的能力。

2.4　安全SoC集成技术

具有多种功能的片上系统(System on Chip, SoC)在应用中占据优势,也需要更加全面的安全防护[17]。研究汇聚多种技术的后量子密码安全IP,将其与SoC中的其他电路与功能模块有机融合,实现芯片性能、功耗、稳定性和安全性的平衡,找到高集成度设计与低功耗方案之间的最佳均衡点。在此基础上,探索SoC所参与的安全认证和实际应用的具体流程,完成后量子密码算法在各种认证协议过程中的替代与兼容,进行后量子密码芯片在安全身份认证的应用示范,并逐步取代现有的各类传统安全芯片。

3　后量子密码芯片设计

信息化时代背景下,个人电子设备与物联网终端节点在社会的方方面面铺展开来,信息通过端点或中继进行加密和掩藏,汇聚到云服务器与高算力数据中心,进行解密还原并处理,再重新加密后传输到无数设备中去。不同应用场景对后量子密码的芯片设计提出不同要求,面向资源受限型的低功耗后量子密码芯片与面向云服务器与数据中心的高性能后量子密码芯片是最具代表性的两大方向。

3.1 低功耗后量子密码芯片

基于格的后量子密码作为轻量化应用的优选方案,适用于低功耗设计与实现。基于格的后量子密码系统中的采样与多项式运算是能耗较高的核心计算,通常从这2部分进行低功耗与低资源开销的优化与改进。针对采样的低功耗设计需要提高采样效率,在硬件结构上采用耦合且紧凑的结构;而多项式运算的优化则更多通过算法改进以及提高存储利用率。

采样器作为加密系统的前端,对伪随机位字符串进行后处理,以从指定的分布中生成系数。多项式系数是通过拒绝采样或从离散分布(通常是二项分布)均匀地生成或"采样"具有适应安全等级的标准偏差系数。Pöppelmann等[18]提出一种面积优化的Bernoulli采样器,使用Bernoulli评估替代复杂的指数评估,但其采样拒绝率过高,最终导致熵耗和采样时间的增加,后续针对二项高斯分布的改进[19]实现同样具有相似缺点。Zhang等[20]实现了在资源受限设备中高效进行离散高斯采样,将概率分布表分布划分成不同子数据块来进行近似采样的方案,减少对概率值中相同的无用信息计算来节约开销,并通过假采样的方式抵抗简单功耗攻击。Zhao等[21]通过将不同格密码算法中差异模数的拒绝采样器进行融合,采用向量化的采样数据结构,并降低其采样过程中的拒绝率,实现了仅占总硬件资源中6.5%的高能效采样。二项式采样器[22]从伪随机数中获取2个比特块,并计算其汉明权重(Hamming Weights)的差异,以生成标准偏差$\sigma = \sqrt{k/2}$的样本。Bisheh-Niasar等[23]通过指令集架构实现高效且高性能的采样过程,设计紧密耦合的Keccak核心与关联采样器,并行化二项采样过程,且使拒绝采样器的延迟降低。

相较于通过优化采样过程与利用节能型采样器以实现高能效设计,对多项式运算模块的能效优化也进一步降低了系统功耗。快速数论变换(number theoretic transform, NTT)是加速多项式乘法并降低计算开销[24]最有效的方法之一,减少该过程中的访存和运算是主要思路。按照NTT的计算顺序将特定的2个数据成对存放在同一地址处[25],进行单次蝶形运算,通过减少对数据存储器访问次数来降低功耗。或采用动态执行倒位序计算的访存方案[26],消除NTT逆运算所需要的倒位序计算步骤,节省整体计算的次数。使用固定模数进行比特流运算简化[27],通过选择器实现取模运算,简化运算单元的电路结构。用基本加法器和减法器设计组合逻辑型NTT[28],可以降低运算过程中的动态翻转功耗。通过消引定理将所需要存储的预计算数据量最小化[29],能够减少迭代过程的中间变量计算次数。采取可扩展的高效多项式乘法器结构动态调整并行的蝶形运算单元数量,可取得局部的功耗优化[30]。

