Descent into Broken Trust

Uncovering ML-DSA Subkeys with Scarce Leakage and Local Optimization

Carsten Schubert[1] Niklas Paskarbeit[2] Jean-Pierre Seifert[1] Marian Margraf[2]

[1]TU Berlin [2]FU Berlin

SAC 2026

The Setting

ML-DSA and the promise of rejection sampling

Signing procedure:

1. uniformly sample masking randomness $\boldsymbol{y} \leftarrow \mathcal{R}_{\gamma_1}^{\ell}$
2. derive random challenge c with exactly τ entries ± 1
3. output $\boldsymbol{z} := \boldsymbol{y} + c\,\boldsymbol{s}_1$ **only if** $\|\boldsymbol{z}\|_\infty \leq \gamma_1 - \tau\eta = \gamma_1 - \beta$

Per coefficient: $z = \langle c, x \rangle + y$, $x =$ one component of $\boldsymbol{s}_1$.

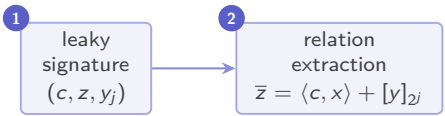
The security argument

Rejection sampling makes z *uniform* on $[\pm(\gamma_1 - \beta)]$ — independent of secret x.

But: rests entirely on y being *completely hidden*.

The attack of LZSWZM20[1] and DKMNT25[2]

From many (c, z, y_j) with y_j as the leaked j-th bit of y, recover $x \in \{-\eta, \ldots, \eta\}^n$

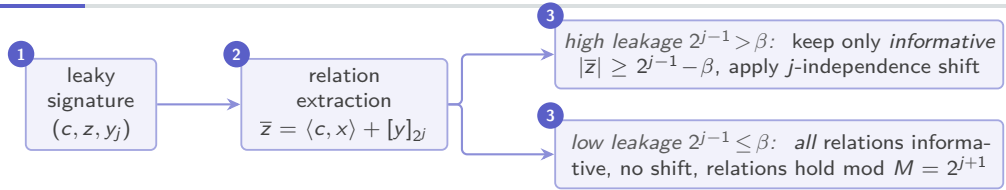


[1] Yuejun Liu, Yongbin Zhou, Shuo Sun, Tianyu Wang, Rui Zhang, and Jingdian Ming. *On the security of lattice-based fiat-shamir signatures in the presence of randomness leakage,* 2020.

[2] Simon Damm, Nicolai Kraus, Alexander May, Julian Nowakowski, and Jonas Thietke. *One Bit to Rule Them All – Imperfect Randomness Harms Lattice Signatures,* 2025.

The attack of LZSWZM20[1] and DKMNT25[2]

From many (c, z, y_j) with y_j as the leaked j-th bit of y, recover $x \in \{-\eta, \dots, \eta\}^n$

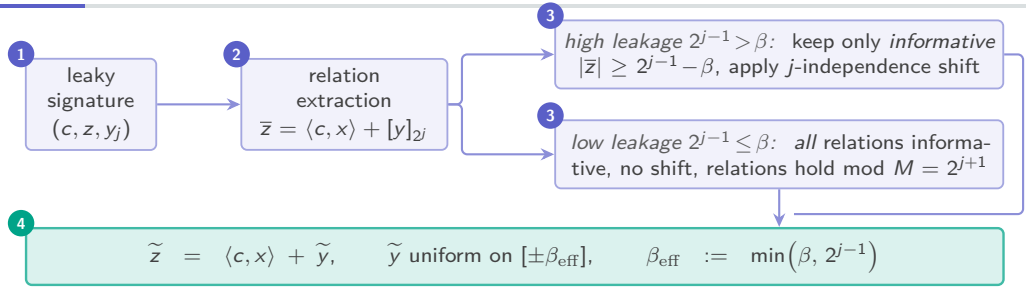


[1] Yuejun Liu, Yongbin Zhou, Shuo Sun, Tianyu Wang, Rui Zhang, and Jingdian Ming. *On the security of lattice-based fiat-shamir signatures in the presence of randomness leakage,* 2020.

