Rare-Event Decorrelation for Linear Tests of Low-Complexity Pseudorandom Functions

Nikolas Melissaris*

IRIF, CNRS & Université Paris Cité
nikolas@irif.fr

Abstract

Ball et al. left open whether their bounded-query strong-PRF candidate in $\mathrm{AC}^0[2]$ resists general non-adaptive linear tests with polynomially many chosen queries. Boyle et al. left open whether their Sipser-based weak-PRF candidate resists linear attacks. We analyze both candidates using the rare-event structure of their predicates. Repeated applications of Cauchy–Schwarz reduce the correlations to parities of rarer events while preserving the available independence.

For the candidate of Ball et al., every fixed polynomial number of distinct chosen queries has linear-test bias $\exp(-\Omega(\lambda))$, with arbitrary linear or affine dependencies. More generally, for every fixed $a \in (0,1)$, $q \le 2^{a(\log_2 \lambda)^2}$ queries have bias $\exp(-\lambda^{1-a-o(1)})$. For the Sipser-based family of Boyle et al., for every fixed $0 < \varepsilon < 2$, with overwhelming probability over $Q \le 2^{(2-\varepsilon)(\log_2 \lambda)^2}$ random inputs, every linear attack has bias $\exp(-\lambda^{\varepsilon-o(1)})$.

We also determine the distance of the Sipser candidate from functions of low rational degree over $\mathbb{F}_2$. With asymptotically balanced noise, almost every key has distance $(1/2+o(1))2^{-d}$ from functions of rational degree at most d, simultaneously for $1 \le d \le (1-\varepsilon)r_\lambda$, where $\varepsilon \in (0,1)$ is fixed and $r_\lambda = \Theta(\lambda)$ is the middle fan-in. This matches the universal upper bound. Here the secret map is sampled uniformly, allowing singular maps. A companion bound applies to every key of the candidate of Ball et al.

Keywords: Pseudorandom functions, linear tests, low-depth circuits, bounded-query security, rational approximation.

1 Introduction

Several candidates for low-complexity pseudorandom functions start with a simple construction: apply a secret linear or affine map to the input, then evaluate a Boolean predicate with a shallow circuit. The map is intended to hide the structure of the predicate, while the predicate keeps evaluation cheap. Low-depth evaluation is useful when a PRF is used inside secure computation or another protocol whose cost reflects the circuit complexity of the primitive [BDE+25, BCG+21]. Such candidates also give concrete examples for the question of how little computation can support pseudorandomness.

The same secret map must be reused across many queries, and this is where relations among the inputs become an obstacle. Linear maps preserve linear relations, while affine maps preserve affine relations. A linear test takes a parity of output bits and asks whether the preserved relations induce a detectable bias.

*Supported by the European Research Council under grant agreement No. 101115790 (OBELiSC).

Even a predicate whose individual evaluations look random can therefore fail when its evaluations are considered together.

We study this problem for two recent candidates with different security notions and query models. In the bounded-query strong-PRF candidate of Ball, Ducros, Erabelli, Kohl, Resch, and Scholl [BDE+25] (BDEKRS), the adversary chooses the queries and can therefore impose arbitrary affine dependencies. In the weak-PRF construction of Boyle, Couteau, Gilboa, Ishai, Kohl, and Scholl [BCG+21] (BCGIKS), the inputs are uniform random samples. Our analysis first chooses the secret matrix uniformly from all binary matrices. In each coordinate of the first candidate's affine image, the bits obtained from distinct queries are pairwise independent. In the second candidate, random samples typically make larger sets of the corresponding bits independent. We later condition the first candidate's matrix on invertibility.

Each TRIBES block in the Ball et al. candidate, and the Sipser predicate in the Boyle et al. candidate, is an OR of many independent local events, each with small probability. In both cases, the OR takes each value with constant probability. Our main observation is that repeated Cauchy–Schwarz can be organized around this rare-event decomposition. Each round introduces an independent copy of one group's randomness, and after expanding the square the condition that no local event occurs in that group is replaced by the event that the corresponding indicators in the two copies disagree. This makes the event that remains in the final parity substantially rarer while preserving the independence supplied by the underlying linear forms. The number of Cauchy–Schwarz rounds is governed by the rarity of the local event rather than by the full dependency structure of the query system.

1.1 Two linear-test problems

We first recall the chosen-query problem of Ball et al. For $z = (z_{j,r}) \in \mathbb{F}_2^{mn}$, write $\mathrm{TRIBES}_{m,n}(z) = \bigvee_{j=1}^{m} \bigwedge_{r=1}^{n} z_{j,r}$. Their candidate applies one secret affine map and then XORs ℓ TRIBES blocks on disjoint coordinate sets:

$$F_{A,b}(x) = \bigoplus_{w=1}^{\ell} \mathrm{TRIBES}_{m,n}\Big((Ax \oplus b)^{(w)}\Big), \tag{1}$$

where $A \leftarrow \mathrm{GL}_N(\mathbb{F}_2)$, $b \leftarrow \mathbb{F}_2^N$, and $N = \ell mn$. A non-adaptive linear test chooses distinct inputs $x_1, \ldots, x_q$ and a nonzero $u \in \mathbb{F}_2^q$ and considers $\bigoplus_{i=1}^{q} u_i F_{A,b}(x_i)$. For example, suppose four distinct queries satisfy $x_1 + x_2 + x_3 + x_4 = 0$. Their images under $x \mapsto Ax \oplus b$ satisfy the same relation, since the four copies of b cancel. A linear test can then ask whether $F_{A,b}(x_1) \oplus F_{A,b}(x_2) \oplus F_{A,b}(x_3) \oplus F_{A,b}(x_4)$ has a detectable bias. For a uniformly random function this parity is unbiased. Since coordinates with $u_i = 0$ can be discarded, it suffices to bound the bias of the output parity on any set of distinct queries. We write $\mathrm{Bias}(x_1, \ldots, x_q)$ for the absolute expectation of $(-1)^{\sum_i F_{A,b}(x_i)}$ over the secret key.

Ball et al. bound the bias by $2^{-\Omega(\lambda \log \lambda)}$ for every fixed $q = O(1)$, with arbitrary query dependencies [BDE+25, Theorem 5.11]. They also note that their analysis handles roughly $q = o(\sqrt{\log \lambda})$ queries [BDE+25, Section 1.4, footnote 6]. Their Theorem 5.13 allows up to N adaptively chosen linearly independent queries with a fixed parity test. The polynomial-query case with arbitrary affine dependencies is posed as an open problem in their Section 1.3.

The weak-PRF problem of Boyle et al. has a different form. Their candidate XORs a secret parity with a depth-three Sipser predicate G_λ applied after a secret linear map:

$$f_{s,K}(x) = \langle s, x \rangle \oplus G_\lambda(Kx). \tag{2}$$

Here the public inputs $x_1, \ldots, x_Q$ are uniform random samples, and a linear attack may choose its coefficient vector after seeing them. If $\sum_i u_i x_i \neq 0$, the uniform secret s makes the corresponding output parity exactly unbiased. Thus only coefficient vectors satisfying $\sum_i u_i x_i = 0$ can give a nonzero bias. Boyle et al. formulate the linear-attack question for this Sipser-noise construction in Section 4.2 of [BCG+21].

For the chosen-query problem, we must allow every affine relation the adversary can impose. In the random-sample problem, every small subfamily is linearly independent with high probability. We use this additional independence to obtain a stronger correlation bound.

The Sipser construction raises a further algebraic question. An attack may seek low-degree polynomials P_0, P_1, not both zero, for which $fP_0 = P_1$ holds on almost every input. Boyle et al. prove a large rational degree, which rules out such an identity holding everywhere, and explicitly distinguish this from proximity to low rational degree [BCG+21, Section 4.1]. Our approximation bound uses the fact that each input variable appears only once in the predicate's formula. We analyze restrictions of this formula to bound how often any such equation must fail.

1.2 Our results

For the balanced parameters of Ball et al., $\ell = \lambda$, $mn = \lambda$, $N = \lambda^2$, and $2^{-n} = \Theta\Big(\frac{\log \lambda}{\lambda}\Big)$, and the key length is $\kappa = N^2 + N = \Theta(\lambda^4)$.

Our first result resolves the open polynomial-query problem of Ball et al. for general non-adaptive linear tests. For every fixed $C > 0$, the bias is at most $\exp(-\Omega_C(\lambda))$ on every set of $q \le \lambda^C$ distinct chosen queries, with arbitrary rank and affine dependencies. The same argument extends beyond every fixed polynomial. For every fixed $a \in (0, 1)$,

$$q \le 2^{a(\log_2 \lambda)^2} \qquad \Longrightarrow \qquad \mathrm{Bias}(x_1, \dots, x_q) \le \exp(-\lambda^{1-a-o(1)}). \tag{3}$$

In particular, writing the bound in terms of the key length, for every fixed $0 < c < 1/4$, $q \le \kappa^{\log^c \kappa}$ implies $\mathrm{Bias}(x_1, \dots, x_q) \le 2^{-\kappa^c}$ for all sufficiently large κ. Thus, within non-adaptive linear tests, the result matches the query/advantage form of the bounded-query conjecture of Ball et al. throughout this range. The theorem with explicit parameter bounds and the full tradeoff are given in Sections 5 and 5.2.

The second result uses the stronger form of the same decorrelation theorem. Suppose every subfamily of at most k of the underlying linear forms is linearly independent. The Cauchy–Schwarz transformation then preserves k-wise independence, and the final parity estimate is exponentially small in k before taking the 2^tth root introduced by the Cauchy–Schwarz iterations. Random samples have no short linear dependencies with overwhelming probability, which makes this stronger bound available in the weak-PRF setting.

For the Sipser-based family of Boyle et al., for every fixed $0 < \varepsilon < 2$ and $Q \le 2^{(2-\varepsilon)(\log_2 \lambda)^2}$, with probability $1 - 2^{-\Omega(n_\lambda)}$ over the Q public samples, *simultaneously for every* nonzero linear attack u,

$$\left| \mathop{\mathbb{E}}_{s,K} \left[(-1)^{\sum_i u_i f_{s,K}(x_i)} \right] \right| \le \exp(-\lambda^{\varepsilon - o(1)}). \tag{4}$$

Here $n_\lambda = \Theta(\lambda^2 \log \lambda)$ is the input dimension. The attack vector may be chosen after seeing the entire sample tuple, so the bound holds for every linear attack on these samples. Extending the guarantee to the conjectured subexponential weak-PRF regime remains open.

We also determine the distance of the Sipser candidate from functions of low rational degree. Write $\Delta_d^{\mathrm{rat}}(f)$ for the minimum fraction of inputs on which f differs from a function of rational degree at most d, over $\mathbb{F}_2$ with the convention that $0/0$ imposes no restriction. For every fixed $\varepsilon \in (0, 1)$, outside a $2^{-\Omega_\varepsilon(\lambda^2)}$ fraction of keys, simultaneously for every $1 \le d \le (1 - \varepsilon) r_\lambda$,

$$(b_\lambda - o(1))2^{-d} \le \Delta_d^{\mathrm{rat}}(f_{s,K}) \le 2^{-d-1}, \tag{5}$$

Here $r_\lambda = \Theta(\lambda)$ is the Sipser middle fan-in, $\beta_\lambda = \mathbb{P}[G_\lambda(U) = 0]$ for uniform U, and $b_\lambda = \min\{\beta_\lambda, 1 - \beta_\lambda\}$. For every pair P_0, P_1 of degree at most d, not both zero, the equation $f_{s,K} P_0 = P_1$ also fails on at least

the fraction in the lower bound. If $\beta_\lambda = 1/2 + o(1)$, the two bounds match to leading constant. The upper bound is classical and holds for every Boolean function [BP05, LL10]; the lower bound is specific to the candidate. It holds without conditioning the secret map on invertibility. For every key of the Ball et al. candidate, a companion argument gives distance $2^{-d-O(d/\log\lambda+1)}$ through degrees of order $\lambda/\log\lambda$. The precise statements are in Section 7.

The correlation bound for the candidate of Ball et al. has consequences beyond parity tests. We show that at the quasipolynomial query scale the vector of responses on any fixed query set fools subexponential-size constant-depth AC^0 distinguishers. We also obtain unrestricted statistical security for a smaller adaptive query range, and a stronger adaptive theorem when every execution path contains only $O(\lambda/\log\lambda)$ independent affine relations. These consequences are developed in Sections 8 and 9.

1.3 Relation to prior work

The two candidates belong to a broader line of work on shallow weak pseudorandom functions; see Boura, Couteau, Perrin, and Rotella [BCPR25] for a recent overview. Akavia, Bogdanov, Guo, Kamath, and Rosen [ABG+14] combined a secret linear transformation with TRIBES. Bogdanov and Rosen later observed that this construction has rational degree $O(\log\lambda)$, yielding a quasipolynomial-time distinguisher [BR17]. Boyle et al. replace the Boolean predicate by a depth-three Sipser function and prove large exact rational degree when the combined secret linear map is invertible [BCG+21]. In the chosen-query setting, Ball et al. identify a different weakness of the earlier TRIBES construction [BDE+25]. They apply a secret affine map, partition its output into blocks, and XOR the TRIBES evaluations on these blocks. We study this candidate. Their constrained Fourier-sum analysis gives the constant-query bound and the slowly growing extension described above. For every fixed $C > 0$, we prove a bias bound of $\exp(-\Omega_C(\lambda))$ for every $q \le \lambda^C$, with arbitrary affine dependencies, and derive the larger query range in (3).

Approximation by functions admitting low-degree annihilators[1] was studied by Braeken and Preneel in their work on probabilistic algebraic attacks [BP05]. Liu and Lin formulate the distance to algebraic immunity and relate it to higher-order nonlinearity [LL10]. We use this established distance notion. Bogdanov and Rosen ask about quantitative distance from low rational degree for AC^0 functions [BR17, Section 7.5], and Boyle et al. distinguish exact degree from proximity for the Sipser candidate. Our result gives the optimal distance scale through degrees proportional to its middle fan-in. For Ball et al., Sections 5.5–5.6 of [BDE+25] establish high exact rational degree and bounds on approximation by low-degree polynomials. Our distance bound concerns the larger class of functions with low rational degree. The proof uses forcing probabilities; the subspace-indicator argument used to handle singular maps is the standard restriction argument for algebraic immunity [Car21, Section 9.1.1] in a quantitative form.

The bounded-query formulation is also related to standard limitations on low-depth circuits. Razborov and Smolensky showed that bounded-depth circuits with modular gates admit low-degree polynomial approximations in the regimes relevant to their lower bounds [Raz87, Smo87], and Krause and Lucks made the corresponding cryptographic implication explicit, showing that sufficiently effective approximation or learning procedures yield distinguishers against PRFs in the class [KL01]. These results motivate the quasipolynomial query regime. Our correlation bounds address the joint distribution of evaluations that share one secret map.

The Boyle et al. problem is similar, but the structure comes from the random samples rather than chosen queries. If the attack coefficients do not define a linear relation among the samples, the parity term makes the chosen output parity unbiased, leaving the deterministic Sipser noise to be analyzed only on linear dependencies among random samples [BCG+21]. Other structured LPN-based constructions obtain pseudorandom correlations using other algebraic or coding techniques [BCG+20, CD23, BCG+22]. We bound the bias of the Sipser output parity on every linear relation among the sampled inputs.

[1] A nonzero polynomial P annihilates f if $fP = 0$ on every input.

Repeated Cauchy–Schwarz for systems of linear forms underlies Gowers' uniformity norms [Gow01], Cauchy–Schwarz complexity [GT10], and true-complexity theory [GW10, HHL16]. Austrin and Mossel study correlations under pairwise independence and record Hatami's observation that a related linear-form correlation can be treated using Gowers–Cauchy–Schwarz and Fourier truncation [AM13]. Manners [Man21] and Candela, González-Sánchez, and Szegedy [CGSS22] organize the sequence of Cauchy–Schwarz steps using the structure of the forms.

Here each step acts on a group of rare local events in the predicate. For an OR of probability-p events that takes each output value with constant probability, $O(1 + \log q/\log(1/p))$ rounds suffice uniformly over distinct nonzero binary forms. The transformed sparse indicators retain the available independence. We tune their total mean to a small constant under pairwise independence, or to $\Theta(k)$ under k-wise independence, and use the corresponding moment bounds to control their parity. For related limited-independence estimates, Benjamini, Gurel-Gurevich, and Peled study Boolean functions including TRIBES [BGGP12], and Gopalan and Yehudayoff obtain general product bounds [GY20].

The nonlinear consequences then follow from standard pseudorandomness results applied to our small-bias bound. We use Alon–Goldreich–Mansour [AGM03] to pass from almost k-wise independence to exact k-wise independence and then apply Bazzi [Baz07] and Tal [Tal17]. For the small-bias notion, we follow Naor and Naor [NN90].

1.4 Scope and organization

Our theorems bound linear-test bias in the query ranges above and the error of low-degree rational equations for fixed keys. Full weak-PRF security remains open. For the candidate of Ball et al., algebraic degree gives a barrier at $2^{\lambda+1}$ queries. The function has degree at most $mn = \lambda$, and an affine cube of this size has output parity zero for every key. Our positive result covers the quasipolynomial regime in (3), leaving an open range between this guarantee and the affine-cube attack. For Sipser, the final exponent is proportional to $k/2^t$, where k is the independence parameter. As the sample count grows, the factor 2^t eventually makes this exponent too small. Extending the bound requires reducing this loss. We return to these limitations in Section 10.

Section 2 gives a high-level account of the proof. Section 3 fixes notation and the basic linear-algebraic facts. In Section 4 we prove the rare-event decorrelation theorem, first for arbitrary systems of forms and then under the stronger assumption that every small subfamily is linearly independent. Section 5 applies Theorem 4.6 to the candidate of Ball et al., and Section 5.2 derives the polynomial and quasipolynomial query ranges. Section 6 applies the higher-independence theorem to the Sipser-based family of Boyle et al. Section 7 proves the rational-distance bounds for both candidates. The remaining sections develop further consequences for the candidate of Ball et al. and give the information-theoretic barriers.

2 Technical overview

Both predicates are read-once: each input variable appears once in the formula. Disjoint subformulas therefore use independent input bits. For linear tests, we duplicate groups of input blocks and track the rare events that survive each Cauchy–Schwarz step. For rational approximation, we leave the variables of a highest-degree monomial free and ask when the other coordinates already determine the predicate's value. This gives inputs on which the polynomial is nonzero and a proposed equation fails.