美国麻省理工学院Utsav Banerjee研究团队于2019年首次提出一种用于量子安全物联网的节能可配置格密码处理器[31],并完成40 nm流片及测试。该处理器通过架构优化实现了高达2个数量级的功耗降低和124 K门电路的减少,可以支持NIST后量子标准化过程的第一轮中的多个基于格的后量子密码算法。如图1[31]所示,其架构采用一个24 KB的LWE高速缓存与一个模块化算术单元接口以执行基于NTT的多项式运算,通过高能效的Keccak核心驱动离散分布采样器并实现散列求值和伪随机数生成。内核和采样器具有专用的时钟门,可以独立配置以实现更好的能耗优化。

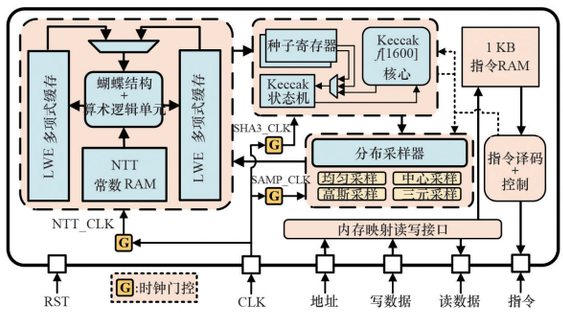

图1 可配置格密码处理器架构

芯片采用40 nm LP CMOS工艺制造,支持从1.1 V降至0.68 V的电压范围,如图2(a)[31]所示。该加密处理器共消耗106 K等效门(gate equivalent, GE),并使用40.25 KB的SRAM。在执行NewHope后量子密钥交换算法时,平均功耗为516

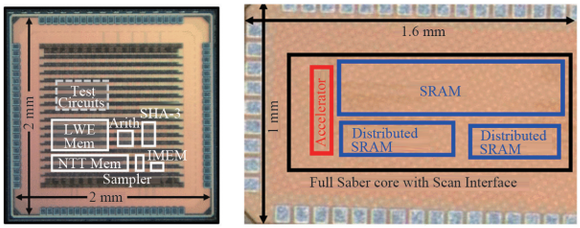

图2　可配置格密码处理器显微图(a)[31]和超低功耗后量子密码加速器显微图(b)[32]

μW(0.68 V@12 MHz)。通过架构和算法优化实现了硬件加速后量子密码算法及协议,可用于保护资源受限的物联网设备。

Ghosh等[32]采用优化多项式乘法实现更低的资源开销。其采用优化的紧凑Toom-Cook乘法器完成多项式运算,所得的多项式在模约$x^{64}+1$之后可以将内存需求降低为目前最有效软件设计的1/2。与其他新型ASIC相比,该乘法的优化可实现1.37倍的能效提高与1.75倍的电路面积缩小。多项式乘法模块经过微编程,可执行矢量乘法,以惰性插值的方法减小数据通路延迟由此减小能耗和改善性能。点值乘法器配合双端口存储器实例化多个乘法和累加单元,进一步并行化点值乘法同时提供功耗、面积和能效方面的改进。为实现更高的内存效率,将多项式乘法拓展为负包裹卷积,仅系数索引的较低有效位用于生成内存地址,便于经串行接口集成片上系统或低功耗终端节点。基于交互式指令集的架构重用构建减少了面积开销,并使其与在SoC中运行的处理器兼容。

图2(b)[32]显示了65 nm工艺下芯片显微图片及内存规划,1.1 V电压下最大频率为160 MHz。在0.7 V下能够以334 μW(@10 MHz)低功耗运行,在仅用10 KB内存的情况下实现了约80倍的算法加速,密钥生成、封装和解封装仅消耗444.1、579.4和724.5 μJ能量。

