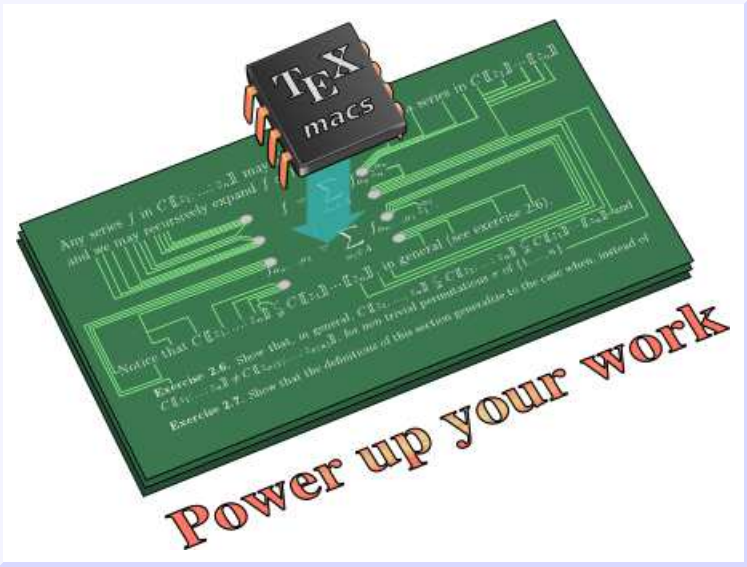
Multiplication rapide d'entiers et de polynômes

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

CNRS, École polytechnique



JNCF, Luminy, 2014

http://www.TEXMACS.org

$\mathsf{I}(n)$: complexité pour multiplier deux entiers de n chiffres

$\mathsf{M}_{\mathbb{K}}(n)$: complexité pour multiplier deux polynômes dans $\mathbb{K}[x]$ de degrés $<n$

r^ω: complexité pour multiplier deux matrices $r \times r$

Opérations plus complexes

Division dans $\mathbb{Z}$	$O(\mathsf{I}(n))$
PGCD dans $\mathbb{Z}$	$O(\mathsf{I}(n) \log n)$
Division dans $\mathbb{K}[X]$	$O(\mathsf{M}_{\mathbb{K}}(n))$
PGCD dans $\mathbb{K}[X]$	$O(\mathsf{M}_{\mathbb{K}}(n) \log n)$
Inverse d'une matrice $r \times r$	$O(r^\omega)$
Multiplication de matrices $r \times r$ d'entiers ($n \gg r$)	$O(\mathsf{I}(n)\, r^2)$
Multiplication de matrices $r \times r$ d'entiers ($r \gg n$)	$O(n\, r^\omega)$
Racine polynôme dans $\mathbb{C}[X]$, précision $p \ggg n$	$O(\mathsf{I}(n\, p) \log n)$
Exponentielle avec une précision de n chiffres	$O(\mathsf{I}(n) \log n)$
Matrices de Stokes d'une fonction holonôme sur $\hat{\mathbb{Q}}$	$O(r^2\, \mathsf{I}(n) \log^3 n)$
Etc.	

Machines de Turing

Machines de Turing avec un nombre fini de bandes [Papadimitriou 94]

Autres modèles pour la complexité binaire

- Opérations sur des nombres de $\log n$ bits pour un coût $O(1)$
- Random access machine (RAM)

« Straight Line Programs » (SLPs)

Graphes acycliques, programmes sans branchements [Bürgisser–Clausen–Shokrollahi 97]

Autres modèles algébriques

- Machines de Turing avec des entrées dans des structures $\mathfrak{S}$ [Friedman 69]
- Machines BSS [Blum–Shub–Smale 89]

