# Wave Parameter Selection

Nicolas Sendrier, Inria

PQCrypto 2023

August 16 2023, College Park, MD, USA

# Wave

Wave is an hash-and-sign digital signature scheme based on codes.

Wave leverages the decoding of ternary generalized $(U|U+V)$ codes, which is easier than the decoding of random codes of same size.

Wave is secure under the following assumptions:

- Hardness of decoding (for large weight),
- Pseudorandomness of permuted generalized ternary $(U|U+V)$ codes.

This talk: relate the security assumptions to hard decoding problems and their solvers, and describe how to select secure parameters

# Decoding Problem

**Decoding Problem** – DP$(q; n, k, t)$

A finite field $\mathbf{F}_q$ and three integers $n, k, t$ such that $n > k > 0$ and $0 \leq t \leq n$.

**Instance:** $(\mathbf{H}, \mathbf{s}) \in \mathbf{F}_q^{(n-k)\times n} \times \mathbf{F}_q^{n-k}$

**Solution:** $\mathbf{e} \in \mathbf{F}_q^n$ such that $|\mathbf{e}| = t$ and $\mathbf{e}\mathbf{H}^\mathsf{T} = \mathbf{s}$.

**Hard** if $\begin{cases} t < \frac{q-1}{q}(n-k) & \text{“small weight”} \\ t > \frac{q-1}{q}(n-k) + k & \text{“large weight”} \end{cases}$.

**Easy** if $0 \leq t - \frac{q-1}{q}(n-k) \leq k$.

$t:$ Hard: from $0$ to $\frac{q-1}{q}(n-k)$; Easy: from $\frac{q-1}{q}(n-k)$ to $k + \frac{q-1}{q}(n-k)$; Hard: from $k + \frac{q-1}{q}(n-k)$ to $n$

# Decoding Problem – Multiple Instances

**DOOM Problem** – $\mathrm{DP}_N(q; n, k, t)$ *Decoding One Out of Many*

A finite field $\mathbf{F}_q$ and three integers $n, k, t$ such that $n > k > 0$ and $0 \leq t \leq n$.

**Instance:** $(\mathbf{H}, \mathbf{s}_1, \dots, \mathbf{s}_N) \in \mathbf{F}_q^{(n-k)\times n} \times \left(\mathbf{F}_q^{n-k}\right)^N$

**Solution:** $\mathbf{e} \in \mathbf{F}_q^n$ such that $|\mathbf{e}| = t$ and $\mathbf{e}\mathbf{H}^\mathsf{T} \in \{\mathbf{s}_1, \dots, \mathbf{s}_N\}$.

$\mathrm{DP}_N$ is not harder when $N$ grows.

$\mathrm{DP}_\infty$ if the adversary is free to choose $N$.

$\mathrm{DP}_\infty$ is hard $\Longleftrightarrow$ DP is hard

# Generalized Ternary $(U|U+V)$ Codes

$n$ an even integer, $k = k_U + k_V$ with $0 < k_U < n/2$ and $0 < k_V < n/2$

A generalized ternary $(U|U+V)$ code admits a parity check matrix

$$\mathbf{H} = \left(\begin{array}{c|c} \mathbf{d} \star \mathbf{H}_U & -\mathbf{b} \star \mathbf{H}_U \\ \hline -\mathbf{c} \star \mathbf{H}_V & \mathbf{a} \star \mathbf{H}_V \end{array}\right) \in \mathbf{F}_3^{(n-k)\times n}$$

where:

- $\mathbf{H}_U \in \mathbf{F}_3^{(n/2-k_U)\times n/2}$ and $\mathbf{H}_V \in \mathbf{F}_3^{(n/2-k_V)\times n/2}$, random
  ($U = \langle \mathbf{H}_U \rangle^\perp$ and $V = \langle \mathbf{H}_V \rangle^\perp$ denote the codes admitting $\mathbf{H}_U$ and $\mathbf{H}_V$ respectively as parity check matrices)

- $\mathbf{a} = (a_i)_{0\le i<n}, \mathbf{b} = (b_i)_{0\le i<n}, \mathbf{c} = (c_i)_{0\le i<n}, \mathbf{d} = (d_i)_{0\le i<n}$ in $\mathbf{F}_3^n$,
  $\forall i, 0 \le i < n, a_i \neq 0, c_i \neq 0, a_i d_i - b_i c_i \neq 0$

- '$\star$' denotes the component-wise product

Wave

We denote $\mathcal{C}$ the generalized $(U|U+V)$ code associated to $(U, V, \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d})$.