3.2　高性能后量子密码芯片

用户数据加密、数字签名与验证过程需要在网络服务器中大量且高速重复,要求芯片具备更高性能,实现高频运行和高吞吐率。并行计算、算法改进、多任务调度及可重构实现是可行的高性能设计方法,流水线结构与平衡的逻辑时序同样可以有效改善芯片性能。

Göttert等[33]提出了首个Ring-LWE公钥加密方案的硬件实现,全面使用并行计算的架构,使其完成NTT运算所需要的时钟周期数达到最小化,因此具有极高的运算性能。但设计仅考虑了运算性能这一指标,其硬件资源开销是极其庞大的,存在很大的优化空间。后续提出的灵活加密引擎[34]可以通过配置来实现密钥生成、加密和解密运算,使得单位面积的性能最大化。按列相乘的基于4-NTT多项式乘法器结构[35]来执行签名方案中的多项式乘法运算保证了低硬件资源开销和高计算速度。进一步优化与增加并行度可以满足在数字签名硬件实现中的高吞吐量。Aikata等[36]通过设计兼容不同位宽多项式乘法器与高效存储管理机制,结合快速哈希伪随机数生成器,实现首个Kyber和Dilithium的硬件设计,能够在28 nm ASIC评估中达到2 GHz时钟频率。Güneysu等[37]设计多路并行结构的NTT运算单元来执行数字签名方案中最为耗时的多项式乘法运算来提高整体性能。

耶鲁大学Mohan等[38]将后量子密码数字签名方案XMSS(与SPHINCS+算法同源)在28 nm下进行ASIC实现,通过新型流水线XMSS Leaf结构加速算法中计算最密集的步骤。XMSS Leaf加速器主要由Merkel哈希树、温特尼茨一次签名(Winternitz One-Time Signature, WOTS)和L-tree模块组成。WOTS本身可以作为密钥拓展(Key Expansion)和链式(Chain)传递,并以流水线或非流水线方式运行。

加速器的关键部分在于特定XMSS的流水线型SHA-256加速模块,称为SHA256XMSS。该模块作为XMSS树中每个圆圈的核心运算,用于计算Merkle树的一个分支页(Leaf),如图3[38]所示。对于SHA256XMSS,轮函数的实现通常是关键路径,因为在单周期实现中必须在2个寄存器级之间放置大量组合逻辑。在组合逻辑轮函数的基础上,使用四级并行流水线以实现寄存器的数量与关键路径之间的平衡,以获得更好的频率性能。并扩展了“预计算”和“固定输入长度”的优化方式从而独立

地支持4个输入,即如果一个输入使用预计算功能,则下一个输入可以不执行该运算,或者它可能使用不同的预先计算的中间状态。灵活的调度方式使得输入的计算与分支Leaf可以开始于任何时候,独立于处理的其他输入进行并行处理,提升计算性能。