2.1 Rare-event decorrelation

Let $\varphi : \mathbb{F}_2^n \to \{0,1\}$ be a local predicate with $p := \mathbb{P}[\varphi(U) = 1] \ll 1$, and form the OR

$$H(z^{(1)}, \dots, z^{(m)}) = \bigvee_{j=1}^{m} \varphi(z^{(j)}), \tag{6}$$

where the blocks are independent and m is chosen so that $\beta := (1-p)^m$ is bounded away from both 0 and 1. TRIBES corresponds to $\varphi = \mathrm{AND}_n$; the Sipser predicate used by BCGIKS has the same form after grouping its middle gates.

Let $L_1, \dots, L_q : V \to \mathbb{F}_2$ be distinct nonzero linear forms. For each input coordinate of H, sample an independent uniform vector in V. Apply every L_i to this same collection of vectors to obtain the input to the ith evaluation of H. For each sampled vector Z, any two of $L_1(Z), \dots, L_q(Z)$ are independent unbiased bits, although larger subfamilies may be dependent. Let J_i be the indicator that none of the m local events occurs in the ith evaluation. The absolute output correlation is then the parity bias of $J_1, \dots, J_q$. Pairwise independence alone gives no useful direct control over this parity.

The key step is to apply Cauchy–Schwarz to groups of the local events rather than to the system of linear forms itself. Partition some of the m blocks into t groups. At each round, Cauchy–Schwarz introduces an independent copy of one group's randomness. After expanding the square, the condition that no event occurs in that group is replaced by the event that the corresponding indicators in the two copies disagree. After t rounds,

$$|\mathcal{R}|^{2^t} \le \mathbb{E}[(-1)^T], \qquad T = \sum_{i=1}^{q} C_i, \tag{7}$$

where $\mathcal{R}$ is the original correlation and C_i records a disagreement in every duplicated group and no local event in the remaining blocks. If the duplicated groups have sizes $r_1, \dots, r_t$, then

$$\mathbb{P}[C_i = 1] = \beta\, 2^t \prod_{g=1}^{t} \left(1 - (1-p)^{r_g}\right). \tag{8}$$

A singleton group contributes a factor $2p$ in (8). Using such groups therefore makes the events rarer at each round.

We take the first $t-1$ groups to be single blocks and use the size of the last group only to tune the final mean. Then $t = O\!\left(1 + \frac{\log q}{\log(1/p)}\right)$ rounds suffice, with fewer rounds when the local events are rarer.

For arbitrary distinct forms, the C_i remain pairwise independent: each is computed from the evaluations of its own form on the original and duplicated vectors. We tune $\mathbb{E}[T]$ to a small constant; then $\mathbb{P}[T=1] \ge \mathbb{E}[T] - \mathbb{E}[T(T-1)] = \Omega(1)$, so $\mathbb{E}[(-1)^T] \le 1 - \Omega(1)$. Taking the 2^tth root in (7) gives a nontrivial one-block correlation bound.

If every subfamily of at most k forms is linearly independent, the C_i are instead k-wise independent. We tune $\mathbb{E}[T] = \Theta(k)$. For every $r < k$, the value of $\mathbb{E}[\binom{T}{r}]$ is then the same as in the independent case, and a truncated expansion of $(-1)^T$ gives $\left|\mathbb{E}[(-1)^T]\right| \le \exp(-\Omega(k))$. Consequently,

$$|\mathcal{R}| \le \exp\!\left(-\Omega\!\left(\frac{k}{2^t}\right)\right). \tag{9}$$

This is the stronger form used in the random-sample application.

2.2 The BDEKRS candidate

For one uniform matrix row a and one uniform shift bit c, the corresponding coordinate of the image of a query x_i is the linear form $L_i(a, c) = \langle a, x_i \rangle + c$. Its coefficient vector is $(x_i, 1)$. Distinct queries therefore give distinct nonzero forms, regardless of how many affine relations the query set contains. Theorem 4.6 applies to a single TRIBES block using only the resulting pairwise independence.

We next need to combine the TRIBES blocks and require the secret matrix to be invertible. A small signed expectation can become large after conditioning, so the one-block correlation alone does not suffice. For a fixed row block R, average over the shift and call the result $z(R)$. We bound $\mathbb{E}_R[|z(R)|]$ using $\mathbb{E}_R[z(R)^2]$. Expanding this square introduces two independent shifts and hence the $2q$ forms $(x_i, 1, 0)$ and $(x_i, 0, 1)$, which are again all distinct. The one-block theorem controls this second moment.

Under a uniform matrix the ℓ row blocks are independent. Independence across these blocks raises the one-block bound on $\mathbb{E}_R[|z(R)|]$ to the ℓth power. Only after this multiplication do we condition the matrix on invertibility, which costs a constant factor because a uniform binary matrix is invertible with constant probability. If t_{2q} denotes the number of Cauchy–Schwarz rounds for the resulting $2q$ forms, the final bound has the shape

$$\mathrm{Bias}(x_1, \ldots, x_q) \le \exp\left(-\Omega\left(\frac{\lambda}{2^{t_{2q}}}\right)\right).$$

The number of rounds satisfies $t_{2q} \le 1 + (\log(2q) + O(1))/\log(1/(2p))$. For the BDEKRS parameters, $p = \Theta(\log \lambda/\lambda)$, so $t_{2q} = O(1)$ for every fixed polynomial query bound. If $q \le 2^{a(\log_2 \lambda)^2}$, then $t_{2q} \le (a + o(1)) \log_2 \lambda$, giving (3).

2.3 The BCGIKS Sipser application

The stronger theorem becomes useful when the forms have no short dependency. For Q independent uniform samples in $\mathbb{F}_2^{n_\lambda}$, a fixed nonempty subset sums to zero with probability 2^{-n_λ}. A union bound therefore shows that, with overwhelming probability, every nonzero dependency has support at least $k_{\mathrm{dep}} = \Omega\left(\frac{n_\lambda}{\log Q}\right)$. In the family (2), if the attack coefficients do not define a linear relation among the samples, the uniform secret s makes the chosen output parity unbiased. Otherwise, $\sum_i u_i x_i = 0$, so $\mathrm{supp}(u)$ has size larger than k_{dep}. Every subfamily of at most k_{dep} corresponding forms is linearly independent.

The Sipser predicate is an OR of rare independent events and takes each output value with constant probability. Its local acceptance probability is $p_\lambda = \Theta(1/\lambda)$. Set $k = k_{\mathrm{dep}}$ in (9). The resulting bias is at most $\exp\left(-\Omega\left(\frac{k_{\mathrm{dep}}}{2^{t_{Q,k_{\mathrm{dep}}}}}\right)\right)$.

For $Q \le 2^{(2-\varepsilon)(\log_2 \lambda)^2}$, the number of rounds satisfies

$$t_{Q,k_{\mathrm{dep}}} \le 1 + \frac{\log(Q/k_{\mathrm{dep}}) + O(1)}{\log(1/(2p_\lambda))} \le (2 - \varepsilon + o(1)) \log_2 \lambda.$$

Together with $n_\lambda = \Theta(\lambda^2 \log \lambda)$, this gives

$$\frac{k_{\mathrm{dep}}}{2^{t_{Q,k_{\mathrm{dep}}}}} \ge \lambda^{\varepsilon - o(1)},$$

which gives (4) simultaneously for every linear attack on the sampled points.

2.4 Rational approximation and forcing

To bound the error of $fP_0 = P_1$, we lower-bound $\mathbb{P}[f = b, R = 1]$ for every nonzero low-degree polynomial R and each $b \in \{0, 1\}$. If $P_1 \ne 0$, the event $f = 0, P_1 = 1$ violates the equation; if $P_1 = 0$, use

$f = 1, P_0 = 1$. To obtain the required probability bound, choose a highest-degree monomial of R, let S be its variable set, and leave those coordinates free. Every fixing outside S leaves a nonzero polynomial: the coefficient of that monomial cannot change, since a term that changed it would have higher degree. Thus $R = 1$ on at least one of the $2^{|S|}$ remaining assignments.

A fixing forces the predicate to equal b if it has that value for every assignment to the free coordinates. Suppose a fraction α of complementary fixings do so. Each of these fixings supplies an input with both $f = b$ and $R = 1$. Consequently, $\mathbb{P}[f = b, R = 1] \geq \alpha 2^{-\deg R}$. Applying this to the appropriate polynomial in a nonzero pair P_0, P_1 gives an error lower bound for $fP_0 = P_1$. The argument reduces rational approximation to a concrete question about the Boolean predicate.

For Sipser, monotonicity lets us test whether a fixing forces zero by setting every free bit to one, and whether it forces one by setting every free bit to zero. For fixed $\xi > 0$ and any set of at most $(1 - \xi)r_\lambda$ free coordinates, a uniform fixing of the remaining coordinates forces each value with probability at least $b_\lambda - o(1)$. This gives the lower bound in (5). To see why the scale is sharp, take a codimension-d coordinate subcube and change f to its majority value there. This changes at most a 2^{-d-1} fraction of the inputs, and the modified function has rational degree at most d.

We still have to pass from G to the keyed function. The intersection bound also holds for $H(z, y) = G(z) + y$, by considering the restrictions $y = 0, 1$. The keyed function is $H(Bx)$, where $B = (K; s)$ is a uniform square binary matrix. An invertible B would allow a change of variables, but the key distribution also includes singular matrices. If B has corank h,[2] its image has codimension h. We extend a polynomial from the image to the full space, then multiply it by the degree-h indicator of the image. The product vanishes outside the image, so its intersection probability in the full space is 2^{-h} times that in the image. The degree increase contributes the same factor 2^{-h} to the lower bound. Cancelling the two factors preserves the error constant and costs h in the available degree range. A uniform binary matrix has corank at least h with probability at most $O(2^{-h^2})$, which gives the claimed bound on the fraction of keys for which the theorem fails.

The BDEKRS predicate requires a different forcing estimate. Free coordinates may leave some TRIBES blocks undecided, but the probability that all blocks are determined is at least $\exp(-O(d/n))$, where n is the tribe width. When fewer than ℓ coordinates are free, an untouched block still supplies both output values with constant probability. The same highest-degree-monomial argument then gives the companion distance bound.

2.5 Further consequences

The correlation theorem for the BDEKRS candidate controls every parity of the responses on a fixed query set, so the response vector has small bias. Small bias makes every marginal on a small set of coordinates close to uniform. Alon–Goldreich–Mansour [AGM03] show that such a distribution is close to an exactly k-wise independent one. Tal's theorem [Tal17] then gives the stated bounds against constant-depth AC^0 circuits of subexponential size. A decision-tree argument also gives statistical security for $O(\lambda)$ adaptive queries.

For the adaptive result with few affine relations, we first choose the matrix uniformly from all binary matrices. Fix a subset S of response coordinates and consider the affine relations among the corresponding queries. If a query occurs in no such relation, its response is independent of the other responses in S and contributes its one-query expectation as a separate factor. We group the resulting Fourier coefficients according to the subspaces of relations. When the full affine relation space has dimension $O(\lambda/\log\lambda)$, this bounds the sum of the absolute Fourier coefficients. A coupling between a uniform affine map and one with an invertible linear part then transfers the resulting pointwise estimate to adaptive queries.

[2] For an $N \times N$ matrix B, its corank is $N - \operatorname{rank} B$.

3 Preliminaries

All vector spaces and linear maps are over $\mathbb{F}_2$. We write $[q] := \{1, \dots, q\}$, use $\log = \log_2$, and reserve $\ln$ for the natural logarithm. Unless otherwise specified, all expectations and probabilities are over all random choices appearing in their subscripts.

For a finite set Ω, the notation $X \leftarrow \Omega$ means that X is sampled uniformly from Ω. We write $\mathrm{GL}_N(\mathbb{F}_2)$ for the group of invertible $N \times N$ matrices over $\mathbb{F}_2$. For a random variable X, write $\mathsf{Law}(X)$ for its distribution. We identify a bit $b \in \{0,1\}$ with its signed encoding $(-1)^b \in \{-1,1\}$. We switch freely between the $\{0,1\}$ and $\{-1,1\}$ representations coordinatewise. This preserves total variation distance and changes an AC^0 circuit only by possible input negations.

3.1 Statistical distance and bounded-query pseudorandomness

For two distributions P, Q on a finite set Ω, their total variation distance is

$$\mathrm{TV}(P,Q) := \frac{1}{2} \sum_{\omega \in \Omega} |P(\omega) - Q(\omega)|.$$

Equivalently, $\mathrm{TV}(P,Q) = \max_{T \subseteq \Omega} |P(T) - Q(T)|$. Thus $\mathrm{TV}(P,Q)$ is precisely the maximum distinguishing advantage of an unbounded distinguisher receiving one sample from either P or Q.

We use the standard PRF experiment of Goldreich, Goldwasser, and Micali [GGM86], with the bounded-query restriction studied by BDEKRS [BDE+25]. Let $\mathcal{F} = \{F_k : \mathcal{X} \to \{0,1\}\}_{k \in \mathcal{K}}$ be a keyed family. In the *real* experiment, a key $k \leftarrow \mathcal{K}$ is sampled and an adversary $\mathcal{A}$ receives oracle access to F_k. In the *random* experiment, $\mathcal{A}$ instead receives oracle access to a uniformly random function $R : \mathcal{X} \to \{0,1\}$. The distinguishing advantage of $\mathcal{A}$ is $\mathrm{Adv}_{\mathcal{F}}(\mathcal{A}) := \left|\mathbb{P}[\mathcal{A}^{F_k} = 1] - \mathbb{P}[\mathcal{A}^{R} = 1]\right|$.

A q-query adversary makes at most q oracle queries in either experiment. It is *non-adaptive* if all queried inputs are chosen before any oracle answers are seen, and *adaptive* otherwise. Fixing its internal randomness turns an adaptive adversary into a deterministic binary decision tree of depth at most q. Its internal nodes are labeled by oracle queries, and its outgoing edges correspond to the two possible answers.

We call a statement computational when it restricts the adversary's running time or circuit size, and statistical when it allows computationally unbounded adversaries. Our main linear-test result is information-theoretic for the specified class of tests. Later, a standard Fourier argument also gives information-theoretic security against arbitrary adaptive adversaries for a smaller query range.

Repeated queries never provide additional information to a deterministic oracle algorithm. We may therefore assume, whenever convenient, that the queries appearing along any root-to-leaf path of an adaptive decision tree are distinct.

3.2 Linear tests and small bias

Following the linear-test model used by BDEKRS [BDE+25], let $F : \mathbb{F}_2^N \to \mathbb{F}_2$. A non-adaptive linear test chooses distinct queries $x_1, \dots, x_q \in \mathbb{F}_2^N$ and a nonzero attack vector $u = (u_1, \dots, u_q) \in \mathbb{F}_2^q$, and considers the parity $\bigoplus_{i=1}^q u_i F(x_i)$. For a uniformly random function, every such nontrivial parity is an unbiased bit.

For the keyed family considered in this paper, the bias of this test is

$$\left| \mathop{\mathbb{E}}_{A,b} \left[(-1)^{\sum_{i=1}^q u_i F_{A,b}(x_i)} \right] \right|.$$

Since the coordinates for which $u_i = 0$ do not contribute, it suffices to take the parity of the outputs on the queries indexed by $\mathrm{supp}(u)$. Accordingly, for distinct $x_1, \dots, x_q \in \mathbb{F}_2^N$, we write

$$\mathrm{Bias}(x_1, \dots, x_q) := \left| \mathop{\mathbb{E}}_{A,b} \left[\prod_{i=1}^{q} (-1)^{F_{A,b}(x_i)} \right] \right|. \tag{10}$$

A uniform bound on (10) therefore controls every non-adaptive linear test. The same statement holds for randomized non-adaptive tests by conditioning on their internal randomness.

The same notion can be stated in the standard language of small-bias distributions [NN90]. Let $Z = (Z_1, \dots, Z_q)$ be a random vector in $\{-1,1\}^q$. We say that Z is *ε-biased* if, for every nonempty $S \subseteq [q]$, $\left|\mathbb{E}\left[\prod_{i\in S} Z_i\right]\right| \le \varepsilon$. Thus our linear-test problem asks for a small-bias bound on the signed outputs of the candidate on an arbitrary chosen query set.

3.3 Fourier analysis on the Boolean cube

For every finite index set I, let U_I denote a uniform random vector in $\{-1,1\}^I$, and abbreviate $U_{[q]}$ by U_q. We use the same symbol for its distribution when it appears as an argument of TV or χ^2. For $S \subseteq [q]$, define the Fourier character $\chi_S(z) := \prod_{i\in S} z_i$, $z \in \{-1,1\}^q$. The characters form an orthonormal basis under U_q. Hence every function $D : \{-1,1\}^q \to \mathbb{R}$ has the expansion

$$D(z) = \sum_{S\subseteq[q]} \widehat{D}(S)\chi_S(z), \qquad \widehat{D}(S) := \mathop{\mathbb{E}}_{U_q}\big[D(U_q)\chi_S(U_q)\big]. \tag{11}$$

Parseval's identity gives $\sum_{S\subseteq[q]} \widehat{D}(S)^2 = \mathbb{E}_{U_q}\big[D(U_q)^2\big]$.

The Fourier degree of D is $\deg_{\mathrm{F}}(D) := \max\{|S| : \widehat{D}(S) \neq 0\}$. We call $\|\widehat{D}\|_1 := \sum_{S\subseteq[q]} |\widehat{D}(S)|$ the *Fourier L_1 norm* of D.[3]

Let P be a distribution on $\{-1,1\}^q$, and define its density relative to uniform by $f_P(z) := 2^q P(z)$. Then

$$\widehat{f_P}(S) = \mathop{\mathbb{E}}_{Z\leftarrow P}\big[\chi_S(Z)\big]. \tag{12}$$

Thus the magnitudes of the nonconstant Fourier coefficients of f_P are the parity biases of P.

We will also use the standard χ^2 identity

$$\chi^2(P\|U_q) = \mathop{\mathbb{E}}_{U_q}\big[(f_P(U_q)-1)^2\big] = \sum_{\emptyset\neq S\subseteq[q]} \widehat{f_P}(S)^2 \tag{13}$$

and the consequence

$$\mathrm{TV}(P, U_q) \le \frac{1}{2}\sqrt{\chi^2(P\|U_q)}. \tag{14}$$

In particular, if P is ε-biased, then $\mathrm{TV}(P, U_q) \le \frac{1}{2}\sqrt{2^q-1}\,\varepsilon$. We will later use the same Fourier viewpoint to control distinguishers with small Fourier L_1 norm and, via decision trees, arbitrary bounded-query adaptive distinguishers.

[3] In Boolean-function analysis this quantity is also commonly called the *spectral norm*. We use "Fourier L_1 norm" to avoid confusion with matrix spectral norms.

3.4 TRIBES and the BDEKRS candidate

We next give the candidate and its parameters in the form used in the proofs. Put $M := mn$. For $z = (z_{j,k})_{j\in[m],\,k\in[n]} \in \mathbb{F}_2^M$, define the TRIBES function of Ben-Or and Linial [BL85] by $\mathrm{TRIBES}_{m,n}(z) := \bigvee_{j=1}^{m} \bigwedge_{k=1}^{n} z_{j,k}$. We refer to the n coordinates $(z_{j,1}, \dots, z_{j,n})$ as the jth *tribe*.