Multiplication des entiers, histoire succincte

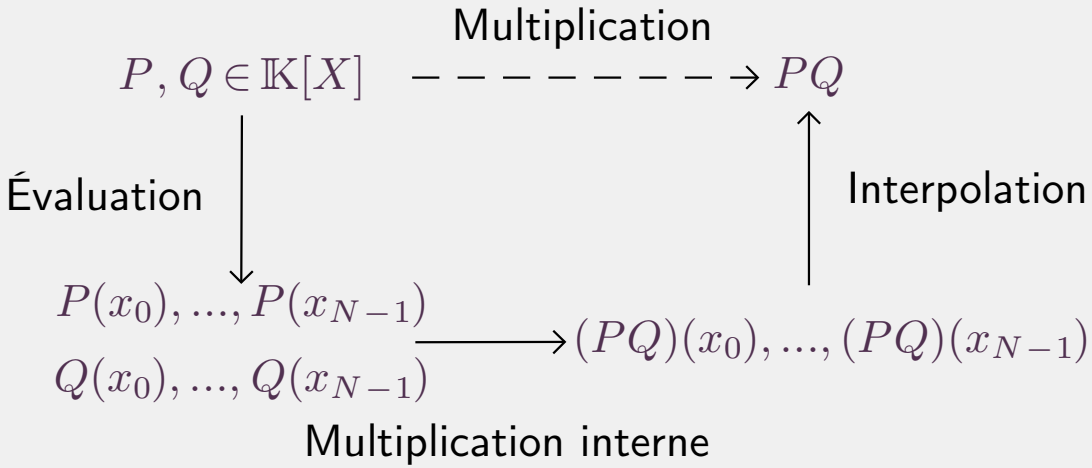
Date	Auteurs	Complexity
<3000 aJC	Unknown	$O(n^2)$
1962	Karatsuba	$O(n^{\log 3/\log 2})$
1963	Toom	$O(n\,2^{5\sqrt{\log n/\log 2}})$
1966	Schönhage	$O(n\,2^{\sqrt{2\log n/\log 2}}\,(\log n)^{3/2})$
1969	Knuth	$O(n\,2^{\sqrt{2\log n/\log 2}}\log n)$
1971	Schönhage–Strassen	$O(n\log n\log\log n)$
2007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

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

Kronecker

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

Évaluation-interpolation



Définition

$\omega \in \mathbb{K}$ racine primitive N-ième d'unité, avec $N \in 2^{\mathbb{N}}$

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

Correspond à évaluer $P = P_0 + \cdots + P_{N-1} X^{N-1}$ en des points $x_i = \omega^i$

Transformation inverse

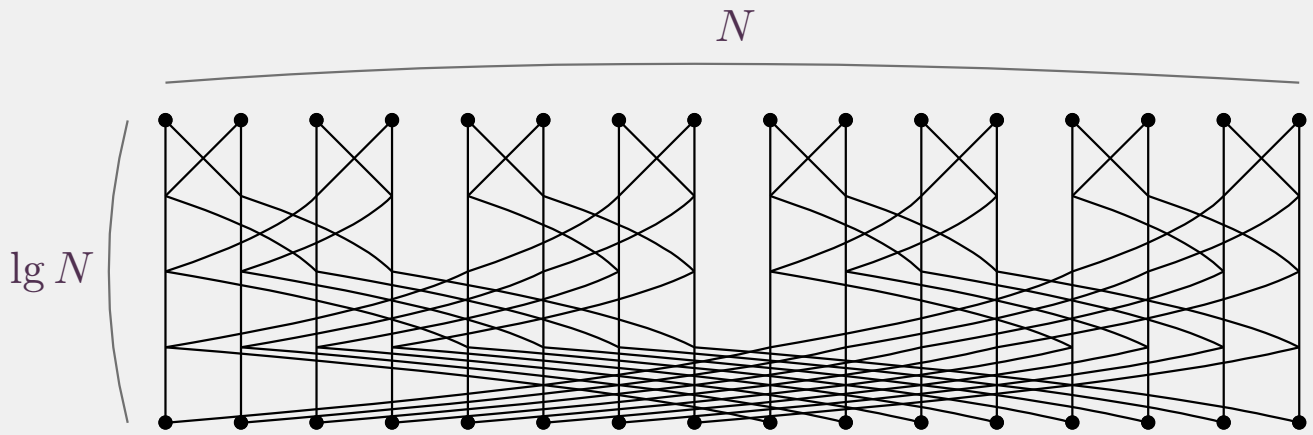
$$\mathrm{DFT}_\omega^{-1} \;=\; \tfrac{1}{N} \mathrm{DFT}_{\omega^{-1}}$$

Interpolation $\rightsquigarrow$ évaluation

Variantes

- $\mathbb{K} = \mathbb{C}_b$: **DFT complexe**, point fixe complexe avec une précision de b bits
- $\mathbb{K} = \mathbb{F}_p$, **DFT modulaire**, avec p un nombre premier de la forme $k\,2^N \pm 1$ [Pollard 71]
- $\mathbb{K} = \mathbb{L}[Y] / \left(Y^{2^N} \pm 1\right)$, **DFT synthétique**, à la Schönhage–Strassen

