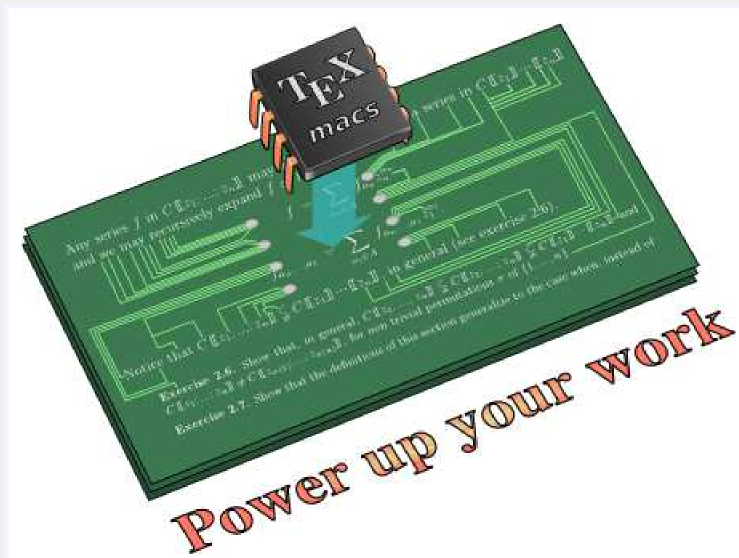
Fast integer multiplication

David Harvey, **Joris van der Hoeven**, Grégoire Lecerf

CNRS, École polytechnique



Bordeaux, February 2, 2015

http://www.TeXmacs.org

$\mathsf{I}(n)$: multiplication of two n-digit integers

$\mathsf{M}_{\mathbb{K}}(n)$: multiplication of two polynomials in $\mathbb{K}[x]$ of degree $<n$

r^ω: multiplication of two $r \times r$ matrices

More involved operations

Division in $\mathbb{Z}$	$O(\mathsf{I}(n))$
GCD in $\mathbb{Z}$	$O(\mathsf{I}(n) \log n)$
Division in $\mathbb{K}[X]$	$O(\mathsf{M}_{\mathbb{K}}(n))$
GCD in $\mathbb{K}[X]$	$O(\mathsf{M}_{\mathbb{K}}(n) \log n)$
Inverting an $r \times r$ matrix	$O(r^\omega)$
Multiplication of $r \times r$ integer matrices ($n \gg r$)	$O(\mathsf{I}(n)\, r^2)$
Multiplication of $r \times r$ integer matrices ($r \gg n$)	$O(n\, r^\omega)$
Roots of polynomial in $\mathbb{C}[X]$, precision $p \ggg n$	$O(\mathsf{I}(n\, p) \log n)$
Exponentiation with n-bit precision	$O(\mathsf{I}(n) \log n)$
Stokes matrix of holonomic function over $\hat{\mathbb{Q}}$	$O(r^2\, \mathsf{I}(n) \log^3 n)$
Etc.	

Turing machines

Turing machines with a finite number of tapes [Papadimitriou 94]

Other bit complexity models

- Operations on $\log n$-bit numbers in time $O(1)$
- Random access machine (RAM)

« Straight Line Programs » (SLPs)

DAGs, non branching programs [Bürgisser–Clausen–Shokrollahi 97]

Other algebraic complexity models

- Turing machines with entries in model-theoretic structures $\mathfrak{S}$ [Friedman 69]
- BSS machines [Blum–Shub–Smale 89]

Integer multiplication, very short history

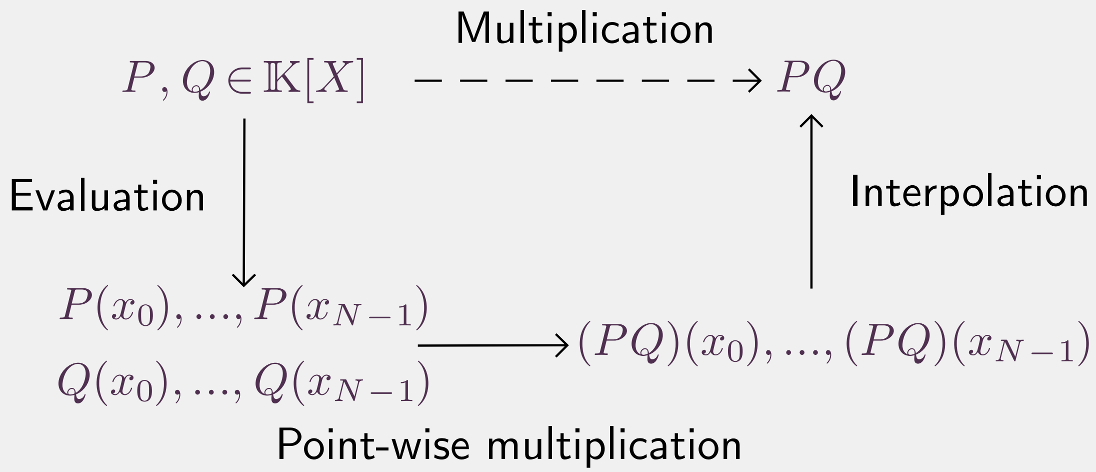
Date	Authors	Complexity
<3000 aJC	Unknown	$O(n^2)$
1962	Karatsuba	$O(n^{\log 3/\log 2})$
1963 (1965)	Toom (Cook)	$O(n\,2^{5\sqrt{\log n/\log 2}})$
1966	Schönhage	$O(n\,2^{\sqrt{2\log n/\log 2}}\,(\log n)^{3/2})$
1969	Knuth	$O(n\,2^{\sqrt{2\log n/\log 2}}\log n)$
1971	Schönhage–Strassen	$O(n\log n\log\log n)$
2007	Fürer	$O\big(n\log n\,2^{O(\log^* n)}\big)$
2014	Harvey–vdH–Lecerf	$O(n\log n\,8^{\log^* n})$

$$\begin{aligned}\log^* x &:= \min\{k\in\mathbb{N}: \log^{\circ k} x \leqslant 1\},\\ \log^{\circ k} &:= \underbrace{\log\circ\cdots\circ\log}_{k\times}.\end{aligned}$$

