AICR · OPEN PROBLEM

The Binary Dual-Distance Bound Is the Worst Case

Sparse-LPN parameters over a large field are chosen using a dual-distance bound computed over the binary field. Is that conservative?

Status. Statement: AI-written from Lennart Braun, Geoffroy Couteau, Kelsey Melissaris, Mahshid Riahinia and Elahe Sadeghi, *Fast Pseudorandom Correlation Functions from Sparse LPN*, IACR ePrint 2025/1644; ASIACRYPT 2025. Not yet checked by a human. Open, and unlike this page's two companions it is a claim the source asserts rather than a question it asks: the source writes that it expects the binary bound to lower-bound the bound over larger fields, gives an intuition, notes that all previous work assumes as much, and proves nothing. Both quantities are explicit finite sums, so the statement is falsifiable by computation — but only where that computation is feasible, which the source says the large-field one is not at the dimensions of interest.

Category. *Information-Theoretic*, *MPC*.

Conjecture (informal). *Every concrete parameter set for a sparse-LPN-based construction rests on one number: the dual distance of the random sparse matrix, the smallest weight of a non-zero vector in its kernel. Nobody computes it. What is computed is a lower bound on it, from a union bound over low-weight kernel vectors, and the computation is only tractable over the binary field, where cancellations are easy to count. Constructions over a larger field then use the binary number anyway, on the belief that a larger field can only help — cancellation needs matching supports* and *matching values. This conjecture is that belief made precise: the bound computed over the binary field never exceeds the bound over a larger field, for the same dimensions, sparsity and failure probability. It is a statement about two explicit finite sums, and every sparse-LPN parameter set over a large field is conservative only if it holds.*

1 The setting

Why the dual distance is the number that matters. Concrete parameters for LPN variants are not chosen by reduction; they are chosen by a two-step heuristic. First one argues that nothing about the special shape of the code matrix helps an attacker, then one prices generic LPN attacks. The first step is where the dual distance enters. Within the linear test framework of Boyle et al. [BCGIKS20] and Couteau, Rindal and Raghuraman [CRR21], an attack that distinguishes by a fixed linear test must run in time roughly $\exp(\mathsf{dd}(A) \cdot t/m)$, where $\mathsf{dd}(A)$ is the dual distance and t the noise weight, so $\mathsf{dd}(A) \cdot t/m$ is the quantity a parameter set is really targeting. The source paper's contribution here is a tighter way to bound $\mathsf{dd}(A)$: rather than the asymptotic estimates of Mossel, Shpilka and Trevisan [MST03] and their successors, it derives an explicit recursive formula for the probability that a weight-w vector lies in the kernel, sums it against a union bound, and computes the resulting D numerically.

The obstacle, and the belief. That computation lives over $\mathbb{F}_2$: "Our analysis above exploits properties of F2 (e.g., cancellations are more likely compared to a general field and only even-weight vectors matter). This allows us to run our implementation to run up to large values of n (see below), but makes it incompatible with general fields Fp." The source supplies a second family of formulae for a general prime field, and then reports why they are not used: "Note that the complexity of the computation is at least cubic in the LPN dimension n and quickly becomes infeasible for the large dimensions that we are interested in."

So the binary number is used for instances over larger fields, and the source states the belief that licenses this: "However, we expect the bound obtained through our analysis over F2 to provide a lower bound on that over larger fields." Its intuition: "Intuitively, as the field grows, canceling rows gets harder; it no longer suffices for balls to land in the same bins but they must have opposite values." And its account of the practice: "All previous works assume that the bound does not degrade as the field size grows and a common heuristic is, therefore, to use the bounds from the analysis over F2" [ADINZ17, BCGI18].

What is at stake. If the conjecture holds, every large-field parameter set chosen this way is conservative, and the source's Table 2 can be read as a safe floor for any $\mathbb{F}_p$. If it fails — at parameters anyone uses — then those parameter sets overstate the dual distance they achieve, and with it the linear-test security level $\mathrm{dd}(A) \cdot t/m$ that the whole selection procedure targets.

2 Notation and parameters

Fix integers $3 \le k \le n$ and $m \ge 1$, a prime p, and a statistical security parameter $\rho > 0$.

- $\omega(v)$ is the Hamming weight of v, and $\mathrm{dd}(A) := \min\{\omega(v) : v \neq 0,\ v^{\top}A = 0\}$ is the *dual distance* of A.
- $\mathsf{SparseCodeGen}(k, m, n, \mathbb{F}_p)$ outputs a uniformly random matrix in $\mathbb{F}_p^{m\times n}$ each of whose rows has exactly k non-zero entries (the source's Definition 3).
- $\mathsf{RegularCodeGen}(k, m, n, \mathbb{F}_2)$, defined when $k \mid n$, outputs a matrix in $\mathbb{F}_2^{m\times n}$ each of whose rows is the concatenation of k blocks of length n/k, each block a uniformly random unit vector (the source's Definition 4).
- q_w is the probability that the sum of $2w$ independent uniformly random unit vectors of length $N := n/k$ over $\mathbb{F}_2$ is zero. This is the source's $p_{w,0}$ of its Section 4.3.
- $\pi_w^{(p)}$ is the probability that the sum of w independent uniformly random weight-k vectors in $\mathbb{F}_p^n$ is zero. This is the source's $p_{w,0}$ of its Section 4.5.