A uniform tribe consists entirely of ones with probability $p := 2^{-n}$. The m tribes use disjoint coordinates, so TRIBES is zero with probability $\beta := (1-p)^m$ and one with probability $1-\beta$. The balanced regime chooses m, n so that β is close to $1/2$.

Let $N := \ell M = \ell mn$, and decompose $y \in \mathbb{F}_2^N$ as $y = (y^{(1)}, \dots, y^{(\ell)})$, where $y^{(w)} \in \mathbb{F}_2^M$. The candidate of BDEKRS [BDE+25] is defined in (1), with $A \leftarrow \mathrm{GL}_N(\mathbb{F}_2)$ and $b \leftarrow \mathbb{F}_2^N$ independent. We write $\mathcal{F}_{\mathrm{BDEKRS}} := \{F_{A,b} : A \in \mathrm{GL}_N(\mathbb{F}_2),\ b \in \mathbb{F}_2^N\}$ for this keyed family.

Let $h := (-1)^{\mathrm{TRIBES}_{m,n}}$. For y partitioned into ℓ blocks of size M, put $g(y) := \prod_{w=1}^{\ell} h(y^{(w)})$. Then $(-1)^{F_{A,b}(x)} = g(Ax \oplus b)$.

For the balanced parameters considered by BDEKRS [BDE+25] and throughout our main asymptotic consequences, $\ell = \lambda$, $M = mn = \lambda$, and $N = \lambda^2$, with $n = \log_2 \lambda - \log_2 \ln \lambda + O(1)$ and $m = \ln(2)\, 2^n + O(1)$. Consequently, $p = \Theta(\log \lambda/\lambda)$, $\beta = 1/2 + o(1)$, and $m = \Theta(\lambda/\log \lambda)$. The key is specified by A and b, so its description length is $\kappa = N^2 + N = \lambda^4 + \lambda^2 = \Theta(\lambda^4)$.

Fix a sufficiently small absolute constant $c_0 > 0$. The proofs below require only upper bounds on its value, so one fixed choice suffices throughout. For $s \geq 1$, we write t_s for the least integer $t \geq 1$ such that

$$s\beta(2p)^t \leq c_0. \tag{15}$$

Thus t_s is the number of Cauchy–Schwarz rounds used below for a system of s forms.

3.5 Linear forms over $\mathbb{F}_2$

We use the following elementary fact about linear forms over $\mathbb{F}_2$.

Lemma 3.1 (Distinct linear forms are pairwise independent)**.** *Let V be a finite-dimensional $\mathbb{F}_2$-vector space, and let $L, L' : V \to \mathbb{F}_2$ be distinct nonzero linear forms. If $Z \leftarrow V$ is uniform, then $(L(Z), L'(Z))$ is uniform on $\mathbb{F}_2^2$. In particular, $L(Z)$ and $L'(Z)$ are independent unbiased bits.*

Proof. Over $\mathbb{F}_2$, distinct nonzero linear forms are linearly independent. Hence the map

$$z \longmapsto (L(z), L'(z))$$

has rank 2 and is therefore surjective onto $\mathbb{F}_2^2$. Every point of $\mathbb{F}_2^2$ has the same number of preimages, so a uniform input induces the uniform distribution on $\mathbb{F}_2^2$. □

More generally, r linearly independent forms give r independent uniform bits on a uniform input. Distinct nonzero forms guarantee only pairwise independence: higher-order relations can prevent their values from being jointly independent.

We will repeatedly distinguish arbitrary systems of forms from systems with no short linear dependency. For nonzero linear forms $L_1, \dots, L_q : V \to \mathbb{F}_2$, we say that there is *no linear dependency of size at most k* if

$$\sum_{i\in I} L_i \neq 0 \qquad \text{for every nonempty } I \subseteq [q] \text{ with } |I| \leq k.$$

Equivalently, every subfamily of at most k forms is linearly independent. Over $\mathbb{F}_2$, every nonzero coefficient in a linear relation equals 1, so the relation is determined by its support.

We use the analogous condition for vectors. In particular, if $L_i(a) = \langle a, x_i \rangle$ for nonzero $x_i \in \mathbb{F}_2^n$, then the forms have no dependency of size at most k exactly when

$$\sum_{i \in I} x_i \neq 0 \qquad \text{for every nonempty } I \subseteq [q] \text{ with } |I| \leq k.$$

Affine expressions arising from the secret shift can be incorporated by adding one coordinate. In particular, for fixed $x \in \mathbb{F}_2^N$, the expression $\langle a, x \rangle + c$, $(a, c) \in \mathbb{F}_2^{N+1}$, is the linear form whose coefficient vector is $(x, 1)$. Hence, for distinct $x_1, \ldots, x_q$, the forms $(a, c) \longmapsto \langle a, x_i \rangle + c$ are distinct and nonzero, even if one of the queries is $x_i = 0$. By Lemma 3.1, their evaluations are therefore pairwise independent under a uniform choice of (a, c).

For a nonempty tuple $X = (x_1, \ldots, x_q)$, we write

$$d_{\text{aff}}(X) := \dim \operatorname{span}\{(x_1, 1), \ldots, (x_q, 1)\} - 1$$

for its affine dimension, and $\nu(X) := q - d_{\text{aff}}(X) - 1$ for the dimension of its affine relation space. Thus $\nu(X)$ is simply the number of independent affine relations among the augmented query vectors.

Distinct queries therefore supply the pairwise independence used in the one-block argument, whatever their affine relations.

4 Rare-event decorrelation for arbitrary linear forms

We first analyze one OR of local events, evaluated on several linear forms. Different input blocks use independent random vectors, while all forms are evaluated on the same vectors within a block. We will keep this independence across blocks throughout the Cauchy–Schwarz argument. The local predicate enters the bound through its acceptance probability.

For a Boolean predicate $\varphi : \mathbb{F}_2^n \to \{0, 1\}$, let $U_n \leftarrow \mathbb{F}_2^n$ be uniform and put $p_\varphi := \mathbb{P}[\varphi(U_n) = 1]$ and $\beta_\varphi := (1 - p_\varphi)^m$, and define

$$H_{\varphi,m}(z^{(1)}, \ldots, z^{(m)}) := \bigvee_{j=1}^{m} \varphi(z^{(j)}), \qquad h_{\varphi,m} := (-1)^{H_{\varphi,m}}.$$

We assume below that β_φ is bounded away from 0 and 1, so the OR takes each value with constant probability. The BDEKRS parameters give the stronger property $\beta = 1/2 + o(1)$. For $s \geq 1$, let $t_s(\varphi)$ be the least integer $t \geq 1$ satisfying $s\beta_\varphi(2p_\varphi)^t \leq c_0$. For $\varphi = \mathrm{AND}_n$, this agrees with the BDEKRS parameter t_s defined in (15).

We begin with a single TRIBES block and distinct linear forms, using only their pairwise independence from Lemma 3.1. After proving this estimate, we state its general form for the Sipser predicate. The outer XOR in BDEKRS is handled in Section 5.

4.1 The one-block correlation problem

Let V be a finite-dimensional $\mathbb{F}_2$-vector space, and let $L_1, \ldots, L_q : V \to \mathbb{F}_2$ be distinct nonzero linear forms. For every tribe $j \in [m]$ and coordinate $k \in [n]$, sample an independent uniform vector $Z_{j,k} \leftarrow V$. For $i \in [q]$, define

$$Y_i := (L_i(Z_{j,k}))_{j \in [m],\, k \in [n]} \in \mathbb{F}_2^M$$

and $\mathcal{R}(L_1, \ldots, L_q) := \mathbb{E}\left[\prod_{i=1}^q h(Y_i)\right]$.

Write $t_q = t_q(\mathrm{AND}_n)$, and let $\eta \in (0, 1)$ be the absolute constant supplied by Lemma 4.4.

Proposition 4.1 (One-block decorrelation). *Assume $\frac{1}{3} \le \beta \le \frac{2}{3}$ and $p \le c_0$. Let $t = t_q$ and suppose $tp \le c_0$ and $t \le \frac{m}{2}$. Then for arbitrary distinct nonzero linear forms $L_1, \dots, L_q : V \to \mathbb{F}_2$,*

$$|\mathcal{R}(L_1, \dots, L_q)| \le \eta^{1/2^{t_q}}.$$

The proof has two stages. We first use repeated Cauchy–Schwarz to reduce the original correlation to the parity of much rarer events while preserving the pairwise independence available from the linear forms. We then choose the number of Cauchy–Schwarz rounds and the tribe groups so that the expected number of those rare events is a small constant, at which point a second-moment argument bounds the resulting parity correlation by a constant below 1.

By Lemma 3.1, for $i \neq i'$ and uniform $Z \leftarrow V$, the pair $(L_i(Z), L_{i'}(Z))$ is uniform on $\mathbb{F}_2^2$.

For tribe j and form i, let $B_{j,i} := \prod_{k=1}^{n} L_i(Z_{j,k}) \in \{0,1\}$. Then $\mathbb{P}[B_{j,i} = 1] = p$, and for $i \neq i'$, the variables $B_{j,i}$ and $B_{j,i'}$ are independent. Different tribes are mutually independent. Thus $B_{j,i}$ is the indicator that tribe j evaluates to 1 for form i.

Define $J_i := \prod_{j=1}^{m}(1 - B_{j,i})$. Thus $J_i = 1$ iff no tribe evaluates to 1 on Y_i, and $h(Y_i) = (-1)^{1-J_i}$. Taking absolute values gives

$$|\mathcal{R}(L_1, \dots, L_q)| = \left| \mathbb{E}\left[(-1)^{\sum_{i=1}^{q} J_i} \right] \right|. \tag{16}$$

4.2 Repeated Cauchy–Schwarz

For an arbitrary system of forms, the variables $J_1, \dots, J_q$ are not guaranteed to be jointly independent, so pairwise independence alone does not control the parity in (16). We instead expose the tribe randomness in groups and apply Cauchy–Schwarz successively to those groups. Each application introduces an independent copy of one group's randomness, and after expanding the square the corresponding factor becomes the XOR of the two copies. After t rounds, the original correlation is bounded by the parity correlation of a family of much rarer events.

Lemma 4.2 (Iterated Cauchy–Schwarz). *Let $Y_0, Y_1, \dots, Y_t$ be mutually independent random variables. For each $i \in [q]$, let $R_i = R_i(Y_0) \in \{0,1\}$ and $I_{g,i} = I_{g,i}(Y_g) \in \{0,1\}$ for $g \in [t]$. Set*

$$\mathcal{R} := \mathbb{E}\left[(-1)^{\sum_{i=1}^{q} R_i \prod_{g=1}^{t} I_{g,i}} \right].$$

For every g, let $Y_g^{(0)}, Y_g^{(1)}$ be independent copies of Y_g and write $\Delta I_{g,i} := I_{g,i}(Y_g^{(0)}) \oplus I_{g,i}(Y_g^{(1)})$. Then

$$|\mathcal{R}|^{2^t} \le \mathbb{E}\left[(-1)^{\sum_{i=1}^{q} R_i \prod_{g=1}^{t} \Delta I_{g,i}} \right].$$

The expectation on the right is nonnegative.

Proof. For $s \in \{0, \dots, t\}$ define

$$\Gamma_s := \mathbb{E}\left[(-1)^{\sum_i R_i (\prod_{g=1}^{s} \Delta I_{g,i})(\prod_{g=s+1}^{t} I_{g,i})} \right].$$

Thus $\Gamma_0 = \mathcal{R}$.

We prove inductively that for every $s \ge 1$, $|\mathcal{R}|^{2^s} \le \Gamma_s$ and $\Gamma_s \ge 0$. For $s = 1$, condition on all variables except Y_1. If $\Phi_0(X, Y_1) = (-1)^{\sum_i R_i I_{1,i} \prod_{g=2}^{t} I_{g,i}}$, then

$$|\mathcal{R}|^2 \le \mathop{\mathbb{E}}_{X}\left[\left| \mathop{\mathbb{E}}_{Y_1}\left[\Phi_0(X, Y_1) \right] \right|^2 \right] = \Gamma_1.$$

The middle quantity is an expectation of a squared absolute value, so $\Gamma_1 \geq 0$.

Now suppose the claim holds for $s < t$. Condition in Γ_s on all variables except the single uncopied variable Y_{s+1}, and call the resulting sign $\Phi_s(X, Y_{s+1})$. By the induction hypothesis and $\Gamma_s \geq 0$,

$$|\mathcal{R}|^{2^{s+1}} \leq \Gamma_s^2.$$

Applying Cauchy–Schwarz to the conditional expectation defining Γ_s gives

$$\Gamma_s^2 \leq \mathop{\mathbb{E}}_{X} \left[\left| \mathop{\mathbb{E}}_{Y_{s+1}} [\Phi_s(X, Y_{s+1})] \right|^2 \right].$$

Expand the square using independent copies $Y_{s+1}^{(0)}, Y_{s+1}^{(1)}$. Their two $I_{s+1,i}$ values sum modulo 2 to $\Delta I_{s+1,i}$, while all previously duplicated groups remain outside the inner conditional expectation. The resulting expression is Γ_{s+1}, and it is nonnegative because it is an expectation of a square. Taking $s = t$ proves the lemma. □

Partition some of the m tribes into disjoint nonempty groups $G_1, \ldots, G_t$, and let the remaining tribes form G_0, which may be empty. For form i, define

$$I_{g,i} := \prod_{j \in G_g} (1 - B_{j,i}), \qquad R_i := \prod_{j \in G_0} (1 - B_{j,i}).$$

Then $J_i = R_i \prod_{g=1}^{t} I_{g,i}$. Applying Lemma 4.2,

$$|\mathcal{R}(L_1, \ldots, L_q)|^{2^t} \leq \mathbb{E}[(-1)^T], \qquad T := \sum_{i=1}^{q} C_i, \tag{17}$$

where $C_i := R_i \prod_{g=1}^{t} \left(I_{g,i}^{(0)} \oplus I_{g,i}^{(1)} \right)$. Thus $C_i = 1$ precisely when no tribe in G_0 evaluates to 1 for form i, and in every duplicated group G_g, one copy has no tribe evaluating to 1 while the other does. The next lemma computes their probability and shows that they retain the pairwise independence needed for the second-moment step.

Lemma 4.3 (Pairwise-independent rare events)**.** *Let $r_g := |G_g|$ and $r_0 := |G_0|$. Then every C_i is Bernoulli with*

$$\alpha := \mathbb{P}[C_i = 1] = \beta\, 2^t \prod_{g=1}^{t} \left(1 - (1-p)^{r_g}\right). \tag{18}$$

For every $i \neq i'$, the variables C_i and $C_{i'}$ are independent.

Proof. Put $a_0 := (1-p)^{r_0}$ and $a_g := (1-p)^{r_g}$. For fixed i, the group G_0 and the duplicated groups depend on disjoint sets of tribe samples. Hence

$$\mathbb{P}[R_i = 1] = a_0, \qquad \mathbb{P}[I_{g,i}^{(0)} \oplus I_{g,i}^{(1)} = 1] = 2a_g(1 - a_g).$$

Therefore $\mathbb{P}[C_i = 1] = a_0 \prod_{g=1}^{t} 2a_g(1 - a_g)$. Since $a_0 \prod_{g=1}^{t} a_g = (1-p)^m = \beta$, this is (18).

Fix $i \neq i'$. Let $\mathcal{Z}$ be the complete list of independent vector samples present after the Cauchy–Schwarz duplication: one sample for every coordinate in G_0 and two independent samples for every coordinate in a duplicated group. Define

$$V_i := (L_i(Z) : Z \in \mathcal{Z}), \qquad V_{i'} := (L_{i'}(Z) : Z \in \mathcal{Z}).$$

For each individual $Z \in \mathcal{Z}$, the pair $(L_i(Z), L_{i'}(Z))$ is uniform on $\mathbb{F}_2^2$. Since the samples in $\mathcal{Z}$ are mutually independent, the joint distribution factors over coordinates, so V_i and $V_{i'}$ are independent. The variable C_i is a deterministic function only of V_i, and $C_{i'}$ only of $V_{i'}$. Hence C_i and $C_{i'}$ are independent. □

Lemma 4.4 (Constant mean bounds the parity correlation away from 1). *There is an absolute constant $\eta \in (0,1)$ such that, whenever $c_0/2 \le \mu := q\alpha \le c_0$, we have*

$$|\mathcal{R}(L_1, \dots, L_q)|^{2^t} \le \eta.$$

Consequently, $|\mathcal{R}(L_1, \dots, L_q)| \le \eta^{1/2^t}$.

Proof. By Lemma 4.3,

$$\mathbb{E}[T] = \mu, \qquad \mathbb{E}[T(T-1)] = q(q-1)\alpha^2 \le \mu^2.$$

Since $r \le r(r-1)$ for every integer $r \ge 2$,

$$\mathbb{P}[T=1] = \mathbb{E}[T] - \mathbb{E}\big[T\mathbf{1}_{\{T\ge 2\}}\big] \ge \mu - \mathbb{E}[T(T-1)] \ge \mu - \mu^2.$$

The mean lies between the fixed positive constants $c_0/2$ and $c_0 < 1$, so $\mathbb{P}[T=1] = \Omega(1)$. Hence $\mathbb{E}[(-1)^T] = 1 - 2\mathbb{P}[T \text{ odd}] \le 1 - \Omega(1)$. The expectation is nonnegative by Lemma 4.2 and Eq. (17), which gives the claimed absolute constant $\eta \in (0,1)$. □

4.3 Choosing the tribe groups

It remains to choose the groups so that $c_0/2 \le \mu \le c_0$, as required by Lemma 4.4. We use t_q Cauchy–Schwarz rounds and vary only the size of the final group.

Lemma 4.5 (Choice of group sizes). *Assume $\frac{1}{3} \le \beta \le \frac{2}{3}$ and $p \le c_0$. Let $t = t_q$ be defined by (15), and suppose $tp \le c_0$ and $t \le \frac{m}{2}$. Then one can choose disjoint nonempty groups $G_1, \dots, G_t$ so that the resulting mean $\mu = q\alpha$ satisfies $c_0/2 \le \mu \le c_0$.*

Proof. Take $r_1 = \cdots = r_{t-1} = 1$ and let $r_t = r$ vary from 1 to $r_{\max} := m - t + 1$. The remaining tribes form G_0. By (18),

$$\mu(r) = q\beta\, 2^t p^{t-1} \left(1 - (1-p)^r\right).$$

At $r = 1$, we have $\mu(1) \le c_0$. For $t > 1$, minimality gives $q\beta(2p)^{t-1} > c_0$. This also holds for $t = 1$, since $q\beta \ge 1/3 > c_0$.