Kronecker

$$
\begin{aligned}
971362651726262537182735 &= 971362\,X^3 + 651726\,X^2 + 262537\,X + 182735 \\
X &= 1000000
\end{aligned}
$$

Evaluation-interpolation



Definition

$\omega \in \mathbb{K}$ primitive N-th root of unity, with $N \in 2^{\mathbb{N}}$

$$\mathrm{DFT}_{\omega}(P_0, \ldots, P_{N-1}) \;=\; (P(1), P(\omega), P(\omega^2), \ldots, P(\omega^{N-1}))$$

Corresponds to evaluating $P = P_0 + \cdots + P_{N-1} X^{N-1}$ at $x_i = \omega^i$

Inverse transform

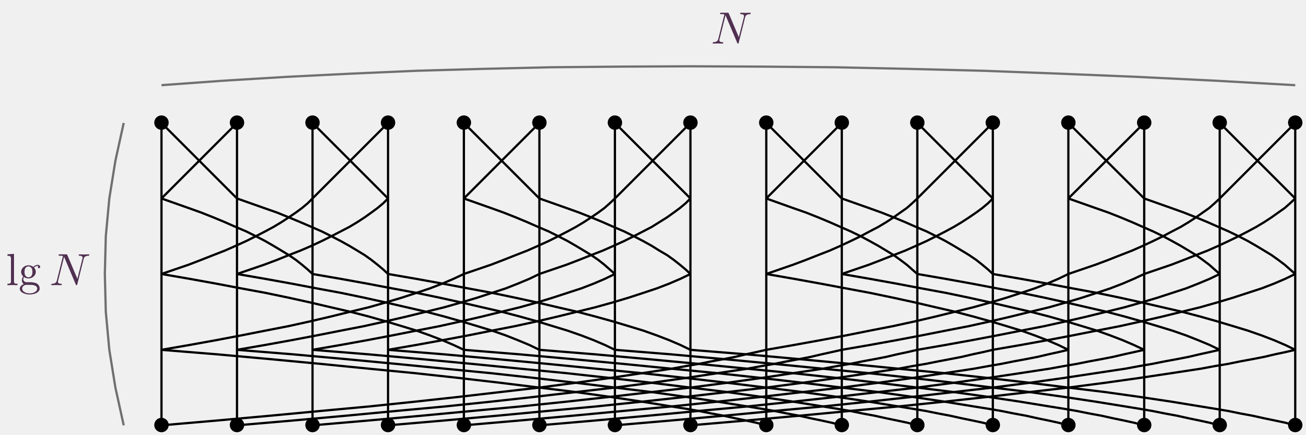
$$\mathrm{DFT}_{\omega}^{-1} \;=\; \frac{1}{N} \mathrm{DFT}_{\omega^{-1}}$$

Interpolation $\rightsquigarrow$ evaluation

Variants

- $\mathbb{K} = \mathbb{C}_b$: **complex DFT**, complex fixed-point arithmetic with b-bit precision
- $\mathbb{K} = \mathbb{F}_p$, **modular DFT**, with p prime number of the form $k\, 2^N \pm 1$ [Pollard 71]
- $\mathbb{K} = \mathbb{L}[Y] / \left(Y^{2^N} \pm 1\right)$, **synthetic DFT**, *à la* Schönhage–Strassen

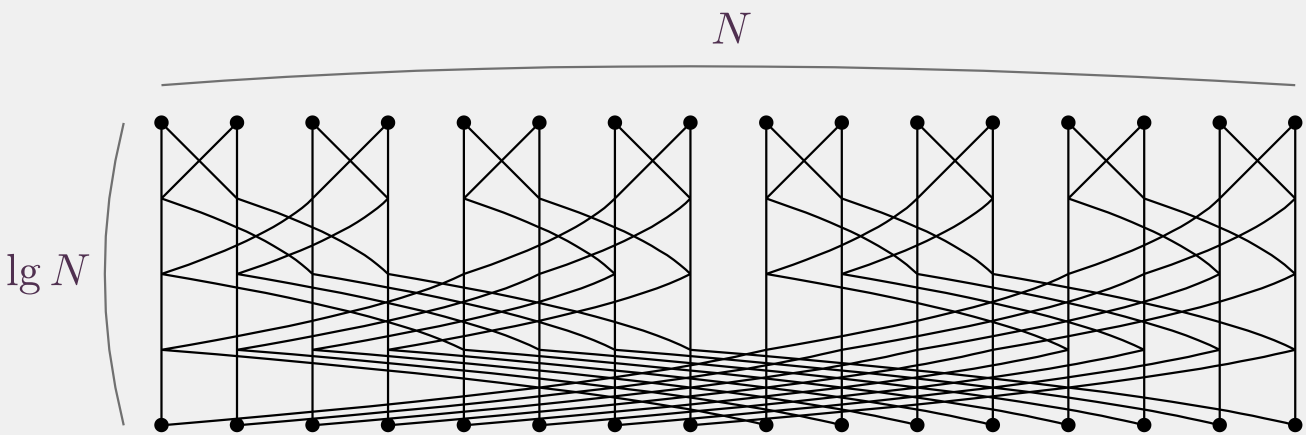
Discrete Fourier Transform (II)



Cost of one DFT : $\frac{1}{2} N \lg N$ "butterflies" $\rightsquigarrow O(N \lg N)$ operations in $\mathbb{K}$