[2] Simon Damm, Nicolai Kraus, Alexander May, Julian Nowakowski, and Jonas Thietke. *One Bit to Rule Them All – Imperfect Randomness Harms Lattice Signatures,* 2025.

The attack of LZSWZM20[1] and DKMNT25[2]

From many (c, z, y_j) with y_j as the leaked j-th bit of y, recover $x \in \{-\eta, \ldots, \eta\}^n$

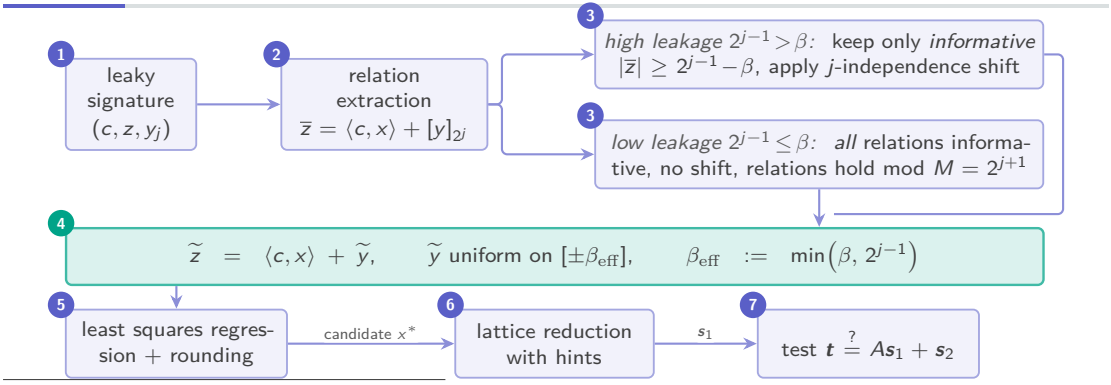


[1] Yuejun Liu, Yongbin Zhou, Shuo Sun, Tianyu Wang, Rui Zhang, and Jingdian Ming. *On the security of lattice-based fiat-shamir signatures in the presence of randomness leakage,* 2020.

[2] Simon Damm, Nicolai Kraus, Alexander May, Julian Nowakowski, and Jonas Thietke. *One Bit to Rule Them All – Imperfect Randomness Harms Lattice Signatures,* 2025.

The attack of LZSWZM20[1] and DKMNT25[2]

From many (c, z, y_j) with y_j as the leaked j-th bit of y, recover $x \in \{-\eta, \dots, \eta\}^n$

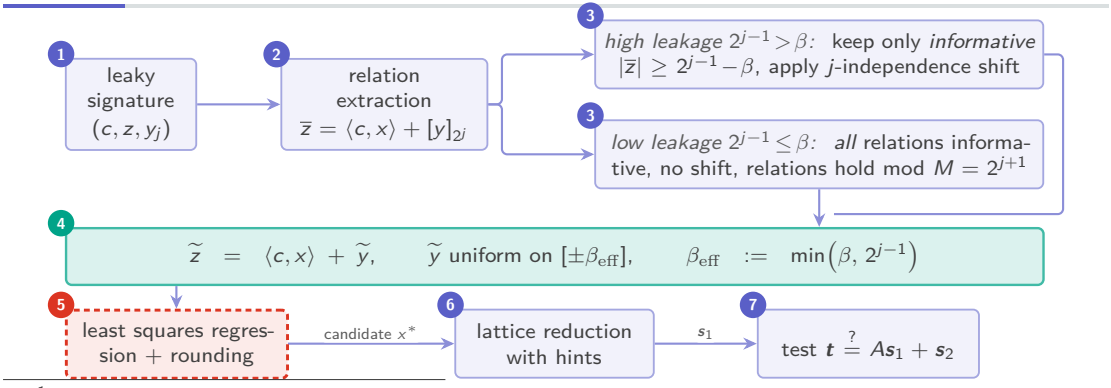


[1] Yuejun Liu, Yongbin Zhou, Shuo Sun, Tianyu Wang, Rui Zhang, and Jingdian Ming. *On the security of lattice-based fiat-shamir signatures in the presence of randomness leakage,* 2020.

