AICR · OPEN PROBLEM

A Subexponential-Query PCF from Sparse LPN

Can a pseudorandom correlation function based on sparse LPN stay secure against an adversary that asks for subexponentially many correlations, rather than only superpolynomially many?

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: the source poses it as the first of its two open questions and takes no position on the answer, so the direction fixed below is this page's reading and not the source's claim. Nothing is settled in either direction; the one follow-up construction from sparse LPN inherits the same superpolynomial ceiling.

Category. *MPC*, *Foundations*.

Conjecture (informal). *A pseudorandom correlation function lets two parties turn a pair of short keys into as much correlated randomness — oblivious transfers, vector-OLEs, Beaver triples — as they will ever need, with no further interaction. The source paper gives the first such construction that is both fast and based on a well-studied assumption, sparse LPN, but pays for it in the one place asymptotics can see: its security survives an adversary that runs for subexponential time only as long as that adversary asks for at most superpolynomially many correlations, where the earlier candidates from newer assumptions plausibly tolerated subexponentially many. This conjecture says the gap is an artefact of the construction rather than of the assumption: some PCF based on sparse LPN is secure against subexponentially many queries as well as subexponential time.*

1 The setting

A *pseudorandom correlation function* (PCF) for a correlation $\mathcal{Y}$ is a pair of algorithms $(\mathsf{Gen}, \mathsf{Eval})$: $\mathsf{Gen}(1^\lambda)$ outputs two keys, and $\mathsf{Eval}(\sigma, k_\sigma, x)$ is deterministic, so that on any common input x the

two parties' outputs look like a fresh sample from $\mathcal{Y}$, and each party's output looks random given the other's key [BCGIKS20]. The quantitative version, which is the one this page is about, is Definition 2: security is graded by three functions, a bound B on the adversary's size, a bound N on how many inputs it may query, and the advantage ε it is allowed.

Every construction in this line is parameterized by an LPN-style assumption, and the source paper's is *sparse* LPN: the code matrix is sampled with k non-zero entries per row, an assumption introduced by Alekhnovich [Ale03] whose cryptanalysis has been studied for two decades. That is what makes the paper's construction the first efficient PCF from an assumption nobody had to invent for the purpose. Its predecessors bought their efficiency with new assumptions — variable-density LPN [BCGIKS20, CD23] and expand-accumulate LPN [BCGIKRS22] — or their standardness with unusable speed, one OT per second from quadratic residuosity [OSY21].

What the source achieves, and where it stops. The construction iterates the primal PCG of Boyle et al. [BCGI18] L times, so that a short seed expands through intermediate sparse-LPN instances $s^{(\ell)} = A^{(\ell)} s^{(\ell-1)} + e^{(\ell)}$, with the rows of each $A^{(\ell)}$ sampled on the fly from a random oracle. Two constraints fight each other in the choice of L. Evaluating one output entry costs k^L operations, and the paper's first parameter constraint is that this stay polynomial, since "evaluation must run in polynomial time"; meanwhile the number of queries the top-level matrix can absorb is the number of its rows. With the parameters the paper settles on, it reports that the PCF is secure for up to $m^{(L)} = n^{O(\log n/\log\log n)}$ queries, and summarizes: "these parameters achieve security with inverse-superpolynomial advantage against subexponential-time adversaries making at most superpolynomially-many queries to the PCF".

The source states the resulting gap as a limitation of its own result: "Compared with previous constructions from VDLPN, EALPN, DCR, or QR, our PCF comes with a theoretical limitation: while previous constructions were plausibly secure against a (subexponential-time) adversary requesting a subexponential number of PCF evaluations," its own is not. It then asks, under the heading "Open Questions. We mention two interesting open questions:", the question this page records: "Is it possible to build a PCF with fully-subexponential

security (in runtime and queries) from sparse LPN? Our PCF only achieves subexponential security up to superpolynomially-many queries, a lower asymptotic security level compared to previous candidates."

What this page fixes and the source does not. The source asks a question; a page here has to state a proposition, so Conjecture 1 fixes the affirmative direction. That choice is this page's, not the source's, and a refutation — a proof that no PCF based on sparse LPN can tolerate subexponentially many queries — would settle the source's question just as well. Three further readings are this page's own and are the first things a reviewer should challenge: that "from sparse LPN" means the assumption of Definition 1 with a standard noise distribution and no new variant; that the random oracle model remains available, since removing it is the source's *second* open question and not this one; and that "fully-subexponential in runtime and queries" is the reading in Conjecture 1, namely that a single exponent δ serves for the adversary's size, its query count and its advantage.