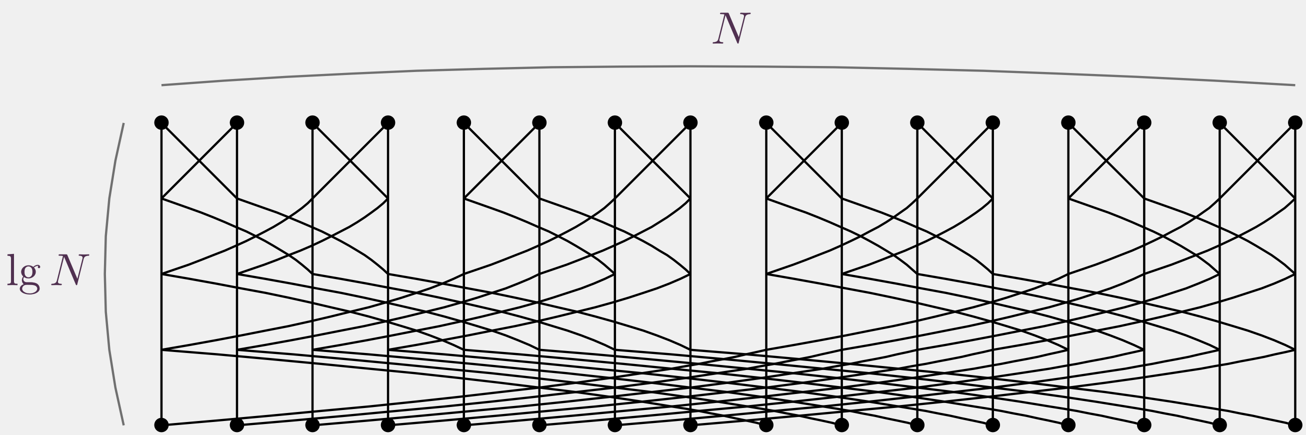
Complex DFT	$N \asymp n/\lg n$	$\mathsf{M}_{\mathbb{K}}(1) = O(\mathsf{I}(\lg n))$	$\mathsf{I}(n) = O(n \lg n\, \mathsf{I}(\lg n) + n \lg n)$
Modular DFT	$N \asymp n/\lg n$	$\mathsf{M}_{\mathbb{K}}(1) = O(\mathsf{I}(\lg n))$	$\mathsf{I}(n) = O(n \lg n\, \mathsf{I}(\lg n) + n \lg n)$
Synthetic DFT	$N \asymp \sqrt{n}$	butterfly $\rightsquigarrow O(\sqrt{n})$	$\mathsf{I}(n) = O(n \lg n \sqrt{n} + \sqrt{n}\, \mathsf{I}(\sqrt{n}))$

Discrete Fourier Transform (II)



Cost of one DFT : $\frac{1}{2} N \lg N$ "butterflies" $\rightsquigarrow O(N \lg N)$ operations in $\mathbb{K}$

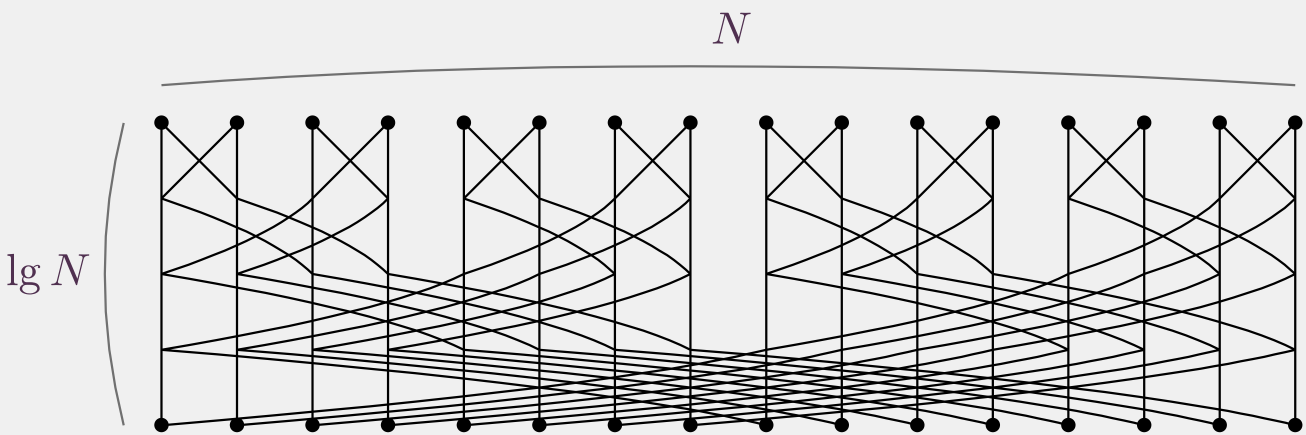
Complex DFT	$N \asymp n / \lg n$	$\mathsf{M}_{\mathbb{K}}(1) = O(\mathsf{I}(\lg n))$	$\mathsf{I}(n) = O(n \lg n\, \mathsf{I}(\lg n) + n \lg n)$
Modular DFT	$N \asymp n / \lg n$	$\mathsf{M}_{\mathbb{K}}(1) = O(\mathsf{I}(\lg n))$	$\mathsf{I}(n) = O(n \lg n\, \mathsf{I}(\lg n) + n \lg n)$
Synthetic DFT	$N \asymp \sqrt{n}$	butterfly $\rightsquigarrow O(\sqrt{n})$	$\mathsf{I}(n) = O(n \lg n \sqrt{n} + \sqrt{n}\, \mathsf{I}(\sqrt{n}))$



Cost of one DFT : $\frac{1}{2} N \lg N$ “butterflies” $\rightsquigarrow O(N \lg N)$ operations in $\mathbb{K}$

