



A Code-based Hash and Sign Signature Scheme

Gustavo Banegas, Kévin Carrier, André Chailloux, Alain Couvreur, **Thomas Debris-Alazard**, Philippe Gaborit, Pierre Karpman, Johanna Loyer, Ruben Niederhagen, Nicolas Sendrier, Benjamin Smith and Jean-Pierre Tillich

Inria, École Polytechnique

1. Wave: standardization candidate (NIST),
2. Next steps,
3. Code-based hash and sign,
4. Design Rationale: Wave Trapdoor,
5. Leakage free signatures,
6. Removing Approximation in Prange.

<https://wave-sign.org>



WAVE: STANDARDIZATION CANDIDATE (NIST)

Wave is a **hash and sign** digital signature scheme.

By proving that signatures are leakage-free,

→ Wave instantiates Gentry-Peikert-Vaikuntanathan (**GPV**) framework
like Falcon, Squirrels, HuFu

But Wave security relies on **coding** problems



Even if parameters are highly conservative

- **Short signatures:** linear scaling in the security

Post-quantum target security	Level I	Level III	Level V
Signature length (Bytes)	822	1249	1644

- **Fast Verification:** (Intel Core i5-1135G7 platform at 2.40GHz)

Post-quantum target security	Level I	Level III	Level V
Verification (MCycles)	1.2	2.5	4.3

- Immune to statistical attacks.
- Proven secure (Q)ROM with tight reductions.



- Big public-key: quadratic scaling in the security

Post-quantum target security	Level I	Level III	Level V
Public-key size (MBytes)	3.6	7.8	13.6

- Signing and key generation rely on Gaussian elimination on large matrices
- Security based on fairly new assumption (2018): distinguishing random and generalized $(U | U + V)$ -codes



NEXT STEPS

Wave parameters are highly **conservative!**

Attack model:

Cost of \mathcal{A} to solve \mathcal{P} :

$$\alpha \stackrel{\text{def}}{=} \lim_{n \rightarrow +\infty} \frac{1}{n} \log_2 \text{Time}(\mathcal{A})$$

Then choose n s.t:

$$\alpha n = \lambda \quad (\alpha \approx 0.0149)$$

→ It ignores (super-)polynomial factors and memory access!

For instance: considered attack to forge a signature

$$\text{Time} = P(\lambda)2^\lambda \quad \text{and} \quad \text{Memory} = Q(\lambda)2^\lambda.$$

Next Step:

Providing parameters for “concrete” security.



Wave reference implementation

- portable C99,
- KeyGen and Sign in constant-time,
- bit-sliced arithmetic over \mathbb{F}_3 .

Bottleneck of Wave: **Gaussian elimination** on big matrices/**memory access**
(it impacts key generation and signing not verification)

Next Step:

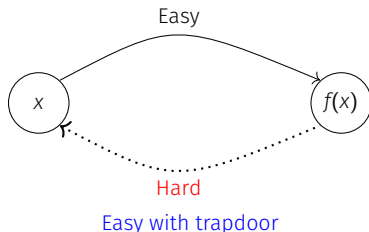
- Providing **optimized** implementation: AVX,
→ Wavelet: AVX2 (intel) & ARM CORTEX M4 **in verification** (2x faster),
- Providing a Wave version with **countermeasures, maskings**,
- Providing (friendly) tools to ensure that Wave is properly implemented.



CODE-BASED HASH AND SIGN

FULL DOMAIN HASH SIGNATURE SCHEME

- ▶ Hash(\cdot) hash function,
- ▶ f **trapdoor one-way** function



- ▶ To sign m :

Compute $\sigma \in f^{-1}(\text{Hash}(m))$.

f needs to be **surjective!**

- ▶ To verify (m, σ) :

Check $f(\sigma) \stackrel{?}{=} \text{Hash}(m)$.



→ Coding theory provides one-way functions!