[2] Simon Damm, Nicolai Kraus, Alexander May, Julian Nowakowski, and Jonas Thietke. *One Bit to Rule Them All – Imperfect Randomness Harms Lattice Signatures,* 2025.

The attack of LZSWZM20[1] and DKMNT25[2]

From many (c, z, y_j) with y_j as the leaked j-th bit of y, recover $x \in \{-\eta, \ldots, \eta\}^n$



[1] Yuejun Liu, Yongbin Zhou, Shuo Sun, Tianyu Wang, Rui Zhang, and Jingdian Ming. *On the security of lattice-based fiat-shamir signatures in the presence of randomness leakage,* 2020.

[2] Simon Damm, Nicolai Kraus, Alexander May, Julian Nowakowski, and Jonas Thietke. *One Bit to Rule Them All – Imperfect Randomness Harms Lattice Signatures,* 2025.

Step 5 is where the cost sits

The regression is polynomial — but it needs enough relations to average the mask away.

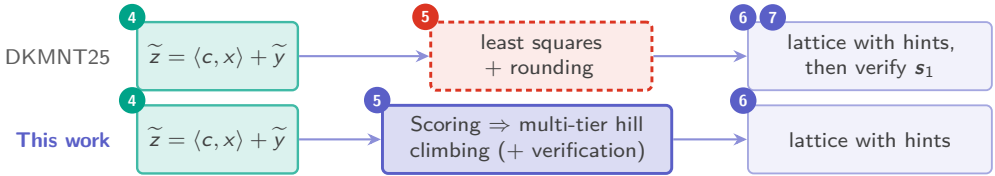
Informative relations needed at high leakage indices	ML-DSA-44	ML-DSA-65	ML-DSA-87
	0.5M	2.4M	0.75M

- Worse: by step 3, the number of *raw signatures* needed grows *exponentially* in j on top of that.
- Lowest leakage index reached in their work: $j = 6$.

Our central question

Can an attacker succeed with fewer informative relations by exploiting structure – **without touching steps 1–4**?

Our contributions: one box, swapped



1. **Subkey verification routine**: decide correctness of subkey with collected relations alone
2. **Scoring function** with monotonicity properties derived from that routine
3. **Hill-climbing attack algorithm** minimizing that score
4. Full analysis of noisy leakage in *both regimes*

Result. 37–68× fewer informative relations (exact, high leakage), $\geq 23\times$ (noisy data), key recovery down to leakage index $j = 4$ at ML-DSA-44 and -87.

Verification Routine

Can we check a single subkey?

Previously, a subkey x could not be verified before step 7 (after $\boldsymbol{s}_1$ is costly assembled).

Our observation: the relations of step 4 already form a verification oracle — making each x^* independently testable and wrong candidates much earlier discardable.

Idea. The true x satisfies *every* constraint $|\langle c_t, x\rangle - \widetilde{z}_t| \leq \beta_{\text{eff}}$ (the error bound)[3].
Wrong candidates violate some of them, with positive probability per relation t.

So we run this one-sided error check on *every one* of our α informative relations and a candidate has to survive all checks.

[3] In the low-leakage regime, this holds after a mod 2^{j+1} centered reduction.

The verification step

Algorithm 1: ML-DSA subkey verification step

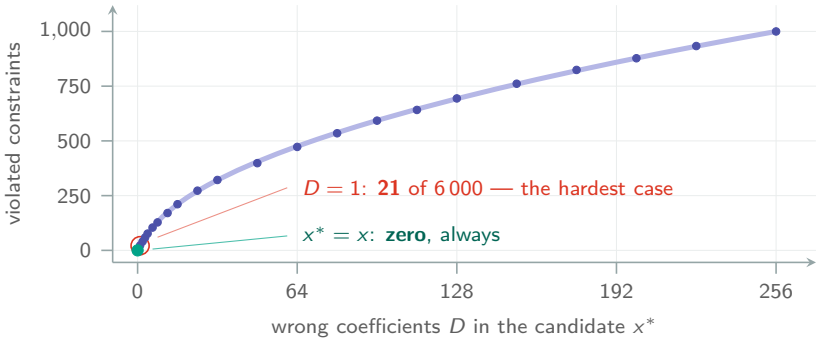
Input: preprocessed informative relation $(c, \widetilde{z})$, bound β_{eff}, index j, candidate x^*
Output: `False` if $x^* \neq x$ (with positive probability), `True` otherwise