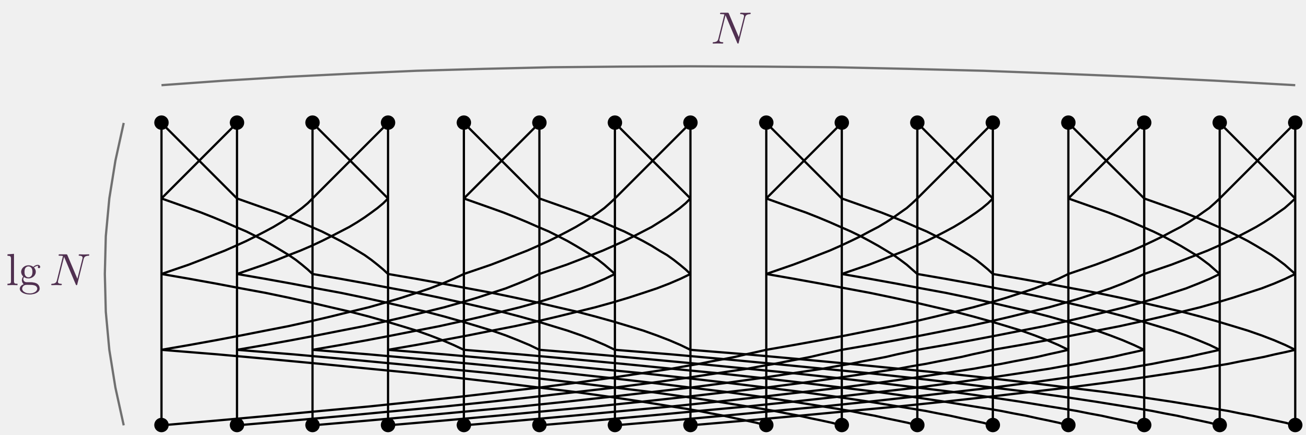
Complex DFT	$N \asymp n/\lg n$	$\mathsf{M}_{\mathbb{K}}(1) = O(\mathsf{I}(\lg n))$	$\mathsf{I}(n) = O(n \lg n\, \mathsf{I}(\lg n) + n \lg n)$
Modular DFT	$N \asymp n/\lg n$	$\mathsf{M}_{\mathbb{K}}(1) = O(\mathsf{I}(\lg n))$	$\mathsf{I}(n) = O(n \lg n\, \mathsf{I}(\lg n) + n \lg n)$
Synthetic DFT	$N \asymp \sqrt{n}$	butterfly $\rightsquigarrow O(\sqrt{n})$	$\mathsf{I}(n) = O(n \lg n \sqrt{n} + \sqrt{n}\, \mathsf{I}(\sqrt{n}))$

Discrete Fourier Transform (II)



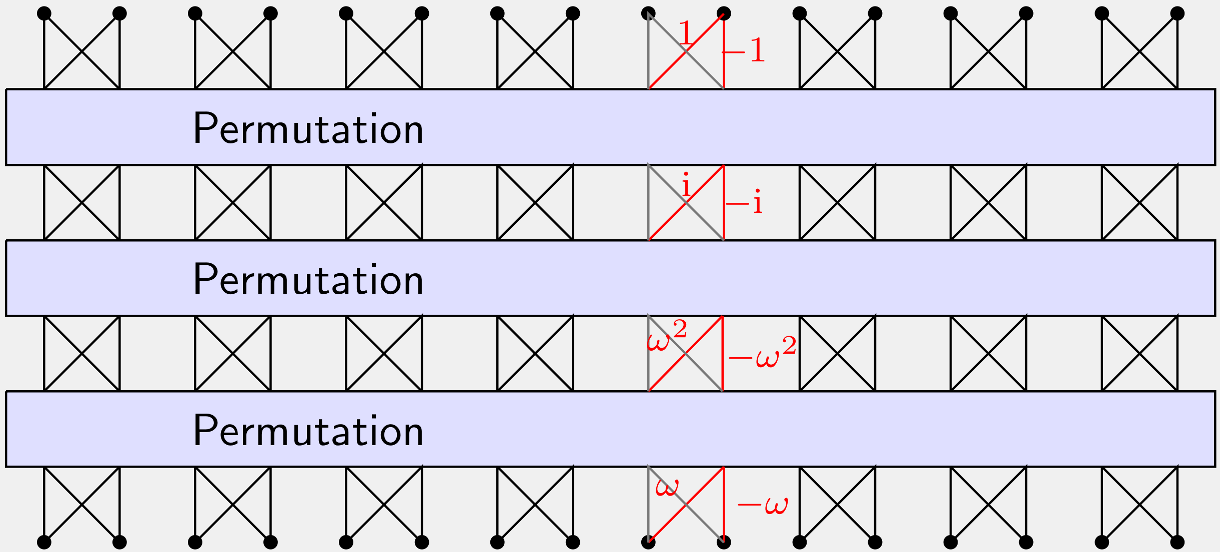
Cost of one DFT : $\frac{1}{2} N \lg N$ "butterflies" $\rightsquigarrow O(N \lg N)$ operations in $\mathbb{K}$

Complex DFT	$N \asymp n / \lg n$	$\mathsf{M}_{\mathbb{K}}(1) = O(\mathsf{I}(\lg n))$	$\mathsf{I}(n) = O(n \lg n \, \mathsf{I}(\lg n) + n \lg n)$
Modular DFT	$N \asymp n / \lg n$	$\mathsf{M}_{\mathbb{K}}(1) = O(\mathsf{I}(\lg n))$	$\mathsf{I}(n) = O(n \lg n \, \mathsf{I}(\lg n) + n \lg n)$
Synthetic DFT	$N \asymp \sqrt{n}$	butterfly $\rightsquigarrow O(\sqrt{n})$	$\mathsf{I}(n) = O(n \lg n \sqrt{n} + \sqrt{n} \, \mathsf{I}(\sqrt{n}))$