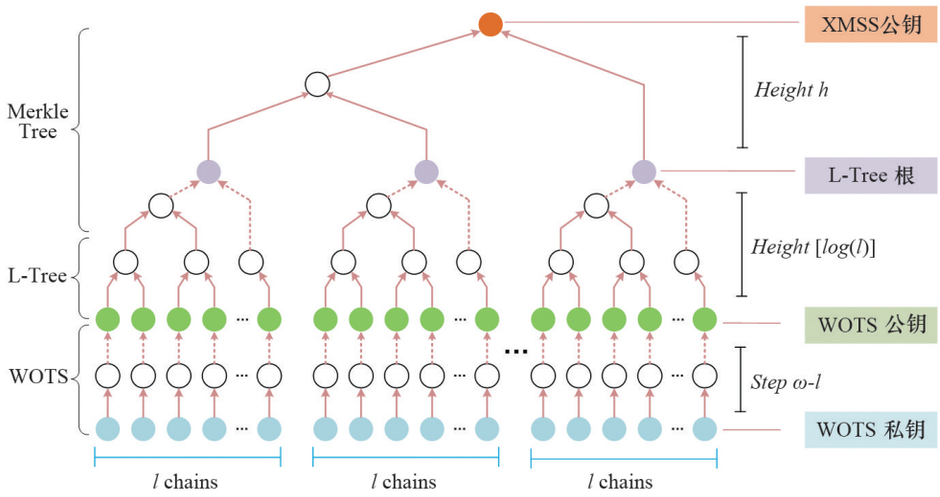

图3 XMSS树结构

上述XMSS加速器在28 nm批量工业CMOS工艺中实现了XMSS Leaf加速器的非流水线和流水线版本并进行了对比,ASIC实现的芯片显微照片如图4(a)[38]所示。多路并行化设计与优秀时序平衡使芯片能够工作在较高的时钟频率,流水线版本能够在时钟1011 MHz(@0.99 V)下运行,非流水线版本运行频率也能够达到823 MHz(@0.99 V)。虽然流水线设计帮助芯片改善了频率性能,但同样会带来额外的电路面积。从图4可以看到流水线XMSS Leaf加速器(0.13 mm^2)比非流水线XMSS Leaf加速器(0.09 mm^2)面积大44%。这是因为当以流水线方式实现时,在逻辑复杂性很高的情况下,由于添加了流水线触发器,时序逻辑面积增加了239%,而组合单元面积仅增加了43%。

更快的频率使芯片能够在固定时间内执行更多操作,可重构的计算结构与并行任务调度也可提高芯片性能。清华大学Zhu等[39]利用用于面向任务的优化型高效任务级调度程序(Task-Level Scheduler, TLS),通过分层执行框架中的可配置混合处理元素阵列(Hybrid Processing Element Array, HPEA)来运行一个并行执行流程并平衡多个独立的操作,提出了一种灵活的后量子密码处理器架构。处理器可支持NIST第三轮标准中的(Kyber、Saber、NTRU、Classic McEliece、Rainbow、Dilithium)6种算法。系统架构由数据生成引擎(Data-Generation Engine, DGE)、数据存储系统(Data-Storage System, DSS)、HPEA和TLS组成。关键设计在于TLS提取特定于任务的特征与前/后处理函数,包括多项式乘法、矩阵运算和高斯消元,它与HPEA协作以完整执行某些任务或执行独立功能,例如排序、多项式求逆、多项式移位和压缩/解压缩。采取并行调度独立的计算任务以进一步提高吞吐量。

该处理器在28 nm HPC CMOS工艺下进行流片,共消耗190万等效门,图4(b)所示其芯片核心区域面积3.6 mm^2,在0.9 V电压下最高工作主频为500 MHz,功耗为368 mW。以数字签名(密钥生成、签名、验证)的一个完整流程作为一个操作单位(Ops),其运行基于格的后量子密码方案可达到6119~556480 Ops的吞吐量,是当前可重构ECC处理器的2.9~7.9倍。

芯片具备高主频和高吞吐量的同时也会带来较大的功耗,如何平衡性能和能效逐渐成为后量子

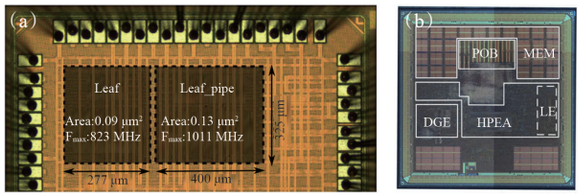

图4　XMSS加速器显微图(a)[38]和可重构后量子密码芯片显微图(b)[39]

