Max Planck Institute for Security and Privacy, National Taiwan University, and Academia Sinica

Algorithmic Views of Vectorized Polynomial Multipliers – NTRU Prime

Vincent Hwang, Chi-Ting Liu, and Bo-Yin Yang

March 6, 2024

Goals

- Polynomial multiplications in NTRU Prime (parameter set `ntrulpr761/sntrup761`)

$$\frac{\mathbb{Z}_{4591}[x]}{\langle x^{761} - x - 1 \rangle}.$$

- Compute products in $\mathbb{Z}_{4591}[x]/\langle \boldsymbol{g} \rangle$ with deg $(\boldsymbol{g}) \geq 2 \cdot 761 - 1 = 1521$.
- Vectorization:
  - No high-power-of-two principal roots of unity: $2^k | 4590 \longrightarrow k = 0, 1$.
  - Vectors contain power-of-two number of elements.
  - Good–Thomas + Schönhage + Bruun's FFTs.
  - Rader's + Good–Thomas + Bruun's FFTs.
- $R = \mathbb{Z}_{4591}$ unless stated otherwise.

Cooley–Tukey FFT

Cooley–Tukey FFT

- Commutative ring R with identity, positive integer n.
- Principal n-th root of unity ω_n:
  - Prime q: n must divide $q-1$.
  - $\mathbb{Z}_q[x]/\langle x^n - 1\rangle \cong \prod_i \mathbb{Z}_q[x]/\left\langle x - \omega_n^i \right\rangle$.
- Radix-2 Cooley–Tukey FFT, $n = 2^k$:
  - Principal 2^k-th root of unity $\omega_{2^k} \in R$, equivalently, $\omega_{2^k}^{2^{k-1}} = -1 \in R$.
  - $\frac{R[x]}{\left\langle x^{2^k} - 1 \right\rangle} \cong \prod \frac{R[x]}{\left\langle x^{2^{k-1}} \pm 1 \right\rangle} \cong \prod_{i_0, i_1 = 0,1} \frac{R[x]}{\left\langle x^{2^{k-2}} - \omega_4^{i_0 + 2i_1} \right\rangle} \cong \cdots \cong R^{2^k}$
  - $O(n \log n)$ operations in R.

Vectorization

Armv8-A Single-Instruction-Multiple-Data (SIMD) instruction set.

- 32 vector registers.
  - Each vector registers holds 128-bit of data $\longrightarrow 8$ coefficients in this talk.
- Component-wise arithmetic:
  - Addition/subtraction: $(a_i) + (b_i) = (a_i + b_i)$
  - Various multiplications: $((a_i), (b_i)) \mapsto \left(a_i b_i \bmod 2^{16}\right), \left(\left\lfloor \frac{2a_i b_i}{2^{16}} \right\rceil\right)$, and more.
- Extending, narrowing, permutation.

Vectorizing Radix-2 Cooley–Tukey

- Suppose we have $R[x]\Big/\Big\langle x^{2^k}-1\Big\rangle \cong \prod_i R[x]/\langle x^8-\omega^i\rangle$:
  - Partition a size-2^k polynomial into several size-8 chunks.
  - Vector load/store/add/sub/mul maps nicely to size-8 chunks.
- We don't have such an ω in NTRU Prime ($R = \mathbb{Z}_{4591}$):
  - 4591 is a prime with $4591 = 2 \cdot 3^3 \cdot 5 \cdot 17 + 1$.
  - We only have ω_2 for the radix-2 Cooley–Tukey.

Schönhage’s and Nussbaumer’s FFTs

Schönhage's and Nussbaumer's FFTs

- Motivation: We don't have roots of unity for (high-dimensional) radix-2 Cooley–Tukey.
- Solution: Craft one by extending.
- How Schönhage works, $R[x]/\left\langle x^{2048}-1\right\rangle$:
  - $\frac{R[x]}{\left\langle x^{2048}-1\right\rangle} \cong \frac{\left(R[x]/\left\langle x^{32}-y\right\rangle\right)[y]}{\left\langle y^{64}-1\right\rangle} \hookrightarrow \frac{\left(R[x]/\left\langle x^{64}+1\right\rangle\right)[y]}{\left\langle y^{64}-1\right\rangle} \cong \prod_i \frac{\left(R[x]/\left\langle x^{64}+1\right\rangle\right)[y]}{\left\langle y-x^{2i}\right\rangle} \cong \left(\frac{R[x]}{\left\langle x^{64}+1\right\rangle}\right)^{64}$
  - Roughly square-root decrease of problem size.
  - Doubly many coeffs.
  - Truncated Schönhage $\frac{R[x]}{\left\langle\left(x^{1024}+1\right)\left(x^{512}-1\right)\right\rangle} \hookrightarrow \left(\frac{R[x]}{\left\langle x^{64}+1\right\rangle}\right)^{48}$ follows similarly.
- Nussbaumer works similarly but only for negacyclic.
- Easily vectorizable.