1 $\delta \leftarrow \langle c, x^* \rangle - \widetilde{z}$;
2 **if** $2^{j-1} \leq \beta$ **then** $\delta \leftarrow [\delta]_{2^{j+1}}$ *// centered reduction, low-leakage regime*;
3 **if** $|\delta| > \beta_{\text{eff}}$ **then return** `False`;
4 **return** `True`;

- **No false negatives:** if $x^* = x$ then $\delta = \widetilde{y}$ and $|\widetilde{y}| \leq \beta_{\text{eff}}$.
- **One-sided error:** a wrong x^* survives a relation if either c misses the differing positions or $\widetilde{y}$ happens to absorb the deviation.
- Implementation note: precompute $\widetilde{z} \pm \beta_{\text{eff}}$; only $\langle c, x^* \rangle$ is evaluated per iteration (and the α checks are trivially parallel).

Does it catch a wrong candidate?

ML-DSA-44 at $j = 6$: how many of $\alpha = 6\,000$ checks fire at Hamming distance D?

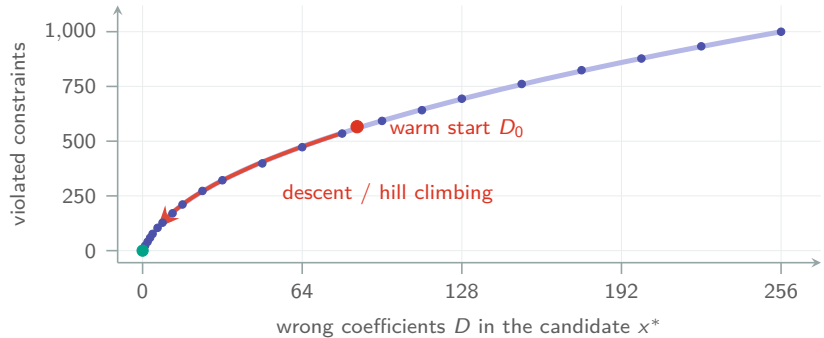


Takeaway. Each x can be verified given enough relations – a near-miss is the worst case. Certainty costs $\Theta(n \cdot \beta_{\text{eff}}/\beta)$ informative relations per security bit. Linear in n, cheaper at low j.

From Verification to a Metric on Keys

The same curve gives a hill descent

Read the same measurement as a landscape: the count *is* a score.



Takeaway. *In expectation*, correcting one position decreases the violation count. Key recovery is thus a task of repeated local optimization, warm-started from the rounded OLS estimate.

Two scores, one blind spot

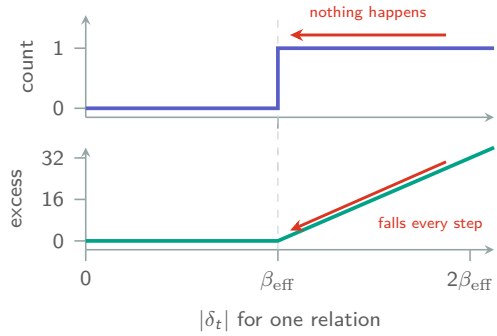
The hill was drawn by *counting*. We can also weigh each violation by how badly it misses.

Count

$$\left|\{t : |\delta_t| > \beta_{\text{eff}}\}\right|$$

Excess

$$\sum_t \max\left(0, |\delta_t| - \beta_{\text{eff}}\right)$$



Takeaway. Excess is the **stronger signal**: correcting a position can push violated constraints closer to satisfaction without satisfying them – nvisible to the count, but visible to excess. This sensitivity later backfires under noise. Minimizing either score is our new step 5.