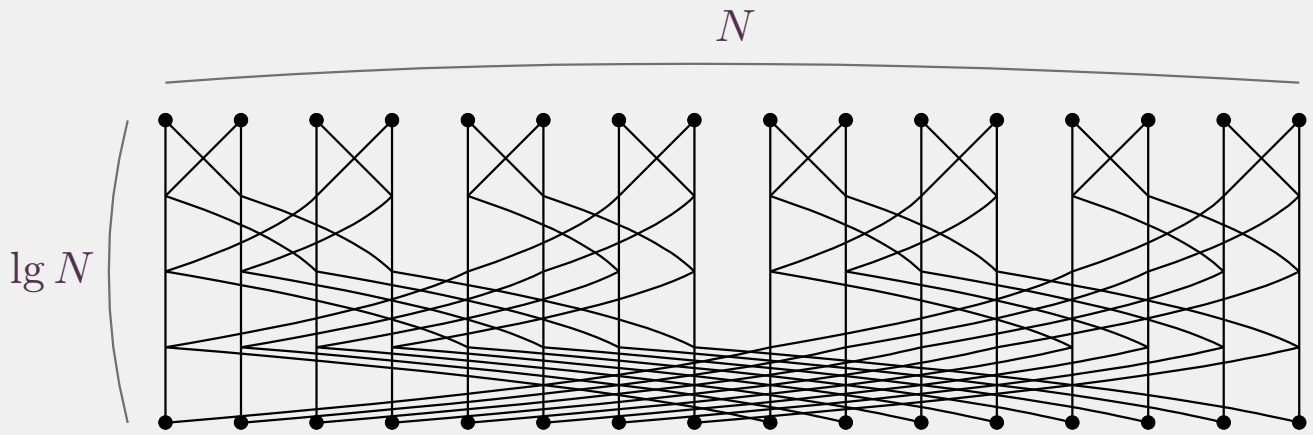
Transformation de Fourier discrète rapide (II)



Côut d'une DFT : $\frac{1}{2} N \lg N$ « papillons » $\rightsquigarrow O(N \lg N)$ opérations dans $\mathbb{K}$

DFT complexe	$N \asymp n/\lg n$	$\mathsf{M}_{\mathbb{K}}(1) = O(\mathsf{I}(\lg n))$	$\mathsf{I}(n) = O(n \lg n\, \mathsf{I}(\lg n) + n \lg n)$
DFT modulaire	$N \asymp n/\lg n$	$\mathsf{M}_{\mathbb{K}}(1) = O(\mathsf{I}(\lg n))$	$\mathsf{I}(n) = O(n \lg n\, \mathsf{I}(\lg n) + n \lg n)$
DFT synthétique	$N \asymp \sqrt{n}$	papillon $\rightsquigarrow O(\sqrt{n})$	$\mathsf{I}(n) = O(n \lg n \sqrt{n} + \sqrt{n}\, \mathsf{I}(\sqrt{n}))$

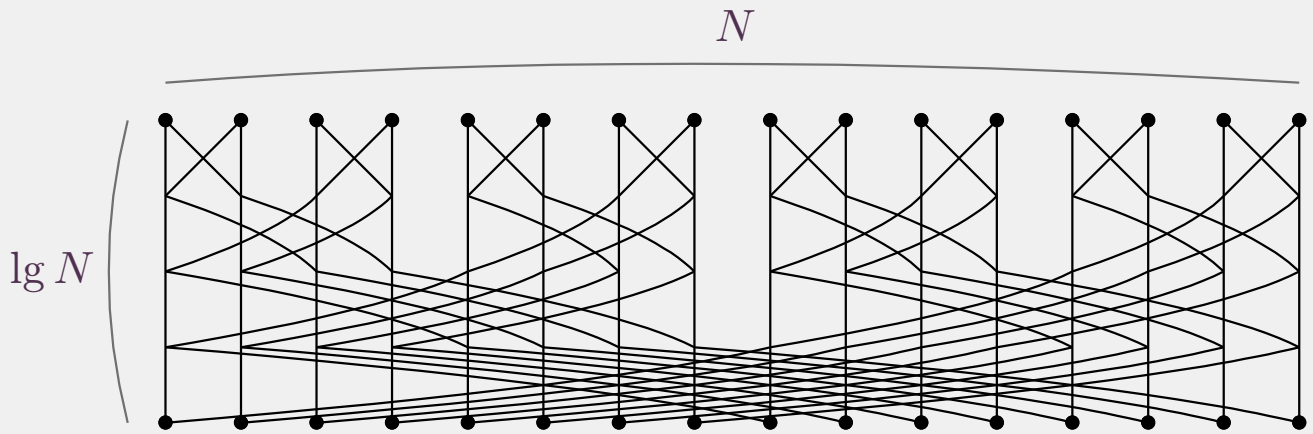
Transformation de Fourier discrète rapide (II)



Côut d'une DFT : $\frac{1}{2} N \lg N$ « papillons » $\rightsquigarrow O(N \lg N)$ opérations dans $\mathbb{K}$