2 Notation and parameters

- λ is the security parameter; n is the LPN dimension, m the number of samples, k the row sparsity of the code matrix and t the noise weight. All are functions of λ.
- $\mathbb{F}_p$ is a finite field and $\mathbb{F}_{p^r}$ an extension of it.
- $\mathsf{SparseCodeGen}(k, m, n, \mathbb{F})$ outputs a uniformly random matrix in $\mathbb{F}^{m\times n}$ whose every row has exactly k non-zero entries.
- $\mathsf{Reg}_t(m, \mathbb{F})$ is the regular noise distribution: a vector in $\mathbb{F}^m$ formed by concatenating t blocks of length m/t, each a uniformly random unit vector.
- A *subfield VOLE correlation* of dimension n over $\mathbb{F}_{p^r}$ is a tuple $((u, v), (\Delta, w))$ with $u \in \mathbb{F}_p^n$, $v, w \in \mathbb{F}_{p^r}^n$, $\Delta \in \mathbb{F}_{p^r}$ and $w - v = \Delta \cdot u$.
- poly, negl and polylog have their usual meanings.

Definition 1 (Sparse LPN, the source's Definition 5). $\mathsf{SparseLPN}(k, n, m, \mathsf{Reg}_t, \mathbb{F})$ is the assumption that

$$(A, A \cdot s + e) \approx_c (A, y),$$

where $A \leftarrow \mathsf{SparseCodeGen}(k,m,n,\mathbb{F})$, $s \leftarrow \mathbb{F}^n$, $e \leftarrow \mathsf{Reg}_t(m,\mathbb{F})$ and $y \leftarrow \mathbb{F}^m$. For a bound $B : \mathbb{N} \to \mathbb{N}$ and $\varepsilon : \mathbb{N} \to [0,1]$, say the assumption is (B,ε)-hard if every adversary of size $B(\lambda)$ distinguishes the two distributions with advantage at most $\varepsilon(\lambda)$.

Definition 2 (PCF, the source's Definition 9, after [BCGIKS20]). Let $\mathcal{Y}$ be a reverse-sampleable correlation with output lengths ℓ_0, ℓ_1 and input length $n(\lambda)$. A pair $(\mathsf{Gen}, \mathsf{Eval})$, with $\mathsf{Gen}(1^\lambda)$ a PPT algorithm outputting keys (k_0, k_1) and $\mathsf{Eval}(\sigma, k_\sigma, x)$ a deterministic polynomial-time algorithm on inputs $x \in \{0,1\}^{n(\lambda)}$, is an (N,B,ε)-secure *strong* PCF for $\mathcal{Y}$ if for every non-uniform adversary $\mathcal{A}$ of size $B(\lambda)$ making at most $N(\lambda)$ oracle queries, and all large enough λ:

- *pseudorandom $\mathcal{Y}$-correlated outputs*: $\mathcal{A}$ distinguishes an oracle answering each fresh query x with $(\mathsf{Eval}(0,k_0,x), \mathsf{Eval}(1,k_1,x))$ from one answering with a fresh sample of $\mathcal{Y}$, with advantage at most $\varepsilon(\lambda)$;
- *security*: for each $\sigma \in \{0,1\}$, given k_σ, $\mathcal{A}$ distinguishes $\mathsf{Eval}(1-\sigma, k_{1-\sigma}, x)$ from $\mathcal{Y}.\mathsf{RSample}(\sigma, \mathsf{Eval}(\sigma, k_\sigma, x))$ with advantage at most $\varepsilon(\lambda)$.

A *weak* PCF is the same with both oracles queried at uniformly random inputs rather than at inputs of $\mathcal{A}$'s choosing.