密码芯片设计中的关键。Imran等[40]基于可重构流水线架构设计了一款密钥交换处理器,构建了关键模块间的共享逻辑,实现高时钟频率和功耗的平衡,并采用4个寄存器组(RegFile)替代常用的静态随机存取存储器(SRAM),降低数据存储开销。在65 nm工艺下进行实现,如图5(a)所示版图面积约为0.39 mm²,后仿真结果显示时钟频率能够达到1000 MHz,功耗187 mW。Xing等[41]提出一种紧凑型硬件结构的安全密钥交换处理器,采用采样-计算同步执行,使用连续寄存器组压缩存算过程,实现低开销与高性能的平衡。Xin等[42]基于RISC-V架构和自定义指令集设计了一种高性能VPQC处理器。笔者研究团队基于动态可重构算术单元与高性能NTT硬件单元核心算子,采用多功能精简微指令与可拓展的密钥-消息-密文管理(Key/Msg/Context management, KMC)策略,提出了一种后量子密码Kyber协处理器。采用快速映射分组查找表(fast look-up table),用于模运算中主要计算中,减少关键路径的深度,提高运行频率。通过奇偶分列式系数并行处理方式,将其应用于模域系数的数论变换迭代计算中,设计合理的双倍位宽乒乓式对称存储结构(ping-pong memory, PP-MEM),降低了其中多项式运算的复杂度,有效地提高了多项式运算的效率,提升了协处理器的性能,相比同类运算操作下最先进的设计快3.95倍。该设计在40 nm LP CMOS工艺下流片,共消耗30万等效门。如图5(b)所示,核心逻辑电路面积为0.34 mm²,芯片在110 MHz时钟频率下功耗仅为15 mW。支持国际最新标准化算法Kyber,实现了综合性能、功耗与面积的均衡优化设计。

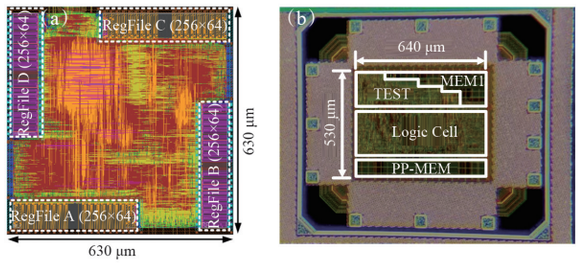

图5　密钥交换处理器版图(a)[40]和高性能后量子密码Kyber芯片显微图像(b)

表2总结了采用不同设计方法的后量子密码芯片的PPA(performance, power, area,即性能、功耗、面积)参数与特点。

表2　典型后量子密码芯片参数对比

属性	优化	工艺/nm	面积/mm²	频率/MHz	功耗/mW	特点
低功耗	采样/计算	40	2*	12	0.516	可配置[31]
	计算	65	1*	10	0.334	超低功耗[32]
高性能	流水线	28	0.13	1011	157	高频率[38]
	可重构架构	28	3.6	500	368	多算法[39]
	计算/访存	65	0.39	1000**	187**	高能效[40]
	可重构架构/计算	40	0.34	110	15	性能均衡***

注:*仅核心电路的面积,根据有效电路在整个芯片die面积进行估算;**该数据均基于后仿真结果,并非实测数据;***该芯片为笔者研究团队研究成果。

4　结论

信息安全防护已经从传统的单点信息加密发展到了以芯片级硬件防护为基础,构建覆盖全网络系统的信息保障体系。基于芯片级的硬件解决方案已经成为保证信息安全最可靠的途径。具有抗

量子计算攻击能力的安全芯片可作为未来信息安全服务、面向安全应用、向微型信息安全设备提供密码服务的重要基础部件。

通过对后量子密码算法以及芯片的发展现状进行分析可知,未来的后量子密码芯片设计将朝向更加产业化与多元化的方向发展。随着知识产权与数字版权交易的发展,核心算子高效IP设计逐渐向着应用化方向推进,为后量子密码的应用奠定坚实基础。目前多算法并存的现状和不同方案的多维属性带来了很多应用壁垒,灵活的多模可重构芯片能够从多个角度实现对大型网络、物联网、数字签名、安全认证等相异场景的兼容性。而在应用过程中,攻击者将会使用各种方式尝试窃密,各式各样的攻击方法对芯片防御强度提出更高的挑战,单一保护手段已远不能满足要求,多元侧信道攻击防御机制的研究将成为重要领域。