- A $[n, k]$ -code \mathcal{C} is defined as a k dimension subspace of \mathbb{F}_q^n .
- \mathbb{F}_q^n embedded with **Hamming weight**,

$$\forall \mathbf{x} \in \mathbb{F}_q^n, \quad |\mathbf{x}| \stackrel{\text{def}}{=} \#\{i, \mathbf{x}(i) \neq 0\}.$$



One-way in code-based crypto:

$$f_w : (\mathbf{c}, \mathbf{e}) \in \mathcal{C} \times \{\mathbf{e} : |\mathbf{e}| = w\} \mapsto \mathbf{c} + \mathbf{e}.$$

(inverting f_w : decoding \mathcal{C} at distance w)

→ To hope f_w surjective: choose noise distance w large enough



One-way in code-based crypto:

$$f_w : (\mathbf{c}, \mathbf{e}) \in \mathcal{C} \times \{\mathbf{e} : |\mathbf{e}| = w\} \mapsto \mathbf{c} + \mathbf{e}.$$

(inverting f_w : decoding \mathcal{C} at distance w)

→ To hope f_w surjective: choose noise distance w large enough

But, be careful...

w parametrizes the hardness of inverting f_w !

→ for some w , it is easy to invert f_w ...



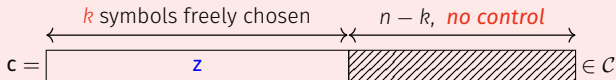
HARD OR EASY TO INVERT? PRANGE ALGORITHM

Inverting f_w :

- Given: $[n, k]$ - \mathcal{C} , y **uniformly distributed** over \mathbb{F}_q^n and w ,
- Find: $c \in \mathcal{C}$ such that $|y - c| = w$.

Fact: by linear algebra (Gaussian elimination)

\mathcal{C} has dimension k : $\forall z \in \mathbb{F}_q^k$, easy to compute $c \in \mathcal{C}$ such that,



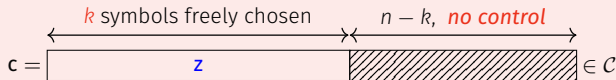
HARD OR EASY TO INVERT? PRANGE ALGORITHM

Inverting f_w :

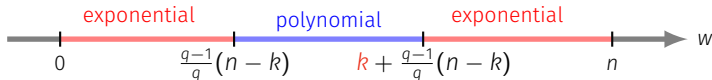
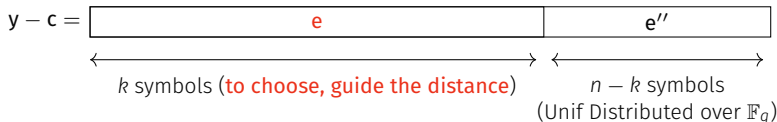
- Given: $[n, k]$ - \mathcal{C} , \mathbf{y} **uniformly distributed** over \mathbb{F}_q^n and w ,
- Find: $\mathbf{c} \in \mathcal{C}$ such that $|\mathbf{y} - \mathbf{c}| = w$.

Fact: by linear algebra (Gaussian elimination)

\mathcal{C} has dimension k : $\forall \mathbf{z} \in \mathbb{F}_q^k$, easy to compute $\mathbf{c} \in \mathcal{C}$ such that,



Given a uniform $\mathbf{y} \in \mathbb{F}_q^n$: compute $\mathbf{c} \in \mathcal{C}$,



- ▶ **Public data:** a hash function $\text{Hash}(\cdot)$, an $[n, k]$ -code \mathcal{C} and,

$$w \notin \left[\frac{q-1}{q}(n-k), k + \frac{q-1}{q}(n-k) \right] \quad (\text{signing distance})$$

- ▶ **Signing m :**

1. Hashing: $m \rightarrow y \stackrel{\text{def}}{=} \text{Hash}(m) \in \mathbb{F}_q^n$,
2. Decoding: find **with a trapdoor** $c \in \mathcal{C}$ such that $|y - c| = w$.