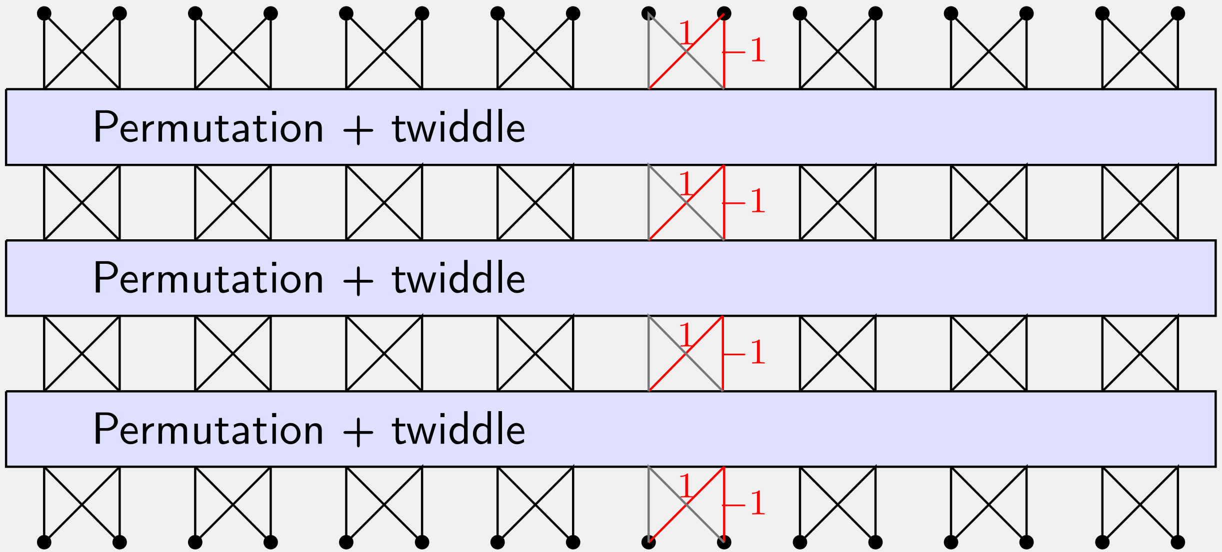


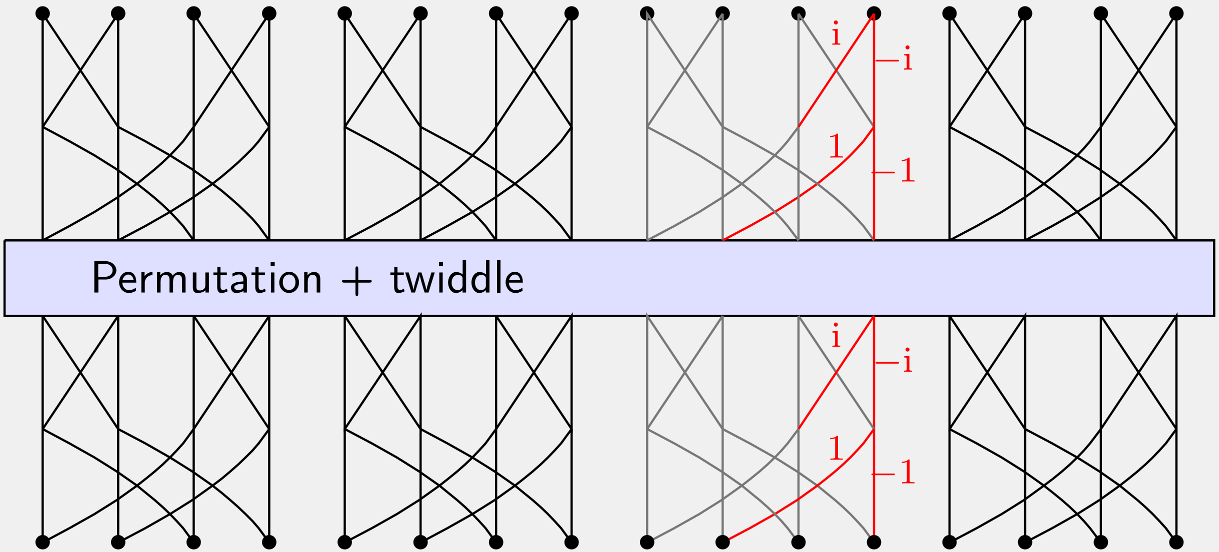
Cost of one DFT : $\frac{1}{2} N \lg N$ “butterflies” $\rightsquigarrow O(N \lg N)$ operations in $\mathbb{K}$

Complex DFT	$N \asymp n / \lg n$	$\mathsf{M}_{\mathbb{K}}(1) = O(\mathsf{I}(\lg n))$	$\mathsf{I}(n) = O(n \lg n \lg\lg n \lg\lg\lg n \cdots)$
Modular DFT	$N \asymp n / \lg n$	$\mathsf{M}_{\mathbb{K}}(1) = O(\mathsf{I}(\lg n))$	$\mathsf{I}(n) = O(n \lg n \lg\lg n \lg\lg\lg n \cdots)$
Synthetic DFT	$N \asymp \sqrt{n}$	butterfly $\rightsquigarrow O(\sqrt{n})$	$\mathsf{I}(n) = O(n \lg n \lg\lg n)$