DFT complexe	$N \asymp n/\lg n$	$\mathsf{M}_{\mathbb{K}}(1) = O(\mathsf{I}(\lg n))$	$\mathsf{I}(n) = O(n \lg n \, \mathsf{I}(\lg n) + n \lg n)$
0 DFT modulaire	$N \asymp n/\lg n$	$\mathsf{M}_{\mathbb{K}}(1) = O(\mathsf{I}(\lg n))$	$\mathsf{I}(n) = O(n \lg n \, \mathsf{I}(\lg n) + n \lg n)$
DFT synthétique	$N \asymp \sqrt{n}$	papillon $\rightsquigarrow O(\sqrt{n})$	$\mathsf{I}(n) = O(n \lg n \sqrt{n} + \sqrt{n} \, \mathsf{I}(\sqrt{n}))$

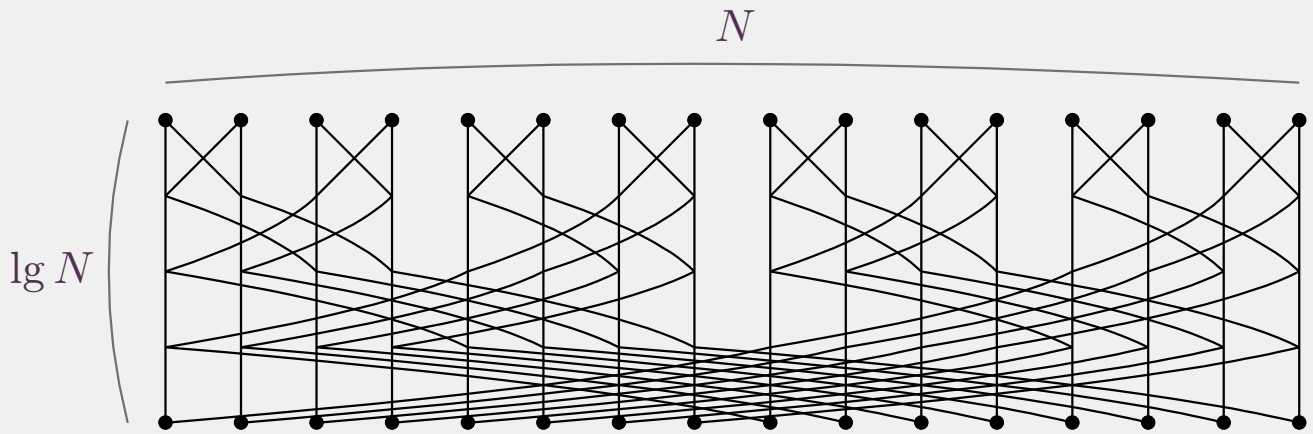
Transformation de Fourier discrète rapide (II)



Côut d'une DFT : $\frac{1}{2} N \lg N$ « papillons » $\rightsquigarrow O(N \lg N)$ opérations dans $\mathbb{K}$

DFT complexe	$N \asymp n/\lg n$	$\mathsf{M}_{\mathbb{K}}(1) = O(\mathsf{I}(\lg n))$	$\mathsf{I}(n) = O(n \lg n\, \mathsf{I}(\lg n) + n \lg n)$
DFT modulaire	$N \asymp n/\lg n$	$\mathsf{M}_{\mathbb{K}}(1) = O(\mathsf{I}(\lg n))$	$\mathsf{I}(n) = O(n \lg n\, \mathsf{I}(\lg n) + n \lg n)$
DFT synthétique	$N \asymp \sqrt{n}$	papillon $\rightsquigarrow O(\sqrt{n})$	$\mathsf{I}(n) = O(n \lg n \sqrt{n} + \sqrt{n}\, \mathsf{I}(\sqrt{n}))$

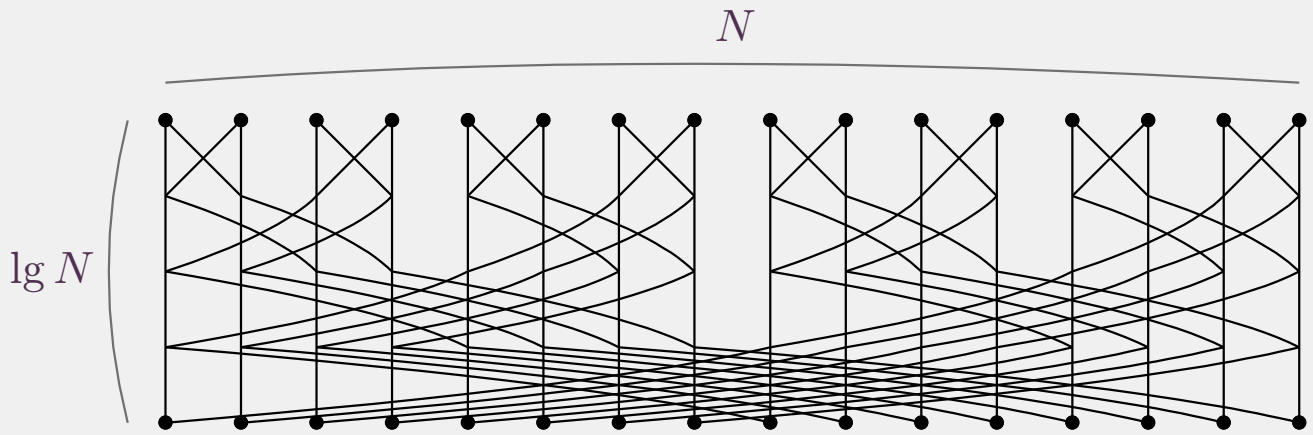
Transformation de Fourier discrète rapide (II)