The code $\mathcal{C}$ admits the following generator matrix

$$\mathbf{G} = \left(\begin{array}{c|c} \mathbf{a} \star \mathbf{G}_U & \mathbf{c} \star \mathbf{G}_U \\ \hline \mathbf{b} \star \mathbf{G}_V & \mathbf{d} \star \mathbf{G}_V \end{array}\right) \in \mathbf{F}_3^{k \times n}$$

where $\mathbf{G}_U \in \mathbf{F}_3^{k_U \times n/2}$ and $\mathbf{G}_V \in \mathbf{F}_3^{k_V \times n/2}$ are any generator matrices of $U$ and $V$ respectively.

Finally note that the dual of $\mathcal{C}$ is also a generalized $(U|U+V)$ code (associated to $(V^{\perp}, U^{\perp}, -\mathbf{c}, \mathbf{d}, \mathbf{a}, -\mathbf{b})$)

## Generalized Ternary $(U|U+V)$ Codes – Trapdoor Decoder

There exists a probabilistic decoding procedure for $\mathcal{C}$

$$\begin{array}{rccll} \Phi_{\mathcal{C},w}: & \mathbf{F}_3^{n-k} & \longrightarrow & \mathbf{F}_3^{n} & \\ & \mathbf{s} & \longmapsto & \mathbf{e} & \text{such that } \mathbf{e}\mathbf{H}^{\intercal} = \mathbf{s}, |\mathbf{e}| = w \end{array}$$

which takes benefit of the $(U|U+V)$ structure and runs successfully in polynomial time for a range of values $w > k + \frac{2}{3}(n-k)$.

(recall that generic decoding is hard for such $w$)

# Wave

Hash-and-Sign signature scheme:

- Public: a ternary $[n, k]$ code $\mathcal{C}_{\text{pub}}$
- Secret: a (trapdoor) decoder for $w$ errors in $\mathcal{C}_{\text{pub}}$
  ($\mathcal{C}_{\text{pub}}$ a permuted generalized ternary $(U|U+V)$ code)
- Signature: the solution of a decoding problem for $w$ errors in $\mathcal{C}_{\text{pub}}$, the instance is obtained by hashing the message

Security:

- Solving $\text{DP}_\infty(3; n, k, w)$ is hard enough
- Distinguishing $\mathcal{C}_{\text{pub}}$ from random is hard enough

# More on Wave Security

Wave is proven EUF-CMA using a GPV-like framework.

Requires the output distribution of the trapdoor function to be independent of the secret $\rightarrow$ immunity to statistical attacks.

$\rightarrow$ an additional parameter $g$, the *gap*, used in the decoder, was introduced to ensure a uniformity condition for the proof.

(The gap is such that, essentially, any $m \times (m+g)$ ternary matrix has rank $m$ with high enough probability,
*e.g.* in NIST's threat model $g = 40 \approx 64/\log_2 3$)

# Selecting Parameters for Wave

1. Choose $n$, $k$ ($k = n/2$ for NIST) and $g$

2. Choose $k_U, w$ (and $k_V = k - k_U$) such that

$$w = \frac{2}{3}(n + k_U - g) \quad \left(\text{and } w > \frac{2}{3}(n-k) + k\right)$$

   $w$ large is best against forgery attacks

   $k_U$ small is best against key attacks.

$\rightarrow$ there is a trade-off to optimize step 2.

