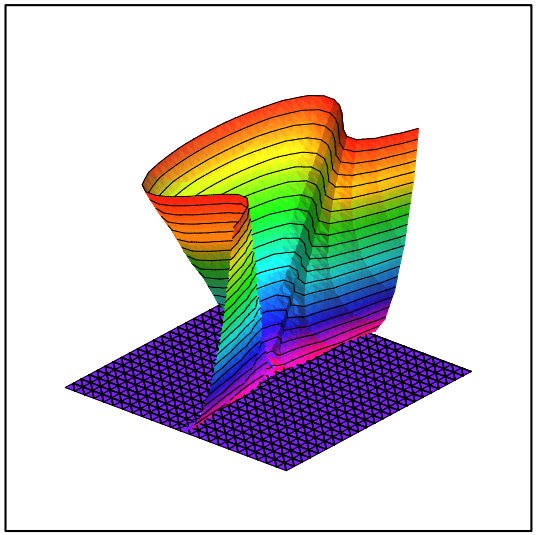
Factorisation de grands nombres à l'aide de courbes elliptiques



Thomas Baignères (thomas.baigneres@epfl.ch)

LASEC

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Au programme

1. Quelques rappels mathématiques (il est toujours temps d'aller chercher un café . . .)
2. La factorisation de 15
3. Premier pas dans la factorisation : l'algorithme $p-1$ de Pollard
4. Courbes elliptique (définition, opérations)
5. Factorisation ECM - Principe
6. Factorisation ECM - L'algorithme
7. La factorisation de $2^{2^{11}}+1$
8. Conclusion

Quelques rappels d'algèbre (1)

- Soient $a, n \in \mathbb{N}$. On dit que a est inversible modulo n lorsqu'il existe b tel que $ab \equiv 1 \pmod{n}$.
- Propriété : a est inversible modulo $n \Leftrightarrow \text{pgcd}(a, n) = 1$

Par exemple :
3 n'est pas inversible modulo 15.
7 est inversible modulo 15 : $7 * 13 \equiv 1 \pmod{15}$

$\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ (p premier) est un corps	$\mathbb{Z}/n\mathbb{Z}$ (n composite) n'est pas un corps

Quelques rappels d'algèbre (2)

Théorème de Lagrange Soit G un groupe d'ordre n. Alors, pour tout élément $a \in G$:

$$\boxed{a^n = e}$$

Comme $\mathbb{F}_p$ est un corps, tous ses éléments (le 0 excepté) sont inversibles. C'est à dire :

$$\mathbb{F}_p^* = \mathbb{F}_p \backslash \{0\} = \{1, 2, \ldots, p-1\}$$

$\mathbb{F}_p$ est un groupe d'ordre $p - 1$.

Petit théorème de Fermat : Soit p un nombre premier. Alors pour tout entier a tel que $\text{pgcd}(a, p) = 1$ on a :

$$\boxed{a^{p-1} \equiv 1 \pmod{p}}$$

LASEC

La factorisation de 15

Supposons ne que personne ici ne connaisse un facteur de 15 (il est encore tôt) . . .

– Tirons un nombre au hasard dans l'ensemble $\{1, 2, \ldots, 14\}$, par exemple 2.

– Calculons pgcd$(2, 15)$, nous obtenons 1 . . .

– Tirons un nouveau nombre, par exemple 6.

– En calculant pgcd$(6, 15)$ nous obtenons 3.

Nous avons découvert que 3 est le pgcd de 6 et 15 . . .

3 est un facteur de 15

LASEC

L'algorithme de factorisation $p-1$ de Pollard (1)

Définitions :

- Un entier n est B-*lisse* lorsque tous les facteurs premiers de n sont inférieurs ou égaux à B.
- Un entier n est B-*superlisse* lorsque toutes les puissances premières divisant n sont inférieures ou égales à B.

Par exemple, $n = 360 = 2^3 3^2 5$ est *5-lisse* et *9-superlisse*.

Soit n un nombre à factoriser, et p un de ces facteurs. L'algorithme trouve p si $p-1$ est B-*superlisse*.
B sera choisi au début de l'algorithme.