- ▶ **Verifying (m, c) :**

$$c \in \mathcal{C} \quad \text{and} \quad |\text{Hash}(m) - c| = w.$$

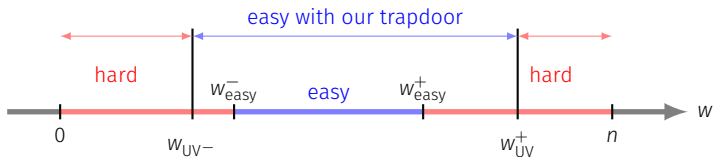
Security:

Signing distance w s.t hard to find $c \in \mathcal{C}$ at distance w

→ Unless to own a secret/trapdoor structure on \mathcal{C} !



DECODING WITH OUR TRAPDOOR



Trapdoor:

An $[n, k]$ -code \mathcal{C} with a peculiar structure enabling to decode at distance

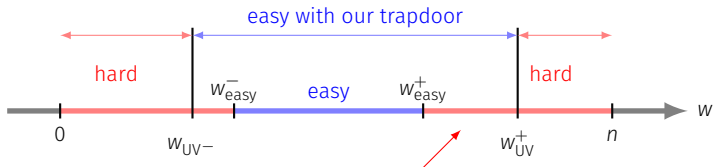
$$w \notin [w_{\text{easy}}^-, w_{\text{easy}}^+]$$

Security:

\mathcal{C} indistinguishable from a random code (unless to know its peculiar structure)



DECODING WITH OUR TRAPDOOR



Trapdoor:

An $[n, k]$ -code \mathcal{C} with a peculiar structure enabling to decode at distance

$$w \notin [w_{easy}^-, w_{easy}^+]$$

Security:

\mathcal{C} indistinguishable from a random code (unless to know its peculiar structure)



DESIGN RATIONALE: WAVE TRAPDOOR

- Vector permutation:

$$\mathbf{x} = (\mathbf{x}(i))_{1 \leq i \leq n} \in \mathbb{F}_q^n ; \pi \text{ permutation of } \{1, \dots, n\}.$$

$$\mathbf{x}^\pi \stackrel{\text{def}}{=} (\mathbf{x}(\pi(i)))_{1 \leq i \leq n}$$

- Component-wise product:

$$\mathbf{a} \star \mathbf{x} \stackrel{\text{def}}{=} (\mathbf{a}(i)\mathbf{x}(i))_{1 \leq i \leq n}$$



Generalized $(U \mid U + V)$ -codes:

Let U and V be $[n/2, k_U]$ and $[n/2, k_V]$ -codes

$$\mathcal{C} \stackrel{\text{def}}{=} \left\{ (\mathbf{x}_U + \mathbf{b} \star \mathbf{x}_V \mid \mathbf{c} \star \mathbf{x}_U + \mathbf{d} \star \mathbf{x}_V)^\pi : \mathbf{x}_U \in U \text{ and } \mathbf{x}_V \in V \right\}$$

where π permutation, $\mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{F}_q^{n/2}$ verify $\mathbf{c}(i) \neq 0$ and $\mathbf{d}(i) - \mathbf{b}(i)\mathbf{c}(i) = 1$.

→ It defines a code with dimension $k \stackrel{\text{def}}{=} k_U + k_V$

Secret-key/Trapdoor: $U, V, \mathbf{b}, \mathbf{c}, \mathbf{d}$ and π .

Security assumption: Distinguishing Wave Key (DWK)

Hard to distinguish random and generalized $(U \mid U + V)$ codes.



OUR DECODING ALGORITHM (1)

Secret-key/Trapdoor: $U, V, \mathbf{b}, \mathbf{c}, \mathbf{d}$ and π .

1. Given $\text{Hash}(\mathbf{m}) = \mathbf{y} \in \mathbb{F}_q^n$: decompose $\mathbf{y} = (\mathbf{y}_L \mid \mathbf{y}_R)^\pi$,
2. Compute any $\mathbf{x}_V \in V$ with **Prange Algorithm**,
3. Using **Prange Algorithm**: compute $\mathbf{x}_U \in U$ by **choosing k_U symbols $x_U(i)$'s** such that

$$\begin{cases} \mathbf{x}_U(i) + \mathbf{b}(i)\mathbf{x}_V(i) \neq \mathbf{y}_L(i) \\ \mathbf{c}(i)\mathbf{x}_U + \mathbf{d}(i)\mathbf{x}_V(i) \neq \mathbf{y}_R(i) \end{cases}$$

(i) $q \geq 3$, (ii) $\mathbf{c}(i) \neq 0$ and (iii) $\mathbf{d}(i) - \mathbf{b}(i)\mathbf{c}(i) = 1$.