At the other endpoint,

$$1 - (1-p)^{r_{\max}} = 1 - \beta(1-p)^{-(t-1)} \ge 1 - \beta e^{2pt} > \frac{1}{4}.$$

Here $-\ln(1-p) \le 2p$ for small p, and the final inequality follows from $\beta \le 2/3$ and $tp \le c_0$ with c_0 sufficiently small. Consequently, $\mu(r_{\max}) > q\beta(2p)^{t-1}/2 > c_0/2$.

If $\mu(1) \ge c_0/2$, take $r = 1$. Otherwise take the first $r > 1$ with $\mu(r) \ge c_0/2$. Consecutive values increase by at most a factor of two, since

$$1 - (1-p)^{r+1} = 1 - (1-p)^r + p(1-p)^r \le 2\left(1 - (1-p)^r\right).$$

Thus the first crossing satisfies $\mu(r) < c_0$, as required. □

Proof of Proposition 4.1. Choose the groups supplied by Lemma 4.5 with $t = t_q$, and apply Lemma 4.4. □

The preceding proof uses φ only through p_φ. We now give the general form needed later.

Theorem 4.6 (Rare-event decorrelation). *Assume $\frac{1}{3} \le \beta_\varphi \le \frac{2}{3}$ and $p_\varphi \le c_0$. Let $L_1, \dots, L_q : V \to \mathbb{F}_2$ be distinct nonzero linear forms, let $t = t_q(\varphi)$, and suppose $tp_\varphi \le c_0$ and $t \le \frac{m}{2}$. For every $j \in [m]$ and $k \in [n]$, sample an independent uniform $Z_{j,k} \leftarrow V$, and for $i \in [q]$ write*

$$Y_i := \big((L_i(Z_{j,1}), \dots, L_i(Z_{j,n}))\big)_{j\in[m]}.$$

Then

$$\left| \mathbb{E}\left[\prod_{i=1}^{q} h_{\varphi,m}(Y_i) \right] \right| \le \eta^{1/2^t}.$$

Proof of Theorem 4.6. For block j and form i, define

$$B_{j,i} := \varphi\big(L_i(Z_{j,1}), \dots, L_i(Z_{j,n})\big),$$

the indicator that the jth local event occurs for form i. For $i \neq i'$, the two n-bit vectors

$$\big(L_i(Z_{j,1}), \dots, L_i(Z_{j,n})\big), \qquad \big(L_{i'}(Z_{j,1}), \dots, L_{i'}(Z_{j,n})\big)$$

are independent and uniform on $\mathbb{F}_2^n$. This follows coordinatewise from Lemma 3.1 and from independence of the $Z_{j,k}$. Hence $B_{j,i}$ and $B_{j,i'}$ are independent Bernoulli-p_φ variables. Different values of j use independent randomness.

The proof of Proposition 4.1 uses no further property of AND_n. After the indicators $B_{j,i}$ are defined, it uses only their Bernoulli parameter, independence across blocks, and pairwise independence across distinct forms. Substituting $p_\varphi, \beta_\varphi, t_q(\varphi)$ for p, β, t_q therefore gives the stated bound. □

4.4 Higher-order independence gives stronger cancellation

For arbitrary distinct forms, the preceding argument guarantees pairwise independence. Suppose now that every subfamily of at most k forms is linearly independent. The same Cauchy–Schwarz steps preserve this k-wise independence. We can then choose a larger mean, proportional to k, and use the extra moments to make the final parity expectation exponentially small in k.

We first show that the transformed rare events preserve the k-wise independence of the original evaluations.

Lemma 4.7 (Rare events inherit k-wise independence). *Consider the repeated Cauchy–Schwarz construction in the proof of Theorem 4.6, for an arbitrary local predicate φ. Suppose that every subfamily of at most k of $L_1, \dots, L_q$ is linearly independent. Then the transformed indicators $C_1, \dots, C_q$ are k-wise independent. Each has the Bernoulli distribution with parameter*

$$\alpha = \beta_\varphi\, 2^t \prod_{g=1}^{t} \left(1 - (1 - p_\varphi)^{r_g}\right).$$

Proof. Fix $I \subseteq [q]$ with $|I| \le k$. The forms $(L_i)_{i\in I}$ are linearly independent. Hence, for every uniform vector Z used anywhere in the construction, $(L_i(Z))_{i\in I}$ is uniform on $\mathbb{F}_2^I$. The vectors Z used for different tribe coordinates and for the independent copies created by Cauchy–Schwarz are mutually independent. It follows that the complete evaluation vectors

$$V_i := (L_i(Z) : Z \in \mathcal{Z}), \qquad i \in I,$$

are mutually independent. Each C_i is a deterministic function of V_i. Thus $(C_i)_{i\in I}$ are mutually independent. Their common marginal is the Bernoulli-α distribution computed in Lemma 4.3. □

Under full independence the parity expectation would be $(1-2\alpha)^q$. To compare with this value using only k-wise independence, we approximate $(-1)^T$ by a polynomial of degree below k. Its expectation is determined by the moments we still control.

Lemma 4.8 (Parity under k-wise independence). *Let $X_1, \dots, X_q$ be k-wise independent Bernoulli-α random variables, let $T := \sum_{i=1}^{q} X_i$ and $\mu := q\alpha$, and assume $4\mu \le k$. Then*

$$\left|\mathbb{E}[(-1)^T] - (1-2\alpha)^q\right| \le \left(O\left(\frac{\mu}{k}\right)\right)^k. \tag{19}$$

The implied constant is absolute. Consequently, there are absolute constants $c > 0$ and $k_0 \in \mathbb{N}$ such that, whenever $k_0 \le k < q$ and

$$\frac{c_0 k}{2} \le \mu \le c_0 k, \tag{20}$$

we have

$$\left|\mathbb{E}[(-1)^T]\right| \le \exp(-ck).$$

Proof. Define the degree-$(k-1)$ polynomial on nonnegative integers $P_{k-1}(x) := \sum_{r=0}^{k-1} (-2)^r \binom{x}{r}$. For integers $0 \le x < k$, the binomial theorem gives $P_{k-1}(x) = (1-2)^x = (-1)^x$.

We first bound the error on the event $T \ge k$. For integers $x \ge k$ and $r < k$,

$$\binom{x}{r} \le \binom{k}{r}\binom{x}{k}.$$

Indeed, $\binom{x}{k}\binom{k}{r} = \binom{x}{r}\binom{x-r}{k-r}$. Therefore

$$|(-1)^x - P_{k-1}(x)| \le \exp(O(k))\binom{x}{k}.$$

Taking expectations and using k-wise independence, $\mathbb{E}\left[\binom{T}{k}\right] = \binom{q}{k}\alpha^k$, so

$$\left|\mathbb{E}[(-1)^T] - \mathbb{E}[P_{k-1}(T)]\right| \le \exp(O(k))\binom{q}{k}\alpha^k \le \left(O\left(\frac{\mu}{k}\right)\right)^k. \tag{21}$$

For every $r < k$, k-wise independence also gives $\mathbb{E}\left[\binom{T}{r}\right] = \binom{q}{r}\alpha^r$. Hence

$$\mathbb{E}[P_{k-1}(T)] = \sum_{r=0}^{k-1} (-2)^r \binom{q}{r}\alpha^r,$$

which is the first k terms in the expansion of $(1-2\alpha)^q$. If $4\mu \le k$, then for every $r \ge k$ the absolute values of consecutive terms in the omitted tail have ratio at most

$$\frac{q-r}{r+1}\,2\alpha \le \frac{2\mu}{k} \le \frac{1}{2}.$$

Thus

$$|(1-2\alpha)^q - \mathbb{E}[P_{k-1}(T)]| \le O\left(\binom{q}{k}(2\alpha)^k\right) \le \left(O\left(\frac{\mu}{k}\right)\right)^k. \tag{22}$$

Combining (21) and (22) proves (19).

Now assume (20). Since $k < q$ and $\mu \le c_0 k$, we have $\alpha = \mu/q < c_0 < 1/2$. Thus $|(1-2\alpha)^q| \le e^{-2\mu} = e^{-\Omega(k)}$. The approximation error in (19) is at most $(O(c_0))^k = e^{-\Omega(k)}$ for sufficiently small c_0. For $k \ge k_0$, the sum of these bounds is at most e^{-ck} for suitable absolute constants $c > 0$ and k_0. □

We now choose the groups so that $\mathbb{E}[T]$ lies in the interval (20), using the same absolute constant c_0. For $1 \le k < q$, define $t_{q,k}(\varphi)$ to be the least integer $t \ge 1$ such that

$$q\beta_\varphi(2p_\varphi)^t \le c_0 k. \tag{23}$$

When φ is clear from context we write $t_{q,k}$.

Lemma 4.9 (Group sizes for the stronger bound). *Assume $\frac{1}{3} \le \beta_\varphi \le \frac{2}{3}$ and $p_\varphi \le c_0$. Let $1 \le k < q$, let $t = t_{q,k}(\varphi)$, and suppose $tp_\varphi \le c_0$ and $t \le \frac{m}{2}$. Then one can choose disjoint nonempty groups $G_1, \dots, G_t$ so that the mean $\mu = q\alpha$ of the transformed rare-event count satisfies*

$$\frac{c_0 k}{2} \le \mu \le c_0 k.$$

Proof. Use the same one-parameter choice as in Lemma 4.5. The first $t-1$ groups are single blocks and the last has size $r \in [1, m-t+1]$. Then

$$\mu(r) = q\beta_\varphi 2^t p_\varphi^{t-1}\left(1-(1-p_\varphi)^r\right).$$

The definition of t gives $\mu(1) \le c_0 k$. Minimality gives $q\beta_\varphi(2p_\varphi)^{t-1} > c_0 k$; when $t = 1$, this follows from $q\beta_\varphi \ge q/3 > c_0 k$. The endpoint estimate from Lemma 4.5 therefore gives $\mu(m-t+1) > c_0 k/2$. If $\mu(1) \ge c_0 k/2$, take $r = 1$. Otherwise take the first $r > 1$ with $\mu(r) \ge c_0 k/2$. Consecutive values increase by at most a factor of two, by the estimate in Lemma 4.5, so this crossing satisfies $\mu(r) < c_0 k$. In either case $c_0 k/2 \le \mu(r) \le c_0 k$. □

Theorem 4.10 (Rare-event decorrelation with no short dependencies). *There are absolute constants $c > 0$ and $k_0 \in \mathbb{N}$ with the following property. Assume $\frac{1}{3} \le \beta_\varphi \le \frac{2}{3}$ and $p_\varphi \le c_0$. Let $L_1, \dots, L_q : V \to \mathbb{F}_2$ be distinct nonzero linear forms such that every subfamily of at most k forms is linearly independent, where $k_0 \le k < q$. Let $t = t_{q,k}(\varphi)$ be defined by (23), and suppose $tp_\varphi \le c_0$ and $t \le \frac{m}{2}$. With Y_i and $h_{\varphi,m}$ as in Theorem 4.6,*

$$\left|\mathbb{E}\left[\prod_{i=1}^{q} h_{\varphi,m}(Y_i)\right]\right| \le \exp\left(-\frac{ck}{2^t}\right). \tag{24}$$

Proof. Choose groups using Lemma 4.9 and perform the same Cauchy–Schwarz iteration as in Lemma 4.2. Writing $\mathcal{R} := \mathbb{E}\left[\prod_{i=1}^{q} h_{\varphi,m}(Y_i)\right]$, we obtain

$$|\mathcal{R}|^{2^t} \le \mathbb{E}[(-1)^T], \qquad T = \sum_{i=1}^{q} C_i.$$

By Lemma 4.7, the C_i are k-wise independent Bernoulli-α variables. Their mean lies in the interval (20), so Lemma 4.8 gives

$$\left|\mathbb{E}[(-1)^T]\right| \le e^{-ck}.$$

The expectation on the right of the Cauchy–Schwarz inequality is nonnegative, by Lemma 4.2. Taking the 2^tth root proves (24). □

From (23),

$$t_{q,k}(\varphi) \le 1 + \frac{\log(q/k) + O(1)}{\log(1/(2p_\varphi))}. \tag{25}$$

We will use this estimate with $p_\varphi = \Theta(1/\lambda)$ and with k coming from the minimum size of a nonzero dependency among the random samples.

5 Application to the BDEKRS candidate

We now apply Theorem 4.6 to the candidate $\mathcal{F}_{\mathrm{BDEKRS}}$.

Theorem 5.1 (General non-adaptive linear tests). *Assume $\frac{1}{3} \le \beta \le \frac{2}{3}$ and $p \le c_0$. Let $x_1, \dots, x_q \in \mathbb{F}_2^N$ be distinct and suppose $t_{2q} p \le c_0$ and $t_{2q} \le \frac{m}{2}$. Then*

$$\left| \mathop{\mathbb{E}}_{\substack{A \leftarrow \mathrm{GL}_N(\mathbb{F}_2) \\ b \leftarrow \mathbb{F}_2^N}} \left[(-1)^{\sum_{i=1}^q F_{A,b}(x_i)} \right] \right| \le 4\eta^{\ell / 2^{t_{2q}+1}}. \tag{26}$$

Here $\eta \in (0,1)$ is the absolute constant from Proposition 4.1. In particular, for an absolute constant $c > 0$,

$$\mathrm{Bias}(x_1, \dots, x_q) \le 4 \exp\left(-\frac{c\ell}{2^{t_{2q}}} \right).$$

5.1 From one block to the full candidate

We first sample the secret matrix uniformly from all binary matrices, so its row blocks are independent. For a fixed row block R, averaging the output parity over its shift gives a real number $z(R)$. Independence will let us multiply bounds on $\mathbb{E}_R[|z(R)|]$ across blocks. Taking the absolute value before averaging also lets us condition on invertibility at the end: this conditioning costs at most the reciprocal of its probability. A second moment supplies the required bound on $\mathbb{E}_R[|z(R)|]$.

Fix distinct PRF queries $x_1, \dots, x_q \in \mathbb{F}_2^N$. For $R \in \mathbb{F}_2^{M \times N}$ define

$$z(R) := \mathop{\mathbb{E}}_{c \leftarrow \mathbb{F}_2^M} \left[\prod_{i=1}^q h(Rx_i \oplus c) \right].$$

Lemma 5.2 (Second moment). *Under the assumptions of Theorem 5.1, for uniform $R \leftarrow \mathbb{F}_2^{M \times N}$,*

$$\mathop{\mathbb{E}}_R[z(R)^2] \le \eta^{1/2^{t_{2q}}}.$$

Consequently,

$$\mathop{\mathbb{E}}_R[|z(R)|] \le \eta^{1/2^{t_{2q}+1}}.$$

Proof. Let $c, c' \leftarrow \mathbb{F}_2^M$ be independent. Expanding the square,

$$\mathop{\mathbb{E}}_R[z(R)^2] = \mathop{\mathbb{E}}_{R,c,c'} \left[\prod_{i=1}^q h(Rx_i \oplus c) h(Rx_i \oplus c') \right].$$

For one row $a \in \mathbb{F}_2^N$ of R and corresponding shift coordinates $\sigma, \sigma' \in \mathbb{F}_2$, define the $2q$ forms on $\mathbb{F}_2^{N+2}$ by

$$L_i^{(0)}(a, \sigma, \sigma') = \langle a, x_i \rangle + \sigma, \qquad L_i^{(1)}(a, \sigma, \sigma') = \langle a, x_i \rangle + \sigma'.$$

Their coefficient vectors are $(x_i, 1, 0)$ and $(x_i, 0, 1)$. They are all nonzero, including when $x_i = 0$, and they are pairwise distinct. Equal superscripts use the distinctness of the x_i, while different superscripts differ in the last two coordinates.

To identify the preceding expectation with the one-block problem, index the $M = mn$ rows of R by pairs $(j,k) \in [m] \times [n]$. If $a_{j,k}$ is the corresponding row and $c_{j,k}, c'_{j,k}$ are the corresponding shift

coordinates, then $Z_{j,k} := (a_{j,k}, c_{j,k}, c'_{j,k})$ is uniform in $\mathbb{F}_2^{N+2}$, and the $Z_{j,k}$ are independent across (j,k). Evaluating the $2q$ forms $L_i^{(0)}, L_i^{(1)}$ on these vectors produces the coordinate bits appearing in $h(Rx_i \oplus c)$ and $h(Rx_i \oplus c')$. Thus the second-moment expression is precisely the one-block correlation of these $2q$ forms.

Applying Proposition 4.1 therefore gives $\mathbb{E}_R[z(R)^2] \le \eta^{1/2^{t_{2q}}}$. Finally,

$$\mathop{\mathbb{E}}_R[|z(R)|] \le \sqrt{\mathop{\mathbb{E}}_R[z(R)^2]} \le \eta^{1/2^{t_{2q}+1}}.$$

□

Lemma 5.3 (Invertibility has constant probability). *If $\widetilde{A} \leftarrow \mathbb{F}_2^{N\times N}$ is uniform, then*

$$\mathbb{P}[\widetilde{A} \in \mathrm{GL}_N(\mathbb{F}_2)] = \prod_{j=1}^{N}(1-2^{-j}) \ge \frac{1}{4}.$$

Proof. Exposing the columns successively gives the product formula. Moreover,

$$\prod_{j=2}^{N}(1-2^{-j}) \ge 1 - \sum_{j=2}^{N} 2^{-j} \ge \frac{1}{2}.$$

The $j=1$ factor is $1/2$, so the full product is at least $1/4$. □

Proof of Theorem 5.1. Write a matrix $A \in \mathbb{F}_2^{N\times N}$ in ℓ row blocks

$$A = \begin{pmatrix} A_1 \\ \vdots \\ A_\ell \end{pmatrix}, \qquad A_w \in \mathbb{F}_2^{M\times N}.$$

For fixed A, averaging over the independent shift blocks $b_1, \dots, b_\ell$ gives

$$\mathop{\mathbb{E}}_b\left[\prod_{i=1}^{q}(-1)^{F_{A,b}(x_i)}\right] = \prod_{w=1}^{\ell} z(A_w) =: W(A).$$

Let $\widetilde{A} \leftarrow \mathbb{F}_2^{N\times N}$ be uniform and let $E := \{\widetilde{A} \in \mathrm{GL}_N(\mathbb{F}_2)\}$. Conditioned on E, the matrix $\widetilde{A}$ is uniform in $\mathrm{GL}_N(\mathbb{F}_2)$. Therefore

$$\mathrm{Bias}(x_1,\dots,x_q) = |\,\mathbb{E}\big[W(\widetilde{A}) \mid E\big]| \le \mathbb{E}\big[|W(\widetilde{A})| \mid E\big] \le \frac{\mathbb{E}[|W(\widetilde{A})|]}{\mathbb{P}[E]}.$$

Before conditioning on E, the row blocks $\widetilde{A}_1, \dots, \widetilde{A}_\ell$ are independent uniform $M \times N$ matrices. Hence, by Lemma 5.2,

$$\mathbb{E}[|W(\widetilde{A})|] = \left(\mathop{\mathbb{E}}_R[|z(R)|]\right)^{\ell} \le \eta^{\ell/2^{t_{2q}+1}}.$$

Combining with Lemma 5.3,

$$\mathrm{Bias}(x_1,\dots,x_q) \le 4\eta^{\ell/2^{t_{2q}+1}}.$$

Writing $c = -\frac{1}{2}\ln\eta > 0$ gives

$$\eta^{\ell/2^{t_{2q}+1}} = \exp\left(-\frac{c\ell}{2^{t_{2q}}}\right).$$

□

Corollary 5.4 (Parity bound before conditioning on invertibility). *Under the hypotheses of Theorem 5.1, if $\widetilde{A} \leftarrow \mathbb{F}_2^{N\times N}$ is uniform, then*

$$\left| \mathop{\mathbb{E}}_{\widetilde{A},b} \left[\prod_{i=1}^{q} (-1)^{F_{\widetilde{A},b}(x_i)} \right] \right| \le \eta^{\ell/2^{t_{2q}+1}}.$$

Here $F_{\widetilde{A},b}$ denotes the same evaluation formula as in (1), without conditioning $\widetilde{A}$ on invertibility.