L'algorithme de factorisation $p-1$ de Pollard (2)

Principe de l'algorithme :

Si $p-1$ est B-*superlisse*, $p-1$ est un facteur de $B!$. Choisissons un x premier avec n et calculons :

$$\begin{aligned} x^{B!} \bmod n &= x^{k_1(p-1)} \bmod n \\ &= (x^{k_1})^{p-1} + k_2 n \quad \text{où} \quad k_1, k_2 \in \mathbb{Z} \end{aligned}$$

D'après le petit théorème de Fermat :

$$x^{B!} \bmod n \equiv 1 \pmod{p}$$

p est donc un facteur de $(x^{B!} \bmod n) - 1$.

Pour trouver p il suffit de calculer $\text{pgcd}((x^{B!} \bmod n) - 1, n)$.

Courbe elliptique - définition (restreinte . . .)

Soit $p > 3$ premier. Considérons l'équation suivante :

$$E : y^2 = x^3 + ax + b \quad \text{où} \quad a, b \in \mathbb{F}_p$$

Une courbe elliptique est l'ensemble suivant :

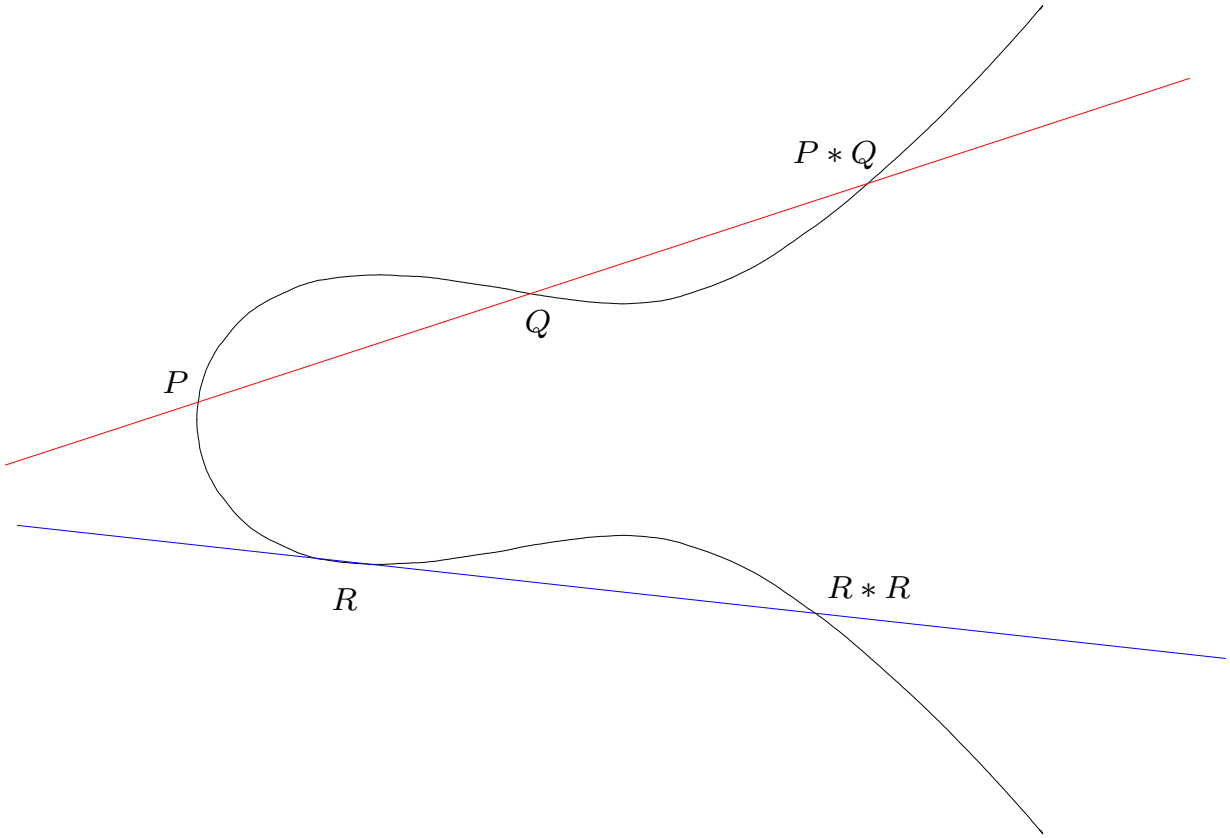
$$\{(x, y) \in \mathbb{F}_p \times \mathbb{F}_p \text{ solutions de } E\} \cup \mathcal{O}$$

- $\mathcal{O}$ est un point particulier, appelé point à l'infini
- Les couples (x, y) seront appelés points de la courbe elliptique E

Nous allons définir une opération d'addition sur notre courbe elliptique, mais avant cela . . .