Côut d'une DFT : $\frac{1}{2} N \lg N$ « papillons » $\rightsquigarrow O(N \lg N)$ opérations dans $\mathbb{K}$

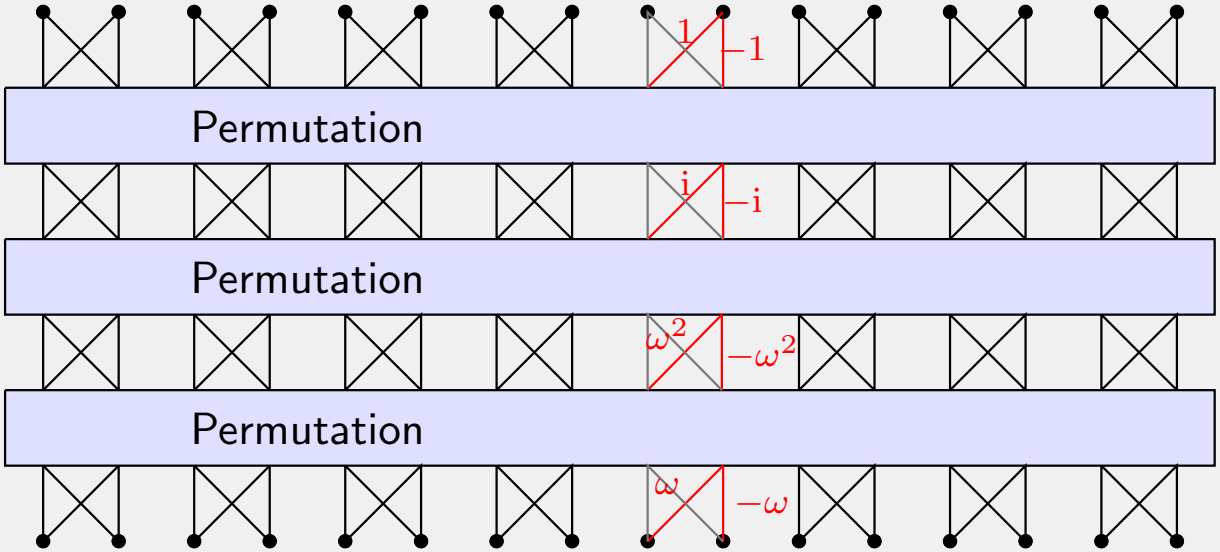
DFT complexe	$N \asymp n/\lg n$	$\mathsf{M}_{\mathbb{K}}(1) = O(\mathsf{I}(\lg n))$	$\mathsf{I}(n) = O(n \lg n\, \mathsf{I}(\lg n) + n \lg n)$
DFT modulaire	$N \asymp n/\lg n$	$\mathsf{M}_{\mathbb{K}}(1) = O(\mathsf{I}(\lg n))$	$\mathsf{I}(n) = O(n \lg n\, \mathsf{I}(\lg n) + n \lg n)$
DFT synthétique	$N \asymp \sqrt{n}$	papillon $\rightsquigarrow O(\sqrt{n})$	$\mathsf{I}(n) = O(n \lg n \sqrt{n} + \sqrt{n}\, \mathsf{I}(\sqrt{n}))$