面向后量子密码芯片在关键信息基础设施中的高性能计算需求,开展从算法、架构到电路等多层次的高性能实现和侧信道防护是未来芯片设计领域的重要发展方向。同时,具有动态监测和安全感知的抗侧信道攻击防护技术也将成为后量子密码芯片研究的重点,对加固物理层安全性和防窃密具有重要意义。此外,针对资源敏感场景的应用难题,围绕低功耗后量子密码SoC的研究工作在物联网领域具有很好的应用价值,可提高安全逻辑单元面积的性能比,缓解性能与资源限定的冲突问题,便于后量子密码芯片嵌入和融合到现有的信息基础设施中。

当前中国在后量子密码领域的研究水平处于国际前列,应加快研制符合国际标准且具有国际竞争力的后量子密码芯片,建立应用示范系统,抢占先机,健全自主的后量子密码理论和应用研究。

参考文献(References)

[1] Shor P W. Algorithms for quantum computation: Discrete logarithms and factoring[C]//Proceedings 35th Annual Symposium on Foundations of Computer Science. Washington, DC: IEEE Computer Society Press, 1994: 124-134.

[2] Joseph D, Misoczki R, Manzano M, et al. Transitioning organizations to post-quantum cryptography[J]. Nature, 2022, 605: 237-243.

[3] Chen L, Jordan S, Liu Y K, et al. Report on post-quantum cryptography[M]. Gaithersburg, MD: National Institute of Standards and Technology, 2016.

[4] Moody D. Post-quantum cryptography: NIST's plan for the future[C]//Proceedings of the 7th International Conference on Post-Quantum Cryptography. Berlin, Heidelberg: Springer, 2016.

[5] Moody D, Alagic G, Apon D, et al. Status report on the second round of the NIST post-quantum cryptography standardization process[R]. Gaithersburg, MD: US Department of Commerce, National Institute of Standards and Technology, 2020.

[6] Peikert C. A decade of lattice cryptography[J]. Foundations and trends in theoretical computer science, 2016, 10 (4): 283-424.

[7] Mceliece R J. A public-key cryptosystem based on algebraic coding theory[J]. DSN Progress Report, 1978, 4244: 114-116.

[8] Berlekamp E R, Mceliece R J, Tilborg H. On the inherent intractability of certain coding problems (Corresp)[J]. IEEE Transactions on Information Theory, 1978, 24(3): 384-386.

[9] Merkle R C. A certified digital signature[C]//Advances in Cryptology: CRYPTO '89 Proceedings. Berlin, Heidelberg: Springer, 1989: 218-238.

[10] Bostan A, Morain F, Salvy B, et al. Fast algorithms for computing isogenies between elliptic curves[J]. Mathematics of Computation, 2008, 77(263): 1755-1778.

[11] Jao D, De Feo L. Towards quantum-resistant cryptosystems from supersingular elliptic curve isogenies[C]//Post-Quantum Cryptography: 4th International Workshop. Berlin Heidelberg: Springer, 2011: 19-34.

[12] Alagic G, Apon D, Cooper D, et al. Status report on the third round of the NIST post-quantum cryptography standardization process[R]. Gaithersburg, MD: National Institute of Standards and Technology, 2022.

[13] Basu K, Soni D, Nabeel M, et al. NIST post-quantum cryptography: A hardware evaluation study[J/OL]. Cryptology ePrint Archive, 2019, https://eprint.iacr.org/2019/047.pdf.

[14] Land G, Sasdrich P, Güneysu T. A hard crystal-implementing dilithium on reconfigurable hardware[C]//International Conference on Smart Card Research and Ad-

vanced Applications. Cham: Springer, 2021: 210-230.

[15] Wang Y C, Paccagnella R, He E T, et al. Hertzbleed: Turning power {side-channel} attacks into remote timing attacks on x86[C]//31st USENIX Security Symposium (USENIX Security 22), Berkeley, California: The Adanced Computing Systems Association, 2022: 679-697. https://www.hertzbleed.com/hertzbleed.pdf.

[16] Maino L, Martindale C. An attack on SIDH with arbitrary starting curve[J/OL]. Cryptology ePrint Archive, 2022, https://eprint.iacr.org/2022/1026.