Courbe elliptique - une propriété

. . .une remarque :

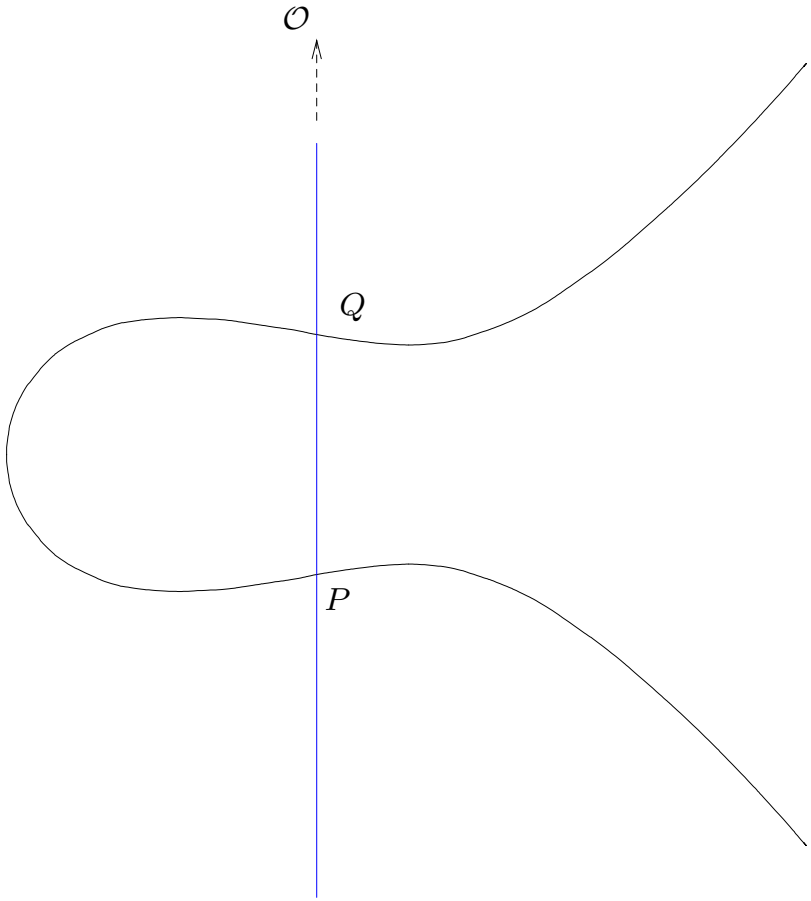


Toute droite qui coupe la courbe en deux points, coupe la courbe en trois points exactement (multiplicité comprise).