Hill climbing — the naive version

Algorithm 2: Naive ML-DSA subkey finder

Input: α relations, β_{eff}, j, η, initial candidate x^*

for $i \in \{1, \ldots, n\}$ **do**
  for $v \in \{-\eta, \ldots, \eta\} \setminus \{x_i^*\}$ **do**
    $x' \leftarrow x^*$ with $x'_i \leftarrow v$;
    if $\text{score}(x') < \text{score}(x^*)$ **then** $x^* \leftarrow x'$, update score;
  end
end
return x^*;

Optimizations

Escaping local minima: an adaptive scheme

At the α we care about not every corrected position cleanly matches a lower score.
For this reason, we integrate additional hill-climbing tactics:

1. **Multiple sweep rounds.**
   Correcting position i reshapes the landscape at j—one pass is not enough.
2. **Increasingly larger block size w.**
   When sweeps stall, try changing w random positions simultaneously (instead of only one).
3. **Perturbation restarts.**
   After some time without progress reset 30 random positions and teleport into a new basin.
4. **Lateral moves.**
   Accept equal-score candidates to traverse plateaus.

Cheap strategies first; expensive ones only on demand.

Handling Noise

Noisy leakage model

Noisy Leaky-Signature-LWE

The attacker receives $y_j^{(p)} = y_j \oplus \zeta$ with $\zeta \sim \mathrm{Ber}(p)$, $p < \frac{1}{2}$.

Theorem (Effect of a flipped bit; DKMNT25[4])

$\overline{z}^{(f)} = \overline{z}^{(c)} - \mathrm{sign}\big(\overline{z}^{(c)}\big) \cdot 2^j$, *over* $\mathbb{Z}$, *in both regimes.*

$\tilde{z}^{(f)} = \tilde{z}^{(c)} - \mathrm{sign}\big(\tilde{z}^{(c)}\big) \cdot 2\beta$, *over* $\mathbb{Z}$, *in the high-leakage regime.*

A flip displaces the relation by exactly 2^j — huge compared to the error β_{eff}.

[4] One Bit to Rule Them All / Less Than a Bit to Rule Them All; Simon Damm, Nicolai Kraus, Alexander May, Julian Nowakowski, and Jonas Thietke; PKC 2025 / eprint 2025/820

Scoring and termination under noise

Excess-based scoring breaks

One flipped relation contributes up to $2\beta_{\text{eff}}$ of excess for the correct key. This drowns the signal and monotonicity is lost.

Count-based scoring survives

Every relation contributes at most $+1$. Noise adds $\approx p\alpha$ to *every* candidate alike, so the clean $(1-p)\,\alpha$ relations still separate them exactly as before.
Since $1-p > p$, the gap grows linearly in α.

Termination needs a new approach

In the exact setting we terminate at a score of 0. In the noisy setting we have to stop after a set number of steps without improvement.