[17] Da Costa V L R, López J, Ribeiro M V. A SoC implementation of a PQC scheme for smart meter[C]. XXXIX Brazilian Symposium on Telecommunications and Signal Processing - SBrT, 2021: 26-29.

[18] Pöppelmann T, Güneysu T. Area optimization of lightweight lattice-based encryption on reconfigurable hardware[C]//2014 IEEE International Symposium on Circuits and Systems (ISCAS). Piscataway NJ: IEEE, 2014: 2796-2799.

[19] Pöppelmann T, Ducas L, Güneysu T. Enhanced lattice-based signatures on reconfigurable hardware[C]//Cryptographic Hardware and Embedded Systems. Berlin, Heidelberg: Springer, 2014: 353-370.

[20] Zhang C, Liu Z L, Chen Y Y, et al. A flexible and generic gaussian sampler with power side-Channel countermeasures for quantum-secure internet of things[J]. IEEE Internet of Things Journal, 2020, 7(9): 8167-8177.

[21] Zhao Y F, Xie R Q, Xin G Z, et al. A high-performance domain-specific processor with matrix extension of RISC-V for module-LWE applications[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2022, 69(7): 2871-2884.

[22] Bos J, Ducas L, Kiltz E, et al. CRYSTALS-Kyber: A CCA-secure module-lattice-based KEM[C]//2018 IEEE European Symposium on Security and Privacy (EuroS&P). Piscataway NJ: IEEE, 2018: 353-367.

[23] Bisheh-Niasar M, Azarderakhsh R, Mozaffari-Kermani M. Instruction-set accelerated implementation of CRYSTALS-kyber[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2021, 68(11): 4648-4659.

[24] Karabulut E, Aysu A. RANTT: A RISC-V architecture extension for the number theoretic transform[C]//2020 30th International Conference on Field-Programmable Logic and Applications (FPL). Piscataway NJ: IEEE, 2020: 26-32.

[25] Zhang C, Liu D S, Liu X J, et al. Towards efficient hardware implementation of NTT for kyber on FPGAs[C]//2021 IEEE International Symposium on Circuits and Systems (ISCAS). Piscataway NJ: IEEE, 2021: 1-5.

[26] Du C H, Bai G Q. Towards efficient polynomial multiplication for lattice-based cryptography[C]//2016 IEEE International Symposium on Circuits and Systems (ISCAS). Piscataway NJ: IEEE, 2016: 1178-1181.

[27] Guo W B, Li S G, Kong L. An efficient implementation of KYBER[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2021, 69(3): 1562-1566.

[28] Liu D S, Zhang C, Lin H, et al. A resource-efficient and side-channel secure hardware implementation of ring-LWE cryptographic processor[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2018, 66(4): 1474-1483.

[29] Du C H, Bai G Q. A family of scalable polynomial multiplier architectures for lattice-based cryptography[C]//2015 IEEE International Conference on Trust, Security and Privacy in Computing and Communications. Piscataway NJ: IEEE, 2015: 392-399.

[30] Lyubashevsky V, Seiler G. NTTRU: Truly fast NTRU using NTT[C]//IACR Transactions on Cryptographic Hardware and Embedded Systems (TCHES). Bochum: IACR, 2019:180-201.

[31] Banerjee U, Pathak A, Chandrakasan A P. An energy-efficient configurable lattice cryptography processor for the quantum-secure Internet of Things[C]//2019 IEEE International Solid-State Circuits Conference(ISSCC). Piscataway NJ: IEEE, 2019: 46-48.

[32] Ghosh A, Mera J M B, Karmakar A, et al. A 334 uW 0.158 mm^2 saber learning with rounding based post-quantum crypto accelerator[C]//2022 IEEE Custom Integrated Circuits Conference (CICC). Piscataway NJ: IEEE, 2022: 1-2.