4. Return $\mathbf{c} \stackrel{\text{def}}{=} (\mathbf{x}_U + \mathbf{b} \star \mathbf{x}_V \mid \mathbf{c} \star \mathbf{x}_U + \mathbf{d} \star \mathbf{x}_V)^\pi \in \mathcal{C}$ (public code).

What is the (typical) distance w between \mathbf{y} and \mathbf{c} ?



OUR DECODING ALGORITHM (2)

Given any valid $x_V = \overbrace{\hspace{10em}}^{n/2} \in V$

$x_U = \overbrace{\hspace{6em}}^{k_U} x_U^{\text{choose}}(i) \overbrace{\hspace{3em}}^{\text{no control}} \in U$

$$c - (y_L | y_R) = \overbrace{x_U^{\text{choose}}(i) + b(i)x_V(i) - y_L(i)} \overbrace{\hspace{2em}} \overbrace{c(i)x_U^{\text{choose}}(i) + d(i)x_V(i) - y_R(i)} \overbrace{\hspace{2em}} \overbrace{\hspace{2em}}^{n/2 - k_U}$$

- Choose k_U symbols $x_U^{\text{choose}}(i)$ such that:
$$\begin{cases} x_U^{\text{choose}}(i) + b(i)x_V(i) - y_L(i) \neq 0 \\ c(i)x_U^{\text{choose}}(i) + d(i)x_V(i) - y_R(i) \neq 0 \end{cases}$$

Typical distance:

$$w = 2k_U + 2 \frac{q-1}{q} (n/2 - k_U) > w_{\text{easy}}^+ = (k_U + k_V) + \frac{q-1}{q} (n - (k_U + k_V))$$

as soon as: $k_U > k_V$ (parameter constraint in Wave)



Collecting signatures:

$$(x_U + \mathbf{b} * x_V \mid \mathbf{c} * x_U + \mathbf{d} * x_V)^\pi$$

may enable to recover the secret, for instance $\pi...$



BE CAREFUL: A HUGE ISSUE

Collecting signatures:

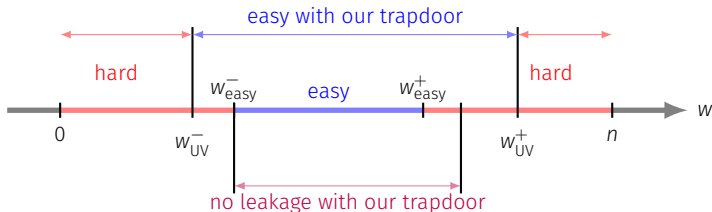
$$(x_U + \mathbf{b} \star x_V \mid \mathbf{c} \star x_U + \mathbf{d} \star x_V)^\pi$$

may enable to recover the secret, for instance π ...

Above procedure **leaks quickly** π ...

Proper Wave specification/implementation:

Choose carefully internal distribution and perform rejection sampling to produce signatures immune to statistical attacks



In what follows:

We will work in \mathbb{F}_3 , $q = 3$.



LEAKAGE FREE SIGNATURES

A signature: $x \in f^{-1}(y)$.

→ x computed via a trapdoor/secret!

Ideal situation:

x distribution **independent** of the secret

→ For instance: x uniform over its domain when y uniform

A hard problem

In our case: **exponential** number of preimages



Given uniform \mathbf{y} : compute $(\mathbf{x}_U + \mathbf{b} \star \mathbf{x}_V \mid \mathbf{c} \star \mathbf{x}_U + \mathbf{d} \star \mathbf{x}_V)^\pi$ such that

$\mathbf{e}^{\text{sgn}} \stackrel{\text{def}}{=} \mathbf{y} - (\mathbf{x}_U + \mathbf{b} \star \mathbf{x}_V \mid \mathbf{c} \star \mathbf{x}_U + \mathbf{d} \star \mathbf{x}_V)^\pi$ uniform over words of Hamming weight w .