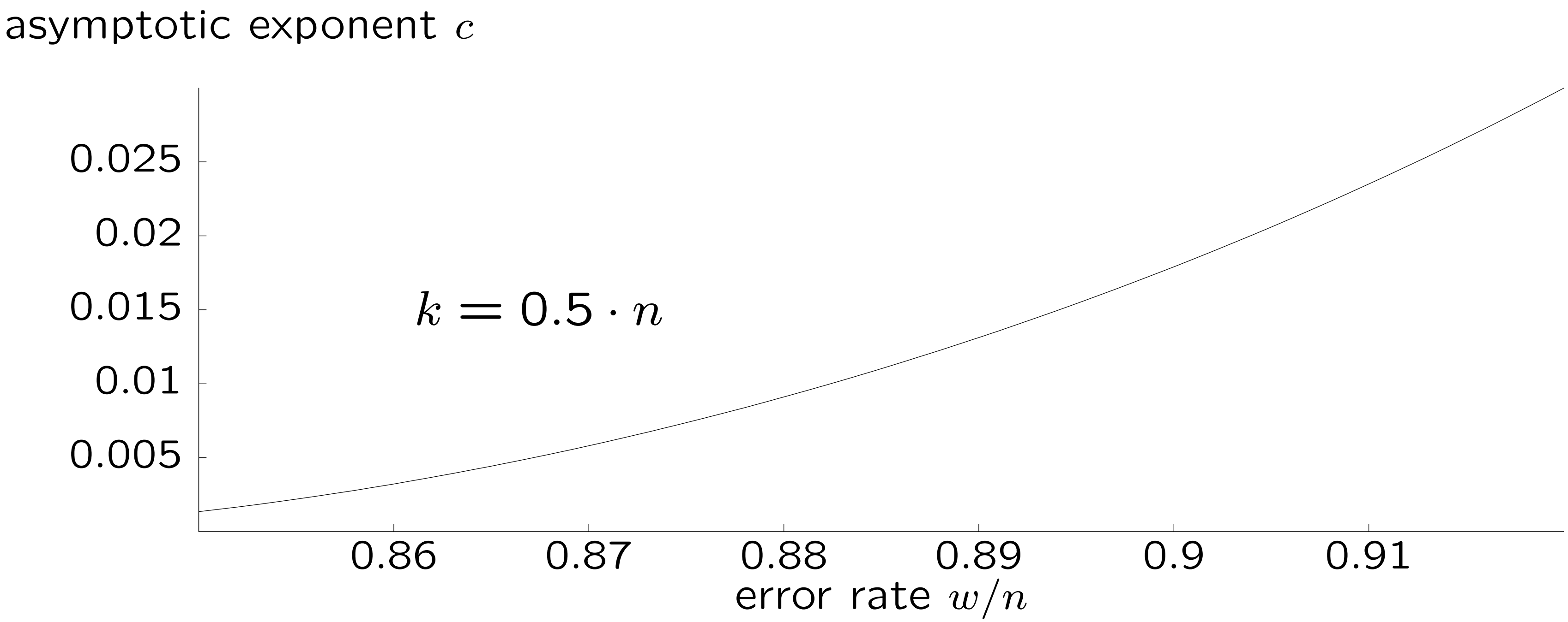
# Best Known Attack Against Computational Assumptions

**Forgery Attacks:** Solve $\mathrm{DP}_\infty(3; n, k, w)$ when $w$ is large. Best known approach [Bricout, Chailloux, Debris-Alazard, Lequesne, 2019] is Information Set Decoding (ISD) + Wagner's Generalized Birthday Algorithm (GBA).

**Key Attacks:** Distinguish $\mathcal{C}_{\text{pub}}$ from random.

Best known approach: find unusual codewords (type-$U$ or type-$V$, definition coming next. . . )

# Forgery Attack




Computational cost (asymptotic) for solving $\mathrm{DP}_\infty(3; n, k, w)$ with ISD+GBA (classical)

$$\mathrm{WF} = 2^{c \cdot n}$$

To reach $\lambda = 128$ bits of (classical) security:

$w = 0.92 \cdot n \rightarrow c = 0.03 \rightarrow n \geq 4\,267$

$w = 0.87 \cdot n \rightarrow c = 0.0058 \rightarrow n \geq 22\,000$

# Weight Distribution of Generalized $(U|U+V)$ Codes

Except for the two following subcodes:

**type-**$U$**:** $\mathcal{U}(\mathcal{C}) = \{(\mathbf{a} \star \mathbf{u}, \mathbf{c} \star \mathbf{u}) \mid \mathbf{u} \in U\}$

**type-**$V$**:** $\mathcal{V}(\mathcal{C}) = \{(\mathbf{b} \star \mathbf{v}, \mathbf{d} \star \mathbf{v}) \mid \mathbf{v} \in V\}$

the weight distribution of a (permuted) generalized $(U|U+V)$ is as for a random code, [Debris-Alazard, PhD, 2019].

# Weight Distribution of Generalized $(U|U+V)$ Codes

$\mathcal{U}(\mathcal{C},j) = \{(\mathbf{a}\star\mathbf{u},\mathbf{c}\star\mathbf{u}) \mid \mathbf{u}\in U, |\mathbf{u}| = j\}$

has cardinality $\frac{\binom{n/2}{j}2^j}{3^{n/2-k_U}}$ and contains words of weight $t = 2j$