LASEC

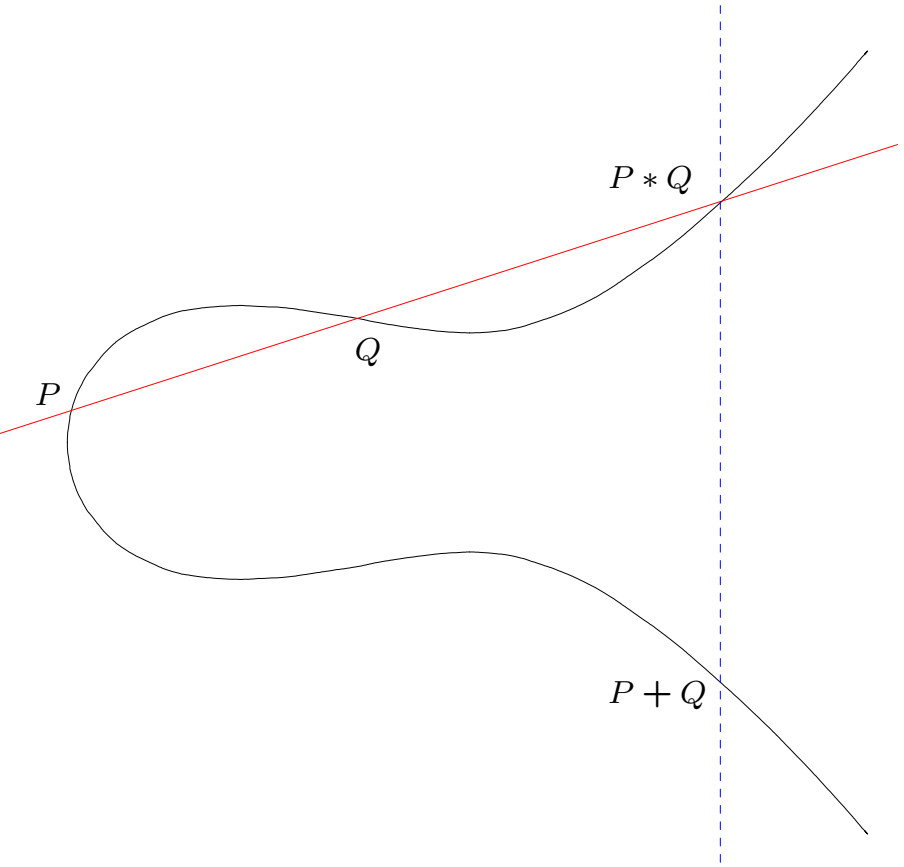
Courbe elliptique - le point à l'infini

Le point à l'infini $\mathcal{O}$ est « le troisième » point d'intersection d'une droite verticale avec la courbe elliptique.



Courbe elliptique - L'opération d'addition (1)

Addition de P et Q



- Si $x_P \neq x_Q$:

$$x_{P+Q} = \lambda^2 - x_P - x_Q$$
$$y_{P+Q} = \lambda(x_P - x_{P+Q} - y_P)$$

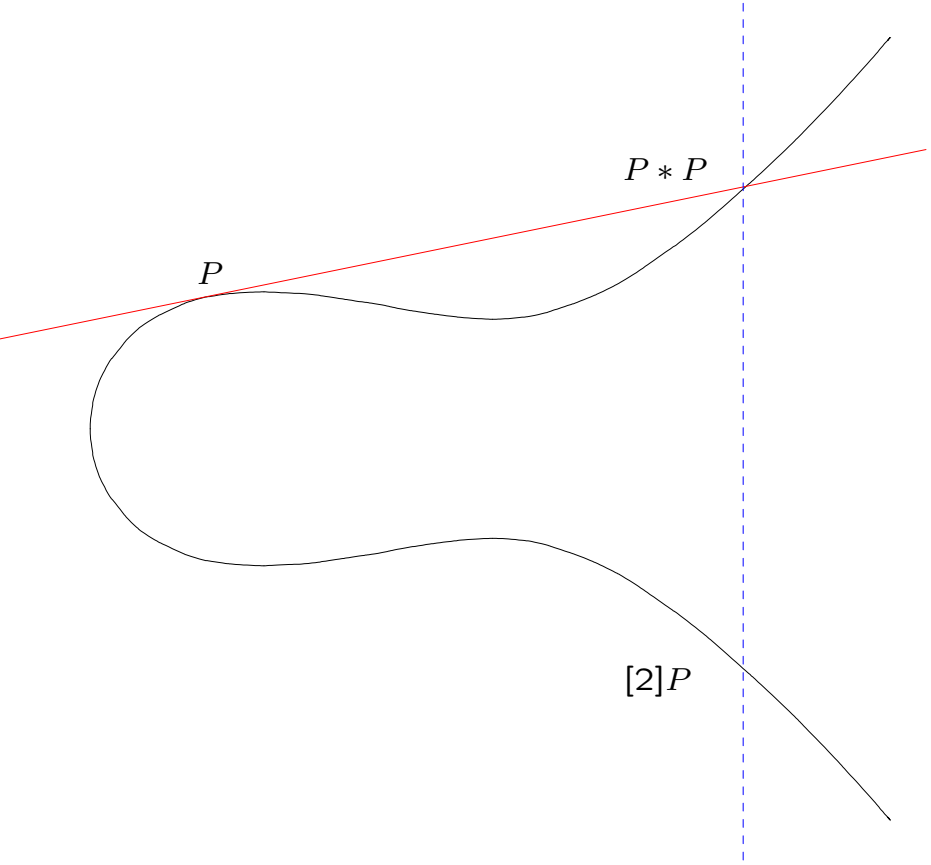
$$\text{avec} \quad \lambda = \frac{y_P - y_Q}{x_P - x_Q}$$

- Si $x_P = x_Q$:

$$P + Q = \mathcal{O}$$

Courbe elliptique - L'opération d'addition (2)