Given uniform \mathbf{y} : compute $(\mathbf{x}_U + \mathbf{b} \star \mathbf{x}_V \mid \mathbf{c} \star \mathbf{x}_U + \mathbf{d} \star \mathbf{x}_V)^\pi$ such that

$\mathbf{e}^{\text{sgn}} \stackrel{\text{def}}{=} \mathbf{y} - (\mathbf{x}_U + \mathbf{b} \star \mathbf{x}_V \mid \mathbf{c} \star \mathbf{x}_U + \mathbf{d} \star \mathbf{x}_V)^\pi$ uniform over words of Hamming weight w .

Important fact: as $d(i) - b(i)c(i) = 1$ for all i ,

$$\varphi : (\mathbf{z}_U, \mathbf{z}_V) \mapsto (\mathbf{z}_U + \mathbf{b} \star \mathbf{z}_V \mid \mathbf{c} \star \mathbf{z}_U + \mathbf{d} \star \mathbf{z}_V)^\pi \text{ *bijection* .}$$

1. Write $\mathbf{y} = (\mathbf{y}_U + \mathbf{b} \star \mathbf{y}_V \mid \mathbf{c} \star \mathbf{y}_U + \mathbf{d} \star \mathbf{y}_V)^\pi$

2. Deduce that $\mathbf{e}^{\text{sgn}} = (\mathbf{e}_U + \mathbf{b} \star \mathbf{e}_V \mid \mathbf{c} \star \mathbf{e}_U + \mathbf{d} \star \mathbf{e}_V)^\pi$ where $\begin{cases} \mathbf{e}_V \stackrel{\text{def}}{=} \mathbf{y}_V - \mathbf{x}_V \\ \mathbf{e}_U \stackrel{\text{def}}{=} \mathbf{y}_U - \mathbf{x}_U \end{cases}$

Here \mathbf{x}_V and \mathbf{x}_U are computed via Prange algorithm...



$e^{\text{sgn}} \stackrel{\text{def}}{=} (e_U + b \star e_V \mid c \star e_U + d \star e_V)^\pi$ and e^{unif} unif word of weight w .

→ Write: $e^{\text{unif}} = (e_U^{\text{unif}} + b \star e_V^{\text{unif}} \mid c \star e_U^{\text{unif}} + d \star e_V^{\text{unif}})^\pi$

We would like,

$$e^{\text{sgn}} \sim e^{\text{unif}}$$

In a first step we want,

$$e_V \sim e_V^{\text{unif}} \quad \text{where} \quad e_V = y_V - x_V = y_V - \text{Prange}(V, y_V)$$

Important remark (function of weight):

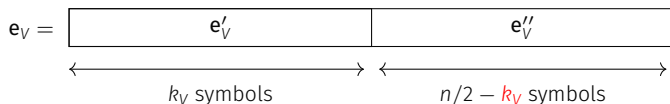
$$\mathbb{P}(e_V^{\text{unif}} = x) = \frac{1}{\#\{y : |y| = t\}} \mathbb{P}(|e_V^{\text{unif}}| = t) \quad \text{when } |x| = t.$$

Approximation: Distribution of Prange algorithm, only function of the weight

$$\mathbb{P}(\text{Prange}(\cdot) = x \mid |\text{Prange}(\cdot)| = t) = \frac{1}{\#\{y : |y| = t\}} \quad \text{when } |x| = t.$$

→ Uniformity property: **enough to reach** $|e_V| \sim |e_V^{\text{unif}}|$ as distribution