Prior Vectorization

[BBCT22]:

$$\frac{R[x]}{\langle (x^{1024}+1)(x^{512}-1)\rangle} \overset{\text{Schönhage}}{\hookrightarrow} \left(\frac{R[x]}{\langle x^{64}+1\rangle}\right)^{48} \overset{\text{Nussbaumer}}{\hookrightarrow} \left(\frac{R[z]}{\langle z^{8}+1\rangle}\right)^{48\cdot 16=768}.$$

1. Schönhage: $1 \times 1536 \to 48 \times 64$.
2. Nussbaumer: $48 \times 64 \to 768 \times 8$.

Removing Nussbaumer

Overview

1. Replace truncated Schönhage by Good–Thomas + Schönhage.
2. Replace Nussbaumer by Bruun.

Replacing Truncated Schönhage

- What we know: radix-2 Schönhage introduces radix-2 roots of unity.
- Question: What if there is already a principal 3rd root of unity?
  - Cooley–Tukey: $\frac{R[x]}{\left\langle x^{3\cdot 2^k}-1\right\rangle} \cong \prod_{i,j} \frac{R[x]}{\left\langle x-\omega_3^i\omega_{2^k}^j\right\rangle}$, twiddles $\omega_3^i\omega_{2^k}^j$.
  - Good–Thomas: $\frac{R[x]}{\left\langle x^{3\cdot 2^k}-1\right\rangle} \overset{x\mapsto yz}{\cong} \frac{R[y,z]}{\left\langle z^3-1, y^{2^k}-1\right\rangle} \cong \prod_{i,j} \frac{R[y,z]}{\left\langle z-\omega_3^i, y-\omega_{2^k}^j\right\rangle}$, twiddles ω_3^i, $\omega_{2^k}^j$.
- Our approach, $R[x]/\left\langle x^{1536}-1\right\rangle$:
  1. Introduce radix-2 roots of unity by Schönhage.
  2. Apply Good–Thomas separating the roots into ω_3^i and $\omega_{2^k}^j$.
  3. Apply radix-3 FFT with **multiplications** since $\exists \omega_3 \in R$.
  4. Apply radix-2 FFT with **data shuffling** where ω_{2^k} is crafted by Schönhage.
  5. $\left(R[x]/\left\langle x^{32}+1\right\rangle\right)^{96}$.
- Benefits:
  - Good–Thomas over Cooley–Tukey: avoid $\omega_3^i\omega_{2^k}^j$ requiring multiplications and shuffling at the same time.
  - Good–Thomas + Schönhage over truncated Schönhage: subproblems have smaller size, we have size-32 instead of size-64.

Replacing Nussbaumer

- What we know: radix-2 Nussbaumer splits $R[x]/\langle x^{32}+1\rangle$ by extending.
  - $R[x]/\langle x^{32}+1\rangle \hookrightarrow \left(R[x]/\langle x^{8}+1\rangle\right)^{8}$.
- Question: What if $x^{32}+1$ factors over R?
- For $q = 4591$, $x^{32}+1$ factors into trinomials of the form $x^4+\gamma x^2-1$ over $\mathbb{Z}_q$.
- Our approach, $R[x]/\langle x^{32}+1\rangle$:
  1. Split $R[x]/\langle x^{32}+1\rangle$ into $\prod_{i=0,\ldots,3} R[x]/\langle x^8+\alpha_i x^4+1\rangle$ (size-8 instead of size-4).
  2. $\prod_{i=0,\ldots,3} R[x]/\langle x^8+\alpha_i x^4+1\rangle$.
- Benefits:
  - Bruun over Nussbaumer: We have 4 size-8 polymuls instead of 8.
- History: [Bru78] introduced the trinomial factorization over $\mathbb{C}$, [BGM93] introduced the finite field case when $q \equiv 3 \bmod 4$.