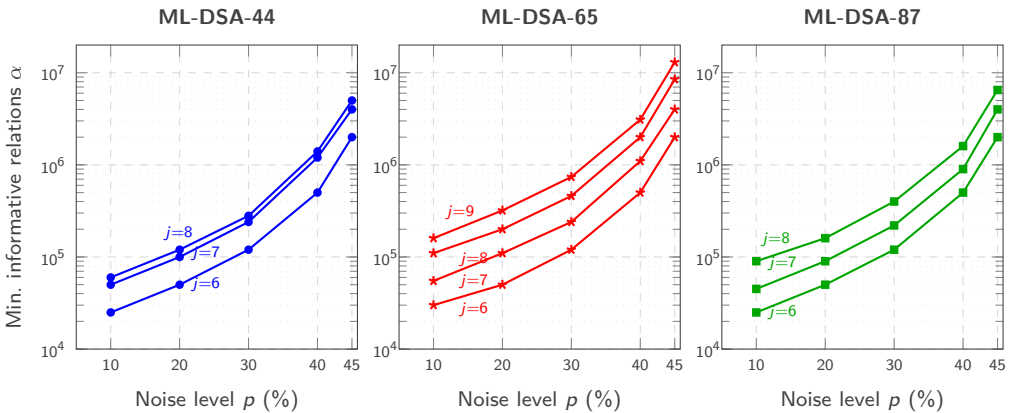
Results

Exact setting: minimum leakage requirements

Leakage j	Informative relations α (signatures N_{sig})		
	ML-DSA-44	ML-DSA-65	ML-DSA-87
4	1 500	—	2 500
5	3 000	3 000	2 500
6	6 000	5 500	5 000
7	11 500	11 500	9 500
8	13 500 (22 329)	22 500	17 500 (18 740)
9	14 500 (48 105)	35 000 (45 818)	17 500 (37 517)

- 10/10 keys recovered at each reported point.
- $\alpha = N_{\mathrm{sig}}$ at $j \leq 7$: in the low-leakage regime *every* signature counts.
- Recovery at $j = 4$ and $j = 5$ — regimes not reached by the previous regression attack.
- At high leakage indices, we improved by factors of **37.0**, **68.5** and **42.8**.

Noisy setting



Smallest α with $\geq 4/5$ successes over five independent keys.

Comparison with DKMNT25

Exact setting:

ML-DSA-44	$j = 6$	$j = 7$	$j \geq 8$
DKMNT25	90 000†	280 000†	500 000
This work	6 000	11 500	13 500
Factor	15.0†	24.3†	**37.0**
ML-DSA-65	$j = 7$	$j = 8$	$j \geq 9$
DKMNT25	375 000†	1 100 000†	2 400 000
This work	11 500	22 500	35 000
Factor	32.6†	48.9†	**68.5**
ML-DSA-87	$j = 6$	$j = 7$	$j \geq 8$
DKMNT25	75 000†	185 000†	750 000
This work	5 000	9 500	17 500
Factor	15.0†	19.5†	**42.8**

ML-DSA-44, $j = 8$, noisy setting:

Noise p	0.10	0.20	0.30	0.40
DKMNT25	1.4M	3.1M	9.0M	47M
This work	0.06M	0.12M	0.28M	1.4M
Factor	23.3	25.8	32.1	33.5

Additionally we recover the subkey for leakage indices as low as 4, reducing the minimum amount of needed signatures.

We also extent the highest noise rate from $p = 0.43$ to $p = 0.45$ while using less signatures in that case as DKMNT25 needed at $p = 0.4$.

Conclusion

Main takeaways

1. Subkey verification $\Rightarrow$ **scoring** $\Rightarrow$ key recovery as **local optimization**.
2. 37–68$\times$ (exact, high leakage) and $\geq 23\times$ (noisy, reported configs) fewer relations than regression; j as low as 4; noise successfully tested up to $p = 45\%$.
3. $\approx$ 1 500–3 000 leaky signatures enough to recover a subkey in most favorable setting

▷ **Research Continues**

Paskarbeit et al. *Statistical Inference from Noisy Randomness Leakage for ML-DSA Attacks.* (`https://eprint.iacr.org/2026/1712`)

Main takeaways

1. Subkey verification $\Rightarrow$ **scoring** $\Rightarrow$ key recovery as **local optimization**.
2. 37–68$\times$ (exact, high leakage) and $\geq 23\times$ (noisy, reported configs) fewer relations than regression; j as low as 4; noise successfully tested up to $p = 45\%$.
3. $\approx$ 1 500–3 000 leaky signatures enough to recover a subkey in most favorable setting

▷ **Research Continues**

Paskarbeit et al. *Statistical Inference from Noisy Randomness Leakage for ML-DSA Attacks.* (`https://eprint.iacr.org/2026/1712`)

Thank you! Any questions?

Backup B1: the monotonicity lemma

Lemma

If x^{**} arises from x^* by correcting *one* wrong position, then $p_{\text{detect}}(x^*) \geq p_{\text{detect}}(x^{**})$ — strictly, for all practically relevant parameter sets.