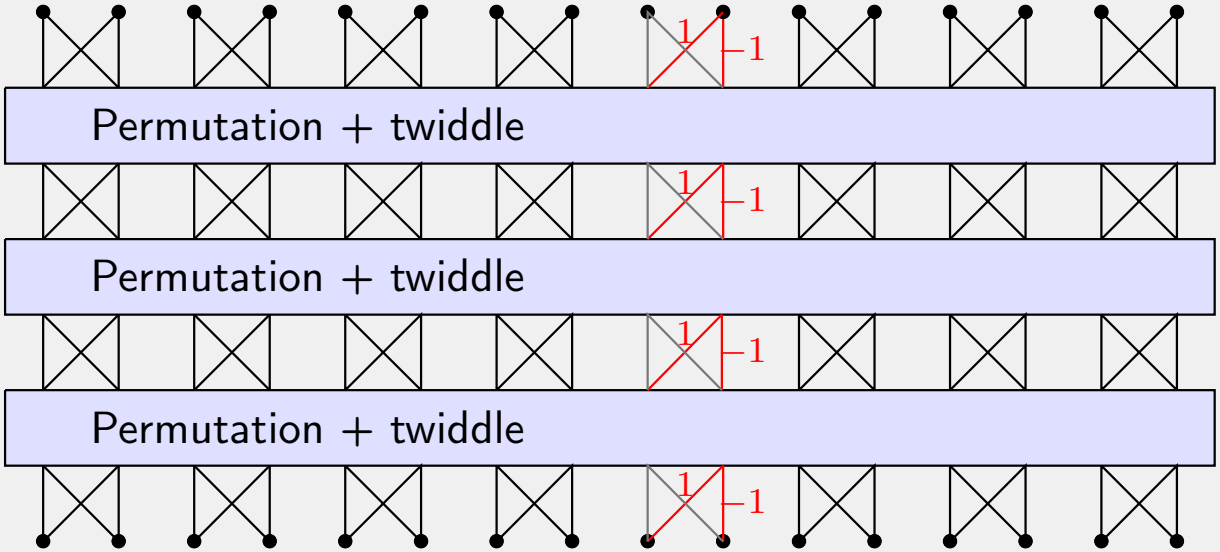
Transformation de Fourier discrète rapide (II)

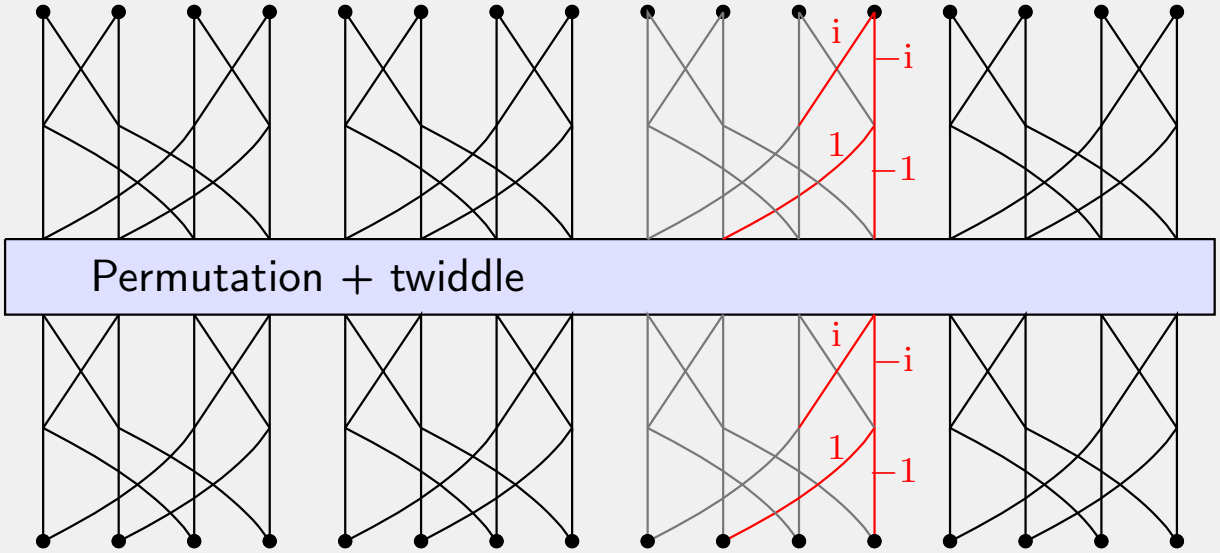


Côut d'une DFT : $\frac{1}{2} N \lg N$ « papillons » $\rightsquigarrow O(N \lg N)$ opérations dans $\mathbb{K}$