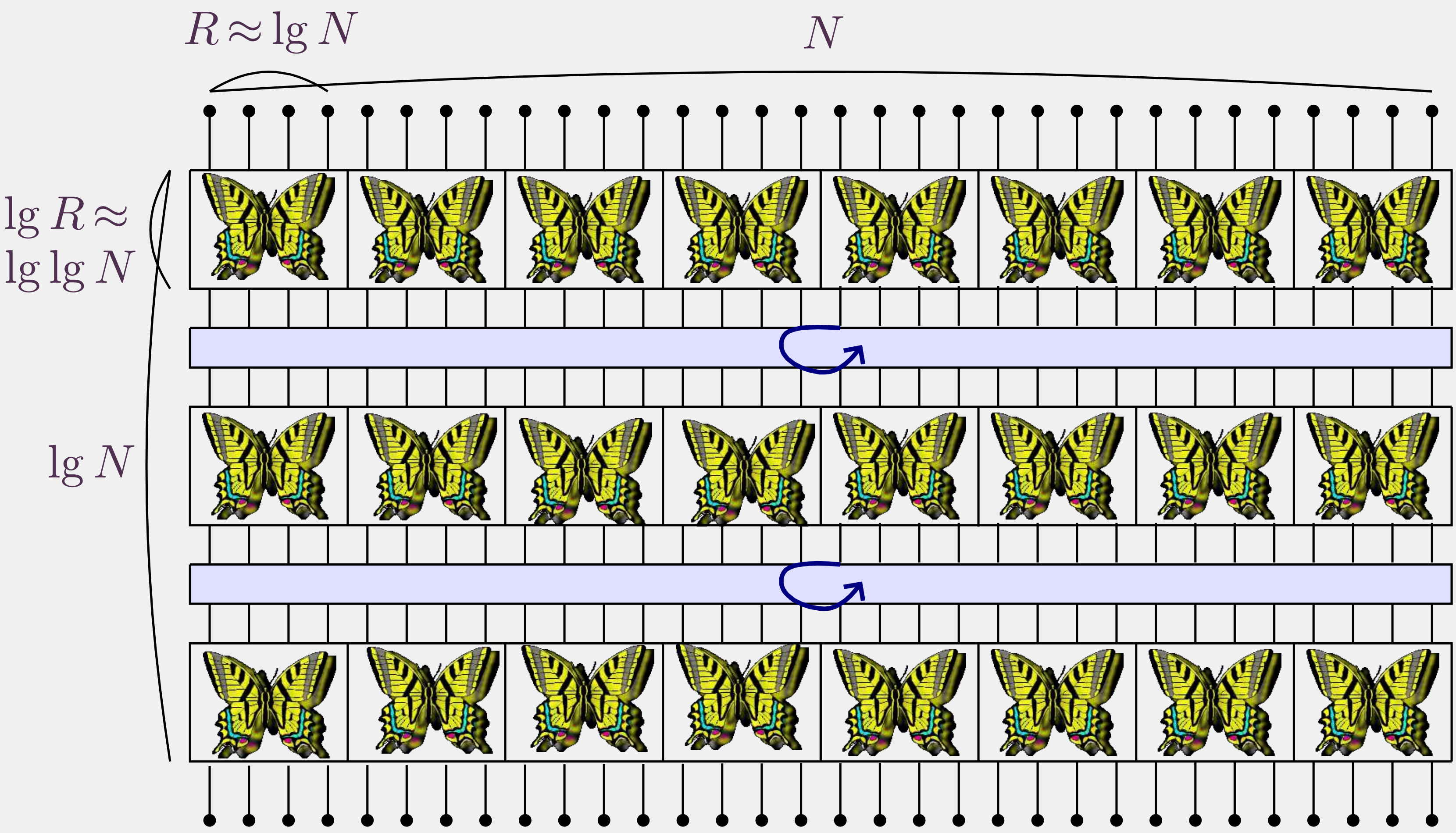

Permutation
Permutation
Permutation
1
−1
i
−i
ω²
−ω²
ω
−ω

Variants of Discrete Fourier Transform




i
−i
1
−1
Permutation + twiddle
i
−i
1
−1

Giant butterflies



Slightly cheating in picture: we should have used 16 butterflies on every line

Fürer's algorithm

Coefficients in $\mathcal{R} = \mathbb{C}_b[X]/(X^{R/2}+1)$

Existence of a "principal" N-th root of unity ω, with $\omega^{2N/R} = X$

Fast giant butterflies (size $R \times \lg R$), but slow *twiddling* $\rightsquigarrow$ multiplications in $\mathcal{R}$

$$\begin{aligned} \mathsf{I}(n) &= O\left(\left(\frac{N}{R}\frac{\lg N}{\lg R}\right)\cdot(b\,R\lg R)\right) + O\left(\left(\frac{N}{R}\frac{\lg N}{\lg R}\right)\cdot \mathsf{M}_{\mathbb{C}_b}(R)\right) \\ \frac{\mathsf{I}(n)}{n\lg n} &= O(1) + O\left(\frac{\mathsf{I}(R\,b)}{(R\,b)\lg(R\,b)}\right) \qquad (R \approx b \approx \lg n) \\ \mathsf{I}(n) &= n\lg n\, 2^{O(\log^* n)} \end{aligned}$$

New algorithm

Ordinary DFT, but accelerate the giant butterflies:

DFT of size $R \times \lg R$ over $\mathbb{C}_b \overset{\text{Bluestein}}{\rightsquigarrow} O(\mathsf{M}_{\mathbb{C}_b[X]}(R)) \overset{\text{Kronecker}}{\rightsquigarrow} O(\mathsf{I}(R\,b))$

Slower giant butterflies, but faster *twiddling*

Ordinary DFT $\Rightarrow$ faster point-wise multiplication (application : integer matrices)