Doublement de P



- Si $y_P \neq 0$

$$x_{[2]P} = \lambda^2 - 2x_P$$
$$y_{[2]P} = \lambda(x_P - x_{[2]P} - y_P)$$

$$\text{avec} \quad \lambda = \frac{3x_P^2 + a}{2y_P}$$

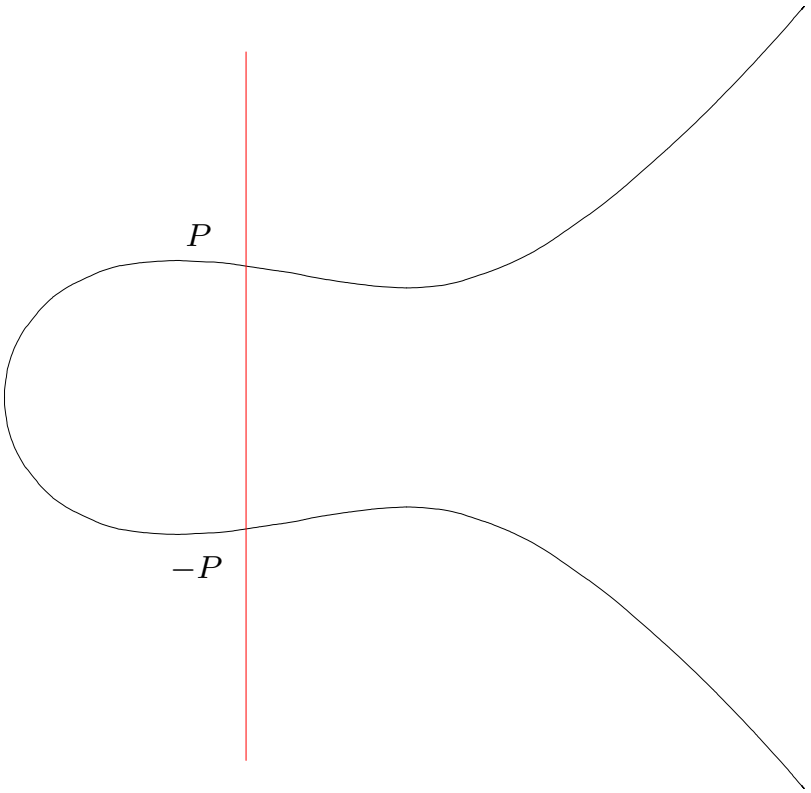
- Si $y_P = 0$:

$$[2]P = \mathcal{O}$$

LASEC

Courbe elliptique - L'opération d'addition (3)

Inverse de P



Comme $P + (-P) = \mathcal{O}$:

$$x_{-P} = x_P$$
$$y_{-P} = y_P$$

Courbe elliptique - L'opération d'addition (4)

Conclusion :

$(E(\mathbb{F}_p), +)$ est un groupe

$\mathcal{O}$ est le neutre pour la loi $+$. Si P est un point de E, d'après le théorème de Lagrange :

$$\underbrace{P + P + \cdots + P}_{\#E(\mathbb{F}_p) \text{ fois}} = \mathcal{O}$$

L'algorithme ECM - L'idée . . .

Nous allons chercher à factoriser n (un nombre non premier).

Pollard $p-1$ effectue des calculs dans $\mathbb{Z}/n\mathbb{Z}$, la méthode ECM les effectue dans $E(\mathbb{Z}/n\mathbb{Z})$, avec :

$$E : y^2 = x^3 + ax + b$$

E n'est *pas* une courbe elliptique ! !

Lors de l'addition (ou du doublement) d'un point, il n'est pas toujours possible de calculer λ. Il nécessite un calcul d'inverse dans $\mathbb{Z}/n\mathbb{Z}$, où tous les éléments ne sont pas inversibles ($\mathbb{Z}/n\mathbb{Z}$ n'est pas un corps).

L'algorithme ECM - Le principe (1)

Si $m = \#E(\mathbb{F}_p)$ est B-*superlisse*, alors m est un facteur de $B!$.
Soit P un point de $E(\mathbb{F}_p)$.

D'après le théorème de Lagrange :

$$[B!]P = \mathcal{O} \quad \text{dans} \quad E(\mathbb{F}_p)$$

Mais nous ne connaissons pas p . . .les calculs se feront dans $E(\mathbb{Z}/n\mathbb{Z})$.

Le point $[B!]P$ ne sera pas défini sur $E(\mathbb{Z}/n\mathbb{Z})$, en effet . . .