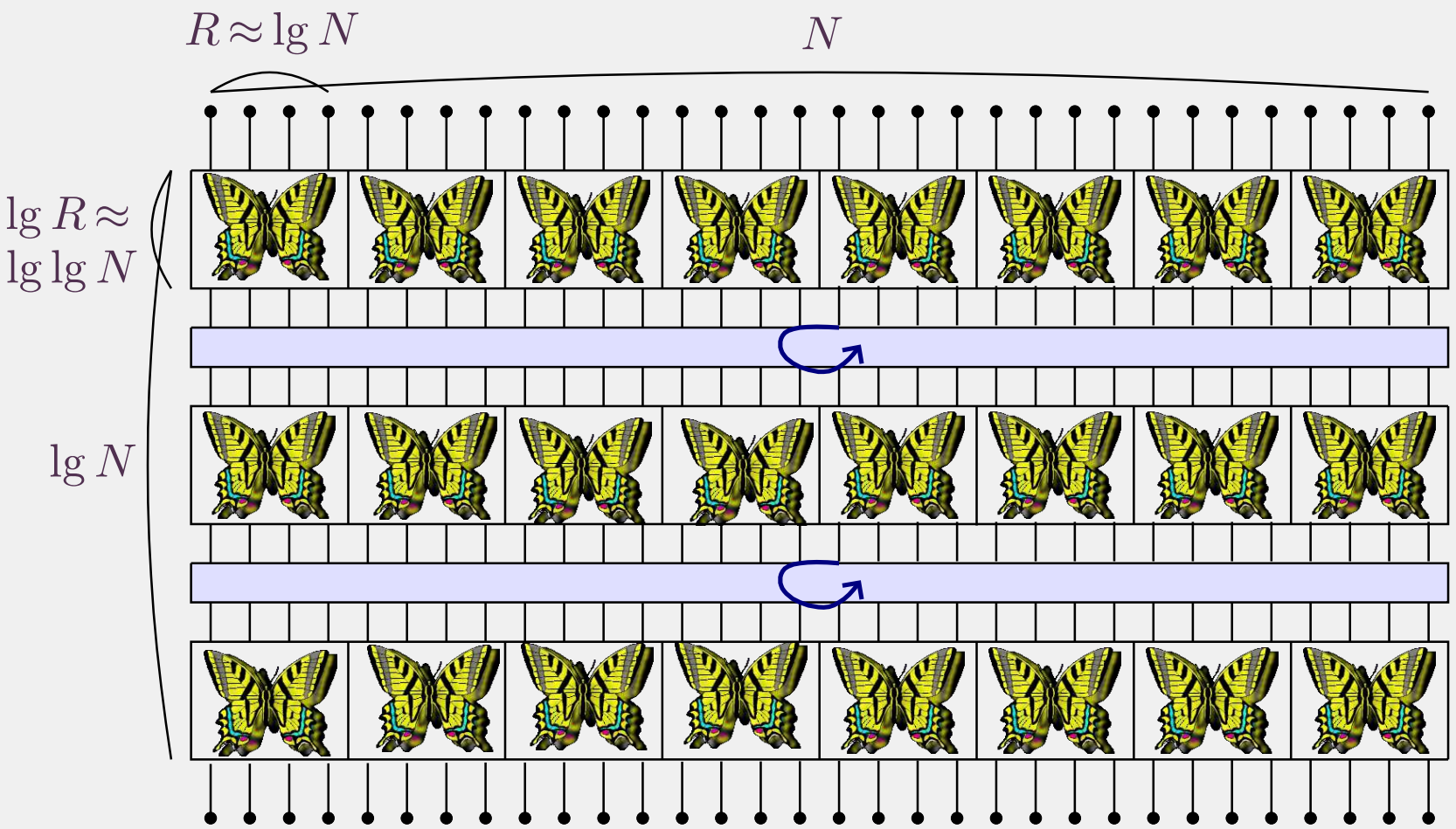
DFT complexe	$N \asymp n/\lg n$	$\mathsf{M}_{\mathbb{K}}(1) = O(\mathsf{I}(\lg n))$	$\mathsf{I}(n) = O(n \lg n \lg\lg n \lg\lg\lg n \cdots)$
DFT modulaire	$N \asymp n/\lg n$	$\mathsf{M}_{\mathbb{K}}(1) = O(\mathsf{I}(\lg n))$	$\mathsf{I}(n) = O(n \lg n \lg\lg n \lg\lg\lg n \cdots)$
DFT synthétique	$N \asymp \sqrt{n}$	papillon $\rightsquigarrow O(\sqrt{n})$	$\mathsf{I}(n) = O(n \lg n \lg\lg n)$


1
−1
Permutation
i
−i
Permutation
ω^2
$-\omega^2$
Permutation
ω
$-\omega$


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


i
i
1
1
Permutation + twiddle
i
i
1
1

Papillons géants



Dessin légèrement trompeur : on aurait dû avoir 16 papillons sur chaque ligne

Algorithme de Fürer

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

Existence d'une racine ω « principale » d'unité d'ordre N, avec $\omega^{2N/R} = X$

Papillons géants (taille $R \times \lg R$) rapides, mais *twiddling* lent $\rightsquigarrow$ multiplications dans $\mathcal{R}$

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

Nouvel algorithme

DFT ordinaire, mais accélération des papillons géants

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

Papillons géants plus lents, mais *twiddling* plus rapide

DFT ordinaire $\Rightarrow$ multiplication interne plus efficace (application : matrices entières)

Convolution cyclique $\rightsquigarrow$ DFT

$P, Q \in \mathbb{K}[X]/(X^n - 1)$, $n \in 2^{\mathbb{N}^>}$, ω racine primitive n-ième d'unité

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

DFT $\rightsquigarrow$ Convolution cyclique [Bluestein 70]

On suppose $\eta \in \mathbb{K}$ tel que $\eta^2 = \omega$.