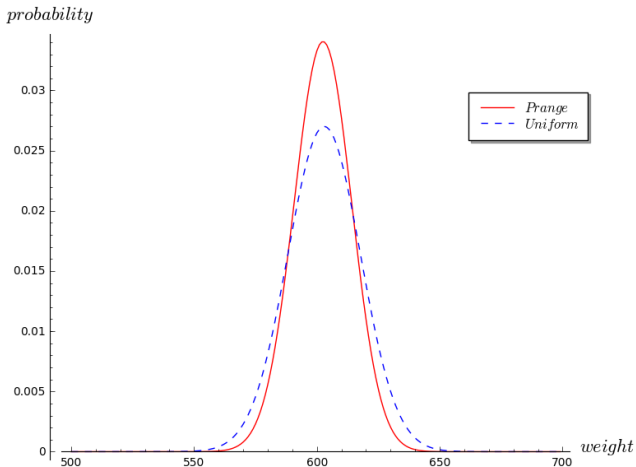
- We first look for $\mathbb{E}(|e_V|) = \mathbb{E}(|e_V^{\text{unif}}|)$



- e''_V follows a uniform law over $\mathbb{F}_3^{n/2 - k_V}$: $\mathbb{E}(|e''_V|) = \frac{2}{3}(n/2 - k_V)$
- e'_V can be chosen.

$$\rightarrow k_V \text{ is fixed as: } \mathbb{E}(|e'_V|) + \frac{2}{3}(n/2 - k_V) = \mathbb{E}(|e_V^{\text{unif}}|)$$

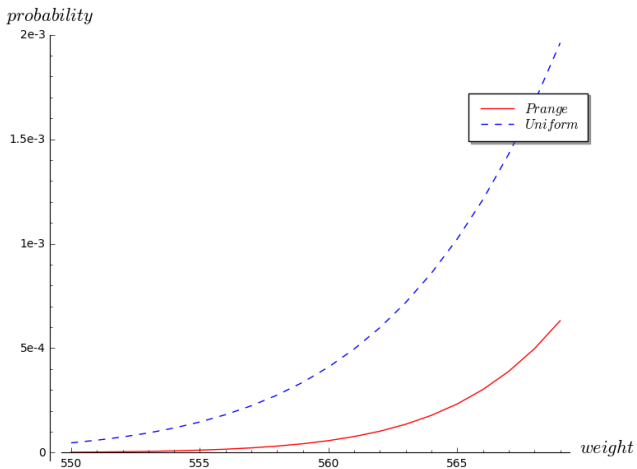
Perform rejection sampling!



$$\mathbb{P}(\text{accept}) = \min_j \frac{\mathbb{P}(|e_V| = j)}{\mathbb{P}(|e_V^{\text{unif}}| = j)} \ll 1.$$

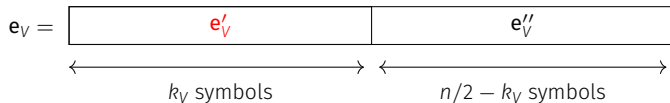


REJECTION SAMPLING: TAIL



$$\mathbb{P}(\text{accept}) = \min_j \mathbb{P}(\text{accept}) = \min_j \frac{\mathbb{P}(|e_V| = j)}{\mathbb{P}(|e_V^{\text{unif}}| = j)} \ll 1.$$





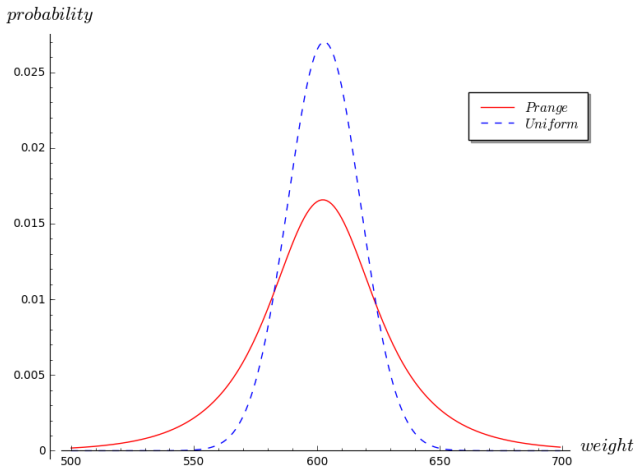
- e''_V follows a uniform law: its variance is fixed,

Choose the weight of e'_V as a random variable!