High-leakage regime ($\tilde{y} \xleftarrow{\$} [\pm\beta]$); $e' := x^* - x$, $\epsilon = \langle c, e' \rangle$:

$$p_{\text{miss}} = \frac{\left|[-\beta, \beta] \cap [\epsilon - \beta, \epsilon + \beta]\right|}{2\beta + 1} = 1 - \frac{|\epsilon|}{2\beta + 1} \implies p_{\text{detect}} = f(\epsilon) = \frac{|\epsilon|}{2\beta + 1}.$$

With $t := |\{k \neq i : c_k \neq 0\}|$: if $t = \tau$ then $c_i = 0$ and nothing changes; if $t = \tau - 1$ then $c_i = \pm 1$ uniformly and Jensen gives $\mathbb{E}_{c_i}[f] \geq f(\epsilon_{-i})$, since f is convex. Averaging over the remaining randomness of c gives the lemma.

Strictness: one challenge with $c_i \neq 0$ and $\epsilon_{-i} = 0$ via *cancellation pairs* ($|e'_j| = |e'_h|$); needs $\tau \leq n - 2\eta - 1$ and τ odd (or one correctly guessed position).

In the low-leakage regime f_{eff} saturates and is *not* convex, so Jensen does not apply directly; Prop. 1 of the paper carries the conclusion across. The hinge of the excess score *is* convex on all of $\mathbb{R}$, so there Jensen applies in both regimes without the extra argument.

Backup B2: the proved verification bound

Per-relation detection probability at Hamming distance 1:

$$\Pr[\text{detect}] \;\geq\; \underbrace{\frac{\tau}{n}}_{c \text{ hits the wrong position}} \cdot \underbrace{\frac{1}{2\beta+1}}_{\widetilde{y} \text{ does not absorb it}} \;=\; \left(2n\eta + \tfrac{n}{\tau}\right)^{-1}.$$

Theorem (Verification)

Given α random informative relations, one can decide $x^* \stackrel{?}{=} x$ in linear time with success probability at least $1 - \left(1 - (2n\eta + \frac{n}{\tau})^{-1}\right)^{\alpha}$, without any information about the remaining subkeys.

Where the slack in "$\geq$" comes from:

- it charges $|\epsilon| = 1$, but $\mathbb{E}|e'| = \frac{2(\eta+1)}{3}$ — factor 2 at $\eta = 2$;
- it uses β, not β_{eff} — factor β/β_{eff} whenever $\beta_{\text{eff}} < \beta$;
- against that: the low-leakage error is uniform on $[-2^{j-1}, 2^{j-1} - 1]$ and *not* symmetric — costs $\frac{1}{2}$ of $\mathbb{E}|e'|$ back.

ML-DSA-44 at $j = 6$: bound $9.7 \cdot 10^{-4}$, measured $3.6 \cdot 10^{-3}$ ($2 \cdot \frac{78}{32} \cdot \frac{3}{4} \approx 3.7$).

Backup B3: how much stronger is the excess?

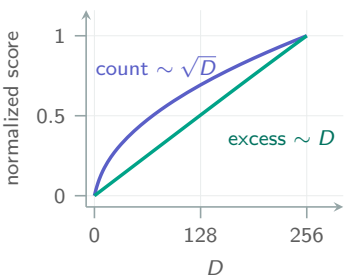
For one relation with deviation $\epsilon = \langle c, x^* - x \rangle$, writing W for the number of values $\widetilde{y}$ can take ($2\beta + 1$ high leakage, 2^j low):

$$\mathbb{E}\big[\text{count} \mid \epsilon\big] = \frac{|\epsilon|}{W}, \qquad \mathbb{E}\big[\text{excess} \mid \epsilon\big] = \frac{1 + 2 + \cdots + |\epsilon|}{W} = \frac{|\epsilon|\big(|\epsilon| + 1\big)}{2W}.$$

Linear vs. quadratic. Shifting the window by ϵ exposes $|\epsilon|$ values of $\widetilde{y}$: the count charges 1 for each, the excess charges $1, 2, \ldots, |\epsilon|$ — which squares the dependence.

