CONJURA · OPEN PROBLEM

Minimum Distance of Expand-Accumulate Codes over Arbitrary Rings

Proved over $\mathbb{F}_2$; conjectured not to degrade, and to improve, over larger rings

Status. Statement: AI-written from Elette Boyle, Geoffroy Couteau, Niv Gilboa, Yuval Ishai, Lisa Kohl, Nicolas Resch and Peter Scholl, *Correlated Pseudorandomness from Expand-Accumulate Codes*, IACR ePrint 2022/1014. Not yet checked by a human. The source's Theorem 3.10 proves the distance bound over the binary field only. Over a ring of size q its proof technique forces the rate to satisfy $R \ll 1/\ln q$, which it states it believes is an artefact, and it supports that belief with small-parameter weight-distribution experiments over $\mathbb{F}_{17}$ and $\mathbb{Z}_{16}$.

Category. *Information-Theoretic Cryptography.*

Conjecture (informal). *An expand-accumulate code is about the simplest linear code that still behaves well: multiply by a sparse random matrix, then take running prefix sums. The second step costs one addition per coordinate, which is why the construction is fast enough to underlie pseudorandom correlation generators for secure computation. What such an application needs is that the code have large minimum distance, and the source proves that over the binary field. Over a larger ring its proof degrades in a suspicious way: the concentration bound it uses forces the rate to shrink like $1/\ln q$ in the ring size. The authors conjecture the distance does not in fact degrade over larger rings — indeed that it improves — and that the rate restriction is an artefact of the argument rather than of the codes.*

1 The setting

Pseudorandom correlation generators compress a long stream of correlated randomness for secure computation into short seeds, and the dual-LPN assumptions they rest on are parameterized by a linear code. What the code must supply is minimum distance: a low-weight dual codeword is exactly a linear test that breaks the

assumption, so the distance controls resistance to that whole class of attack. What the application also needs is speed, since the code's encoding sits on the critical path.

The source's answer to both is the expand-accumulate family. The accumulator matrix A has ones on and below the diagonal, so computing $A\vec{x}$ takes $N-1$ sequential ring additions — over $\mathbb{F}_2$, $N-1$ xors — and parallelizes well. An EA generator matrix is BA for a sparse random B whose rows are independent Bernoulli vectors of density p (Definition 2).

Over the binary field the source proves the distance bound it needs. Its Theorem 3.10 fixes rate $R = n/N$ constant, sets $p = C \ln N/N$, takes $\delta \in (0,1/2)$ and $\beta = 1/2-\delta$, and shows that for sufficiently large N the code sampled by $\mathsf{EAGen}(n,N,p)$ has minimum distance at least δN except with probability $2\sum_{r=1}^{n}\binom{n}{r}\exp(-2\frac{1-\xi_r}{1+\xi_r})$ where $\xi_r = (1-2p)^r$, provided R and C satisfy two explicit inequalities, among them $C > 1/\beta^2$.

That theorem is binary-only, and the source says why the extension resists: "the expander Hoeffding bound that we apply does not appear to work well in this case; in particular, we are forced to require the rate $R \ll \frac{1}{\ln q}$, where q is the size of the ring $\mathcal{R}$." It then states its belief — "However, we believe that this is an artefact of the proof" — and justifies it by computing average weight distributions for small codes sampled over $\mathbb{F}_{17}$ and over $\mathbb{Z}_{16}$, alongside its comparison of row distributions over $\mathbb{F}_2$.

2 Notation and parameters

- $\mathcal{R}$ is a ring, $q = |\mathcal{R}|$; the binary case is $\mathcal{R} = \mathbb{F}_2$.
- $n \le N$ are the code's dimension and blocklength, $R = n/N$ the rate, assumed constant.
- $p \in (0,1)$ is the density of the sparse matrix B; the source's setting is $p = C \ln N/N$ for a constant $C > 0$.
- $A \in \mathcal{R}^{N\times N}$ is the accumulator matrix, with ones on and below the main diagonal and zeros elsewhere.
- $\mathsf{Ber}_p^N(\mathcal{R})$ is the distribution on $\mathcal{R}^N$ with each coordinate independently non-zero with probability p, as in the source.
- $d(H)$ is the minimum distance of the code generated by H; $\delta \in (0,1/2)$ is the relative distance target and $\beta = 1/2-\delta$.

3 The conjecture

Definition 1 (Accumulator matrix, [BCGIKRS22, Definition 3.1]). For $N \in \mathbb{N}$ and a ring $\mathcal{R}$, the accumulator matrix $A \in \mathcal{R}^{N\times N}$ has ones on and below the main diagonal and zeros elsewhere; so $A\vec{x} = \vec{y}$ means $y_i = \sum_{j\le i} x_j$ for all i, computable with $N-1$ sequential ring additions.