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

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

Alors $\omega^{ij} = f_i\, f_j\, g_{i-j}$, donc pour tout $a \in \mathbb{K}^n$:

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

On reconnaît une convolution cyclique

Fonction logarithmiquement lente

Fonction $\Phi\colon [x_0,\infty)\to\mathbb{R}$ telle qu'il existe $\ell\in\mathbb{N}$ avec

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

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

Itérateurs [voir aussi Écalle 92, Schmeling 01]

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

Lemme. *Φ logarithmiquement lente, Φ^* itérateur de Φ. Alors*

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

Lemme. *Φ logarithmiquement lente. Constantes K, B, L, ℓ et fonction T telle que*

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

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

Fonction logarithmiquement lente

Fonction $\Phi\colon [x_0,\infty)\to\mathbb{R}$ telle qu'il existe $\ell\in\mathbb{N}$ avec

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

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

Itérateurs [voir aussi Écalle 92, Schmeling 01]

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

Lemme. *Φ logarithmiquement lente, Φ^* itérateur de Φ. Alors*

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

Lemme. *$\Phi_1,\ldots,\Phi_k$ logarithmiquement lentes. Constantes K, B, L, ℓ, $c_1+\cdots+c_k=1$ et fonction T telle que*

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

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

Quelle approche donne l'algorithme le plus rapide ?

Nous avons différentes méthodes pour démontrer

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

Quel est le meilleur K possible ?

Fürer, après optimisation : $K = 16$ (?)
Nous, après optimisation : $K = 8$

Ingrédients

- Multiplication dans $\mathbb{Z} \rightsquigarrow$ multiplication dans $(\mathbb{Z}/(2^n - 1)\,\mathbb{Z})[\mathrm{i}]$.
- Un argument fréquemment partagé dans appels récursifs $\rightsquigarrow$ 2 DFTs au lieu de 3.
- Convolution de longueur N avec coefficients de b bits $\rightsquigarrow$ sorties de taille $2\,b + O(\lg N)$.

  Prendre $b \asymp (\lg n)^2$ au lieu de $b \asymp \lg n$ améliore le ratio $(2\,b + O(\lg N))\,/\,b$.
- Augmentation $R \approx \lg N \rightsquigarrow R \approx (\lg N)^{\lg\lg N + O(1)}$.

  Coût Bluestein–Kronecker $\gg$ coût twiddling et autres.

D'où vient le coût ?

a) Facteur 2 pour segmentation Kronecker ($\mathbb{Z}[\mathrm{i}] \rightsquigarrow \mathbb{C}_b[X]$, découpage en morceaux de $\frac{b}{2}$ bits)
b) Facteur 2 pour DFT directe et inverse
c) Facteur 2 pour substitution de Kronecker ($\mathbb{C}_b[X]/(X^R-1) \rightsquigarrow \mathbb{Z}/(2^{2bR}-1)\mathbb{Z}$)

Nombres premiers de Fermat

Et *si, si, si*, s'il y avait suffisamment de nombres premiers de la forme $p = 2^{2^k}+1$
Approche de Fürer pour $\mathbb{K} = \mathbb{F}_p$ (et optimisée) donne $K = 4$
Malheureusement, $p = 2^{16}+1$ est le plus grand tel nombre connu

Nombres premiers de Mersenne

Conjecture 1. *Soit* $\pi_m(x) = \{p \leqslant x : p = 2^q - 1, p \text{ premier}, q \text{ premier}\}$. *Alors* $\exists a < b$,

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

Algorithme de Crandall–Fagin

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 dans $\mathbb{F}_q[X]$ et dans $\mathbb{K}[X]$

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

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

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

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

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

Théorème. *On a, de façon* ***uniforme*** *en* p *:*

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

Théorème. *Modulo « conjectures vraisemblables », on a de façon* ***uniforme*** *en* p *:*

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

Théorème. *Soit* $\mathbb{A}$ *un* $\mathbb{F}_p$*-algèbre. Alors* $\mathsf{M}^{\mathrm{alg}}_{\mathbb{A}}(n) = O(n \lg n\, 8^{\log^* n})$*, uniformément en* $\mathbb{A}$*. En outre, on n'a besoin que de* $O(n\, 4^{\log^* n})$ *multiplications (non scalaires) dans* $\mathbb{A}$*.*

Éléments de la démonstration

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

2. k tel que $\mathbb{F}_{p^k}[X]$ admet une racine primitive ω d'unité d'ordre N élevé et « friable »

3. Écrire $N = N_1 \cdots N_r$ avec $N_1, ..., N_r$ « maîtrisés » et utiliser 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]
```