3 The conjecture

Conjecture 1 (Subexponential-query PCF from sparse LPN). *There exist a constant $\delta > 0$, a field $\mathbb{F}_{p^r}$, parameter functions $n(\lambda) = \mathrm{poly}(\lambda)$, $m(\lambda)$, $t(\lambda)$ and $k(\lambda)$ with $3 \le k(\lambda) \le n(\lambda)$, and a pair of algorithms* $(\mathsf{Gen}, \mathsf{Eval})$ *with keys of size* $\mathrm{poly}(\lambda)$*, such that* $(\mathsf{Gen}, \mathsf{Eval})$ *is an (N, B, ε)-secure strong PCF for the subfield VOLE correlation over $\mathbb{F}_{p^r}$ (Definition 2) with*

$$N(\lambda) = B(\lambda) = 2^{\lambda^\delta}, \qquad \varepsilon(\lambda) = 2^{-\lambda^\delta},$$

assuming only that $\mathsf{SparseLPN}(k, n, m, \mathsf{Reg}_t, \mathbb{F}_p)$ *of Definition 1 is $(2^{\lambda^{\delta'}}, 2^{-\lambda^{\delta'}})$-hard for some constant $\delta' > 0$, together with a random oracle.*

Remark 1 (What the quantifiers are doing). δ is existentially quantified, so the conjecture asks for *some* subexponential query bound and not for a particular exponent; this is the weakest reading of "fully-subexponential" that still separates the statement from what the source proves, where $N(\lambda)$ is capped at $n^{O(\log n / \log\log n)}$ for a polynomially-bounded n and is therefore superpolynomial but not 2^{λ^δ} for any $\delta > 0$. The correlation is fixed to subfield VOLE because that is the source's headline correlation; the OT correlation, which the source also achieves, would be an equally faithful reading and is not a different problem.

Two further choices are worth arguing with. First, the source's phrase constrains *runtime and queries*, and says nothing about the advantage; its own construction achieves "inverse-superpolynomial advantage", not $2^{-\lambda^\delta}$. The statement above asks for a single exponent to serve all three, which is the symmetric reading; a reader who wants only what the source's phrase demands should replace ε by an arbitrary negligible function, giving a weaker conjecture that this one implies. Second, the sparsity k is left free within Definition 1's range rather than pinned to $\mathrm{polylog}(n)$. The source's own conservative flavour of the assumption does take $k = \mathrm{polylog}(n)$, so a construction that needed a much larger sparsity would answer this statement while answering the source's question less well; pinning it here would instead make the conjecture strictly harder than the question asked.

Remark 2 (Why the obvious route stalls). Neither the source nor

this page knows an obstruction, and this is the weakest point of the problem as posed: the source names no barrier, only its own shortfall. What can be said is where the parameters bind. In the source's recursion the query bound is the row count $m^{(L)}$ of the outermost matrix, which grows with the number of levels L, while evaluation of a single output costs k^L operations, so polynomial evaluation time forces $k^L = \mathrm{poly}(n)$ and hence $L = O(\log n / \log k)$. With $k = \mathrm{polylog}(n)$ that ceiling on L is what makes $m^{(L)}$ superpolynomial rather than subexponential. Escaping it means either evaluating an entry of a subexponentially long pseudorandom correlation in polynomial time without a subexponential-depth recursion, or a route to a PCF from sparse LPN that is not this recursion at all.

4 Bibliography

[Ale03] Michael Alekhnovich. *More on average case vs approximation complexity.* 44th Annual Symposium on Foundations of Computer Science (FOCS), 2003.

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

[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. The PCF notion and the template every later candidate follows.

[BCGIKRS22] Elette Boyle, Geoffroy Couteau, Niv Gilboa, Yuval Ishai, Lisa Kohl, Nicolas Resch and Peter Scholl. *Correlated pseudorandomness from expand-accumulate codes.* CRYPTO 2022, Part II.

[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 of this conjecture: the two open questions are on p. 10, the limitation on p. 5, the parameter constraints and the resulting query bound on p. 16, and Definitions 5 and 9 on pp. 11 and 13.

[CD23] Geoffroy Couteau and Clément Ducros. *Pseudorandom correlation functions from variable-density LPN, revisited.* PKC 2023, Part II.

[HR25] Sebastian Hasler and Pascal Reisert. *Pseudorandom correlation functions for multiparty Beaver triples from sparse LPN.* IACR ePrint 2025/2002, December 2025. The one follow-up that reuses the source's recursion,

and records the same ceiling: “the expansion is still asymptotically smaller than the subexponential expansion of previous PCFs for Beaver triples”. [UNVERIFIED]

[OSY21] Claudio Orlandi, Peter Scholl and Sophia Yakoubov. *The rise of Paillier: homomorphic secret sharing and public-key silent OT*. EUROCRYPT 2021, Part I.