- $|e'_V|$ s.t. $\begin{cases} \mathbb{E}(|e'_V|) + \frac{2}{3}(n/2 - k_V) = \mathbb{E}(|e_V^{\text{unif}}|) \\ |e'_V| \text{ high variance!} \end{cases}$



REJECTION SAMPLING



$$\mathbb{P}(\text{accept}) = \min_j \mathbb{P}(\text{accept}) = \min_j \frac{\mathbb{P}(|e_V| = j)}{\mathbb{P}(|e_V^{\text{unif}}| = j)} \approx C^{\text{ste}}.$$



→ Distribution $|e_V|'$ can be **precisely** chosen s.t. $\mathbb{P}(\text{accept}) \approx 1$

Using Renyi divergence argument: **removing rejection sampling!**

Signing algorithm: signatures **don't leak** any information on the secret-key!



REMOVING APPROXIMATION IN PRANGE

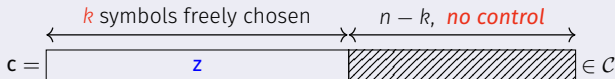
PRANGE ALGORITHM: GAUSSIAN ELIMINATION

To represent \mathcal{C} : use a basis/**generator-matrix** $\mathbf{G} \in \mathbb{F}_q^{k \times n}$,

$$\mathcal{C} = \{ \mathbf{x}\mathbf{G} : \mathbf{x} \in \mathbb{F}_q^k \} \quad (\text{rows of } \mathbf{G} \text{ form a basis of } \mathcal{C}).$$

Prange algorithm: by linear algebra (Gaussian elimination)

\mathcal{C} has dimension k : $\forall \mathbf{z} \in \mathbb{F}_q^k$, easy to compute $\mathbf{c} \in \mathcal{C}$ such that,



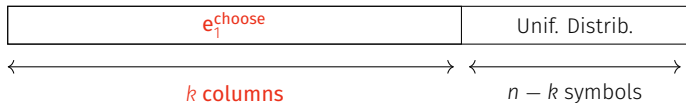
Where the k symbols are picked is not uniform!

1. Pick $\mathcal{I} \subseteq \{1, \dots, n\}$ such that $\mathbf{G}_{\mathcal{I}}$ has rank k (columns of \mathbf{G} indexed by \mathcal{I}),
2. Compute the codeword $\mathbf{x}\mathbf{G}$ where $\mathbf{x} \stackrel{\text{def}}{=} \mathbf{z}\mathbf{G}_{\mathcal{I}}^{-1}$.



$$\mathbb{P}(\text{Prange}(\cdot) = \mathbf{x} \mid |\text{Prange}(\cdot)| = t) = \frac{1}{\#\{\mathbf{y} : |\mathbf{y}| = t\}} \quad : \text{only } \approx$$

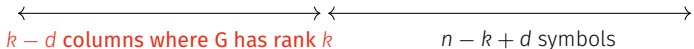
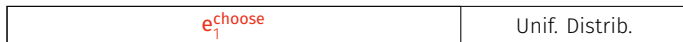
→ Only \approx as **we cannot invert** the system for all k coordinates!



NON-UNIFORMITY OF PRANGE

$$\mathbb{P}(\text{Prange}(\cdot) = \mathbf{x} \mid |\text{Prange}(\cdot)| = t) = \frac{1}{\#\{\mathbf{y} : |\mathbf{y}| = t\}} \quad : \text{only } \approx$$

→ Only \approx as **we cannot invert** the system for all k coordinates!



true with proba. $\approx 1 - 1/3^{k-(k-d)}$



Signing algorithm: signatures **don't leak** any information on the secret-key!

→ It enables to reduce the security (EUF-CMA in (Q)ROM) to the hardness of:

Security reduction ((Q)ROM):

- Decoding a random linear code at distance $w \approx 0.9n$,
- Distinguishing random and generalized $(U | U + V)$ -codes.