Cyclic convolution $\rightsquigarrow$ DFT

$P, Q \in \mathbb{K}[X]/(X^n - 1)$, $n \in 2^{\mathbb{N}^>}$, primitive n-th root of unity ω

$$((PQ)_0, \ldots, (PQ)_{n-1}) \;=\; \mathrm{DFT}_\omega^{-1}(\mathrm{DFT}_\omega(P_0, \ldots, P_{n-1})\,\mathrm{DFT}_\omega(Q_0, \ldots, Q_{n-1}))$$

DFT $\rightsquigarrow$ Cyclic convolution [Bluestein 70]

Assume $\eta \in \mathbb{K}$ given with $\eta^2 = \omega$.

$$f_i := \eta^{i^2}, \qquad g_i := \eta^{-i^2}$$

$$f_{i+n} = \eta^{(i+n)^2} = \eta^{i^2 + n^2 + 2ni} = \eta^{i^2}\omega^{(\frac{n}{2}+i)n} = f_i, \qquad g_{i+n} = g_i$$

Then $\omega^{ij} = f_i\, f_j\, g_{i-j}$, so for all $a \in \mathbb{K}^n$:

$$\hat{a}_i = \mathrm{DFT}_\omega(a)_i = \sum_{j=0}^{n-1} a_j\,\omega^{ij} = f_i \sum_{j=0}^{n-1} (a_j\, f_j)\, g_{i-j}$$

One recognizes a cyclic convolution

Logarithmically slow function

Function $\Phi\colon [x_0, \infty) \to \mathbb{R}$ such that there exists a $\ell \in \mathbb{N}$ with

$$(\log^{\circ \ell} \circ \Phi \circ \exp^{\circ \ell})(x) \;=\; \log x + O(1) \qquad\qquad (x \to \infty).$$

Examples: $\Phi(x) = \log x$, $\Phi(x) = \log^2 x$, $\Phi(x) = (\log x)^{\log \log x}$, $\Phi(x) = \mathrm{e}^{\mathrm{e}^{2014 \log \log \log x}}$

Iterateurs [see also Écalle 92, Schmeling 01]

$$\begin{aligned} \Phi^*(\Phi(x)) &= \Phi(x) - 1 \\ \Phi^*(x) &= \min \{k \in \mathbb{N} \colon \Phi^{\circ k}(x) \leqslant \sigma\}. \end{aligned}$$

Lemma. *Φ logarithmically slow, Φ^* iterator of Φ. Then*

$$\Phi^*(x) \;=\; \log^* x + O(1)$$

Lemma. *Φ logarithmically slow. Constants K, B, L, ℓ and function T such that*

$$T(x) \;\leqslant\; K \left(1 + \frac{B}{\log^{\circ \ell} x}\right) T(\Phi(x)) + L.$$

Then $T(x) = O(K^{\log^* n})$.

Logarithmically slow function

Function $\Phi: [x_0, \infty) \to \mathbb{R}$ such that there exists a $\ell \in \mathbb{N}$ with

$$(\log^{\circ \ell} \circ \Phi \circ \exp^{\circ \ell})(x) \;=\; \log x + O(1) \qquad (x \to \infty).$$

Examples: $\Phi(x) = \log x$, $\Phi(x) = \log^2 x$, $\Phi(x) = (\log x)^{\log \log x}$, $\Phi(x) = \mathrm{e}^{\mathrm{e}^{2014 \log \log \log x}}$

Iterateurs [see also Écalle 92, Schmeling 01]

$$\begin{aligned} \Phi^*(\Phi(x)) &= \Phi(x) - 1 \\ \Phi^*(x) &= \min \{k \in \mathbb{N}: \Phi^{\circ k}(x) \leqslant \sigma\}. \end{aligned}$$

Lemma. Φ *logarithmically slow,* Φ^* *iterator of* Φ*. Then*

$$\Phi^*(x) \;=\; \log^* x + O(1)$$

Lemma. $\Phi_1, \ldots, \Phi_k$ *logarithmically slow. Constants* K, B, L, ℓ, $c_1 + \cdots + c_k = 1$ *and function* T *such that*

$$T(x) \;\leqslant\; K \left(1 + \frac{B}{\log^{\circ \ell} x}\right) (c_1 T(\Phi_1(x)) + \cdots + c_k T(\Phi_k(x))) + L.$$

Then $T(x) = O(K^{\log^* n})$.

We now have several ways to show that

$$\mathsf{I}(n) \;=\; O(n \lg n \, K^{\log^* n}).$$

What is the best K we can get?

Fürer, after optimisation : $K = 16$ (?)
We, after optimisation : $K = 8$

Ingredients

- Multiplication in $\mathbb{Z} \rightsquigarrow$ multiplication in $(\mathbb{Z}/(2^n - 1)\,\mathbb{Z})[\mathrm{i}]$.
- One argument shared many times in recursive calls $\rightsquigarrow$ 2 DFTs instead of 3.
- Convolution of length N with b-bit coefficients $\rightsquigarrow$ output of size $2\,b + O(\lg N)$.

  Taking $b \asymp (\lg n)^2$ instead of $b \asymp \lg n$ improves the ratio $(2\,b + O(\lg N))\,/\,b$.
- Increase $R \approx \lg N \rightsquigarrow R \approx (\lg N)^{\lg \lg N + O(1)}$.

  Cost Bluestein–Kronecker $\gg$ cost twiddling and other.

