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`)

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

- Compute products in $\mathbb{Z}_{4591}[x]/\langle \boldsymbol{g} \rangle$ with deg $(\boldsymbol{g}) \geq 2 \cdot 761 - 1 = 1521$.
- Vectorization:
  - Vectors contain power-of-two number of elements.
  - High-dimensional power-of-two-multiple transformation.
  - Approaches in this talk:
    1. Good–Thomas + Schönhage + Bruun's FFTs.
    2. Rader's + Good–Thomas + Bruun's FFTs.
- $R = \mathbb{Z}_{4591}$ unless stated otherwise.

Vectorization

Vectorization

Armv8-A Neon 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: $((a_i), (b_i)) \mapsto \left(a_i b_i \bmod 2^{16}\right), \left(\left\lfloor \frac{2 a_i b_i}{2^{16}} \right\rceil\right)$, and more.
- Extending, narrowing, permutation.

Cooley–Tukey FFT

Cooley–Tukey FFT

- Principal n-th root of unity ω_n:
  - $R = \mathbb{Z}_q$, prime q: n must divide $q-1$.
- Radix-2, $n = 2^k$:

$$
\begin{tikzcd}[column sep=tiny]
 & & & \frac{R[x]}{\langle x^{2^k}-1 \rangle} \arrow[dll] \arrow[drr] & & & \\
 & \frac{R[x]}{\langle x^{2^{k-1}}-1 \rangle} \arrow[dl] \arrow[dr] & & & & \frac{R[x]}{\langle x^{2^{k-1}}+1 \rangle} \arrow[dl] \arrow[dr] & \\
\frac{R[x]}{\langle x^{2^{k-2}}-1 \rangle} & & \frac{R[x]}{\langle x^{2^{k-2}}+1 \rangle} & & \frac{R[x]}{\langle x^{2^{k-2}}-\omega_4 \rangle} & & \frac{R[x]}{\langle x^{2^{k-2}}+\omega_4 \rangle}
\end{tikzcd}
$$

- Unfortunately, we don't have ω_{2^k} with a high-power 2^k in $\mathbb{Z}_{4591}$:
  - $4591 - 1 = 2 \cdot 3^3 \cdot 5 \cdot 17 \longrightarrow k = 0, 1.$

Schönhage’s and Nussbaumer’s FFTs

Schönhage's and Nussbaumer's FFTs

- Schönhage

$$
\begin{tikzcd}
 & & \frac{R[x]}{\langle (x^{1024}+1)(x^{512}-1) \rangle} \arrow[dll] \arrow[dl] \arrow[d] \arrow[dr] \arrow[drr] & & \\
\frac{R[x]}{\langle x^{32}-y \rangle} \arrow[d] & \cdots \arrow[d] & \cdots \arrow[d] & \cdots \arrow[d] & \frac{R[x]}{\langle x^{32}-y \rangle} \arrow[d] \\
\frac{R[x]}{\langle x^{64}+1 \rangle} & \cdots & \cdots & \cdots & \frac{R[x]}{\langle x^{64}+1 \rangle}
\end{tikzcd}
$$

1. Chopping: $x^{32} \mapsto y$.
2. Extending: replace $x^{32} - y$ by $x^{64} + 1$ (zero-padding).
3. Transform with x as the root of unity.
4. $\left(R[x]/\langle x^{64} + 1 \rangle \right)^{48}$.

- Nussbaumer works similarly.

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: $1 \times 1536 = 1536 \to 48 \times 64 = 3072$.
2. Nussbaumer: $48 \times 64 = 3072 \to 768 \times 8 = 6144$.

Removing Nussbaumer

Overview

1. Apply Good–Thomas.
2. Replace Nussbaumer by Bruun.

Applying Good–Thomas

- What we know: radix-2 Schönhage introduces radix-2 roots of unity.
- Question: What if there is already a principal 3rd root of unity?
- Apply Good–Thomas first (simplified):

$$\begin{tikzcd}
& \frac{R[x]}{\langle x^{1536}-1\rangle} \arrow[dl] \arrow[d] \arrow[dr] & \\
\frac{R[x]}{\langle x^{512}-1\rangle} \arrow[d] & \frac{R[x]}{\left\langle x^{512}-\omega_3^2\right\rangle} \arrow[d] & \frac{R[x]}{\langle x^{512}-\omega_3\rangle} \arrow[d] \\
\left(\frac{R[x]}{\langle x^{32}+1\rangle}\right)^{32} & \left(\frac{R[x]}{\langle x^{32}+1\rangle}\right)^{32} & \left(\frac{R[x]}{\langle x^{32}+1\rangle}\right)^{32}
\end{tikzcd}$$

- Good–Thomas.
- Schönhage.
- Size-32 polymuls instead of size-64.