Proof. With $\widetilde{A}$ uniform over all matrices, the bias is $|\,\mathbb{E}[W(\widetilde{A})]|$. The block-product bound in the proof of Theorem 5.1 gives

$$|\,\mathbb{E}[W(\widetilde{A})]| \le \mathbb{E}[|W(\widetilde{A})|] \le \eta^{\ell/2^{t_{2q}+1}}.$$

□

5.2 Parameter consequences

Each Cauchy–Schwarz round takes a square root of the final bound. To obtain the query ranges in the introduction, we therefore estimate how many rounds are needed as the query count grows.

Lemma 5.5 (Growth of t_s). *Assume $p \le c_0$. For every $s \ge 1$, $t_s \le 1 + \frac{\log s + O(1)}{\log(1/(2p))}$. In particular,*

$$t_s = O\left(1 + \frac{\log s}{\log(1/(2p))}\right), \tag{27}$$

with an absolute constant in the $O(\cdot)$ notation.

Proof. Put $D := \log(1/(2p)) = \Omega(1)$ and take $t := \lceil (\log s + \log(1/c_0))/D \rceil$. Since $\beta \le 1$, we have $s\beta(2p)^t \le c_0$. Thus $t_s \le t \le 1 + (\log s + O(1))/D$, where the additive constant is absolute because c_0 is fixed. This also gives (27). □

Corollary 5.6 (Every fixed polynomial number of queries). *For every fixed $C > 0$, uniformly for all $q \le \lambda^C$ and every set of q distinct chosen queries,*

$$\mathrm{Bias}(x_1, \dots, x_q) \le \exp(-\Omega_C(\lambda)).$$

Equivalently, the candidate resists every non-adaptive linear test making an arbitrary fixed polynomial number of queries in the key length κ.

Proof of Corollary 5.6. For the balanced parameters, $p = \Theta(\log\lambda/\lambda)$, $\beta = 1/2 + o(1)$, $\ell = \lambda$, and $m = \Theta(\lambda/\log\lambda)$. If $q \le \lambda^C$ for fixed C, then $\log(1/(2p)) = (1 + o(1))\log\lambda$, and (27) gives $t_{2q} = O_C(1)$. Hence $t_{2q}p = o(1)$ and $t_{2q}/m = o(1)$, so the hypotheses of Theorem 5.1 hold for all sufficiently large λ. Therefore

$$\mathrm{Bias}(x_1, \dots, x_q) \le \exp(-\Omega_C(\lambda)).$$

Finally, $\kappa = \lambda^4 + \lambda^2$, so every fixed polynomial in κ is a fixed polynomial in λ. □

Corollary 5.7 (Quasipolynomial linear-test regime). *Fix $\theta \in (0,1)$ and put $Q := \kappa^{\log^\theta \kappa}$. Uniformly for all $q \le Q$ and every set of q distinct chosen queries,*

$$\mathrm{Bias}(x_1, \dots, x_q) \le \exp(-\lambda^{1-o(1)}).$$

Consequently, for every fixed $0 < \rho < 1/4$ and all sufficiently large κ,

$$q \le Q \qquad \Longrightarrow \qquad \mathrm{Bias}(x_1, \dots, x_q) \le 2^{-\kappa^\rho}.$$

Proof of Corollary 5.7. Let $Q = \kappa^{\log^\theta \kappa}$. Since $\kappa = \Theta(\lambda^4)$,

$$\log Q = (\log \kappa)^{1+\theta} = \Theta((\log \lambda)^{1+\theta}).$$

Hence $t_{2Q} = O((\log \lambda)^\theta)$. For every fixed $0 < \theta < 1$, $t_{2Q}p = o(1)$ and $t_{2Q}/m = o(1)$, so Theorem 5.1 applies uniformly for every $q \le Q$. Moreover,

$$2^{t_{2Q}} = \exp(O((\log \lambda)^\theta)) = \lambda^{o(1)}.$$

Therefore

$$\mathrm{Bias}(x_1, \dots, x_q) \le \exp(-\lambda^{1-o(1)}).$$

Finally fix $0 < \rho < 1/4$. Since $\kappa^\rho = \Theta(\lambda^{4\rho}) = o(\lambda^{1-o(1)})$, the same estimate is eventually at most $2^{-\kappa^\rho}$. □

Keeping the leading coefficient of $(\log \lambda)^2$ in the query exponent gives the following more precise tradeoff.

Corollary 5.8 (Explicit query/advantage tradeoff)**.** *Fix $a \in (0,1)$ and put $Q_a := 2^{a(\log_2 \lambda)^2}$. Uniformly for all $q \le Q_a$ and every set of q distinct chosen queries,*

$$\mathrm{Bias}(x_1, \dots, x_q) \le \exp\left(-\lambda^{1-a-o(1)}\right).$$

In key-length notation, if $a_{\mathrm{qry}}, \rho > 0$ are fixed and $16a_{\mathrm{qry}} + 4\rho < 1$, then for all sufficiently large κ,

$$q \le 2^{a_{\mathrm{qry}}(\log_2 \kappa)^2} \qquad \Longrightarrow \qquad \mathrm{Bias}(x_1, \dots, x_q) \le 2^{-\kappa^\rho}.$$

Proof of Corollary 5.8. Fix $a \in (0,1)$ and let $Q_a = 2^{a(\log \lambda)^2}$. For the balanced parameters, $D_\lambda := \log \frac{1}{2p} = (1 + o(1)) \log \lambda$. By Lemma 5.5, uniformly for $q \le Q_a$,

$$t_{2q} \le 1 + \frac{\log(2Q_a) + O(1)}{D_\lambda} = (a + o(1)) \log \lambda.$$

Hence $2^{t_{2q}} \le \lambda^{a+o(1)}$. Also

$$t_{2q}p = O\left(\frac{(\log \lambda)^2}{\lambda}\right) = o(1), \qquad \frac{t_{2q}}{m} = O\left(\frac{(\log \lambda)^2}{\lambda}\right) = o(1),$$

so the hypotheses of Theorem 5.1 hold for all sufficiently large λ. Since $\ell = \lambda$, Theorem 5.1 gives

$$\mathrm{Bias}(x_1, \dots, x_q) \le 4 \exp\left(-\frac{c\lambda}{2^{t_{2q}}}\right) \le \exp\left(-\lambda^{1-a-o(1)}\right).$$

For the key-length formulation, fix $a_{\mathrm{qry}}, \rho > 0$ with $16a_{\mathrm{qry}} + 4\rho < 1$. Since $\log \kappa = (4 + o(1)) \log \lambda$, choose a fixed a satisfying $16a_{\mathrm{qry}} < a < 1 - 4\rho$. For all sufficiently large λ, $2^{a_{\mathrm{qry}}(\log \kappa)^2} \le 2^{a(\log \lambda)^2}$. The first part therefore gives

$$\mathrm{Bias}(x_1, \dots, x_q) \le \exp\left(-\lambda^{1-a-o(1)}\right).$$

Because $1 - a > 4\rho$ and $\kappa^\rho = \Theta(\lambda^{4\rho})$, this is eventually at most $2^{-\kappa^\rho}$. □

6 Application to the BCGIKS Sipser-noise family

For the Sipser candidate, the adversary receives uniform random inputs and chooses which output bits to XOR after seeing those inputs. BCGIKS [BCG+21] propose a weak-PRF candidate in $\mathrm{AC}^0 \circ \mathrm{MOD}_2$ that uses a biased depth-three Sipser predicate as deterministic noise after a secret linear map. They analyze its algebraic structure and leave resistance of this candidate to linear attacks open. We show that random samples supply the additional independence needed by Theorem 4.10.

6.1 The Sipser noise family and linear attacks

Let $m_\lambda, r_\lambda = \Theta(\lambda)$, $w_\lambda = \Theta(\log \lambda)$, and $M_\lambda := m_\lambda r_\lambda w_\lambda$. For $z = (z_{a,b,c}) \in \mathbb{F}_2^{M_\lambda}$, with $a \in [m_\lambda]$, $b \in [r_\lambda]$, and $c \in [w_\lambda]$, define

$$G_\lambda(z) := \bigvee_{a=1}^{m_\lambda} \bigwedge_{b=1}^{r_\lambda} \bigvee_{c=1}^{w_\lambda} z_{a,b,c}. \tag{28}$$

BCGIKS [BCG+21] use this depth-three Sipser formula with $w_\lambda \simeq \log \lambda - \log\log \lambda$ and choose its integer fan-ins so that each output value has constant probability. We use the following parameter bounds. Put

$$\varphi_\lambda(z) := \bigwedge_{b=1}^{r_\lambda} \bigvee_{c=1}^{w_\lambda} z_{b,c}, \qquad p_\lambda := \mathbb{P}_{U \leftarrow \mathbb{F}_2^{r_\lambda w_\lambda}}[\varphi_\lambda(U) = 1].$$

We choose the fan-ins so that $p_\lambda = \Theta(1/\lambda)$ and $\frac{1}{3} \le \beta_\lambda := (1-p_\lambda)^{m_\lambda} \le \frac{2}{3}$ for all sufficiently large λ. In particular, $M_\lambda = \Theta(\lambda^2 \log \lambda)$. These bounds suffice for the linear-test result. For the sharp leading constant in the rational-distance result, we choose the fan-ins more precisely so that $\beta_\lambda = 1/2 + o(1)$.

Following BCGIKS [BCG+21], consider their family

$$f_{s,K}(x) := \langle s, x \rangle \oplus G_\lambda(Kx), \tag{29}$$

where $n_\lambda := M_\lambda + 1$, $s \leftarrow \mathbb{F}_2^{n_\lambda}$, and $K \leftarrow \mathbb{F}_2^{M_\lambda \times n_\lambda}$ are independent and uniform. This is an instance of their framework of *learning parity with simple deterministic noise*: $G_\lambda(Kx)$ is the noise added to the secret parity. The public inputs are uniform random samples, and an attack may choose a linear combination of the output bits after seeing those inputs.

For a fixed sample tuple $X = (x_1, \dots, x_Q)$ and $u \in \mathbb{F}_2^Q \setminus \{0\}$, write

$$\mathrm{Bias}_{\mathrm{Sip}}(X; u) := \left| \mathop{\mathbb{E}}_{s,K} \left[(-1)^{\sum_{i=1}^Q u_i f_{s,K}(x_i)} \right] \right|.$$

If $\sum_i u_i x_i \neq 0$, then

$$\mathrm{Bias}_{\mathrm{Sip}}(X; u) = 0, \tag{30}$$

because the expectation over the uniform s vanishes. Thus only coefficient vectors in the linear relation space of the public samples can contribute.

Our main result for this family is the following.

Theorem 6.1 (Quasipolynomial linear tests for the Sipser-based family). *Fix $0 < \varepsilon < 2$ and let $Q \le 2^{(2-\varepsilon)(\log_2 \lambda)^2}$. Let $x_1, \dots, x_Q \leftarrow \mathbb{F}_2^{n_\lambda}$ be independent and uniform. With probability $1 - 2^{-\Omega(n_\lambda)}$ over these public samples,* simultaneously for every *nonzero* $u \in \mathbb{F}_2^Q$,

$$\left| \mathop{\mathbb{E}}_{s,K} \left[(-1)^{\sum_{i=1}^Q u_i f_{s,K}(x_i)} \right] \right| \le \exp(-\lambda^{\varepsilon - o(1)}).$$

6.2 Random samples have no short linear dependencies

We use a union bound to rule out short dependencies in the entire sample tuple. On this event, the stronger correlation bound applies to every attack vector chosen after seeing the tuple.

Lemma 6.2 (No short dependencies among random binary samples). *Let $x_1, \dots, x_Q \leftarrow \mathbb{F}_2^{n_0}$ be independent and uniform. For every integer $k \geq 1$,*

$$\mathbb{P}\left[\exists \emptyset \neq I \subseteq [Q],\ |I| \leq k : \sum_{i \in I} x_i = 0\right] \leq 2^{-n_0} \sum_{r=1}^{k} \binom{Q}{r}. \tag{31}$$

In particular, put $k := \left\lfloor \frac{c_0 n_0}{\log_2(eQ)} \right\rfloor$. Whenever $k \geq 1$, the probability on the left-hand side of (31) is at most $2^{-\Omega(n_0)}$.

Proof. For a fixed nonempty $I \subseteq [Q]$, the sum $\sum_{i \in I} x_i$ is uniform on $\mathbb{F}_2^{n_0}$, and hence vanishes with probability 2^{-n_0}. Summing over all I of size at most k gives (31).

For the stated choice of k,

$$2^{-n_0} \sum_{r=1}^{k} \binom{Q}{r} \leq 2^{-n_0 + O(k \log(eQ))} = 2^{-\Omega(n_0)},$$

where the last step uses the smallness of c_0. □

6.3 Quasipolynomial linear-test security

We first fix a sample tuple with no short dependency. If the attack coefficients do not define a relation, averaging over the secret parity gives zero bias. Otherwise, that relation must be long, and every small subfamily of the corresponding linear forms is linearly independent. This is the hypothesis needed for the stronger correlation theorem. Since the property holds for the whole tuple, the attack vector can be chosen after the samples are seen.

Proof of Theorem 6.1. Put $k_{\mathrm{dep}} := \left\lfloor \frac{c_0 n_\lambda}{\log_2(eQ)} \right\rfloor$. By Lemma 6.2, except with probability $2^{-\Omega(n_\lambda)}$, the sample tuple has no nonzero linear dependency of size at most k_{dep}. Fix a tuple in this event and a nonzero attack vector u.

If $\sum_i u_i x_i \neq 0$, (30) gives zero bias. Suppose therefore that $\sum_{i=1}^{Q} u_i x_i = 0$. Let $S := \mathrm{supp}(u)$ and $q := |S|$. The absence of short dependencies implies $q > k_{\mathrm{dep}}$. For each row $a \in \mathbb{F}_2^{n_\lambda}$ of K, the corresponding bit of Kx_i is the linear form $L_i(a) := \langle a, x_i \rangle$. On the support S, every subfamily of at most k_{dep} of these forms is linearly independent.

The Sipser predicate in (28) is an OR of $m = m_\lambda = \Theta(\lambda)$ independent copies of φ_λ. We will apply Theorem 4.10 with $k = k_{\mathrm{dep}}$, $p_\varphi = p_\lambda = \Theta(1/\lambda)$, and $\beta_\varphi = \beta_\lambda \in [1/3, 2/3]$. Let $t = t_{q,k_{\mathrm{dep}}}$. Since $q \leq Q$, monotonicity in the number of forms gives $t_{q,k_{\mathrm{dep}}} \leq t_{Q,k_{\mathrm{dep}}}$. The growth estimate (25) yields

$$t_{Q,k_{\mathrm{dep}}} \leq 1 + \frac{\log(Q/k_{\mathrm{dep}}) + O(1)}{\log(1/(2p_\lambda))}. \tag{32}$$

Because $n_\lambda = \Theta(\lambda^2 \log \lambda)$ and $\log Q \leq (2 - \varepsilon)(\log \lambda)^2$, we have

$$k_{\mathrm{dep}} = \Omega\left(\frac{\lambda^2}{\log \lambda}\right). \tag{33}$$

Also $\log(1/(2p_\lambda)) = (1+o(1))\log\lambda$, so

$$t_{Q,k_{\mathrm{dep}}} \le (2-\varepsilon+o(1))\log_2\lambda \tag{34}$$

and therefore

$$2^{t_{Q,k_{\mathrm{dep}}}} \le \lambda^{2-\varepsilon+o(1)}.$$

In particular, $t_{Q,k_{\mathrm{dep}}}p_\lambda = o(1)$ and, since $m_\lambda = \Theta(\lambda)$, $t_{Q,k_{\mathrm{dep}}} = o(m_\lambda)$. Thus the hypotheses of Theorem 4.10 hold for all sufficiently large λ.

Combining (33) with the preceding bound gives

$$\frac{k_{\mathrm{dep}}}{2^{t_{q,k_{\mathrm{dep}}}}} \ge \frac{k_{\mathrm{dep}}}{2^{t_{Q,k_{\mathrm{dep}}}}} \ge \lambda^{\varepsilon-o(1)}.$$

Applying Theorem 4.10 therefore gives

$$\left|\mathop{\mathbb{E}}_{K}\left[(-1)^{\sum_{i\in S}G_\lambda(Kx_i)}\right]\right| \le \exp(-\lambda^{\varepsilon-o(1)}).$$

Since $\sum_{i\in S}x_i = 0$, the secret parity term cancels. This gives the claimed bound. □

7 Distance from low rational degree

The preceding bounds average output parities over the key. We now fix a key and bound the error of a nontrivial low-degree equation $fP_0 = P_1$. Boyle et al. identify proximity to low rational degree as an open question for their Sipser candidate [BCG+21, Section 4.1]. We determine the optimal distance scale for this candidate and give a companion bound for the candidate of Ball et al.

We identify polynomials over $\mathbb{F}_2$ with their multilinear representations on the Boolean cube.