Where does the cost come from?

a) Factor $2 \leftrightsquigarrow$ Kronecker segmentation ($\mathbb{Z}[\mathrm{i}] \rightsquigarrow \mathbb{C}_b[X]$, cutting into pieces of $\frac{b}{2}$ bits)
b) Factor $2 \leftrightsquigarrow$ direct and inverse DFT
c) Factor $2 \leftrightsquigarrow$ Kronecker substitution ($\mathbb{C}_b[X]/(X^R - 1) \rightsquigarrow \mathbb{Z}/(2^{2bR} - 1)\,\mathbb{Z}$)

Fermat primes

And *if*, *if*, *if* there were sufficiently many prime numbers of the form $p = 2^{2^k} + 1$
(Optimized) Fürer approach for $\mathbb{K} = \mathbb{F}_p$ yields $K = 4$
Unfortunately..., $p = 2^{16} + 1$ is the largest known prime number of this form

Mersenne primes

Conjecture 1. *Let* $\pi_m(x) = \{p \leqslant x : p = 2^q - 1, p \textit{ prime}, q \textit{ prime}\}$. *Then* $\exists a < b$,

$$a \log \log x < \pi_m(x) < b \log \log x$$

Crandall–Fagin algorithm

Multiplication $\mathbb{F}_p[\mathrm{i}][X]/(X^M - 1) \rightsquigarrow \mathbb{F}_{p'}[\mathrm{i}][X, Y]/(X^M - 1, Y^N - 1)$, $p' \lll p$
Conjecture $1 \Rightarrow K = 4$

Multiplication in $\mathbb{F}_q[X]$ and in $\mathbb{K}[X]$

Kronecker : $\mathsf{M}_{\mathbb{F}_p}(n) = O(\mathsf{I}(n \log p))$ if $\log n = O(\log p)$

Schönhage–Strassen : $\mathsf{M}_{\mathbb{F}_q}(n) = O(n \log n \log\log n\, \mathsf{M}_{\mathbb{F}_q}(1))$ if $\mathrm{char}\, \mathbb{F}_q > 2$

Schönhage : $\mathsf{M}_{\mathbb{F}_q}(n) = O(n \log n \log\log n\, \mathsf{M}_{\mathbb{F}_q}(1))$ for all q

Cantor–Kaltofen : for any $\mathbb{K}$-algebra $\mathbb{A}$, $\mathsf{M}_{\mathbb{A}}^{\mathrm{alg}}(n) = O(n \log n \log\log n)$

Kronecker : $\mathsf{M}_{\mathbb{F}_{p^k}}(n) \asymp \mathsf{M}_{\mathbb{F}_p}(k\, n)$, modulo $O(k\, n \log p)$ operations

Theorem. *We have,* ***uniformly*** *in p:*

$$\mathsf{M}_p(n) \;=\; O\big((n \log p) \log(n \log p)\, 8^{\log^*(n \log p)}\big)$$

Theorem. *Modulo "plausable conjectures", we have,* ***uniformly*** *in p:*

$$\mathsf{M}_p(n) \;=\; O\big((n \log p) \log(n \log p)\, 4^{\log^*(n \log p)}\big)$$

Theorem. *Let $\mathbb{A}$ be an $\mathbb{F}_p$-algebra. Then $\mathsf{M}_{\mathbb{A}}^{\mathrm{alg}}(n) = O(n \lg n\, 8^{\log^* n})$, uniformly in $\mathbb{A}$. Moreover, we only need $O(n\, 4^{\log^* n})$ (non scalar) multiplications in $\mathbb{A}$.*

Proof elements

1. Multiplication in $\mathbb{F}_p[X] \rightsquigarrow$ multiplication in $\mathbb{F}_{p^k}[X]$

2. k such that $\mathbb{F}_{p^k}[X]$ admits an N-th primitive root of unity ω, with N large and smooth

3. Write $N = N_1 \cdots N_r$ with' $N_1, \ldots, N_r$ "under control" and use Bluestein-Kronecker

```
Pari] factor (2^60 - 1)
```

$$\%1 = \begin{pmatrix} 3 & 2 \\ 5 & 2 \\ 7 & 1 \\ 11 & 1 \\ 13 & 1 \\ 31 & 1 \\ 41 & 1 \\ 61 & 1 \\ 151 & 1 \\ 331 & 1 \\ 1321 & 1 \end{pmatrix}$$

```
Pari] factor (7^60 - 1)
```

$$\%2 = \begin{pmatrix} 2 & 5 \\ 3 & 2 \\ 5 & 3 \\ 11 & 1 \\ 13 & 1 \\ 19 & 1 \\ 31 & 1 \\ 43 & 1 \\ 61 & 1 \\ 181 & 1 \\ 191 & 1 \\ 281 & 1 \\ 2801 & 1 \\ 4021 & 1 \\ 159871 & 1 \\ 6568801 & 1 \\ 555915824341 & 1 \end{pmatrix}$$

```
Pari] factor (2^210 - 1)
```

%3 =

$$\begin{pmatrix} 3 & 2 \\ 7 & 2 \\ 11 & 1 \\ 31 & 1 \\ 43 & 1 \\ 71 & 1 \\ 127 & 1 \\ 151 & 1 \\ 211 & 1 \\ 281 & 1 \\ 331 & 1 \\ 337 & 1 \\ 5419 & 1 \\ 29191 & 1 \\ 86171 & 1 \\ 106681 & 1 \\ 122921 & 1 \\ 152041 & 1 \\ 664441 & 1 \\ 1564921 & 1 \end{pmatrix}$$

```
Pari] factor (37^60 - 1)
```

%4 =

$$\begin{pmatrix} 2 & 4 \\ 3 & 3 \\ 5 & 2 \\ 7 & 1 \\ 11 & 1 \\ 13 & 1 \\ 19 & 1 \\ 31 & 1 \\ 41 & 1 \\ 43 & 1 \\ 61 & 1 \\ 67 & 1 \\ 137 & 1 \\ 601 & 1 \\ 2671 & 1 \\ 4021 & 1 \\ 4271 & 1 \\ 144061 & 1 \\ 318211 & 1 \\ 1824841 & 1 \\ 239020081 & 1 \\ 6002229721 & 1 \\ 11507920001 & 1 \\ 51654756031569841 & 1 \end{pmatrix}$$

```
Pari]
```