Replacing Nussbaumer by Bruun

- 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 irreducible trinomials of the form $x^4+\gamma x^2-1$ over $\mathbb{Z}_q$.

$$
\begin{tikzcd}[column sep=small]
 & & & \frac{R[x]}{\langle x^{32}+1\rangle} \arrow[dlll] \arrow[dl] \arrow[dr] \arrow[drrr] & & & \\
\frac{R[x]}{\langle x^{8}+\alpha_0 x^4+1\rangle} & & \frac{R[x]}{\langle x^{8}+\alpha_1 x^4+1\rangle} & & \frac{R[x]}{\langle x^{8}+\alpha_2 x^4+1\rangle} & & \frac{R[x]}{\langle x^{8}+\alpha_3 x^4+1\rangle}
\end{tikzcd}
$$

- Four size-8 polymuls instead of eight.

What We Have Now

- [BBCT22]:

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

- `Good-Schönhage-Bruun`:

$$\frac{R[x]}{\left\langle x^{1536}-1\right\rangle} \overset{\text{Good–Thomas + Schönhage}}{\hookrightarrow} \left(\frac{R[x]}{\left\langle x^{32}+1\right\rangle}\right)^{96} \overset{\text{Bruun}}{\cong} \text{384 size-8.}$$

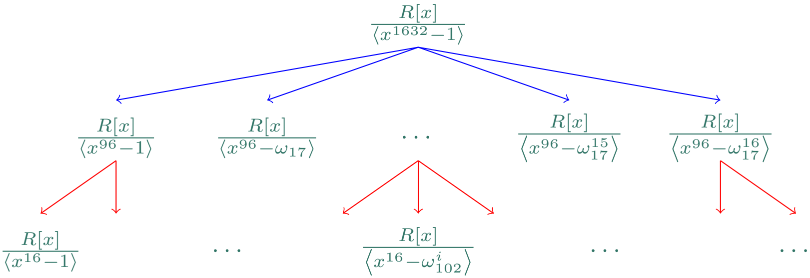
Removing Schönhage

Overview

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

Replacing Schönhage with Rader

- $4591 - 1 = 2 \cdot 3^3 \cdot 5 \cdot 17 \longrightarrow \exists \omega_{17}, \omega_3, \omega_2$.
- Size-17 transformation via Rader (size-16 cyclic convolution).
- Our approach



- Rader, Good–Thomas.
- Good–Thomas.
- Cooley–Tukey + Bruun for most of the size-16 (omitted).

What We Have Now

- [BBCT22]:

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

- `Good-Schönhage-Bruun`:

$$\frac{R[x]}{\langle x^{1536}-1\rangle} \overset{\text{Good–Thomas + Schönhage}}{\hookrightarrow} \left(\frac{R[x]}{\langle x^{32}+1\rangle}\right)^{96} \overset{\text{Bruun}}{\cong} \text{384 size-8.}$$

- `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: Polymuls. with blow-up factors. Blow-up factor (BF): $\frac{\#\text{coeff. after transformation}}{\#\text{coeff. before transformation}}$.

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

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

Take Away

- Cryptographers choose structures admitting efficient implementations.
  - $R[x]\Big/\Big\langle x^{2^k}+1\Big\rangle$ with $\omega_{2^{k+1}} \in R$ for Cooley–Tukey.
- Vectorization when there is no ω_{2^k}:
  - Prior [BBCT22]: radix-2 Schönhage and Nussbaumer.
  - This work: Rader, Good–Thomas, and Bruun.
- Many more choices of polynomial rings with efficient implementations other than Cooley–Tukey.

Thanks for listening

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

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

Slides: `https://vincentvbh.github.io/slides/ACNS2024_1_21_slide.pdf`

Reference I

[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.

[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.

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).