周报-向嘉豪 (2025 年 3 月 24 日)

摘要: 本周主要完成了自适应线程分配（ATA）方法的实验验证与扩展应用。通过对 80 种不同 block 和 thread 配置的系统性测试，成功建立了性能优化函数模型，将 ATA 从公钥生成扩展至签名过程。实验结果显示，我们的方法在签名性能上实现了 17.4% 的提升（从 0.0605 降至 0.0493）。同时，完成了包含自适应线程分配和函数级并行两个核心组件的优化架构图设计，为论文框架提供了清晰的视觉表达。

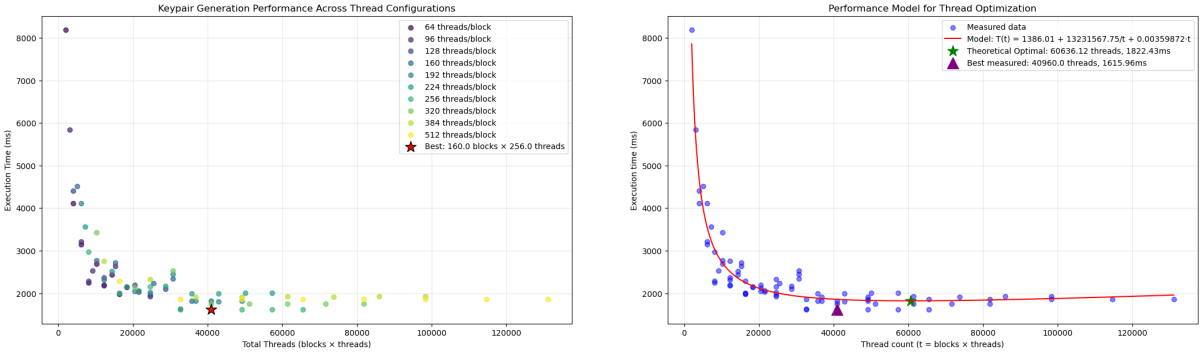
下周计划： 1）开始论文实验章节书写；2）完善创新点 2，函数级并行优化

1 论文实验

本周我们将上周提出的自适应线程函数进行实验验证，其中我们将其从公钥的生成扩展到签名过程，并在此基础上进行性能分析。

1.1 线程数与性能

实验系统性测试了 8 种不同 block 数量（32 至 256）和 10 种不同 thread 数量（64 至 512），共 80 种配置组合对密钥生成和签名操作的影响。如图1a所示。通过对实验数据进行函数拟合，我们得到了图1b中的性能预测曲线。该模型成功预测出最优线程配置，与实际性能测量高度吻合，证实了我们方法的有效性。与 [WDC+25] 的基准实现相比，我们在签名处理时间上从 0.0605 秒降至 0.0493 秒，实现了 17.4% 的性能提升。



(a) 不同 block 和 thread 配置下的性能对比　　(b) 性能函数拟合曲线

图 1: 线程配置性能分析

1.2 thread 和 block 的选择

我们进一步分析了在固定总并行度条件下，block 和 thread 分配比例对性能的影响。图2展示了不同配置组合的性能对比。数据表明，当每个 block 包含 256 个 threads 时，签名和公钥生成操作均达到最佳性能。这一发现具有显著实用价值，因为它表明在我们的 SPHINCS+ 实现中存在一个明确的线程组织最优点，该配置能够最有效地平衡线程管理开销与并行计算效率。我们推测这与 GPU 内存层次结构和 warp 执行模式的交互有关，将在论文中进行分析。

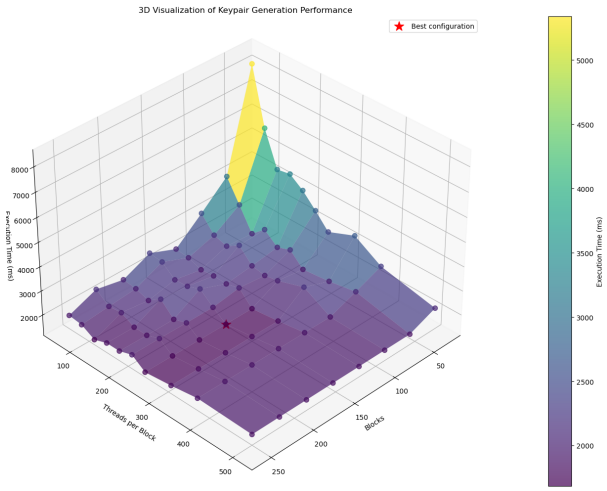


图 2: 不同 block 和 thread 配置下的性能对比

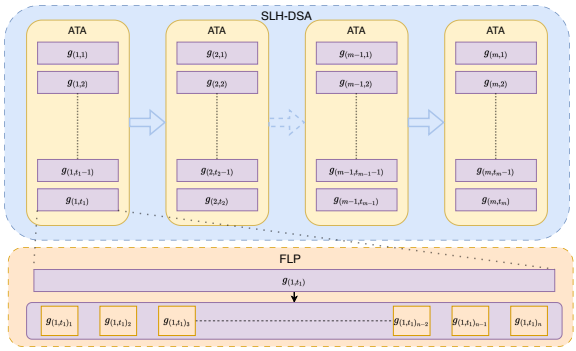


图 3: 优化架构图

2 论文写作

2.1 优化架构图

设计并完成了优化框架的整体架构图（图3）。该图展示了我们提出的双组件优化架构：自适应线程分配（Adaptive Thread Allocation, ATA）和函数级并行（Function-Level Parallelism, FLP）。ATA 组件负责动态确定最优的 GPU 线程配置，而 FLP 组件则实现算法中不同函数的并行执行。这种层次化优化策略能够在不同级别上提升计算效率，显著优于单一优化方法。该架构图将作为论文的核心图表之一，有助于清晰传达我们的方法论。

参考文献

[WDC+25] Ziheng Wang, Xiaoshe Dong, Heng Chen, Yan Kang, and Qiang Wang. Cuspx: Efficient gpu implementations of post-quantum signature sphincs⁺. *IEEE Transactions on Computers*, 74(1):15–28, 2025.