What We Have Now

- [BBCT22]:

$$\frac{R[x]}{\langle (x^{1024}+1)(x^{512}-1)\rangle} \overset{\text{Schönhage}}{\hookrightarrow} \left(\frac{R[x]}{\langle x^{64}+1\rangle}\right)^{48} \overset{\text{Nussbaumer}}{\hookrightarrow} \left(\frac{R[z]}{\langle z^{8}+1\rangle}\right)^{768}.$$

- `Good-Schönhage-Bruun`:

$$\frac{R[x]}{\langle x^{1536}-1\rangle} \overset{\text{Good–Thomas + Schönhage}}{\hookrightarrow} \left(\frac{R[x]}{\langle x^{32}+1\rangle}\right)^{96} \overset{\text{Bruun}}{\cong} \left(\prod_{i=0,\dots,3}\frac{R[x]}{\langle x^{8}+\alpha_i x^{4}+1\rangle}\right)^{96}.$$

Removing Schönhage

Overview

1. Replace Schönhage by Rader.
2. Generalize Bruun (omitted).

Rader's FFT

For a prime p, $R[x]/\langle x^p - 1\rangle \cong \prod_i R[x]/\langle x - \omega_p^i\rangle$ can be implemented with the aid of a size-$(p-1)$ cyclic convolution. Consider

$$\begin{pmatrix}\hat{a}_0\\ \hat{a}_1\\ \hat{a}_2\\ \hat{a}_3\\ \hat{a}_4\end{pmatrix} = \begin{pmatrix}1 & 1 & 1 & 1 & 1\\ 1 & \omega_5 & \omega_5^2 & \omega_5^3 & \omega_5^4\\ 1 & \omega_5^2 & \omega_5^4 & \omega_5 & \omega_5^3\\ 1 & \omega_5^3 & \omega_5 & \omega_5^4 & \omega_5^2\\ 1 & \omega_5^4 & \omega_5^3 & \omega_5^2 & \omega_5\end{pmatrix} \begin{pmatrix}a_0\\ a_1\\ a_2\\ a_3\\ a_4\end{pmatrix}.$$

We have:

$$\begin{pmatrix}\hat{a}_2 - a_0\\ \hat{a}_4 - a_0\\ \hat{a}_3 - a_0\\ \hat{a}_1 - a_0\end{pmatrix} = \begin{pmatrix}\omega_5 & \omega_5^2 & \omega_5^4 & \omega_5^3\\ \omega_5^3 & \omega_5 & \omega_5^2 & \omega_5^4\\ \omega_5^4 & \omega_5^3 & \omega_5 & \omega_5^2\\ \omega_5^2 & \omega_5^4 & \omega_5^3 & \omega_5\end{pmatrix} \begin{pmatrix}a_3\\ a_4\\ a_2\\ a_1\end{pmatrix},$$

a size-4 cyclic convolution of $(\omega_5, \omega_5^3, \omega_5^4, \omega_5^2)$ and (a_3, a_4, a_2, a_1).

Replacing Schönhage with Rader

- $4591 = 2 \cdot 3^3 \cdot 17 + 1 \longrightarrow \exists \omega_{17}, \omega_3, \omega_2$.
- $R[x]/\langle x^{17m} - 1 \rangle \cong \prod_i R[x]/\langle x^m - \omega_{17}^i \rangle$ via Rader.
- Good–Thomas + Schönhage reduce the problem size by $\frac{1536}{32} = 48$ times.
- Choose $\omega_{51} = \omega_3 \omega_{17}$ for a factor of 51 problem size reduction.
- Extend it to a size-102 transform with $\omega_{102} = \omega_2 \omega_3 \omega_{17}$.
- Our approach:
  1. Apply Good–Thomas so $\frac{R[x]}{\langle x^{1632} - 1 \rangle} \cong \frac{R[u,w,v]}{\langle x^{16} - uvw, u^{17} - 1, w^3 - 1, v^2 - 1 \rangle}$.
  2. Apply Rader to size-17 transformation.
  3. Apply size-3 and size-2 transformations straightforwardly.
  4. We have $\prod_i \frac{R[x]}{\langle x^{16} \pm \omega_{102}^{2i} \rangle}$.
  5. Apply Cooley–Tukey to 48 instances of the form $\frac{R[x]}{\langle x^{16} - \omega_{102}^{2i} \rangle}$.
  6. Apply Bruun to 48 instances of the form $\frac{R[x]}{\langle x^{16} + \omega_{102}^{2i} \rangle}$.
  7. We have 192 size-8 polymuls. and 6 size-16 polymuls.