L'algorithme ECM - Le principe (2)

Si $P + Q = \mathcal{O}$ dans $E(\mathbb{F}_p)$ $(P \neq Q)$:

$$\begin{aligned} x_P \equiv x_Q \pmod{p} &\Rightarrow x_P - x_Q \equiv 0 \pmod{p} \\ &\Rightarrow p \mid (x_P - x_Q) \end{aligned}$$

Donc pgcd$(x_P - x_Q, n) > 1$, $x_P - x_Q$ est non inversible modulo n.

Comme $\lambda = \frac{y_P - y_Q}{x_P - x_Q}$, son calcul provoque une erreur dans $E(\mathbb{Z}/n\mathbb{Z})$.

Même résultat pour le calcul de $[2]P$.

Cette erreur de calcul peut nous permettre de trouver un facteur de n !

L'algorithme ECM - L'algorithme (1)

Algorithme ECM :

- N un entier composé
- B une borne bien choisie

[Initialisation]

1. Calculer le tableau $p[1], \ldots, p[k]$ de tous les premiers jusqu'à B
2. Calculer le tableau $q[1], \ldots, q[k]$ tel que :
   - $q[i] = p[i]^{s_i}$
   - $q[i] \leq B < q[i] \cdot p[i]$

[Initialisation courbe]

3. $a \leftarrow 0$ et soit E la courbe $y^2 = x^3 + ax + 1$

[Initialisation point]

4. $P \leftarrow (0, 1)$ et $i \leftarrow 0$

L'algorithme ECM - L'algorithme (2)

[Puissance première suivante]

5. $i \leftarrow i + 1$
6. Si $i > k$
7. $\quad a \leftarrow a + 1$ et aller à [Initialisation du point]

[Multiplication du point]

8. $P \leftarrow [q[i]]P$
9. Si le calcul réussit, aller à [Puissance première suivante]

[Fin]

10. Le calcul a échoué. Soir α l'élément non inversible.
11. $g \leftarrow \mathrm{pgcd}(\alpha, N)$
12. Si $g < N$, g est un facteur de N.
13. Sinon faire $a \leftarrow a + 1$ et aller à [Initialisation du point]

L'algorithme ECM - L'algorithme (3)

Une dernière explication :

Décomposons $m = \#E(\mathbb{F}_p)$ en produit de puissances premières :

$$m = p_1^{n_1} \dots p_r^{n_r}$$

Dans l'algorithme, nous calculons progressivement le point :

$$[p[1]^{s_1} \dots p[k]^{s_k}]P$$

Si m est B-*superlisse*, nous pouvons affirmer que :

$$p_1^{n_1} \dots p_r^{n_r} \mid p[1]^{s_1} \dots p[k]^{s_k}$$

Donc :

$$[p[1]^{s_1} \dots p[k]^{s_k}]P = \mathcal{O}$$

LASEC

Complexité (1)

La complexité de notre algorithme dépends de la taille de p (pas de la taille de n).

La complexité de l'algorithme ECM :

$$O\left(e^{(\sqrt{2}+o(1))\sqrt{\log p \log\log p}}\right)$$

Le calcul de la complexité nous permet de fixer une valeur optimale pour B :

$$B = \left(e^{\sqrt{\log p \log\log p}}\right)^{(1/\sqrt{2})}$$

Complexité (2)

Comparaison des complexités de différents algorithmes de factorisation :

- Pollard $p-1$ en

$$O\left(n^{1/4}\right) \rightsquigarrow O\left(e^{\frac{1}{4}\log n}\right) \rightsquigarrow O\left(e^{\frac{1}{2}\log p}\right)$$

- ECM et Quadratic Sieve en

$$O\left(e^{(\sqrt{2}+o(1))\sqrt{\log p \log\log p}}\right)$$

- Number Field Sieve en

$$e^{O\left((\log n)^{1/3}(\log\log n)^{2/3}\right)}$$

La factorisation du 11^e nombre de Fermat

L'algorithme ECM nous a permis de factoriser le 11^e nombre de Fermat :

$$2^{2^{11}}+1 = 319489 \cdot 974849 \cdot 167988556341760475137 \cdot 3560841906445833920513 \cdot p564$$

Conclusion

Mais ça ne s'arrête pas là . . .

Il est possible d'apporter des améliorations :

- Optimiser les additions/doublements de points
- Paralléliser les calculs sur plusieurs courbes
- Agrémenter l'algorithme d'une deuxième phase
- Choisir la courbe pour que l'ordre de $E(\mathbb{F}_p)$ soit divisible par 12
- . . .