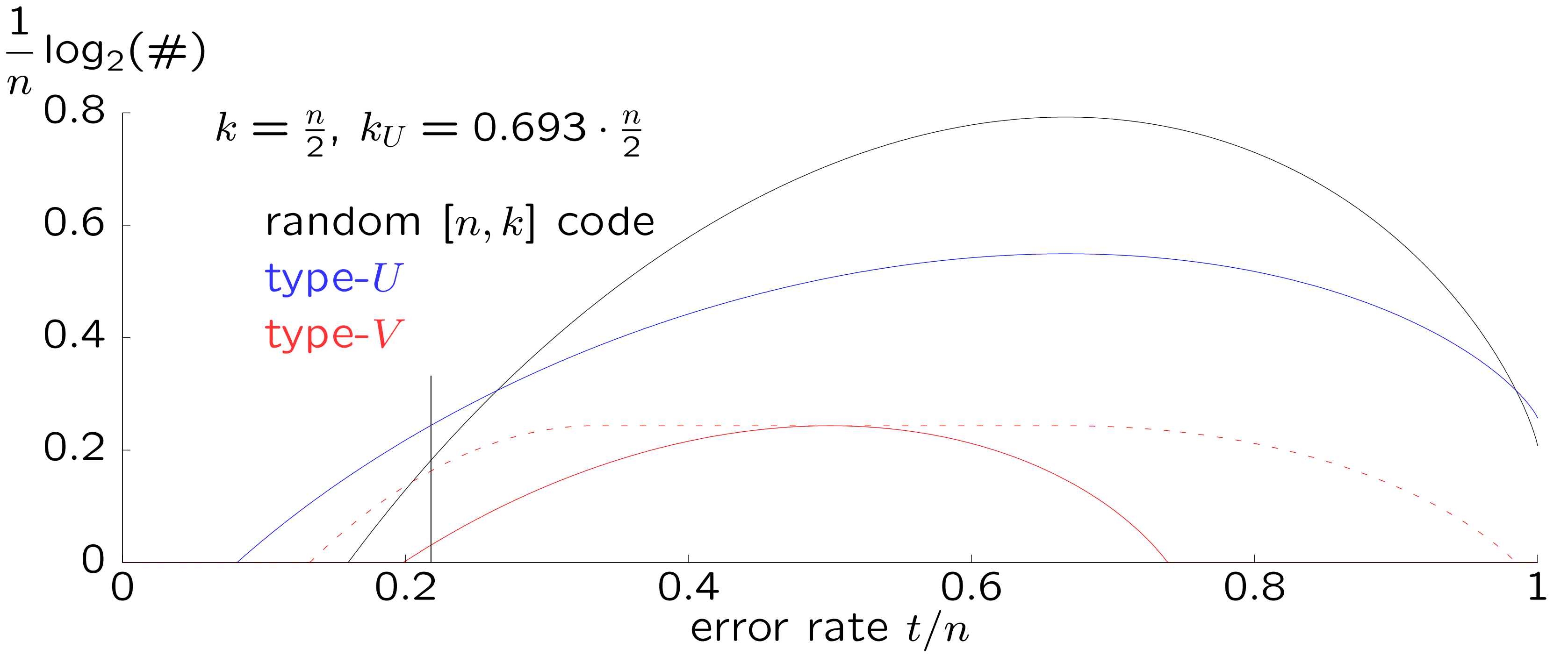
$\mathcal{V}(\mathcal{C},j) = \{(\mathbf{b}\star\mathbf{v},\mathbf{d}\star\mathbf{v}) \mid \mathbf{v}\in V, |\mathbf{v}| = j\}$

has cardinality $\frac{\binom{n/2}{j}2^j}{3^{n/2-k_V}}$ and contains words of weight $t \in [j, 2j]$

(average weight is $\frac{4}{3}\cdot j$)

A random ternary $[n,k]$ code contains $\frac{\binom{n}{t}2^t}{3^{n-k}}$ words of weight $t$

# Weight Distribution of Generalized $(U|U+V)$ Codes




Example: for $t = 0.209 \cdot n$ in the above figure:

- the number of "random" codewords is $2^{0.156 \cdot n}$
- the number of type-$U$ codewords is $2^{0.231 \cdot n}$
- the number of type-$V$ codewords is $2^{0.0169 \cdot n}$