What We Have Now

- [BBCT22]:

$$\frac{R[x]}{\langle (x^{1024}+1)(x^{512}-1)\rangle} \overset{\text{Schönhage}}{\hookrightarrow} \left(\frac{R[x]}{\langle x^{64}+1\rangle}\right)^{48} \overset{\text{Nussbaumer}}{\hookrightarrow} \left(\frac{R[z]}{\langle z^{8}+1\rangle}\right)^{768}.$$

- `Good-Schönhage-Bruun`:

$$\frac{R[x]}{\langle x^{1536}-1\rangle} \overset{\text{Good–Thomas + Schönhage}}{\hookrightarrow} \left(\frac{R[x]}{\langle x^{32}+1\rangle}\right)^{96} \overset{\text{Bruun}}{\cong} \left(\prod_{i=0,\dots,3}\frac{R[x]}{\langle x^{8}+\alpha_i x^4+1\rangle}\right)^{96}.$$

- `Good-Rader-Bruun`:

$$\frac{R[x]}{\langle x^{1632}-1\rangle} \overset{\text{Good–Thomas + Rader}}{\cong} \prod_i \frac{R[x]}{\langle x^{16}\pm\omega_{102}^{2i}\rangle} \overset{\text{Cooley–Tukey + Bruun}}{\cong} \text{192 size-8 + 6 size-16.}$$

Results

Polynomial Multiplications

Table: Overview of polynomial multiplications in `ntrulpr761/sntrup761` with blow-up factors. Blow-up factor: $\frac{\#\text{coeff. after transformation}}{\#\text{coeff. before transformation}}$.

Armv8-A Neon		x86 AVX2	
Implementation	Cycles	Implementation	Cycles
Big-by-small polynomial multiplications			
`Good-Thomas` $(1\times)$	47 696	[BBCT22] $(1\times)$	16 992
[Haa21] $(1\times)$	242 585		
Big-by-big polynomial multiplications			
`Good-Rader-Bruun` $(1\times)$	39 788	[BBCT22] $(4\times)$	25 113
`Good-Schönhage-Bruun` $(2\times)$	50 398		

- Similar transformations, but not covered in this talk (see paper for more details).
- Transformations we just went through.
- Reducing $\#$ small-dimensional transform is effective.

Scheme

sntrup761			
Operation	Key generation	Encapsulation	Decapsulation
Ref	273 598 470	29 750 035	89 968 342
`Good-Rader-Bruun`	6 333 403	147 977	158 233
`Good-Thomas`	6 340 758	153 465	182 271
`Good-Schönhage-Bruun`	6 345 787	163 305	193 626
ntrulpr761			
Operation	Key generation	Encapsulation	Decapsulation
Ref	29 853 635	59 572 637	89 185 030
[Haa21]	775 472	1 150 294	1 417 394
`Good-Rader-Bruun`	260 606	412 629	461 250
`Good-Thomas`	269 590	422 102	471 014
`Good-Schönhage-Bruun`	272 738	436 965	499 559

Follow up Work

[Hwa23] gave a systematic study of vectorization:

- $R[x]/\left\langle \Phi_{17}\left(x^{96}\right)\right\rangle$.
- $1.29 \sim 1.36$ times faster compared to `Good-Rader-Bruun` with Neon.
- $1.99 \sim 2.16$ times faster compared to [BBCT22] with AVX2.

Thanks for listening

Paper (IACR ePrint): `https://eprint.iacr.org/2023/1580`

Artifact: `https://github.com/vector-polymul-ntru-ntrup/NTRU_Prime`

Slides: TBA

[BBCT22] Daniel J. Bernstein, Billy Bob Brumley, Ming-Shing Chen, and Nicola Tuveri, *OpenSSLNTRU: Faster post-quantum TLS key exchange*, 31st USENIX Security Symposium (USENIX Security 22), 2022, https://www.usenix.org/conference/usenixsecurity22/presentation/bernstein, pp. 845–862.

[BGM93] Ian F. Blake, Shuhong Gao, and Ronald C. Mullin, *Explicit Factorization of* $x^{2^k}+1$ *over* $\mathbb{F}_p$ *with Prime* $p \equiv 3 \bmod 4$, Applicable Algebra in Engineering, Communication and Computing **4** (1993), no. 2, 89–94, https://link.springer.com/article/10.1007/BF01386832.

[Bru78] Georg Bruun, *z-transform DFT Filters and FFT's*, IEEE Transactions on Acoustics, Speech, and Signal Processing **26** (1978), no. 1, 56–63, https://ieeexplore.ieee.org/document/1163036.

Reference II

[Haa21] Jasper Haasdijk, *Optimizing NTRU LPRime on the ARM Cortex - A72*, 2021, `https://github.com/jhaasdijk/KEMobi`.

[Hwa23] Vincent Hwang, *Pushing the Limit of Vectorized Polynomial Multiplication for NTRU Prime*, `https://eprint.iacr.org/2023/604`.