Why the shapes differ. At distance D the challenge hits $\approx \tau D/n$ wrong positions and ϵ is their *signed* sum, so $\mathbb{E}|\epsilon| \sim \sqrt{D}$ while $\mathbb{E}[\epsilon^2] = \frac{\tau D}{n}\mathbb{E}[e'^2] \propto D$: magnitudes cancel, variances add.

So how much? Relative signal per corrected coefficient (ML-DSA-44, $j = 6$): $1.05\times$ at $D = 1$, $1.56\times$ at $D = 16$, $\mathbf{1.84}\times$ at the warm start $D_0 \approx 86$ — the advantage sits *far* from the solution, where the search stalls.



Backup B4: low-leakage residual under a bit flip

For $x^* = x$, a flipped relation and $\mathrm{sign}(\overline{z}^{(c)}) = +1$: $\delta^{(f)} = 2^j - [y]_{2^j}$, $M = 2^{j+1}$.

- $[y]_{2^j} > 0$: $\delta^{(f)} \in [2^{j-1} + 1, 2^j - 1]$, already centered.
- $[y]_{2^j} \leq 0$: $\delta^{(f)} \in [2^j, 3 \cdot 2^{j-1}]$, so $[\delta^{(f)}]_M = \delta^{(f)} - M \in [-2^j, -2^{j-1}]$.

In both cases $|[\delta^{(f)}]_M| = 2^j - |[y]_{2^j}| \in [2^{j-1}, 2^j]$, which exceeds $\beta_{\mathrm{eff}} = 2^{j-1}$ except for the single value $[y]_{2^j} = -2^{j-1}$.

Backup B5: experimental setup

- Python reference implementation, fixed seed 42; simulated key generation.
- Block size: start $w = 1$, $\rightarrow w = 2$ after one full sweep, $+1$ every 100 iterations, capped at $w = 4$ ($w = 3$ for $p \in \{0.1, \ldots, 0.4\}$). Improvements reset w.
- Perturbation: 30 positions, triggered after 150 consecutive iterations at max block size.
- Exact setting: sweep m^* downward in steps of 500; success $= 10/10$ keys.
- Noisy setting: patience 1 000 iterations; success $= 4/5$ keys; step sizes 5k/10k/20k/100k/500k for $p = 0.1/0.2/0.3/0.4/0.45$.
- Success also counts *ambiguous recovery*: ≤ 20 feasible solutions, true key among them.

Backup B6: ML-DSA parameters

Parameter	ML-DSA-44	ML-DSA-65	ML-DSA-87
(k, ℓ)	$(4, 4)$	$(6, 5)$	$(8, 7)$
η	2	4	2
τ	39	49	60
γ_1	2^{17}	2^{19}	2^{19}
$\beta = \tau\eta$	78	196	120
low-leakage regime ($2^{j-1} \leq \beta$)	$j \leq 7$	$j \leq 8$	$j \leq 7$

$n = 256$, $q = 8\,380\,417$ throughout.

Backup B7: Noise consequences per regime

High leakage ($2^{j-1} > \beta$)

- If $\overline{z}^{(c)}$ was *non*-informative: $|\overline{z}^{(f)}| > 2^{j-1} + \beta$ — impossible without noise $\Rightarrow$ identifiable, discard. (Bonus: estimates p for free.)
- If $\overline{z}^{(c)}$ was informative: the flip lands in the opposite half of the window, $\widetilde{z}^{(f)} = \widetilde{z}^{(c)} - \mathrm{sign}(\widetilde{z}^{(c)}) \cdot 2\beta$ — invisible.

Low leakage ($2^{j-1} \leq \beta$)

- All relations informative, everything mod $M = 2^{j+1}$: no value is impossible $\Rightarrow$ no filtering at all.
- Residual at the *correct* key: $\left|[\delta^{(f)}]_M\right| = 2^j - |[y]_{2^j}|$, violating the constraint with probability $\frac{2^j-1}{2^j}$ ($\approx 98.4\%$ at $j = 6$).

Common conclusion

In both regimes: a flipped relation almost always violates the constraint — *even for the correct key*. There is a noise floor.