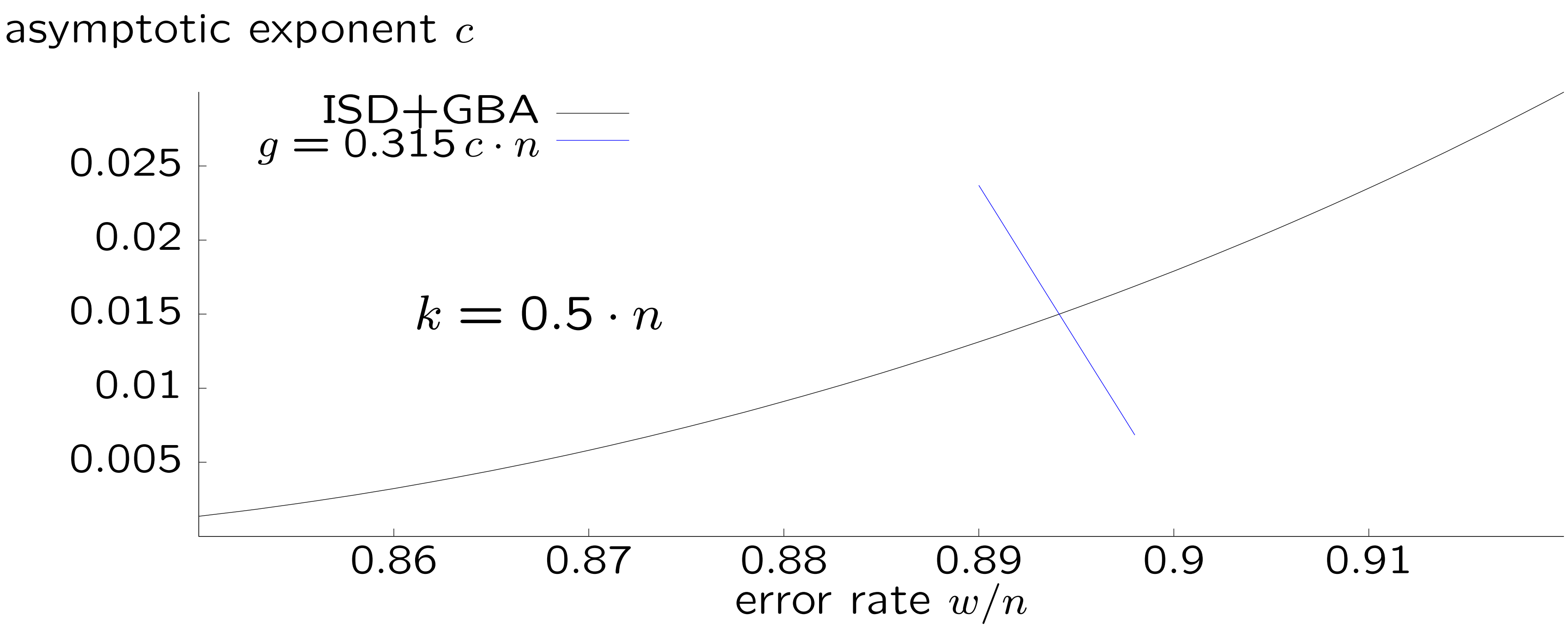
# Key Attack: Finding Unexpected Codewords

For Wave relevant parameters, there are always fewer type-$V$ than type-$U$ codewords.

For extremal weights type-$U$ codewords may dominate, and the cost for finding words of that weight in $\mathcal{C}_{\text{pub}}$ (or $\mathcal{C}_{\text{pub}}^{\perp}$) will be smaller than expected in a random code. This provides a distinguisher whose cost is obtained by minimizing over all weights.

To estimate this cost, codewords are searched with the variant of ISD due to [May, Meurer, Thomae, 2011].

# Forgery and Key Attacks – Trade-off




For fixed $n, k, g$, the cost for finding type-$U$ codewords depends on $k_U$. Using the relation $w = \frac{2}{3}(n + k_U - g)$, this cost can be viewed as a function of $w$ plotted above in blue together with the forgery cost.

The intersection of the curves corresponds to the optimal parameters.

# Wave Parameters

NIST parameters are for $k = n/2$

The security parameter $\lambda$ corresponds to classical security bits

Quantum security is always $\geq \lambda/2$ bits

| NIST | $\lambda$ | $n$ | $k$ | $w$ | $k_U$ | $k_V$ | $g$ |
|---|---|---|---|---|---|---|---|
| Level I | 128 | 8 576 | 4 288 | 7 668 | 2 966 | 1 322 | 40 |
| Level III | 192 | 12 544 | 6 272 | 11 226 | 4 335 | 1 937 | 40 |
| Level V | 256 | 16 512 | 8 256 | 14 784 | 5 704 | 2 552 | 40 |

| | Signature length (bytes) | | Key size |
|---|---|---|---|
| | avg. entropy | max length | (MBytes) |
| Level I | 772.5 | 822 | 3.68 |
| Level III | 1129.8 | 1249 | 7.87 |
| Level V | 1487.0 | 1644 | 13.63 |

# Conclusion

- Signature length scales linearly with security
- Key size scales quadratically with security
- The parameter selection process is easy to adapt if/when forgery or key attacks improve
- Code rate 1/2 features a good trade-off between signature length and key size
  (higher rates reduce the signature length)
  (lower rates reduce the key size)