Definition 2 (Expand-accumulate code, [BCGIKRS22, Definition 3.2]). For $n \le N$, a ring $\mathcal{R}$ and density $p \in (0,1)$, sample row vectors $\vec{r}_1^{\top}, \ldots, \vec{r}_n^{\top} \leftarrow \mathrm{Ber}_p^N(\mathcal{R})$ independently, let B be the matrix with those rows, and output $H = BA$. Write $H \leftarrow \mathsf{EAGen}(n, N, p, \mathcal{R})$; when $\mathcal{R}$ is omitted it is $\mathbb{F}_2$.

Theorem 1 (What is proved, over $\mathbb{F}_2$ only, [BCGIKRS22, Theorem 3.10]). *Let $n \le N$ with $R = n/N$ constant, let $C > 0$ and $p = C \ln N/N \in (0,1/2)$, fix $\delta \in (0,1/2)$ and put $\beta = 1/2 - \delta$. Then for sufficiently large N,*

$$\Pr\left[d(H) \ge \delta N \,\middle|\, H \leftarrow \mathsf{EAGen}(n,N,p)\right] \;\ge\; 1 - 2\sum_{r=1}^{n} \binom{n}{r} \exp\left(-2\,\frac{1-\xi_r}{1+\xi_r}\right),$$

where $\xi_r = (1-2p)^r$, provided R and C satisfy the source's two stated inequalities, including $C > 1/\beta^2$.

Conjecture 1 (Distance does not degrade over larger rings). *Theorem 1 holds with $\mathbb{F}_2$ replaced by an arbitrary ring $\mathcal{R}$, with no restriction of the rate R in terms of $q = |\mathcal{R}|$: that is, under the same conditions on R, C, δ as in the binary case, a code sampled from $\mathsf{EAGen}(n, N, p, \mathcal{R})$ has minimum distance at least δN with at least the same probability.*

Remark 1 (The stronger form the source also believes). The source states two things, and Conjecture 1 is the weaker: "We conjecture that the minimum distance should at the very least not degrade over larger rings; in fact, we believe that it should increase commensurately." The second clause — that the distance *improves* with q — is not quantified anywhere, so it is recorded here as the source's belief rather than promoted to the statement. A resolution

establishing an improvement should report the improvement it gets.

Remark 2 (Where exactly the proof breaks). The named obstruction is the expander Hoeffding bound. Over $\mathbb{F}_2$ the argument bounds, for each weight class, the probability that a codeword falls below the distance target, using concentration for a sum indexed by the sparse rows. Over a ring of size q the same bound is available only after shrinking the rate to $R \ll 1/\ln q$, which is what makes the theorem vacuous for large rings at constant rate. So the substance of Conjecture 1 is not a new distance argument from scratch but a concentration inequality that does not pay $\ln q$ — or a different route to the same conclusion.

Remark 3 (Why larger rings should if anything help). The intuition the source offers, and which its small-parameter experiments support, is that a non-zero entry of the sparse matrix carries more information over a larger ring, so cancellations that produce low-weight codewords become rarer rather than more common. Its experiments compare codes sampled by $\mathsf{EAGen}(4,20,4,\mathbb{F}_{17})$ and $\mathsf{EAGen}(4,20,4,\mathbb{Z}_{16})$ — deliberately one field and one ring with zero divisors, since the latter is where cancellation is easiest and a counterexample would most plausibly live.

Remark 4 (A separate open problem in the same discussion). Immediately before the passage this statement comes from, the source compares row distributions for B over $\mathbb{F}_2$ — Bernoulli, exact weight, regular — reports that regular noise appears best in putting the most mass on heavier weights, and says "[p]roviding theoretical guarantees for these other row distributions remains an open problem meriting further study." That is a different question, about which distribution on B to analyse rather than which ring to work over, and it is not this statement.

4 Bibliography

[BCGIKRS22] Elette Boyle, Geoffroy Couteau, Niv Gilboa, Yuval Ishai, Lisa Kohl, Nicolas Resch and Peter Scholl. *Correlated pseudorandomness from expand-accumulate codes.* IACR ePrint 2022/1014. The source. Definitions 3.1 and 3.2 are on pages 13–14, Theorem 3.10 on page 18, and the arbitrary-rings discussion with the conjecture on page 21.

[BCG+19b] Elette Boyle, Geoffroy Couteau, Niv Gilboa, Yuval Ishai, Lisa Kohl and Peter Scholl. *Efficient pseudorandom correlation generators: silent OT extension and more.* CRYPTO 2019, Part III. The pseudorandom correlation generator line the source's codes are built for.