Definition 1 (The binary bound, the source's Section 4.3).

$$D_2(k, m, n, \rho) := \max\left\{ d > 0 \ : \ \sum_{w=1}^{\lfloor (d-1)/2 \rfloor} \binom{m}{2w} (q_w)^k \le 2^{-\rho} \right\}.$$

This is the quantity the source computes for $A \leftarrow \mathsf{RegularCodeGen}(k, m, n, \mathbb{F}_2)$ and tabulates in its Table 2; it is a lower bound on $\mathrm{dd}(A)$ except with probability $2^{-\rho}$.

Definition 2 (The general-field bound, the source's Section

4·5)·

$$D_p(k,m,n,\rho) := \max\Big\{d > 0 \;:\; \sum_{w=1}^{d-1} \binom{m}{w} \pi_w^{(p)}\,(p-1)^w \le 2^{-\rho}\Big\}.$$

This is the quantity the source's Section 4.5 defines for $A \leftarrow \mathsf{SparseCodeGen}(k,m,n,\mathbb{F}_p)$, and does not evaluate at the dimensions its construction uses.

In both definitions, if no $d > 0$ satisfies the constraint then the value is 0, following the source's own convention for its table: "D = 0 indicates that we can only guarantee a trivial bound".

3 The conjecture

Conjecture 1 (The binary bound is the worst case). *For every prime $p > 2$, every $\rho > 0$, and all integers $3 \le k \le n$ with $k \mid n$ and $m \ge 1$,*

$$D_2(k,m,n,\rho) \;\le\; D_p(k,m,n,\rho),$$

with D_2 as in Definition 1 and D_p as in Definition 2.

Remark 1 (The two bounds are computed for different ensembles). This is the first thing to check about the statement, and it is a feature of the source's sentence rather than of this transcription. D_2 is computed for the *regular* ensemble, one unit vector per block; D_p for the *sparse* ensemble, an arbitrary weight-k row. The source gives an extension of its binary analysis to the sparse ensemble in one line — replace $(q_w)^k$ by the probability that $2wk$ unit vectors of length n sum to zero — so a variant of the conjecture with both sides on the sparse ensemble is available and is arguably the more natural claim. It is not the one the source's sentence makes, so it is not the one stated above. Note also the structural consequence of the difference: over the regular ensemble an odd number of rows can never cancel, so D_2's sum runs over even weights only, while D_p's runs over all weights.

Remark 2 (Two effects pull against each other). The source's

intuition covers one of them. Raising p shrinks $\pi_w^{(p)}$, since supports must now match *and* values must cancel. But raising p also multiplies the number of candidate kernel vectors of each support by $(p-1)^w$, and that factor appears in D_p's summand. The conjecture is the assertion that the first effect wins, uniformly in the other parameters, and nothing in the source addresses the second.

Remark 3 (What is checkable, and what this page checked). Both sides are explicit finite sums of rational numbers, so any instance is decidable by computation; the source's own script is public for the $\mathbb{F}_2$ side. That makes the conjecture falsifiable in the strongest sense — a single parameter triple would do it — and it is why the statement is stated for *all* parameters rather than asymptotically. The obstruction is cost, exactly as the source says: the $\mathbb{F}_p$ side is at least cubic in n, so the dimensions where the source's construction lives, $n = 2^{15}$ and up, are out of reach of a direct check.

4 Bibliography

[ADINZ17] Benny Applebaum, Ivan Damgård, Yuval Ishai, Michael Nielsen and Lior Zichron. *Secure arithmetic computation with constant computational overhead.* CRYPTO 2017, Part I. Cited by the source as one of the works that use the binary bound for a larger field.

[BCGI18] Elette Boyle, Geoffroy Couteau, Niv Gilboa and Yuval Ishai. *Compressing vector OLE.* ACM CCS 2018. Likewise.

[BCGIKS20] Elette Boyle, Geoffroy Couteau, Niv Gilboa, Yuval Ishai, Lisa Kohl and Peter Scholl. *Correlated pseudorandom functions from variable-density LPN.* 61st Annual Symposium on Foundations of Computer Science (FOCS), 2020. Introduces the linear test framework in which the dual distance is priced.

[BCM25] Dung Bui, Geoffroy Couteau and Nikolas Melissaris. *Structured-seed local pseudorandom generators and their applications.* APPROX/RANDOM 2025. Source of the asymptotic large-dual-distance lemma the source reproduces, and of a quantified form of the "high dual distance suffices" conjecture.

[BCMRS25] Lennart Braun, Geoffroy Couteau, Kelsey Melissaris, Mahshid Riahinia and Elahe Sadeghi. *Fast pseudorandom correlation functions from sparse LPN.* IACR ePrint 2025/1644; ASIACRYPT 2025. The source: the claim and its intuition are on p. 26, the binary bound on pp. 24–26,

Table 2 on p. 28, the general-field formulae in Section 4.5 on pp. 30–32, and the cost remark on p. 32.

[CRR21] Geoffroy Couteau, Peter Rindal and Srinivasan Raghuraman. *Silver: silent VOLE and oblivious transfer from hardness of decoding structured LDPC codes.* CRYPTO 2021, Part III. The general heuristic for LPN variants, and asymptotic dual-distance bounds for arbitrary fields.

[Kha26] Majid Khabbazian. *Linear distance for fixed-row-weight expand-accumulate codes over arbitrary fields.* IACR ePrint 2026/1753. A different code ensemble, and the nearest available data point on whether bounds of this kind are field-uniform: it proves a conjectured constant-relative-distance bound for expand-accumulate codes "in a stronger, field-uniform form", with "the same constants" for every prime power. [UNVERIFIED]

[MST03] Elchanan Mossel, Amir Shpilka and Luca Trevisan. *On ϵ-biased generators in NC0.* 44th Annual Symposium on Foundations of Computer Science (FOCS), 2003. The original analysis of the dual distance of random sparse matrices.