Write $\mathrm{RD}(v)$ for the least d for which there are polynomials P_0, P_1, not both zero, of degree at most d such that $vP_0 = P_1$ everywhere. An identity imposes no restriction where both polynomials vanish. Equivalently, $\mathrm{RD}(v) \le d$ if v or $1+v$ has a nonzero annihilator of degree at most d. This minimum degree is also called the algebraic immunity of v. Distance to functions with low-degree annihilators was studied in probabilistic algebraic attacks [BP05] and as distance to algebraic immunity [LL10, Sections 4–6]. For a Boolean function f, let U be uniform on its domain and define

$$\Delta_d^{\mathrm{rat}}(f) := \min_{v:\,\mathrm{RD}(v)\le d}\mathbb{P}[f(U)\ne v(U)].$$

Use the Sipser parameters in (28), abbreviating $m = m_\lambda, r = r_\lambda, w = w_\lambda, M = mrw$, and $n_\lambda = M+1$. Put $p = (1-2^{-w})^r, \beta = (1-p)^m$, and $b_\lambda = \min\{\beta, 1-\beta\}$. Thus $m, r = \Theta(\lambda), w = \Theta(\log\lambda), p = \Theta(1/\lambda)$, and $\beta \in [1/3, 2/3]$.

Theorem 7.1 (Distance from low rational degree). *Fix $\varepsilon \in (0,1)$ and sample s, K as in (29). There is $\delta_\lambda = o(1)$ such that, outside a $2^{-\Omega_\varepsilon(\lambda^2)}$ fraction of keys, simultaneously for every integer $1 \le d \le (1-\varepsilon)r$ and every pair P_0, P_1 of degree at most d, not both zero,*

$$\mathbb{P}_x[f_{s,K}(x)P_0(x) \ne P_1(x)] \ge (b_\lambda - \delta_\lambda)2^{-d}, \tag{35}$$

$$(b_\lambda - \delta_\lambda)2^{-d} \le \Delta_d^{\mathrm{rat}}(f_{s,K}) \le 2^{-d-1}. \tag{36}$$

Here x is uniform on $\mathbb{F}_2^{n_\lambda}$. If $\beta = 1/2 + o(1)$, then $\Delta_d^{\mathrm{rat}}(f_{s,K}) = (1/2+o(1))2^{-d}$ uniformly in this range.

The secret map is sampled uniformly, without conditioning on invertibility. We first find inputs on which a polynomial is nonzero and the Sipser predicate has a specified value. We then pass through the parity bit and the secret map, keeping track of the degree lost when the map is singular.

7.1 Forcing probabilities and nonzero polynomials

We seek inputs on which both the polynomial is nonzero and the predicate has a specified value. Leaving the variables of a highest-degree monomial free guarantees that the restricted polynomial is nonzero, whatever values are fixed elsewhere. We therefore ask how often those other values already determine the predicate.

For $S \subseteq [N_0]$ and a Boolean function v on N_0 bits, let $v(a, U_{\overline{S}})$ denote the evaluation with assignment a on S and independent uniform bits elsewhere. An assignment outside S *forces* the value b if $v = b$ for every assignment on S.

Lemma 7.2 (Forcing and low-degree polynomials)**.** *Let $D \geq 0$ be an integer and $\alpha \geq 0$. Suppose that for every $S \subseteq [N_0]$ with $|S| \leq D$ and every $b \in \{0, 1\}$,*

$$\mathbb{P}_{U_{\overline{S}}}[v(a, U_{\overline{S}}) = b \text{ for every } a \in \mathbb{F}_2^S] \geq \alpha.$$

Then every nonzero polynomial P of degree $e \leq D$ satisfies

$$\mathbb{P}[v(U) = b,\ P(U) = 1] \geq \alpha 2^{-e} \qquad (b \in \{0, 1\}). \tag{37}$$

Proof. Take a monomial $\prod_{i \in S} x_i$ of degree e with coefficient 1 in the multilinear representation of P. For every fixing of the complementary coordinates, its coefficient in the restriction remains 1: a monomial that could change it would have support strictly containing S and hence degree greater than e. Thus each restricted polynomial is nonzero. On the e free coordinates it is 1 on at least one of the 2^e assignments. For the fraction at least α of complementary fixings that force $v = b$, all these assignments also satisfy $v = b$. □

Lemma 7.3 (Forcing the Sipser predicate)**.** *Put $a = 1 - 2^{-w}$. For $0 \leq D \leq r$, define*

$$\alpha_D := \min\left\{\beta - p(a^{-D} - 1),\quad 1 - \beta - \frac{pD}{w}\right\}.$$

For every $S \subseteq [M]$ with $|S| \leq D$, a uniform fixing of the complementary coordinates forces each of the two values of G with probability at least α_D. In particular, if $D \leq (1 - \xi)r$ for a fixed $\xi > 0$, then

$$\alpha_D \geq b_\lambda - \max\left\{p^\xi, \frac{pr}{w}\right\} = b_\lambda - o(1). \tag{38}$$

Proof. Monotonicity implies that a complementary fixing forces 0 precisely when $G(1_S, U_{\overline{S}}) = 0$, and forces 1 precisely when $G(0_S, U_{\overline{S}}) = 1$.

First set all coordinates in S to 1. For conjunction j, let t_j be the number of its bottom OR gates containing a coordinate of S. Then $\sum_j t_j \leq D$, and this conjunction is true with probability $p_j = pa^{-t_j}$. Different conjunctions remain independent. By telescoping the products and using $p_j \geq p$,

$$\mathbb{P}[G(1_S, U_{\overline{S}}) = 0] = \prod_j (1 - p_j) \geq \beta - \sum_j (p_j - p).$$

For $x, y \geq 1$, $(x - 1) + (y - 1) \leq xy - 1$. Iterating this inequality gives $\sum_j (a^{-t_j} - 1) \leq a^{-\sum_j t_j} - 1 \leq a^{-D} - 1$. This proves the first lower bound.

Now set the coordinates in S to 0. If u coordinates of a bottom OR gate are fixed, its acceptance probability changes from $1 - 2^{-w}$ to $1 - 2^{-w}2^u$. Its relative loss is

$$\theta(u) := \frac{2^{-w}(2^u - 1)}{1 - 2^{-w}} \leq \frac{u}{w} \qquad (0 \leq u \leq w).$$

Indeed, the function $2^u - 1$ is convex and lies below the line joining its values at 0 and w. If $u_{j,b}$ bits are fixed in the bth bottom OR of conjunction j, the new acceptance probability is $p'_j = p\prod_b(1-\theta(u_{j,b}))$. Hence

$$0 \le p - p'_j \le p\sum_b \theta(u_{j,b}) \le \frac{p}{w}\sum_b u_{j,b}.$$

Telescoping the outer products now gives

$$\mathbb{P}[G(0_S, U_{\overline{S}}) = 1] \ge 1 - \beta - \sum_j (p - p'_j) \ge 1 - \beta - \frac{pD}{w}.$$

Finally, $a^{-D} \le p^{-(1-\xi)}$ when $D \le (1-\xi)r$. Thus $p(a^{-D}-1) \le p^{\xi}$, while $pD/w \le pr/w = o(1)$. □

7.2 The parity bit and the secret linear map

The forcing estimate concerns the unkeyed predicate. To pass to the candidate, we must add the parity bit and precompose with the secret map. Forcing itself depends on coordinates, so we instead transfer the bound on $\mathbb{P}[v(U) = b,\ P(U) = 1]$ in (37).

Lemma 7.4 (Adding a parity bit)**.** *Suppose v satisfies (37) for every nonzero polynomial of degree at most D. Then $H(z, y) := v(z) + y$ satisfies the same bound with the same α and D.*

Proof. Write a nonzero polynomial of degree $e \le D$ as $P(z, y) = A(z) + yB(z)$, where $\deg A \le e$ and $\deg B \le e - 1$. If both A and $A + B$ are nonzero, apply (37) to the two slices $y = 0, 1$ and average; each slice contributes at least $\alpha 2^{-e}$ before averaging. If exactly one slice is nonzero, that slice has degree at most $e - 1$, since it equals B. It contributes at least $\frac{1}{2}\alpha 2^{-(e-1)} = \alpha 2^{-e}$. The case $e = 0$ has two identical nonzero slices. □

We will use the elementary fact that a nonzero binary polynomial of degree t is 1 on at least a 2^{-t} fraction of the cube. To see this, take the difference in a variable from a highest-degree monomial and induct on the degree. Wherever the difference is 1, at least one of the two original evaluations is 1.

Lemma 7.5 (Composition with a linear map)**.** *Suppose $H : \mathbb{F}_2^{N_0} \to \mathbb{F}_2$ satisfies (37) for every nonzero polynomial of degree at most D. Let $B : \mathbb{F}_2^{N_0} \to \mathbb{F}_2^{N_0}$ be linear with corank h. For every nonzero polynomial P of degree e with $e + h \le D$,*

$$\mathbb{P}_x[H(Bx) = b,\ P(x) = 1] \ge \alpha 2^{-e} \qquad (b \in \{0, 1\}).$$

Proof. We first restrict the probability bound for H to the image of B. We then account for polynomials that also depend on the kernel coordinates. Let $V = \operatorname{im} B$, and take a nonzero polynomial R on V of degree $j \le D - h$, using linear coordinates on V. Extend it to a polynomial $\widetilde{R}$ of the same degree on the ambient space. We use the standard indicator argument for restricting algebraic immunity to affine subspaces [Car21, Section 9.1.1]. If V is defined by the independent linear equations $L_1(z) = \cdots = L_h(z) = 0$, its indicator is $1_V(z) = \prod_{i=1}^{h}(1 + L_i(z))$ and has degree at most h. The polynomial $1_V\widetilde{R}$ is nonzero and has degree at most $j + h$. Applying (37) in the ambient space yields

$$2^{-h}\mathbb{P}_{v \leftarrow V}[H(v) = b,\ R(v) = 1] \ge \alpha 2^{-(j+h)}.$$

Thus the restriction of H to V satisfies the same bound for polynomials of degree at most $D - h$.

The polynomial P may also depend on coordinates in the kernel of B. Choose linear coordinates $x \leftrightarrow (v, k)$ identifying the input with $V \times \ker B$, so $Bx = v$. Express P in these coordinates and let t be its largest degree in the k coordinates. Choose a nonzero coefficient $R(v)$ of a k-monomial of degree t.

Then $\deg R \le e - t$. Whenever $R(v) = 1$, the polynomial $k \mapsto P(v,k)$ is nonzero and has degree at most t, so it is 1 on at least a 2^{-t} fraction of inputs. Consequently,

$$\begin{aligned}\mathbb{P}_{v,k}[H(v) = b,\ P(v,k) = 1] &\ge 2^{-t}\mathbb{P}_v[H(v) = b,\ R(v) = 1] \\ &\ge 2^{-t}\alpha 2^{-(e-t)} = \alpha 2^{-e}.\end{aligned}$$

□

Lemma 7.6 (Corank of a uniform binary matrix)**.** *For a uniform $N_0 \times N_0$ binary matrix B and every integer $1 \le h \le N_0$,*

$$\mathbb{P}[\operatorname{corank} B \ge h] = O(2^{-h^2}).$$

Proof. If the corank is at least h, the kernel contains an h-dimensional subspace. There are $O(2^{h(N_0-h)})$ such subspaces in $\mathbb{F}_2^{N_0}$. Indeed, there are at most $2^{N_0 h}$ ordered independent h-tuples in the ambient space, and each such subspace has $2^{h^2}\prod_{i=1}^{h}(1-2^{-i}) = \Omega(2^{h^2})$ ordered bases. The product is bounded below uniformly in h by Lemma 5.3. For any fixed such subspace, a uniform matrix vanishes on it with probability $2^{-N_0 h}$. A union bound proves the claim. □

7.3 Rational equations and optimality

Lemma 7.7 (Rational equations and approximation)**.** *Suppose f satisfies (37) through degree d. Then every pair P_0, P_1 of degree at most d, not both zero, satisfies $\mathbb{P}[fP_0 \ne P_1] \ge \alpha 2^{-d}$, and $\Delta_d^{\mathrm{rat}}(f) \ge \alpha 2^{-d}$. For every Boolean f on $N_0 \ge d$ bits, one always has $\Delta_d^{\mathrm{rat}}(f) \le 2^{-d-1}$.*

Proof. If $P_1 = 0$, then $P_0 \ne 0$ and the event $f = 1, P_0 = 1$ witnesses a failure of the equation. If $P_1 \ne 0$, the event $f = 0, P_1 = 1$ does so. Apply (37) with the corresponding value of b. If v has rational degree at most d, choose a nonzero pair with $vP_0 = P_1$. The equation $fP_0 = P_1$ can fail only where $f \ne v$. This proves the distance lower bound.

The upper bound is classical [BP05, LL10]. For completeness, take any coordinate subcube C of codimension d. Choose a majority value b of f on C, and define $v = b$ on C and $v = f$ elsewhere. Then $\mathbb{P}[f \ne v] \le \mathbb{P}[C]/2 = 2^{-d-1}$. The nonzero polynomial 1_C has degree d and satisfies $(v+b)1_C = 0$, so $\mathrm{RD}(v) \le d$. □

Proof of Theorem 7.1. Reserve $h_0 = \lfloor \varepsilon r/2 \rfloor$ degrees to account for a possible rank deficiency of the secret map. Put $D = \lfloor (1-\varepsilon)r \rfloor + h_0 \le (1-\varepsilon/2)r$. By Lemmas 7.2 and 7.3, G satisfies (37) for polynomials of degree at most D, with

$$\alpha \ge b_\lambda - \max\{p^{\varepsilon/2}, pr/w\} =: b_\lambda - \delta_\lambda.$$

Here $\delta_\lambda = o(1)$, so $\alpha > 0$ for sufficiently large λ. By Lemma 7.4, $H(z,y) = G(z) + y$ has the same bound.

The combined matrix $B = (K; s)$ is uniform in $\mathbb{F}_2^{m_\lambda \times n_\lambda}$. By Lemma 7.6,

$$\mathbb{P}[\operatorname{corank} B > h_0] = O(2^{-(h_0+1)^2}) = 2^{-\Omega_\varepsilon(\lambda^2)}.$$

For every other key, Lemma 7.5 gives the same bound for $H(Bx) = f_{s,K}(x)$ for polynomials of degree at most $\lfloor (1-\varepsilon)r \rfloor$. Apply Lemma 7.7. If $\beta = 1/2 + o(1)$, the lower and universal upper bounds have the same leading constant $1/2$. □

7.4 The XOR-of-TRIBES candidate

For BDEKRS, an invertible affine change of variables preserves both degree and uniform measure, so it suffices to study the unkeyed XOR of TRIBES blocks. Leaving d coordinates free may keep several blocks undecided. We bound the probability that every block is forced; an untouched block then supplies either XOR value with constant probability. This introduces a factor $\exp(-O(d/n))$, where n is the tribe width.

Theorem 7.8 (Distance from low rational degree for BDEKRS)**.** *Let F_0 be the XOR of ℓ disjoint copies of $\mathrm{TRIBES}_{m,n}$, and suppose $\beta = (1-2^{-n})^m \in [1/3, 2/3]$. There is an absolute constant $C > 0$ such that, for every key (A, b) with A invertible, every integer $1 \le d \le \min\{\lfloor m/2 \rfloor, \ell - 1\}$, and every pair P_0, P_1 of degree at most d, not both zero,*

$$\mathbb{P}_x[F_0(Ax+b)P_0(x) \neq P_1(x)] \ge \frac{1}{3}\, 2^{-d} \exp(-Cd/n).$$

In particular,

$$\frac{1}{3}\, 2^{-d} \exp(-Cd/n) \le \Delta_d^{\mathrm{rat}}(F_{A,b}) \le 2^{-d-1}.$$

For the balanced BDEKRS parameters, this gives

$$2^{-d-O(d/\log\lambda+1)} \le \Delta_d^{\mathrm{rat}}(F_{A,b}) \le 2^{-d-1}$$

through a degree range of order $\lambda/\log\lambda$.

Proof. Fix a set S of $e \le \min\{\lfloor m/2 \rfloor, \ell - 1\}$ free input coordinates of F_0. For each TRIBES block v, let s_v be the number of its free coordinates, and let q_v be the probability that a uniform fixing of its other coordinates does not force the block's value.

Consider a block and let t be the number of tribes containing a free coordinate. For each such tribe j, let u_j be its number of free coordinates. By monotonicity, the probability of a nonconstant restriction is the difference between the acceptance probabilities obtained by setting the free bits to 1 and to 0. With $p = 2^{-n}$, this is exactly

$$q_v = (1-p)^{m-t}\left(1 - \prod_{j:\, u_j > 0} (1 - p2^{u_j})\right).$$

Since $t \le s_v \le e \le m/2$, $q_v \le \sqrt{\beta} \le 1 - \Omega(1)$. Also, $2^{u-n} \le u/n$ for integers $1 \le u \le n$. Indeed, $2^u/u$ is nondecreasing on the positive integers. The union bound for the product above therefore gives

$$q_v \le \sum_{j:\, u_j > 0} p2^{u_j} \le \frac{s_v}{n}.$$

Thus $\ln(1-q_v) \ge -Cq_v$ for an absolute constant $C > 0$, uniformly over the blocks. Independence across blocks gives

$$\mathbb{P}[\text{every block is forced}] = \prod_v (1-q_v) \ge \exp\left(-C\sum_v q_v\right) \ge \exp(-Ce/n).$$

Because $e \le \ell - 1$, at least one block contains no free coordinate. Its output is independent of all the forcing events for the other blocks, and takes each value with probability at least $1/3$. Consequently, a random fixing of the complement of S forces each value of the entire XOR with probability at least $\frac{1}{3}\exp(-Ce/n)$. By Lemma 7.2, every nonzero polynomial R of degree e satisfies

$$\mathbb{P}[F_0(U) = c,\ R(U) = 1] \ge \frac{1}{3}\, 2^{-e} \exp(-Ce/n) \qquad (c \in \{0,1\}).$$

The bound decreases with e, so it applies uniformly through degree d with the lower bound in the theorem. Apply Lemma 7.7. Finally, an invertible affine change of variables preserves both the uniform measure and algebraic degree. It therefore preserves the minimum error of equations $fP_0 = P_1$ with $\deg P_0, \deg P_1 \le d$ and $(P_0, P_1) \neq (0, 0)$, and the distance to functions of rational degree at most d. □

7.5 Rational equations on uniform samples

Return to the Sipser family, with input dimension n_λ. The error bound under a uniform input also controls equations chosen after seeing the samples, provided there are enough samples for a union bound over all coefficient pairs.

Corollary 7.9 (No low-degree equation fits all sampled values)**.** *Fix a key satisfying Theorem 7.1, a degree d in its range, and $\tau \in (0, 1)$. Put $L_d = \sum_{j=0}^{d} \binom{n_\lambda}{j}$ and $\gamma = (b_\lambda - \delta_\lambda)2^{-d} > 0$. There is an absolute constant $C > 0$ such that, for Q independent uniform inputs, if*

$$Q \ge \frac{C}{\gamma}\left(L_d + \ln\frac{1}{\tau}\right),$$

then, with probability at least $1 - \tau$, every equation $f_{s,K}P_0 = P_1$, where P_0, P_1 have degree at most d and are not both zero, fails on at least a $\gamma/2$ fraction of the samples. In particular, the matrix with rows

$$\left(f_{s,K}(x_i)\,(x_i^S)_{|S|\le d}, \quad (x_i^S)_{|S|\le d}\right)$$

has full column rank $2L_d$.