[33] Göttert N, Feller T, Schneider M, et al. On the design of hardware building blocks for modern lattice-based encryption schemes[C]//Cryptographic Hardware and Embedded Systems-CHES 2012. Berlin, Heidelberg: Springer, 2012: 512-529.

[34] Pöppelmann T, Güneysu T. Towards practical lattice-based public-key encryption on reconfigurable hardware [C]//Selected Areas in Cryptography-SAC 2013. Berlin, Heidelberg: Springer, 2013: 68-85.

[35] Wang T F, Zhang C, Cao P, et al. Efficient Implementation of Dilithium Signature Scheme on FPGA SoC Platform[J]. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2022, 30(9): 1158-1171.

[36] Aikata A, Mert A C, Imran M, et al. KaLi: A crystal for

post-quantum security using Kyber and Dilithium[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2022, 70(2): 747-758.

[37] Güneysu T, Lyubashevsky V, Pöppelmann T, et al. Lattice-based signatures: Optimization and implementation on reconfigurable hardware[J]. IEEE Transactions on Computers, 2015, 64(7): 1954-1967.

[38] Mohan P, Wang W, Jungk B, et al. ASIC accelerator in 28 nm for the post-quantum digital signature scheme XMSS[C]//2020 IEEE 38th International Conference on Computer Design (ICCD). Piscataway NJ: IEEE, 2020: 656-662.

[39] Zhu Y, Zhu W, Zhu M, et al. A 28nm 48KOPS 3.4 μJ/OP agile crypto-processor for post-quantum cryptography on multi-mathematical problems[C]//2022 IEEE International Solid-State Circuits Conference (ISSCC). Piscataway NJ: IEEE, 2022, 65: 514-516.

[40] Imran M, Almeida F, Raik J, et al. Design space exploration of saber in 65 nm ASIC[C]//Proceedings of the 5th Workshop on Attacks and Solutions in Hardware Security. New York: Association for Computing Machinery, 2021: 85-90.

[41] Xing Y F, Li S G. A compact hardware implementation of CCA-secure key exchange mechanism CRYSTALS-KYBER on FPGA[J]. IACR Transactions on Cryptographic Hardware and Embedded Systems, 2021, 2021(2): 328-356.

[42] Xin G Z, Han J, Yin T Y, et al. VPQC: A domain-specific vector processor for post-quantum cryptography based on RISC-V architecture[J]. IEEE transactions on circuits and systems I: Regular papers, 2020, 67(8): 2672-2684.

The development of post-quantum cryptography algorithm and chip design

LIU Dongsheng, LI Aobo, HU Ang, LU Jiahao, HUANG Tianze, YANG Shuo, LI Xiang, ZHAGN Jiaming

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

Abstract Post-quantum cryptography is a new generation of cryptography technology for defending quantum computer attacks. It is regarded as a reliable alternative to traditional cryptography systems, and relevant international standards are gradually emerging. This paper briefly describes the development of post-quantum cryptography, and analyzes the latest development, mathematical principles and characteristics of current algorithm research. On this basis, the analysis is carried out from the three levels of algorithm, hardware architecture, and specific circuit implementation. Then we indicate key technologies that future research needs to overcome, such as efficient hardware implementation, dynamic reconfigurability, side channel attack defense, and secure SoC integration. Moreover, the low-power post-quantum cryptographic chip, the high-performance post-quantum cryptographic chip and core modules such as hashing, random sampling, operation acceleration and logic processing in the chip are described in detail. Finally, we summarize the application status and research value of the current chip implementation in terms of efficient IP design for core circuits, multi-scenario application compatibility, multiple defense mechanisms, and information infrastructure integration, and cover the future development trend of industrialization and diversification. By studying the post-quantum cryptography algorithm and its key technologies, then exploring efficient chip design and implementation methods, it is conducive to promoting the research on the theory and application of public key cryptosystems against quantum attacks, and provides guarantee for China's information security strategy in the quantum era.

Keywords post-quantum cryptography; information security; algorithm principle; efficient implementation; chip design

（责任编辑　王志敏）