Proof. For a fixed nonzero pair, the failure indicators are independent Bernoulli variables with common mean at least γ. A multiplicative Chernoff bound shows that the fraction of failures is below $\gamma/2$ with probability at most $e^{-\Omega(\gamma Q)}$. There are fewer than 2^{2L_d} nonzero pairs, so a union bound gives an overall probability at most $\exp(O(L_d) - \Omega(\gamma Q))$. This is at most τ under the stated sample bound for sufficiently large C. A nonzero vector in the kernel of the displayed matrix would give an equation that holds on every sample. □

The scale 2^{-d} also appears in the universal upper bound: every Boolean function admits an equation of degree at most d that fails on at most a 2^{-d-1} fraction of all inputs.

8 Further consequences for the BDEKRS candidate

A distinguisher may combine the response bits by an arbitrary function. Expanding that function in the Fourier basis expresses its advantage in terms of the parity biases we have bounded. This gives statistical security for a smaller number of queries. For constant-depth circuits, bounded independence lets us retain the quasipolynomial query range.

8.1 Fixed non-adaptive query sets

Fix distinct inputs $X = (x_1, \dots, x_q)$ and define

$$Z_X := \left((-1)^{F_{A,b}(x_1)}, \dots, (-1)^{F_{A,b}(x_q)}\right) \in \{-1, 1\}^q.$$

Put $P_X := \mathsf{Law}(Z_X)$. For $S \subseteq [q]$, write $\mu_X(S) := \mathbb{E}\left[\prod_{i\in S}(-1)^{F_{A,b}(x_i)}\right]$. Thus $\mu_X(\emptyset) = 1$. The coefficients $\mu_X(S)$ for nonempty S are the nonconstant Fourier coefficients of the density of P_X relative to uniform.

Lemma 8.1 (Parity bias and Fourier L_1 norm). *Let $D : \{-1,1\}^q \to [-1,1]$. If $|\mu_X(S)| \le \varepsilon$ for every nonempty $S \subseteq [q]$, then*

$$|\mathbb{E}[D(Z_X)] - \mathbb{E}[D(U_q)]| \le \varepsilon \sum_{\emptyset \neq S \subseteq [q]} |\widehat{D}(S)|.$$

In particular,

$$\mathrm{TV}(P_X, U_q) \le \frac{1}{2}\sqrt{2^q - 1}\,\varepsilon.$$

Proof. By (11),

$$\mathbb{E}[D(Z_X)] - \mathbb{E}[D(U_q)] = \sum_{\emptyset \neq S} \widehat{D}(S)\mu_X(S),$$

which proves the first claim. For the second, use (12) and (13),

$$\chi^2(P_X \| U_q) = \sum_{\emptyset \neq S} \mu_X(S)^2 \le (2^q - 1)\varepsilon^2.$$

Apply (14). □

Whenever the hypotheses of Theorem 5.1 hold for q, put $\varepsilon_q := 4\eta^{\ell/2^{t_{2q}+1}}$. Since t_s is nondecreasing, Theorem 5.1 gives $|\mu_X(S)| \le \varepsilon_q$ for every nonempty $S \subseteq [q]$. Thus every fixed set of q distinct chosen queries satisfies

$$\mathrm{TV}(P_X, U_q) \le \frac{1}{2}\sqrt{2^q - 1}\,\varepsilon_q,$$

so the same quantity bounds the advantage of every non-adaptive distinguisher on that query set.

The same bound also controls distinguishers of low Fourier degree. Fix $0 < \theta < 1$ and $0 < \rho < 1/4$, let $q \le \kappa^{\log^\theta \kappa}$, and suppose $D : \{-1,1\}^q \to [-1,1]$ has Fourier degree at most d. By Cauchy–Schwarz, Parseval, and Lemma 8.1 and Corollary 5.7,

$$|\mathbb{E}[D(Z_X)] - \mathbb{E}[D(U_q)]| \le 2^{-\kappa^\rho} \sqrt{\sum_{j=1}^{d} \binom{q}{j}}.$$

Thus $d \log(eq/d) = o(\kappa^\rho)$ gives advantage $2^{-\kappa^\rho + o(\kappa^\rho)}$.

For constant-depth circuits we use a stronger bounded-independence argument below.

8.2 Adaptive adversaries

An adaptive adversary chooses its next query from the answers it has already received. After fixing its internal randomness, represent these choices by a binary decision tree of depth at most q, deleting repeated queries. Each leaf indicator expands into parities on subsets of the queries on that path. The parity bound therefore applies to every nonconstant term in this expansion.

Lemma 8.2 (Fourier L_1 norm of a query tree). *Let D be the $\{0,1\}$-valued acceptance function of a deterministic binary decision tree of depth at most q. Then*

$$\sum_{\emptyset \neq S} |\widehat{D}(S)| \le 2^{q-1}.$$

The input variables of D are oracle truth-table bits, one for each domain point queried anywhere in the tree.

Proof. Each leaf is represented by the indicator of its root-to-leaf path, a product of at most q literals. Its Fourier L_1 norm is exactly 1. Let a and r be the numbers of accepting and rejecting leaves. Writing D as the sum of accepting-leaf indicators gives $\sum_{S\neq\emptyset} |\widehat{D}(S)| \le a$; applying the same argument to $1 - D$ gives the bound by r. Hence

$$\sum_{S\neq\emptyset} |\widehat{D}(S)| \le \min\{a, r\} \le \frac{a+r}{2} \le 2^{q-1}.$$

□

Proposition 8.3 (From parity bias to adaptive distinguishers; cf. [BR17])**.** *Suppose that for every set of at most q distinct domain points and every nonempty subset of those points, the corresponding parity bias is at most ε_q. Then every randomized adaptive q-query adversary has advantage at most $2^{q-1}\varepsilon_q$.*

Proof. Fix the adversary's internal randomness. The resulting deterministic algorithm is a depth-q query tree. Every nonconstant Fourier monomial of a leaf indicator is supported on a subset of the fixed domain points queried along that path, so its expectation changes by at most ε_q. Summing over the nonconstant Fourier coefficients and applying Lemma 8.2 gives the claim. Average over the adversary's coins. □

Theorem 8.4 (Unrestricted adaptive statistical security)**.** *For the balanced BDEKRS parameters, there is an absolute $\delta > 0$ such that every randomized adaptive q-query adversary has advantage at most $\exp(-\Omega(\lambda))$ whenever $q \le \delta\lambda$. Since $\kappa = \Theta(\lambda^4)$, this gives information-theoretic security for $\Theta(\kappa^{1/4})$ unrestricted adaptive queries.*

Proof of Theorem 8.4. For $q \le \delta\lambda$, we have $t_{2q} \le 2$ for all sufficiently large λ, since

$$2q\,\beta(2p)^2 = O\left(\delta\frac{\log^2\lambda}{\lambda}\right) = o(1).$$

Hence $\varepsilon_q \le 4\eta^{\lambda/8}$. By Proposition 8.3,

$$\mathrm{Adv} \le 2^{q-1}\varepsilon_q \le 2\exp\left(q\ln 2 + \frac{\lambda}{8}\ln\eta\right).$$

Choosing $\delta > 0$ sufficiently small makes the exponent $-\Omega(\lambda)$. Since $\kappa = \Theta(\lambda^4)$, this is the $\Theta(\kappa^{1/4})$ query scale. □

8.3 AC^0 distinguishers on the response vector

For quasipolynomially many responses, the factor $2^{q/2}$ in the general statistical-distance bound is too large. To handle constant-depth circuits, we first use small bias to make the marginals close to uniform. We then find a nearby exactly k-wise independent distribution and apply the bounded-independence theorem for AC^0.

Theorem 8.5 (AC^0 distinguishers on the response vector)**.** *Fix $\theta \in (0,1)$, an integer $d \ge 1$, and $0 < a_{\mathrm{circ}} < \frac{1}{12d+12}$. For the balanced BDEKRS parameters, let $Q = \kappa^{\log^\theta \kappa}$. Uniformly for all fixed sets of $q \le Q$ distinct non-adaptive queries and every depth-d AC^0 circuit D of size at most $2^{\kappa^{a_{\mathrm{circ}}}}$ applied to the q response bits,*

$$|\mathbb{E}[D(Z_X)] - \mathbb{E}[D(U_q)]| \le 2^{-\Omega(\kappa^{a_{\mathrm{circ}}})}.$$

The threshold on a_{circ} is $a_{\mathrm{circ}}(3d+3) < 1/4$: the factor $3d+3$ comes from the bounded-independence theorem below, and $1/4$ comes from the bias bound in terms of the key length.

The two results we need concern distance to exact independence and the amount of independence needed to fool a circuit. We say that a distribution P on q bits is (δ, k)-wise independent if every marginal on at most k coordinates is within total variation δ of uniform.

Lemma 8.6 (Almost k-wise to exact k-wise; [AGM03])**.** *There is an absolute constant $C > 0$ such that every (δ, k)-wise independent distribution P on q bits is within total variation $q^{Ck}\delta$ of some exactly k-wise independent distribution.*

Theorem 8.7 (Bounded independence fools AC^0; [Tal17])**.** *For every fixed depth d there is a constant $C_d > 0$ such that the following holds. Every*

$$k \geq C_d\big(\log(s/\varepsilon)\big)^{3d+3}$$

-wise independent distribution ε-fools every depth-d, size-s AC^0 circuit.

Lemma 8.8 (Small bias gives accurate marginals)**.** *If every nonempty parity of a distribution P on q bits has bias at most ξ, then P is $(2^{k/2-1}\xi, k)$-wise independent for every $1 \leq k \leq q$.*

Proof. Fix a set I of at most k coordinates. Every nonconstant Fourier coefficient of the marginal P_I has magnitude at most ξ. Therefore

$$\chi^2(P_I \| U_I) \leq (2^{|I|} - 1)\xi^2,$$

and (14) gives the result. □

Proof of Theorem 8.5. Choose ρ so that $a_{\mathrm{circ}}(3d+3) < \rho < \frac{1}{4}$. By Corollary 5.7, uniformly over the entire query range in the theorem, every nonempty response parity has bias at most $\xi := 2^{-\kappa^\rho}$ for all sufficiently large κ.

Let $s \leq 2^{\kappa^{a_{\mathrm{circ}}}}$ be the circuit size and set $\varepsilon_0 := 2^{-2\kappa^{a_{\mathrm{circ}}}}$. Choose

$$k_0 := \left\lceil C_d\big(\log(s/\varepsilon_0)\big)^{3d+3} \right\rceil = O\Big(\kappa^{a_{\mathrm{circ}}(3d+3)}\Big),$$

where C_d is from Theorem 8.7. We separate the case in which the number of response bits is smaller than the bounded-independence parameter.

If $q < k_0$, the direct Fourier bound following (14) gives

$$\mathrm{TV}(P_X, U_q) \leq \frac{1}{2}\sqrt{2^q - 1}\,\xi \leq 2^{q/2-1-\kappa^\rho}.$$

Since $q < k_0 = O(\kappa^{a_{\mathrm{circ}}(3d+3)}) = o(\kappa^\rho)$, this is $2^{-\Omega(\kappa^\rho)}$. The claimed bound follows for every distinguisher, without using bounded independence.

Suppose now that $q \geq k_0$. By Lemma 8.8, the response distribution is (δ_{k_0}, k_0)-wise independent with $\delta_{k_0} \leq 2^{k_0/2}\xi$. Apply Lemma 8.6. There is an exactly k_0-wise independent $P_X^\star$ satisfying

$$\mathrm{TV}(\mathsf{Law}(Z_X), P_X^\star) \leq q^{O(k_0)}2^{k_0/2}\xi.$$

Since $\log q = O((\log \kappa)^{1+\theta})$ and $a_{\mathrm{circ}}(3d+3) < \rho$,

$$O(k_0 \log q) + \frac{k_0}{2} - \kappa^\rho = -\Omega(\kappa^\rho).$$

Thus

$$\mathrm{TV}(\mathsf{Law}(Z_X), P_X^\star) = 2^{-\Omega(\kappa^\rho)}.$$

By the definition of k_0, the distribution $P_X^\star$ fools the circuit to error ε_0 by Theorem 8.7. The triangle inequality gives

$$|\mathbb{E}[D(Z_X)] - \mathbb{E}[D(U_q)]| \le 2^{-2\kappa^{a_{\mathrm{circ}}}} + 2^{-\Omega(\kappa^\rho)} = 2^{-\Omega(\kappa^{a_{\mathrm{circ}}})}.$$

□

Remark 8.9 (DNF/CNF distinguishers). Bazzi's depth-two theorem gives a slightly better numerical range for DNF and CNF formulas. Under the same quasipolynomial query bound as in Theorem 8.5, formulas with at most $2^{\kappa^{a_{\mathrm{circ}}}}$ terms or clauses have distinguishing advantage $2^{-\Omega(\kappa^{a_{\mathrm{circ}}})}$ for every fixed $0 < a_{\mathrm{circ}} < 1/8$. Indeed, choose ρ with $2a_{\mathrm{circ}} < \rho < 1/4$ and apply the same approximation argument using Bazzi's $O(\log^2(s/\varepsilon))$ bounded-independence bound [Baz07].

9 Query sets with few affine relations

The generic Fourier bound treats all 2^q Fourier characters alike. We can do better when each transcript has only a few independent affine relations. For example, suppose an adversary starts with affinely independent queries and then adds only a few queries from their affine span. Each such query adds at most one independent affine relation.

We first choose the matrix uniformly from all binary matrices. For a fixed subset S of response coordinates, any response whose query occurs in no affine relation among the queries indexed by S is independent of the other responses in S. These coordinates contribute separate one-coordinate expectations to the Fourier coefficient. We group coefficients by the space of relations supported on their index sets. Once that space is fixed, the remaining indices contribute only powers of the small one-query expectation. Summing these powers replaces the loss over all query subsets by a loss governed by the relation dimension.

Theorem 9.1 (Adaptive statistical security with few affine relations). *Fix $\gamma \in (0,1)$. There is an absolute constant $c_* > 0$ such that, for the balanced BDEKRS parameters and all sufficiently large λ, the following holds. Let $\mathcal{A}$ be an arbitrary computationally unbounded adaptive adversary making at most $q \le (1-\gamma)N = (1-\gamma)\lambda^2$ distinct queries. Suppose that on every root-to-leaf path of its query tree the affine relation space of the queried points has dimension at most $\nu \le c_*\lambda/\log\lambda$. Then*

$$\mathrm{Adv}_{\mathcal{F}_{\mathrm{BDEKRS}}}(\mathcal{A}) \le \exp(-\Omega(\lambda)).$$

9.1 Relation spaces and their supports

Fix distinct $X = (x_1, \dots, x_q)$ and put $v_i := (x_i, 1) \in \mathbb{F}_2^{N+1}$. Define the affine relation space

$$\mathcal{K}_X := \left\{ c \in \mathbb{F}_2^q : \sum_{i=1}^q c_i v_i = 0 \right\}.$$

Recall that

$$d_{\mathrm{aff}}(X) = \dim \operatorname{span}\{v_1, \dots, v_q\} - 1,$$
$$\nu(X) = \dim \mathcal{K}_X = q - \dim \operatorname{span}\{v_1, \dots, v_q\}.$$

To factor the Fourier coefficient indexed by S, we need to know which queries participate in relations within S. For $S \subseteq [q]$, let $W_X(S) := \{c \in \mathcal{K}_X : \operatorname{supp}(c) \subseteq S\}$, the space of affine relations whose support is contained in S. For any subspace $W \subseteq \mathcal{K}_X$, define its support by $\operatorname{supp}(W) := \bigcup_{c \in W} \operatorname{supp}(c)$, and abbreviate $C_X(S) := \operatorname{supp}(W_X(S))$. Thus $C_X(S) = \emptyset$ when $W_X(S) = 0$.

For intuition, take five queries $x_1 = 0, x_2 = e_1, x_3 = e_2, x_4 = e_1 + e_2$, and $x_5 = e_3$, where e_1, e_2, e_3 are distinct standard basis vectors. For $S = \{1,2,3,4,5\}$, the only nonzero affine relation is $v_1 + v_2 + v_3 + v_4 =$

0, while $v_5 \notin \operatorname{span}\{v_1, v_2, v_3, v_4\}$. Thus $W_X(S) = \operatorname{span}\{(1,1,1,1,0)\}$ and $C_X(S) = \{1,2,3,4\}$. The fifth coordinate occurs in no relation supported on S, so its contribution to the corresponding Fourier coefficient factors as a separate one-coordinate expectation.

Let $\widetilde{A} \leftarrow \mathbb{F}_2^{N\times N}$ be uniform and $b \leftarrow \mathbb{F}_2^N$. Write $\widetilde{P}_X$ for the distribution of the signed response vector produced by the same evaluation formula as the candidate with $\widetilde{A}$ in place of A, and write

$$\widetilde{\mu}_X(S) := \mathop{\mathbb{E}}_{\widetilde{A},b}\left[\prod_{i\in S}(-1)^{F_{\widetilde{A},b}(x_i)}\right].$$

Finally put $\zeta := \mathbb{E}_{Y\leftarrow\mathbb{F}_2^N}\left[g(Y)\right] = (2\beta-1)^\ell$. For the balanced parameters, $|2\beta-1| = O(p)$ and hence

$$|\zeta| = 2^{-\Omega(\lambda\log\lambda)}. \tag{39}$$

Lemma 9.2 (Factorization outside the relation support)**.** *For every $S \subseteq [q]$, put $W = W_X(S)$ and $C = C_X(S)$. There is a quantity b_W, depending only on the queries indexed by C, such that*

$$\widetilde{\mu}_X(S) = b_W\zeta^{|S\setminus C|}.$$

We have $b_0 = 1$. If $W \neq 0$, then

$$|b_W| \le \varepsilon_q^{\mathrm{unif}} := \eta^{\ell/2^{t_{2q}+1}}$$

whenever the hypotheses of Theorem 5.1 hold for q.

The two error bounds satisfy $\varepsilon_q = 4\varepsilon_q^{\mathrm{unif}}$; the factor 4 is the cost of conditioning on invertibility.

Proof. The affine relation space among the augmented vectors indexed by S is $W_X(S)$. An index $i \in S\setminus C$ occurs in no relation supported in S. Equivalently, $v_i \notin \operatorname{span}\{v_j : j \in S\setminus\{i\}\}$. We may therefore choose a basis of $\operatorname{span}\{v_j : j \in S\}$ that contains every vector indexed by $S\setminus C$. The evaluations of a uniform linear functional on these basis vectors are independent unbiased bits and are independent of its evaluations on the remaining basis vectors. Applying this independently to each of the N rows of the random affine map shows that the vectors corresponding to indices in $S\setminus C$ are independent uniform elements of $\mathbb{F}_2^N$ and are independent of those corresponding to indices in C. Each therefore contributes the factor ζ.

The remaining correlation is b_W. If $W \neq 0$, then C is a nonempty set of distinct queries and $|C| \le q$. Monotonicity of t_s and Corollary 5.4 give $|b_W| \le \varepsilon_q^{\mathrm{unif}}$. □

9.2 Counting subspaces of relations

Although S ranges over 2^q subsets, the subspaces $W_X(S)$ take far fewer distinct values when the affine relation space has small dimension.

Lemma 9.3 (Number of distinct subspaces $W_X(S)$)**.** *Let $\nu = \dim \mathcal{K}_X$. The number $\mathcal{N}_X$ of distinct subspaces $W_X(S)$, as S ranges over all subsets of $[q]$, satisfies*

$$\mathcal{N}_X \le \sum_{j=0}^{\nu}\binom{q}{j}.$$

In particular, if $1 \le \nu \le q/2$, then

$$\mathcal{N}_X \le \left(\frac{eq}{\nu}\right)^\nu.$$

For $\nu = 0$, $\mathcal{N}_X = 1$.

Proof. Choose a full-row-rank generator matrix $H \in \mathbb{F}_2^{\nu \times q}$ for $\mathcal{K}_X$ and let its columns be $h_1, \dots, h_q \in \mathbb{F}_2^{\nu}$. The map $u \mapsto u^\top H$ identifies $\mathbb{F}_2^{\nu}$ with $\mathcal{K}_X$. Under this identification, $W_X(S)$ corresponds to

$$\{u : u \cdot h_i = 0 \text{ for every } i \notin S\} = \operatorname{span}\{h_i : i \notin S\}^{\perp}.$$

Thus $W_X(S)$ is determined by a subspace spanned by a subset of the q columns of H. Every such span has a basis consisting of at most ν of those columns, so the number of possible spans is at most $\sum_{j=0}^{\nu} \binom{q}{j}$. The standard binomial bound gives the second claim. □

Theorem 9.4 (Fourier L_1 norm with few affine relations)**.** *Under the hypotheses of Theorem 5.1, let X be a fixed set of q distinct queries whose affine relation space has dimension ν. When $\widetilde{A}$ is uniform over all $N \times N$ matrices,*

$$1 + \sum_{\emptyset \neq S \subseteq [q]} |\widetilde{\mu}_X(S)| \leq (1 + |\zeta|)^q \left(1 + (\mathcal{N}_X - 1)\varepsilon_q^{\mathrm{unif}}\right),$$

where $\mathcal{N}_X$ is bounded as in Lemma 9.3. In particular, for $1 \leq \nu \leq q/2$,

$$1 + \sum_{S \neq \emptyset} |\widetilde{\mu}_X(S)| \leq (1 + |\zeta|)^q \left[1 + \left(\frac{eq}{\nu}\right)^{\nu} \varepsilon_q^{\mathrm{unif}}\right].$$

Proof. Partition the subsets $S \subseteq [q]$ according to $W = W_X(S)$. By Lemma 9.2, for fixed W, put $C = \operatorname{supp}(W)$. Then

$$\sum_{S : W_X(S) = W} |\widetilde{\mu}_X(S)| \leq |b_W| \sum_{T \subseteq [q] \setminus C} |\zeta|^{|T|} \leq |b_W|(1 + |\zeta|)^q.$$

For $W = 0$ we have $b_W = 1$, and for every nonzero W we have $|b_W| \leq \varepsilon_q^{\mathrm{unif}}$. Summing over the $\mathcal{N}_X$ possible subspaces $W_X(S)$ proves the claim. □

Bounding the sum of absolute Fourier coefficients controls the probability of each individual response string relative to its uniform probability. This relative bound will let us sum over adaptive transcripts without paying for the number of possible transcripts.

Corollary 9.5 (Pointwise closeness to uniform with few affine relations)**.** *There exists an absolute $c_* > 0$ such that the following holds for the balanced parameters. If $q \leq N = \lambda^2$ and $\nu(X) \leq c_* \lambda / \log \lambda$, then, when $\widetilde{A}$ is uniform over all $N \times N$ matrices,*

$$\sum_{S \neq \emptyset} |\widetilde{\mu}_X(S)| \leq \exp(-\Omega(\lambda)).$$

Consequently, for every $z \in \{-1, 1\}^q$,

$$\widetilde{P}_X(z) = 2^{-q} \left(1 \pm \exp(-\Omega(\lambda))\right).$$

Proof. For $q \leq \lambda^2$, (39) gives $q|\zeta| = 2^{-\Omega(\lambda \log \lambda)}$. Moreover, Corollary 5.4 and Lemma 5.5 with $q \leq \lambda^2$ give an absolute constant $c_{\mathrm{corr}} > 0$ such that $\varepsilon_q^{\mathrm{unif}} \leq e^{-c_{\mathrm{corr}} \lambda}$ for all sufficiently large λ. If $\nu \leq q/2$, then $\nu \log(eq/\nu) \leq O(c_* \lambda)$ under the stated bound on the relation dimension. Choosing c_* sufficiently small compared with c_{corr} makes the contribution from nonzero relation subspaces in Theorem 9.4 exponentially small. If instead $\nu > q/2$, then $q < 2\nu = O(c_* \lambda / \log \lambda)$, and the crude bound $\mathcal{N}_X \leq 2^q$ gives the same conclusion. In both cases the factor $(1 + |\zeta|)^q$ is $1 + 2^{-\Omega(\lambda \log \lambda)}$.

For the pointwise statement, Fourier inversion gives

$$2^q \widetilde{P}_X(z) - 1 = \sum_{S \neq \emptyset} \widetilde{\mu}_X(S) \chi_S(z),$$

whose absolute value is bounded by the preceding sum of absolute Fourier coefficients. □

9.3 Adaptive queries and the invertible matrix

For a query tree satisfying the relation bound, each leaf has its random-function probability times $1 \pm e^{-\Omega(\lambda)}$. These random-function probabilities sum to 1, so the relative errors do not accumulate with the number of leaves. To return to an invertible secret matrix, we couple the first query image and then reveal images of new directions as the query span grows. The following coupling makes the two maps agree until a new independent direction is mapped into the span of previous images.

Lemma 9.6 (Adaptive affine-map coupling)**.** *Let a deterministic adaptive algorithm make distinct queries $x_1, x_2, \dots$ in $\mathbb{F}_2^N$. Couple two experiments: in the first, $\widetilde{A} \leftarrow \mathbb{F}_2^{N\times N}$ and $b \leftarrow \mathbb{F}_2^N$ are uniform; in the second, $A \leftarrow \mathrm{GL}_N(\mathbb{F}_2)$ and $b' \leftarrow \mathbb{F}_2^N$ are uniform. If the affine dimension of every query prefix is at most r, the query images $\widetilde{A}x_i + b$ and $Ax_i + b'$ in the two experiments can be coupled to agree throughout the execution except with probability less than 2^{r-N}. The same bound holds after applying any deterministic function to these images, in particular the candidate output predicate.*

Proof. Couple the image of the first query to be the same uniform point in both experiments. Relative to that point, expose images of direction vectors as the affine span grows. In the first experiment, when a new direction is linearly independent of the previous directions, its image is uniform in $\mathbb{F}_2^N$. In the second experiment, its image is uniform outside the span of the images of the previously exposed directions. If j independent directions have already been exposed, these two choices can be coupled except when the image in the first experiment falls in their j-dimensional span, an event of probability 2^{j-N}. A query in the existing affine span introduces no new randomness: its affine image is determined by the previous images in both experiments. Summing the failure probabilities over $j = 0, \dots, r-1$ gives

$$\sum_{j=0}^{r-1} 2^{j-N} < 2^{r-N}.$$

As long as the coupling has not failed, the observed oracle answers agree, so the adaptive algorithm asks the same next query in both experiments. □

Proof of Theorem 9.1. Fix the adversary's internal randomness. Each leaf τ of the resulting query tree determines a fixed sequence X_τ of distinct queries and a fixed response string z_τ. By hypothesis, $|X_\tau| \le q \le N$ and $\nu(X_\tau) \le c_*\lambda/\log\lambda$. For the experiment with $\widetilde{A}$ uniform over all matrices, Corollary 9.5 gives, uniformly over leaves,

$$\mathbb{P}[\tau] = 2^{-|X_\tau|}\left(1 \pm e^{-\Omega(\lambda)}\right).$$

Under a random function, the same leaf has probability exactly $2^{-|X_\tau|}$. Summing over accepting leaves, whose random-function probabilities sum to at most 1, shows that this experiment and the random-function experiment differ in acceptance probability by $e^{-\Omega(\lambda)}$.

Every path has affine dimension at most $q - 1 \le (1-\gamma)N - 1$. By Lemma 9.6, replacing $\widetilde{A}$ by a uniform invertible matrix changes the entire adaptive experiment by at most $2^{-\Omega_\gamma(N)}$. Averaging over the adversary's internal randomness proves

$$\mathrm{Adv}_{\mathcal{F}_{\mathrm{BDEKRS}}}(\mathcal{A}) \le e^{-\Omega(\lambda)} + 2^{-\Omega_\gamma(N)} = e^{-\Omega(\lambda)}.$$

□

10 Limitations and information-theoretic barriers

Algebraic degree gives a linear attack with sufficiently many queries. On some low-dimensional query sets, the responses take too few possible values to be statistically close to uniform.

10.1 A linear attack from algebraic degree

Proposition 10.1 (Affine-cube attack). *For every fixed key (A, b), the Boolean function $F_{A,b} : \mathbb{F}_2^N \to \mathbb{F}_2$ has algebraic degree at most $M = mn$. Consequently, whenever $N \geq M + 1$, there are 2^{M+1} distinct non-adaptive queries whose output parity is identically zero for every key. The corresponding linear-test bias is* 1.

Proof. Each tribe AND has degree n. Over $\mathbb{F}_2$,

$$\bigvee_{j=1}^{m} a_j = 1 + \prod_{j=1}^{m} (1 + a_j),$$

so $\mathrm{TRIBES}_{m,n}$ has algebraic degree at most $mn = M$. Composition with an affine map does not increase degree, and the outer XOR does not increase it either. Hence $\deg_{\mathbb{F}_2}(F_{A,b}) \leq M$.

Choose linearly independent directions $v_1, \ldots, v_{M+1} \in \mathbb{F}_2^N$ and a base point x_0. On the affine cube

$$\mathcal{Q} = \left\{ x_0 + \sum_{r=1}^{M+1} u_r v_r : u \in \mathbb{F}_2^{M+1} \right\},$$

the $(M + 1)$st discrete derivative of every degree-M polynomial vanishes. Equivalently,

$$\bigoplus_{x \in \mathcal{Q}} F_{A,b}(x) = 0.$$

□

For the balanced parameters, $M = \lambda$ and $N = \lambda^2$, so the attack uses $2^{\lambda+1}$ queries. Corollary 5.8 proves negligible bias through $2^{a(\log_2 \lambda)^2}$ queries for every fixed $0 < a < 1$. Determining the bias in the intermediate range remains open. The number of Cauchy–Schwarz rounds grows as $t = \Theta(\log q / \log \lambda)$; the resulting 2^tth root governs the query range of our bound.

10.2 A support-size barrier for low-dimensional query sets

Proposition 10.2 (Support and affine dimension). *Fix distinct queries $X = (x_1, \ldots, x_q)$ with $d_{\mathrm{aff}}(X) = d$. Let P_X be the distribution of their signed response vector under a uniform key. Then*

$$|\mathrm{supp}(P_X)| \leq 2^{N(d+1)}$$

and hence

$$\mathrm{TV}(P_X, U_q) \geq 1 - 2^{N(d+1)-q}.$$

Proof. Choose a base point x_0 in the affine span of X and direction vectors $v_1, \ldots, v_d$ spanning its direction space. The entire tuple $(Ax_i + b)_{i \leq q}$ is determined by $Ax_0 + b$ and $Av_1, \ldots, Av_d$. There are at most $2^{N(d+1)}$ possible values of these $d+1$ vectors, so the response vector, being determined by these vectors, has at most this many possible values. A distribution supported on at most K points of $\{-1, 1\}^q$ has total variation distance at least $1 - K/2^q$ from uniform. □

The lower bound is positive once $q > N(d + 1)$. A set of q distinct points can have affine dimension as small as $\lceil \log_2 q \rceil$. Thus there are query sets of size $\Theta(N \log N)$ for which the support bound is already nontrivial. By contrast, the trivial support bound from the 2^κ possible keys becomes informative only around $q = \kappa = N^2 + N$.

References

[ABG+14] Adi Akavia, Andrej Bogdanov, Siyao Guo, Akshay Kamath, and Alon Rosen. Candidate weak pseudorandom functions in $\mathsf{AC}^0 \circ \mathrm{MOD}_2$. In Moni Naor, editor, *ITCS 2014*, pages 251–260. ACM, January 2014.

[AGM03] Noga Alon, Oded Goldreich, and Yishay Mansour. Almost k-wise independence versus k-wise independence. *Information Processing Letters*, 88(3):107–110, 2003.

[AM13] Per Austrin and Elchanan Mossel. Noise correlation bounds for uniform low degree functions. *Arkiv för Matematik*, 51(1):29 – 52, 2013.

[Baz07] Louay Bazzi. Polylogarithmic independence can fool DNF formulas. In *48th FOCS*, pages 63–73. IEEE Computer Society Press, October 2007.

[BCG+20] Elette Boyle, Geoffroy Couteau, Niv Gilboa, Yuval Ishai, Lisa Kohl, and Peter Scholl. Correlated pseudorandom functions from variable-density LPN. In *61st FOCS*, pages 1069–1080. IEEE Computer Society Press, November 2020.

[BCG+21] Elette Boyle, Geoffroy Couteau, Niv Gilboa, Yuval Ishai, Lisa Kohl, and Peter Scholl. Low-complexity weak pseudorandom functions in $\mathtt{AC0}[\mathtt{MOD2}]$. In Tal Malkin and Chris Peikert, editors, *CRYPTO 2021, Part IV*, volume 12828 of *LNCS*, pages 487–516, Virtual Event, August 2021. Springer, Cham.

[BCG+22] Elette Boyle, Geoffroy Couteau, Niv Gilboa, Yuval Ishai, Lisa Kohl, Nicolas Resch, and Peter Scholl. Correlated pseudorandomness from expand-accumulate codes. In Yevgeniy Dodis and Thomas Shrimpton, editors, *CRYPTO 2022, Part II*, volume 13508 of *LNCS*, pages 603–633. Springer, Cham, August 2022.

[BCPR25] Christina Boura, Geoffroy Couteau, Léo Perrin, and Yann Rotella. SoK: On shallow weak PRFs A common symmetric building block for MPC protocols. *IACR Trans. Symm. Cryptol.*, 2025(3):289–336, 2025.

[BDE+25] Marshall Ball, Clément Ducros, Saroja Erabelli, Lisa Kohl, Nicolas Resch, and Peter Scholl. Strong pseudorandom functions in $AC^0[2]$ in the bounded-query setting. Cryptology ePrint Archive, Paper 2025/2085, 2025.

[BGGP12] Itai Benjamini, Ori Gurel-Gurevich, and Ron Peled. On k-wise independent distributions and boolean functions, 2012.

[BL85] Michael Ben-Or and Nathan Linial. Collective coin flipping, robust voting schemes and minima of banzhaf values. In *26th FOCS*, pages 408–416. IEEE Computer Society Press, October 1985.

[BP05] An Braeken and Bart Preneel. Probabilistic algebraic attacks. In Nigel P. Smart, editor, *10th IMA International Conference on Cryptography and Coding*, volume 3796 of *LNCS*, pages 290–303. Springer, Berlin, Heidelberg, December 2005.

[BR17] Andrej Bogdanov and Alon Rosen. Pseudorandom functions: Three decades later. Cryptology ePrint Archive, Report 2017/652, 2017.

[Car21] Claude Carlet. *Boolean Functions for Cryptography and Coding Theory*. Cambridge University Press, 2021.

[CD23] Geoffroy Couteau and Clément Ducros. Pseudorandom correlation functions from variable-density LPN, revisited. In Alexandra Boldyreva and Vladimir Kolesnikov, editors, *PKC 2023, Part II*, volume 13941 of *LNCS*, pages 221–250. Springer, Cham, May 2023.

[CGSS22] Pablo Candela, Diego González-Sánchez, and Balázs Szegedy. A refinement of cauchy–schwarz complexity. *European Journal of Combinatorics*, 106:103592, 2022.

[GGM86] Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct random functions. *Journal of the ACM*, 33(4):792–807, October 1986.

[Gow01] W. T. Gowers. A new proof of szemerédi's theorem. *GAFA, Geom. funct. anal.*, 11(3):465–588, 2001.

[GT10] Benjamin Green and Terence Tao. Linear equations in primes. *Ann. Math. (2)*, 171(3):1753–1850, 2010.

[GW10] W. T. Gowers and J. Wolf. The true complexity of a system of linear equations. *Proceedings of the London Mathematical Society*, 100(1):155–176, 2010.

[GY20] Parikshit Gopalan and Amir Yehudayoff. Concentration for limited independence via inequalities for the elementary symmetric polynomials. *Theory of Computing*, 16(17):1–29, 2020.

[HHL16] Hamed Hatami, Pooya Hatami, and Shachar Lovett. General systems of linear forms: Equidistribution and true complexity. *Advances in Mathematics*, 292:446–477, 2016.

[KL01] Matthias Krause and Stefan Lucks. Pseudorandom functions in TC^0 and cryptographic limitations to proving lower bounds. *Comput. complex.*, 10(4):297–313, 2001.

[LL10] Meicheng Liu and Dongdai Lin. On the immunity of Boolean functions against probabilistic algebraic attacks. In Carlos Cid and Jean-Charles Faugère, editors, *Proceedings of the Second International Conference on Symbolic Computation and Cryptography*, pages 203–207, 2010. Extended abstract.

[Man21] Freddie Manners. True complexity and iterated cauchy–schwarz, 2021.

[NN90] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. In *22nd ACM STOC*, pages 213–223. ACM Press, May 1990.

[Raz87] Alexander A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical Notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987.

[Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In Alfred Aho, editor, *19th ACM STOC*, pages 77–82. ACM Press, May 1987.

[Tal17] Avishay Tal. Tight Bounds on the Fourier Spectrum of AC0. In Ryan O'Donnell, editor, *32nd Computational Complexity Conference (CCC 2017)*, volume 79 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 15:1–15:31, Dagstuhl, Germany, 2017. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.